

Kaiser, Caspar

Preprint

Using memories to assess the intrapersonal comparability of wellbeing reports

Suggested Citation: Kaiser, Caspar (2020) : Using memories to assess the intrapersonal comparability of wellbeing reports, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/226218>

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.

Using memories to assess the intrapersonal comparability of wellbeing reports[†]

Caspar Kaiser*

November 2020

Abstract

Research on subjective wellbeing typically assumes that responses to survey questions are comparable across respondents and across time. However, if this assumption is violated, standard methods in empirical research may mislead. I address this concern with three contributions. First, I give a theoretical analysis of the extent and direction of bias that may result from violations of this assumption. Second, I propose to use respondents' stated memories of their past wellbeing to estimate and thereby to correct for differentials in scale use. Third, using the proposed approach, I test whether wellbeing reports are intrapersonally comparable across time. Using BHPS data, I find that the direction in which explanatory variables affect latent satisfaction is typically the same as the direction in which scale use is affected. Unemployment and bereavement appear to have particularly strong effects on scale use. Although discussed in the context of life satisfaction scales, the proposed approach for anchoring response scales is applicable to a wide range of other subjectively reported constructs.

Keywords: *subjective wellbeing; life satisfaction; BHPS; vignettes; memories; differential item functioning*

[†] I thank Mark Fabian, Joel McGuire, Brian Nolan, Andrew Oswald, Michael Plant, Alberto Prati, Nhat An Trinh, Maarten Vendrik, Jakob Zeitler, and participants of the Oxford Wellbeing Research Centre seminar series, as well as the 2020 GLO/EHERO special session on Happiness Economics for helpful comments and suggestions. Funding from the Oxford Wellbeing Research Centre is gratefully acknowledged.

*Nuffield College, Department of Social Policy & Intervention, Wellbeing Research Centre, University of Oxford.
Email: caspar.kaiser@nuffield.ox.ac.uk

1 Introduction

Most empirical research on subjective wellbeing – including happiness and life satisfaction – is concerned with establishing how explanatory variables are associated with the quality of ‘latent’ mental states. In order to establish such associations, respondents are asked to rate their mental states by means of a numerical scale. Analyses based on this type of data typically rely on the assumption that ratings of different respondents at different times are comparable. Call this the assumption of *common scale use*. Previous work has questioned this assumption (Adler 2013; Fleurbaey and Blanchet 2013; Viscusi 2020). Nevertheless, despite many studies analysing subjective wellbeing and other subjective constructs,¹ few papers constructively engage with this issue. Are wellbeing statements interpersonally and intertemporally comparable? If such statements are not comparable, standard methods in wellbeing research may mislead. In this paper, I attempt to answer this question.

The paper has five sections. In section 2, I introduce the standard set of assumptions adopted in empirical research on happiness and life satisfaction. I then focus on the assumption of *common scale use* and discuss several arguments as to why this assumption may be violated. In the next step, I formally explore the biases that result when *common scale use* is violated. In section 3, I first note that the use of vignettes (i.e. descriptions of fictional persons) is, to my knowledge, the only existing method to correct for differences in scale use. As an alternative, I then propose a new method to correct for differences in scale use. That method relies on variation in people’s stated memories of past wellbeing. I discuss the plausibility of the method’s identifying assumptions and derive the direction of bias in cases where my assumptions are violated. In section 4, I provide an empirical application and investigate whether satisfaction reports are intrapersonally comparable across time. British Household Panel Study (BHPS) data are used for this purpose. Section 5 concludes.

My empirical findings are as follows. Variables that positively affect² latent satisfaction are found to also be associated with more ‘stringent’ scale use,³ and *vice versa* for negative effects. Scale use is especially affected by unemployment and bereavement. Fortunately, compared to an analysis that maintains *common scale use*, differences in results are never so large as to change our conclusions regarding the direction and statistical significance of effects. Although this is generally good news for research on happiness and life satisfaction, ratios of coefficients – which may indicate the relative importance of explanatory variables – are affected by differences in scale use.

Moreover, lags and leads of explanatory variables have only limited effects on both latent satisfaction and scale use. The same holds for a set of social reference variables. Non-parallel shifts in scale use may be possible for some variables. However, such non-parallel scale shift effects turn out to be empirically indistinguishable from heterogeneous effects of covariates across the distribution of satisfaction.

¹ While the ensuing discussion is stated in the context of subjective wellbeing, and life satisfaction in particular, it should be stressed that my theoretical points are in principle applicable to any type of data relating to subjective states of respondents.

² In my empirical estimations, I will often use phrases like “the effect of”, “ x affects y ”, etc. I am aware that these terms suggest causal relations which I cannot guarantee to hold. My estimates are only suggestive of such causal relations. However, all of my substantive points continue to hold if all estimated coefficients are interpreted as merely indicating descriptive differences in conditional means.

³ I.e. making it harder to report a higher response category at a given level of latent satisfaction. See the next section for more details.

2 The problem

2.1 Standard assumptions in wellbeing research

To analyse the determinants of life satisfaction, researchers typically use responses to questions of the form:

How dissatisfied or satisfied are you with your life overall?

When answering this question, each respondent i at time t chooses a response $r_{it} = k$ from a set of K options. In the British Household Panel Survey ('BHPS'), where this particular question is asked, there are seven response options with extremes labelled as '*Not at all satisfied*' (for $k = 1$) to '*Completely satisfied*' (for $k = 7$). We assume that respondents make their choice by adhering to a rule of the form:⁴

$$r_{it} = k \leftrightarrow \tau_{it,k-1} < s_{it} \leq \tau_{it,k}, \quad (1)$$

where s_{it} is the latent level of satisfaction that we are ultimately interested in. The τ s are 'cutoffs' that need to be crossed to start reporting the next category. When cutoffs are shifted to take higher (smaller) values, I say that scale use becomes more (less) stringent (see figure 1).⁵ The literature standardly assumes that scale use is common across all respondents and across all times, and that cut-offs are equidistant. The former assumption can be stated as:

$$\tau_{it,k} = \tau_k, \quad (2)$$

which I call the assumption of *common scale use*. This assumption of *common scale use* will be the main subject of this paper. The assumption that cutoffs are equidistant can be written as:

$$\tau_k - \tau_{k-1} = c, \quad (3)$$

where c is some constant. We might call (3) the assumption of *equidistance*, or *linearity of scale use*. The plausibility and importance of this assumption is analysed in Schröder and Yitzhaki (2017), Bond and Lang (2019) and Kaiser and Vendrik (2020).

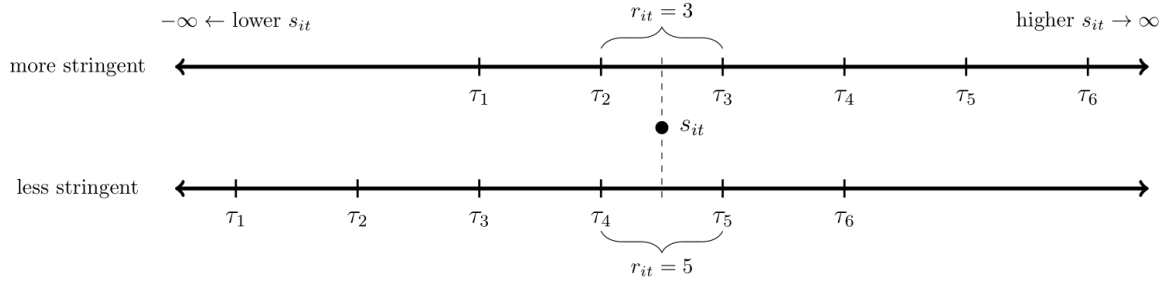
Jointly adopting assumptions (1)-(3) is taken to motivate the use of OLS regressions of r_{it} on a vector of variables \mathbf{X} to infer the associations between s_{it} and \mathbf{X} (Ferrer-i-Carbonell & Frijters, 2004).⁶ If we are unwilling to assume *linearity of scale use*, we may infer differences $\tau_k - \tau_{k-1}$ by postulating a structural model for s_{it} and a distribution for the stochastic error of such a model. When assuming a normally distributed error, this yields the ordered probit approach.

⁴ Equation (1) is sometimes called the 'reporting function' (Oswald 2008). The applicability of this equation itself may be questioned. We might doubt that s_{it} indeed determines the choice of r_{it} . Perhaps some other quantity, one that does not carry the sort of presumed normative content that motivates our analysis, drives the choice of r_{it} instead. Another possibility is that latent s_{it} does not really exist. Instead respondents only behave as though s_{it} exists when prompted (Bertrand and Mullainathan 2001). I leave these further doubts aside.

⁵ The idea of more or less stringent scale use in answers to wellbeing questions is analogous to differences in the stringency of academic grading systems: For a given level of student achievement, a less stringent grading system (e.g. that of the United States) assigns a comparatively high grade, whereas a more stringent grading system (e.g. that of France) assigns a comparatively low grade.

⁶ However, OLS is not necessarily sufficiently motivated even when assumptions (1)-(3) hold. Instead, in order for OLS to yield correct ratios of coefficients, we require the more demanding assumption that $E(s_{it}|r_{it} = k; \mathbf{X}) = E(s_{it}|r_{it} = k) = a + bk$, where a and b are constants. This assumption entails that independent variables \mathbf{X} do not affect the average level of latent satisfaction within each response category.

Figure 1. Examples of more and less stringent scale use



Note: For a given level of s_{it} , we observe a lower level of reported satisfaction r_{it} ($=3$) with more stringent scale use and a higher level of reported satisfaction ($=5$) with less stringent scale use.

Standard ordered probit regressions rely on *common scale use*.⁷ Unfortunately, violations of this assumption can lead to biases in the estimated effects of explanatory variables. These biases are formally shown in section 2.3. Before that, I will discuss some reasons for believing that response scales are not common over time and across respondents.

2.2 Are response scales common over time and across respondents?

An argument in favour of assuming *common scale use* is that respondents' own satisfaction ratings and the ratings of others about a respondent's satisfaction correlate (Diener, Inglehart, and Tay 2013). However, correlations between a respondent's own report and the ratings of others can be observed even if scale use differed across respondents and third-party raters. Such correlations only establish that third-party raters are reasonably accurate at inferring interpersonal variation in s_{it} .⁸ To defend *common scale use* one would have to show that there is no systematic variation across third-party raters in their ratings of subjects.

Another argument notes that for any evaluative label (such as 'very satisfied' or '8 out of 10') to elicit appropriate behavioural responses, such evaluative labels must have shared meanings (Van Praag 1993; Ferrer-i-Carbonell and Frijters 2004; Plant 2020).

However, Diamond (2008), Schwarz and Strack (1999), and Fleurbaey and Blanchet (2013) assert that respondents likely report their life satisfaction in terms of their relative position in some reference distribution. Fleurbaey and Blanchet (2013) motivate this by noting that the scale with which respondents report their satisfaction is bounded, while many aspects of life are unbounded (like income). Thus, satisfaction reports may rather indicate the quantile of a respondent's life in a reference distribution of possible lives. If people's judgements are relative to a reference distribution, and if these distributions vary, then response scales cannot be common.

Schwarz and Strack (1999) note that reference distributions likely depend on intrapersonal and interpersonal comparisons. Indeed, Steffel and Oppenheimer (2009) suggest that people calibrate their scale against their past. In that case, better past outcomes will make today's scale use more

⁷ Bond & Lang (2019) show that even if the assumption of *common scale use* is satisfied, violations of the normality assumption can lead to reversals of the signs of coefficient estimates. In Kaiser and Vendrik (2020) we show that such reversals imply assumptions about scale use that strongly depart from the assumption of *linearity*.

⁸ To observe a positive correlation between third-party raters' responses and respondents' own responses, two conditions are sufficient. First, locations of thresholds and latent satisfaction of respondents must not be correlated to the extent that r_{it} is expected to decrease when s_{it} increases. Second, third-party raters' perceptions of respondents' latent satisfaction must be positively correlated with respondents' latent satisfaction.

stringent. Similarly, if people calibrated their scale by means of fixed social comparisons, better outcomes of those in the comparison group will cause scale use to become more stringent.

However, a respondent's own contemporaneous characteristics will also influence comparisons. One such case are expectations, where respondent's contemporaneous characteristics determine her outlook of what is possible in her future.⁹ This may have two effects. First, better contemporaneous outcomes may improve expectations about the future, thus increasing current latent life satisfaction (Clark and Georgellis 2013). Second, altered expectations may change the standard against which current satisfaction is evaluated. This may cause respondents to adjust their scale such that future events can be accommodated by the updated scale.

To illustrate, consider *Lukas*, a final year PhD student. He would rate his life a 7 out of 10. If someone were to give him a paid academic position, one of his life's major goals would be accomplished. Some months later, Lukas obtains a three-year position as a postdoc. Consequently, he feels a great sense of satisfaction and feels inclined to rate himself a 9 out of 10. However, the satisfaction of a tenure track position, an achievement he did not dare dream of when a graduate student, is now no longer impossible. Lukas wants to reserve a 10 out of 10 for a truly exceptional outcome. Thus, a tenure-track position can be no better than a 9 out of 10. However, a postdoc is worse than a tenure-track position. He must therefore score his life with an 8 out of 10. Thus, on Lukas' old scale, as a student, a postdoc position would be a 9 out of 10. However, on his new scale, as a postdoc, a postdoc position is only rated an 8 out of 10.¹⁰

A second reason why current circumstances may determine response scales is that contemporaneous attributes likely determine the set of members of one's comparison group (Falk and Knell 2004). On the one hand, such endogenously updated reference groups could reduce the magnitudes of the effects of explanatory variables on latent satisfaction.¹¹ On the other hand, if respondents indeed interpret satisfaction scales as being relative to people like them, then response scales should shift following variation in personal characteristics.

To illustrate, consider *Mary*. Last year, she rated her life a 7 out of 10. She imagined that losing her partner would be devastating, resulting in a satisfaction of 2 out of 10. This year, Mary's partner does die. Consequently, life indeed feels much less satisfying. However, when comparing herself to others that also lost a partner, Mary thinks that she is doing relatively well. Hence, taking into account the newly acquired comparison, she rates her life a 5 out of 10. Therefore, in response to acquiring a new standard of comparison, her scale use has become less stringent. Suggestive evidence of such behaviour is given by Gibbons (1999) and Dermer et al. (1979).

To summarise, interpersonal and intrapersonal comparisons plausibly affect scale use. When considering variation in a respondent's past characteristics, or the characteristics of a fixed social reference group, changes that are expected to decrease respondents' own satisfaction are expected to make the response scale more stringent. However, when considering contemporaneous or future values (via anticipation) of respondents' personal characteristics, we should expect the direction in which scale use is impacted to concur with the direction in which latent satisfaction is impacted.

⁹ See Dominitz and Manski (1997) for the case of income expectations.

¹⁰ See Sprangers and Schwartz (1999) for similar reasoning.

¹¹ This might help to explain why people tend to overestimate the impacts of future events on their wellbeing (Wilson and Gilbert 2005).

2.3 Biases arising from violations of the common scale use assumption

In the ordered probit approach we usually estimate a model of the following form:¹²

$$s_{it} = \mathbf{X}_{it*}\boldsymbol{\beta}_* + \varepsilon_{it} \quad (4)$$

Here, $\mathbf{X}_{it*} \equiv (\mathbf{X}_{it+L}, \dots, \mathbf{X}_{it+l}, \dots, \mathbf{X}_{it+F})$, where each element is a row vector of J independent variables, defined as $\mathbf{X}_{it+l} \equiv (X_{1,it+l}, X_{2,it+l}, \dots, X_{J,it+l})$. Moreover, $\boldsymbol{\beta}_* \equiv (\boldsymbol{\beta}_L, \dots, \boldsymbol{\beta}_l, \dots, \boldsymbol{\beta}_F)'$, where each element is a column vector $\boldsymbol{\beta}_l \equiv (\beta_{1,l}, \dots, \beta_{j,l}, \dots, \beta_{J,l})'$ of parameters. Constants $L \leq 0$ and $F \geq 0$ respectively denote the numbers of included lags and leads. When no intrapersonal comparisons and anticipation effects are considered, no such lags and leads are included (i.e. $L = F = 0$). ε_{it} is assumed to be a normally distributed error with mean zero and standard deviation 1. A constant is omitted from equation (4). \mathbf{X}_{it*} may contain personal characteristics as well as attributes of a (fixed) social comparison group.

Assume that reports r_{it} for $k = 1, 2, \dots, K$ are generated by a rule of the form given in equation (1), imposing $\tau_0 = -\infty$, $\tau_K = \infty$. To specify all other cutoffs, write:

$$\tau_{it,k} = \gamma_k + \mathbf{X}_{it*}\boldsymbol{\delta}_* \quad (5)$$

For *common scale use* to hold, it is necessary that $\boldsymbol{\delta}_* = \mathbf{0}$, i.e. that neither contemporaneous nor lagged, nor leaded values of covariates affect the location of the thresholds. The contribution of each observation to the log-likelihood of the ordered probit model with the extension of equation (5) is then given by:

$$\ell_{it} = \ln \sum_{k=1}^K \mathbf{I}(r_{it} = k) [\Phi(\gamma_k + \mathbf{X}_{it*}\boldsymbol{\delta}_* - \mathbf{X}_{it*}\boldsymbol{\beta}_*) - \Phi(\gamma_{k-1} + \mathbf{X}_{it*}\boldsymbol{\delta}_* - \mathbf{X}_{it*}\boldsymbol{\beta}_*)]. \quad (6)$$

Here $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. $\mathbf{I}(\cdot)$ denotes the indicator function. Throughout we maximise $\ell = \sum_{i=1}^N \sum_{t=1}^{T_i} \ell_{it}$ to estimate parameters. T_i denotes the number of observations for respondent i , and N is the number of respondents. Since $\mathbf{X}_{it*}\boldsymbol{\delta}_* - \mathbf{X}_{it*}\boldsymbol{\beta}_* = \mathbf{X}_{it*}(\boldsymbol{\delta}_* - \boldsymbol{\beta}_*)$, equation (6) implies that $\boldsymbol{\delta}_*$ and $\boldsymbol{\beta}_*$ are not separately identified (King et al. 2004). Only the difference $\boldsymbol{\beta}_* - \boldsymbol{\delta}_*$ is identified.

A severe consequence of violating *common scale use* is that we might estimate incorrect signs for some explanatory variables. Suppose $\beta_{j,0} > 0$ and $\delta_{j,0} > 0$ for some explanatory variable $X_{j,it}$. An increase in the value of $X_{j,it}$ is here associated with a higher level of s_{it} and more stringent scale use. As discussed in the previous section, this may occur when $X_{j,it}$ records personal characteristics, variation of which could change a respondent's relevant comparison group. Now suppose that $\beta_{j,0} < \delta_{j,0}$. Hence, variation in $X_{j,it}$ impacts the location of the thresholds more strongly than it impacts latent satisfaction. In large samples, our estimate $(\widehat{\beta_{j,0}} - \widehat{\delta_{j,0}})$ would then be negative. If we interpreted this estimate as an estimate of $\beta_{j,0}$, we would falsely conclude that the effect of $X_{j,it}$ on s_{it} is negative, when in fact it is positive.

The opposite case is likely to occur when considering lagged values of personal covariates. If people adapt to changes in covariates (Frederick and Loewenstein 1999; Clark and Georgellis 2013), we should expect a variable whose contemporaneous effect is positive to have a negative lagged effect.

¹² Analogous analyses could also be given for an OLS model of rank-order r_{it} , but are omitted here for brevity.

However, the lagged effect of such a variable on scale use is likely positive. Hence, we expect that if $\beta_{j,-1} < 0$, then $\delta_{j,-1} > 0$. This means that if we are only able to estimate the difference $\beta_{j,-1} - \delta_{j,-1}$, and interpret this directly as an estimate of adaptation, we may overstate the extent of adaptation observed in empirical data.¹³ Conversely, in cases where we expect lags of variables to reinforce a smaller contemporaneous effect, such reinforcement will be underestimated.

For leads of explanatory variables, the previous section suggested equal signs for $\beta_{j,1}$ and $\delta_{j,1}$. If I expect a good to occur in the future, I should both be more satisfied this year and use the scale more stringently, so as to accommodate the future good. It then follows that anticipation effects on latent satisfaction will be underestimated. However, if X_j were to refer to an attribute of a fixed social comparison group (e.g. mean income in the reference group), then the previous section suggested opposite signs for $\beta_{j,*}$ and $\delta_{j,*}$.¹⁴ Consequently, an analysis which does not account for scale shifts likely overstates the effects of fixed reference groups on latent satisfaction.

Finally, wellbeing research is increasingly interested how variables X_j and X_p can be traded off while maintaining constant levels of s_{it} . Such analyses are particularly useful for computing shadow prices (e.g. Luechinger 2009; Danzer and Danzer 2016; Dolan et al. 2019) or for estimating equivalence scales (e.g. Borah, Keldenich, and Knabe 2019; Kaiser 2020). To obtain trade-offs, we must estimate the ratio β_j/β_p (dropping the ‘0’ subscript). However, we only obtain an estimate of $(\beta_j - \delta_j)/(\beta_p - \delta_p)$. Unless $\beta_j = \beta_p$ and $\delta_j = \delta_p$, biases occur in this context. Assuming that coefficients are ordered as $\beta_j \geq \beta_p > \delta_p \geq \delta_j \geq 0$ implies that our estimated trade-off ratio will be (weakly) positively biased. Likewise, assuming that $\beta_p \geq \beta_j \geq 0$, $\delta_j \geq \delta_p \geq 0$ and $\beta_p > \delta_p$ implies that our trade-off ratio will be (weakly) negatively biased.¹⁵

Thus, when the assumption of common response scales is violated and when response scales are correlated with explanatory variables, we might obtain incorrect signs of coefficients, false impressions of the magnitudes of effects, and biased trade-off ratios between variables. However, variation in scale use is entirely unproblematic if one is merely interested in analysing marginal effects of explanatory variables on the probability of responding with a certain category. This, together with some remarks on interpretations of absolute magnitudes of coefficients more generally, is shown in Appendix A.

3 Anchoring response scales

In order to separately identify scale shifts from effects on latent satisfaction, previous research anchored response scales by presenting respondents with short vignettes that describe an imaginary person’s life (King et al., 2004). In that approach, respondents are asked to rate the life satisfaction of the persons described in the vignettes. On the assumption that the vignettes’ perceived levels of satisfaction are uncorrelated with respondents’ personal characteristics, identification can be achieved. This assumption is known as the *vignette equivalence* assumption.

¹³ Since, if there is indeed adaptation and a positive lagged scale shift, $\beta_{j,-1} - \delta_{j,-1} < \beta_{j,-1} < 0$.

¹⁴ This should hold for lagged, contemporaneous and leaded values.

¹⁵ The difference $(\beta_j - \delta_j)/(\beta_p - \delta_p) - \beta_j/\beta_p$ can be rearranged to give $(\beta_j\delta_p - \beta_p\delta_j)/\beta_p(\beta_p - \delta_p)$. The numerator is positive or zero if $\beta_j \geq \beta_p \geq 0$ and $\delta_p \geq \delta_j \geq 0$. The denominator is positive if $\beta_p > \delta_p \geq 0$. The whole expression is therefore positive or zero if $\beta_j \geq \beta_p > \delta_p \geq \delta_j \geq 0$. Similarly, the numerator is negative or zero if $\beta_p \geq \beta_j \geq 0$ and $\delta_j \geq \delta_p \geq 0$.

In the case of life satisfaction research, Kapteyn, Smith, and Van Soest (2010), Corrado and Weeks (2010), Angelini et al. (2013), Bertoni (2015) and Montgomery (2017) have adopted the vignette approach. In line with my arguments of sections 2.2 and 2.3, these studies generally find that correcting for differences in scale use increases estimates of the effects of explanatory variables on latent satisfaction. A more extensive discussion of that approach and an appraisal of the associated literature is given in Appendix B.

However, the central assumption of the vignette approach, i.e. *vignette equivalence*, is untestable. As argued in Appendix B, violations of this assumption will likely lead to downwardly biased coefficients. Moreover, vignettes are expensive to administer, and the SHARE survey, which only samples individuals aged 50 or above, is the only widely accessible survey that contains vignettes designed for life satisfaction. Finally, there are currently no panel datasets that contain vignettes on life satisfaction. Thus, simultaneously accounting for time-invariant characteristics and differentials in scale use is currently impossible. Previous studies on life satisfaction that employ a vignette methodology are therefore hard to compare with studies employing a panel-data approach. Unobserved time-constant variables likely cause biases of their own. Therefore, previous work on life satisfaction that used vignettes may be biased even if *vignette equivalence* holds.

We should thus consider alternative approaches.

3.1 Using memories

This section sets out a strategy by which memories of life satisfaction can be used in a similar manner as vignettes. The idea of using memories to identify shifts in scale use was first proposed by Howard and colleagues (Howard et al. 1979; Howard and Dailey 1979; Howard, Schmeck, and Bray 1979). These studies first obtained a self-report at t_1 (the ‘pre-test’). At t_2 they performed some treatment. At t_3 , they asked respondents to recall their performance on the measure of interest at t_1 (the ‘then-test’). They also asked respondents to evaluate their current performance at t_3 (the ‘post-test’). A change in scale use is interpreted to occur if the treatment is correlated with the mean difference between ‘pre-test’ and ‘then-test’ scores. Other studies employing such ‘then-tests’ include Hoogstraten (1985), Sprangers et al. (1999), Ahmed et al. (2004) or Townsend and Wilton (2003).

Fabian (2019) is a more recent inspiration. He first asked respondents about levels of both their current life satisfaction and their satisfaction ten years ago. He then asked respondents to also draw how their satisfaction has changed in the last ten years. Inconsistencies between drawings of changes in satisfaction versus reports about levels of past and present satisfaction may indicate scale shifts. For example, consider a respondent reporting a ‘7’ for the present year, a ‘6’ concerning life ten years ago, and returning a drawing that indicates a decreasing trajectory. Such a respondent either has inconsistent memories about changes and levels of satisfaction, or the meaning of a ‘6’ ten years ago indicates a level of satisfaction higher than a level of ‘7’ this year. In the latter case, scale use has become less stringent. Fabian (2019) finds that about 8.5% of respondents are ‘reversers’ of this sort, and about 14% conform to a looser criterion of probable scale change.

3.2 Outline of memory approach

One possible question about memories of life satisfaction is:

‘Thinking about last year rather than this year, how dissatisfied or satisfied were you with your life overall?’

Such a question would be closest to the ‘then-test’ methodology of Howard and colleagues. In the framework of the present paper, a natural rule by which a response $r_{it}^{(m)}$ to this question is chosen mirrors equation (1):

$$r_{it}^{(m)} = k \leftrightarrow \tau_{it,k-1} < s_{it}^{(m)} \leq \tau_{it,k}. \quad (7)$$

Here, $s_{it}^{(m)}$ denotes the respondent's memory of satisfaction, which is not directly observed. Respondents are assumed to use the same set of thresholds for their current satisfaction (equation (1)) and their memories of past satisfaction (equation (7)). Following the literature on vignettes, call this the assumption of *response consistency*.

The BHPS, to be used in the empirical analysis of section 4, asks a question about change in satisfaction:

Would you say that you are more satisfied with life, less satisfied or feel about the same as you did a year ago?

Here, respondents have only three response options (*more satisfied, less satisfied, about the same*). In this context, let $r_{it}^{(m)} = -1$ denote that a respondent reports to be less satisfied this year, let $r_{it}^{(m)} = 0$ denote that a respondent reports ‘about the same’ satisfaction, and let $r_{it}^{(m)} = 1$ denote that a respondent reports to be more satisfied this year than last year. Notice that $r_{it}^{(m)} = -1$ means $s_{it} < s_{it}^{(m)}$ and $r_{it}^{(m)} = 1$ means $s_{it} > s_{it}^{(m)}$. Furthermore, $r_{it}^{(m)} = 0$ means $s_{it} = s_{it}^{(m)} + \epsilon_{it}$, where ϵ_{it} can be positive or negative, and we do not know its size. Recall that equation (1) states that $r_{it} = k$ implies $\tau_{it,k-1} < s_{it} \leq \tau_{it,k}$. Hence, we find the following implied reporting rule for memories:¹⁶

$$\begin{aligned} r_{it}^{(m)} = -1 \quad \& \quad r_{it} = k \quad \leftrightarrow \quad s_{it}^{(m)} > \tau_{it,k-1} \\ r_{it}^{(m)} = 1 \quad \& \quad r_{it} = k \quad \leftrightarrow \quad s_{it}^{(m)} < \tau_{it,k} \end{aligned} \quad (8)$$

This data can potentially be used to identify threshold locations. To do so, assume that one’s memory at t about last period’s life satisfaction is generated by:

$$s_{it}^{(m)} = \theta^{(m)} + \mathbf{X}_{it-1}\boldsymbol{\alpha} + \mathbf{X}_{it}\boldsymbol{\lambda}^{(m)} + \varepsilon_{it}^{(m)} \quad (9)$$

Here, $\varepsilon_{it}^{(m)}$ is a normally distributed error term with mean 0 and standard deviation 1. $\theta^{(m)}$ is an intercept to be estimated. For ease of exposition assume that there are no lags and leads in equation (1). This will be relaxed in section 3.4. In that case, and if $\boldsymbol{\lambda}^{(m)} = \mathbf{0}$, the structure of the equation determining memories of satisfaction is the same as the structure of the equation determining current satisfaction. In so far as memories of satisfaction are determined by how satisfied a respondent in fact was, this is a natural choice.

Indeed, in order to point-identify the effects of contemporaneous values of covariates on scale shifts, we must assume $\boldsymbol{\lambda}^{(m)} = \mathbf{0}$. Call this assumption *present independence*. Substantively this assumption states that, conditional on values of last year’s covariates, one’s memories about last year’s satisfaction do not depend on values for this year’s covariates.

¹⁶ $r_{it} = k$ & $r_{it}^{(m)} = 0$ implies $\tau_{it,k-1} < s_{it}^{(m)} + \epsilon_{it} \leq \tau_{it,k}$. Since we do not know the value of ϵ_{it} , cases where respondents answer ‘about the same’ are not informative about which reporting thresholds are crossed.

Present independence does not entail that reports $r_{it}^{(m)}$ are unbiased reports of past satisfaction s_{it-1} . This would imply the constraint that $\theta^{(m)} = \mathbf{0}$ and $\alpha = \beta$. I therefore allow for respondents to have biased memories of their past satisfaction and to misjudge the direction of the effects of past variables on past satisfaction (see e.g. Wilson, Meyers, and Gilbert 2003).

Using equations (1), (4), (6) and (9), together with the assumed distributions for each error term and maintaining $\lambda^{(m)} = \mathbf{0}$, individual contributions to the log-likelihood of this model are given by:

$$\begin{aligned} \ell_{it} = & \ln \sum_{k=1}^K I(r_{it} = k) [\Phi(\gamma_k + \mathbf{X}_{it}\delta - \mathbf{X}_{it}\beta) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}\delta - \mathbf{X}_{it}\beta)] \\ & + \sum_{k=1}^K I(r_{it}^{(m)} = k) [\Phi(\gamma_k + \mathbf{X}_{it}\delta - \theta^{(m)} - \mathbf{X}_{it-1}\alpha) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}\delta - \theta^{(m)} - \mathbf{X}_{it-1}\alpha)] \end{aligned} \quad (10)$$

When using data to which equation (8) applies, as is the case for BHPS data, the contributions are defined similarly:

$$\begin{aligned} \ell_{it} = & \ln \sum_{k=1}^K I(r_{it} = k) [\Phi(\gamma_k + \mathbf{X}_{it}\delta - \mathbf{X}_{it}\beta) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}\delta - \mathbf{X}_{it}\beta) \\ & + I(r_{it}^{(m)} = -1)(1 - \Phi(\gamma_{k-1} + \mathbf{X}_{it}\delta - \theta^{(m)} - \mathbf{X}_{it-1}\alpha)) \\ & + I(r_{it}^{(m)} = 1)\Phi(\gamma_k + \mathbf{X}_{it}\delta - \theta^{(m)} - \mathbf{X}_{it-1}\alpha)] \end{aligned} \quad (11)$$

Parameters β (effects on current satisfaction), α (effects on memories of past satisfaction) and δ (scale shift effects) are all separately identified by equations (10) or (11).

3.3 Plausibility of identifying assumptions

Since it seems implausible for respondents to vary their scale use when answering two questions about themselves, *response consistency* appears to be a weak assumption. In contrast, *present independence* may be a strong assumption. Although *present independence* does not require that there be no memory biases, it requires that the magnitude of any memory bias does not depend on values of current covariates. However, there are at least three mechanisms by which memory biases could depend on current values of covariates.

First, respondents might react to differences in objective circumstances between the present and the past and infer that a concordant change in the subjective quality of their lives must have taken place. A study by Conway and Ross (1984) is suggestive of this. There, participants were given a program to improve their study skills. When asked to recall their abilities before the treatment, participants recalled their abilities as worse than what they had self-reported before the treatment. However, participants did not in fact perform better on a final exam, nor did they differ in their self-assessment after the treatment. This suggests that the treatment had no actual effect on study skills. Conway and Ross (1984) interpret this as respondents biasing their memories to be less favourable, such that the expected gain in objective performance is at least subjectively realised. Their results suggest opposite signs for $\lambda^{(m)}$ and β .

Second, memory biases may also be caused by ‘mood-congruency’. Mood-congruency is a general tendency of positive (negative) moods to bias memories to be more positive (negative). As

evidence, Urban et al. (2018) find that depressive symptoms predict an upward bias in respondents' recall of negative emotions and a downward bias in the recall of positive emotions. Prati (2017) is an example from economic who shows that respondents who are subjectively more satisfied with their incomes recalled their incomes to have been larger than in reality, while those with low satisfaction had a negatively biased recall of their incomes. Further experimental evidence of mood-congruency is given by e.g. Lewis et al. (2005) and Klaassen et al. (2002). Given these findings, we should thus expect that people in more positive moods recall their past satisfaction as more positive. Explanatory variables that affect life satisfaction typically impact affect and domain-specific satisfaction in the same direction. Consequently, people's tendencies towards mood-congruent memories may cause an explanatory variable's effect on $\lambda^{(m)}$ to have the same direction as that variable's effect on β .

Anticipation, the phenomenon that satisfaction at t is in part driven by expectations about explanatory variables' values at $t + 1$ (Clark & Georgellis, 2013; Frijters et al., 2011; Vendrik, 2013), may be a third source of bias. Given anticipation and assuming that memories of satisfaction depend on actual past satisfaction, equation (9) for memories of satisfaction should include a term for future values of \mathbf{X} .¹⁷ However, since memories of satisfaction during $t - 1$ are modelled, that anticipation term concerns values of covariates at t . Consequently, such anticipation terms are given by \mathbf{X}_{it} . To observe the likely direction of bias, we can add \mathbf{X}_{it+1} to the equation for s_{it} and observe the estimated coefficients.¹⁸ If the associated coefficients are positive (negative), we may expect that elements in \mathbf{X}_{it+1} positively (negatively) impact memories $s_{it}^{(m)}$. In that case, *present independence* is violated. Previous studies find that anticipation effects typically have the same direction as contemporaneous effects, implying that $\lambda^{(m)}$ and β likely have the same signs.

Taking an overview, the argument based on Conway & Ross (1986) predicts opposing signs for $\lambda^{(m)}$ and β . The arguments based on mood-congruent reporting and anticipation effects predict equal signs for $\lambda^{(m)}$ and β . Rewriting equation (10) as follows helps us to assess the implications of either bias.¹⁹

$$\begin{aligned} \ell_{it} = & \ln \sum_{k=1}^K \mathbb{I}(r_{it} = k) \times \\ & [\Phi(\gamma_k + \mathbf{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \mathbf{X}_{it}(\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)})) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \mathbf{X}_{it}(\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)}))]_{(12)} \\ & + \sum_{k=1}^K \mathbb{I}(r_{it}^{(m)} = k) \times \\ & [\Phi(\gamma_k + \mathbf{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \theta^{(m)} - \mathbf{X}_{it-1}\boldsymbol{\alpha}) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \theta^{(m)} - \mathbf{X}_{it-1}\boldsymbol{\alpha})] \end{aligned}$$

We see that only the parameter differences $\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}$ and $\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)}$ are identified. If *present independence* for a particular variable X_j is satisfied, $\lambda_j^{(m)} = 0$ holds. In that case, $\delta_j - \lambda_j^{(m)} = \delta_j$ and $\beta_j - \lambda_j^{(m)} = \beta_j$. We can then correctly interpret these respective estimates as indicating effects on scale use and latent satisfaction. However, if $\lambda_j^{(m)}$ and β_j have opposite signs, then $|\beta_j - \lambda_j^{(m)}| >$

¹⁷ Since expectations are unobserved, we proxy current expectations with future realizations.

¹⁸ However, note that such coefficients may be overestimated via reverse causality of current satisfaction causing future values. See e.g. Vendrik (2013) or De Neve and Oswald (2012) for such endogeneity in the case of incomes.

¹⁹ The argument to follow would be analogous for the case in which equation (11) applies and is therefore omitted. Likewise, the following discussion also applies to the vignettes approach.

$|\beta_j|$. Estimates are then biased away from zero. In contrast, if the signs of $\lambda_j^{(m)}$ and β_j are equal and $|\lambda_j^{(m)}| < |\beta_j|$, we find that $|\beta_j - \lambda_j^{(m)}| < |\beta_j|$. Hence, in that case, estimate $(\beta_j - \lambda_j^{(m)})$, when interpreted as an estimate of β_j , is biased toward zero. Analogous analyses apply to the relationship between δ_j and $\lambda_j^{(m)}$. The results presented below can therefore be interpreted in two ways. More positive (negative) estimates $(\beta_j - \lambda_j^{(m)})$ are either interpreted as indicating a variable's positive (negative) effect on latent satisfaction or as indicating a variable's negative (positive) effect on memories of past satisfaction.

The possible biases described by Conway & Ross (1986) *versus* those arising from anticipation and mood-congruency may partially offset each other. Moreover, the emotional intensity of memories tends to fade with time (e.g. Levine and Bluck 2004). It therefore seems implausible that the absolute magnitude of the present-induced memory bias $\lambda^{(m)}$ should exceed the effect β on latent satisfaction.

We may therefore assume $|\lambda_j^{(m)}| \leq |\beta_j|$ for any particular variable X_j . Further assuming that $\lambda_j^{(m)}$ and β_j have opposite signs (which is the more worrying bias), we find that $|\lambda_j^{(m)}| \leq 0.5|(\beta_j - \lambda_j^{(m)})|$. On the basis of this bound on $\lambda_j^{(m)}$ we can assess whether the implied scale shift has a determinate direction. We may do so by inspecting the sign of $(\delta_j - \lambda_j^{(m)}) - 0.5(\beta_j - \lambda_j^{(m)})$. We can also assess whether the effects on latent satisfaction and scale use are statistically significant. This is done by respectively testing $h_0: 0.5(\beta_j - \lambda_j^{(m)}) = 0$ and $h_0: (\delta_j - \lambda_j^{(m)}) - 0.5(\beta_j - \lambda_j^{(m)}) = 0$.²⁰

3.4 Interpretation of lags and leads

I now turn to the case in which latent satisfaction also depends on lagged and leaded values of \mathbf{X} . For simplicity I only consider 1-year lags and leads, which is a special case of equation (4):

$$s_{it} = \mathbf{X}_{it-1}\boldsymbol{\beta}_{-1} + \mathbf{X}_{it}\boldsymbol{\beta}_0 + \mathbf{X}_{it+1}\boldsymbol{\beta}_{+1} + \varepsilon_{it} \quad (13)$$

Further lags and leads make the notation more arduous but present no further problem. Given the arguments of section 2.2, we may expect that scale use depends on the same lags and leads:

$$\tau_{it,k} = \gamma_k + \mathbf{X}_{it-1}\boldsymbol{\delta}_{-1} + \mathbf{X}_{it}\boldsymbol{\delta}_0 + \mathbf{X}_{it+1}\boldsymbol{\delta}_{+1} \quad (14)$$

If the equation for latent memories is to have the same structure as the equation for latent contemporaneous satisfaction, then equation (9) must be extended, too:

$$s_{it}^{(m)} = \theta^{(m)} + \mathbf{X}_{it-2}\boldsymbol{\alpha}_{-2} + \mathbf{X}_{it-1}\boldsymbol{\alpha}_{-1} + \mathbf{X}_{it}\boldsymbol{\alpha}_0 + \varepsilon_{it}^{(m)} \quad (15)$$

Note that $\boldsymbol{\alpha}_0$ will contain the effects of memory biases ($\lambda^{(m)}$) discussed in the previous section. The extended expression for the contributions to the log-likelihood can then be written as:

²⁰ Assuming that $\lambda_j^{(m)}$ and β_j have opposite signs, and assuming that $|\lambda_j^{(m)}| \leq |\beta_j|$, yields a worst case of $\beta_j = -\lambda_j^{(m)}$. In that worst case $(\beta_j - \lambda_j^{(m)}) = 2\beta_j \leftrightarrow 0.5(\beta_j - \lambda_j^{(m)}) = \beta_j$. Moreover, in that case, $(\delta_j - \lambda_j^{(m)}) = (\delta_j + \beta_j)$. Given these identities, $(\delta_j - \lambda_j^{(m)}) - 0.5(\beta_j - \lambda_j^{(m)}) = (\delta_j + \beta_j) - \beta_j = \delta_j$.

$$\begin{aligned}
\ell_{it} = & \ln \sum_{k=1}^K \mathbb{I}(r_{it} = k) \times \\
& [\Phi(\gamma_k + \mathbf{X}_{it+1}\boldsymbol{\delta}_{+1} + \mathbf{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \mathbf{X}_{it+1}\boldsymbol{\beta}_{+1} - \mathbf{X}_{it}(\boldsymbol{\beta}_0 - \boldsymbol{\alpha}_0) - \mathbf{X}_{it-1}(\boldsymbol{\beta}_{-1} - \boldsymbol{\delta}_{-1})) \\
& - \Phi(\gamma_{k-1} + \mathbf{X}_{it+1}\boldsymbol{\delta}_{+1} + \mathbf{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \mathbf{X}_{it+1}\boldsymbol{\beta}_{+1} - \mathbf{X}_{it}(\boldsymbol{\beta}_0 - \boldsymbol{\alpha}_0) - \mathbf{X}_{it-1}(\boldsymbol{\beta}_{-1} - \boldsymbol{\delta}_{-1})))] \\
& + \sum_{k=1}^K \mathbb{I}(r_{it}^{(m)} = k) \times \\
& [\Phi(\gamma_k + \mathbf{X}_{it+1}\boldsymbol{\delta}_{+1} + \mathbf{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \theta^{(m)} - \mathbf{X}_{it-1}(\boldsymbol{\alpha}_{-1} - \boldsymbol{\delta}_{-1}) - \mathbf{X}_{it-2}\boldsymbol{\alpha}_{-2}) \\
& - \Phi(\gamma_{k-1} + \mathbf{X}_{it+1}\boldsymbol{\delta}_{+1} + \mathbf{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \theta^{(m)} - \mathbf{X}_{it-1}(\boldsymbol{\alpha}_{-1} - \boldsymbol{\delta}_{-1}) - \mathbf{X}_{it-2}\boldsymbol{\alpha}_{-2})]
\end{aligned} \tag{16}$$

Only the leaded effects on latent satisfaction ($\boldsymbol{\beta}_{+1}$) and scale use ($\boldsymbol{\delta}_{+1}$), as well as the two-year lagged effect on memories of satisfaction ($\boldsymbol{\alpha}_{-2}$) are now uniquely identified. In contrast, estimates of the contemporaneous effects on latent satisfaction ($\boldsymbol{\beta}_0$) and scale use ($\boldsymbol{\delta}_0$) are contaminated by the effects on memories ($\boldsymbol{\alpha}_0$).

Fortunately, as also touched on in the previous section, since $\boldsymbol{\beta}_{+1}$ is identified (barring reverse causality), we can get some indication for the likely signs of $\boldsymbol{\alpha}_0$. This may aid us forming beliefs about the likely signs and magnitudes of $\boldsymbol{\delta}_0$. Moreover, only the difference of lagged effects on latent satisfaction and scale use ($\boldsymbol{\beta}_{-1} - \boldsymbol{\delta}_{-1}$) is identified. Given the arguments of section 2.2, $\boldsymbol{\beta}_{-1}$ and $\boldsymbol{\delta}_{-1}$ likely have opposing signs in the presence of adaptation, thus leading to an over-estimation of adaptation (or, equivalently, an underestimation of a lagged scale shift effect).

4 Empirical application

Having introduced the memory approach, I now turn to an empirical application using BHPS data. However, before doing so I will note a final difficulty for estimation.

4.1 Accounting for fixed effects

Unfortunately, failure to account for individual fixed effects severely biases coefficients in research on life satisfaction (Ferrer-i-Carbonell & Frijters, 2004). The standard approach to deal with this issue is to demean both sides of equation (4). However, since s_{it} is not observed, this is not an option.²¹ Alternatively estimating separate parameters for each individual is known to yield inconsistent estimates for finite T (Greene, 2004). To nevertheless correct for this issue, I add individual averages of all explanatory variables to the estimation. Such a specification was proposed by Mundlak (1978).

Van Praag (2015) recently argued that a specification which includes such terms yields estimates of $\boldsymbol{\beta}_*$ that are identical to estimates that would be obtained if it were possible to demean equation (4). Hence, although this is normally thought to be infeasible (Riedl and Geishecker 2014), adequately accounting for individual fixed-effects in an ordered probit context seems possible. Appendix B reconstructs Van Praag's argument. Appendix C successfully evaluates its veracity in a Monte Carlo study. My preferred estimates therefore include individual averages of all explanatory variables.

Finally, if latent satisfaction is partly determined by an individual fixed effect, then memories of satisfaction and scale use are also likely affected by individual fixed effects. To account for these, averages $\bar{\mathbf{X}}_i$ are added to these equations, too. Thus, the estimates shown below control for fixed effects in scale use. They are therefore primarily concerned with intrapersonal variation in scale

²¹ In the OLS context, applied researchers demean r_{it} , which is observed.

use. This is in contrast to vignette studies, which are always cross-sectional and therefore estimate interpersonal differences in scale use.

4.2 Data

I include waves 6-10 and 12-18 of the BHPS, corresponding to the years 1996-2000 and 2002-2008. These are all the years in which data on (memories of) life satisfaction are available. Life satisfaction is measured by the question given in section 2.1. Memories of satisfaction are elicited with the second question discussed in section 3.2.

I focus on the effects of log post-tax household income²², employment status, marital status, childbirth, and log household size. I also include a measure of reference income. Reference income is constructed as the log average income of those with similar age (i.e. respondent's age ± 5 years), and similar education (having the same education on a four category CASMIN scale). I additionally control for $\ln(1 + \text{working hours})$, homeownership, having a temporary job, age(-squared),²³ and wave dummies.

4.3 Descriptive evidence

In table 1, I provide some descriptive evidence. Panel A shows that more respondents state that last year was worse (27.9%), than that last year was better (15.3%). Moreover, the higher reported satisfaction, the higher the probability of stating that last year was worse. This is intuitive: People with higher latent satisfaction are, *ceteris paribus*, more likely to state that their life satisfaction has increased. Only 0.8% of respondents state that they 'don't know' whether last year was better or worse. Thus, respondents generally feel able to recall their past wellbeing.

Panel B cross-tabulates the observed change in r_{it} against respondents' remembered change. There are two situations that are evidently inconsistent. The first of these occurs when observing that reported life satisfaction decreased in subsequent years, but respondents nevertheless state that the past was worse (4.6% of all observations). The opposite situation occurs when observing an increase in reported satisfaction across subsequent years, but respondents report that last year was better (2.6% of all observations). The former (latter) case indicates that scale use has become more (less) stringent.

Panel C shows possible shifts in scale use by levels of reported satisfaction. Shifts are here counted as possible when people either report a change in satisfaction, but none is observed, or when remembered and observed changes are strongly inconsistent in the way discussed above. Counting possible scale shifts in this manner follow the mechanics of the model proposed in section 3 (which ignores information on respondents stating 'about same'). Using these definitions, 24.8% of all cases record a possible scale shift. The higher the level of reported satisfaction, the more likely it is that scale use has become more stringent. This is also intuitive: *Ceteris paribus*, those with less stringent scale use are more likely to respond with high response categories.

²² This variable is provided by Levy and Jenkins (2012). I additionally apply the modified OECD equivalence scale and deflate to 2005 prices.

²³ When simultaneously accounting for fixed effects and including wave dummies, the linear term for age drops out due to collinearity.

Table 1. Descriptive evidence on possible scale shifts

Last year was:’ ($r_{it}^{(m)}$)				
Panel A: <i>Memories of life satisfaction by level of reported life satisfaction</i>				
r_{it}	‘Worse’	‘About the same’	‘Better’	‘Don’t know’
1	116 (5.6%)	506 (24.2%)	1,371 (65.6%)	96 (4.6%)
2	197 (6.5%)	758 (24.8%)	2,031 (66.5%)	68 (2.2%)
3	688 (8.1%)	3,058 (36.0%)	4,587 (54.1%)	153 (1.8%)
4	2,758 (14.2%)	10,682 (54.8%)	5,761 (29.6%)	293 (1.5%)
5	11,183 (27.4%)	23,940 (58.7%)	5,359 (13.1%)	297 (0.7%)
6	17,090 (37.7%)	26,315 (58.1%)	1,745 (3.85%)	171 (0.4%)
7	6,724 (33.8%)	12,666 (63.6%)	418 (2.10%)	94 (0.5%)
Total	38,756 (27.9%)	77,925 (56.0%)	21,272 (15.3%)	1,172 (0.8%)
Panel B: <i>Memories of life satisfaction by direction of change in reported life satisfaction</i>				
Δr_{it}	‘Worse’	‘About the same’	‘Better’	‘Don’t know’
Decrease	4,620 (16.3%)	14869 (54.0%)	7792 (28.3%)	277 (1.0%)
Steady	12,649 (27.3%)	28571 (61.6%)	4,924 (10.6%)	233 (0.5%)
Increase	8,848 (34.2%)	14,296 (55.2%)	2,613 (10.1%)	151 (0.6%)
Panel C: <i>Possible shifts in scale use by levels of reported life satisfaction</i>				
r_{it}	Less stringent	No change	More stringent	
1	299 (21.8%)	1,007 (73.5%)	65 (4.7%)	
2	406 (19.0%)	1,626 (76.0%)	109 (5.1%)	
3	1,165 (19.3%)	4,510 (74.9%)	350 (5.8%)	
4	1,847 (13.3%)	10,680 (77.0%)	1,340 (9.7%)	
5	2,433 (8.2%)	21,608 (72.7%)	5,684 (19.1%)	
6	1,114 (3.4%)	24,476 (74.1%)	7,462 (22.6%)	
7	273 (2.0%)	11,185 (81.5%)	2,259 (16.5%)	
Total	7,537 (7.5%)	75,092 (75.2%)	17,269 (17.3%)	

Note: Percentages are relative to row-totals. Bold cells indicate possible shifts in scale use. Observations are counted as ‘less stringent’ when respondents’ reported satisfaction either increased or remained steady, but respondents reported that life last year was better. Observations are counted as ‘more stringent’ when respondents’ reported satisfaction either decreased or remained steady, but respondents reported that life last year was worse. Case numbers for panels B and C are lower because first differences are used, thus dropping waves 6 and 12.

A recent study by Prati and Senik (2020) uses the same data. Fortunately, they observe the same descriptive patterns. They offer an alternative interpretation of the data, primarily invoking memory biases. The present paper instead interprets these patterns as being indicative of scale shifts. As explained in section 3.2, the interpretation invoked in this paper does not require a complete absence of the sorts of memory biases suggested by Prati & Senik (2020). Both interpretations may therefore be (partly) true. Unfortunately, the data at hand does not allow us to take an ultimate stance on this. Nevertheless, the conclusion will offer some suggestion on how future research may decide between scale shifts and memory biases.

Unfortunately, these descriptive results do not show which variables are associated with changes in scale use.²⁴ They also do not show how scale shifts affect estimates of the determinants of latent satisfaction. Therefore, I now turn to analyses using the model described in section 3.2.

4.4 Baseline results

Table 2 presents my baseline results. Columns (1) and (2) give results without individual averages of explanatory variables. Columns (3) and (4) include such averages and thereby account for individual fixed effects. Column (2) and (4) account for possible shifts in scale use.

²⁴ Table E1 in appendix E shows results from multinomial logistic regressions in which indicators of whether possible scale shifts occurred are regressed on the same set of socio-economic variables as used in section 4.4. These estimates indicate that variables known to be associated with higher (lower) reported satisfaction are also associated with higher odds of shifting toward more (less) stringent scale use. This is in line with the theoretical expectations of section 2.2 and the results of section 4.4.

Table 2. Main Results for BHPS data

	Without averages	Without averages, with scale shift	With averages	With averages, with scale shift
	(1)	(2)	(3)	(4)
Panel A: Current life satisfaction				
ln(HH income)	0.089*** (0.010)	0.127*** (0.020)	0.037*** (0.007)	0.097*** (0.023)
Employed (base)				
<i>Self-employed</i>	0.043 (0.027)	0.092 (0.057)	0.022 (0.021)	0.072 (0.061)
<i>Unemployed</i>	-0.316*** (0.033)	-0.709*** (0.063)	-0.197*** (0.025)	-0.615*** (0.069)
<i>Retired</i>	0.120*** (0.030)	0.137** (0.068)	0.070*** (0.022)	0.192*** (0.073)
<i>Family care</i>	-0.026 (0.027)	-0.107** (0.052)	-0.002 (0.020)	-0.048 (0.055)
<i>Lt. sick / disabled</i>	-0.799*** (0.036)	-1.104*** (0.072)	-0.319*** (0.030)	-0.542*** (0.082)
<i>In education</i>	0.130*** (0.025)	0.121** (0.055)	0.095*** (0.022)	0.056 (0.063)
Married (base)				
<i>Widowed</i>	-0.317*** (0.032)	-1.112*** (0.113)	-0.336*** (0.044)	-1.007*** (0.137)
<i>Divorced</i>	-0.440*** (0.025)	-0.592*** (0.058)	-0.262*** (0.028)	-0.369*** (0.069)
<i>Never married</i>	-0.203*** (0.019)	-0.417*** (0.053)	-0.114*** (0.021)	-0.287*** (0.064)
ln(HH size)	-0.086*** (0.014)	-0.199*** (0.042)	-0.087*** (0.015)	-0.216*** (0.048)
Childbirth	0.143*** (0.018)	0.351*** (0.041)	0.081*** (0.015)	0.224*** (0.044)
Reference income	-0.271*** (0.035)	0.149 (0.148)	-0.254*** (0.055)	-0.220 (0.187)
Panel B: Memory of last year's life satisfaction				
ln(HH income)	0.005 (0.016)	0.034* (0.017)	0.013 (0.020)	0.027 (0.020)
Employed (base)				
<i>Self-employed</i>	-0.013 (0.045)	0.046 (0.051)	-0.003 (0.057)	0.010 (0.058)
<i>Unemployed</i>	-0.084 (0.053)	-0.198*** (0.054)	-0.132** (0.062)	-0.135** (0.063)
<i>Retired</i>	0.016 (0.057)	0.032 (0.065)	-0.124* (0.068)	-0.081 (0.073)
<i>Family care</i>	-0.060 (0.045)	-0.100** (0.047)	-0.106** (0.051)	-0.115** (0.053)
<i>Lt. sick / disabled</i>	-0.050 (0.058)	-0.249*** (0.065)	-0.271*** (0.076)	-0.314*** (0.078)
<i>In education</i>	0.008 (0.044)	-0.006 (0.048)	0.042 (0.054)	-0.008 (0.056)
Married (base)				
<i>Widowed</i>	-0.244*** (0.057)	-0.969*** (0.113)	-0.627*** (0.109)	-0.968*** (0.135)
<i>Divorced</i>	-0.313*** (0.034)	-0.466*** (0.055)	-0.328*** (0.054)	-0.419*** (0.063)
<i>Never married</i>	-0.105*** (0.027)	-0.309*** (0.050)	-0.180*** (0.048)	-0.285*** (0.057)
ln(HH size)	0.108*** (0.021)	0.007 (0.039)	0.081** (0.035)	-0.008 (0.046)
Childbirth	0.042 (0.036)	0.047 (0.036)	0.109*** (0.038)	0.093** (0.039)
Reference income	-0.316*** (0.048)	0.134 (0.145)	0.165 (0.144)	0.216 (0.173)
Constant	-0.884* (0.495)	-1.020** (0.510)	-1.086 (0.675)	-0.919 (0.782)
Panel C: Scale shift				
ln(HH income)		0.037** (0.018)		0.059*** (0.021)
Employed (base)				
<i>Self-employed</i>		0.049 (0.050)		0.051 (0.056)
<i>Unemployed</i>		-0.393*** (0.056)		-0.418*** (0.065)
<i>Retired</i>		0.017 (0.062)		0.122* (0.070)
<i>Family care</i>		-0.080* (0.046)		-0.047 (0.051)
<i>Lt. sick / disabled</i>		-0.306*** (0.064)		-0.223*** (0.077)
<i>In education</i>		-0.009 (0.050)		-0.039 (0.058)
Married (base)				
<i>Widowed</i>		-0.795*** (0.110)		-0.671*** (0.132)
<i>Divorced</i>		-0.152*** (0.053)		-0.107* (0.060)
<i>Never married</i>		-0.214*** (0.050)		-0.173*** (0.060)
ln(HH size)		-0.113*** (0.040)		-0.129*** (0.047)
Childbirth		0.208*** (0.038)		0.142*** (0.042)
Reference inc.		0.420*** (0.145)		0.034 (0.179)
Panel D: Cutoff constants				
γ_1	-4.794*** (0.310)	-4.791*** (0.310)	-4.828*** (0.350)	-4.827*** (0.350)
γ_2	-4.382*** (0.025)	-4.379*** (0.025)	-4.413*** (0.025)	-4.412*** (0.025)
γ_3	-3.837*** (0.014)	-3.835*** (0.014)	-3.865*** (0.015)	-3.865*** (0.015)
γ_4	-3.200*** (0.010)	-3.199*** (0.010)	-3.224*** (0.010)	-3.224*** (0.010)
γ_5	-2.329*** (0.007)	-2.329*** (0.007)	-2.348*** (0.007)	-2.348*** (0.007)
γ_6	-1.283*** (0.007)	-1.283*** (0.007)	-1.297*** (0.007)	-1.297*** (0.007)
Obs. (Respondents)	118, 440 (20,572)	118, 440 (20,572)	118,440 (20,572)	118,440 (20,572)

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Being unemployed or having a long-term sickness reduces satisfaction, compared to being employed. Being in education has a small positive effect. The self-employed, those retired and those who are in family care are not estimated to have a different level of satisfaction compared to the employed. In line with previous literature, accounting for fixed effects reduces the size of estimated coefficients (compare columns 1 and 3).²⁵ However, upon accounting for scale shifts (see columns 2 and 4), coefficients again become larger. Of those relating to job status, the coefficient on unemployment is increased the most. Thus, while a standard analysis would show that the effect of long-term sickness is stronger than that of unemployment, adjusting for scale use yields the opposite result. These increases in magnitudes of the positive (negative) β estimates²⁶ are driven by corresponding positive (negative) δ estimates relating to scale use. For example, the unemployed are estimated to use the response scale much less stringently. In contrast, those in retirement use the scale much more stringently.

Higher incomes are associated with higher life satisfaction. The size of the effect is reduced when accounting for fixed effects. When additionally accounting for changes in scale use, the income effect is again increased. Hence respondents' scale use is more stringent with higher incomes. Regarding reference income, the association with scale use is positive. This is in line with the theoretical expectation of section 2.2. In the case of column (2) this association is quite extreme, causing the coefficient on latent satisfaction to switch signs. However, this result is not found in the preferred specification of column (4), where I account for fixed effects. Although the coefficient on scale use is small there, controlling for scale shifts nevertheless causes the coefficient on latent satisfaction to become insignificant.

Those who are neither married nor in a stable partnership are much less satisfied. Comparing columns (1) and (3), these estimates are generally reduced when accounting for fixed effects, though less so than what was observed for employment statuses. Upon accounting for changes in scale use, the negative effects of not being married are increased. This change is particularly large for widowhood, where the associated coefficient is tripled in size, and becomes the largest coefficient in the model. This is also reflected by widowhood having the largest coefficient in the scale shift equation.

Finally, scale use becomes significantly more stringent following childbirth and less stringent in larger households. The respectively positive and negative effects of these variables on satisfaction are consequently magnified, too. Overall, my results agree with the theoretical expectations of section 2.2. Personal characteristics that are associated with higher (lower) levels of latent satisfaction are also associated with more (less) stringent scale use. Unsurprisingly, this is especially so for bereavement and unemployment, both of which have particularly profound impacts upon people's lives and their outlook upon life.

As noted in section 2.3, trade-offs between variables may be affected by scale shifts. Table 3 therefore records the estimated ratios $\beta_j / \beta_{\ln(HH \text{ inc.})}$ for all explanatory variables j of interest. This exercise shows more clearly that, relative to the effect of income, the effects of unemployment,

²⁵ Differences between cut-off constants γ_k are almost identical across all models. Thus, the sizes of coefficients are directly comparable. See Appendix A for further discussion on this.

²⁶ When writing variants of 'estimates of β ', I implicitly assume that $\lambda^{(m)} = \mathbf{0}$. It would be more accurate, but more cumbersome, to always write 'estimates of $\beta - \lambda^{(m)}$ '. The implications of $\lambda^{(m)}$ potentially not being equal to $\mathbf{0}$ are further explored in section 4.5

family care and widowhood become larger when accounting for scale shifts. In contrast, the effects of long-term sickness, education and divorce are smaller when accounting for scale shifts.

Table 3. Ratios of coefficients on each explanatory variable against coefficient on income

	Without averages		With averages	
	No shift	With shift	No shift	With shift
ln(HH income)	1.000	1.000	1.000	1.000
Employed (base)				
<i>Self-employed</i>	0.480	0.723	0.579	0.750
<i>Unemployed</i>	-3.539	-5.607	-5.291	-6.371
<i>Retired</i>	1.347	1.082	1.877	1.983
<i>Family care</i>	-0.296	-0.843	-0.048	-0.501
<i>Lt. sick / disabled</i>	-8.940	-8.727	-8.565	-5.613
<i>In education</i>	1.460	0.959	2.541	0.576
Married (base)				
<i>Widowed</i>	-3.549	-8.787	-9.004	-10.423
<i>Divorced</i>	-4.925	-4.680	-7.035	-3.826
<i>Never married</i>	-2.276	-3.299	-3.054	-2.970
ln(HH size)	-0.965	-1.576	-2.331	-2.234
Childbirth	1.604	2.778	2.178	2.315
Reference income	-3.038	1.177	-6.818	-2.280

Note: Ratios of coefficients computed on the basis of table 2.

4.5 Sensitivity to violations of *present independence*

With the data at hand, the validity of the *present independence* assumption cannot be tested. However, we do obtain estimates of the lagged effects of explanatory variables on memories of satisfaction (parameter α in equation (9)). These are shown in panel B of table 2. Looking at the preferred model in column (4), signs of coefficients on current *versus* remembered satisfaction typically equal.²⁷ For example, just as being unemployed reduces contemporaneous satisfaction, having been unemployed reduces remembered satisfaction. Thus, estimated signs of coefficients suggest that respondents are generally able to recall the direction in which past personal circumstances affected their past satisfaction. Moreover, almost all coefficients on memories have smaller magnitudes compared to those for current satisfaction. Though not conclusive, both these findings suggest that violations of *present independence* may be small.

It is nevertheless important to evaluate how violations of *present independence* could affect my conclusions. If the contemporaneous effect on latent satisfaction has the same sign as a potential contemporaneous effect on memory bias, then coefficients will be biased towards zero.²⁸ The same relation holds with respect to memory biases and effects on scale use. Note that when accounting for differences in scale use, the magnitudes of estimates on latent satisfaction almost always exceed those obtained when failing to account for scale shifts. This implies that if *present independence* is violated in a manner in which β and $\lambda^{(m)}$ have the same signs, the resulting biases (which lead to an underestimation of the true β) are smaller than the biases resulting from failing to account for changes in scale use (which also lead to an underestimation of the true β).

However, we do not know whether the opposite case, i.e. opposing signs for β and $\lambda^{(m)}$ occurs. Given the small lagged coefficients on memories of satisfaction, and the arguments of section 3.3,

²⁷ The cases for which signs differ are retirement, education, and reference income. However, in the case of education, the coefficient on latent satisfaction is close to zero, making it easy to observe sign switches.

²⁸ In an extreme case where both effects have the same sign, but the memory bias effect has a larger magnitude, it is possible to obtain inverted signs of coefficients. Since all coefficients have the expected sign, this does not appear to be the case here.

which suggested that the signs of $\lambda^{(m)}$ are indeterminate, it seems implausible that the absolute magnitudes of $\lambda^{(m)}$ could exceed those of β . Thus, adopting the assumption that the magnitudes of $\lambda^{(m)}$ are at most equal to those of β , we can test whether effects of covariates on scale use, i.e. δ , remain significant and have a determinate sign in such a worst-case. As explained in section 3.3, this can be done by testing $h_0: (\widehat{\delta_j} - \widehat{\lambda_j^{(m)}}) - 0.5(\widehat{\beta_j} - \widehat{\lambda_j^{(m)}}) = 0$ for every variable of interest. Table 4 presents results from such an exercise on the basis of the estimates of column (4) of table 2.

Table 4. Robustness of scale shift effects

	Worst-case estimate	Standard error	p-value
ln(HH income)	0.011	0.011	0.332
Employed (base)			
<i>Self-employed</i>	0.015	0.029	0.615
<i>Unemployed</i>	-0.111***	0.035	0.002
<i>Retired</i>	0.026	0.037	0.489
<i>Family care</i>	-0.022	0.028	0.415
<i>Lt. sick / disabled</i>	0.048	0.042	0.249
<i>In education</i>	-0.067**	0.031	0.030
Married (base)			
<i>Widowed</i>	-0.168**	0.070	0.017
<i>Divorced</i>	0.077**	0.032	0.014
<i>Never married</i>	-0.030	0.032	0.353
ln(HH size)	-0.021	0.025	0.391
Childbirth	0.031	0.023	0.177
Reference income	0.144	0.094	0.126

Note: Estimates are based on column (4) of table 2. Test computed under assumption that $\lambda_j^{(m)}$ is not larger than β_j and of opposite sign. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The signs of most estimated scale shift effects are robust to this exercise. Own income, self-employment, being retired, and childbirth continue to be associated with more stringent scale use. Unemployment, family care, widowhood, never being married, and household size remain associated with less stringent scale use. However, only the estimates for widowhood and unemployment remain statistically significant. Moreover, contrary to the baseline assumption of $\lambda^{(m)} = \mathbf{0}$, disability and divorce are now seen to be associated with more stringent scale use.²⁹ Finally note that assuming $|\lambda^{(m)}| = |\beta|$ (and opposing signs) causes all coefficients on latent satisfaction to be halved. Table E2 in Appendix E shows that the statistical significance of each relevant coefficient is unaffected by this reduction in estimated magnitudes.

4.6 Comparison to standard ordered probit and OLS models

The models of table 2 are unusual. It is therefore instructive to compare them to standard ordered probit and OLS models. Results from these are shown in table 5. Recall that in all columns of table 2, the effects of explanatory variables on memories of satisfaction and current satisfaction are estimated by assuming that the same set of thresholds is used for memories as for current satisfaction. If that assumption were violated, we should expect the standard ordered probit results given in columns (1) and (2) of table 5 to strongly differ from those in columns (1) and (3) of table 2. However, signs and magnitudes of estimated coefficients are similar across the models, and the estimated cutoff locations are in agreement, too. These results support the assumption of *response consistency*.

²⁹ In the cases of reference income and being in education, $\widehat{\delta_j}$ and $\widehat{\beta_j}$ have opposing signs. This leads to an exacerbation of the estimated scale use effect in this test.

Table 5. Comparison to Ordered Probit and OLS models

	OP	OP-Avg.	OP, equidistant cutoffs imposed	OP-Avg., equidistant cutoffs imposed	OLS	OLS-FE
	(1)	(2)	(3)	(4)	(5)	(6)
ln(HH income)	0.087*** (0.010)	0.036*** (0.007)	0.088*** (0.010)	0.036*** (0.007)	0.109*** (0.011)	0.042*** (0.009)
Employed (base)						
<i>Self-employed</i>	0.041 (0.026)	0.021 (0.020)	0.051** (0.026)	0.025 (0.020)	0.070** (0.030)	0.032 (0.024)
<i>Unemployed</i>	-0.309*** (0.032)	-0.193*** (0.025)	-0.343*** (0.033)	-0.226*** (0.026)	-0.420*** (0.040)	-0.278*** (0.032)
<i>Retired</i>	0.119*** (0.030)	0.069*** (0.022)	0.100*** (0.029)	0.058*** (0.022)	0.108*** (0.033)	0.062** (0.026)
<i>Family care</i>	-0.025 (0.026)	-0.001 (0.019)	-0.040 (0.026)	-0.014 (0.020)	-0.053* (0.031)	-0.020 (0.023)
<i>Lt. sick / disabled</i>	-0.782*** (0.036)	-0.313*** (0.029)	-0.874*** (0.038)	-0.368*** (0.032)	-1.066*** (0.046)	-0.450*** (0.038)
<i>In education</i>	0.127*** (0.025)	0.093*** (0.021)	0.123*** (0.025)	0.086*** (0.022)	0.153*** (0.029)	0.101*** (0.026)
Married (base)						
<i>Widowed</i>	-0.310*** (0.032)	-0.330*** (0.043)	-0.308*** (0.031)	-0.323*** (0.044)	-0.362*** (0.035)	-0.368*** (0.049)
<i>Divorced</i>	-0.431*** (0.024)	-0.257*** (0.028)	-0.446*** (0.025)	-0.263*** (0.029)	-0.538*** (0.030)	-0.316*** (0.035)
<i>Never married</i>	-0.199*** (0.019)	-0.111*** (0.021)	-0.199*** (0.019)	-0.111*** (0.021)	-0.240*** (0.022)	-0.136*** (0.025)
ln(HH size)	-0.084*** (0.014)	-0.085*** (0.014)	-0.082*** (0.014)	-0.082*** (0.014)	-0.099*** (0.017)	-0.097*** (0.017)
Childbirth	0.140*** (0.018)	0.080*** (0.014)	0.132*** (0.018)	0.073*** (0.015)	0.156*** (0.021)	0.084*** (0.017)
Reference income	-0.310*** (0.032)	-0.330*** (0.043)	-0.308*** (0.031)	-0.323*** (0.044)	-0.362*** (0.035)	-0.368*** (0.049)
Cutoffs						
τ_1	-4.745*** (0.304)	-4.780*** (0.344)	-5.077*** (0.301)	-5.108*** (0.340)		
τ_2	-4.326*** (0.024)	-4.358*** (0.024)	-4.286*** (0.005)	-4.312*** (0.005)		
τ_3	-3.789*** (0.014)	-3.817*** (0.014)	-3.496*** (0.005)	-3.516*** (0.005)		
τ_4	-3.172*** (0.010)	-3.196*** (0.010)	-2.705*** (0.005)	-2.720*** (0.005)		
τ_5	-2.334*** (0.007)	-2.352*** (0.007)	-1.915*** (0.005)	-1.924*** (0.005)		
τ_6	-1.294*** (0.007)	-1.307*** (0.007)	-1.124*** (0.005)	-1.127*** (0.005)		
Observations	118,773	118,773	118,773	118,773	118,773	118,773

Note: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Furthermore, signs and significance levels of the OLS estimates shown in columns (5) and (6) - the latter of which account for fixed effects - match the corresponding results of the ordered probit models. Ratios of coefficients are also in agreement between the ordered probit and OLS models, though they do not match perfectly. One example is the ratio of the income to unemployment coefficients. This ratio is $0.036/-0.193 = -0.187$ in the ordered model with individual averages (“OP-Avg.”), but $0.042/-0.278 = -0.151$ in the OLS-FE model. Part of the discrepancy is driven by the ordered probit model allowing cutoffs to not be equidistant, while the OLS model assumes constant differences in expected s_{it} across response categories. To verify this, I constrained cutoffs to be equidistant in the ordered probit models of columns (3) and (4). When doing so, the estimated coefficient ratios are even closer to the corresponding ratios estimated in the OLS models (e.g. for the income to unemployment ratio in column (4), I now obtain $0.036/-0.226 = -0.159$). Thus, it indeed appears that adding individual averages to the ordered probit model accurately accounts for fixed effects. These results therefore agree with the extended theoretical discussion of Appendix C and the Monte Carlo results of Appendix D.

Finally, note that the magnitudes of coefficients in the OLS model are uniformly higher than in the ordered probit models. This is because the OLS model assumes constant differences of 1 in expected s_{it} across response categories. The constrained ordered probit model with $\sigma = 1$ instead estimates differences of 0.796 between thresholds. Hence, coefficients of the OLS model can be expected to be larger by a factor of about $1/0.796 = 1.256$.

4.7 Lags, leads, and reference effects

We may be worried that the estimates of table 2 are affected by lags and leads of explanatory variables. Effects of past and future values of explanatory variables on latent satisfaction and scale use may also be of substantive interest. Table E3 in the appendix therefore presents such results.

For each of the variables relating to marital status, I obtain positive lagged effects on latent satisfaction. The corresponding contemporaneous coefficients are all negative (and of larger magnitude). We might interpret this result as implying that deviating from being married is subject to partial adaptation. This interpretation follows Clark and Georgellis (2013). However, the discussion of section 3.4 showed that lagged coefficients in this model (and any other model in the literature) are contaminated by lagged effects on scale use. For each variable relating to marital status, I do obtain negative contemporaneous effects on scale use. There may therefore also be negative lagged effects on scale use. In that case, the estimated lagged coefficients on latent satisfaction are biased away from zero, suggesting more adaptation than really occurs. Unfortunately, the data cannot decide on this. Similar dynamics are also observed for self-employment (c.f. Hanglberger and Merz 2015) and being in education. For all other variables, no such dynamics are observed.

To capture anticipation, I also consider the addition of lead terms. Their addition reduces the magnitudes of the contemporaneous coefficients on latent satisfaction relating to retirement, divorce and being single. For all variables relating to marital status, I find negative lead coefficients. Hence, a part of the total effect relating to these variables may be driven by anticipation. A similar dynamic is also observed for childbirth. More generally, the signs of coefficients on latent satisfaction associated with each lead variable are always the same as the corresponding signs of coefficients of each contemporaneous variable. As explained in section 3.4, this equality of signs reduces the likelihood that β and $\lambda^{(m)}$ have opposite signs.

The estimated effects of leaded variables on scale use are similar to those on latent satisfaction. For example, being married in the next year is associated with both more stringent scale use and greater latent satisfaction. Adding lags and leads to the model jointly does not yield substantially different results.

Scale use may also be impacted via comparisons to a fixed reference group. However, although the sign of the reference income effect on scale use was positive in my baseline estimations, it was insignificant. In order to investigate this further, I computed a set of additional ‘reference’ variables. These variables include the percentages of those in the reference group who are: employed, unemployed, retired, disabled, married, divorced, widowed, or had a birth in their household. In addition, I constructed variables measuring mean log household size and mean reported life satisfaction of those in the reference group.

When jointly adding these variables and accounting for fixed effects (i.e. akin to column (4) of table 2), almost all of these variables’ coefficients on latent satisfaction, memories, and scale use are statistically insignificant and small. See table E4 in the appendix. Mean reported satisfaction in the reference group is the only major exception to this. That variable shows a large positive coefficient on latent satisfaction, which suggests that having a more satisfied reference group makes respondents themselves more satisfied. However, mean reported satisfaction neither affects memories nor scale use.³⁰

4.8 Parallel and non-parallel shifts

So far I only considered shifts with equal magnitude and direction for all cutoffs. Xu & Xie (2016), call this a *parallel shift* in the cut-offs. However, it may be that independent variables impact the location of each cut-off differently. Thus, variables may cause scale use to become more stringent at high levels of reported satisfaction, but less stringent at low levels of reported satisfaction. This idea is illustrated in figure 2. In order to allow for non-parallel shifts, it is common to postulate a nested model for the set of thresholds:³¹

$$\begin{aligned}\tau_{it,1} &= \gamma_1 + \mathbf{X}_{it}\boldsymbol{\delta}_1, \\ \tau_{it,2} &= \tau_{it,1} + \exp(\gamma_2 + \mathbf{X}_{it}\boldsymbol{\delta}_2), \\ &\dots \\ \tau_{it,K} &= \tau_{it,K-1} + \exp(\gamma_K + \mathbf{X}_{it}\boldsymbol{\delta}_K).\end{aligned}\tag{17}$$

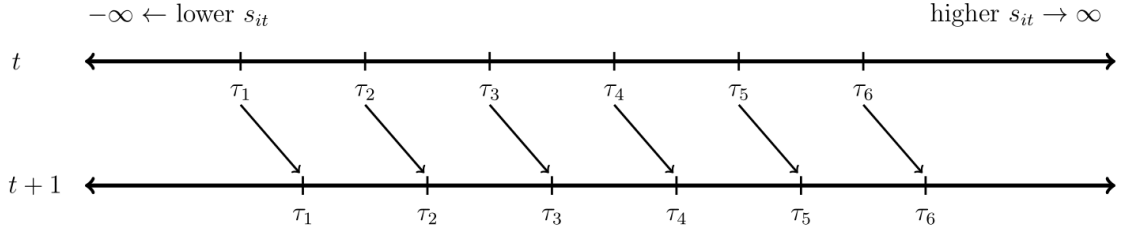
This non-linear specification has the advantage that the ordering of cutoffs is always preserved, no matter the signs of $\boldsymbol{\delta}_k$ (Greene and Hensher 2010). However, this specification is inelegant since we assume a non-linear functional form for all but the first cutoff. A linear specification is therefore more natural.

³⁰ Mean reported satisfaction might also pick up effects from personal characteristics that are not included in the model. Furthermore, as can be gleaned from columns (2)-(4) of Table E4, these results are robust to the addition of lags and leads. One reason for these surprisingly insignificant results may be collinearity between the reference variables. In table E5, I therefore estimated sparser models in which only reference income and percentages of employed, disabled, widowed, having birth, and mean satisfaction were added (these variables were closest to statistical significance in the full model). However, results were largely the same in these sparser models. I also estimated further models for reference satisfaction in which reference groups were defined depending on personal characteristics. This yielded similar results. See table E6 in the appendix.

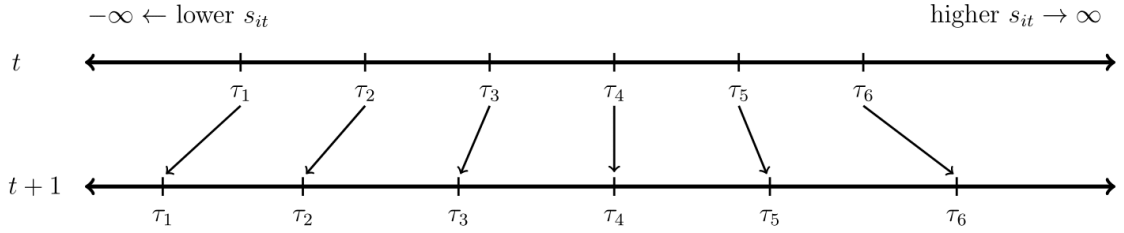
³¹ In virtue of this ‘nesting’ property, this model is sometimes referred to as a ‘Hierarchical Ordered Probit’, or ‘HOPIT’ model.

Figure 2. Parallel and non-parallel shifts in scale use.

Panel A. Parallel Shift



Panel B. Non-parallel Shift



Such a specification may be given by:

$$\begin{aligned}
 \tau_{it,1} &= \gamma_1 + \mathbf{X}_{it}\boldsymbol{\delta}_1, \\
 \tau_{it,2} &= \tau_{it,1} + \gamma_2 + \mathbf{X}_{it}\boldsymbol{\delta}_2, \\
 &\dots \\
 \tau_{it,K} &= \tau_{it,K-1} + \gamma_K + \mathbf{X}_{it}\boldsymbol{\delta}_K.
 \end{aligned} \tag{18}$$

Here $\boldsymbol{\delta}_k$ for $k > 1$ may be interpreted as increasing the distance between adjacent thresholds (when $\boldsymbol{\delta}_k > 0$) or as decreasing the distance between adjacent thresholds (when $\boldsymbol{\delta}_k < 0$). In the former case, scale use is ‘stretched’, while in the latter case it is ‘squeezed’.

Note that in the specification of equation (18) it may occur that $\tau_{it,k-1} > \tau_{it,k}$ for some values of \mathbf{X}_{it} and some k when not all corresponding elements in $\boldsymbol{\delta}_k$ and $\boldsymbol{\delta}_{k-1}$ have the same sign. Although theoretically undesirable, empirically this does not appear to be the case. Given its greater elegance and ease of interpretation, I prefer the specification of equation (18). Results for such a model are thus presented in table 6.³² That model accounts for fixed effects. We can therefore compare these results to those of column (4) in table 2.

Results are almost unchanged, including those on scale use (compare the estimates of $\boldsymbol{\delta}$ in table 2 to the corresponding estimates of $\boldsymbol{\delta}_1$ in table 6). Consequently, I find little evidence of non-parallel shifts in scale use. The only clear such cases are observed for various employment statuses. In these, coefficients on the higher thresholds tend to be significantly negative. This may indicate that for those who are not employed, scale use is ‘squeezed’ at the upper end of the response scale.

³² Results using the more traditional specification of equation (17) can be found in appendix table E7. These are very similar to those given in table 6.

Table 6. Allowing for non-parallel shifts in scale use when adding individual averages (linear specification)

	Current Sat. (1)	Memory of Sat. (2)	Cutoff $\tau_{it,1}$ (3)	Cutoff $\tau_{it,2}$ (4)	Cutoff $\tau_{it,3}$ (5)	Cutoff $\tau_{it,4}$ (6)	Cutoff $\tau_{it,5}$ (7)	Cutoff $\tau_{it,6}$ (8)
ln(HH income)	0.098*** (0.023)	0.028 (0.020)	0.090** (0.038)	-0.013 (0.026)	-0.032* (0.019)	0.016 (0.014)	-0.007 (0.012)	0.016 (0.012)
Employed (base)								
<i>Self-employed</i>	0.080 (0.065)	0.007 (0.062)	0.001 (0.121)	-0.019 (0.095)	0.045 (0.064)	0.005 (0.044)	0.036 (0.038)	0.027 (0.042)
<i>Unemployed</i>	-0.628*** (0.069)	-0.129** (0.064)	-0.254** (0.111)	-0.029 (0.082)	0.045 (0.058)	-0.137*** (0.042)	-0.155*** (0.040)	-0.053 (0.043)
<i>Retired</i>	0.151** (0.072)	-0.081 (0.071)	0.115 (0.121)	0.075 (0.082)	-0.034 (0.059)	-0.018 (0.041)	-0.077** (0.036)	-0.019 (0.036)
<i>Family care</i>	-0.073 (0.057)	-0.111** (0.054)	-0.000 (0.105)	0.027 (0.082)	0.018 (0.057)	-0.045 (0.039)	-0.121*** (0.034)	-0.013 (0.034)
<i>Lt. sick / disabled</i>	-0.549*** (0.079)	-0.298*** (0.075)	-0.131 (0.120)	0.062 (0.081)	-0.027 (0.057)	-0.087** (0.044)	-0.149*** (0.043)	-0.101** (0.047)
<i>In education</i>	0.042 (0.066)	-0.013 (0.058)	0.117 (0.119)	-0.119 (0.090)	-0.023 (0.065)	0.019 (0.046)	-0.103*** (0.038)	0.058 (0.044)
Married (base)								
<i>Widowed</i>	-0.941*** (0.125)	-0.937*** (0.121)	-0.581*** (0.164)	-0.065 (0.082)	-0.024 (0.065)	0.078 (0.053)	0.003 (0.046)	-0.074 (0.052)
<i>Divorced</i>	-0.375*** (0.070)	-0.413*** (0.064)	-0.116 (0.100)	0.026 (0.067)	-0.042 (0.050)	0.075* (0.041)	-0.030 (0.039)	-0.079* (0.047)
<i>Never married</i>	-0.296*** (0.067)	-0.289*** (0.060)	-0.100 (0.110)	0.002 (0.081)	-0.095 (0.060)	0.053 (0.043)	-0.004 (0.036)	-0.055 (0.042)
ln(HH size)	-0.224*** (0.051)	-0.009 (0.047)	-0.106 (0.079)	-0.028 (0.054)	-0.029 (0.038)	0.044 (0.028)	-0.001 (0.024)	-0.033 (0.027)
Childbirth	0.223*** (0.046)	0.087** (0.041)	0.228*** (0.086)	-0.018 (0.067)	-0.017 (0.046)	-0.030 (0.034)	-0.055** (0.026)	0.022 (0.028)
Reference inc.	-0.200 (0.189)	0.241 (0.176)	0.188 (0.305)	0.021 (0.207)	-0.077 (0.133)	-0.057 (0.096)	-0.120 (0.080)	0.030 (0.089)
Constant		-0.500 (0.547)	-2.356*** (0.099)	0.539*** (0.063)	0.532*** (0.042)	0.706*** (0.033)	0.869*** (0.030)	1.018*** (0.039)
Observations	118,440							

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

In both the linear and the nonlinear case, models remain identified when omitting the information from memories, but constraining all elements in δ_1 to 0. Estimates for such models are presented in appendix tables E8 and E9.³³ With respect to latent satisfaction, the results for these models are rather similar to those obtained in column (3) of table 2, i.e. when omitting the estimation of scale shift effects. This is because estimates for β capture the main scale shift effects contained in δ_1 in this specification. Therefore, these models only capture possible distortions in the ‘width’ of the reporting scale, but do not capture shifts in the overall location of the scale.

Moreover, the interpretation of δ_k for $k > 1$ is generally ambiguous. Instead of interpreting these δ_k parameters as indicating whether explanatory variables squeeze or expand the response scale, these parameters can equally well be interpreted as indicating heterogeneities in the effects of explanatory variables across the distribution of reported satisfaction. Indeed, via a change of notation, the linear specification of equation (18), together with equations (1) and (4), and when constraining δ_1 to 0, can be seen to be equivalent to the generalized ordered logit/probit model

³³ Variants of such a test are also used by Bond & Lang (2019). See e.g. section A3.2 of their online appendix.

described by Williams (2006).³⁴ That model explicitly allows for heterogenous effects across the distribution of the response variable. Taking unemployment as an example, the negative coefficient of -0.137 for the 4th cutoff can thus either be interpreted as unemployment bringing the 4th cutoff closer to the third (i.e. unemployment squeezing the response scale), or as indicating that the effect of unemployment is less severe for respondents whose level of life satisfaction is larger than the third response cutoff.

5 Conclusions

I hope to have made three main contributions. First, I showed how estimates of the effects of explanatory variables will be biased if the assumption of *common scale use* is violated. This analysis showed that violations of *common scale use* could in principle lead to sign reversals of the effects of explanatory variables. However, they are unlikely to do so in practice. Instead it is more likely that such violations lead to downward biases in the sizes of effect estimates. My theoretical analysis also showed that lagged scale shift effects may lead to the overestimation of adaptation effects. Second, using variation in respondents' stated memories, I proposed a method by which contemporaneous scale shift effects may be separately identified from effects on latent satisfaction. Third, using the proposed memory approach, I analysed whether intrapersonal scale shift effects can plausibly be observed in British BHPS data. In my baseline analyses, I find, in line with theoretical expectation, that the direction in which explanatory variables affect scale use is typically the same as the direction in which explanatory variables affect latent satisfaction. I find particularly pronounced such effects for unemployment and bereavement. Under plausible assumptions on the upper bound of memory biases, these scale effects turn out to be significant for the latter two variables, but not for my other explanatory variables.

In addition to these contributions, I evaluated Van Praag's (2015) approach of accounting for fixed effects in Appendices C and D. There, I showed that the method outperforms the more classical approach of using OLS to account for fixed effects. In my empirical analyses using BHPS data I further corroborated the viability of that approach. Given the ease at which Van Praag's approach can be implemented, it therefore seems useful to adopt his approach more widely.

The proposed method of using memories may be viewed as a complement to the only other viable method in the literature, i.e. the use of vignettes. Compared to the vignette approach, advantages of this approach are that the assumption of *response consistency* is more likely to be satisfied, that the approach is less costly to administer, and that it can be applied in the context of well-known panel data like BHPS or GSOEP.

However, *present independence* may be less plausible than the assumption of *vignette equivalence*. Future research may address this concern in three ways: First, it would help to obtain data on memories

³⁴ In the probit version of that model, the probability of choosing response category k is given by $\Pr(r_{it} = k | \mathbf{X}_{it}) = \Phi(\omega_k - \mathbf{X}_{it}\boldsymbol{\rho}_k) - \Phi(\omega_{k-1} - \mathbf{X}_{it}\boldsymbol{\rho}_{k-1})$, where we impose $\omega_0 = -\infty$, $\omega_K = \infty$, and $\boldsymbol{\rho}_0 = \mathbf{0}$. This formulation allows the effects of explanatory variables on the latent variable to differ across response categories (hence the subscript k for $\boldsymbol{\rho}_k$). In the linear specification of the model discussed in the present paper, the same probability for $k > 1$ is given by:

$$\Pr(r_{it} = k | \mathbf{X}_{it}) = \Phi\left(\left[\sum_{m=1}^k \gamma_m + \mathbf{X}_{it}\boldsymbol{\delta}_m\right] - \boldsymbol{\beta}\mathbf{X}_{it}\right) - \Phi\left(\left[\sum_{m=1}^{k-1} \gamma_m + \mathbf{X}_{it}\boldsymbol{\delta}_{m-1}\right] - \boldsymbol{\beta}\mathbf{X}_{it}\right)$$

Where we impose $\gamma_0 = -\infty$, $\gamma_K = \infty$, $\gamma_K = \infty$, and $\boldsymbol{\delta}_0 = \boldsymbol{\delta}_1 = \mathbf{0}$. Since we can write $\boldsymbol{\rho}_k = \boldsymbol{\beta} - \sum_{m=1}^k \boldsymbol{\delta}_m$ and $\omega_k = \sum_{m=1}^k \gamma_m$ these formulations are equivalent. This equivalence does not hold in the case of the non-linear specification of equation (17). Nevertheless, the underlying point that we cannot separate heterogenous effects across the distribution of reported happiness from distortions of the response scale remains.

from multiple years in the past. With such data one could implement a version of d’Uva et al.’s (2011) test of vignette equivalence to test *present independence*.³⁵ Second, data in which questions on vignettes and memories are asked simultaneously would allow for a test of whether the two approaches yield equivalent results. Such cross-validation could yield strong evidence in favour of both approaches. Third, it would help to ask respondents about their current life satisfaction, about last year’s (or earlier years’) life satisfaction, and then – crucially – to ask respondents whether they can recall the rating they gave for their contemporaneous life satisfaction last year. Following Howard and Dailey (1979), one would thereby force people to report their past wellbeing with their past scale. Using such data, one could then perform an alternative formal test of *present independence* by observing whether people’s recollection of their past ratings is impacted by contemporaneous values of explanatory variables. Hence, although it was not possible to validate the assumption of *present independence* with the data at hand, more detailed data would allow for such an exercise.

³⁵ In particular, if one had data on memories about multiple years ago, say $q = 1, 2, \dots, Q$ years ago, one could extend equation (9) to read $s_{it}^{(m_q)} = \theta^{(m_q)} + \mathbf{X}_{it-q} \boldsymbol{\alpha}^{(m_q)} + \mathbf{X}_{it} \boldsymbol{\lambda}^{(m_q)} + \varepsilon_{it}^{(m_q)}$ and test if $\boldsymbol{\lambda}^{(m_q)} = \mathbf{0}$ for all $q > 1$.

References

- Adler, Matthew D. 2013. 'Happiness Surveys and Public Policy: What's the Use?' *Duke Law Journal* 62 (8): 1509–1601.
- Ahmed, Sara, Nancy E. Mayo, Sharon Wood-Dauphinee, James A Hanley, and S Robin Cohen. 2004. 'Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke'. *Journal of Clinical Epidemiology* 57 (6): 561–70. <https://doi.org/10.1016/j.jclinepi.2003.11.003>.
- Angelini, Viola, Danilo Cavapozzi, Luca Corazzini, and Omar Paccagnella. 2013. 'Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases'. *Oxford Bulletin of Economics and Statistics* 5. <https://doi.org/10.1111/obes.12039>.
- Bago d'Uva, Teresa, Eddy Van Doorslaer, Maarten Lindeboom, and Owen O'Donnell. 2008. 'Does Reporting Heterogeneity Bias the Measurement of Health Disparities?' *Health Economics* 17 (3): 351–375.
- Bell, Andrew, and Kelvyn Jones. 2015. 'Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data'. *Political Science Research and Methods* 3 (1): 133–153.
- Bertoni, Marco. 2015. 'Hungry Today, Unhappy Tomorrow? Childhood Hunger and Subjective Wellbeing Later in Life'. *Journal of Health Economics* 40 (March): 40–53. <https://doi.org/10.1016/j.jhealeco.2014.12.006>.
- Bertrand, Marianne, and Sendhil Mullainathan. 2001. 'Do People Mean What They Say? Implications for Subjective Survey Data'. *American Economic Review* 91 (2): 67–72.
- Bond, Timothy N, and Kevin Lang. 2018. 'The Sad Truth About Happiness Scales'. *Journal of Political Economy*. <https://doi.org/10.1086/701679>.
- Bonsang, Eric, and Arthur Van Soest. 2012. 'Satisfaction with Job and Income Among Older Individuals Across European Countries'. *Social Indicators Research* 105 (2): 227–54. <https://doi.org/10.1007/s11205-011-9879-5>.
- Borah, Melanie, Carina Keldenich, and Andreas Knabe. 2019. 'Reference Income Effects in the Determination of Equivalence Scales Using Income Satisfaction Data'. *Review of Income and Wealth* 65 (4): 736–70. <https://doi.org/10.1111/roiw.12386>.
- Clark, Andrew E., and Yannis Georgellis. 2013. 'Back to Baseline in Britain: Adaptation in the British Household Panel Survey'. *Economica* 80 (319): 496–512. <https://doi.org/10.1111/ecca.12007>.
- Conway, Michael, and Michael Ross. 1984. 'Getting What You Want by Revising What You Had.' *Journal of Personality and Social Psychology* 47 (4): 738.
- Corrado, Luisa, and Melvyn Weeks. 2010. 'Identification Strategies in Survey Response Using Vignettes'. *CWPE Working Paper* 1031.
- Crettaz, Eric, and Christian Suter. 2013. 'The Impact of Adaptive Preferences on Subjective Indicators: An Analysis of Poverty Indicators'. *Social Indicators Research* 114 (1): 139–152. <https://doi.org/10.1007/s11205-013-0388-6>.
- Danzer, Alexander M., and Natalia Danzer. 2016. 'The Long-Run Consequences of Chernobyl: Evidence on Subjective Well-Being, Mental Health and Welfare'. *Journal of Public Economics* 135 (March): 47–60. <https://doi.org/10.1016/j.jpubeco.2016.01.001>.
- Datta Gupta, Nabanita, Nicolai Kristensen, and Dario Pozzoli. 2010. 'External Validation of the Use of Vignettes in Cross-Country Health Studies'. *Economic Modelling* 27 (4): 854–65. <https://doi.org/10.1016/j.econmod.2009.11.007>.
- 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 (49): 19953–19958.
- Dermer, Marshall, Sidney J. Cohen, Elaine Jacobsen, and Erling A. Anderson. 1979. 'Evaluative Judgments of Aspects of Life as a Function of Vicarious Exposure to Hedonic Extremes.' *Journal of Personality and Social Psychology* 37 (2): 247.

- Diamond, Peter. 2008. 'Behavioral Economics'. *Journal of Public Economics* 92 (8–9): 1858–62. <https://doi.org/10.1016/j.jpubeco.2008.03.003>.
- Diener, Ed, Ronald Inglehart, and Louis Tay. 2013. 'Theory and Validity of Life Satisfaction Scales'. *Social Indicators Research* 112 (3): 497–527. <https://doi.org/10.1007/s11205-012-0076>.
- Dolan, Paul, Georgios Kavetsos, Christian Krekel, Dimitris Mavridis, Robert Metcalfe, Claudia Senik, Stefan Szymanski, and Nicolas R. Ziebarth. 2019. 'Quantifying the Intangible Impact of the Olympics Using Subjective Well-Being Data'. *Journal of Public Economics* 177 (September): 104043. <https://doi.org/10.1016/j.jpubeco.2019.07.002>.
- Dominitz, Jeff, and Charles F. Manski. 1997. 'Using Expectations Data to Study Subjective Income Expectations'. *Journal of the American Statistical Association* 92 (439): 855–67. <https://doi.org/10.1080/01621459.1997.10474041>.
- Falk, Armin, and Markus Knell. 2004. 'Choosing the Joneses: Endogenous Goals and Reference Standards'. *Scandinavian Journal of Economics* 106 (3): 417–435.
- Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–659.
- Fleurbay, Marc, and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press.
- Forgas, Joseph P. 1995. 'Mood and Judgment: The Affect Infusion Model (AIM)'. *Psychological Bulletin* 117 (1): 39.
- Frederick, Shane, and George Loewenstein. 1999. 'Hedonic Adaptation'. In *Well-Being: Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener, and Norbert Schwarz.
- Frijters, Paul, David W. Johnston, and Michael A. Shields. 2011. 'Happiness Dynamics with Quarterly Life Event Data'. *Scandinavian Journal of Economics* 113 (1): 190–211. <https://doi.org/10.1111/j.1467-9442.2010.01638.x>.
- Gibbons, F. X. 1999. 'Social Comparison as a Mediator of Response Shift'. *Social Science & Medicine* 48 (11): 1517–30. [https://doi.org/10.1016/S0277-9536\(99\)00046-5](https://doi.org/10.1016/S0277-9536(99)00046-5).
- Greene, William. 2004. 'The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects'. *The Econometrics Journal* 7 (1): 98–119. <https://doi.org/10.1111/j.1368-423X.2004.00123.x>.
- Greene, William H., and David A. Hensher. 2010. *Modeling Ordered Choices: A Primer and Recent Developments*. Cambridge University Press. <https://papers.ssrn.com/abstract=1213093>.
- Hanglberger, Dominik, and Joachim Merz. 2015. 'Does Self-Employment Really Raise Job Satisfaction? Adaptation and Anticipation Effects on Self-Employment and General Job Changes'. *Journal for Labour Market Research* 48 (4): 287–303.
- Hoogstraten, Joh. 1985. 'Influence of Objective Measures on Self-Reports in a Retrospective Pretest-Posttest Design'. *The Journal of Experimental Education* 53 (4): 207–10. <https://doi.org/10.1080/00220973.1985.10806383>.
- Howard, George S., and Patrick R. Dailey. 1979. 'Response-Shift Bias: A Source of Contamination of Self-Report Measures'. *Journal of Applied Psychology* 64 (2): 144–50.
- Howard, George S., Kenneth M. Ralph, Nancy A. Gulanick, Scott E. Maxwell, Don W. Nance, and Sterling K. Gerber. 1979. 'Internal Invalidity in Pretest-Posttest Self-Report Evaluations and a Re-Evaluation of Retrospective Pretests'. *Applied Psychological Measurement* 3 (1): 1–23.
- Howard, George S., Ronald R. Schmeck, and James H. Bray. 1979. 'Internal Invalidity in Studies Employing Self-Report Instruments: A Suggested Remedy'. *Journal of Educational Measurement*, 129–135.
- Kaiser, Caspar. 2020. 'People Do Not Adapt. New Analyses of the Dynamic Effects of Own and Reference Income on Life Satisfaction'. *Journal of Economic Behavior & Organization* 177: 494–513. <https://doi.org/10.1016/j.jebo.2020.06.003>.
- Kaiser, Caspar, and Maarten CM Vendrik. 2020. 'How Threatening Are Transformations of Happiness Scales to Subjective Wellbeing Research?' *INET Working Paper*.

- <https://www.inet.ox.ac.uk/publications/no-2020-19-how-threatening-are-transformations-of-happiness-scales-to-subjective-wellbeing-research/>.
- Kapteyn, Arie, James P Smith, and Arthur Van Soest. 2007. 'Vignettes and Self-Reports of Work Disability in the United States and the Netherlands'. *The American Economic Review* 97 (1): 461–73.
- Kapteyn, Arie, James P. Smith, and Arthur Van Soest. 2010. 'Life Satisfaction'. In *International Differences in Wellbeing*, edited by Ed Diener, John F. Helliwell, and Daniel Kahneman. Oxford University Press.
- King, Gary, Christopher J L Murray, Joshua A Salomon, Ajay Tandon, Gary King, Christopher J L Murray, World Health, and Joshua A Salomon. 2004. 'Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research'. *The American Political Science Review* 98 (1): 191–207.
- Klaassen, Tineke, W. J. Riedel, N. E. P. Deutz, and H. M. Van Praag. 2002. 'Mood Congruent Memory Bias Induced by Tryptophan Depletion'. *Psychological Medicine* 32 (1): 167.
- Kristensen, Nicolai, and Edvard Johannson. 2008. 'New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes'. *Labour Economics* 15 (1): 96–117. <https://doi.org/10.1016/j.labeco.2006.11.001>.
- Levine, Linda J., and Susan Bluck. 2004. 'How Emotions Fade: Valence, Appraisals, and the Emotional Impact of Remembered Events.' In *Advances in Psychology Research*, edited by Serge P. Shohov, 30:3–20. Nova Science Publishers.
- Levy, Horacio, and Stephen P. Jenkins. 2012. 'Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-18'. *UK Data Archive Study Number 3909*.
- Lewis, P. A., H. D. Critchley, A. P. Smith, and R. J. Dolan. 2005. 'Brain Mechanisms for Mood Congruent Memory Facilitation'. *NeuroImage* 25 (4): 1214–23. <https://doi.org/10.1016/j.neuroimage.2004.11.053>.
- Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.
- Mayer, John D., Yvonne N. Gaschke, Debra L. Braverman, and Temperance W. Evans. 1992. 'Mood-Congruent Judgment Is a General Effect.' *Journal of Personality and Social Psychology* 63 (1): 119.
- Montgomery, Mallory. 2017. 'Reversing the Gender Gap in Happiness: Validating the Use of Life Satisfaction Self-Reports Worldwide'. *Working Paper*.
- Mundlak, Yair. 1978. 'On the Pooling of Time Series and Cross Section Data'. *Econometrica*, 69–85.
- Murray, Christopher JL, Emre Ozaltin, Ajay Tandon, Joshua Salomon, Ritu Sadana, and Somnath Chatterji. 2003. 'Empirical Evaluation of the Anchoring Vignette Approach in Health Surveys'. *Health Systems Performance Assessment: Debates, Methods and Empiricism*, 369–99.
- Oswald, Andrew J. 2008. 'On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings'. *Economics Letters* 100 (3): 369–72. <https://doi.org/10.1016/j.econlet.2008.02.032>.
- Plant, Michael. 2020. 'A Happy Possibility About Happiness (And Other) Scales: Understanding Why the Cardinality Assumption Is Both Defensible and Unnecessary'. *Unpublished Working Paper*.
- Prati, Alberto. 2017. 'Hedonic Recall Bias. Why You Should Not Ask People How Much They Earn'. *Journal of Economic Behavior & Organization* 143: 78–97.
- Prati, Alberto, and Claudia Senik. 2020. 'Feeling Good or Feeling Better?' *PSE Working Papers*, no. 17. <https://econpapers.repec.org/paper/halpsewpa/halshs-02545228.htm>.
- Ravallion, Martin, Kristen Himelein, and Kathleen Beegle. 2016. 'Can Subjective Questions on Economic Welfare Be Trusted?' *Economic Development and Cultural Change* 64 (4): 697–726.
- Riedl, Maximilian, and Ingo Geishecker. 2014. 'Keep It Simple: Estimation Strategies for Ordered Response Models with Fixed Effects'. *Journal of Applied Statistics* 41 (11): 2358–74. <https://doi.org/10.1080/02664763.2014.909969>.
- Schröder, Carsten, and Shlomo Yitzhaki. 2017. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. <https://doi.org/10.1016/j.euroecorev.2016.12.011>.

- Schwarz, Norbert, and Fritz Strack. 1999. 'Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications'. In *Well-Being: The Foundations of Hedonic Psychology*, 7:61–84.
- Sprangers, Mirjam A. G., and Carolyn E. Schwartz. 1999. 'Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model'. *Social Science & Medicine* 48 (11): 1507–15. [https://doi.org/10.1016/S0277-9536\(99\)00045-3](https://doi.org/10.1016/S0277-9536(99)00045-3).
- Sprangers, Mirjam A. G., Frits S. A. Van Dam, Jenny Broersen, Litanja Lodder, Lidwina Wever, Mechteld R. M. Visser, Paul Oosterveld, and Ellen M.A. Smets. 1999. 'Revealing Response Shift in Longitudinal Research on Fatigue: The Use of the Thentest Approach'. *Acta Oncologica* 38 (6): 709–18. <https://doi.org/10.1080/028418699432860>.
- Steffel, Mary, and Daniel M. Oppenheimer. 2009. 'Happy by What Standard? The Role of Interpersonal and Intrapersonal Comparisons in Ratings of Happiness'. *Social Indicators Research* 92 (1): 69–79. <https://doi.org/10.1007/s11205-008-9289-5>.
- Townsend, Michael, and Keri Wilton. 2003. 'Evaluating Change in Attitude towards Mathematics Using the "Then-Now" Procedure in a Cooperative Learning Programme'. *British Journal of Educational Psychology* 73 (4): 473–87. <https://doi.org/10.1348/000709903322591190>.
- Urban, Emily J., Susan T. Charles, Linda J. Levine, and David M. Almeida. 2018. 'Depression History and Memory Bias for Specific Daily Emotions'. *PLoS ONE* 13 (9). <https://doi.org/10.1371/journal.pone.0203574>.
- Uva, Teresa Bago d', Maarten Lindeboom, Owen O'Donnell, and Eddy Van Doorslaer. 2011. 'Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity'. *Journal of Human Resources* 46 (4): 875–906.
- Van Praag, Bernard M. S. 1971. 'The Welfare Function of Income in Belgium: An Empirical Investigation'. *European Economic Review* 2 (3): 337–69.
- . 1993. 'The Relativity of the Welfare Concept'. In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen, 362–85. Oxford University Press. <https://doi.org/10.1093/0198287976.003.0027>.
- . 2015. 'A New View on Panel Econometrics: Is Probit Feasible After All?' *IZA Discussion Paper*, no. 9345.
- Van Praag, Bernard M. S., and Ada Ferrer-i-Carbonell. 2008. *Happiness Quantified: A Satisfaction Calculus Approach*. 2nd ed. Oxford University Press.
- Van Praag, Bernard M. S., and Paul Frijters. 1999. 'The Measurement of Welfare and Well-Being: The Leyden Approach'. In *Well-Being: Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener, and Norbert Schwarz, 302.
- Vendrik, Maarten C. M. 2013. 'Adaptation, Anticipation and Social Interaction in Happiness: An Integrated Error-Correction Approach'. *Journal of Public Economics* 105: 131–149. <https://doi.org/10.1016/j.jpubeco.2013.06.009>.
- Viscusi, W. Kip. 2020. 'Wellbeing Measures of Mortality Risks: Life-Cycle Contradictions and Ordinal Index Challenges'. *Behavioural Public Policy* 4 (2): 245–53. <https://doi.org/10.1017/bpp.2019.47>.
- Williams, Richard. 2006. 'Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables'. *Stata Journal* 6 (1): 58–82. <https://doi.org/st0097>.
- Wilson, Timothy D., and Daniel T. Gilbert. 2005. 'Affective Forecasting'. *Current Directions in Psychological Science* 14 (3): 131.
- Wilson, Timothy D., Jay Meyers, and Daniel T. Gilbert. 2003. "'How Happy Was I, Anyway?": A Retrospective Impact Bias'. *Social Cognition* 21 (6): 421–46. <https://doi.org/10.1521/soco.21.6.421.28688>.
- Xu, Hongwei, and Yu Xie. 2016. 'Assessing the Effectiveness of Anchoring Vignettes in Bias Reduction for Socioeconomic Disparities in Self-Rated Health among Chinese Adults'. *Sociological Methodology* 46 (1): 84–120. <https://doi.org/10.1177/0081175015599808>.

Appendix

A Interpretation of absolute magnitudes of coefficients

Life satisfaction, like most other subjective constructs, has no natural cardinal scale, with no well-defined physical phenomena to ground it and no natural zero point. Despite this feature of subjective constructs, some interpretations of the absolute magnitudes of coefficients may nevertheless be both useful and meaningful.

It might be natural to interpret the magnitudes of coefficients in relation to how a (marginal) change in covariates changes the probability of exceeding some particular response level. Given the model described by equations (1), (4) and (5), the probability of responding category $r > k$ conditional on \mathbf{X}_{it*} is given by $\Pr(r_{it} > k | \mathbf{X}_{it*}) = 1 - \Phi(\gamma_k + \mathbf{X}_{it*}\boldsymbol{\delta}_* - \mathbf{X}_{it*}\boldsymbol{\beta}_*) = \Phi(\mathbf{X}_{it*}(\boldsymbol{\beta}_* - \boldsymbol{\delta}_*) - \gamma_k)$. Taking the partial derivative with respect to some particular variable, say $X_{j,it}$, we get $\partial \Pr(r_{it} > k | \mathbf{X}_{it*}) / \partial X_{j,it} = \varphi(\mathbf{X}_{it*}(\boldsymbol{\beta}_* - \boldsymbol{\delta}_*) - \gamma_k)(\beta_{j,0} - \delta_{j,0})$, where $\varphi(\cdot)$ denotes the standard normal p.d.f. As explained in the main text, an analysis which does not account for shifts in scale use already estimates $\boldsymbol{\beta}_* - \boldsymbol{\delta}_*$. Therefore, if we are only interested in estimating expected effects on response probabilities, accounting for changes in scale use would serve no purpose.

One could average $\varphi(\mathbf{X}_{it*-j}(\boldsymbol{\beta}_{*-j} - \boldsymbol{\delta}_{*-j}) + X_{j,it}\beta_{j,0} - \gamma_k)\beta_{j,0}$ over all observed combinations of \mathbf{X}_{it*} . Here, \mathbf{X}_{it*-j} is the same as \mathbf{X}_{it*} but has element $X_{j,it}$ removed. Vectors $\boldsymbol{\beta}_{*-j}$ and $\boldsymbol{\delta}_{*-j}$ are defined analogously. Such an exercise can be interpreted as estimating the average marginal effect of $X_{j,it}$ on the probability of responding category $r > k$ conditional on \mathbf{X}_{it*} under a counterfactual in which $X_{j,it}$ did not also affect scale use. However, this is cumbersome to compute and not a very intuitive interpretation of the absolute magnitude.

In the context of OLS regressions, one oft-used intuitive interpretation is to discuss by how many categories a respondent can be expected to change her response given a unit change in an explanatory variable's value. This is very straightforward in this context and given by $E(r_{it} | X_{j,it} + 1; \mathbf{X}_{it*-j}) - E(r_{it} | X_{j,it}; \mathbf{X}_{it*-j}) = \beta_{j,0}$.

However, in the ordered probit context, differences between thresholds are not equal. Nevertheless, an analogue may usefully indicate how much latent satisfaction changes for a unit change in $X_{j,it}$ as a fraction of the normalized width of the satisfaction scale. Here, the natural normalization is over the number of estimated thresholds. The width of the scale can be given by the difference between the highest and lowest estimated threshold.

Formally this is given by:

$$\begin{aligned} & [E(s_{it}/(\tau_{K-1} - \tau_1) | X_{j,it} + 1; \mathbf{X}_{it*-j}) - E(s_{it}/(\tau_{K-1} - \tau_1) | X_{j,it}; \mathbf{X}_{it*-j})](K - 1) \\ & = \beta_{j,0}(K - 1)/(\gamma_{K-1} - \gamma_1) \end{aligned} \quad (\text{A1})$$

Thus, the effect of $X_{j,it}$ on scale use in this expression cancels in the case of parallel shifts. Fortunately, this expression is directly obtained from our parameter estimates. We may hence use this as an intuitive interpretation of the absolute magnitudes of effects on latent satisfaction.

B Using vignettes to correct for violations of common scale use

The vignette approach tackles violations of common scale use by presenting respondents with short descriptions of imaginary persons' lives (King et al., 2004). An example from the SHARE survey for such a vignette is the following:

John is 63 years old. His wife died 2 years ago, and he still spends a lot of time thinking about her. He has four children and ten grandchildren who visit him regularly. John can make ends meet but has no money for extras such as expensive gifts to his grandchildren. He has had to stop working recently due to heart problems. He gets tired easily. Otherwise, he has no serious health conditions. How satisfied with his life do you think John is?'

Respondents are then asked to rate the life satisfaction of these imaginary persons. On the assumption that all respondents perceive the described person's satisfaction equally, identification of differences in scale use become possible.

No study to date used vignettes to analyse the separate effects of lags and leads of explanatory variables on both latent life satisfaction and scale use. To ease the exposition, I will therefore assume that latent life satisfaction and scale use only depend on contemporaneous values of covariates. Parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ (without subscripts) are thus only associated with contemporaneous values of covariates. If a panel dataset containing vignettes existed, an analysis of the effect of lags and leads would pose no special problem. Unfortunately, I am not aware of such data. This partly motivates the approach proposed in section 3.

B.1 Formal approach

Assume that, from a set of V vignettes, respondents give a report $r_{it}^{(v)}$ about a particular vignette v on the basis of a perceived satisfaction level $s_{it}^{(v)}$ of that vignette. Also assume that respondents use the same response scale for their own satisfaction as for rating the vignettes. This yields a reporting rule analogous to equation (A1):

$$r_{it}^{(v)} = k \leftrightarrow \tau_{it,k-1} < s_{it}^{(v)} \leq \tau_{it,k} \quad (\text{A2})$$

Since the same thresholds are used in equations (1) and (A2), we assume *response consistency*. Further assume that the perceived level of $s_{it}^{(v)}$ does not depend on \mathbf{X}_{it} . This assumption is known as the *vignette equivalence* assumption. Now consider:

$$s_{it}^{(v)} = \theta^{(v)} + \mathbf{X}_{it}\boldsymbol{\lambda}^{(v)} + \varepsilon_{it}^{(v)} \quad (\text{A3})$$

Here, $\theta^{(v)}$ is a vignette-specific parameter and $\varepsilon_{it}^{(v)}$ is a normally distributed error term with mean 0 and standard deviation 1. *Vignette equivalence* entails that $\boldsymbol{\lambda}^{(v)} = \mathbf{0}$. Incorporating these additional equations, the individual contributions to the likelihood then become:

$$\begin{aligned} \ell_{it} = & \ln \sum_{k=1}^K \mathbf{I}(r_{it} = k) [\Phi(\gamma_k + \mathbf{X}_{it}\boldsymbol{\delta} - \mathbf{X}_{it}\boldsymbol{\beta}) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}\boldsymbol{\delta} - \mathbf{X}_{it}\boldsymbol{\beta})] \\ & + \sum_v \sum_{k=1}^K \mathbf{I}(r_{it}^{(v)} = k) [\Phi(\gamma_k + \mathbf{X}_{it}\boldsymbol{\delta} - \theta^{(v)} - \mathbf{X}_{it}\boldsymbol{\lambda}^{(v)}) - \Phi(\gamma_{k-1} + \mathbf{X}_{it}\boldsymbol{\delta} - \theta^{(v)} - \mathbf{X}_{it}\boldsymbol{\lambda}^{(v)})] \end{aligned} \quad (\text{A4})$$

When assuming $\boldsymbol{\lambda}^{(v)} = \mathbf{0}$, $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are separately identified from maximising equation (A4).

B.2 Plausibility of *vignette equivalence* and *response consistency*

The vignette approach exchanges the assumption of *common scale use* with the assumptions of *response consistency* and *vignette equivalence*. Regarding *response consistency*, and extending the worries of section 2.2, respondents might rate vignettes relative to reference points that vary with characteristics given in the vignette: Persons described to be richer might be rated with a higher reference point appropriate for richer persons, and persons described as poor might be rated less stringently. Unfortunately, statistical tests of *response consistency* rely on having a further objective indicator for the dependent variable. In the case of entirely subjective variables, no objective indicators can exist, making such tests impossible. However, in cases where *response consistency* with respect to health was tested, *response consistency* was rejected (d’Uva et al., 2011; Datta Gupta et al., 2010).

Regarding *vignette equivalence*, some studies test whether respondents give a common ordering to multiple vignettes (Angelini et al., 2013; d’Uva et al., 2011; Montgomery, 2017; Murray et al., 2003). Although this test is typically passed, failure to reject does not imply that vignette equivalence holds. D’Uva et al. (2011) propose a stronger test, noting that vignette equivalence is required for only one vignette, say for $v = 1$. For all other vignettes, *vignette equivalence* can be relaxed, thus allowing a test of $\lambda^{(v)} = \mathbf{0}$ in $s_{it}^{(v)} = \theta^{(v)} + \mathbf{X}_{it}\lambda^{(v)} + \varepsilon_{it}^{(v)}$ for $v = 2, \dots, V$. They find that vignette equivalence is rejected for almost all considered covariates.³⁶

Vignettes are incomplete descriptions of a person’s life. Some aspects relevant to life satisfaction must therefore be imagined by the respondent. For example, in the case of the vignette quoted above, respondents must make some assumptions about what it means to make ‘ends meet’. It is plausible that a person’s current circumstances (i.e. values of \mathbf{X}_{it}), will impact how such assumptions are made. In particular, for a respondent with low incomes or who is unemployed, making ‘ends meet’ likely means a lower level of economic welfare compared to respondents in stable employment or for those with high incomes (Van Praag and Frijters 1999; Van Praag 1971; Crettaz and Suter 2013). Thus, while income and being employed are positively associated with latent satisfaction, they will also be positively associated with respondents’ perceptions of vignettes. This implies that $\lambda_j^{(v)}$ and β_j have the same sign for such variables. As shown in section 3.3, estimates of β_j are then biased towards zero. Moreover, individuals tend to make mood-congruent judgements in both self- and non-self-relevant domains (Mayer et al. 1992; Forgas 1995).³⁷ This means that respondents with more positive affect can be expected to judge others more positively, too. Since affect and life satisfaction are driven by similar variables, this also suggests that $\lambda_j^{(v)}$ and β_j have the same sign.

Hence, vignette equivalence might fail. The direction in which it is likely to fail suggests that magnitudes of β will be biased towards zero.

³⁶ Rejection of this test can be interpreted as indicating that there are differences across vignettes in how perceptions of vignettes’ satisfaction depend on covariates. Passing this test does not imply that vignette equivalence holds. Instead, failure to pass indicates that vignette equivalence does not hold for at least one vignette.

³⁷ For example, in study 3 of Mayer et al. (1992), respondents in a representative sample of the population of New Hampshire (USA) were asked to rate both their mood, as well as to make a number of other-regarding judgements. Those judgements included assessments of e.g. ‘the likelihood of a marriage resulting in long-term happiness for both members of the couple?’, ‘the likelihood that there will be an atomic war in the next five years’, or ‘the likelihood that the average 30-year-old in New Hampshire will be involved in a happy, loving romance?’. For all such cases, those with a better mood provided significantly higher probabilities and thus made mood-congruent judgements.

B.3 Previous results using vignettes

The vignettes methodology was previously applied to investigate e.g. variation in the reporting of subjective health (Bago d’Uva et al. 2008; Xu and Xie 2016), work disability (Kapteyn, Smith, and Van Soest 2007), subjective poverty (Ravallion, Himelein, and Beegle 2016), as well as income and job satisfaction (Kristensen and Johansson 2008; Bonsang and Van Soest 2012).

Five studies applied the vignette approach to studying life satisfaction. Kapteyn et al. (2010) use Dutch ‘CenterPanel’ data from 2006 and American ‘Rand American Life’ data from 2006-2007. For the USA, they find that accounting for scale use increases the estimated effects of income and household size (in the USA). In the Dutch sample, the effects of being employed, and being married are increased in size. Montgomery (2017) uses global Gallup World Poll data for the years 2011-2014. She shows that the relative gender gap in life satisfaction (with women being more satisfied) is driven by differential scale use. Furthermore, compared to being married, accounting for differential scale use marginally increases the magnitudes of the negative effects of being single, separation, divorce, and widowhood. The effects of unemployment and log income are also marginally increased. However, among two sets of vignettes available to her, one of these sets fails D’Uva et al.’s (2011) test of *vignette equivalence*. Angelini et al. (2013) use European SHARE data. They find large differences in scale use across European countries. Moreover, compared to having a partner, they also find slightly increased effects for being single and widowhood. Increased effects are also observed for not working and retirement. Finally, Bertoni (2015) and Corrado and Weeks (2010) use the same SHARE dataset. Bertoni (2015) finds that childhood deprivation is associated with less stringent scale use.

Apart from the case of gender investigated by Montgomery (2017), whenever there was a change in estimated coefficients due to accounting for variation in scale use, estimated magnitudes on latent satisfaction increased. This is despite the possible bias toward zero discussed above.

C Van Praag's (2015) argument

Van Praag's argument may be stated as follows. First, extend equation (4) by writing $\varepsilon_{it} = u_i + v_{it}$. This yields:

$$s_{it} = \mathbf{X}_{it*}\boldsymbol{\beta}_* + u_i + v_{it} \quad (\text{A5})$$

The fixed effect u_i may be correlated with \mathbf{X}_{it*} . Assume that v_{it} , after purging it of the individual fixed effect, is normally distributed with constant variance and uncorrelated with \mathbf{X}_{it*} .

Now stack observations over the T_i periods for each individual i :

$$\mathbf{s}_i = \mathbf{X}_i\boldsymbol{\beta}_* + \mathbf{u}_i + \mathbf{v}_i. \quad (\text{A6})$$

Here, $\mathbf{s}_i = (s_{i1}, \dots, s_{iT_i})'$ and analogously for $\mathbf{X}_i, \mathbf{u}_i$, and \mathbf{v}_i . Now consider the matrix $\mathbf{M}_i = \frac{1}{T_i}\mathbf{1}\mathbf{1}'$, where $\mathbf{1}$ is a vector of T_i ones. For any matrix \mathbf{A}_i , define $\mathbf{M}_i\mathbf{A}_i \equiv \bar{\mathbf{A}}_i$. Each column of $\bar{\mathbf{A}}_i$ contains the average of each column of \mathbf{A}_i . The demeaned matrix $\check{\mathbf{A}}_i$ is then obtained as $\check{\mathbf{A}}_i = \mathbf{A}_i - \mathbf{M}_i\mathbf{A}_i$. Equation (A6) is a constrained version of:

$$\mathbf{s}_i = \check{\mathbf{X}}_i\check{\boldsymbol{\beta}}_* + \bar{\mathbf{X}}_i\bar{\boldsymbol{\beta}}_* + \mathbf{u}_i + \mathbf{v}_i. \quad (\text{A7})$$

Equation (A7) can be recovered from (A6) when $\check{\boldsymbol{\beta}}_* = \bar{\boldsymbol{\beta}}_*$. Since \mathbf{M}_i is idempotent, we find that $\check{\mathbf{X}}_i'\mathbf{u}_i = \mathbf{0}$.³⁸ Thus, $\check{\mathbf{X}}_i$ and \mathbf{u}_i are uncorrelated, and the exclusion or inclusion of \mathbf{u}_i does not affect estimates of $\check{\boldsymbol{\beta}}_*$. Also note that $\check{\mathbf{X}}_i'\mathbf{v}_i = \check{\mathbf{X}}_i'\check{\mathbf{v}}_i$,³⁹ where $\check{\mathbf{v}}_i$ is obtained by demeaning equation (A5), as is normally done when accounting for fixed effects. Therefore, even when omitting \mathbf{u}_i , OLS estimates of $\check{\boldsymbol{\beta}}_*$ obtained from estimating equation (A7), would be identical to OLS estimates obtained by demeaning equation (A5). Also see Bell and Jones (2015) for an extended treatment of this point.

Although van Praag (2015) does not take this step, we can rewrite equation (A7) as:

$$\mathbf{s}_i = \mathbf{X}_i\check{\boldsymbol{\beta}}_* + \bar{\mathbf{X}}_i\bar{\boldsymbol{\beta}}_* + \mathbf{u}_i + \mathbf{v}_i \quad (\text{A8})$$

Here, $\bar{\boldsymbol{\beta}}_* - \check{\boldsymbol{\beta}}_* \equiv \tilde{\boldsymbol{\beta}}_*$. This yields the specification proposed by Mundlak (1978).⁴⁰ Van Praag (2015) makes a further claim. When \mathbf{s}_i is not observed, and only \mathbf{r}_i is, an estimation of equation (A7) by ordered probit will also yield a consistent estimate of $\check{\boldsymbol{\beta}}_*$ when omitting \mathbf{u}_i .

However, estimates of $\check{\boldsymbol{\beta}}_*$ will only be consistent up to a proportionality factor. Since that factor is constant across all elements in $\check{\boldsymbol{\beta}}_*$, estimation should nevertheless yield consistent estimates of the signs, ratios of coefficients, and normalized magnitudes (c.f. appendix A).⁴¹

³⁸ A matrix \mathbf{A} is idempotent iff $\mathbf{A}\mathbf{A} = \mathbf{A}$. We can therefore write: $\check{\mathbf{X}}_i'\mathbf{u}_i = \check{\mathbf{X}}_i'\mathbf{M}_i\mathbf{u}_i = ((\mathbf{I}_i - \mathbf{M}_i)\mathbf{X}_i)'\mathbf{M}_i\mathbf{u}_i = (\mathbf{X}_i'(\mathbf{I}_i - \mathbf{M}_i))'\mathbf{M}_i\mathbf{u}_i = (\mathbf{X}_i' - \mathbf{X}_i'\mathbf{M}_i)\mathbf{M}_i\mathbf{u}_i = \mathbf{X}_i'\mathbf{M}_i\mathbf{u}_i - \mathbf{X}_i'\mathbf{M}_i\mathbf{M}_i\mathbf{u}_i = \mathbf{0}$.

³⁹ Because: $\check{\mathbf{X}}_i'\mathbf{v}_i = \check{\mathbf{X}}_i'(\check{\mathbf{v}}_i + \bar{\mathbf{v}}_i) = \check{\mathbf{X}}_i'\check{\mathbf{v}}_i + \check{\mathbf{X}}_i'\bar{\mathbf{v}}_i = \check{\mathbf{X}}_i'\check{\mathbf{v}}_i + (\mathbf{X}_i' - \mathbf{X}_i'\mathbf{M}_i)\bar{\mathbf{v}}_i = \check{\mathbf{X}}_i'\check{\mathbf{v}}_i + \mathbf{X}_i'\bar{\mathbf{v}}_i - \mathbf{X}_i'\mathbf{M}_i\bar{\mathbf{v}}_i = \check{\mathbf{X}}_i'\check{\mathbf{v}}_i$.

⁴⁰ Wooldridge (2002) considers Chamberlain's random effects estimator in this context. That estimator is largely identical to the approach of Van Praag (2015). It merely adds a further random effect that is assumed to be uncorrelated with \mathbf{X}_{it*} (when conditioning on $\bar{\mathbf{X}}_i$). Since this addition is more complicated to implement and only presents an efficiency gain, I do not pursue this approach here.

⁴¹ The proportionality factor depends on our choice of the assumed standard deviation σ and the true standard deviation σ^* of ε_{it} . In ordered probit we consistently estimate $\check{\boldsymbol{\beta}}_*\sigma/\sigma^*$ (Wooldridge 2002). The value of σ^* positively

D Monte Carlo study

This small Monte Carlo study evaluates how well the proposed approach of using memories works at identifying variables' separate effects on latent satisfaction and scale use in the presence of fixed effects.

D.1 Experimental design

The set-up is as follows. I set $N = 5,000$ and $T = 6$. With respect to T , this is close to the BHPS data used in the main text. N is somewhat smaller than what is available in BHPS to reduce computing time. Latent satisfaction is generated by:

$$s_{it} = \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + u_i + v_{it}, \quad (\text{A9})$$

where $u_i \sim N(0,1)$ and $v_{it} \sim N(0,1)$. Furthermore, $X_{1,it} = X_{1,it-1} + e_{1,it}$ for $t > 1$ and $X_{1,it} = u_i$ for $t = 1$, with $e_{1,it} \sim N(0,1)$. $X_{2,it} \sim N(u_i, 1)$. Finally, $X_{3,it} = 1$ if $u_i + e_{2,it} > 0$ and $X_{3,it} = 0$ otherwise, with $e_{2,it} \sim N(0,1)$. Thus, each independent variable is correlated with the fixed effect u_i and is strongly autocorrelated (as would be the case for most demographic variables). $X_{1,it}$ and $X_{2,it}$ are continuous, while $X_{3,it}$ is a dummy. I further set $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5$. I evaluate two different processes by which the cutoffs $\tau_{it,k}$ are generated to determine r_{it} (which in turn is generated as in equation (1)). As a first case, I generate $\tau_{it,k} = -7 + \bar{s} + 2k$, where \bar{s} is the average of s_{it} across i and t . As a second case, I generate $\tau_{it,k} = -2.75 + \delta_1 X_{1,it} + \delta_2 X_{2,it} + \delta_3 X_{3,it} + \bar{s} + 1.25k$. With $\delta_1 = 0.2, \delta_2 = 0.8, \delta_3 = 0.8$. In both cases, the number of response categories K is set to 7. Both cases yield equidistant cutoffs and roughly even shares across each of the seven response categories for r_{it} . However, in the first case all cutoffs are independent of \mathbf{X}_{it} , while in the latter case cutoffs strongly depend on \mathbf{X}_{it} . Memories of satisfaction are generated by $s_{it}^{(m)} = \alpha_1 X_{1,it-1} + \alpha_2 X_{2,it-1} + \alpha_3 X_{3,it-1} + u_i + v_{it}^{(m)}$, with $\alpha_1 = 0.8, \alpha_2 = 0.4, \alpha_3 = 0.6$ and $v_{it}^{(m)} \sim N(0,1)$.

I draw 250 independent sets of observations and estimate six different models on each set. Those models are:

- An OLS regression of r_{it} ('OLS'),
- An OLS regression of demeaned r_{it} ('OLS-FE'),
- A standard ordered probit regression ('OP'),
- An ordered probit regression with individual averages for each independent variable added (as discussed in Appendix C; 'OP-Avg.')
- An ordered probit regression with individual averages, extended by an equation for memories of past satisfaction (as discussed in section 3, 'MOP-Avg.').
- An ordered probit regression with individual averages, extended by an equation for memories of past satisfaction and using data for memories as available for the UK (as discussed in section 3; 'UKMOP-Avg.')

depends on $\text{Var}(v_{it})$ and $\text{Var}(u_i)$, but negatively on $\text{Cov}(\bar{X}_{it}, u_i)$. When $\text{Cov}(\bar{X}_{it}, u_i) = 0$, then $\text{Var}(\epsilon_{it}) = \text{Var}(v_{it}) + \text{Var}(u_i)$. Thus, $\text{SD}(v_{it}) \leq \sigma^* \leq \sqrt{\text{Var}(v_{it}) + \text{Var}(u_i)}$. If we choose $\sigma = 1$ and $\text{SD}(v_{it}) = 1$, we will get estimates with negatively biased magnitudes by a factor $1/\sigma^*$. Estimates of the thresholds will be biased by the same factor $1/\sigma^*$ as the coefficients on latent satisfaction. Hence magnitudes of coefficients that are standardized by the width of the response scale (see appendix A) are not affected by the proportionality factor.

For both MOP-Avg. and the UKMOP-Avg. model, I add individual averages to the memory and the cutoff equations as well as the equation on latent satisfaction.

D.2 Results in case of cutoffs being independent of X_{it}

Table D1 presents mean estimates of coefficients and their ratios from 250 independent trials.

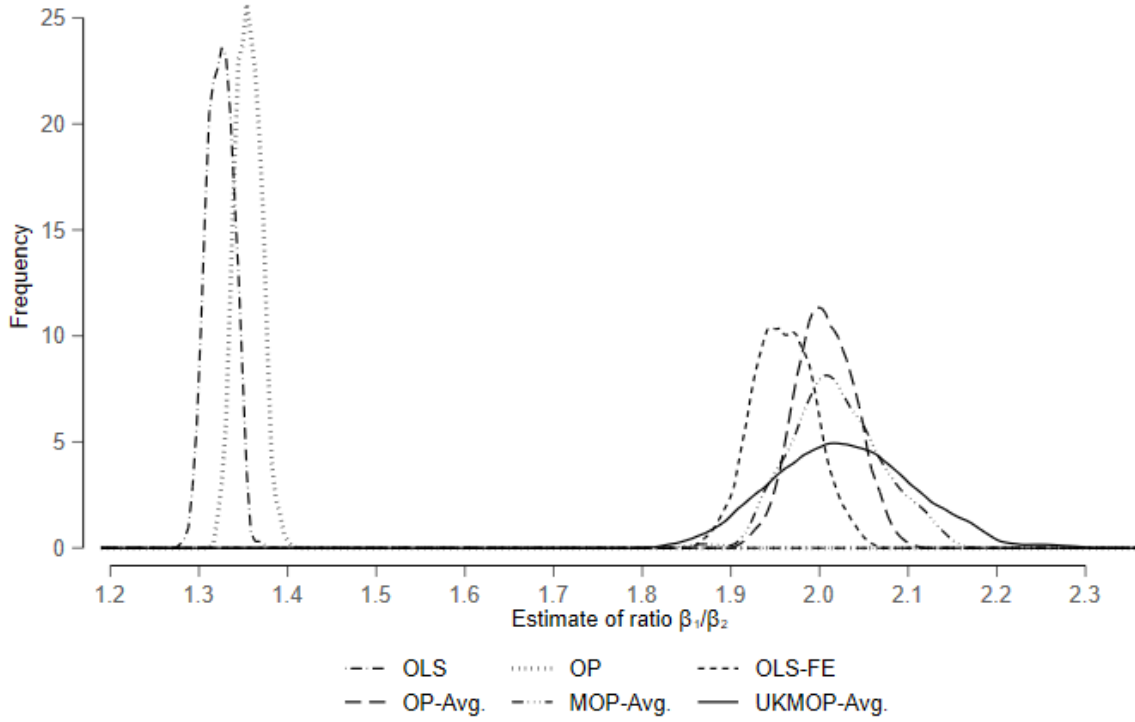
Table D1. Results from Monte Carlo simulation with cutoffs independent of X_{it}

Parameter estimates	(1) OLS	(2) OP	(3) OLS-FE	(4) OP-Avg.	(5) MOP-Avg.	(6) UKMOP-Avg.
β_1	0.526	0.988	0.459	0.958	0.964	0.938
β_2	0.398	0.729	0.234	0.478	0.478	0.463
β_3	0.544	0.930	0.245	0.479	0.480	0.466
β_1/β_2	1.324 (0.676)	1.355 (0.645)	1.960 (0.052)	2.007 (0.033)	2.019 (0.053)	2.027 (0.081)
β_1/β_3	0.968 (1.032)	1.063 (.938)	1.875 (0.142)	2.001 (0.069)	2.015 (0.116)	2.029 (0.183)
β_2/β_3	0.731 (0.269)	0.784 (0.216)	0.957 (0.055)	0.998 (0.036)	0.998 (0.059)	1.002 (0.094)
Norm. β_1	0.526 (0.074)	0.687 (0.088)	0.459 (0.141)	0.602 (0.005)	0.606 (0.010)	0.560 (0.042)
Norm. β_2	0.398 (0.098)	0.507 (0.207)	0.234 (0.066)	0.300 (0.004)	0.300 (0.007)	0.276 (0.025)
Norm. β_3	0.544 (0.244)	0.647 (0.347)	0.245 (0.056)	0.301 (0.010)	0.302 (0.017)	0.278 (0.033)
α_1					0.783	0.767
α_2					0.383	0.376
α_3					0.576	0.559
α_1/α_2					2.043 (0.064)	2.045 (0.102)
α_1/α_3					1.361 (0.059)	1.377 (0.099)
α_2/α_3					0.666 (0.027)	0.675 (0.052)
δ_1					0.030	-0.043
δ_2					0.000	-0.039
δ_3					0.001	-0.038
δ_1/δ_2					-4.724	-56.582
δ_1/δ_3					-0.434	-0.292
δ_2/δ_3					0.275	-0.147

Note: $T = 6$ and $N = 5000$. Numbers show mean estimates of coefficients and their ratios from 250 independent trials. Root mean squared error (RMSE) shown in parentheses. True values for each parameter are: $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\alpha_1 = 0.8$, $\alpha_2 = 0.4$, $\alpha_3 = 0.6$, $\delta_1 = 0$, $\delta_2 = 0$, $\delta_3 = 0$. For ratios of δ the RMSE is not defined due to division by zero. For OP regressions ‘Norm. β ’ give normalized coefficients, which are given by $(K - 1)\beta/(\gamma_{K-1} - \gamma_1) = 6\beta/(\gamma_6 - \gamma_1)$. For OLS regressions, ‘Norm. β ’ is identical to β since r is rank-order coded. True values are Norm. $\beta_1 = 0.6$, Norm. $\beta_2 =$ Norm. $\beta_3 = 0.3$.

Absolute magnitudes of coefficients are affected by arbitrary scaling choices for σ (also see Van Praag and Ferrer-i-Carbonell 2008). As motivated in Appendix A, I therefore present estimates of normalized coefficients on latent satisfaction. These are given by Norm. $\beta = 6\beta/(\gamma_6 - \gamma_1)$. True values are Norm. $\beta_1 = 0.6$ and Norm. $\beta_2 =$ Norm. $\beta_3 = 0.3$. With respect to these targets, the OLS and OP models perform worst overall. However, interestingly, the OLS-FE model is furthest from the target value for the normalized coefficient for β_1 . This is despite the fact cutoffs are specified to be equidistant, which is normally taken to motivate the use of OLS regressions (see section 2.1), but which is in line with footnote 6 of the main text. In contrast, the OP-Avg. and MOP-Avg. models both perform very well. The UKMOP-Avg. also performs much better than OLS-FE, though coefficients seem somewhat biased towards zero.

Figure D1. Density plots of estimates of β_1/β_2 when scale use is independent of \mathbf{X}_{it}



Note: The OP-Avg., the MOP-Avg. and UKMOP-Avg. model give estimates which are centred close to the true value ($=2$), while the OLS-FE model is somewhat negatively biased, but with little variance. The OLS and OP model also have little variance but yield a strongly negatively biased ratio of coefficients.

Moreover, from the experimental setup, it follows that the true ratios β_1/β_2 and β_1/β_3 are equal to 2, while the ratio β_2/β_3 equals 1. With respect to these target values, the OLS regressions in column (1) perform worst in terms of both the average prediction and the root mean squared error (RMSE). Simple ordered probit regressions shown in column (2) perform almost as badly, with all ratios of coefficients being far below the target. All other models perform rather similar and quite well. Of these, the OLS-FE models shows the worst average estimate of each ratio of coefficients, with a slight negative bias compared to the true value.

Hence, and encouragingly, the ordered probit model with individual averages added (see column (4)) performs best with almost no bias in either direction and low RMSE. Thus, the argument of Appendix C, namely that accounting for fixed effects in an ordered probit setting is possible, appears to be correct. Models which are extended by an equation for memories of past satisfaction ('MOP' and 'UKMOP') perform almost as well as the OP-Avg. model, with the average estimates being close to the true values. However, RMSEs are larger than for the OP-Avg. model because these models estimate the β coefficients using one fewer wave. On top, by explicitly modelling dependence of scale use on \mathbf{X}_{it} ,⁴² as well as memories of past satisfaction, these models estimate three times as many free parameters as the other models, leading to a loss in efficiency. One reason why the UKMOP model may perform worse than the MOP model is because each observation of $r_{it}^{(m)}$ in that model carries much less information than $r_{it}^{(m)}$ in the MOP model (cf. section 3.2).

⁴² Although the coefficients are explicitly modelled, all true coefficients relating to scale use are zero in the present example.

The location of the cutoffs is thus less precisely estimated. This can also be seen in the estimates of α and δ , which are somewhat less precise in the UKMOP model compared to the MOP model.

D.3 Results in case of cutoffs depending on X_{it}

I now turn to the case in which higher values of each independent variable makes scale use more stringent. The assumption of common scale use is thus violated. Table D2 presents results from the same 250 trials as in table D1, with only the data-generating process for the cutoffs altered. Because true differences between thresholds are now smaller, the true values for normalized coefficients are given by $\text{Norm.}\beta_1 = 0.96$, $\text{Norm.}\beta_2 = \text{Norm.}\beta_3 = 0.48$.

Table D2. Results from Monte Carlo simulation with cutoffs depending on values of X_{it}

Parameter estimates	(1) OLS	(2) OP	(3) OLS-FE	(4) OP-Avg.	(5) MOP-Avg.	(6) UKMOP-Avg.
β_1	0.672	0.818	0.564	0.767	0.964	0.968
β_2	0.039	0.039	-0.219	-0.286	0.479	0.5
β_3	0.222	0.235	-0.227	-0.288	0.478	0.495
β_1/β_2	17.367 (15.546)	21.551 (19.86)	-2.58 (4.581)	-2.683 (4.683)	2.015 (0.048)	1.940 (0.086)
β_1/β_3	3.034 (1.047)	3.490 (1.504)	-2.491 (4.493)	-2.668 (4.670)	2.020 (0.111)	1.967 (0.162)
β_2/β_3	0.178 (0.823)	0.166 (0.835)	0.966 (0.063)	0.995 (0.053)	10.003 (0.054)	10.015 (0.087)
Norm. β_1	0.672 (0.288)	0.910 (0.050)	0.564 (0.396)	0.771 (0.189)	0.968 (0.015)	0.929 (0.037)
Norm. β_2	0.039 (0.441)	0.043 (0.437)	-0.219 (0.699)	-0.288 (0.768)	0.481 (0.010)	0.479 (0.015)
Norm. β_3	0.222 (0.258)	0.262 (0.219)	-0.227 (0.707)	-0.290 (0.770)	0.480 (0.025)	0.475 (0.036)
α_1					0.783	0.804
α_2					0.383	0.392
α_3					0.575	0.585
α_1/α_2					2.044 (0.062)	2.053 (0.097)
α_1/α_3					1.362 (0.056)	1.380 (0.098)
α_2/α_3					0.666 (0.026)	0.673 (0.047)
δ_1					0.221	0.193
δ_2					0.765	0.799
δ_3					0.767	0.797
δ_1/δ_2					0.289	0.242
δ_1/δ_3					0.289	0.243
δ_2/δ_3					0.998	1.004

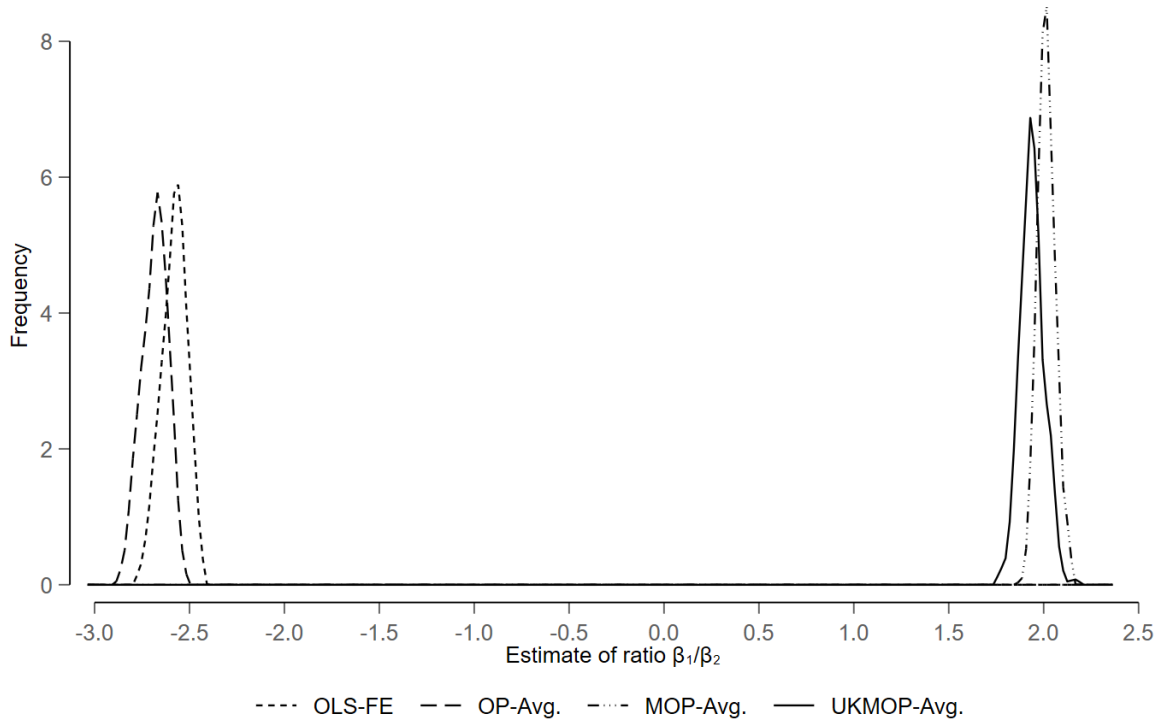
Note: $T = 6$ and $N = 5000$. Numbers show mean estimates of coefficients and their ratios from 250 independent trials. Root mean squared error (RMSE) shown in parentheses. True values for each parameter are: $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\alpha_1 = 0.8$, $\alpha_2 = 0.4$, $\alpha_3 = 0.6$, $\delta_1 = 0.2$, $\delta_2 = 0.8$, $\delta_3 = 0.8$. For OP regressions ‘Norm. β ’ give normalized coefficients, which are given by $(K - 1)\beta/(\gamma_{K-1} - \gamma_1) = 6\beta/(\gamma_6 - \gamma_1)$. For OLS regressions, ‘Norm. β ’ is identical to β since r is rank-order coded. True values are $\text{Norm.}\beta_1 = 0.96$, $\text{Norm.}\beta_2 = \text{Norm.}\beta_3 = 0.48$.

With respect to both target values for ratios and normalized magnitudes, the OLS and OP model continue to be strongly biased. When considering estimated ratios of coefficients, the OLS-FE and OP-Avg. model are now similarly biased.⁴³ When considering normalized magnitudes, these models are even more biased than the simple OLS and OP models. Importantly, since the true effects of $X_{2,it}$ and $X_{3,it}$ on scale use are stronger than on latent satisfaction (and of the same sign), these models fail to yield the correct signs for these coefficient estimates. In contrast, the MOP-Avg. and the UKMOP-Avg. models continue to yield correct coefficient ratios for each β

⁴³ However, on average the OP-Avg. provides the correct estimate for the ratio β_1/β_2 . As flagged in section 2.3., this is because the special case $\beta_2 = \beta_3$ and $\delta_2 = \delta_3$ obtains here.

coefficient.⁴⁴ Indeed, these estimates of ratios are almost unaffected by the change in how cutoffs are determined. Moreover, MOP and UKMOP models also yield largely correct normalized magnitudes of coefficients, though the coefficient on β_1 estimated by the UKMOP model seems to be underestimated slightly. Finally, ratios δ_1/δ_2 and δ_1/δ_3 , which indicate relative shifts in scale use, appear positively biased for the MOP model, and negatively biased for the UKMOP model, though in the latter case that bias is negligible. The ratio δ_2/δ_3 is well estimated by both models. Signs of coefficients are always correctly estimated by both the MOP and the UKMOP model.

Figure D2. Density plots of estimates of β_1/β_2 when scale use depends on X_{it}



Note: The MOP-Avg. and UKMOP-Avg. model continue to give estimates which are centred close to the true value ($=2$). The OLS-FE and OP-Avg. model now yield a strongly negatively biased estimate. The OP and OLS model are omitted from this figure.

I draw three conclusions. First, the approach of adding individual averages to an ordered probit model in order to account for individual fixed effects works well and appears to yield consistent estimates of the normalized coefficients and trade-off ratios. Remarkably, the OLS-FE model, which is standard in empirical happiness research, performs worse than the OP-Avg. model. This is despite the fact that cutoffs are equidistant, which is an assumption normally maintained in order to motivate the OLS approach. Second, in cases where independent variables affect scale use, both the OP-Avg. and OLS-FE model yield biased estimates. Third, the MOP and UKMOP models appear to yield close to consistent estimates whether or not scale use is affected by any of the independent variables. However, the MOP and even more so the UKMOP model are less efficient than the OLS-FE and the OP-Avg. model.

⁴⁴ And, on average, ratios of coefficients for memories of satisfaction are also estimated correctly. RMSEs for the UKMOP model are even improved compared to Table D1.

E Additional results tables

Table E1. Multinomial logistic regressions of changes in stringency of scale use (exponentiated coefficients)

	(1) Not controlling for levels of reported satisfaction		(2) Controlling for levels of reported satisfaction	
	Less stringent	More stringent	Less stringent	More stringent
ln(HH income)	0.945** (0.024)	1.175*** (0.026)	1.000 (0.027)	1.131*** (0.024)
Employed (base)				
<i>Self-employed</i>	1.044 (0.089)	1.171*** (0.070)	1.087 (0.093)	1.125** (0.067)
<i>Unemployed</i>	1.683*** (0.155)	0.702*** (0.052)	1.405*** (0.132)	0.790*** (0.058)
<i>Retired</i>	1.039 (0.085)	1.258*** (0.091)	1.101 (0.092)	1.227*** (0.088)
<i>Family care</i>	1.107 (0.089)	0.987 (0.057)	1.067 (0.087)	1.012 (0.058)
<i>Lt. sick/ disabled</i>	2.477*** (0.205)	0.559*** (0.047)	1.662*** (0.138)	0.756*** (0.063)
<i>In education</i>	0.994 (0.094)	1.177*** (0.063)	1.062 (0.101)	1.131** (0.061)
Married (base)				
<i>Widowed</i>	1.413*** (0.097)	0.995 (0.070)	1.154** (0.076)	1.094 (0.076)
<i>Divorced</i>	1.526*** (0.081)	1.049 (0.049)	1.183*** (0.060)	1.208*** (0.056)
<i>Never married</i>	1.086 (0.056)	0.773*** (0.026)	0.950 (0.047)	0.830*** (0.028)
ln(# adults)	1.118*** (0.043)	0.790*** (0.021)	1.063* (0.039)	0.800*** (0.021)
Childbirth	0.858* (0.073)	1.706*** (0.071)	0.916 (0.078)	1.671*** (0.070)
Reference income	1.708*** (0.144)	2.211*** (0.144)	1.584*** (0.126)	2.160*** (0.139)
$r_{it}=1$ (base)				
$r_{it}=2$			0.878 (0.088)	0.833 (0.143)
$r_{it}=3$			0.941 (0.082)	0.926 (0.137)
$r_{it}=4$			0.652*** (0.055)	1.475*** (0.206)
$r_{it}=5$			0.431*** (0.036)	2.952*** (0.406)
$r_{it}=6$			0.173*** (0.015)	3.517*** (0.484)
$r_{it}=7$			0.093*** (0.010)	3.405*** (0.476)
Observations	99,260		99,260	

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01 (against null of odds ratio=1).

Table E2. Robustness of effect on latent satisfaction

	Estimate	Standard error	p-value
ln(HH income)	0.048***	0.011	0.000
Employed (base)			
<i>Self-employed</i>	0.036	0.031	0.235
<i>Unemployed</i>	-0.308***	0.035	0.000
<i>Retired</i>	0.096***	0.037	0.009
<i>Family care</i>	-0.024	0.028	0.380
<i>Lt. sick / disabled</i>	-0.271***	0.041	0.000
<i>In education</i>	0.028	0.031	0.376
Married (base)			
<i>Widowed</i>	-0.503***	0.069	0.000
<i>Divorced</i>	-0.185***	0.034	0.000
<i>Never married</i>	-0.143***	0.032	0.000
ln(HH size)	-0.108***	0.024	0.000
Childbirth	0.112***	0.022	0.000
Reference income	-0.110	0.093	0.238

Note: Test computed under assumption that $\lambda_j^{(m)}$ is not larger than β_j and of opposite sign. * p<0.1, ** p<0.05, *** p<0.01.

Table E3. Addition of Lags and Leads

	No lags no leads (1)	One-year lag in current satisfaction, two-year lags in memories (2)	One-year lead in current satisfaction and one-year lead in scale use (3)	One-year lag and lead in current satisfaction, two-year lags in memories, one-year lead scale use (4)
Current life satisfaction				
Contemporaneous effects				
ln(HH income)	0.097*** (0.023)	0.108*** (0.024)	0.108*** (0.026)	0.113*** (0.028)
Employed (base)				
<i>Self-employed</i>	0.072 (0.061)	0.063 (0.065)	0.020 (0.070)	-0.008 (0.075)
<i>Unemployed</i>	-0.615*** (0.069)	-0.694*** (0.076)	-0.613*** (0.079)	-0.672*** (0.088)
<i>Retired</i>	0.192*** (0.073)	0.184** (0.078)	0.110 (0.086)	0.081 (0.092)
<i>Family care</i>	-0.048 (0.055)	-0.086 (0.060)	-0.075 (0.063)	-0.123* (0.068)
<i>Lt. sick / disabled</i>	-0.542*** (0.082)	-0.578*** (0.090)	-0.601*** (0.093)	-0.620*** (0.102)
<i>In education</i>	0.056 (0.063)	0.006 (0.072)	0.007 (0.075)	-0.058 (0.086)
Married (base)				
<i>Widowed</i>	-1.007*** (0.137)	-1.192*** (0.149)	-0.618*** (0.171)	-0.797*** (0.184)
<i>Divorced</i>	-0.369*** (0.069)	-0.478*** (0.074)	-0.105 (0.083)	-0.229*** (0.089)
<i>Never married</i>	-0.287*** (0.064)	-0.368*** (0.071)	-0.129 (0.082)	-0.227** (0.091)
ln(HH size)	-0.216*** (0.048)	-0.163*** (0.052)	-0.204*** (0.061)	-0.163** (0.066)
Childbirth	0.224*** (0.044)	0.238*** (0.049)	0.256*** (0.050)	0.282*** (0.056)
Reference inc.	-0.220 (0.187)	-0.045 (0.203)	-0.085 (0.230)	0.065 (0.250)
1-year lagged effects				
ln(HH income)		0.009 (0.008)		0.006 (0.009)
Employed (base)				
<i>Self-employed</i>		-0.052** (0.023)		-0.053** (0.025)
<i>Unemployed</i>		-0.001 (0.028)		-0.028 (0.031)
<i>Retired</i>		-0.009 (0.024)		-0.021 (0.027)
<i>Family care</i>		0.001 (0.021)		0.001 (0.023)
<i>Lt. sick / disabled</i>		-0.038 (0.032)		-0.042 (0.036)
<i>In education</i>		0.063** (0.026)		0.071** (0.029)
Married (base)				
<i>Widowed</i>		0.199*** (0.045)		0.247*** (0.049)
<i>Divorced</i>		0.181*** (0.030)		0.192*** (0.033)
<i>Never married</i>		0.128*** (0.025)		0.117*** (0.028)
ln(HH size)		-0.092*** (0.019)		-0.077*** (0.021)
Childbirth		0.019 (0.016)		0.021 (0.018)
Reference inc.		-0.232*** (0.067)		-0.126* (0.073)
1-year lead effects				
ln(HH income)			0.008 (0.024)	0.001 (0.026)
Employed (base)				
<i>Self-employed</i>			-0.034 (0.070)	-0.016 (0.076)
<i>Unemployed</i>			-0.056 (0.078)	-0.111 (0.087)
<i>Retired</i>			0.073 (0.083)	0.064 (0.088)
<i>Family care</i>			-0.041 (0.064)	-0.007 (0.069)
<i>Lt. sick / disabled</i>			-0.290*** (0.093)	-0.302*** (0.100)
<i>In education</i>			0.045 (0.074)	0.087 (0.085)
Married (base)				
<i>Widowed</i>			-0.674*** (0.152)	-0.685*** (0.166)
<i>Divorced</i>			-0.456*** (0.078)	-0.504*** (0.086)
<i>Never married</i>			-0.281*** (0.077)	-0.249*** (0.085)
ln(HH size)			-0.042 (0.052)	-0.089 (0.057)
Childbirth			0.232*** (0.053)	0.257*** (0.058)
Reference inc.			-0.180 (0.219)	-0.339 (0.240)
Memory of last year's life satisfaction				
1-year lagged effects				
ln(HH income)	0.027 (0.020)	0.029 (0.022)	0.042* (0.023)	0.045* (0.025)
Employed (base)				
<i>Self-employed</i>	0.010 (0.058)	0.018 (0.065)	0.102 (0.064)	0.096 (0.070)
<i>Unemployed</i>	-0.135** (0.063)	-0.118* (0.070)	-0.153** (0.073)	-0.148* (0.082)
<i>Retired</i>	-0.081 (0.073)	-0.113 (0.084)	0.025 (0.083)	-0.014 (0.095)
<i>Family care</i>	-0.115** (0.053)	-0.089 (0.057)	-0.082 (0.059)	-0.086 (0.065)
<i>Lt. sick / disabled</i>	-0.314*** (0.078)	-0.300*** (0.087)	-0.302*** (0.087)	-0.310*** (0.096)
<i>In education</i>	-0.008 (0.056)	-0.025 (0.064)	-0.010 (0.065)	-0.061 (0.074)
Married (base)				
<i>Widowed</i>	-0.968*** (0.135)	-0.556*** (0.152)	-1.047*** (0.167)	-0.683*** (0.185)

Table E3. Addition of Lags and Leads (continued)

<i>Divorced</i>	-0.419*** (0.063)	-0.352*** (0.075)	-0.401*** (0.073)	-0.362*** (0.084)
<i>Never married</i>	-0.285*** (0.057)	-0.238*** (0.069)	-0.273*** (0.066)	-0.283*** (0.079)
ln(HH size)	-0.008 (0.046)	-0.048 (0.058)	-0.043 (0.052)	-0.090 (0.065)
Childbirth	0.093** (0.039)	0.097** (0.044)	0.075* (0.043)	0.066 (0.050)
Reference inc.	0.216 (0.173)	0.046 (0.208)	0.173 (0.196)	-0.026 (0.232)
2-year lagged effects				
ln(HH income)		0.052** (0.022)		0.056** (0.025)
Employed (base)				
<i>Self-employed</i>		0.139** (0.062)		0.116* (0.068)
<i>Unemployed</i>		-0.086 (0.067)		-0.074 (0.074)
<i>Retired</i>		0.213** (0.083)		0.179** (0.091)
<i>Family care</i>		-0.017 (0.056)		0.013 (0.063)
<i>Lt. sick / disabled</i>		-0.054 (0.087)		-0.006 (0.096)
<i>In education</i>		0.033 (0.063)		0.038 (0.072)
Married (base)				
<i>Widowed</i>		-0.712*** (0.136)		-0.678*** (0.153)
<i>Divorced</i>		-0.078 (0.066)		-0.101 (0.076)
<i>Never married</i>		-0.043 (0.060)		-0.003 (0.069)
ln(HH size)		0.078 (0.047)		0.027 (0.054)
Childbirth		-0.028 (0.041)		-0.036 (0.045)
Reference inc.		0.211 (0.191)		0.201 (0.208)
Constant	-0.920 (0.782)	-0.937 (1.035)	-1.202 (0.989)	-2.020 (1.296)
Scale shift				
contemporaneous effects				
ln(HH income)	0.059*** (0.021)	0.076*** (0.023)	0.071*** (0.025)	0.075*** (0.027)
Employed (base)				
<i>Self-employed</i>	0.051 (0.056)	0.037 (0.060)	0.027 (0.065)	-0.003 (0.070)
<i>Unemployed</i>	-0.418*** (0.065)	-0.471*** (0.071)	-0.399*** (0.075)	-0.445*** (0.083)
<i>Retired</i>	0.122* (0.070)	0.116 (0.075)	0.086 (0.083)	0.043 (0.090)
<i>Family care</i>	-0.047 (0.051)	-0.066 (0.056)	-0.052 (0.059)	-0.091 (0.064)
<i>Lt. sick / disabled</i>	-0.223*** (0.077)	-0.259*** (0.086)	-0.280*** (0.089)	-0.329*** (0.099)
<i>In education</i>	-0.039 (0.058)	-0.062 (0.067)	-0.092 (0.070)	-0.147* (0.080)
Married (base)				
<i>Widowed</i>	-0.671*** (0.132)	-0.703*** (0.143)	-0.482*** (0.167)	-0.511*** (0.177)
<i>Divorced</i>	-0.107* (0.060)	-0.104 (0.064)	0.034 (0.074)	0.042 (0.078)
<i>Never married</i>	-0.173*** (0.060)	-0.175*** (0.067)	-0.083 (0.078)	-0.112 (0.086)
ln(HH size)	-0.129*** (0.047)	-0.127** (0.050)	-0.136** (0.058)	-0.122* (0.063)
Childbirth	0.142*** (0.042)	0.145*** (0.046)	0.153*** (0.048)	0.162*** (0.053)
Reference inc.	0.034 (0.179)	0.076 (0.194)	-0.020 (0.223)	0.040 (0.241)
1-year lead effects				
ln(HH income)			0.008 (0.023)	0.000 (0.025)
Employed (base)				
<i>Self-employed</i>			-0.032 (0.067)	-0.019 (0.072)
<i>Unemployed</i>			0.039 (0.075)	0.007 (0.082)
<i>Retired</i>			0.038 (0.082)	0.051 (0.087)
<i>Family care</i>			-0.031 (0.061)	0.015 (0.066)
<i>Lt. sick / disabled</i>			-0.150* (0.088)	-0.171* (0.096)
<i>In education</i>			0.049 (0.071)	0.083 (0.080)
Married (base)				
<i>Widowed</i>			-0.386** (0.151)	-0.367** (0.165)
<i>Divorced</i>			-0.210*** (0.072)	-0.275*** (0.078)
<i>Never married</i>			-0.183** (0.073)	-0.152* (0.082)
ln(HH size)			-0.031 (0.050)	-0.084 (0.054)
Childbirth			0.127** (0.050)	0.157*** (0.055)
Reference inc.			0.085 (0.211)	-0.050 (0.231)
Cutoff constants				
γ_1	-4.794*** (0.310)	-4.753*** (0.407)	-5.312*** (0.386)	-5.334*** (0.477)
γ_2	-4.382*** (0.025)	-4.338*** (0.027)	-4.898*** (0.027)	-4.920*** (0.030)
γ_3	-3.837*** (0.014)	-3.787*** (0.016)	-4.346*** (0.016)	-4.365*** (0.017)
γ_4	-3.200*** (0.010)	-3.145*** (0.010)	-3.702*** (0.011)	-3.718*** (0.011)
γ_5	-2.329*** (0.007)	-2.263*** (0.008)	-2.826*** (0.008)	-2.836*** (0.008)
γ_6	-1.283** (0.007)	-1.198*** (0.008)	-1.770*** (0.008)	-1.766*** (0.008)
Observations	118,440	102,924	97,939	85,472

Note: Column (1) is identical to column (4) of table 2. Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E4. Estimates for fixed reference groups (with lags and leads)

	No lags no leads (1)	One-year lag in current satisfaction, two- year lags in memories (2)	One-year lead in current satisfaction and one-year lead in scale use (3)	One-year lag and lead in current satisfaction, two-year lags in memories, one-year lead in scale use (4)
Current life satisfaction				
Contemporaneous effects				
ln(HH income)	0.094*** (0.023)	0.103*** (0.026)	0.107*** (0.027)	0.115*** (0.031)
Employed (base)				
<i>Self-employed</i>	0.102 (0.063)	0.131* (0.072)	0.055 (0.074)	0.052 (0.085)
<i>Unemployed</i>	-0.615*** (0.072)	-0.697*** (0.084)	-0.592*** (0.084)	-0.657*** (0.099)
<i>Retired</i>	0.169** (0.078)	0.177** (0.086)	0.140 (0.090)	0.138 (0.102)
<i>Family care</i>	-0.070 (0.057)	-0.101 (0.065)	-0.097 (0.066)	-0.160** (0.074)
<i>Lt. sick / disabled</i>	-0.543*** (0.085)	-0.598*** (0.099)	-0.591*** (0.097)	-0.664*** (0.112)
<i>In education</i>	0.012 (0.066)	-0.068 (0.079)	-0.046 (0.077)	-0.130 (0.093)
Married (base)				
<i>Widowed</i>	-1.038*** (0.141)	-1.219*** (0.162)	-0.711*** (0.182)	-0.908*** (0.207)
<i>Divorced</i>	-0.337*** (0.072)	-0.448*** (0.083)	-0.090 (0.089)	-0.252** (0.099)
<i>Never married</i>	-0.329*** (0.066)	-0.410*** (0.078)	-0.182** (0.083)	-0.290*** (0.097)
ln(HH size)	-0.184*** (0.051)	-0.142** (0.057)	-0.177*** (0.063)	-0.153** (0.072)
Childbirth	0.209*** (0.046)	0.229*** (0.054)	0.239*** (0.052)	0.295*** (0.064)
Reference income	0.065 (0.231)	0.246 (0.262)	0.014 (0.280)	0.141 (0.317)
Reference employed	-1.211* (0.641)	-0.995 (0.754)	-0.528 (0.783)	-0.229 (0.918)
Reference unemployed	-0.237 (0.816)	-0.853 (0.979)	-0.359 (0.997)	-0.171 (1.202)
Reference retired	0.459 (0.659)	-0.353 (0.779)	-0.064 (0.907)	-0.503 (1.050)
Reference % disabled	0.512 (0.936)	-0.204 (1.079)	2.264** (1.095)	1.407 (1.246)
Reference % married	0.290 (0.440)	0.223 (0.531)	-0.495 (0.599)	-0.855 (0.702)
Reference % divorced	0.138 (0.846)	0.408 (0.988)	-0.426 (1.029)	-0.587 (1.207)
Reference % widowed	-1.378 (1.115)	-1.418 (1.272)	-1.979 (1.534)	-2.421 (1.657)
Reference % birth	-0.249 (0.651)	0.390 (0.794)	-0.263 (0.779)	0.561 (0.956)
Reference ln(HHsize)	-0.056 (0.336)	-0.085 (0.416)	0.615 (0.459)	0.437 (0.538)
Reference ln(Job hours)	0.163 (0.204)	-0.033 (0.239)	0.099 (0.241)	0.005 (0.282)
Reference life satisfaction	0.659*** (0.125)	0.610*** (0.148)	0.632*** (0.144)	0.468*** (0.173)
1-year lagged effects				
ln(HH income)		0.006 (0.009)		-0.000 (0.010)
Employed (base)				
<i>Self-employed</i>		-0.033 (0.025)		-0.033 (0.028)
<i>Unemployed</i>		0.009 (0.031)		-0.027 (0.035)
<i>Retired</i>		-0.013 (0.027)		-0.023 (0.030)
<i>Family care</i>		0.000 (0.023)		-0.003 (0.026)
<i>Lt. sick / disabled</i>		-0.041 (0.035)		-0.064 (0.039)
<i>In education</i>		0.058** (0.029)		0.065** (0.033)
Married (base)		0.000 (.)		0.000 (.)
<i>Widowed</i>		0.216*** (0.049)		0.258*** (0.055)
<i>Divorced</i>		0.191*** (0.033)		0.193*** (0.036)
<i>Never married</i>		0.106*** (0.028)		0.087*** (0.031)
ln(HH size)		-0.075*** (0.021)		-0.053** (0.024)
Childbirth		-0.001 (0.018)		0.011 (0.020)
Reference income		-0.171** (0.087)		-0.116 (0.099)
Reference employed		-0.294 (0.268)		-0.104 (0.323)
Reference unemployed		0.328 (0.327)		0.476 (0.378)
Reference retired		0.034 (0.218)		0.164 (0.245)
Reference % disabled		-0.194 (0.338)		0.090 (0.393)
Reference % married		0.192 (0.184)		0.210 (0.223)
Reference % divorced		-0.282 (0.320)		-0.316 (0.364)
Reference % widowed		-0.443 (0.324)		-0.235 (0.371)
Reference % birth		-0.094 (0.272)		-0.063 (0.316)
Reference ln(HHsize)		-0.159 (0.134)		-0.169 (0.162)
Reference ln(Job hours)		0.058 (0.081)		0.020 (0.095)

Table E4. Estimates for fixed reference groups (with lags and leads) (continued)

Reference life satisfaction		-0.074* (0.039)		-0.079* (0.045)
1-year leaded effects				
ln(HH income)		0.006 (0.025)		-0.003 (0.029)
Employed (base)				
<i>Self-employed</i>		-0.033 (0.073)		-0.001 (0.082)
<i>Unemployed</i>		-0.062 (0.083)		-0.061 (0.098)
<i>Retired</i>		0.052 (0.089)		0.035 (0.097)
<i>Family care</i>		-0.081 (0.067)		0.015 (0.077)
<i>Lt. sick / disabled</i>		-0.309*** (0.097)		-0.331*** (0.109)
<i>In education</i>		-0.003 (0.079)		0.033 (0.095)
Married (base)				
<i>Widowed</i>		-0.543*** (0.163)		-0.524*** (0.189)
<i>Divorced</i>		-0.421*** (0.082)		-0.483*** (0.097)
<i>Never married</i>		-0.287*** (0.079)		-0.264*** (0.093)
ln(HH size)		-0.014 (0.056)		-0.063 (0.064)
Childbirth		0.234*** (0.057)		0.280*** (0.066)
Reference income		0.055 (0.269)		-0.027 (0.317)
Reference employed		-0.453 (0.765)		0.027 (0.902)
Reference unemployed		0.526 (1.003)		-1.078 (1.206)
Reference retired		0.746 (0.768)		0.822 (0.909)
Reference % disabled		-2.460** (1.089)		-2.613** (1.270)
Reference % married		0.791 (0.536)		0.779 (0.626)
Reference % divorced		0.023 (0.968)		0.890 (1.113)
Reference % widowed		0.309 (1.336)		1.228 (1.503)
Reference % birth		0.448 (0.797)		1.109 (0.951)
Reference ln(HHsize)		-0.405 (0.432)		-0.199 (0.518)
Reference ln(Job hours)		0.098 (0.239)		-0.135 (0.277)
Reference life satisfaction		-0.046 (0.132)		0.067 (0.152)
Memory of last year's life satisfaction				
1-year lagged effects				
ln(HH income)	0.020 (0.021)	0.027 (0.024)	0.032 (0.024)	0.041 (0.027)
Employed (base)				
<i>Self-employed</i>	-0.035 (0.061)	-0.021 (0.071)	0.070 (0.068)	0.094 (0.078)
<i>Unemployed</i>	-0.117* (0.065)	-0.080 (0.076)	-0.125 (0.077)	-0.092 (0.090)
<i>Retired</i>	-0.149* (0.078)	-0.157* (0.090)	-0.056 (0.089)	-0.059 (0.102)
<i>Family care</i>	-0.119** (0.055)	-0.096 (0.062)	-0.084 (0.062)	-0.064 (0.072)
<i>Lt. sick / disabled</i>	-0.324*** (0.081)	-0.257*** (0.092)	-0.303*** (0.091)	-0.226** (0.104)
<i>In education</i>	-0.011 (0.059)	-0.018 (0.071)	0.007 (0.070)	-0.017 (0.084)
Married (base)				
<i>Widowed</i>	-0.985*** (0.137)	-0.503*** (0.163)	-1.072*** (0.168)	-0.672*** (0.198)
<i>Divorced</i>	-0.398*** (0.065)	-0.374*** (0.081)	-0.372*** (0.075)	-0.401*** (0.090)
<i>Never married</i>	-0.314*** (0.059)	-0.287*** (0.075)	-0.302*** (0.069)	-0.343*** (0.086)
ln(HH size)	-0.024 (0.047)	-0.096* (0.058)	-0.045 (0.053)	-0.139** (0.063)
Childbirth	0.095** (0.041)	0.133*** (0.048)	0.066 (0.045)	0.078 (0.055)
Reference income	0.277 (0.214)	-0.104 (0.269)	0.012 (0.246)	-0.377 (0.304)
Reference employed	-0.566 (0.606)	-0.733 (0.765)	-0.375 (0.718)	-0.106 (0.904)
Reference unemployed	-0.049 (0.735)	-0.626 (0.921)	-0.097 (0.855)	-0.198 (1.046)
Reference retired	1.091* (0.638)	0.136 (0.876)	1.522** (0.714)	1.244 (0.958)
Reference % disabled	-0.141 (0.854)	-1.377 (1.054)	-0.387 (0.983)	-1.604 (1.218)
Reference % married	-0.443 (0.424)	-0.527 (0.583)	-0.424 (0.520)	-0.438 (0.701)
Reference % divorced	-0.360 (0.779)	-0.033 (0.959)	-0.635 (0.897)	-0.507 (1.103)
Reference % widowed	-0.856 (1.073)	-0.577 (1.388)	-0.909 (1.289)	-0.577 (1.596)
Reference % birth	0.242 (0.615)	0.173 (0.762)	0.419 (0.735)	0.419 (0.893)
Reference ln(HHsize)	0.204 (0.334)	-0.005 (0.431)	0.146 (0.399)	-0.060 (0.504)
Reference ln(Job hours)	0.206 (0.187)	0.247 (0.237)	0.273 (0.219)	0.281 (0.275)
Reference life satisfaction	0.073 (0.107)	0.045 (0.121)	0.087 (0.122)	-0.005 (0.138)
2-year lagged effects				
ln(HH income)		0.044* (0.024)		0.042 (0.027)

Table E4. Estimates for fixed reference groups (with lags and leads) (continued)

Employed (base)				
<i>Self-employed</i>		0.155** (0.069)		0.120 (0.076)
<i>Unemployed</i>		-0.071 (0.074)		-0.064 (0.084)
<i>Retired</i>		0.145 (0.090)		0.120 (0.100)
<i>Family care</i>		-0.000 (0.061)		0.039 (0.068)
<i>Lt. sick / disabled</i>		-0.094 (0.095)		-0.011 (0.106)
<i>In education</i>		0.052 (0.071)		0.057 (0.081)
Married (base)		0.000 (.)		0.000 (.)
<i>Widowed</i>		-0.780*** (0.152)		-0.775*** (0.172)
<i>Divorced</i>		-0.091 (0.074)		-0.105 (0.087)
<i>Never married</i>		-0.041 (0.067)		0.018 (0.079)
ln(HH size)		0.098* (0.051)		0.051 (0.058)
Childbirth		-0.008 (0.045)		-0.036 (0.050)
Reference income		-0.020 (0.261)		-0.154 (0.290)
Reference employed		1.001 (0.739)		0.610 (0.870)
Reference unemployed		-1.195 (0.887)		-1.809* (0.999)
Reference retired		-0.171 (0.716)		-0.769 (0.822)
Reference % disabled		-0.328 (0.969)		-1.105 (1.098)
Reference % married		0.105 (0.522)		0.137 (0.604)
Reference % divorced		-1.544* (0.923)		-1.728 (1.058)
Reference % widowed		-1.798 (1.279)		-1.710 (1.490)
Reference % birth		-0.293 (0.748)		0.112 (0.866)
Reference ln(HHsize)		-0.119 (0.398)		-0.326 (0.456)
Reference ln(Job hours)		-0.275 (0.226)		-0.265 (0.267)
Reference life satisfaction		-0.163 (0.126)		-0.166 (0.140)
Constant	-0.664 (2.032)	2.485 (2.475)	0.274 (2.434)	1.139 (2.874)
Scale shift				
contemporaneous effects				
ln(HH income)	0.062*** (0.022)	0.079*** (0.025)	0.075*** (0.026)	0.087*** (0.029)
Employed (base)				
<i>Self-employed</i>	0.068 (0.058)	0.094 (0.066)	0.057 (0.069)	0.050 (0.079)
<i>Unemployed</i>	-0.414*** (0.067)	-0.503*** (0.078)	-0.377*** (0.078)	-0.454*** (0.092)
<i>Retired</i>	0.154** (0.075)	0.135 (0.084)	0.141 (0.088)	0.094 (0.100)
<i>Family care</i>	-0.056 (0.053)	-0.085 (0.061)	-0.066 (0.062)	-0.129* (0.069)
<i>Lt. sick / disabled</i>	-0.207*** (0.080)	-0.266*** (0.093)	-0.260*** (0.093)	-0.352*** (0.108)
<i>In education</i>	-0.052 (0.061)	-0.108 (0.073)	-0.117 (0.073)	-0.194** (0.086)
Married (base)				
<i>Widowed</i>	-0.727*** (0.136)	-0.702*** (0.156)	-0.599*** (0.180)	-0.564*** (0.202)
<i>Divorced</i>	-0.077 (0.064)	-0.063 (0.072)	0.046 (0.080)	0.049 (0.088)
<i>Never married</i>	-0.173*** (0.062)	-0.175** (0.072)	-0.102 (0.078)	-0.147* (0.089)
ln(HH size)	-0.114** (0.049)	-0.108** (0.054)	-0.120** (0.061)	-0.107 (0.069)
Childbirth	0.134*** (0.044)	0.137*** (0.051)	0.141*** (0.051)	0.157*** (0.061)
Reference income	0.074 (0.223)	0.205 (0.253)	-0.030 (0.271)	0.068 (0.308)
Reference employed	-0.547 (0.608)	-0.257 (0.715)	0.066 (0.746)	0.496 (0.877)
Reference unemployed	-0.338 (0.779)	-0.537 (0.937)	-0.503 (0.963)	0.094 (1.158)
Reference retired	0.526 (0.645)	0.079 (0.760)	0.039 (0.895)	0.119 (1.033)
Reference % disabled	0.544 (0.896)	0.078 (1.034)	2.196** (1.064)	1.839 (1.202)
Reference % married	0.291 (0.420)	0.266 (0.498)	-0.669 (0.572)	-0.863 (0.672)
Reference % divorced	-0.096 (0.809)	0.292 (0.942)	-0.949 (0.992)	-0.709 (1.168)
Reference % widowed	-0.830 (1.096)	-1.019 (1.248)	-1.803 (1.532)	-2.066 (1.651)
Reference % birth	-0.117 (0.613)	0.568 (0.742)	-0.220 (0.739)	0.888 (0.897)
Reference ln(HHsize)	0.078 (0.329)	0.008 (0.401)	0.636 (0.448)	0.527 (0.523)
Reference ln(Job hours)	0.089 (0.194)	-0.076 (0.227)	0.046 (0.229)	-0.003 (0.269)
Reference life satisfaction	0.023 (0.121)	0.008 (0.143)	0.012 (0.140)	-0.120 (0.167)
1-year leaded effects				
ln(HH income)			0.010 (0.024)	-0.003 (0.027)
Employed (base)				
<i>Self-employed</i>			-0.052 (0.070)	-0.021 (0.079)
<i>Unemployed</i>			0.029 (0.079)	0.037 (0.092)

Table E4. Estimates for fixed reference groups (with lags and leads) (continued)

Table B.V. Estimates for fixed reference groups (with high and ready) (continued)				
<i>Retired</i>		0.040 (0.087)	0.049 (0.095)	
<i>Family care</i>		-0.063 (0.063)	0.023 (0.073)	
<i>Lt. sick / disabled</i>		-0.154* (0.092)	-0.196* (0.104)	
<i>In education</i>		0.006 (0.075)	0.044 (0.088)	
Married (base)				
<i>Widowed</i>		-0.239 (0.162)	-0.243 (0.188)	
<i>Divorced</i>		-0.174** (0.076)	-0.251*** (0.088)	
<i>Never married</i>		-0.160** (0.075)	-0.124 (0.087)	
ln(HH size)		-0.007 (0.053)	-0.050 (0.061)	
Childbirth		0.132** (0.054)	0.165*** (0.062)	
Reference income		0.002 (0.001)	0.000 (0.001)	
Reference employed		-0.242 (0.734)	0.118 (0.864)	
Reference unemployed		0.148 (0.977)	-1.342 (1.139)	
Reference retired		0.831 (0.750)	0.704 (0.881)	
Reference % disabled		-2.231** (1.057)	-2.685** (1.221)	
Reference % married		0.816 (0.515)	0.714 (0.598)	
Reference % divorced		0.045 (0.939)	0.528 (1.074)	
Reference % widowed		0.592 (1.335)	1.162 (1.500)	
Reference % birth		0.898 (0.767)	1.555* (0.900)	
Reference ln(HHsize)		-0.291 (0.419)	-0.300 (0.496)	
Reference ln(Job hours)		0.060 (0.228)	-0.144 (0.263)	
Reference life satisfaction		-0.069 (0.131)	0.078 (0.150)	
Cutoff constants				
γ_1	2.212** (1.035)	3.190** (1.316)	2.451** (1.214)	3.412** (1.458)
γ_2	-0.873*** (0.025)	-0.865*** (0.028)	-0.876*** (0.028)	-0.872*** (0.031)
γ_3	-0.590*** (0.015)	-0.577*** (0.016)	-0.581*** (0.016)	-0.573*** (0.018)
γ_4	-0.439*** (0.010)	-0.435*** (0.011)	-0.434*** (0.011)	-0.426*** (0.012)
γ_5	-0.122*** (0.007)	-0.113*** (0.008)	-0.123*** (0.008)	-0.113*** (0.008)
γ_6	0.063*** (0.008)	0.080*** (0.008)	0.068*** (0.008)	0.085*** (0.009)
Observations	111,683	90,280	91,631	73,654

Note: All percentages divided by 100. Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E5. Sparse estimates for fixed reference group (with lags and leads)

	No lags no leads (1)	One-year lag in current satisfaction, two-year lags in memories (2)	One-year lead in current satisfaction and one-year lead in scale use (3)	One-year lag and lead in current satisfaction, two-year lags in memories, one-year lead in scale use (4)
Current life satisfaction				
Contemporaneous effects				
ln(HH income)	0.094*** (0.023)	0.103*** (0.026)	0.106*** (0.027)	0.115*** (0.031)
Employed (base)				
<i>Self-employed</i>	0.100 (0.063)	0.128* (0.071)	0.052 (0.074)	0.053 (0.084)
<i>Unemployed</i>	-0.613*** (0.072)	-0.695*** (0.084)	-0.594*** (0.084)	-0.658*** (0.099)
<i>Retired</i>	0.177** (0.077)	0.183** (0.086)	0.133 (0.090)	0.133 (0.101)
<i>Family care</i>	-0.070 (0.057)	-0.099 (0.065)	-0.093 (0.066)	-0.160** (0.074)
<i>Lt. sick / disabled</i>	-0.539*** (0.085)	-0.601*** (0.098)	-0.588*** (0.097)	-0.663*** (0.112)
<i>In education</i>	0.017 (0.065)	-0.064 (0.078)	-0.018 (0.077)	-0.111 (0.092)
Married (base)				
<i>Widowed</i>	-1.041*** (0.141)	-1.221*** (0.163)	-0.705*** (0.183)	-0.888*** (0.207)
<i>Divorced</i>	-0.338*** (0.072)	-0.442*** (0.082)	-0.088 (0.088)	-0.248** (0.099)
<i>Never married</i>	-0.331*** (0.066)	-0.411*** (0.077)	-0.183** (0.084)	-0.289*** (0.098)
ln(HH size)	-0.187*** (0.050)	-0.144** (0.057)	-0.181*** (0.063)	-0.155** (0.072)
Childbirth	0.211*** (0.046)	0.228** (0.054)	0.241*** (0.052)	0.294*** (0.064)
Reference income	0.139 (0.217)	0.256 (0.246)	0.051 (0.261)	0.122 (0.292)
Reference employed	-0.599** (0.284)	-0.776** (0.344)	-0.225 (0.380)	-0.281 (0.456)
Reference % disabled	0.691 (0.792)	0.338 (0.925)	2.552** (0.992)	1.846* (1.114)
Reference % widowed	-1.573 (1.067)	-1.591 (1.206)	-1.662 (1.444)	-1.981 (1.556)
Reference % birth	0.261 (0.573)	0.500 (0.729)	0.059 (0.694)	0.478 (0.877)
Reference life satisfaction	0.690*** (0.123)	0.627*** (0.146)	0.622*** (0.142)	0.456*** (0.170)
1-year lagged effects				
ln(HH income)		0.006 (0.009)		-0.000 (0.010)
Employed (base)				
<i>Self-employed</i>		-0.031 (0.025)		-0.033 (0.028)
<i>Unemployed</i>		0.010 (0.030)		-0.025 (0.035)
<i>Retired</i>		-0.014 (0.027)		-0.024 (0.030)
<i>Family care</i>		0.002 (0.023)		-0.002 (0.026)
<i>Lt. sick / disabled</i>		-0.039 (0.034)		-0.061 (0.039)
<i>In education</i>		0.059** (0.028)		0.061* (0.033)
Married (base)				
<i>Widowed</i>		0.214*** (0.049)		0.260*** (0.055)
<i>Divorced</i>		0.186*** (0.033)		0.190*** (0.036)
<i>Never married</i>		0.105*** (0.028)		0.085*** (0.031)
ln(HH size)		-0.079*** (0.021)		-0.054** (0.023)
Childbirth		0.001 (0.018)		0.012 (0.020)
Reference income		-0.127 (0.080)		-0.109 (0.092)
Reference employed		0.016 (0.121)		0.063 (0.142)
Reference % disabled		0.056 (0.273)		0.167 (0.329)
Reference % widowed		-0.490* (0.286)		-0.205 (0.315)
Reference % birth		-0.005 (0.237)		0.052 (0.282)
Reference life satisfaction		-0.057 (0.039)		-0.065 (0.044)
1-year lead effects				
ln(HH income)			0.007 (0.025)	-0.002 (0.029)
Employed (base)				
<i>Self-employed</i>			-0.030 (0.072)	0.006 (0.082)
<i>Unemployed</i>			-0.059 (0.083)	-0.068 (0.098)
<i>Retired</i>			0.050 (0.088)	0.036 (0.096)
<i>Family care</i>			-0.082 (0.067)	0.015 (0.077)
<i>Lt. sick / disabled</i>			-0.300*** (0.097)	-0.324*** (0.109)
<i>In education</i>			0.010 (0.078)	0.043 (0.093)

Table E5. Sparse estimates for fixed reference group (with lags and leads) (continued)

Married (base)				
Widowed			-0.549*** (0.164)	-0.529*** (0.190)
Divorced			-0.419*** (0.082)	-0.475*** (0.097)
Never married			-0.289*** (0.079)	-0.271*** (0.093)
ln(HH size)			-0.004 (0.055)	-0.052 (0.064)
Childbirth			0.229*** (0.057)	0.270*** (0.067)
Reference income			0.086 (0.255)	-0.056 (0.306)
Reference employed			-0.048 (0.346)	-0.223 (0.421)
Reference % disabled			-2.536*** (0.943)	-2.893*** (1.097)
Reference % widowed			-0.027 (1.260)	0.682 (1.431)
Reference % birth			0.197 (0.722)	0.548 (0.859)
Reference life satisfaction			-0.022 (0.130)	0.054 (0.148)
Memory of last year's life satisfaction				
1-year lagged effects				
ln(HH income)	0.019 (0.021)	0.027 (0.024)	0.033 (0.024)	0.042 (0.027)
Employed (base)				
Self-employed	-0.034 (0.061)	-0.020 (0.071)	0.065 (0.068)	0.096 (0.078)
Unemployed	-0.120* (0.065)	-0.082 (0.076)	-0.134* (0.077)	-0.096 (0.091)
Retired	-0.132* (0.077)	-0.151* (0.089)	-0.027 (0.088)	-0.042 (0.101)
Family care	-0.119** (0.055)	-0.095 (0.062)	-0.086 (0.062)	-0.066 (0.071)
Lt. sick / disabled	-0.320*** (0.080)	-0.252*** (0.092)	-0.301*** (0.091)	-0.223** (0.104)
In education	-0.014 (0.058)	-0.022 (0.070)	-0.005 (0.068)	-0.040 (0.082)
Married (base)				
Widowed	-0.993*** (0.137)	-0.511*** (0.163)	-1.088*** (0.169)	-0.667*** (0.199)
Divorced	-0.398*** (0.065)	-0.377*** (0.081)	-0.378*** (0.075)	-0.410*** (0.090)
Never married	-0.305*** (0.059)	-0.280*** (0.075)	-0.296*** (0.069)	-0.337*** (0.086)
ln(HH size)	-0.026 (0.047)	-0.103* (0.058)	-0.052 (0.054)	-0.147** (0.063)
Childbirth	0.096** (0.041)	0.135*** (0.048)	0.069 (0.045)	0.086 (0.055)
Reference income	0.350* (0.198)	0.049 (0.250)	0.147 (0.229)	-0.186 (0.284)
Reference employed	-0.382 (0.282)	-0.305 (0.366)	0.024 (0.324)	0.302 (0.419)
Reference % disabled	-0.793 (0.747)	-1.517 (0.949)	-1.201 (0.872)	-2.113* (1.102)
Reference % widowed	-0.562 (1.023)	-0.222 (1.297)	-0.413 (1.181)	-0.041 (1.460)
Reference % birth	-0.343 (0.527)	-0.105 (0.657)	-0.108 (0.652)	0.170 (0.789)
Reference life satisfaction	0.077 (0.105)	0.057 (0.119)	0.114 (0.119)	0.068 (0.134)
2-year lagged effects				
ln(HH income)				
Employed (base)				
Self-employed		0.153** (0.068)		0.116 (0.076)
Unemployed		-0.076 (0.074)		-0.073 (0.083)
Retired		0.139 (0.089)		0.116 (0.099)
Family care		-0.001 (0.061)		0.038 (0.068)
Lt. sick / disabled		-0.095 (0.095)		-0.013 (0.106)
In education		0.049 (0.069)		0.048 (0.080)
Married (base)				
Widowed		-0.785*** (0.152)		-0.791*** (0.173)
Divorced		-0.090 (0.074)		-0.106 (0.087)
Never married		-0.032 (0.067)		0.024 (0.078)
ln(HH size)		0.103** (0.051)		0.050 (0.058)
Childbirth		-0.008 (0.045)		-0.033 (0.050)
Reference income		0.071 (0.238)		-0.050 (0.264)
Reference employed		0.350 (0.312)		0.265 (0.359)
Reference % disabled		-0.082 (0.864)		-0.594 (0.977)
Reference % widowed		-1.747 (1.217)		-1.911 (1.399)
Reference % birth		-0.812 (0.665)		-0.096 (0.755)
Reference life satisfaction		-0.109 (0.123)		-0.077 (0.138)

Table E5. Sparse estimates for fixed reference group (with lags and leads) (continued)

Constant	0.691 (1.314)	1.134 (1.641)	0.165 (1.563)	-0.699 (1.936)
Scale shift				
contemporaneous effects				
ln(HH income)	0.061*** (0.022)	0.079*** (0.025)	0.073*** (0.026)	0.087*** (0.029)
Employed (base)				
<i>Self-employed</i>	0.064 (0.058)	0.090 (0.066)	0.053 (0.069)	0.050 (0.078)
<i>Unemployed</i>	-0.414*** (0.067)	-0.502*** (0.078)	-0.382*** (0.078)	-0.455*** (0.093)
<i>Retired</i>	0.162** (0.074)	0.146* (0.083)	0.136 (0.087)	0.096 (0.099)
<i>Family care</i>	-0.056 (0.053)	-0.084 (0.061)	-0.064 (0.062)	-0.129* (0.069)
<i>Lt. sick / disabled</i>	-0.203** (0.080)	-0.269*** (0.093)	-0.257*** (0.093)	-0.349*** (0.107)
<i>In education</i>	-0.044 (0.060)	-0.109 (0.072)	-0.090 (0.072)	-0.179** (0.085)
Married (base)				
<i>Widowed</i>	-0.728*** (0.137)	-0.701*** (0.157)	-0.593*** (0.181)	-0.544*** (0.202)
<i>Divorced</i>	-0.076 (0.063)	-0.056 (0.072)	0.048 (0.080)	0.053 (0.088)
<i>Never married</i>	-0.174*** (0.062)	-0.176** (0.072)	-0.102 (0.079)	-0.146 (0.089)
ln(HH size)	-0.112** (0.048)	-0.107** (0.054)	-0.123** (0.061)	-0.108 (0.069)
Childbirth	0.135*** (0.044)	0.134*** (0.051)	0.142*** (0.051)	0.154** (0.061)
Reference income	0.109 (0.209)	0.164 (0.237)	0.011 (0.253)	0.038 (0.283)
Reference employed	-0.192 (0.275)	-0.313 (0.329)	0.106 (0.366)	0.191 (0.438)
Reference % disabled	0.451 (0.776)	0.166 (0.895)	2.218** (0.980)	1.805* (1.087)
Reference % widowed	-0.974 (1.053)	-1.213 (1.191)	-1.299 (1.450)	-1.633 (1.557)
Reference % birth satisfaction	0.435 (0.543)	0.620 (0.689)	0.047 (0.665)	0.570 (0.836)
	0.051 (0.119)	0.020 (0.141)	0.006 (0.138)	-0.131 (0.165)
1-year leaded effects				
ln(HH income)			0.011 (0.024)	-0.002 (0.027)
Employed (base)				
<i>Self-employed</i>			-0.049 (0.069)	-0.013 (0.079)
<i>Unemployed</i>			0.029 (0.079)	0.031 (0.092)
<i>Retired</i>			0.041 (0.086)	0.051 (0.094)
<i>Family care</i>			-0.065 (0.063)	0.024 (0.073)
<i>Lt. sick / disabled</i>			-0.144 (0.092)	-0.190* (0.104)
<i>In education</i>			0.019 (0.073)	0.055 (0.087)
Married (base)				
<i>Widowed</i>			-0.246 (0.163)	-0.250 (0.189)
<i>Divorced</i>			-0.172** (0.076)	-0.243*** (0.088)
<i>Never married</i>			-0.162** (0.075)	-0.128 (0.088)
ln(HH size)			0.004 (0.053)	-0.039 (0.061)
Childbirth			0.127** (0.054)	0.156** (0.063)
Reference income			0.058 (0.246)	-0.093 (0.292)
Reference employed			-0.005 (0.333)	-0.133 (0.404)
Reference % disabled			-2.551*** (0.929)	-3.008*** (1.070)
Reference % widowed			0.210 (1.267)	0.753 (1.440)
Reference % birth satisfaction			0.615 (0.694)	0.848 (0.812)
			-0.040 (0.128)	0.071 (0.146)
Cutoff constants				
γ_1	0.262 (0.652)	0.870 (0.835)	0.484 (0.737)	1.358 (0.946)
γ_2	-0.873*** (0.025)	-0.866*** (0.028)	-0.876*** (0.028)	-0.872*** (0.031)
γ_3	-0.590*** (0.015)	-0.577*** (0.016)	-0.582*** (0.016)	-0.573*** (0.018)
γ_4	-0.439*** (0.010)	-0.435*** (0.011)	-0.434*** (0.011)	-0.426*** (0.012)
γ_5	-0.122*** (0.007)	-0.114*** (0.008)	-0.123*** (0.008)	-0.114*** (0.008)
γ_6	0.062*** (0.008)	0.079*** (0.008)	0.067*** (0.008)	0.084*** (0.009)
Observations	111,683	90,280	91,631	73,654

Note: All percentages divided by 100. Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E6. Inclusion of reference satisfaction variables

	Average (1)	Each individually (2)
Current life satisfaction		
ln(HH income)	0.055** (0.024)	0.056** (0.026)
Employed (base)		
<i>Self-employed</i>	0.085 (0.064)	0.076 (0.064)
<i>Unemployed</i>	-0.467*** (0.073)	-0.408*** (0.079)
<i>Retired</i>	0.175** (0.076)	0.187** (0.077)
<i>Family care</i>	-0.042 (0.057)	-0.030 (0.058)
<i>Lt. sick / disabled</i>	-0.239*** (0.088)	-0.121 (0.103)
<i>In education</i>	0.060 (0.065)	0.069 (0.065)
Married (base)		
<i>Widowed</i>	-0.930*** (0.143)	-0.923*** (0.145)
<i>Divorced</i>	-0.160** (0.073)	-0.126 (0.081)
<i>Never married</i>	-0.243*** (0.067)	-0.224*** (0.069)
ln(HH size)	-0.213*** (0.051)	-0.234*** (0.051)
Childbirth	0.179*** (0.047)	0.253*** (0.050)
Reference income	0.037 (0.196)	0.044 (0.198)
Ref. sat. income quintile		0.226*** (0.051)
Ref. sat. employment		0.325*** (0.051)
Ref. sat. marital status		0.259*** (0.058)
Ref. sat childbirth		-0.172 (0.108)
Ref. sat HH size		0.340*** (0.060)
Avg. reference satisfaction	1.127*** (0.088)	
Memory of last year's life satisfaction		
ln(HH income)	0.010 (0.021)	0.005 (0.023)
Employed (base)		
<i>Self-employed</i>	-0.041 (0.061)	-0.046 (0.062)
<i>Unemployed</i>	-0.093 (0.066)	-0.044 (0.071)
<i>Retired</i>	-0.120 (0.076)	-0.118 (0.077)
<i>Family care</i>	-0.116** (0.055)	-0.108* (0.055)
<i>Lt. sick / disabled</i>	-0.294*** (0.083)	-0.203** (0.094)
<i>In education</i>	-0.006 (0.058)	-0.002 (0.059)
Married (base)		
<i>Widowed</i>	-0.985*** (0.139)	-1.012*** (0.140)
<i>Divorced</i>	-0.363*** (0.066)	-0.378*** (0.072)
<i>Never married</i>	-0.291*** (0.060)	-0.301*** (0.062)
ln(HH size)	-0.023 (0.048)	-0.020 (0.049)
Childbirth	0.086** (0.041)	0.101** (0.044)
Reference income	0.238 (0.182)	0.269 (0.183)
Ref. sat. income quintile		0.064 (0.046)
Ref. sat. employment		0.108*** (0.040)
Ref. sat. marital status		0.011 (0.051)
Ref. sat childbirth		-0.047 (0.093)
Ref. sat HH size		-0.002 (0.053)
Avg. reference satisfaction	0.169** (0.075)	
Constant	3.518*** (0.945)	5.327*** (0.974)
Scale shift		
ln(HH income)	0.063*** (0.023)	0.065** (0.025)
Employed (base)		
<i>Self-employed</i>	0.061 (0.059)	0.062 (0.059)
<i>Unemployed</i>	-0.418*** (0.069)	-0.420*** (0.075)
<i>Retired</i>	0.164** (0.074)	0.162** (0.075)
<i>Family care</i>	-0.051 (0.054)	-0.050 (0.054)
<i>Lt. sick / disabled</i>	-0.207** (0.083)	-0.213** (0.098)
<i>In education</i>	-0.033 (0.061)	-0.032 (0.061)
Married (base)		
<i>Widowed</i>	-0.741*** (0.138)	-0.773*** (0.140)
<i>Divorced</i>	-0.088 (0.065)	-0.133* (0.073)
<i>Never married</i>	-0.179*** (0.063)	-0.195*** (0.065)
ln(HH size)	-0.114** (0.049)	-0.125** (0.050)

Table E6. Inclusion of reference satisfaction variables (continued)

Childbirth	0.143*** (0.045)	0.148*** (0.049)
Reference income	0.064 (0.188)	0.067 (0.189)
Ref. sat. income quintile		-0.023 (0.050)
Ref. sat. employment		-0.019 (0.050)
Ref. sat. marital status		-0.112** (0.054)
Ref. sat childbirth		-0.017 (0.110)
Ref. sat HH size		0.081 (0.057)
Avg. reference satisfaction	-0.068 (0.086)	
γ_1	0.984** (0.418)	3.117*** (0.428)
γ_2	-0.851*** (0.025)	-0.829*** (0.025)
γ_3	-0.569*** (0.015)	-0.551*** (0.015)
γ_4	-0.421*** (0.010)	-0.406*** (0.010)
γ_5	-0.107*** (0.007)	-0.096*** (0.007)
γ_6	0.075*** (0.007)	0.084*** (0.007)
Observations	118, 440	118, 440

Note: ‘Ref. sat. income’ measures mean reported life satisfaction of those in the pre-defined reference group (see section 4.1) who are additionally in the same quintile of the income distribution. Likewise, ‘Ref. sat. marital status’ measures mean reported life satisfaction of those in the predefined reference group who additionally share the same marital status. Variables ‘Ref. sat. employment’, ‘Ref. sat childbirth’, and ‘Ref. sat HH size’ are defined analogously. Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E7. Allowing for non-parallel shifts in scale use when adding individual averages (non-linear specification)

	Current sat. (1)	Memory of sat. (2)	Cutoff $\tau_{it,1}$ (3)	Cutoff $\tau_{it,2}$ (4)	Cutoff $\tau_{it,3}$ (5)	Cutoff $\tau_{it,4}$ (6)	Cutoff $\tau_{it,5}$ (7)	Cutoff $\tau_{it,6}$ (8)
ln(HH income)	0.102*** (0.023)	0.027 (0.021)	0.090*** (0.030)	-0.050* (0.026)	-0.025 (0.025)	0.018 (0.019)	-0.003 (0.013)	0.015 (0.013)
Employed (base)								
<i>Self-employed</i>	0.082 (0.065)	-0.001 (0.062)	0.068 (0.124)	-0.079 (0.175)	0.027 (0.125)	-0.007 (0.071)	0.050 (0.040)	0.024 (0.036)
<i>Unemployed</i>	-0.631*** (0.069)	-0.134** (0.064)	-0.327*** (0.125)	0.098 (0.198)	0.125 (0.124)	-0.301*** (0.074)	-0.177*** (0.051)	-0.057 (0.049)
<i>Retired</i>	0.159** (0.072)	-0.080 (0.071)	0.134 (0.100)	0.136 (0.109)	-0.074 (0.086)	-0.028 (0.060)	-0.080** (0.040)	-0.000 (0.034)
<i>Family care</i>	-0.072 (0.056)	-0.113** (0.054)	-0.020 (0.081)	0.052 (0.088)	0.061 (0.081)	-0.070 (0.055)	-0.137*** (0.037)	-0.015 (0.031)
<i>Lt. sick / disabled</i>	-0.558*** (0.080)	-0.301*** (0.075)	-0.238* (0.133)	0.381** (0.189)	-0.034 (0.124)	-0.201*** (0.077)	-0.180*** (0.063)	-0.138** (0.063)
<i>In education</i>	0.038 (0.065)	-0.022 (0.058)	0.103 (0.117)	-0.203 (0.203)	-0.069 (0.142)	0.038 (0.073)	-0.100** (0.042)	0.057 (0.037)
Married (base)								
<i>Widowed</i>	-0.910*** (0.125)	-0.912*** (0.121)	-0.659*** (0.163)	-0.165 (0.197)	0.139 (0.212)	0.114 (0.105)	0.017 (0.067)	-0.083 (0.065)
<i>Divorced</i>	-0.371*** (0.070)	-0.427*** (0.065)	0.041 (0.124)	-0.352** (0.158)	0.061 (0.146)	0.115 (0.072)	-0.010 (0.048)	-0.091* (0.048)
<i>Never married</i>	-0.303*** (0.067)	-0.296*** (0.060)	-0.164 (0.117)	0.104 (0.183)	-0.107 (0.128)	0.065 (0.067)	0.000 (0.038)	-0.046 (0.038)
ln(HH size)	-0.219*** (0.050)	-0.003 (0.047)	-0.166** (0.076)	0.073 (0.082)	-0.062 (0.068)	0.051 (0.042)	-0.006 (0.026)	-0.026 (0.025)
Childbirth	0.220*** (0.046)	0.086** (0.041)	0.199*** (0.077)	0.034 (0.090)	-0.045 (0.085)	-0.041 (0.050)	-0.054* (0.030)	0.006 (0.024)
Reference income	-0.117 (0.188)	0.255 (0.175)	0.276 (0.242)	-0.288 (0.227)	-0.091 (0.237)	-0.063 (0.152)	-0.098 (0.096)	0.019 (0.088)
Constant		0.057 (0.798)	-1.188 (0.771)	-3.440*** (1.050)	-1.516** (0.747)	1.094** (0.481)	-1.968*** (0.340)	-6.816*** (0.371)
Observations	118,440							

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E8. Allowing for non-parallel shifts in scale use when adding individual averages (non-linear specification, omitting first cut-off)

	Current sat. (1)	Memory of sat. (2)	Cutoff $\tau_{it,1}$ (3)	Cutoff $\tau_{it,2}$ (4)	Cutoff $\tau_{it,3}$ (5)	Cutoff $\tau_{it,4}$ (6)	Cutoff $\tau_{it,5}$ (7)	Cutoff $\tau_{it,6}$ (8)
ln(HH income)	0.009 (0.021)			-0.048* (0.025)	-0.029 (0.024)	0.014 (0.018)	-0.001 (0.013)	0.019 (0.013)
Employed (base)								
<i>Self-employed</i>	0.035 (0.114)			0.014 (0.169)	-0.003 (0.113)	-0.017 (0.067)	0.040 (0.039)	0.029 (0.036)
<i>Unemployed</i>	-0.276*** (0.100)			0.092 (0.169)	0.145 (0.109)	-0.274*** (0.068)	-0.184*** (0.048)	-0.065 (0.048)
<i>Retired</i>	0.013 (0.072)			0.101 (0.100)	-0.069 (0.080)	-0.041 (0.056)	-0.087** (0.039)	0.002 (0.034)
<i>Family care</i>	-0.052 (0.064)			0.047 (0.082)	0.055 (0.075)	-0.075 (0.051)	-0.139*** (0.035)	-0.016 (0.031)
<i>Lt. sick / disabled</i>	-0.273*** (0.103)			0.448** (0.175)	-0.031 (0.115)	-0.209*** (0.073)	-0.170*** (0.061)	-0.145** (0.061)
<i>In education</i>	-0.076 (0.099)			-0.231 (0.175)	-0.065 (0.125)	0.033 (0.069)	-0.097** (0.040)	0.054 (0.036)
Married (base)								
<i>Widowed</i>	-0.231** (0.111)			-0.209 (0.176)	0.196 (0.214)	0.125 (0.104)	0.010 (0.066)	-0.081 (0.064)
<i>Divorced</i>	-0.414*** (0.106)			-0.335** (0.155)	0.050 (0.136)	0.108 (0.066)	-0.001 (0.046)	-0.088* (0.047)
<i>Never married</i>	-0.161* (0.096)			0.109 (0.156)	-0.138 (0.116)	0.048 (0.062)	-0.000 (0.037)	-0.049 (0.037)
ln(HH size)	-0.029 (0.059)			0.101 (0.082)	-0.055 (0.067)	0.051 (0.040)	-0.008 (0.025)	-0.029 (0.025)
Childbirth	-0.016 (0.062)			-0.014 (0.083)	-0.062 (0.079)	-0.037 (0.048)	-0.060** (0.028)	0.004 (0.024)
Reference income	-0.385** (0.167)			-0.276 (0.203)	-0.073 (0.225)	-0.076 (0.144)	-0.108 (0.093)	0.006 (0.087)
Constant			-2.068*** (0.776)	-3.608*** (0.991)	-1.552** (0.710)	1.133** (0.471)	-1.943*** (0.334)	-6.778*** (0.370)
Observations	118,440							

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E9. Allowing for non-parallel shifts in scale use when adding individual averages (linear specification, omitting first cut-off)

	Current sat.	Memory of sat.	Cutoff	Cutoff	Cutoff	Cutoff	Cutoff	Cutoff
	(1)	(2)	$\tau_{it,1}$ (3)	$\tau_{it,2}$ (4)	$\tau_{it,3}$ (5)	$\tau_{it,4}$ (6)	$\tau_{it,5}$ (7)	$\tau_{it,6}$ (8)
ln(HH income)	0.010 (0.031)			-0.009 (0.026)	-0.032* (0.018)	0.013 (0.013)	-0.006 (0.011)	0.019 (0.012)
Employed (base)								
<i>Self-employed</i>	0.087 (0.109)			0.004 (0.097)	0.045 (0.061)	-0.004 (0.041)	0.025 (0.036)	0.032 (0.042)
<i>Unemployed</i>	-0.331*** (0.093)			-0.003 (0.081)	0.056 (0.055)	-0.130*** (0.039)	-0.156*** (0.037)	-0.055 (0.042)
<i>Retired</i>	0.039 (0.098)			0.081 (0.081)	-0.025 (0.057)	-0.031 (0.039)	-0.078** (0.034)	-0.016 (0.035)
<i>Family care</i>	-0.056 (0.092)			0.047 (0.080)	0.020 (0.054)	-0.052 (0.036)	-0.120*** (0.031)	-0.014 (0.034)
<i>Lt. sick / disabled</i>	-0.402*** (0.096)			0.070 (0.080)	-0.021 (0.055)	-0.095** (0.041)	-0.139*** (0.041)	-0.104** (0.046)
<i>In education</i>	-0.069 (0.106)			-0.107 (0.088)	-0.023 (0.061)	0.011 (0.042)	-0.097*** (0.036)	0.053 (0.044)
Married (base)								
<i>Widowed</i>	-0.334*** (0.114)			-0.055 (0.080)	-0.010 (0.064)	0.082 (0.052)	-0.002 (0.044)	-0.074 (0.051)
<i>Divorced</i>	-0.259*** (0.078)			0.028 (0.065)	-0.047 (0.047)	0.066* (0.038)	-0.025 (0.037)	-0.077 (0.047)
<i>Never married</i>	-0.223** (0.093)			0.003 (0.080)	-0.114** (0.056)	0.039 (0.040)	-0.005 (0.034)	-0.061 (0.042)
ln(HH size)	-0.123** (0.062)			-0.032 (0.052)	-0.028 (0.036)	0.044* (0.026)	-0.005 (0.022)	-0.037 (0.026)
Childbirth	-0.032 (0.074)			-0.036 (0.063)	-0.027 (0.042)	-0.028 (0.031)	-0.054** (0.024)	0.019 (0.028)
Reference income	-0.361 (0.251)			0.048 (0.206)	-0.062 (0.130)	-0.068 (0.090)	-0.126* (0.076)	0.017 (0.088)
Constant			-2.324*** (0.097)	0.548*** (0.061)	0.526*** (0.039)	0.686*** (0.031)	0.839*** (0.028)	1.009*** (0.038)
Observations	118,440							

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.