How much can we learn from happiness data? †

Caspar Kaiser* and Maarten C.M. Vendrik**

22nd June 2022

Abstract

Although survey data on happiness are increasingly used in economics and are now an official part of UK Government data collection, their reliability is a long-standing issue. Two recent studies argue that results based on such data might be reversible under certain kinds of monotonically increasing transformations of the associated happiness scale (Bond and Lang 2019; Schröder and Yitzhaki 2017). These studies raise a widely applicable question about the extent to which subjective data can be used for economic research. In response, we make three contributions. First, we derive a simple test of whether OLS coefficients can be reversed by relabelling the categories with which happiness is reported. In this context, we also deduce bounds for ratios of coefficients under any relabelling and discuss the commonalities between reversals in OLS regressions and reversals in ordered probit models. Second, we show that respondents would have to be using response scales in a strongly non-linear fashion for such reversals to appear. Yet, we point out, the existing empirical evidence suggests approximately linear scale use. Third, using several datasets, we empirically show that, in practice, sign reversals seem to be exceptionally rare or even impossible for a number of core socio-economic variables. Based on these analyses, we suggest new robustness tests that may be helpful for future work using subjective data.

JEL Codes: I31, C25

Keywords: ordinal reports, transformations of cardinal scales, happiness, subjective wellbeing, life satisfaction, utility, General Social Survey, German Socio-Economic Panel, Longitudinal Internet studies for the Social Sciences

---

† This paper is based on an earlier manuscript available on the Open Science Framework with a different title (see previous versions on https://osf.io/gzt7a/). We thank Timothy Bond and Kevin Lang for kindly sharing their replication files. We further thank Jan-Emmanuel De Neve, Richard Easterlin, Martijn Hendriks, Micah Kaats, Brian Nolan, Ekaterina Oparina, Andrew Oswald, Juan Palomino, Michael Plant, Alberto Prati, Marco Ranaldi, and participants of the Oxford Wellbeing Research Centre seminar series and the ISQOLS 2019 (Granada) Conference for helpful comments and suggestions. Funding from Nuffield College, the Department of Social Policy & Intervention, the Wellbeing Research Centre (Oxford), and the Institute for New Economic Thinking (via the ERC Grant no. 856455) is gratefully acknowledged.

* Institute for New Economic Thinking and Wellbeing Research Centre, University of Oxford. Email: caspar.kaiser@hmc.ox.ac.uk

** Department of Macro, International and Labour Economics and ROA, SBE, Maastricht University; IZA, Bonn; EHERO, Erasmus University, Rotterdam. Email: m.vendrik@maastrichtuniversity.nl

1 Introduction

There is now a vast literature on the determinants and consequences of life satisfaction, happiness, and subjective wellbeing. Over the last three decades, that literature has investigated a broad set of variables, ranging from classical economic variables like income, inflation, or unemployment, to more demographic variables like age, childbirth, marriage, or disability. For overviews of this large literature, see e.g. MacKerron (2012), Clark (2018), or Nikolova & Graham (2020). On the basis of this work, both national governments (e.g. HM Treasury, 2021) and international organisations (e.g. OECD, 2020) have begun to incorporate these measures into policy-making.

Most of this work relies on survey data, specifically questions like *"Taking all things together, how satisfied are you with your life?"*. Typically, answers are recorded using a small number of ordered response categories. OLS regressions or ordered probit models are then used to analyse such data. In the more prevalent case of OLS regressions, responses are normally recorded in their rank-order, i.e., assigning a "1" to the 1st response option, a "2" to the 2nd response option, and so on.

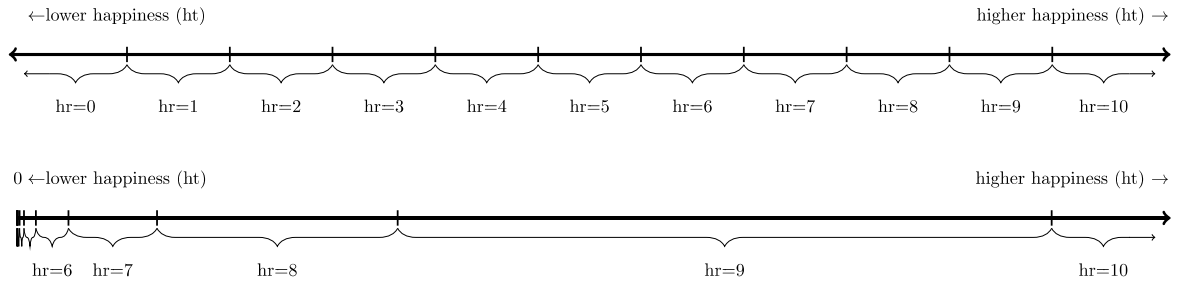
Strikingly, Schröder and Yitzhaki (2017; henceforth S&Y) showed that the signs of coefficients from such OLS regressions can sometimes be reversed by relabelling each of the observed response categories. For example, while labelling responses in their rank-order (i.e., 1, 2, 3, 4, etc.) may yield a positive coefficient for a given variable, applying a labelling in which differences between response categories are increasing (e.g., 1, 2, 4, 8, etc.) might yield a negative coefficient. Bond and Lang (2019; henceforth B&L) provide a similar result in an ordered probit setting, showing that results from such models can almost always be reversed.

The papers by B&L and S&Y raise an important issue. Their arguments apply to many kinds of subjective data and therefore have implications beyond the literature on subjective wellbeing. Moreover, subjective data is increasingly used in policy making and collected by national statistical agencies (ONS 2021). It is therefore vital to give a comprehensive assessment of the reasons why, and the conditions under which, such reversals are possible. This paper provides such an analysis.

Fundamentally, the issue we tackle has the following intuition: If wellbeing data merely records ordinal information, then the difference in underlying wellbeing between (say) the 1st and 2nd response category can be arbitrarily larger or smaller than the difference between (say) the 9th and 10th response category. In turn, when the effect of some variable X is positive in one part of the distribution of reported wellbeing, but negative in another, then the sign of the average effect of X can be flipped by rescaling the different parts of the response scale. For example, if the effect of X were negative at the bottom of the response scale, but positive at the top, we could adopt the assumption that differences between response categories are miniscule at the bottom of the scale and extremely large at the top. With that assumption, we could obtain a positive average effect of X. However, if we were instead to assume that differences between response categories are extremely large at the bottom of the scale, but miniscule at the top, we could obtain a negative average effect of X. Thus, so long as the effect of X is heterogenous across response categories, we can flip the sign of the average effect of X by changing our assumptions about how respondents interpret the meaning of each response category. Such heterogeneities can plausibly occur for any variable, making this a rather general concern about the potential limits of wellbeing data.

Our first contribution is to show how to test whether such reversals are possible for OLS regression coefficients. This test is quite simple and only requires estimating regressions of

Figure 1. Illustration of linear and non-linear scale use.



Caption: We use hr to denote reported happiness and ht to indicate underlying happiness (which is not directly observed). Linear scale use is shown at the top. An example of non-linear scale use is shown in the bottom. Specifically, in the bottom picture, differences between boundaries of response categories ("thresholds"), increase by a constant factor equal to 2.718 (i.e. e). Most of the variables we analyse require that respondents use the response scale in a way that is even more non-linear than shown here.

dummies indicating each possible dichotomisation of the response scale (i.e., dummies indicating whether the respondent chose the 1st category, the 2nd or lower category, the 3rd or lower category, etc., for all but the top response category). We also derive bounds on ratios of OLS regression coefficients for any positive monotonic transformation of how response options are labelled.

In that part of the analysis we assume that mean wellbeing within each response category does not depend on X. This assumption may be a shortcoming of the otherwise appealing OLS approach. However, given the widespread use of OLS in the field, it seems especially important to clarify the degree to which estimating such regressions can be informative. Unfortunately, without this assumption, reversals are almost always possible. For example, in the case of ordered probit models, which do not rely on this assumption, reversals are possible whenever estimated scale parameters depend on X. Practically, this is always the case in large samples.

As a second contribution, therefore, we demonstrate that for reversals to occur in practice, respondents must interpret response scales in a non-linear manner (see Figure 1). In contrast, we point out, the available empirical work suggests that respondents interpret response scales as approximately linear. Moreover, by analysing data in which happiness is recorded on both a continuous and a discrete scale, we present some first evidence indicating that mean wellbeing within response categories does not depend on standard socio-economic characteristics. This evidence provides some justification for the OLS approach.

Our third contribution is to present empirical evidence on the possibility and plausibility of reversals in life satisfaction data. In most cases, we find that reversals of OLS regressions coefficients are impossible. Similarly, reversals of effects estimated by ordered probit typically require respondents to use the response scale in a manner that is more non-linear than what is suggested by previous works and our evidence. However, we do find that ratios of coefficients, which are key for using wellbeing data in e.g. policy-analysis, are affected by relatively mild transformations. Hence, it appears that concerns about non-linear response scales are especially relevant when comparing magnitudes of effects across variables.

Overall, our analyses should be helpful for researchers looking to perform robustness tests in response to the questions raised by B&L and S&Y. For that purpose, our supplementary materials provide replication files and additional Stata codes to implement tests of whether sign reversals in OLS regressions are possible.[1] We also provide Stata code to find bounds on ratios of coefficients and code that numerically finds reversal conditions for ordered probit and OLS regressions. These, and our findings, should inform future work using subjective data.

1.1 Background

Before embarking on our analyses, it may help to contextualise the papers by S&Y and B&L into the wider literature on the feasibility of measuring and analysing subjective wellbeing.

The literature standardly distinguishes between affective measures of "happiness" and evaluative measures of "life satisfaction". Sometimes, a third concept of "eudemonic" wellbeing is considered, too (Clark and Senik 2011; OECD 2013). We are agnostic about the relative merits of these aspects of subjective wellbeing. See e.g. Sumner, (1996) or Kahneman et al. (1997) for discussion. There is also some debate on whether these measures capture respondents' feelings or cognitive judgments (Schwarz and Strack 1999). This likely depends on the particular questions asked and we are not committed to either view.

The intensity of these feelings/judgements is only directly accessible to respondents themselves. This is why we have to turn to self-reports. We will denote the underlying intensities of these feelings/judgements as ht ("**h**appiness **t**rue") and the rank-order coded self-reports as hr ("**h**appiness **r**eport"). To simplify our terminology, we will use the term "happiness" in the more theoretical sections (sections 2 and 3). However, we emphasise that our analyses also apply to other subjectively reported constructs, such as those relating to life satisfaction or subjective health.

Levels of ht are typically motivating the analysis of these self-reports. One standard motivation is classically utilitarian. A social planner might want to maximise total or average utility and might equate people's utility with their ht. This is a substantive ethical position which has been debated at length within economics and philosophy. For discussion, see e.g. Bentham (1789), Sidgwick (1874), Broome (1991), Harsanyi (1996), Ng (1997), Adler (2013), or Frijters & Krekel (2021). However, even if we are not merely interested in average or total utility, e.g. if we have a prioritarian social welfare function (Adler and Holtug 2019; Parfit 1997), or if we think that values other than welfare matter, we still want to learn about quantities of ht as a core consideration in social choice.

Of course, estimating the probability of responding with a certain response category of hr carries less epistemological baggage. However, for many applications, such analyses seem difficult to motivate when not interpreted in terms of being indicative of ht. That said, Bertrand & Mullainathan (2001) prominently questioned whether respondents do keep something like ht in their minds, suggesting that respondents might not have stable attitudes about their wellbeing. Citing previous work in psychology, they show that the order in which survey questions are posed, as well as a survey question's specific wording can have a strong impact on people's response behaviour. In addition, respondents often seem to exert little effort towards introspection and are prone to social desirability biases.

[1] See https://osf.io/sv4zb/.

Since Bertrand & Mullainathan's (2001) work, several studies responded to these issues.[2] For example, wellbeing reports correlate with regional compensating differentials (Oswald & Wu, 2010), neural activity (Kong et al. 2015; Urry et al. 2004), age patterns in anti-depressant use (Blanchflower and Oswald 2016), and suicide risk (Koivumaa-Honkanen et al. 2001). Self-reported wellbeing levels also correlate with third-party assessments of people's satisfaction (Schneider and Schimmack 2009), and responses are strongly autocorrelated, indicating test-retest reliability of the underlying measurement (Krueger and Schkade 2008; Lucas and Donnellan 2012; Michalos and Kahlke 2010). Moreover, individuals' (anticipated) self-reported wellbeing, and people's subsequent choices correlate strongly (Adler, Dolan, and Kavetsos 2017; Benjamin et al. 2012; 2014; Clark, Senik, and Yamada 2017; Perez-Truglia 2015).

A second type of issue is raised by questions over the comparability of wellbeing data (Adler 2013; Diamond 2008; Viscusi 2020). For example, Fleurbaey & Blanchet (2013) argue that questions about wellbeing are likely to be answered relative to some reference group. As another example, Barrington-Leigh (2021) argues that respondents with different levels of numeracy will differ in their reporting style. If so, the meaning of a given response category will not be the same across people. Several papers sought to empirically evaluate the importance of this kind of worry, using either data on 'vignettes' (Angelini et al. 2014; Kapteyn, Smith, and Van Soest 2010; Montgomery 2017) or people's memories (Fabian 2021; Kaiser 2022; Odermatt and Stutzer 2019). Although these studies often find differences in scale use across people and over time, the differences are typically too small to change substantive conclusions.

Having noted these largely separate issues, and to contain the scope of the paper, we set these issues aside for the analysis to follow. Instead, we assume that scale is common across respondents and focus on whether self-reported data can be interpreted cardinally.

Of course, B&L and S&Y are not the only previous works to examine the cardinal interpretation of wellbeing data. In a seminal paper, Ferrer-i-Carbonell & Frijters (2004) showed that OLS regressions, which make use of assumptions about the cardinal information in self-reported data, and ordered probit regression, which treat the data as merely ordinal, yield strikingly similar results. Their findings provided a justification for the wide-spread use of OLS in applied work on self-reported wellbeing. Some others have sought to give more direct empirical arguments in favour of a cardinal interpretation of wellbeing data. Those papers include Van Praag (1991), Layard et al. (2007), Oswald (2008), and Kristoffersen (2017) and are reviewed in section 4.

Providing complementary takes on the matter, there are a few other reactions to B&L's and S&Y's arguments. Bloem & Oswald (2021) and Chen et al. (2019) propose to focus on the median instead of the mean, pointing out that rankings of medians are invariant against all positive monotonic transformations of the underlying data. Bloem (2021) empirically shows that signs of estimates reported in several previous works using ordinal scales are robust to a range of convex and concave transformations. Finally, Liu & Netzer (2020) show how identification of rankings of mean happiness in two groups can be achieved with data on response times.

[2] This is not to say that there is no work on the validity of such data prior to this. Indeed, even in Easterlin's classic work from 1974, significant attention is given to these questions (see e.g. p.96-99 of his paper).

1.2 Outlook

In Section 2, we start with a concrete example of a reversal of an OLS regression coefficient. We then explain why this reversal occurs. Thereafter, we provide a general condition with which to test whether reversals of OLS regression coefficients are possible.

Section 3 first explores the general implications of allowing mean wellbeing to vary within each response category. Building on B&L, we subsequently analyse reversals of marginal effects in ordered probit models. We show that although ordered probit reversals are always possible, they depend, like OLS, on assuming that respondents interpret response scales in a non-linear manner.

In section 4, we then present evidence to suggest that respondents interpret response scales in a roughly linear fashion and show that mean wellbeing within response categories does not appear to strongly vary with standard socio-economic variables (which, as shown in section 2, turns out to be an assumption that can motivate the OLS approach).

Section 5 provides empirical evidence on the possibility and plausibility of reversals in life satisfaction data. We primarily use the German Socio-Economic Panel (SOEP). In the SOEP, life satisfaction is recorded with 11 possible response categories. B&L's analyses rely on questions with only three to seven response categories, which are less common in current empirical work. Such shorter scales are often viewed as being less informative than 10- or 11-point scales (OECD 2013). It is thus particularly useful to study whether plausible reversals can be obtained with such more typical data.

A final section concludes.

2 Reversals of OLS regression coefficients

2.1 Intuitions based on a special case

In this section we first give an example in which an OLS regression coefficient is reversed. We then explain why this reversal occurs. Our example is based on a simple special case, using aggregate data, only three response categories, and a single explanatory variable. In section 2.2 we generalise our analysis to questions with arbitrarily many response options and any number of individual-level covariates.

Here, we use responses to the question *"Taken all together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?"*. These are taken from the 1972–2006 biannual waves of the GSS. We regress these responses on yearly real log GDP per capita. Specifically, we begin by labelling the three response categories of the happiness question in their rank-order, i.e. with a 1 for "*not too happy*", a 2 for "*pretty happy*", and a 3 for "*very happy*". Labelling response categories in this manner is almost universally adopted in the literature that uses OLS regressions of such data. Column (1) of Table 1 shows that such a regression of rank-order coded reported happiness on log GDP per capita yields a negative but insignificant coefficient.[3]

However, if our data only contains ordinal information, we can also use any other positive monotonic transformation of this variable. For example, we could (say) assign a 1 to the 1st

[3] This could be interpreted as supporting the "Easterlin Paradox", which states that there are no long-term effects of changes in GDP per capita on mean happiness. However, tests of the paradox would require us to account for business cycle fluctuations. See Easterlin & O'Connor (2020), Kaiser & Vendrik (2019), and Stevenson & Wolfers (2008b) for discussion. The data used here is also used by B&L and is based on replication data by Stevenson & Wolfers (2008a).

Table 1. An example of reversing an OLS regression coefficient using GSS data.

	(1) Yearly mean of rank-order coded reported happiness	(2) Yearly mean of concave transformation of reported happiness
Log GDP per capita	-0.028	0.019
	(0.029)	(0.031)
Constant	2.200*** (0.006)	2.535*** (0.006)
Observations	26	26

Note: Results from an OLS regression of mean rank-order reported happiness (column 1) and an OLS regression of the mean of a concave transformation of reported happiness (column 2). Specifically, in column (1), the first response category is labelled with a 1, the second category with a 2 and the third category with a 3. In column (2), the first response category is labelled with a 1, the second category with a 2.6 and the third category with a 3. Yearly means of both variables are regressed on yearly log GDP per capita. Data are from the 1972–2006 waves of the GSS, as provided in the replication files of Stevenson & Wolfers (2008a). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

response category ("not too happy"), a 2.6 to the 2nd response category ("pretty happy"), and a 3 to the 3rd response category ("very happy"). Column (2) shows that a regression of the mean of such a concave transformation of reported happiness yields a positive coefficient. Hence, we obtain a reversal of the estimated OLS coefficient. Why does this occur?

To explain this, we need some notation. Let hr_{it} be respondent i's rank-order-coded happiness in year t. Here, we assign a 1 to the first category, a 2 to the second category, and a 3 to the third category. Let $\widetilde{hr}_{it} = f(hr_{it})$ be some positive monotonic transformation of hr_{it} and let $l = (l_1, l_2, l_3)$ denote a "labelling scheme", which records how each response category is coded. Each labelling scheme l corresponds to a particular transformation f (i.e. the elements in l are the image of f). For example, for rank-order coded reported happiness hr_{it}, as shown in column (1) of Table 1, we had $l = (1, 2, 3)$. Hence, in this case f is just the identity function. For the concave transformation in column (2) we had $l = (1, 2.6, 3)$, in which case f is concave.

In columns (1) and (2) we estimated regressions of the form:

$$\widetilde{hr}_t = \alpha + \beta y_t + \varepsilon_{it} \tag{1}$$

Here, $\widetilde{hr}_t$ denotes the yearly sample average of $\widetilde{hr}_{it}$, i.e. $\widetilde{hr}_t = N_t^{-1} \sum_i \widetilde{hr}_{it} = N_t^{-1} \sum_i f(hr_{it})$, where N_t records the number respondents in year t. Yearly log GDP per capita is denoted by y_t.

We can decompose mean reported happiness as the sum of shares in each response category times the label we attach to each response category. Let s_{kt} denotes the share of respondents in the k^{th} response category. We can then write:

$$\widetilde{hr}_t = s_{1t} l_1 + s_{2t} l_2 + s_{3t} l_3, \tag{2}$$

Differentiating equations (1) and (2) with respect to y_t, we get:

$$\beta = \frac{d\widetilde{hr}_t}{dy_t} = \frac{ds_{1t}}{dy_t} l_1 + \frac{ds_{2t}}{dy_t} l_2 + \frac{ds_{3t}}{dy_t} l_3 \tag{3}$$

Since the sum of all derivatives in equation (3) must equal 0 (as $s_{1t} + s_{2t} + s_{3t} = 1$), we can rewrite this as:

$$\beta = \frac{d\widetilde{hr_t}}{dy_t} = (l_1 - l_2)\frac{ds_{1t}}{dy_t} + (l_2 - l_3)\frac{d(s_{1t} + s_{2t})}{dy_t} \quad (4)$$

Here, the expression $d(s_{1t} + s_{2t})/dy_t$ denotes the derivative of the cumulative response share of the first and second response category (i.e. $s_{1t} + s_{2t}$) with respect to y_t.

The terms $l_1 - l_2$ and $l_2 - l_3$ are negative for all positive monotonic transformations of hr. Hence, if both derivatives in equation (4) are positive, the effect of y_t on mean reported happiness will remain negative for all positive monotonic transformation of hr. However, if the two derivatives do not share the same sign, the sign of β will depend on the relative magnitudes of the two derivatives and the relative magnitudes of $l_1 - l_2$ and $l_2 - l_3$. By setting equation (4) to zero and rearranging, we can see how much smaller the magnitude of $l_3 - l_2$ needs to be than the magnitude of $l_1 - l_2$ in order to achieve a reversal of β:

$$\frac{l_3 - l_2}{l_2 - l_1} = -\frac{ds_{1t}/dy_t}{d(s_{1t} + s_{2t})/dy_t} \quad (5)$$

In Appendix Table A3, we estimate ds_{1t}/dy_t and $d(s_{1t} + s_{2t})/dy_t$ using OLS regressions of s_{1t} and of $s_{1t} + s_{2t}$ on y_t. We estimate that $ds_{1t}/dy_t = -0.025$ and $d(s_{1t} + s_{2t})/dy_t = 0.054$. These results imply that when the difference between the 2nd and 3rd response category is smaller than difference between the 1st and 2nd category by a factor equal to $0.025/0.054 \approx 0.46$, a reversal of the OLS regression coefficient will occur. Thus, as shown in column (4) of Appendix Table A3, a concave transformation that assigns a 1 to the 1st category, a 2.37 to the 2nd category, and a 3 to the 3rd category yields a coefficient of zero (since $(3 - 2.37)/(2.37 - 1) \approx 0.46$). However, since our estimate of ds_{1t}/dy_t is statistically insignificant (with $p = 0.149$), a statistically significant reversal is not feasible.

Finally, notice that $d(s_{1t} + s_{2t})/dy_t = -ds_{3t}/dy_t$. Therefore, an increase in the cumulative shares of the 1st and 2nd response category entails a decline of the share of the 3rd response category. Since increases in GDP per capita moved some respondents from the 1st to the 2nd response category, some "not too happy" people became "pretty happy" (i.e. happier). Conversely, because increases in GDP per capita also moved some respondents from the 3rd to the 2nd response category, some "very happy" people became "pretty happy" (i.e. less happy). In that sense, the sign of the effect of GDP per capita on happiness is heterogeneous across the distribution of reported happiness. This may be seen as the underlying cause of the possibility of sign reversals.[4]

2.2 A *non-reversal condition* for OLS regressions

Building on the intuition of the preceding section, we now state a general *non-reversal condition* for OLS regression coefficients. This *non-reversal condition* applies to individual-level data for discrete dependent variables, and in settings where an arbitrary number of controls is included. Although we continue with our notation in terms of $\widetilde{hr_i}$ and continue to refer to it as transformed reported happiness, the condition is a statement about OLS regressions of discrete variables generally and

[4] S&Y and B&L emphasize violations of first-order stochastic dominance ("FOSD") between groups (cf. S&Y's Condition 1 and B&L's Section 2) as the root cause of the possibility of reversals. It is therefore worth noting that heterogeneity in the signs of derivatives on the (cumulative) response shares is equivalent to a violation of first-order stochastic dominance (FOSD) in the cumulative categories between groups with low or high *per capita* GDP.

is not specific to happiness data. Section 2.3 discusses the additional assumptions needed to apply this condition to happiness data.

To fix notation, consider the following individual-level regression equation:

$$\widetilde{hr}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i, \tag{6}$$

Here, $\boldsymbol{X}_i$ is a $1 \times M$ vector of explanatory variables with the first element set to 1 (to record a constant), and $\boldsymbol{\beta}$ is a $M \times 1$ vector of coefficients. Let $\widehat{\boldsymbol{\beta}}$ be the OLS estimate of $\boldsymbol{\beta}$ in equation (6). We are interested in learning whether each element $\hat{\beta}_m$ of $\widehat{\boldsymbol{\beta}}$ has the same sign for every positive monotonic transformation $\widetilde{hr}_i$ of hr_i.

To do so, define the dummies $hd_{k,i} \equiv \mathbb{1}(hr_i \leq k)$, where k indexes the ordered response categories, and consider the following regression equation:

$$hd_{k,i} = \boldsymbol{X}_i\boldsymbol{\beta}_k^{(d)} + \varepsilon_{k,i}^{(d)} \tag{7}$$

Let $\widehat{\boldsymbol{\beta}}_k^{(d)}$ be the OLS estimate of $\boldsymbol{\beta}_k^{(d)}$ in equation (7). We assume that all permissible labelling schemes have strictly increasing labels. More formally:

> **Assumption A1.** A labelling scheme $l = (l_1, l_2, \ldots, l_K)$ to record transformed reported happiness $\widetilde{hr}_i$ is *permissible* if and only if $l_1 < l_2 < \ldots < l_K$. Equivalently, only monotonically increasing transformations f are *permissible*.

This assumption is motivated by the idea that although happiness reports may not record happiness cardinally, these reports nevertheless record happiness in an ordinal sense. We can now state our *non-reversal condition*:

> **Proposition 1 (*non-reversal condition*).** Given Assumption A1, all but the first element of $\widehat{\boldsymbol{\beta}}$ obtained from a regression of $\widetilde{hr}_i$ on $\boldsymbol{X}_i$ will have the same sign for all permissible labellings schemes if and only if the corresponding elements of $\widehat{\boldsymbol{\beta}}_k^{(d)}$ have the same sign for all $k = 1,2, \ldots., K-1$.

A proof of Proposition 1 is given in Appendix A. A key part of the proof is that for any particular explanatory variable X_m, the associated estimated OLS coefficients $\hat{\beta}_m$ can be written as a sum over $\hat{\beta}_{k,m}^{(d)}$. Specifically, it turns out that $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$.

Substantively, the *non-reversal condition* allows us to test whether the sign of the estimated OLS coefficient $\hat{\beta}_m$ of some explanatory variable X_m remains the same under every labelling scheme. All we need to do is to estimate $K-1$ regressions of $hd_{k,i}$ and observe whether the corresponding coefficients $\hat{\beta}_{k,m}^{(d)}$ on X_m all carry the same sign. If they do, reversals are impossible.

Intuitively, the dummies $hd_{k,i}$ are the individual-level analogues to the cumulative response shares of section 2.1. Likewise, estimates $\hat{\beta}_{k,m}^{(d)}$ are the individual-level counterparts to the estimated derivatives $\mathrm{d}s_{1,t}/\mathrm{d}y_t$ and $d(s_{1t} + s_{2t})/\mathrm{d}y_t$ in section 2.1. When not all $\hat{\beta}_{k,m}^{(d)}$ share the same sign across k, the conditional association of hr_i with X_m is heterogeneous across response categories. Hence, as in section 2.1, such heterogeneities are the cause of the possibility of sign reversals.

For the special case where Equation (6) includes only a single explanatory variable, S&Y also provide a sufficient condition for the possibility of a sign reversal. Their condition is stated in terms of "*Line of independence Minus Absolute concentration"* curves ("LMA"; see Definitions 1 and 2 in Yitzhaki 1990), showing that if the LMA curve of hr with respect to a single explanatory variable X does not intersect 0, no reversals of estimated OLS regression coefficients are possible. Although S&Y hint at the possibility of extending their approach to multivariable regressions, they do not provide an explicit non-reversal condition in this setting. Following S&Y's suggestion for future research (p. 342), our *non-reversal condition* fills this gap.

Finally, for the special case of comparing two groups A and B our *non-reversal condition* can be shown to reduce to the condition $\sum_{q=1}^{k} s_{A,q} < \sum_{q=1}^{k} s_{B,q}$ or $\sum_{q=1}^{k} s_{A,q} > \sum_{q=1}^{k} s_{B,q}$ for all $k = 1, \dots, K-1$, where $s_{A,q}$ and $s_{B,q}$ respectively denote observed shares in groups A and B in category q. This condition is just a statement of first-order stochastic dominance (FOSD) in the cumulative distribution functions of hr between these groups. That rankings of means between groups are invariant under all positive monotonic transformations whenever FOSD holds is well known (e.g. Hadar and Russell 1969). Our non-reversal condition simply restates the need for FOSD in a multivariable OLS setting.

2.3 OLS-based reversals of effects on underlying happiness

The *non-reversal condition* is merely a statement about the properties of OLS regression coefficients. In order to make a stronger statement about whether our data can identify the effects of explanatory variables on underlying happiness ht_i we need two additional assumptions.

As in the introduction, let ht_i denote the unobservable cardinal quantity of underlying happiness, that we are ultimately interested in. A key assumption is then the following:

> **Assumption A2.** There exists some permissible labelling scheme l or, equivalently, some permissible transformation f of rank-order coded hr_i, such that $ht_i = f(hr_i) + \zeta_i = \widetilde{hr_i} + \zeta_i$ with $\mathrm{E}(\zeta_i | \boldsymbol{X}_i) = 0$.

This kind of assumption is rarely made explicit among studies that do use OLS. For a notable exception, see e.g. Layard et al. (2007). Substantively, Assumption A2 maintains that there is some set of labels for reported happiness so that underlying happiness ht_i is proxied by $\widetilde{hr_i}$ with a measurement error ζ_i that is mean-independent of $\boldsymbol{X}_i$. As shown in Appendix A2, Assumption A2 is satisfied whenever $\mathrm{E}(ht_i | hr_i = k; \boldsymbol{X}_i) = \mathrm{E}(ht_i | hr_i = k)$ holds (but not vice versa). Substantively, this stricter condition states that mean ht_i within each response category does not depend on $\boldsymbol{X}_i$. Section 4 will provide a tentative test of this stricter condition.

Additionally, we need a standard assumption adopted in much empirical research:

> **Assumption A3.** Underlying happiness ht_i is linearly related to $\boldsymbol{X}_i$ i.e., $ht_i = \boldsymbol{X}_i \boldsymbol{\beta} + \eta_i$, where $E(\eta_i | \boldsymbol{X}_i) = 0$.

Variants of Assumption A3 are standardly made in empirical research – both within and beyond the literature on subjective wellbeing. Nevertheless, although a linear specification can be quite flexible in the sense of allowing for an arbitrary number of covariates, including interactions and

higher-order terms, this assumption may not always be met. It may therefore be a useful avenue for future research to extend the present analysis to non-linear models.

With these assumptions in place, we can state the following:

> **Proposition 2**. Given Assumptions A1-A3, and whenever the conditional of the *non-reversal condition* holds, the sign of the effect of any variable $X_{i,m}$ in $\boldsymbol{X}_i$ on underlying happiness ht_i is unbiasedly and consistently estimated by an OLS regression of any permissible labelling scheme of $\widetilde{hr_i}$ on $\boldsymbol{X}_i$.

Proposition 2 follows from the linearity of the expectation operator and our assumptions about η_i and ζ_i. A proof is given in Appendix A2. Substantively, Proposition 2 makes clear that, under some assumptions, ordinal data on self-reported happiness can indeed be used to study the sign of effects of explanatory variables on underlying happiness.

2.4 Bounds on ratios of OLS coefficients

When assessing the substantive implications of estimates – especially regarding the relative importance of effects of explanatory variables – we are often more interested in ratios of estimated coefficients than in their magnitude. For example, the estimation of shadow prices (Bertram and Rehdanz 2015; Levinson 2012; Luechinger 2009) or equivalence scales (Biewen and Juhasz 2017; Borah, Keldenich, and Knabe 2019; Rojas 2007) principally relies on ratios of coefficients. As noted in e.g. Frijters et al. (2020), using wellbeing data for policy analysis also requires making comparisons across coefficient magnitudes.

Unfortunately, when effects are not perfectly homogenous across the distribution of reported happiness, ratios of coefficients are affected by positive monotonic transformations of reported happiness. Specifically, consider any two coefficients from the vector $\widehat{\boldsymbol{\beta}}$, say $\hat{\beta}_m$ and $\hat{\beta}_n$, which correspond to explanatory variables X_m and X_n. We can then state the following two propositions:

> **Proposition 3.** Given Assumption A1, the ratio $\hat{\beta}_m/\hat{\beta}_n = \rho$ obtained from an OLS regression of $\widetilde{hr_i}$ is the same for all permissible labelling schemes of $\widetilde{hr_i}$ if the corresponding ratios $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$ take the same value for all $k = 1, \dots, K-1$.

As will become clear in section 5, the conditional of Proposition 3 is practically never satisfied, meaning that ratios of coefficients are almost always affected by changing how reported happiness is labelled. However, we can bound how much this ratio will vary:

> **Proposition 4.** Given Assumption A1, the infimum of the ratio $\hat{\beta}_m/\hat{\beta}_n$ across all permissible labelling schemes for $\widetilde{hr_i}$ is given by the smallest of all estimated ratios $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$. Vice versa, the supremum of $\hat{\beta}_m/\hat{\beta}_n$ across all permissible labelling schemes for $\widetilde{hr_i}$ is given by the largest estimated ratio $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$.

Both propositions follow from the fact that $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$. Proofs are given in Appendix A. Practically, Proposition 4 enables evaluating the relative impact of explanatory variables on happiness across all positive monotonic transformations of $\widetilde{hr_i}$. We illustrate this in section 5.2.

2.5 OLS reversals using exponential transformations

The technique used to achieve reversals in our proof of the *non-reversal condition* makes use of highly irregular labelling schemes. A more regular approach to obtaining reversals is to impose that the differences between adjacent response categories grow or decline by some constant multiplicative factor $w > 0$. To do so, set $l_2 - l_1 = 1$ and impose that $(l_{k+2} - l_{k+1})/(l_{k+1} - l_k) = w$ for $k = 1, \dots, K-1$. Since $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$, we can then write $\hat{\beta}_m$ as a polynomial with coefficients $\hat{\beta}_{k,m}^{(d)}$:

$$\hat{\beta}_m = \sum_{k=1}^{K-1} -w^{k-1}\hat{\beta}_{k,m}^{(d)}\,. \tag{8}$$

For $w < 1$ differences between adjacent categories decline and for $w > 1$ differences increase. By Descartes' Rule of Signs (Weisstein 2021a), when $\hat{\beta}_{k,m}^{(d)}$ switches signs once across all k, equation (8) will be guaranteed to be zero for exactly one unique value of w.[5] Our empirical results of section 5 indicate that this is the prevalent case whenever any sign switches occur.

For $w > 1$, a labelling scheme like that used in equation (8) can be obtained with the convex transformation $\widetilde{hr} = f(hr) = e^{chr}$ where c is a positive constant. For $w < 1$, a concave transformation of the form $\widetilde{hr} = f(hr) = -e^{chr}$, where c is a negative constant, can be used. In both cases, we have $w = e^c$.[6] In an ordered probit context, B&L use the same transformations, facilitating the comparison between the ordered probit and the OLS approach.

Of course, these are not the only kinds of transformation that one might consider. For example, transformations of the form $\widetilde{hr} = hr^c$ would also yield a positive monotonic transformation that is either concave (for $0 < c < 1$) or convex (for $c > 1$). However, such transformations would not be guaranteed to yield a reversal for some c.

Transformations that are neither globally concave nor globally convex are possible, too. For example, Ng (2008) suggests that the boundedness of response scales may lead respondents to adopt scales in which differences in happiness between categories are large for top and bottom response categories, but small for categories in the middle. In other words, ht may be concave for low hr but convex for high hr. In contrast, Kristoffersen (2010) suggests the opposite, i.e. transformations in which differences in happiness between categories are small for extreme response options, but large for response categories in the middle. In principle, when there are at least two sign switches of $\hat{\beta}_{k,m}^{(d)}$ across k, such transformations with an inflection point could lead to simpler reversals of coefficient signs than are feasible with an exponential transformation. However, our empirical results of section 5 show that this never occurs in the SOEP data. Instead, we see at most one sign switch in $\hat{\beta}_{k,m}^{(d)}$ across k. In such cases with a single sign flip, any transformation with an inflection point would be more extreme than a transformation that is either

[5] Descartes' Rule of Signs generally states that the number of positive real roots of this equation is either equal to the number of times $\hat{\beta}_{k,m}^{(d)}$ switches sign across k or less than that by an even number. Hence, when the number of sign switches of $\hat{\beta}_{k,m}^{(d)}$ is even, no positive real root might exist. Moreover, roots of polynomials with degree higher than four cannot in general be found analytically (Weisstein 2021c). In these cases we search for roots numerically.

[6] Since $(e^{c(k+2)} - e^{c(k+1)})/(e^{c(k+1)} - e^{ck}) = e^c(e^{c(k+1)} - e^{ck})/(e^{c(k+1)} - e^{ck}) = e^c$.

globally concave or globally convex. For this reason, and to contain the length of the paper, we focus on exponential transformations.

3 Relaxing Assumption A2 and reversals based on ordered probit

As in noted in section 2.3, OLS regressions of happiness data can be motivated on the basis of Assumption A2. This assumption may not always be met. We therefore analyse the consequences of relaxing this assumption. Section 3.1 discusses the general implications of doing so and notes that even when scale use is linear, reversals of rankings of means between two groups remain possible. Section 3.2 discusses how reversals in ordered probit models – which do not rely on Assumption A2 – are obtained. In that section we also compare the OLS and ordered probit approach, showing that reversals in ordered probit models also require that respondents use response scales in a non-linear manner.

3.1 Relaxing Assumption A2

Assumption A2 may be seen as restrictive. To analyse the consequences of relaxing this assumption, we can replace assumption A2 by a weaker assumption:

> **Assumption A4.** There exists a collection of thresholds $\boldsymbol{\iota}$ such that $hr_i = k \leftrightarrow \iota_{k-1} < ht_i \leq \iota_k$, where $\iota_0 < \iota_1 < \ldots < \iota_{K+1}$.

Here, ι_k are thresholds that a respondent's level of underlying happiness ht_i needs to cross in order for respondents to report a certain response category. In keeping with the idea that happiness is recorded ordinally, these thresholds are strictly increasing, but no restrictions are made on how ht may vary within each category of hr.[7]

Consider two groups A and B, where group membership may be determined by combinations of values of the covariates in $\boldsymbol{X}$. As in section 2.2, let $s_{j,q}$ denote the share of members of group j responding with response category q. B&L then state a variant of the following (c.f. B&L, p. 1632):

> **Proposition 5:** Without maintaining Assumptions A2, whether $\mathrm{E}(ht_i|j = A) > \mathrm{E}(ht_i|j = B)$ is not identified from data on hr_i unless Assumption A4 and the following conditions hold:
>
> P5.I. $s_{A,1} = 0$
>
> P5.II. $s_{B,K} = 0$
>
> P5.III. $\sum_{q=1}^{k} s_{A,q} < \sum_{q=1}^{k-1} s_{B,q}$ for all $k = 2, \ldots, K-1$.

Mutatis mutandis, analogous conditions hold for identifying whether $\mathrm{E}(ht_i|j = A) < \mathrm{E}(ht_i|j = B)$. A proof of Proposition 5 in our notation is given in Appendix A5. The conditionals of this proposition are almost never satisfied. Hence, replacing Assumption A2 with Assumption A4 implies that we can almost never study differences in mean wellbeing between groups. The validity of Assumption A2 is thus key in considering how much we can learn from happiness data. In section 4, we therefore give an initial assessment of its validity. Although we do find preliminary evidence in favour of it being approximately met, more ought to be done to test it.

[7] Assumption A2 allows ht to vary within each category of hr, but in a very restricted way as implied by the identity $\mathrm{E}(\zeta_i|\boldsymbol{X}_i) = \sum_{k=1}^{K} s_k * \mathrm{E}(\zeta_i|hr_i = k; \boldsymbol{X}_i) = 0$ (see Appendix A2).

An interesting special case occurs when differences between thresholds $\iota_{k+1} - \iota_k$ are assumed to be constant for all k, i.e. when the response scale is assumed to be linear. As shown in Appendix B1, the conditions for identification of rankings of means are less demanding in this special case. Specifically, we only require that the difference in mean rank-order coded hr exceeds the value 1 between groups. For scales with few response options, e.g. in the GSS with only three response options, this is practically never satisfied. Even in the SOEP, where 11 response categories are available, a difference of a full point in hr between two groups occurs rarely. Indeed, as shown in Table 2 (see section 5), only the difference between the unemployed and all others (=1.273) exceeds this amount. Hence, upon relaxing Assumption A2, reversals of rankings of means are theoretically feasible even when respondents use the scale linearly.[8] This underlines the importance of Assumption A2. Nevertheless, as also shown in Appendix B1, increasing the number of available response categories makes identification easier to achieve. This may be an important consideration for future data-collection efforts.

3.2 Reversals based on ordered probit models

In the case where we do not want to maintain Assumption A2 we may consider the use of an ordered probit model. However, B&L show that signs of effects of explanatory variables are almost never identified with an ordered probit model. To understand why this is the case and to make a comparison with the OLS approach feasible, we will now reconstruct and extend B&L's argument. As part of this analysis, we provide a reversal condition for marginal effects and show that such reversals rely, as was also true in the OLS case, on assuming non-linear scale use.

To motivate the ordered probit model, we require the following :

Assumption A5.

A5.I There is a latent index hp_i ("**h**appiness **p**robit"), which is given by:

$$hp_i = \boldsymbol{X}_i\boldsymbol{\beta}^{(\boldsymbol{p})} + \varepsilon p_i, \tag{9}$$

where $\boldsymbol{X}_i$ contains a constant.

A5.II Reported happiness and the latent index are related as $hr_i = k \leftrightarrow \tau_{k-1} < hp_i \leq \tau_k$, where $\tau_0 = -\infty$, $\tau_K = \infty$, $\tau_1 = 0$, and $\tau_2 = 1$.

A5.III The error εp_i is normally distributed[9] with mean zero and standard deviation σ_i, where the log of σ_i is given by:

$$\ln(\sigma_i) = \boldsymbol{X}_i\boldsymbol{\beta}^{(\boldsymbol{s})}. \tag{10}$$

A5.IV There exists some positive monotonic function g, such that $g(hp_i) = ht_i$.

Jointly, these assumptions motivate a heteroskedastic ordered probit (HOP) model which can be estimated using maximum likelihood.

Some remarks about these assumptions may be useful. First, setting $\tau_1 = 0$, and $\tau_2 = 1$ selects a particular linear transformation of hp_i. We could alternatively drop the constants from equations (9) and (10) and explicitly estimate the thresholds τ_1 and τ_2 to yield an equivalent model. Second, it is typically assumed that $\boldsymbol{\beta}^{(\boldsymbol{s})} = \boldsymbol{0}$, i.e. that the error εp_i is homoscedastic. The HOP model

[8] However, these kinds of reversals cannot directly be obtained with ordered probit models. As shown below, sign reversals in ordered probit models instead depend on assuming non-linear scale use.

[9] εp_i could be logistically distributed, yielding an ordered logit model. All our arguments can be adapted to that case.

relaxes this assumption. The functional form in equation (10) is chosen for convenience, ensuring that σ_i is positive (Wooldridge 2010). Third, the function g, which relates the probit index to underlying happiness, is analogous to the function f discussed in section 2. There is nothing in the data that informs us about this function. Yet, most research implicitly assumes that g is linear.

When g is linear, the marginal effect of any particular variable X_m on mean ht, i.e. $\partial \mathrm{E}(ht_i|\boldsymbol{X})/\partial X_m$, is directly given by our estimate of $\beta_m^{(p)}$ (times a constant). However, when g is non-linear, $\partial \mathrm{E}(ht_i|\boldsymbol{X})/\partial X_m$ will depend on both $\beta_m^{(p)}$ and $\beta_m^{(s)}$. Sign reversals of $\partial \mathrm{E}(ht_i|\boldsymbol{X})/\partial X_m$ then become possible.

B&L focus on two kinds of exponential functions as choices for g, namely $ht_i = e^{chp_i}$ for some $c > 0$, and $ht_i = -e^{chp_i}$ for some $c < 0$. The former function is convex in hp and the latter function is concave in hp. We already introduced these functions when reversing OLS coefficients in section 2.5. When $c > 0$, the assumed model for ht_i is given by $ht_i = e^{c(\boldsymbol{X}_i\boldsymbol{\beta}^{(p)} + \varepsilon p_i)}$.[10] Here, ht_i will have a conditional distribution that is log-normal with mean (e.g. Weisstein, 2021b):

$$\mathrm{E}(ht_i|\boldsymbol{X_i}) = e^{c\mu_i + 0.5c^2{\sigma_i}^2}, \tag{11}$$

where $\mu_i \equiv \mathrm{E}(hp_i|\boldsymbol{X_i})$. In the case where $c < 0$, the conditional mean of ht_i is given by:

$$\mathrm{E}(ht_i|\boldsymbol{X_i}) = -e^{c\mu_i + 0.5c^2{\sigma_i}^2} \tag{12}$$

Notice that as the magnitude of c increases, the weight we place on the ${\sigma_i}^2$ term in determining $\mathrm{E}(ht_i|\boldsymbol{X_i})$ increases. As a consequence, and since $\mu_i = \boldsymbol{X_i}\boldsymbol{\beta^{(p)}}$ while ${\sigma_i}^2 = e^{\boldsymbol{X_i}\boldsymbol{\beta}^{(s)}}$, it follows for the case of (11) that if μ_i rises with $X_{i,m}$ ($\beta_m^{(p)}$>0), but ${\sigma_i}^2$ falls with $X_{i,m}$ ($\beta_m^{(s)}$<0), the effect of $X_{i,m}$ on $\mathrm{E}(ht_i|\boldsymbol{X_i})$ will change sign and become negative for sufficiently large c. Analogously for the case of (12), if $\beta_m^{(p)}$ and $\beta_m^{(s)}$ have the same sign, the effect of $X_{i,m}$ on $\mathrm{E}(ht_i|\boldsymbol{X_i})$ will change sign for a sufficiently negative c. This thought motivates the following proposition:[11]

Proposition 6: Given Assumptions A4 and A5, when $ht_i = e^{chp_i}$ for some $c > 0$ or $ht_i = -e^{chp_i}$ for some $c < 0$, the value of c at which the marginal effect of $X_{m,i}$ on $\mathrm{E}(ht_i|\boldsymbol{X})$ would switch its sign is given by $c = -\beta_m^{(p)}/e^{2\boldsymbol{X}_i\boldsymbol{\beta}^{(s)}}\beta_m^{(s)}$.

Proposition 6 can be obtained by differentiating equation (11) with respect to $X_{i,m}$:

$$\frac{\partial E(ht_i|\boldsymbol{X_i})}{\partial X_{i,m}} = e^{c\mu_i + 0.5c^2{\sigma_i}^2}\left(c\beta_m^{(p)} + c^2 e^{2\boldsymbol{X}_i\boldsymbol{\beta}^{(s)}}\beta_m^{(s)}\right) \tag{13}$$

Setting equation (13) to 0 and solving for c yields the condition in Proposition 6. The same expression is obtained when differentiating equation (12). When $\beta_m^{(p)}$ and $\beta_m^{(s)}$ have opposite (the same) sign, Proposition 6 predicts the sign-reversing value of c to be positive (negative). The only case in which no sign-reversing c exists occurs when $\beta_m^{(s)} = 0$.

[10] Generally, any non-linear choice for g implies a substantively different assumption about the functional form by which ht_i and $\boldsymbol{X}_i$ relate. The choice of g discussed here implies that $\ln(ht_i)/c$ is assumed to be linear in $\boldsymbol{X}_i$.

[11] Instead of deriving Proposition 6, B&L discuss a special case suitable for comparing two groups A and B. In that case, the reversal condition for the difference $\mathrm{E}(ht_i|A) - \mathrm{E}(ht_i|B)$ is given by $c = 2(\mu_A - \mu_B)/(\sigma_B^2 - \sigma_A^2)$.

A crucial characteristic of obtaining sign reversals of $\mathrm{E}(ht_i|\boldsymbol{X_i})$ in ordered probit models is that they rely – as was the case in the OLS setting – on assuming that individuals use response scales in a non-linear manner. To show this, our argument goes as follows: Initially, it appears as though the ordered probit model produces estimates of how people use the response scale. The estimated thresholds $(\tau_0, \tau_1, \ldots, \tau_K)$ seem to do just that. However, from Assumptions A4 and A5 we see that $hr_i = k \leftrightarrow g(\tau_{k-1}) < g(hp_i) = ht_i \leq g(\tau_k)$, where $g(\tau_k) = \iota_k$. Thus, to obtain the true thresholds ι_k, we must transform each estimated threshold τ_k with the function g. Our beliefs about scale use therefore depend on the estimated probit thresholds and our choice of g. This is analogous to the OLS case, where our beliefs about scale use depend on our choice of f.

In the special case of $ht_i = e^{chp_i}$, we obtain $(\iota_0, \iota_1, \ldots, \iota_K) = (e^{c\tau_0}, e^{c\tau_1}, \ldots, e^{c\tau_K}) = (0, 1, e^{c\tau_2}, \ldots, \infty)$. When differences between the estimated thresholds do not vary, and are thus equal to some constant $\Delta\tau$, then differences between the corresponding thresholds for ht_i, i.e. $(\iota_0, \ldots, \iota_K)$ will increase (for $c > 0$) or decrease (for $c < 0$) by a factor $e^{c\Delta\tau}$.[12] This is again analogous to the OLS case, where an exponential transformation implied that differences between response categories grow by a constant factor e^c.

Moreover, in Appendix B2 we show that ordered probit reversals are driven by sign heterogeneities of effects of explanatory variables across the distribution of hp_i. This is again in close analogy to the OLS case, where reversals made possible by heterogeneities across the distribution of hr_i.

Finally, in Appendix B3 we estimate a heteroskedastic ordered probit model on the same data as used in section 2.1. We find that a transformation $\widetilde{hp}_i = -e^{-0.73hp_i}$ suffices for a reversal. Such a transformation implies that differences between subsequent thresholds shrink by a factor $w = 0.48$. This is close to what we obtained in the OLS case, where $w = 0.46$. We also find that a transformation $\widetilde{hp}_i = -e^{-2.59hp_i}$ yields a marginal effect of log GDP per capita that is significant at the 5% level. In contrast, no statistically significant reversals were possible in the OLS case.

More generally, it remains a key difference that OLS reversals are rarely possible, while ordered probit reversals are almost always possible. Fundamentally, this is due to the fact that OLS maintains Assumption A2, while ordered probit does not. Consequently, the ordered probit approach requires first-order stochastic dominance (FOSD) in the conditional cumulative distribution function of hp_i (which occurs when $\beta_m^{(s)} = 0$), while OLS merely requires FOSD in the response shares for each level of hr_i (which occurs when all $\beta_{k,m}^{(d)}$ share the same sign).

4 Are response scales interpreted linearly? Is Assumption A2 valid?

Thus far, the discussion established that reversals in both the ordered probit and the OLS case are driven by assuming non-linear scale use. Hence, sections 4.1 and 4.2 will investigate whether strongly non-linear scale use is the norm. Thereafter, section 4.3 will probe the validity of Assumption A2, which we showed to be important in motivating the OLS approach.

[12] For $c > 0$, the differences in ht_i between thresholds ι_k and ι_{k-1} differ from the difference between thresholds ι_{k-1} and ι_{k-2} by a factor $(e^{c\tau_k} - e^{c\tau_{k-1}})/(e^{c\tau_{k-1}} - e^{c\tau_{k-2}})$. When $\tau_k - \tau_{k-1} = \Delta\tau$ for all $k = 1, 2, \ldots, K-1$ (i.e. excluding the outer thresholds τ_0 and τ_K) we can write $(e^{c(\tau_{k-1}+\Delta\tau)} - e^{c(\tau_{k-2}+\Delta\tau)})/(e^{c\tau_{k-1}} - e^{c\tau_{k-2}}) = e^{c\Delta\tau}$. By analogous reasoning, the same factor is obtained for $c < 0$. Of course, differences between estimated thresholds are never exactly constant. However, deviations from this approximation typically imply even more non-linear scale use.

4.1 Intuitions on scale use

The exponential transformations we consider entail that differences in happiness between response categories grow or decline by a factor $w = e^c$ (or $w = e^{\Delta\tau c}$ in the ordered probit case). Thus, these transformations move us from assuming (approximately) linear scale use to multiplicative scale use.

For response scales with just three categories, as in the GSS, this is not too problematic. For instance, to just reverse the coefficient on GDP per capita in the example of section 2, we only required a factor $w = e^{-0.78} = 0.46$, which implies that a jump in happiness intensity from the 2nd to the 3rd response category is approximately half as big as a jump from the 1st to the 2nd response category. It seems possible that respondents use the response scale in this manner.

However, the question on life satisfaction in the often-used German SOEP survey has 11 response categories. For this case, consider a transformation $\widetilde{hr}_i = e^{chr}$ with $c = 1$ as an initial benchmark. This value for c is less extreme than almost all of the sign-reversing c's obtained in our empirical results in section 5. With $c = 1$, the difference in $\widetilde{hr}_i$ between subsequent response categories grows by a factor of $e^1 \approx 2.78$. Consequently, the difference in $\widetilde{hr}$ between the top two levels of hr, (i.e. $hr = 11$ and $hr = 10$), is roughly $8{,}100$ times larger than the difference between the bottom two levels, $hr = 2$ and $hr = 1$ (since $(e^{11} - e^{10})/(e^2 - e^1) \approx 8{,}100$). Substantively, this would mean that there are almost no differences in respondents' underlying feelings across low levels of hr (see Figure 1 on page 3 for an illustration). To assess whether such scale use may indeed be common, the next section will survey the previous literature on this question.

4.2 Previous work on linearity of scale use

It seems that only four studies empirically investigated whether respondents use response scales for feelings or judgements linearly.[13] Each of these takes a different approach to the question. Yet, all of them conclude that scale use is not far from linear.

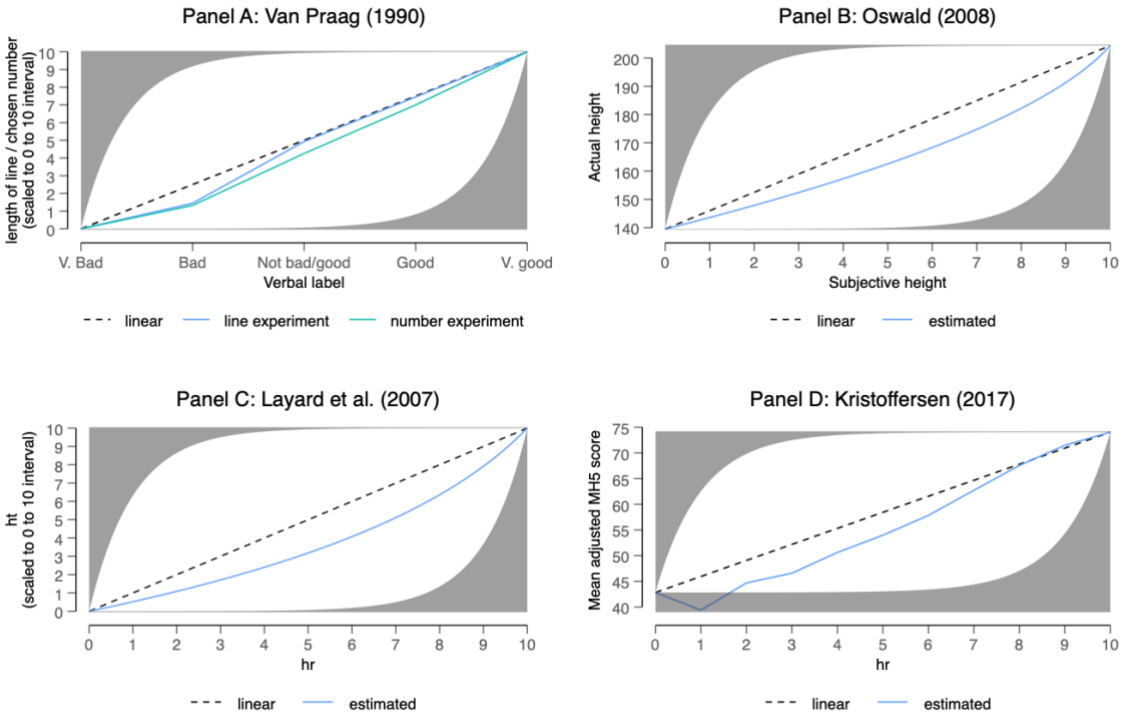
Initial empirical evidence is given in Van Praag (1991). There, Van Praag tested how individuals translate five ordered verbal labels (*very bad; bad; not bad; not good; good; very good*) into cardinal quantities in a context-free setting. In a first experiment he asked respondents to assign numbers between 1 and 1000 to each of the five verbal labels. In a second experiment, he asked respondents to produce lines of certain length corresponding to each of the verbal labels. As shown in panel A of Figure 2, Van Praag finds roughly linear scale use across both experiments.[14] This is especially clear when comparing his results with a multiplicative scale with $c = 1$, as also shown in Figure 2.

Oswald (2008) pursued a rather different idea. He asked a sample of respondents to report on their height using only a bounded slider. The extremes of the slider were labelled as "very short" and "very tall". He then regressed these responses on respondents' actual and squared height (as measured in centimetres). He finds a small but statistically significant negative coefficient on the

[13] To clarify further: in this section we are interested in whether the underlying cardinal quantities that respondents are asked about in survey questions (like ht) are linear in respondents' choices of response options (like hr).

[14] Together with Kapteyn and Van Herwaarden, Van Praag also made a theoretical argument in favour of respondents adopting a linear response scale (Kapteyn 1977; Kapteyn, Van Praag, and Van Herwaarden 1978; Van Praag 1971; 1991; 1993). In particular, they show that a linear response scale is minimising the loss in information resulting from discretizing a continuous quantity (like ht) with a finite number of response categories. Similar arguments are also offered by Parducci (1995) and Plant (2020).

Figure 2. Previous evidence on linearity of scale use.



Note: Solid lines indicate estimates of scale use obtained in previous work. Dotted lines show linear scale use. Shaded regions indicate scales implied by exponential transformations with $c < -1$ or $c > 1$. Panel A shows Van Praag's results (1990), specifically his Table 1. Panel B is based on Oswald (2008), and shows the result from inverting the equation displayed on page 371. Panel C shows the result of Layard et al. (2007), specifically the lower left panel of Figure 4. Panel D shows Kristoffersen's results (2017), specifically her Table 5.

squared term. In turn, when inverting the equation he estimated, this implies a small amount of convexity when transforming subjectively reported height into actual height. However, as shown in panel B of Figure 2 this estimate is much closer to linearity than a multiplicative scale with $c = 1$. Moreover, when distinguishing between genders, the squared term is no longer significant, suggesting that the observed convexity in the pooled sample may have been driven by a reporting difference across genders.

Layard et al.'s (2007) strategy to assess linearity is different yet again. Using SOEP data, they estimate an OLS regression of rank-order-coded life satisfaction on a wide set of explanatory variables and individual fixed effects. They then assume that the error of a similar model with actual life satisfaction (i.e. ht) as the dependent variable, is homoscedastic. Any heteroskedasticity in their OLS regression of rank-order coded hr would then indicate a non-linear response scale. They indeed find the residual variance to be larger for low predicted hr than for high hr. As they show, if their assumption of homoscedasticity with respect to ht indeed holds, this pattern implies that the response scale is convex. However, as illustrated in panel C of Figure 2, the amount of convexity they infer is much less than that of our benchmark scale with $c = 1$.

Kristoffersen (2017) assumes that psychometrically adjusted scores from the MH5 index of mental health are a cardinal measure of ht. Using Australian HILDA data, she regresses these MH5 scores on dummies for each of the 11 response categories of the life satisfaction question asked in the HILDA survey. As shown in panel D of Figure 2, she finds a largely linear pattern, albeit with the

dummy for $hr = 1$ being an outlier. Nevertheless, if her assumption of the cardinality of the MH5 index holds, her results also suggest that strongly non-linear scales are unlikely to be the norm.

Of course, each of these papers had to make assumptions about the relationship between observable quantities and ht. These assumptions may not be met: In Van Praag's (1991) experiments, respondents' behaviour with respect to context-free verbal labels may not be representative of how they might use scales to report their happiness. Similarly, Oswald's finding may be particular to people's subjective height, and might not generalise. Layard et al.'s (2007) homoscedasticity assumption may not hold, and Kristoffersen's (2017) assumption that psychometrically adjusted MH5 scores are cardinal has been questioned (cf. Bond & Lang, 2013).

Unfortunately, since intensities of subjective feelings and judgements are unobservable, none of these assumptions can be tested. Nevertheless, despite each paper making fundamentally different *sorts* of assumptions, the results from each paper suggest that such response scales are interpreted as approximately linear. Hence, in order for strongly non-linear scale use to be the norm, each of these studies' assumptions would have to fail in the same direction. Therefore, in so far as these studies present convergent evidence, we should place less credence on strongly non-linear scales.

However, the reviewed studies are based on samples from Dutch, German, or English-speaking countries. Since it is possible that linearity of scale use depends on respondent's particular language culture, research assessing intercultural and interlinguistic differences in scale use would be an important next step (see e.g. Angelini et al. (2014) for initial work in this direction).

Moreover, the evidence shown here only concerns questions with at least five categories. This leaves open that scales with three or four response options, such as those in the GSS and World Values Survey, are interpreted in a strongly non-linear manner. In Appendix B4 we investigate this possibility by comparing response behavour for ten and eleven-points scales with behaviour for three and four-points scales. Using comparable samples for the US and several European countries, we find that although three-points scales may be interpreted non-linearly, this does not appear to be the case for four-points scales.

4.3 A tentative test of Assumption A2

Based on the previous literature, it seems that approximately linear scale use is a reasonable description of how respondents answer subjective survey questions. However, whether Assumption A2 holds – which would motivate the OLS approach – does not appear to have yet been investigated. By utilising data in which respondents reported their data on both a discrete and a continuous scale, this section begins to fill this gap

Specifically, we use the March and April 2011 waves of the Dutch LISS panel, which surveys a representative sample of the Dutch population. In March, one randomly selected half of respondents reported their happiness on a ten-points discrete scale. All other respondents reported their happiness on a continuous scale, using a slider on a computer screen.[15] Although the slider appeared as continuous to respondents, the data was only recorded with a resolution of 100 distinct values (which we scale to range from 1 to 10, like the discrete scale). In April, roles were reversed and the first half of respondents answered on the continuous scale and the other half answered on

[15] The question reads *"Alles bij elkaar genomen, hoe gelukkig zou u zeggen dat u bent?" (Taking all things together, how happy would you say you are?*), with extremes labelled *"helemaal ongelukkig"* (*completely unhappy*) and *"helemaal gelukkig" (completely happy).* This data was originally collected for a project by Raphael Studer and Rainer Winkelmann (see Studer, 2012).

the discrete scale. Thus, for a total of 8,538 respondents we observe their happiness on both types of scales. The LISS data also contains information on several socio-economic characteristics. Among others, these include household income, whether the respondent is employed, is married, has children, and the respondent's disability status.

Recall from section 2.2 that Assumption A2 is satisfied if mean ht does not vary with X within each response category of hr. If this condition were to hold in the data at hand, we should expect that the mean of continuously reported happiness does not vary across socio-economic characteristics within each of the ten discrete response categories. We test this by estimating a regression of continuously reported happiness on dummies for each of the ten discrete response options. Each of these dummies are interacted with each of the aforementioned socio-economic characteristics. If these interactions terms are small and not statistically significantly different from zero, we interpret this as evidence in favour of Assumption A2.

The results from this exercise are presented in Figure 3. We see that within each discrete response category, predicted levels of continuous happiness are relatively homogenous across socio-economic characteristics. A hypothesis test of equality in predicted continuous happiness across groups cannot be rejected in most cases.[16] This is tentative evidence in favour of Assumption A2.

However, there are very few respondents who chose discrete options 1-4 (2.1%), which is the reason for the wide confidence intervals associated with these categories. Indeed, for several subgroups in the sample, no observations were available for the first and second response categories, making a test of Assumption A2 impossible for these categories. Therefore, only results for categories 3 to 10 are shown in Figure 3. Although this is not ideal, the fact that so few respondents (or none) are observed in these categories implies that the extent to which Assumption A2 is satisfied in these categories has less practical importance than for categories with many respondents.

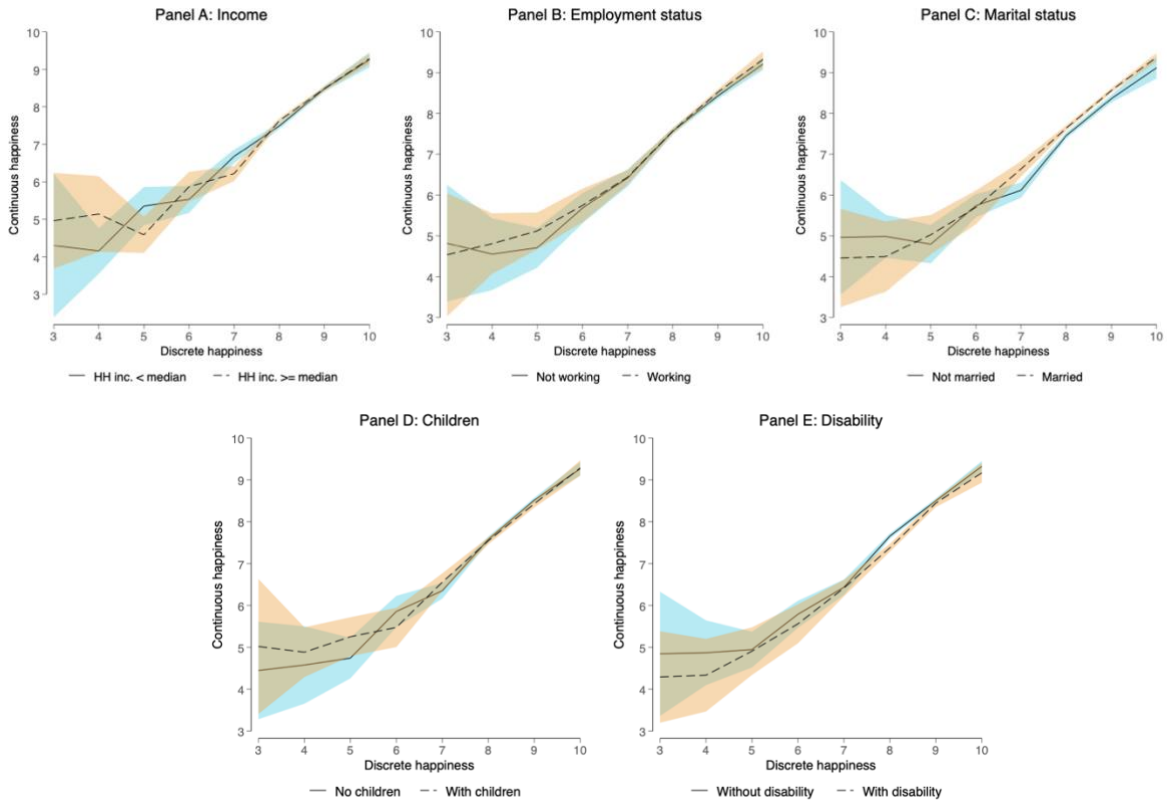
Moreover, for violations of Assumption A2 to facilitate reversals of OLS regression coefficients, we require that groups with higher mean discrete hr have lower continuous hr within each discrete response category (see Appendix B). This not the case in Figure 3. Instead, we see a largely inconsistent pattern. At most, groups with higher discrete hr tend to also have greater continuous hr within each discrete category. Such violations of Assumption A2 bias OLS coefficients from regressions of discrete hr towards zero, but cannot reverse their signs. Indeed, the only difference between regressions of continuous hr and discrete hr is that coefficients on the former are slightly larger than those on the latter (see Appendix Table A4).

Violations of Assumption A2 thus seem to be mild in our data. This result gives some justification for the use of OLS regressions and our *non-reversal condition*. We would nevertheless welcome the collection of more data using a continuous slider to see if our observations can be replicated in other data, e.g. in other large-scale surveys like the German SOEP.

Finally, for discrete response options larger than 4, continuous hr increases approximately linearly in discrete hr. This is also shown in Studer (2012). For response options lower or equal to 4, we observe that mean continuous hr increases less steeply, but this non-linearity is not as pronounced

[16] This test is rejected at the 5% level for the 7th and 8th category when comparing income above or below the median, for the 7th, 8th, and 9th category when considering marital status, and the 8th category when considering disability.

Figure 3. Continuously reported happiness hr does not systematically vary with socio-economic variables within each discrete category of hr.



Note: Based on an OLS regression using LISS data, each panel shows predicted values of continuous hr conditional on responding with a certain discrete response category of hr and conditional on being in a certain socio-economic group. In each panel, all other variables are set to their mean. Within each discrete category, mean continuous happiness does not tend to systematically vary with socio-economic variables, lending support to Assumption A2. 95% confidence intervals are given by the shaded regions (based on robust standard errors).

as would be the case for a multiplicative scale with $c = 1$. Hence, if we are willing to assume that the continuous scale allows respondents to report their ht cardinally (up to a linear transformation), this could be interpreted as additional evidence against strongly non-linear scale use in discrete scales. This idea is pursued further in Appendix B4.

5 Empirical Applications

We now turn to further assessing the empirical relevance of the points of the preceding sections. Primarily, we do so by evaluating the possibility and plausibility of reversals for a range of socio-economic variables using waves 1 (1984) to 32 (2015) of the German Socio-Economic Panel (SOEP), which is among the most commonly used dataset in empirical work on happiness. However, we can largely replicate our conclusions using the LISS and GSS surveys.

Similar to the previous section, our explanatory variables of interest are household income, unemployment, marriage, having children, and self-reported disability. These variables are also similar to those investigated by B&L. Answers to the question *"How satisfied are you with your life, all things considered"* (*"Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?")* are used as our dependent variable. This question is typically taken to elicit evaluative judgements from respondents about

their lives as a whole (OECD 2013; Pavot and Diener 2008). Answers are recorded with eleven response categories, here labelled from 1 to 11.

For income we use log net (post-tax) household incomes, deflated to 2005 prices. We equivalize incomes using the modified OECD scale.[17] Regarding unemployment, we code a dummy that is 1 when a person reports to be unemployed, and 0 for any other possible employment status. We code similar dummies for being married, living with children in the household, and reporting a disability. Next to reporting results in which these variables are entered separately, we also report results in which all variables are entered jointly along with a set of additional control variables. The additional control variables include region and wave dummies, age (linear and squared), a tertiary education dummy, a home- ownership dummy, log(household size) and log(1+working hours).

We first present OLS results. Thereafter, we focus on results from the ordered probit model.

5.1 OLS reversals using relabelling

Table 2 shows results for pooled and fixed-effects OLS regressions of hr_{it} on each explanatory variables of interest.[18] Column (1) shows results from separate regressions in which each variable is entered individually (being married and having children are always entered jointly), column (2) shows results from a pooled regression in which all variables of interest, along with the additional controls discussed above, are entered jointly. Column (3) adds individual fixed effects.

In all specifications, household income, being married, and having children are associated with higher life satisfaction, while unemployment and reporting a disability are associated with lower life satisfaction. Accounting for fixed effects generally reduces the magnitudes of our estimates.

To evaluate whether the sign of these coefficients can be reversed, we estimate OLS regressions of $hd_{k,it}$ for $k = 1,2,\dots,10$ when entering variables separately, when including controls, and when adding fixed effects. Figure 4 illustrates our results. For most variables and specifications, estimates of $\hat{\beta}_{k,m}^{(d)}$ have the same sign across all k. In these cases, reversals are impossible with the data at hand. In other words, the sign of the regression coefficient $\hat{\beta}_m$ will remain the same under all positive monotonic transformations of hr_{it}.

However, there are four exceptions to this. First, when failing to include controls, the coefficient $\hat{\beta}_{10,income}^{(d)}$ for income has a positive sign (though its magnitude is small), while the coefficients $\hat{\beta}_{k,income}^{(d)}$ for $k = 1,2,..,9$ all have a negative sign. Hence, a sufficiently convex transformation in which the difference between labels l_{11} and l_{10} is much larger than the differences between all other labels can reverse the sign of the overall effect of income $\hat{\beta}_{income}$. A numerical search shows that a multiplicative scale in which spaces between adjacent response categories grow by a factor $w = 24.1$ (with implied $c = \ln(24.1) = 3.18$) is required to achieve such a reversal.

[17] We exclude respondents in the top and bottom percentiles of the income distribution since there may be substantial measurement error in these observations (Berthoud and Bryan 2011). However, when we include these respondents, results are nearly unchanged.

[18] Ferrer-i-Carbonell & Frijters (2004) showed that time-invariant unobserved heterogeneity, due to e.g. individual personality traits (Boyce, 2010), biases pooled regression of hr. Fixed-effects regressions are thus standard in the literature. Unfortunately, a fixed-effects estimator is not available for ordered probit. In contrast, by demeaning each regression of $hd_{k,it}$ over respondents i, our non-reversal condition is directly applicable to the fixed-effects model.

Table 2. An application of the *non-reversal condition* for several socio-economic variables available the SOEP data.

	(1) No controls	(2) Full controls	(3) Full controls, with fixed effects
Log HH income	0.691*** (0.011)	0.568*** (0.012)	0.296*** (0.011)
	reversal occurs at c=3.18	**reversal impossible**	**reversal impossible**
Unemployed	-1.273*** (0.019)	-0.917*** (0.018)	-0.638*** (0.015)
	reversal impossible	**reversal impossible**	**reversal impossible**
Married	0.189*** (0.012)	0.290*** (0.013)	0.168*** (0.014)
	reversal impossible	**reversal impossible**	**reversal impossible**
Children	0.175*** (0.012)	0.132*** (0.012)	0.008 (0.012)
	reversal impossible	***reversal occurs at c=−2.83***	***reversal occurs at c=0.13***
Disability	-0.857*** (0.021)	-0.766*** (0.020)	-0.306*** (0.018)
	reversal impossible	**reversal impossible**	***reversal occurs at c=2.53***
Respondents	77,039	77,039	77,039
Observations	557,999	557,999	557,999

Note: All coefficients are obtained from OLS regressions of rank-order coded hr. The results of column (1) are based on separate regressions for each explanatory variable. The possibility of reversals is assessed based on OLS regressions of $hd_{k,it}$ for $k = 1, 2, \ldots, 10$ (see Figure 4 and Table A5 for results). Where reversals are possible, just-reversing c values have been obtained numerically. Model titles indicate the specification estimated in each column. Data are from the 1984-2015 waves of the SOEP. Standard errors in parentheses (clustered by respondents). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Second, the effect of having children can be reversed in a pooled regression when including controls. Here, the sign of $\hat{\beta}^{(d)}_{1,children}$ is positive while all other $\hat{\beta}^{(d)}_{k,children}$ are negative. Hence, a sufficiently concave transformation may yield a reversal. However, the magnitude of $\hat{\beta}^{(d)}_{1,children}$ is small and hardly visible in Figure 4. This entails that we a need a rather extreme transformation. Indeed, a transformation $\widetilde{hr}_{it} = -e^{chr_{it}}$ with $c = -2.83$ or lower is needed, which corresponds to differences between adjacent response categories shrinking by a factor $w = 0.06$.

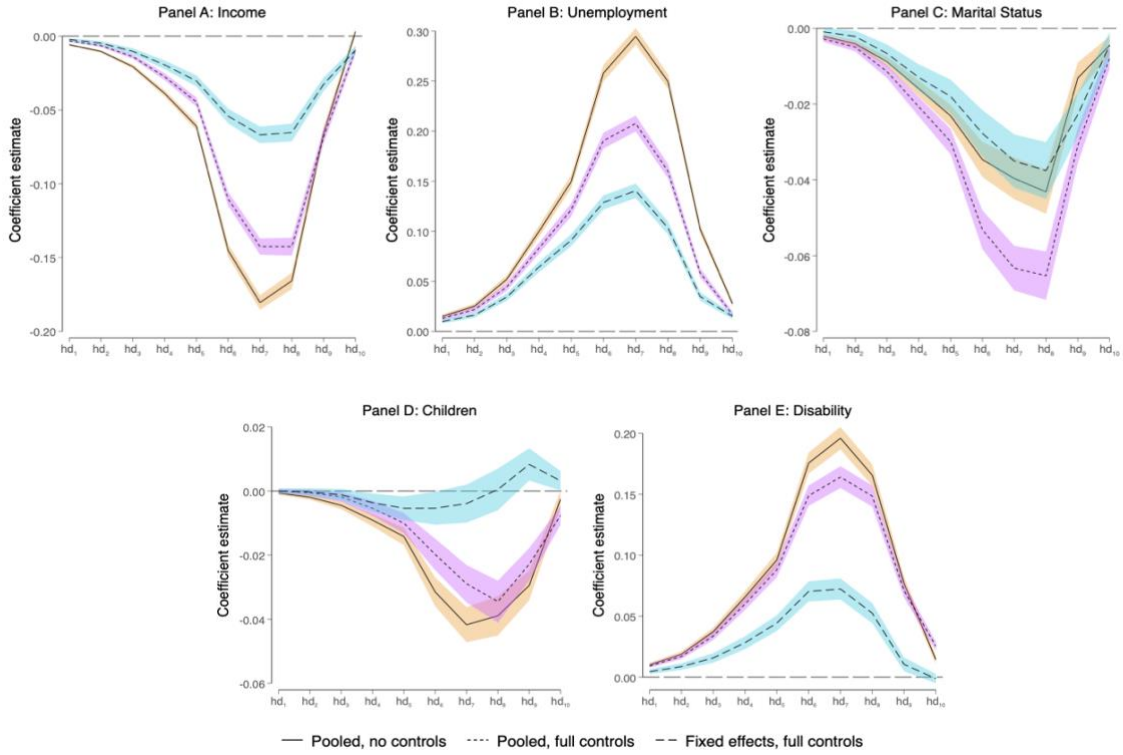
Third, in a fixed-effects regression with full controls, the sign of $\hat{\beta}^{(d)}_{k,children}$ is negative for $k \leq 7$, but positive for $k > 7$. In the regression of rank-order hr, the corresponding overall coefficient $\hat{\beta}_{children}$ was positive. Hence, a mild convex transformation with $c = 0.13$, corresponding to a factor $w = e^{0.13} = 1.14$ is sufficient for a sign reversal.

Fourth, the coefficient $\hat{\beta}^{(d)}_{10,disability}$ for disability in a regression with fixed effects is just negative, while all other coefficients are positive. Hence, a sufficiently convex transformation can yield a reversal. In this case, $c = 2.53$ corresponding to $w = e^{2.53} = 12.5$, would be sufficient.

The effects of unemployment and marriage cannot be reversed in any of our specifications. We thus conclude that the common finding that unemployment is associated with lower life satisfaction, and that marriage is associated with higher satisfaction is especially robust.

How likely are any of the theoretically possible reversals? In light of the arguments given in section 4, scales with c values in the order of at least 1 ($r \geq e^1 = 2.72$) or at most -1 ($r \leq e^{-1} = 0.37$) appear inconsistent with previous work and our evidence based on LISS data. Almost all the just-sign-reversing transformations found above are outside these values. The only exception was the effect of having children in a fixed-effects regression (where $c = 0.13$). Such a transformation is reasonably close to linear. We thus conclude that while reversals are possible for several variables

Figure 4. Illustration of the *non-reversal condition* (coefficient estimates for each regression of $hd_{k.it}$).



Note: Each line shows coefficient estimates for different specifications of OLS regressions of $hd_{k,it}$ using SOEP data. The *non-reversal condition* is not satisfied when a given line crosses zero. This is rarely the case. Shaded regions show 95% confidence intervals (based on standard errors clustered by respondents).

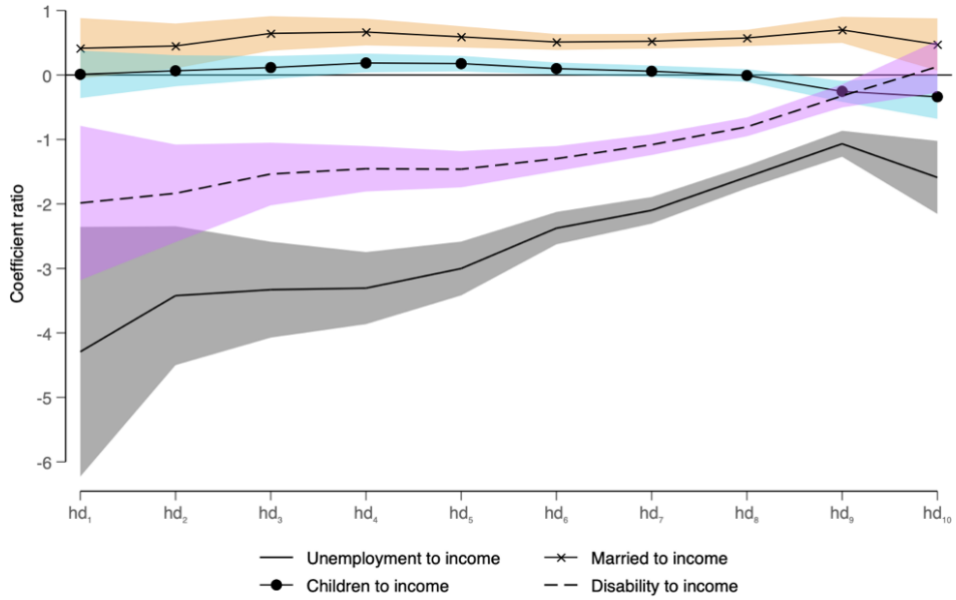
in at least some specifications, the only clearly plausible reversal is that of the effect of having children in a fixed-effects regression. Since that result was strongly insignificant and close to zero in Table 2, this does not look to be a particularly striking result. Moreover, the effect of having children was the only one for which a statistically *significant* reversal was possible (to achieve a reversal significant at the 5% level, we require $c = 0.65$).

Appendix D, Figure A3 replicates the above analysis with LISS data. We assess the possibility of reversals for both continuously reported happiness and discrete happiness. Three results stand out. First, transformations required for continuously reported happiness tend to be more extreme than the corresponding transformations required for discrete happiness. Second, in specifications with controls, the only case in which reversals are possible with $|c| < 1$ occurs for the effect of having children on happiness reported on a discrete scale ($c = -0.81$). Third, when including controls, no reversals are possible for working and for being married. Finally, we performed similar tests for available variables in the GSS data (see Appendix D, Table A6). Reversals where only feasible for the effect of having children (with $c = -1.73$).

5.2 Shadow prices and bounds on ratios of coefficients

A stated earlier, often we are not merely interested in the absolute magnitudes of coefficients. Instead, e.g. when estimating shadow prices, equivalence scales, or when assessing the cost-effectiveness of policy, we wish to learn about ratios of coefficients. However, Proposition 3 asserted that, when ratios of coefficients from regressions of $hd_{k,it}$ differ, these ratios will not be invariant under all transformations of hr_{it}.

Figure 5. Ratios of coefficients across regressions of $hd_{k,it}$



Note: Displayed are ratios of coefficients for unemployment, being married, children, and disability with respect to income. Coefficient ratios would be invariant across all positive monotonic transformations of hr if each line were exactly horizontal. This is most clearly not the case for ratios relating to unemployment and disability. All displayed ratios are based on OLS regressions of $hd_{k,it}$ with individual fixed effects (using SOEP data, corresponding to the bottom panel of Table A5). Shaded regions represent 95% confidence intervals (obtained using the Delta method).

In our data, these ratios indeed differ substantially. To illustrate, Figure 5 plots the ratio of the coefficients for unemployment, being married, having children and disability against the coefficient for income in each of the fixed effects regressions of $hd_{k,it}$ (corresponding to the bottom panel of Table A5). For unemployment and disability, the ratios of their estimated coefficients with the estimated income coefficient tend to increase with k. Therefore, the absolute magnitudes of the ratios of the effects of unemployment and disability on hr_{it} to the effect of income on hr_{it} will decrease (increase) for increasingly convex (concave) transformations of hr_{it} (because convex (concave) transformations give relatively more weight to higher (lower) levels of hr_{it}).

To illustrate that changes in these ratios are indeed of practical importance, we calculated shadow prices of each of the variables under consideration. We define the shadow price of e.g. unemployment as the amount of additional income needed for an unemployed person with an income level y to be as satisfied as someone who is not unemployed. This amount is given by $\left(e^{-\beta_{unemployed}/\beta_{\ln(income)}} - 1\right)y$.[19] Shadow price ranges for being married, having children, and disability can be found analogously. Given Proposition 4, each shadow price fall in a range determined by the largest and smallest ratio of coefficients obtained from regressions of $hd_{k,it}$.

Table 3 shows the results of this exercise. We find that the ranges of estimated shadow prices for unemployment and disability cover particularly wide ranges. For example, for unemployment the

[19] To see this, solve $\left[\beta_{\ln(y)}\ln(y+\Delta y)+\beta_{ue}\right]-\beta_{\ln(y)}\ln(y)=0$ for Δy.

Table 3. Shadow prices for each explanatory variable based on OLS regressions using SOEP data

Scale	Unemployment	Marriage	Children	Disability
Rank-order hr_{it}	€146,050 $(\frac{-0.638}{0.296})$	-€8,278 $(\frac{0.168}{0.296})$	-€500 $(\frac{0.008}{0.296})$	€34,677 $(\frac{-0.306}{0.296})$
Lower bound shadow price	€36,376 $(\frac{-0.035}{0.033})$	-€9,598 $(\frac{0.023}{0.033})$	€7,669 $(\frac{-0.003}{0.033})$	-€2,296 $(\frac{0.001}{0.009})$
Upper bound shadow price	€1,376,076 $(\frac{-0.010}{0.002})$	-€6,478 $(\frac{0.001}{0.002})$	-€3,249 $(\frac{0.004}{0.002})$	€119,769 $(\frac{-0.005}{0.002})$
$\widetilde{hr_{it}} = e^{chr_{it}}$ with $c = 0.4$	€92,782 $(\frac{-3.135}{1.772})$	-€8,370 $(\frac{1.024}{1.772})$	€1,108 $(\frac{-0.100}{1.772})$	€21,968 $(\frac{-1.358}{1.722})$
$\widetilde{hr_{it}} = -e^{chr_{it}}$ with $c = -0.4$	€273,885 $(\frac{-0.037}{0.014})$	-€8,227 $(\frac{0.008}{0.014})$	-€1,590 $(\frac{0.001}{0.014})$	€53,100 $(\frac{-0.018}{0.014})$

Note: Shadow prices are estimated at the sample mean of household income. Calculations are based on Table 2, column (4), the lower panel of Table A5, as well as fixed-effects regression of $\widetilde{hr_{it}}$ with $c = 0.4$ or $c = -0.4$. Corresponding ratios of coefficients in parentheses. Negative shadow prices imply that a variable is estimated to benefit respondents. For example, at the sample mean of household income and when using rank-order hr_{it}, a person who is *not* married needs to be compensated with €8,278€ of additional household income to be as satisfied as a married person. Since sign reversals were possible for having children and disability, the signs of their shadow prices also depend on the chosen scale. Notably, our estimates of the effects of income, unemployment, and marriage, obtained from regressions of rank-order hr_{it}, fall within the ranges reported by Frijters et al. (2020) as the current best estimates these variables' effects (no such estimates are provided for having children or a disability).

estimated shadow price can range from €36,376 to €1,376,076 across all possible transformations of hr. Likewise shadow prices for disability may range from -€2,296 (which, counterintuitively, implies that a *non-disabled* person should be compensated) to €119,769.

These ranges seem too wide to be useful. However, these maximal ranges of possible shadow prices rely on extreme transformations of hr_{it} in which differences between response categories approach zero except for some particular chosen response category. We therefore also evaluate how shadow prices change for a transformation $\widetilde{hr_{it}} = \pm e^{chr_{it}}$, with $c = 0.4$ and $c = -0.4$. These levels of c imply that differences in life satisfaction intensity between adjacent response categories increase or decrease by a factor of $e^{0.4} \approx 1.5$, which we take to still be plausible. That exercise shows that shadow prices for unemployment and disability still cover a rather wide range. Indeed, based on these figures, we do not know whether an unemployed (disabled) person can be compensated with as little as €93,000 (€22,000) or requires as much as €274,000 (€53,000).

We thus conclude that although sign reversals of the effects of explanatory variables on life satisfaction tend to either be impossible or unlikely, ratios of coefficients are substantially affected under even mild transformations.

5.3 Reversals using ordered probit

We now turn to the case of searching for sign reversals in the context of the heteroskedastic ordered probit model. Table 4 shows our results. In column (1) we enter each variable in a separate bivariate model. In column (2), all explanatory variables are entered jointly, including all previously mentioned controls. To reduce the bias from individual fixed effects being correlated with our explanatory variables, we add individual averages of all explanatory variables to the specification in column (3) (cf. Ferrer-i-Carbonell and Frijters, 2004; Mundlak, 1978; Van Praag, 2015).

As in the OLS case, higher incomes, being married, and having children are associated with a higher mean of the latent probit index (i.e. hp). Unemployment and disability are associated with

Table 4. Heteroskedastic ordered probit (HOP) models for hr and reversal conditions for each explanatory variable using SOEP data.

	(1) HOP, variables entered separately	(2) HOP, full controls	(3) HOP, full controls and individual averages
$\boldsymbol{\mu_{it}}$			
Log HH income	1.453*** (0.043)	1.209*** (0.039)	0.605*** (0.027)
	c=0.66	**c=1.43**	**c=2.34**
Unemployed	-2.711*** (0.075)	-1.759*** (0.055)	-1.181*** (0.040)
	c=2.33	**c=1.84**	**c=2.08**
Married	0.408*** (0.029)	0.577*** (0.031)	0.349*** (0.031)
	c=0.61	**c=0.90**	**c=2.01**
Children	0.409*** (0.030)	0.320*** (0.028)	-0.002 (0.024)
	c=1.14	**c=26.08**	**c=-0.02**
Disability	-1.804*** (0.062)	-1.525*** (0.055)	-0.573*** (0.038)
	c=1.06	**c=1.53**	**c=1.15**
Constant		10.336*** (0.224)	10.281*** (0.223)
$\boldsymbol{\ln(\sigma_{it})}$			
Log HH income	-0.140*** (0.004)	-0.065*** (0.005)	-0.021*** (0.005)
Unemployed	0.066*** (0.006)	0.069*** (0.006)	0.044*** (0.006)
Married	-0.038*** (0.004)	-0.049*** (0.005)	-0.014** (0.006)
Children	-0.021*** (0.004)	-0.001 (0.005)	-0.007 (0.006)
Disability	0.097*** (0.007)	0.073*** (0.007)	0.039*** (0.007)
Constant		1.281*** (0.024)	1.262*** (0.024)
Thresholds			
τ_0		-∞ (assumed)	-∞ (assumed)
τ_1		0.000 (assumed)	0.000 (assumed)
τ_2		1.000 (assumed)	1.000 (assumed)
τ_3		2.377*** (0.036)	2.373*** (0.036)
τ_4		3.775*** (0.068)	3.763*** (0.068)
τ_5		4.875*** (0.094)	4.855*** (0.093)
τ_6		7.032*** (0.145)	6.998*** (0.144)
τ_7		8.366*** (0.177)	8.324*** (0.176)
τ_8		10.499*** (0.228)	10.442*** (0.226)
τ_9		13.781*** (0.306)	13.703*** (0.305)
τ_{10}		16.257*** (0.365)	16.169*** (0.364)
τ_{11}		∞ (assumed)	∞ (assumed)
Observations	557,999	557,999	557,999

Note: In most cases, except for the coefficient on children in column (3), required magnitudes of c are larger than what is consistent with the evidence of section 4. Reversal conditions are evaluated at the sample means of all explanatory variables. Column (1) displays results from separate models for each explanatory variable. Since constants and thresholds vary (slightly) across regressions in column 1, they are not reported there. Model titles denote the specification estimated in each column. Data are from the 1984-2015 waves of the SOEP. Standard errors in parentheses (clustered by respondents). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

a lower mean. Analogously to the OLS fixed-effects specification, adding individual averages of all explanatory variables markedly reduces the magnitudes of our estimates. Across specifications, the magnitudes of these coefficients are roughly twice the magnitudes obtained in the corresponding OLS regressions of hr shown in Table 2. This is because differences between thresholds are estimated to be somewhat above 2 for high response categories and somewhat below 2 for low response categories. Coefficients are therefore scaled by a factor of approximately 2 when compared to the rank-order coding used in Table 2.

Concerning the estimated standard deviation of the error term, being married, having children, and higher incomes reduce σ_{it}. Unemployment and disability increase σ_{it}. Since no coefficient on

$\ln(\sigma_{it})$ is precisely zero, reversals are always possible. In Table 4, the level of c needed to reverse the sign of marginal effects is shown below each variable's estimated coefficients.

Two points are worth noting. First, in most cases the required level of c is clearly larger than our benchmark of $c = 1$. In these cases, reversals would require assumptions about scale use that are not currently supported by the work reviewed in section 4. Second, for income and marriage we sometimes find required levels of c that are smaller than 1. However, estimated differences between thresholds are typically above 2 for the majority of our sample (more than 75% of the sample report a level of hr above 4). Therefore, for any given c, a transformation $\widetilde{hp} = e^{chp}$ is typically more than twice as extreme as a similar transformation of rank-order-coded hr. Hence, for a more reasonable comparison with the latter transformations, we should multiply the c values in Table 4 by roughly 2.[20] After this multiplication, none of these required levels of c are within our benchmark of $c < 1$ except for the insignificant estimate of the effect of having children in column 3. As was true in the OLS setting, this is the only case in which a mild transformation would reverse the estimated marginal effect.

Finally, we replicated all these analyses with GSS and LISS data. See Appendix Tables A7 and A8. These results largely agree with those shown here.

6 Conclusions

This paper has made three main contributions. First, our *non-reversal condition* provided a test of whether reversals of OLS regression coefficients are possible. In this context, we showed that reversals are caused by variables having heterogenous effects across the distribution of reported happiness. We also compared the ordered probit approach with the OLS approach and argued that reversals in both approaches share the same underlying causes.

Second, we showed that reversals either require analysts to assume that respondents use response scales in a strongly non-linear fashion or that Assumption A2 is violated in a particularly severe manner. We then presented arguments and evidence to suggest that respondents use the response scale in a roughly linear fashion and that extreme violations of A2 are not observed in the available data.

Third, we empirically investigated the possibility and plausibility of reversals for a set of socio-economic variables. We found that reversals of OLS coefficients are impossible or unlikely in most of the cases we considered. Similarly, in the data and models we analysed, reversals using ordered probit require extreme transformations of the underlying scale. Although our main analyses rely on German SOEP data, which is one of the most common sources of data in the field, we obtain similar results using Dutch LISS and American GSS data.

Practically, our results suggest that when researchers merely wish to identify the direction of effects of explanatory variables, standard methods are reasonably robust to the worries flagged by B&L and S&Y. However, since effects of explanatory variables are rarely homogenous across the

[20] As stated in section 3.2, after a transformation $\widetilde{hp} = e^{chp}$, differences between transformed thresholds approximately grow by a factor $e^{\Delta\tau c}$, where $\Delta\tau$ is the typical difference between untransformed thresholds. In our case, $\Delta\tau \approx 2$. Therefore, differences thresholds approximately grow by a factor e^{2c}. In contrast, for rank-order order coded hr, differences between transformed categories only grow by a factor of e^{c}. Thus, to allow for a comparison between the OLS and the ordered probit approach, we should multiply the ordered probit values for c by 2.

distribution of happiness, ratios of coefficients can be affected even by mild transformations of reported happiness. Of course, it is unclear in how far our results would also hold for other datasets, other variables, or different model specifications. We therefore believe that future work should further investigate these issues on both a theoretical and empirical level.

We also suggest two new kinds of robustness test that may be applied in related future work.

First, researchers may want to verify the sensitivity of their results against plausible transformations (e.g. for $-0.4 < c < 0.4$). Practically, this means that researchers may want to test if signs of coefficients, significance levels, and ratios of coefficients remain the same when estimating regressions of rank-order hr and when estimating regressions of $\widetilde{hr} = \pm e^{chr}$ for $c = 0.4$ and $c = -0.4$. Of course, these suggested values for c are tentative, and may depend on the specific application.

Second, by ascertaining whether the *non-reversal condition* is satisfied, future work may attempt to verify that the signs of estimated OLS coefficients are immune to reversals. As stated in section 2.2. the *non-reversal condition* is satisfied when the signs of $\hat{\beta}_{m,k}^{(d)}$ are the same for all $k = 1,2, \dots, K-1$. If one is willing and able to defend Assumptions A2 and A3, satisfying the *non-reversal condition* practically means that signs of estimated coefficients are particularly robust against the questions raised by B&L and S&Y. Stata code that performs these tests are provided in our replication files.

Moreover, our Proposition 4 enabled evaluating how ratios of coefficients can change across all permissible labelling schemes. This may be particularly useful when assessing the relative impact of explanatory variables.

There are at least two important gaps in our analysis. First, our defence of Assumption A2, which helped to justify the use of OLS regressions, is tentative and was based on a single dataset. An extended investigation into the validity of this assumption would thus be welcome. Second, we set potential issues arising from heterogeneities in scale use aside. As argued in e.g. Angelini et al. (2014) or Kaiser (2022), such heterogeneities can bias estimates. Hence, future work should seek to analyse these issues jointly.

Lastly, our finding that the relative effects of explanatory variables are not homogenous across the distribution of reported happiness shows that estimating mean effects on happiness hides patterns in the data that are interesting and informative in their own right. As was previously done using quantile regressions (Binder and Coad 2011; 2015; Gupta et al. 2015), such patterns should be investigated more broadly.

References

Adler, Matthew D. 2013. 'Happiness Surveys and Public Policy: What's the Use?' *Duke Law Journal* 62 (8): 1509–1601.

Adler, Matthew D., Paul Dolan, and Georgios Kavetsos. 2017. 'Would You Choose to Be Happy? Tradeoffs between Happiness and the Other Dimensions of Life in a Large Population Survey'. *Journal of Economic Behavior & Organization* 139 (July): 60–73. https://doi.org/10.1016/j.jebo.2017.05.006.

Adler, Matthew D, and Nils Holtug. 2019. 'Prioritarianism: A Response to Critics'. *Politics, Philosophy & Economics* 18 (2): 101–44. https://doi.org/10.1177/1470594X19828022.

Angelini, Viola, Danilo Cavapozzi, Luca Corazzini, and Omar Paccagnella. 2014. '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* 76 (5): 643–66. https://doi.org/10.1111/obes.12039.

Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.

Barrington-Leigh, C. 2021. 'A Critique of the Econometrics of Happiness: Are We Underestimating the Returns to Education and Income?' *Working Paper*, 42.

Benjamin, Daniel J, Ori Heffetz, Miles S Kimball, and Alex Rees-Jones. 2012. 'What Do You Think Would Make You Happier? What Do You Think You Would Choose?' *American Economic Review* 102 (5): 2083–2110. https://doi.org/10.1257/aer.102.5.2083.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2014. 'Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices'. *American Economic Review* 104 (11): 3498–3528. https://doi.org/10.1257/aer.104.11.3498.

Bentham, Jeremy. 1789. *An Introduction to the Principle of Morals and Legislations*. Blackwell.

Berthoud, Richard, and Mark Bryan. 2011. 'Income, Deprivation and Poverty: A Longitudinal Analysis'. *Journal of Social Policy* 40 (1): 135–56. https://doi.org/10.1017/S0047279410000504.

Bertram, Christine, and Katrin Rehdanz. 2015. 'The Role of Urban Green Space for Human Well-Being'. *Ecological Economics* 120: 139–52. https://doi.org/10.1016/j.ecolecon.2015.10.013.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. 'Do People Mean What They Say? Implications For Subjective Survey Data'. *American Economic Review* 91 (2): 67–72.

Biewen, Martin, and Andos Juhasz. 2017. 'Direct Estimation of Equivalence Scales and More Evidence on Independence of Base'. *Oxford Bulletin of Economics and Statistics* 79 (5): 875–905. https://doi.org/10.1111/obes.12166.

Binder, Martin, and Alex Coad. 2011. 'From Average Joe's Happiness to Miserable Jane and Cheerful John: Using Quantile Regressions to Analyze the Full Subjective Well-Being Distribution'. *Journal of Economic Behavior & Organization* 79 (3): 275–90. https://doi.org/10.1016/j.jebo.2011.02.005.

———. 2015. 'Heterogeneity in the Relationship Between Unemployment and Subjective Wellbeing: A Quantile Approach'. *Economica* 82 (328): 865–91. https://doi.org/10.1111/ecca.12150.

Blanchflower, David G., and Andrew J. Oswald. 2016. 'Antidepressants and Age: A New Form of Evidence for U-Shaped Well-Being through Life'. *Journal of Economic Behavior & Organization* 127: 46–58.

Bloem, Jeffrey R. 2021. 'How Much Does the Cardinal Treatment of Ordinal Variables Matter? An Empirical Investigation'. *Political Analysis*, February, 1–17. https://doi.org/10.1017/pan.2020.55.

Bloem, Jeffrey R., and Andrew J. Oswald. 2021. 'The Analysis of Human Feelings: A Practical Suggestion for a Robustness Test'. *Review of Income and Wealth*. https://doi.org/10.1111/roiw.12531.

Bond, Timothy N, and Kevin Lang. 2013. 'The Evolution of the Black-White Test Score Gap in Grade K-3: The Fragility of Results'. *The Review of Economics and Statistics* 95 (1468–1479): 12.

Bond, Timothy N., and Kevin Lang. 2019. 'The Sad Truth about Happiness Scales'. *Journal of Political Economy* 127 (4): 1629–40. https://doi.org/10.1086/701679.

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.

Broome, John. 1991. '"Utility"'. *Economics & Philosophy* 7 (1): 1–12. https://doi.org/10.1017/S0266267100000882.

Chen, Le-Yu, Ekaterina Oparina, Nattavudh Powdthavee, and Sorawoot Srisuma. 2019. 'Have Econometric Analyses of Happiness Data Been Futile? A Simple Truth About Happiness Scales'. *ArXiv Preprint ArXiv:1902.07696*.

Clark, Andrew E. 2018. 'Four Decades of the Economics of Happiness: Where Next?' *Review of Income and Wealth* 64 (2): 245–69. https://doi.org/10.1111/roiw.12369.

Clark, Andrew E., and Claudia Senik. 2011. 'Is Happiness Different from Flourishing? Cross-Country Evidence from the ESS'. *Revue d'economie Politique*, 17–34.

Clark, Andrew E., Claudia Senik, and Katsunori Yamada. 2017. 'When Experienced and Decision Utility Concur: The Case of Income Comparisons'. *Journal of Behavioral and Experimental Economics* 70: 1–9.

Diamond, Peter. 2008. 'Behavioral Economics'. *Journal of Public Economics* 92 (8–9): 1858–62. https://doi.org/10.1016/j.jpubeco.2008.03.003.

Easterlin, Richard A. 1974. 'Does Economic Growth Improve the Human Lot? Some Empirical Evidence'. *Nations and Households in Economic Growth* 89: 89–125.

Easterlin, Richard A., and Kelsey J. O'Connor. 2020. 'The Easterlin Paradox'.

Fabian, Mark. 2021. 'Scale Norming Undermines the Use of Life Satisfaction Scale Data for Welfare Analysis'. *Journal of Happiness Studies*, October. https://doi.org/10.1007/s10902-021-00460-8.

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): 641–59. https://doi.org/10.1111/j.1468-0297.2004.00235.x.

Fleurbaey, Marc, and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press.

Frijters, Paul, Andrew E. Clark, Christian Krekel, and Richard Layard. 2020. 'A Happy Choice: Wellbeing as the Goal of Government'. *Behavioural Public Policy* 4 (2): 126–65.

Frijters, Paul, and Christian Krekel. 2021. *A Handbook for Wellbeing Policy-Making: History, Theory, Measurement, Implementation, and Examples*. New product. New York: Oxford University Press.

Gupta, Prashant, Tapas Mishra, Nigel O'Leary, and Mamata Parhi. 2015. 'The Distributional Effects of Adaption and Anticipation to Ill Health on Subjective Wellbeing'. *Economics Letters* 136 (November): 99–102. https://doi.org/10.1016/j.econlet.2015.09.010.

Hadar, Josef, and William R. Russell. 1969. 'Rules for Ordering Uncertain Prospects'. *The American Economic Review* 59 (1): 25–34.

Harsanyi, John C. 1996. 'Utilities, Preferences, and Substantive Goods'. *Social Choice and Welfare* 14 (1): 129–45.

HM Treasury. 2021. 'Wellbeing Guidance for Appraisal: Supplementary Green Book Guidance'. 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005388/Wellbeing_guidance_for_appraisal_-_supplementary_Green_Book_guidance.pdf.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. 'Back to Bentham? Explorations of Experienced Utility'. *The Quarterly Journal of Economics* 112 (2): 375–406.

Kaiser, Caspar. 2022. 'Using Memories to Assess the Intrapersonal Comparability of Wellbeing Reports'. *Journal of Economic Behavior & Organization* 193 (January): 410–42. https://doi.org/10.1016/j.jebo.2021.11.009.

Kaiser, Caspar, and Maarten CM Vendrik. 2019. 'Different Versions of the Easterlin Paradox: New Evidence for European Countries'. In *The Economics of Happiness*, 27–55. Springer. https://doi.org/10.1007/978-3-030-15835-4_2.

Kapteyn, Arie. 1977. 'A Theory of Preference Formation'. Tilburg University. https://research.tilburguniversity.edu/en/publications/a-theory-of-preference-formation-2.

Kapteyn, Arie, Bernard M. S. van Praag, and Floor G. Van Herwaarden. 1978. 'Individual Welfare Functions and Social Reference Spaces'. *Economics Letters* 1 (2): 173–77.

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.

Koivumaa-Honkanen, Heli, Risto Honkanen, Heimo Viinamäki, Kauko Heikkilä, Jaakko Kaprio, and Markku Koskenvuo. 2001. 'Life Satisfaction and Suicide: A 20-Year Follow-Up Study'. *American Journal of Psychiatry* 158 (3): 433–39. https://doi.org/10.1176/appi.ajp.158.3.433.

Kong, Feng, Ke Ding, Zetian Yang, Xiaobin Dang, Siyuan Hu, Yiying Song, and Jia Liu. 2015. 'Examining Gray Matter Structures Associated with Individual Differences in Global Life Satisfaction in a Large Sample of Young Adults'. *Social Cognitive and Affective Neuroscience* 10 (7): 952–60. https://doi.org/10.1093/scan/nsu144.

Kristoffersen, Ingebjørg. 2010. 'The Metrics of Subjective Wellbeing: Cardinality, Neutrality and Additivity*'. *Economic Record* 86 (272): 98–123. https://doi.org/10.1111/j.1475-4932.2009.00598.x.

———. 2017. 'The Metrics of Subjective Wellbeing Data: An Empirical Evaluation of the Ordinal and Cardinal Comparability of Life Satisfaction Scores'. *Social Indicators Research* 130 (2): 845–65. https://doi.org/10.1007/s11205-015-1200-6.

Krueger, Alan B., and David A. Schkade. 2008. 'The Reliability of Subjective Well-Being Measures'. *Journal of Public Economics* 92 (8–9): 1833–45.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2007. 'The Marginal Utility of Income'. *CEP Discussion Paper*. No. 784

Levinson, Arik. 2012. 'Valuing Public Goods Using Happiness Data: The Case of Air Quality'. *Journal of Public Economics* 96 (9): 869–80. https://doi.org/10.1016/j.jpubeco.2012.06.007.

Liu, Shuo, and Nick Netzer. 2020. 'Happy Times: Identification from Ordered Response Data'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3752581.

Lucas, Richard E., and M. Brent Donnellan. 2012. 'Estimating the Reliability of Single-Item Life Satisfaction Measures: Results from Four National Panel Studies'. *Social Indicators Research* 105 (3): 323–31.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

MacKerron, George. 2012. 'Happiness Economics from 35 000 Feet'. *Journal of Economic Surveys* 26 (4): 705–35. https://doi.org/10.1111/j.1467-6419.2010.00672.x.

Michalos, Alex C., and P. Maurine Kahlke. 2010. 'Stability and Sensitivity in Perceived Quality of Life Measures: Some Panel Results'. *Social Indicators Research* 98 (3): 403–34.

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.

Ng, Yew-Kwang. 1997. 'A Case for Happiness, Cardinalism, and Interpersonal Comparability'. *The Economic Journal* 107 (445): 1848–58.
———. 2008. 'Happiness Studies: Ways to Improve Comparability and Some Public Policy Implications'. *Economic Record* 84 (265): 253–66. https://doi.org/10.1111/j.1475-4932.2008.00466.x.
Nikolova, Milena, and Carol Graham. 2020. 'The Economics of Happiness'. In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmermann, 1–33. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57365-6_177-2.
Odermatt, Reto, and Alois Stutzer. 2019. '(Mis-)Predicted Subjective Well-Being Following Life Events'. *Journal of the European Economic Association* 17 (1): 245–83. https://doi.org/10.1093/jeea/jvy005.
OECD. 2013. *OECD Guidelines on Measuring Subjective Well-Being*. https://doi.org/10.1787/9789264191655-en.
———. 2020. *How's Life? 2020: Measuring Well-Being*. How's Life? OECD. https://doi.org/10.1787/9870c393-en.
ONS. 2021. 'Well-Being - Office for National Statistics'. 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing.
Oswald, A. J., and S. Wu. 2010. 'Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A.' *Science* 327 (5965): 576–79. https://doi.org/10.1126/science.1180606.
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.
Parducci, Allen. 1995. *Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications*. Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
Parfit, Derek. 1997. 'Equality and Priority'. *Ratio* 10 (3): 202–21. https://doi.org/10.1111/1467-9329.00041.
Pavot, William, and Ed Diener. 2008. 'The Satisfaction With Life Scale and the Emerging Construct of Life Satisfaction'. *The Journal of Positive Psychology* 3 (2): 137–52. https://doi.org/10.1080/17439760701756946.
Perez-Truglia, Ricardo. 2015. 'A Samuelsonian Validation Test for Happiness Data'. *Journal of Economic Psychology* 49: 74–83. https://doi.org/10.1016/j.joep.2015.05.002.
Plant, Michael. 2020. 'A Happy Possibility About Happiness (And Other) Scales: Understanding Why the Cardinality Assumption Is Both Defensible and Unnecessary'. *Happier Lives Institute Working Paper*.
Rojas, Mariano. 2007. 'A Subjective Well-Being Equivalence Scale for Mexico: Estimation and Poverty and Income-Distribution Implications'. *Oxford Development Studies* 35 (3): 273–93.
Schneider, Leann, and Ulrich Schimmack. 2009. 'Self-Informant Agreement in Well-Being Ratings: A Meta-Analysis'. *Social Indicators Research* 94 (3): 363.
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.
Sidgwick, Henry. 1874. *The Methods of Ethics*. Good Press.
Stevenson, Betsey, and Justin Wolfers. 2008a. 'Happiness Inequality in the United States'. *The Journal of Legal Studies* 37 (S2): S33–79.

———. 2008b. 'Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox'. Working Paper 14282. National Bureau of Economic Research. https://doi.org/10.3386/w14282.

Studer, Raphael. 2012. 'Does It Matter How Happiness Is Measured? Evidence from a Randomized Controlled Experiment'. *Journal of Economic and Social Measurement* 37 (4): 317–36. https://doi.org/10.3233/JEM-120364.

Sumner, L. W. 1996. *Welfare, Happiness, and Ethics*. Oxford University Press.

Urry, Heather L., Jack B. Nitschke, Isa Dolski, Daren C. Jackson, Kim M. Dalton, Corrina J. Mueller, Melissa A. Rosenkranz, Carol D. Ryff, Burton H. Singer, and Richard J. Davidson. 2004. 'Making a Life Worth Living: Neural Correlates of Well-Being'. *Psychological Science* 15 (6): 367–72.

Van Praag, Bernard M. S. 1971. 'The Welfare Function of Income in Belgium: An Empirical Investigation'. *European Economic Review* 2 (3): 337–69.

———. 1991. 'Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept'. *Journal of Econometrics* 50 (1–2): 69–89.

———. 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.

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.

Weisstein, Eric W. 2021a. 'Descartes' Sign Rule'. 2021. https://mathworld.wolfram.com/DescartesSignRule.html.

———. 2021b. 'Log Normal Distribution'. 2021. https://mathworld.wolfram.com/LogNormalDistribution.html.

———. 2021c. 'Polynomial'. 2021. https://mathworld.wolfram.com/Polynomial.html.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Yitzhaki, Shlomo. 1990. 'On the Sensitivity of a Regression Coefficient to Monotonic Transformations'. *Econometric Theory* 6 (2): 165–69.

Appendix

A Proofs

A1 Proposition 1

We can write $\widetilde{hr_i} = \sum_{k=1}^{K-1}(l_k - l_{k+1})hd_{k,i} + l_K$. To see this, suppose that respondent i chooses some arbitrary category a. We then have $\widetilde{hr_i} = l_a$. Recall that we defined $hd_{k,i} \equiv \mathbb{1}(hr_i \leq k)$, implying $hd_{k,i} = 0$ for all $k < a$ and $hd_{k,i} = 1$ for all $k \geq a$. We therefore get:

$$\begin{aligned}\widetilde{hr_i} &= (l_1 - l_2)0 + \cdots + (l_{a-1} - l_a)0 + (l_a - l_{a+1})1 + (l_{a+1} - l_{a+2})1 + \cdots + (l_{K-1} - l_K)1 + l_K \\ &= (l_a - l_{a+1}) + (l_{a+1} - l_{a+2}) + \cdots + (l_{K-1} - l_K) + l_K \\ &= l_a\end{aligned} \tag{A1}$$

Hence, all terms except l_a in the above expression for $\widetilde{hr_i}$ cancel out.

Stacking over all N individuals i, we get $\widetilde{\boldsymbol{hr}} = \sum_{k=1}^{K-1}(l_k - l_{k+1})\boldsymbol{hd_k} + l_K\boldsymbol{I}$, where $\boldsymbol{I}$ is a $N \times 1$ vector of 1s. Also stacking equations (6) and (7) over i, we get $\widetilde{\boldsymbol{hr}} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ and $\boldsymbol{hd_k} = \boldsymbol{X\beta}_k^{(d)} + \boldsymbol{\varepsilon}_k^{(d)}$. The estimated coefficient vector $\widehat{\boldsymbol{\beta}}$ can then be written as:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\widetilde{\boldsymbol{hr}} \\ &= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\left(\sum_{k=1}^{K-1}(l_k - l_{k+1})\,\boldsymbol{hd_k} + l_K\boldsymbol{I}\right) \\ &= (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\left(\sum_{k=1}^{K-1}(l_k - l_{k+1})\,((\boldsymbol{X'X})^{-1}\boldsymbol{X'})^{-1}\widehat{\boldsymbol{\beta}}_k^{(d)} + l_K\boldsymbol{I}\right) \\ &= \sum_{k=1}^{K-1}(l_k - l_{k+1})\,\widehat{\boldsymbol{\beta}}_k^{(d)} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'}l_K\boldsymbol{I}\end{aligned} \tag{A2}$$

In moving from the second to the third line above, we used the fact that $\widehat{\boldsymbol{\beta}}_k^{(d)} = ((\boldsymbol{X'X})^{-1}\boldsymbol{X'})\boldsymbol{hd_k}$, and hence $((\boldsymbol{X'X})^{-1}\boldsymbol{X'})^{-1}\widehat{\boldsymbol{\beta}}_k^{(d)} = \boldsymbol{hd_k}$. The term $(\boldsymbol{X'X})^{-1}\boldsymbol{X'}l_K\boldsymbol{I}$ equals an OLS estimate of a regression of a vector of constants $l_K\boldsymbol{I}$ on $\boldsymbol{X}$. Such a regression yields a vector with the first element equal to l_K and all other elements equal to 0. Hence, all but the first element of $\sum_{k=1}^{K-1}(l_k - l_{k+1})\widehat{\boldsymbol{\beta}}_k^{(d)}$ equal the corresponding elements of $\widehat{\boldsymbol{\beta}}$. For each element $\hat{\beta}_m$ of $\widehat{\boldsymbol{\beta}}$ (except the first), this entails that $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$. Assumption A1 entails that $l_k - l_{k+1}$ is negative for all k and all permissible labelling schemes. Therefore, when the estimate $\hat{\beta}_{k,m}^{(d)}$ is negative for all $k = 1, \ldots, K-1$, the corresponding estimate $\hat{\beta}_m$ will be positive for all permissible labelling schemes. Vice versa, when $\hat{\beta}_{k,m}^{(d)}$ is positive for all $k = 1, \ldots, K-1$, $\hat{\beta}_m$ will be negative for all permissible labelling schemes. However, when $\hat{\beta}_{k,m}^{(d)}$ is positive for some k, but negative for at least one other k, we can set $l_k - l_{k+1}$ to -1 for all k where $\hat{\beta}_{k,m}^{(d)}$ is positive, and set $l_k - l_{k+1}$ to some negative constant c for all k where $\hat{\beta}_{k,m}^{(d)}$ is negative. Define Δ^+ to be the sum of all positive $\hat{\beta}_{k,m}^{(d)}$ and Δ^- to be the sum of all negative $\hat{\beta}_{k,m}^{(d)}$. We can then write $\hat{\beta}_m = c\Delta^- - \Delta^+$. Setting this expression to 0, we obtain $c = \Delta^+/\Delta^-$. This will always yield a negative, and hence permitted, value of c. For values of c below $c = \Delta^+/\Delta^-$, $\hat{\beta}_m$ will be positive and for values of c above $c = \Delta^+/\Delta^-$,

$\hat{\beta}_m$ will be negative. This implies the possibility of sign reversal via changes in the relative size of Δ^+ and Δ^-.

A2 Proposition 2 and condition to satisfy Assumption A2

A2.1 Proposition 2

When Assumptions A2 and A3 hold, there exists some labelling scheme such that $\widetilde{hr}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i$, where $\varepsilon_i = \eta_i - \zeta_i$. By linearity of the expectation operator, $\mathrm{E}(\varepsilon_i|\boldsymbol{X}_i) = \mathrm{E}(\eta_i - \zeta_i|\boldsymbol{X}_i) = \mathrm{E}(\eta_i|\boldsymbol{X}_i) - \mathrm{E}(\zeta_i|\boldsymbol{X}_i) = 0$. The OLS estimate $\hat{\beta}_m$ is thus an unbiased and consistent estimate of β_m (including its sign). By the *non-reversal condition*, and given A1, the sign of $\hat{\beta}_m$ will be the same for all permissible labelling schemes. Hence, a regression of any other permissible labelling for $\widetilde{hr}_i$ also yields an unbiased and consistent estimate of the sign of $\hat{\beta}_m$.

A2.2 Condition to satisfy Assumption A2

In section 2.3 we stated that Assumption A2 is satisfied whenever $\mathrm{E}(ht_i|hr_i = k; \boldsymbol{X}_i) = \mathrm{E}(ht_i|hr_i = k)$ holds (but not vice versa).

Assumption A2 states that some permissible transformation of hr_i exists such that $ht_i = \widetilde{hr}_i + \zeta_i$ with $\mathrm{E}(\zeta_i|\boldsymbol{X}_i) = 0$. We can write $\mathrm{E}(ht_i|hr_i = k) = \widetilde{hr}_i + \mathrm{E}(\zeta_i|hr_i = k) = l_k + \mathrm{E}(\zeta_i|hr_i = k)$. In this case, by setting $l_k = \mathrm{E}(ht_i|hr_i = k)$, it follows that $\mathrm{E}(\zeta_i|hr_i = k) = 0$ for all $k = 1, \dots, K$. If $\mathrm{E}(ht_i|hr_i = k; \boldsymbol{X}_i)$=$\mathrm{E}(ht_i|hr_i = k)$, we also have $\mathrm{E}(\zeta_i|hr_i = k; \boldsymbol{X}_i) = \mathrm{E}(\zeta_i|hr_i = k) = 0$.

Now note that we can write $\mathrm{E}(\zeta_i|\boldsymbol{X}_i) = \sum_{k=1}^{K} s_k * \mathrm{E}(\zeta_i|hr_i = k; \boldsymbol{X}_i)$, which is analogous to equation (2) in section 2.1. If $\mathrm{E}(\zeta_i|hr_i = k; \boldsymbol{X}_i) = 0$ for all $k = 1, \dots, K$, we also have that $\mathrm{E}(\zeta_i|\boldsymbol{X}_i) = 0$. In that case, Assumption A2 is satisfied.

Notably, the reverse is not true: If $\mathrm{E}(\zeta_i|\boldsymbol{X}_i) = 0$, the identity can be satisfied if $\mathrm{E}(\zeta_i|hr_i = k; \boldsymbol{X}_i) > 0$ for some k and $\mathrm{E}(\zeta_i|hr_i = k'; \boldsymbol{X}_i) < 0$ for some other k'.

A3 Proposition 3

The proof of Proposition 1 established that $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$. For the ratio $\hat{\beta}_m/\hat{\beta}_n$ we thus get: $\frac{\hat{\beta}_m}{\hat{\beta}_n} = \frac{\sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}}{\sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,n}^{(d)}}$. If $\frac{\hat{\beta}_{k,m}^{(d)}}{\hat{\beta}_{k,n}^{(d)}} = \rho$ for all $k = 1, \dots, K-1$, we can substitute $\hat{\beta}_{k,m}^{(d)} = \rho\hat{\beta}_{k,n}^{(d)}$ into the expression for $\frac{\hat{\beta}_m}{\hat{\beta}_n}$, yielding $\frac{\hat{\beta}_m}{\hat{\beta}_n} = \frac{\sum_{k=1}^{K-1}(l_k - l_{k+1})\rho\hat{\beta}_{k,n}^{(d)}}{\sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,n}^{(d)}} = \rho\frac{\sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,n}^{(d)}}{\sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,n}^{(d)}} = \rho$.

A4 Proposition 4

In general, we can write the ratio $\frac{\hat{\beta}_m}{\hat{\beta}_n}$ as $\frac{\hat{\beta}_m}{\hat{\beta}_n} = \sum_{k=1}^{K-1}\frac{(l_k - l_{k+1})\hat{\beta}_{k,n}^{(d)}}{\sum_{j=1}^{K-1}(l_j - l_{j+1})\hat{\beta}_{j,n}^{(d)}}\frac{\hat{\beta}_{k,m}^{(d)}}{\hat{\beta}_{k,n}^{(d)}}$, i.e. as a weighted average of all ratios $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$. Suppose $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$ for $k = a$ is smallest among all $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$. By recoding $\widetilde{hr}_i$ such that $l_a - l_{a+1} < 0$ and $l_k - l_{k+1} = 0$ for all other $k \neq a$, we can assign all weight to $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$. In that case, $\hat{\beta}_m/\hat{\beta}_n = \hat{\beta}_{a,m}^{(d)}/\hat{\beta}_{a,n}^{(d)}$. However, by Assumption A1, $l_k - l_{k+1} = 0$ is just

not permissible. Therefore, $\hat{\beta}_m/\hat{\beta}_n > \hat{\beta}_{a,m}^{(d)}/\hat{\beta}_{a,n}^{(d)}$ and $\lim_{l_k - l_{k+1} \to 0 \text{ for } k \neq a} \hat{\beta}_m/\hat{\beta}_n \to \hat{\beta}_{a,m}^{(d)}/\hat{\beta}_{a,n}^{(d)}$. Hence, $\hat{\beta}_{a,m}^{(d)}/\hat{\beta}_{a,n}^{(d)}$ is the infimum of $\hat{\beta}_m/\hat{\beta}_n$.

Now suppose $\hat{\beta}_{b,m}^{(d)}/\hat{\beta}_{b,n}^{(d)}$ is largest among all $\hat{\beta}_{k,m}^{(d)}/\hat{\beta}_{k,n}^{(d)}$. By the same argument $\hat{\beta}_m/\hat{\beta}_n < \hat{\beta}_{b,m}^{(d)}/\hat{\beta}_{b,n}^{(d)}$ and $\lim_{l_k - l_{k+1} \to 0 \text{ for } k \neq b} \hat{\beta}_m/\hat{\beta}_n \to \hat{\beta}_{b,m}^{(d)}/\hat{\beta}_{b,n}^{(d)}$. Hence, $\hat{\beta}_{b,m}^{(d)}/\hat{\beta}_{b,n}^{(d)}$ is the supremum of $\hat{\beta}_m/\hat{\beta}_n$.

A5 Proposition 5

Assumption 4 states that $hr_i = k \leftrightarrow \iota_{k-1} < ht_i \leq \iota_k$. Let $s_{j,k}$ be the share in group $j \in \{A, B\}$ that responded with response category $hr = k$. Mean ht for groups A and B will then be given by the following inequalities:

$$\begin{aligned} \sum_{k=1}^{K} \iota_{k-1}\, s_{A,k} < \mathrm{E}[ht_A] \leq \sum_{k=1}^{K} \iota_k\, s_{A,k} \\ \sum_{k=1}^{K} \iota_{k-1}\, s_{B,k} < \mathrm{E}[ht_B] \leq \sum_{k=1}^{K} \iota_k\, s_{B,k} \end{aligned} \tag{A2}$$

Consequently, the difference $\mathrm{E}[ht_A] - E[ht_B]$ between the two groups is given by:

$$\sum_{k=1}^{K} \iota_{k-1} s_{A,k} - \iota_k s_{B,k} < \mathrm{E}[ht_A] - E[ht_B] < \sum_{k=1}^{K} \iota_k s_{A,k} - \iota_{k-1} s_{B,k} \tag{A3}$$

Suppose we want to ascertain whether $\mathrm{E}[ht_A] - E[ht_B] > 0$. To do so, it is sufficient to evaluate whether the lower part of the above inequality is positive. We can then write:

$$\begin{aligned} \mathrm{E}[ht_A] - E[ht_B] > \sum_{k=1}^{K} \iota_{k-1} s_{A,k} - \iota_k s_{B,k} \\ = (\iota_0 - \iota_1) s_{A,1} + (\iota_{K-1} - \iota_K) s_{B,K} + \sum_{k=2}^{K-1} (\iota_{k-1} - \iota_k) \left(\sum_{m=1}^{k} s_{A,m} - \sum_{m=1}^{k-1} s_{B,m} \right) \end{aligned} \tag{A4}$$

The equality in this relation can be shown to hold by expanding the terms relating to group A in the latter expression. Doing so yields (we repeatedly draw out terms from the summation over k and let some terms cancel):

$$\begin{aligned} (\iota_0 - \iota_1) s_{A,1} + \sum_{k=2}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ = \iota_0 s_{A,1} - \iota_1 s_{A,1} + \sum_{k=2}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ = \iota_0 s_{A,1} + \iota_1 s_{A,2} - \iota_2 s_{A,1} - \iota_2 s_{A,2} + \sum_{k=3}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ = \iota_0 s_{A,1} + \iota_1 s_{A,2} + \iota_2 s_{A,3} - \iota_3 s_{A,1} - \iota_3 s_{A,2} - \iota_3 s_{A,3} + \sum_{k=4}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ = \cdots = \sum_{k=1}^{K-1} \iota_{k-1} s_{A,k} - \iota_{K-1} \sum_{k=1}^{K-1} s_{A,k} \\ = \sum_{k=1}^{K-1} \iota_{k-1} s_{A,k} - \iota_{K-1}(1 - s_{A,K}) \\ = -\iota_{K-1} + \sum_{k=1}^{K} \iota_{k-1} s_{A,k} \end{aligned} \tag{A5}$$

By an analogous process, the terms relating to group B can be expanded to yield:

$$(\iota_{K-1} - \iota_K)s_{B,K} - \sum_{k=2}^{K-1}(\iota_{k-1} - \iota_k)\sum_{m=1}^{k-1} s_{B,m} = \iota_{K-1} - \sum_{k=1}^{K} \iota_k s_{B,k} \qquad \text{(A6)}$$

Combining the results of (A5) and (A6) yields the expression in the first line of (A4).

The expression in (A4) is only guaranteed to be positive for any permissible set of thresholds when $s_{A,1} = s_{B,K} = 0$ and $\sum_{m=1}^{k} s_{A,m} < \sum_{m=1}^{k-1} s_{B,m}$ for all $k = 2, \dots, K-1$. To see this, note that when $s_{A,1} \neq 0$, we can set the magnitude of $\iota_0 - \iota_1$ to be arbitrarily large and the magnitude of all other $\iota_k - \iota_{k+1}$ to be arbitrarily small, yielding a negative value for this expression. Conversely, when $s_{B,K} \neq 0$, we can choose the magnitude of $\iota_{K-1} - \iota_K$ to be arbitrarily large, and the magnitude of all other $\iota_k - \iota_{k+1}$ to be arbitrarily small. Finally, when $s_{A,1} = s_{B,K} = 0$ the sign of expression (A4) only depends on $\sum_{k=2}^{K-1}(\iota_{k-1} - \iota_k)\left(\sum_{m=1}^{k} s_{A,m} - \sum_{m=1}^{k-1} s_{B,m}\right)$. Recall that $(\iota_{k-1} - \iota_k)$ is negative for all permissible sets of thresholds. Therefore, if $\sum_{m=1}^{k} s_{A,m} - \sum_{m=1}^{k-1} s_{B,m}$ is negative for all k the entire expression in (A4) will be positive. However, when $\sum_{m=1}^{k} s_{A,m} - \sum_{m=1}^{k-1} s_{B,m}$ is positive for some k and negative for some other k' we can set the corresponding $\iota_{k-1} - \iota_k$ to be arbitrarily large and the corresponding $\iota_{k'-1} - \iota_{k'}$ to be arbitrarily small, yielding a negative sign for the entire expression. An analogous result to ascertain whether $\mathrm{E}[ht_B] - E[ht_A] > 0$ can be obtained by switching indices for groups B and A.

B Further discussion

B1 Violations of Assumption A2 when assuming a linear response scale

This appendix assesses the implications of dropping Assumption A2 while maintaining that response scales are linear.

When assuming that the response scale is linear in ht, we have $\iota_k - \iota_{k-1} = d$ for all k, where d is some constant. As usual, denote the number of available response categories with K. As in the proof of proposition 5, when allowing for mean ht to vary between two groups A and B within each response category, the expected difference in ht between the two groups is given by the interval provided in relation (A3). In order to verify whether $\mathrm{E}[ht_A] - E[ht_B] > 0$, it is sufficient to determine the sign of the lower bound of the interval, as given in relation (A4). The reverse case is obtained by swapping the indices for the two groups.

Since, in the present case, $\iota_k - \iota_{k-1} = d$ for all k we can rewrite relation (A4) as:

$$E[ht_A] - E[ht_B] \geq \sum_{k=1}^{K}(\iota_k - d)s_{A,k} - \iota_k s_{B,k} = \sum_{k=1}^{K} \iota_k(s_{A,k} - s_{B,k}) - d s_{A,k} \qquad \text{(B1)}$$

Note that we can write $d = K/(\iota_K - \iota_0)$, i.e. as the number of response options divided by the difference between the upper and lower limits of ht. Set $\iota_0 = 0$ and $\iota_K = 1$. We then have $d = 1/K$, and equation (B1) becomes:

$$E[ht_A] - E[ht_B] \geq \sum_{k=1}^{K} \iota_k(s_{A,k} - s_{B,k}) - 1/K \qquad \text{(B2)}$$

The first term in this expression $\sum_{k=1}^{K} \iota_k \left(s_{A,k} - s_{B,k}\right)$ is just the difference in mean ht between the two groups when maintaining that ht does not vary within response categories. Its value can be obtained by noting that $\iota_k = kd = k/K$, and calculating $\sum_{k=1}^{K} \frac{k}{K} \left(s_{A,k} - s_{B,k}\right)$. This value can also be obtained by simply labelling each k^{th} response category of hr as k/K, and taking the mean of this labelling of hr.

The second term in relation (B2), $1/K$, decreases with the number of available response categories K. For example, when $K = 3$ as in the GSS, we require that the difference in mean hr between the groups must exceed $1/3$ in order to be identified when dropping Assumption 2. For $K = 11$, as in German SOEP, we only require that this difference exceeds $1/11$. Thus, the amount by which group A must be happier than group B in order for the sign of this difference to be identified (while maintaining that the response scale is linear), is inversely proportional to the number of response categories. From this point of view, offering respondents a greater number of response categories is preferable.

Finally note that we could also set $\iota_0 = 0$ and $\iota_K = K$, which would correspond to a rank-order labelling of hr. In that case, we would have $d = K/(K-0) = 1$ and relation (B2) would read $E[ht_A] - E[ht_B] \geq \sum_{k=1}^{K} k(s_{A,k} - s_{B,k}) - 1$, which is the case we discuss in the main text.

B2 Ordered probit reversals rely on heterogeneities of effects across hp_i

In section 3.2 we asserted that ordered probit reversals are driven by effect heterogeneities across hp_i. This can be shown as follows. First write the normally distributed error εp_i in Equation (9) as $\sigma_i \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0,1)$ and σ_i is the standard deviation of εp_i as estimated on the basis of equation (10). We can then express the marginal effect of some variable $X_{i,m}$ on $\mathrm{E}\left(\widetilde{hp_i} \middle| \boldsymbol{X_i}\right)$ as an integral over error ϵ_i. We then obtain:

$$\frac{\partial E\left(\widetilde{hp_i} \middle| \boldsymbol{X_i}\right)}{\partial X_{i,m}} = \frac{\partial \int_{-\infty}^{\infty} \widetilde{hp_i} \varphi(\epsilon_i) d\epsilon_i}{\partial X_{i,m}} = \int_{-\infty}^{\infty} \frac{\partial \widetilde{hp_i}}{\partial X_{i,m}} \varphi(\epsilon_i) d\epsilon_i = E\left(\frac{\partial \widetilde{hp_i}}{\partial X_{i,m}} \middle| \boldsymbol{X_i}\right). \tag{B3}$$

Hence, the marginal effect of $X_{i,m}$ on mean $\widetilde{hp_i}$ (i.e. $\mathrm{E}\left(\widetilde{hp_i} \middle| \boldsymbol{X_i}\right)$) equals the mean effect of $X_{i,m}$ on individual $\widetilde{hp_i}$. Now suppose that coefficient $\beta_m^{(p)}$ of $X_{i,m}$ on hp_i is positive while coefficient $\beta_m^{(s)}$ of $X_{i,m}$ on $\ln(\sigma_i)$ is negative. In that case, a transformation $\widetilde{hp_i} = e^{chp_i}$ for some $c > 0$ will yield a sign reversal. Analogous arguments can also be given for each of the other possible cases, but are omitted for brevity.

To now show that such reversals are indeed driven by effect heterogeneities, we can elaborate the last integral in equation (B3) as:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\partial \widetilde{hp_i}}{\partial X_{i,m}} \varphi(\epsilon_i) d\epsilon_i &= \int_{-\infty}^{\infty} \frac{\partial e^{chp_i}}{\partial X_{i,m}} \varphi(\epsilon_i) d\epsilon_i = \int_{-\infty}^{\infty} \frac{de^{chp_i}}{dhp_i} \frac{\partial hp_i}{\partial X_{i,m}} \varphi(\epsilon_i) d\epsilon_i \\ &= c \int_{-\infty}^{\infty} e^{chp_i} \frac{\partial hp_i}{\partial X_{i,m}} \varphi(\epsilon_i) d\epsilon_i. \end{aligned} \tag{B4}$$

The derivative $\partial hp_i/\partial X_{i,m}$ in this expression indicates the "local" effect of a unit change in $X_{i,m}$ on hp_i for a given value of error ϵ_i. By virtue of equation (9), the fact that $\varepsilon p_i = \sigma_i \epsilon_i$, and the relation $\sigma_i = e^{\ln(\sigma_i)} = e^{\boldsymbol{X_i}\boldsymbol{\beta}^{(s)}}$, this local effect equals $\beta_m^{(p)} + \beta_m^{(s)}\sigma_i\epsilon_i$. Hence, we can write:

$$\frac{\partial E\left(\widetilde{hp}_i \middle| \boldsymbol{X_i}\right)}{\partial X_{i,m}} = c e^{c\boldsymbol{X_i}\boldsymbol{\beta}^{(p)}} \int_{-\infty}^{\infty} e^{c\sigma_i\epsilon_i}\left(\beta_m^{(p)} + \beta_m^{(s)}\sigma_i\epsilon_i\right)\varphi(\epsilon_i)d\epsilon_i. \tag{B5}$$

Thus, the marginal effect of $X_{i,m}$ on $\mathrm{E}\left(\widetilde{hp}_i \middle| \boldsymbol{X_i}\right)$ is proportional to an integral of the "local" effects $\beta_m^{(p)} + \beta_m^{(s)}\sigma_i\epsilon_i$, each of which are weighted by the term $e^{c\sigma_i\epsilon_i}\varphi(\epsilon_i)$. The sign of these local effects depends on ϵ_i and it changes from positive to negative beyond $\epsilon_i = -\frac{\beta_m^{(p)}}{\beta_m^{(s)}\sigma_i} > 0$. The weight on each local effect, as given by $e^{c\sigma_i\epsilon_i}\varphi(\epsilon_i)$, increases with c. Hence, for sufficiently large c the negative local effects will start to dominate the positive local effects, yielding a negative mean effect, and thus a sign reversal. Hence, reversals in the ordered probit approach are caused by heterogeneities in the sign of the local effects of $X_{i,m}$ on hp_i.[21] This is analogous to the OLS case, where sign reversals were caused by heterogeneities in the effects of $X_{i,m}$ across the distribution of hr_i.

Notably, the point at which reversals occur in equation (B5) is given by our reversal condition of Proposition 6. To see this, first observe that the term in front of the integrand never changes sign and can thus be ignored. Now expand the integral in equation (B5) as:

$$\beta_m^{(p)} \int_{-\infty}^{\infty} e^{c\sigma_i\epsilon_i}\varphi(\epsilon_i)d\epsilon_i + \beta_m^{(s)}\sigma_i \int_{-\infty}^{\infty} e^{c\sigma_i\epsilon_i}\,\epsilon_i\varphi(\epsilon_i)d\epsilon_i = 0 \tag{B6}$$

The first integral equals $E(\mathrm{e}^{\mathrm{c}\sigma_\mathrm{i}\epsilon_\mathrm{i}}) = e^{0.5c^2\sigma_i^2}$. The second integral ($I$) can be evaluated using integration by parts. Note that $\epsilon_i\varphi(\epsilon_i) = \epsilon_i(2\pi)^{-0.5}e^{-0.5e^2} = -\varphi'(\epsilon_i)$, and let $u = e^{c\sigma_i\epsilon_i}$ and $v'(\epsilon_i) = \epsilon_i\varphi(\epsilon_i)$. Hence, $u'(\epsilon_i) = e^{c\sigma_i\epsilon_i}c\sigma_i$ and $v(\epsilon_i) = -\varphi(\epsilon_i)$, yielding:

$$I = \int_{-\infty}^{\infty} e^{c\sigma_i\epsilon_i}\,\epsilon_i\varphi(\epsilon_i)d\epsilon_i = -e^{c\sigma_i\epsilon_i}\varphi(\epsilon_i)|_{-\infty}^{\infty} + c\sigma_i \int_{-\infty}^{\infty} e^{c\sigma_i\epsilon_i}\varphi(\epsilon_i)\,d\epsilon_i \tag{B7}$$

Evaluating the first term at either limit of integration leads to:

$$\begin{aligned}\lim_{\epsilon_i\to\pm\infty} -e^{c\sigma_i\epsilon_i}\,\varphi(\epsilon_i) &= \lim_{\epsilon_i\to\pm\infty} -e^{c\sigma_i\epsilon_i}(2\pi)^{-0.5}e^{-0.5\epsilon_i^2} \\ &= -(2\pi)^{-0.5}\lim_{\epsilon_i\to\pm\infty} e^{c\sigma_i\epsilon_i - 0.5\epsilon_i^2} = 0.\end{aligned} \tag{B8}$$

Hence, $I = c\sigma_i E(e^{c\sigma_i\epsilon_i}) = c\sigma_i e^{0.5c^2\sigma_i^2}$. We therefore obtain:

$$\beta_m^{(p)}e^{0.5c^2\sigma_i^2} + \beta_m^{(s)}c\sigma_i^2 e^{0.5c^2\sigma_i^2} = \left(\beta_m^{(p)} + \beta_m^{(s)}c\sigma_i^2\right)e^{0.5c^2\sigma_i^2} = 0. \tag{B9}$$

[21] As noted in e.g. Angrist and Pischke (2009, p.46) when using a linear model to approximate a nonlinear conditional expectation function, the non-linearity reveals itself as heteroscedasticity of the error term.

Solving for c yields $c = -\beta_m^{(p)} / e^{2\boldsymbol{X}_i\boldsymbol{\beta}^{(s)}}\beta_m^{(s)}$, which is the expression in Proposition 6.

B3 Illustration of ordered-probit-based reversals

One of B&L's examples concerns the Easterlin Paradox, which, using the same data, we also analysed in section 2. For comparison, Table B1 shows the results from a heteroskedastic ordered probit (HOP) model.

Columns (1) and (2) yield estimates of -0.045 and -0.165 for the marginal effects of log GDP per capita on μ_i and $\ln(\sigma_i)$. Applying the condition of Proposition 6 at the mean of log GDP per capita yields that with $c = -0.727$ the effect of log GDP per capita on the mean of the transformed index $\widetilde{hp}_i = -e^{-0.727hp_i}$ becomes 0. Such a c implies a response scale with transformed thresholds $\tilde{\tau}_0 = -\infty$, $\tilde{\tau}_1 = -1$, $\tilde{\tau}_2 = -0.483$, and $\tilde{\tau}_3 = 0$. Here, the difference between thresholds $\tilde{\tau}_3$ and $\tilde{\tau}_2$ is smaller than the difference between thresholds $\tilde{\tau}_2$ and $\tilde{\tau}_1$ by a multiplicative factor $e^{-0.727} = 0.483 = w$. This is close to the value $w = 0.46$ obtained in section 2.1. However, the sign-reversing value of c in Proposition 6 depends on the level of $\boldsymbol{X}_i$. Therefore, in the present example, the sign-reversing level of c lies between the value in Proposition 6 for the highest level of log GDP per capita ($= 10.80$, yielding $c = -0.81$) and its lowest observed level ($= 10.13$, yielding $c = -0.65$).

Moreover, as shown in column (4), we find that for $c < -2.588$, the effect of log GDP per capita on mean $\widetilde{hp}_i$ is significantly positive at the 5% level. Thus, a statistically significant reversal is feasible for the ordered probit approach, but not for the OLS approach. Finally, column (5) shows that for $c > 0.222$, the effect on mean $\widetilde{hp}_i$ is significantly negative, which implies a ratio $w = e^{0.222} = 1.249$ that is much less than what was needed in the OLS case.

Finally, despite using the same data, our sign-reversing value of c is more negative than B&L's (who obtain $c = -0.67$). This is for three reasons. First, because B&L do not derive a condition that is suitable for continuous explanatory variables, they use a heuristic numerical search where predictions of $\mathrm{E}(\widetilde{hp}_i|\boldsymbol{X})$ for a given value of c are regressed by OLS on X for different c. Second, B&L use a more flexible specification for their ordered probit model, estimating μ_i and $\ln(\sigma_i)$ separately for each year. The value of c for which we obtain a statistically significant reversal is more extreme than that of B&L because they do not cluster standard errors across years, thus potentially downwardly biasing the standard errors they report.

Table B1. Results for μ, $\ln(\sigma)$, and $E(\widetilde{hp}_i)$ based on heteroskedastic ordered probit models of hr using GSS data

	(1) μ	(2) $\ln(\sigma)$	(3) Effect on mean $\widetilde{hp}_i$ for $c = -0.727$	(4) Effect on mean $\widetilde{hp}_i$ for $c = -2.588$	(5) Effect on mean $\widetilde{hp}_i$ for $c = 0.222$
Log GDP per capita	-0.045 (0.030)	-0.165***(0.050)	0.000 (0.016)	0.163**(0.083)	-0.015**(0.008)
Constant	0.716***(0.006)	-0.493***(0.010)	n.a.	n.a.	n.a.
Waves			26		
Observations			46,303		

Note: Untransformed thresholds are set to $\tau_0 = -\infty$, $\tau_1 = 0$, $\tau_2 = 1$, $\tau_3 = \infty$. Marginal effects in columns (3) to (5) are evaluated at the mean of Log GDP per capita. Data are from the 1972–2006 GSS waves, as provided in the replication files of Stevenson & Wolfers (2008a). Standard errors in parentheses (clustered by respondents); standard errors in columns (3)-(5) are obtained using the Delta method. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B4 Comparing scales with different numbers of response options.

It seems natural to think of response scales with fewer response options as being interpreted as collapsed versions of response scales with more response options. In the most extreme case, we might think of discrete response scales as collapsed versions of a continuous response scale.

As noted in section 4.3, we may find it plausible that continuous response scales allow for ht to be recorded cardinally. In that case, if a discrete response scale is observed to be a linear collapse of the continuous scale, it would be further evidence in favour of believing that respondents interpret discrete response scales linearly.

Such an analysis is possible with the LISS data used in section 4.3. Specifically, we can evaluate which partitioning of the continuous scale would reproduce the observed cumulative response shares on the discrete scale. More formally, we find the set of thresholds $\tau_k, k = 0, \ldots, 10$, that satisfy $\tau_0 = 0$ and $F_{cont.}(\tau_k) = \sum_{p=0}^{k-1} s_p$, where $F_{cont.}$ is the empirical CDF of responses for the question using the continuous scale and s_p denotes the share of respondents that report response category p on the question using the discrete scale. The result of this exercise is shown in Figure A1. Figure 5 in Studer (2012) shows the result of the same procedure. There this figure served a more descriptive purpose. Based on our figure, it indeed seems as though the discrete response scale is a linear collapse of the continuous response scale. This is further evidence of approximately linear scale use for discrete response scales.

Figure A1. Partitioning of continuous happiness to match response shares for discrete happiness (based on LISS data).

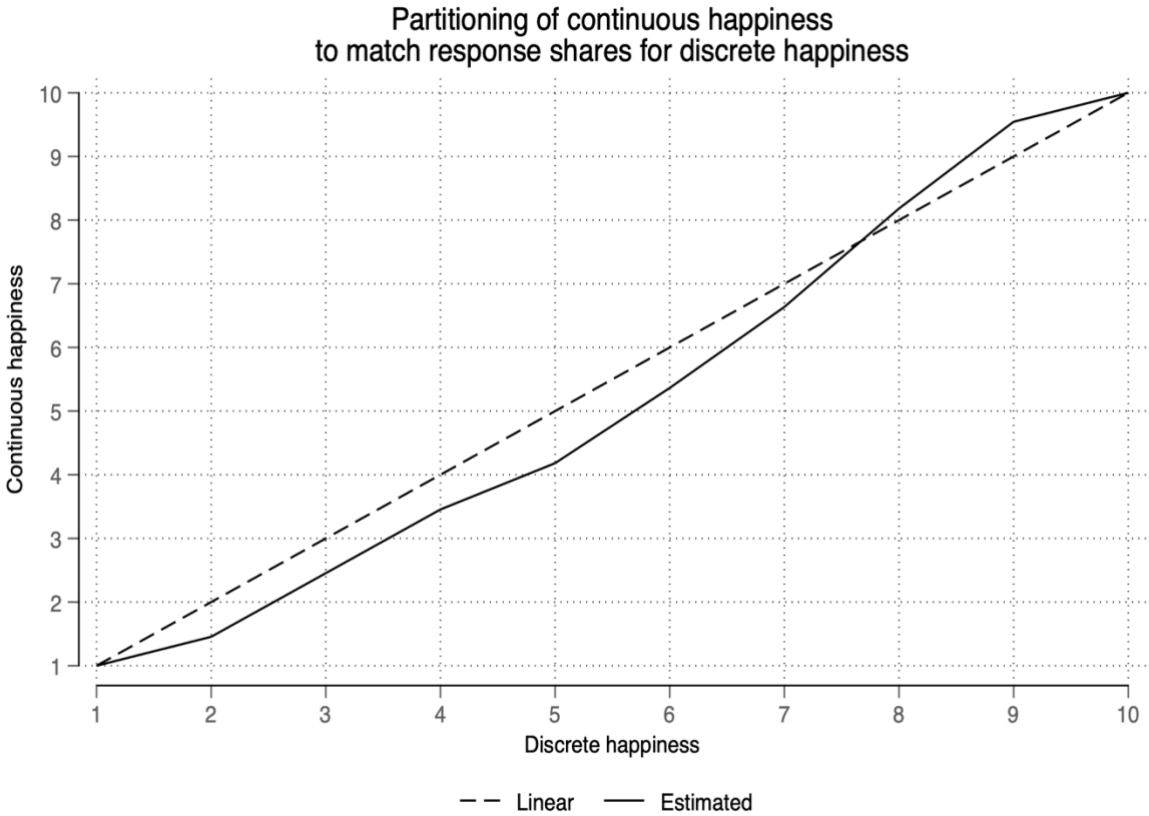


Table A1. Cumulative response shares for happiness and life satisfaction in GSS and WVS

GSS		WVS		Mean hr after collapse
hr	Share in % (cum.)	hr	Share in % (cum.)	
1 ("Not too happy")	11.98 (11.98)	1 ("Completely dissatisfied")	0.46 (0.46)	4.14
		2	0.90 (1.36)	
		3	2.05 (3.41)	
		4	3.89 (7.30)	
		5	7.32 (14.61)	
2 ("Pretty happy")	55.80 (67.78)	6	9.71 (24.32)	7.30
		7	23.06 (47.38)	
		8	28.27 (75.65)	
3 ("Very happy")	32.22 (1.00)	9	17.65 (93.29)	9.28
		10 ("Completely dissatisfied")	6.71 (100.00)	

Note: Data are taken from the 2006 waves of the GSS and WVS.

The same idea can also be applied to the issue of whether questions with only three or four response options are interpreted linearly by respondents. Thus, we now compare the three-points scale of the 2006 GSS wave with responses to the ten-points scale of the 2006 wave of the United States sample in the World Values survey (WVS).

The GSS asks about respondents' general happiness, while the WVS asks about life satisfaction. The comparison is therefore not ideal, but we are unfortunately not aware of a publicly available dataset that has a ten-points or an eleven-points scale for a question on happiness in the United States. Nevertheless, both samples are representative of the same population, and we hope that the two questions measure strongly correlated concepts of ht.

Table A1 shows cumulative response shares in each category of the two datasets. The observed cumulative response shares in these samples suggest that the 1st category ("not too happy") in the GSS questions most closely corresponds to categories 1-5 on a ten-points scale. Likewise, the 2nd category ("pretty happy") seems most likely to correspond to categories 6-8 and the 3rd category ("very happy") corresponds to categories 9-10 on a ten-points scale.

Assume now that the relative distribution of responses across the ten-points scale in the WVS sample (measuring life satisfaction) is a reasonable approximation of the distribution of responses we would observe had the GSS sample (measuring happiness) been given a ten-points scale. Further assuming that the 10-points scale measures ht roughly cardinally (as argued in the main text section), we can then take mean hr across categories 1-5 of the WVS variable as indicative of mean ht in the "not too happy" response category of the GSS variable. This yields a mean of 4.14. Same arguments apply to mean hr of categories 6-8 (mean = 7.30) and 9-10 (mean = 9.28) of WVS as being indicative of mean ht in categories "pretty happy" and "very happy" of GSS. See panel A of Figure A2 for an illustration of this analysis.

Furthermore, using WVS (four-points scale for happiness) and ESS (eleven-points scale for happiness) data, we also applied a similar procedure to a set of 14 European countries. As shown in Table A2 below, that exercise shows that differences between responses on the four-points WVS scale collapse in a roughly linear manner onto the eleven-points ESS scale. The figures for mean hr in the fifth column of Table A2 imply adjacent happiness differences from "not at all happy" to "very happy" of 2.71, 3.56, and 2.64. The subsequent ratios of these differences are

given by 1.31 and 0.74, which reveals no obvious pattern, and is not suggestive of a clear concave or convex response scale. See panel B of Figure A2 for an illustration of this analysis.

Taken together, the analysis of this appendix suggests that the convex/concave scales of the degree B&L require (see e.g. their section A3.4) may be plausible for questions with three response options, but less so for questions with more response options.

Figure A2. Illustration of results of Tables A1 and A2.

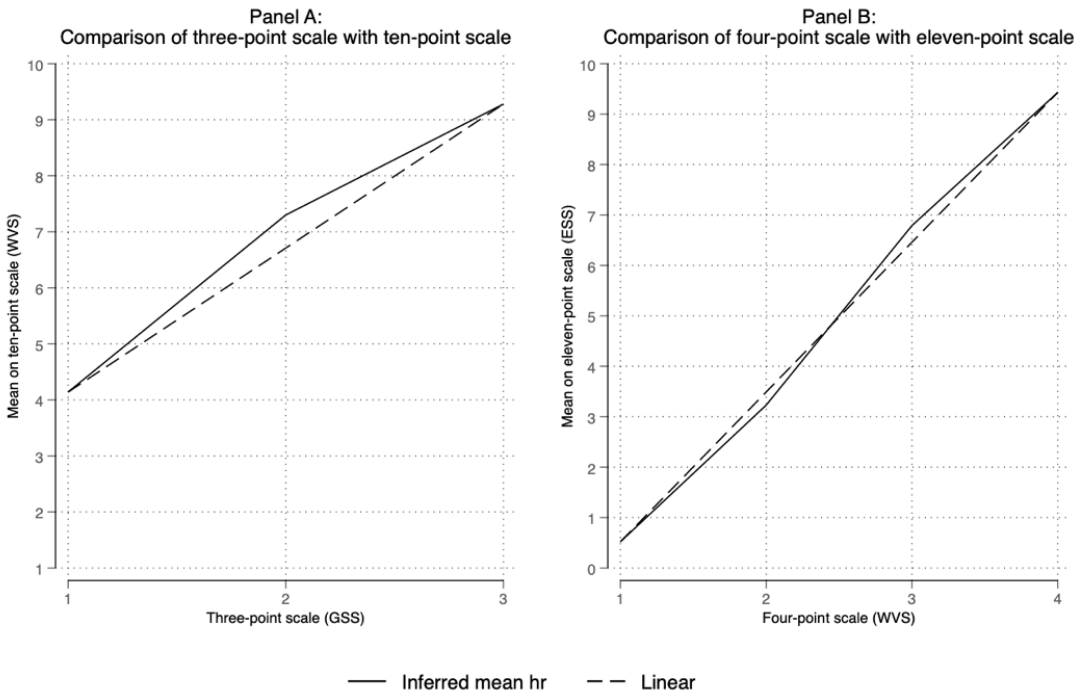


Table A2. Cumulative response shares for happiness and life satisfaction in ESS and WVS for European countries

WVS (Happiness)		ESS (Happiness)			ESS (Life Satisfaction)		WVS (Life Satisfaction)	
hr	% share (cumulative)	hr	% share (cumulative)	$\overline{hr}$ after collapse	hr	% share (cumulative)	hr	% share (cumulative)
1	2.45 (2.45)	0	0.97 (0.97)	0.52	0	3.26 (3.26)	1	2.40 (2.40)
		1	1.04 (2.01)		1	2.15 (5.41)	2	1.99 (4.39)
2	13.06 (15.51)	2	2.11 (4.12)	3.23	2	3.37 (8.78)	3	4.16 (8.55)
		3	3.88 (8.00)		3	6.10 (14.88)	4	4.55 (13.11)
		4	4.47 (12.47)		4	5.92 (20.80)	5	11.93 (25.04)
3	58.98 (74.50)	5	14.60 (27.06)	6.79	5	14.78 (35.58)	6	10.83 (35.87)
		6	9.24 (36.30)		6	9.45 (45.03)		
		7	18.70 (55.00)		7	16.25 (61.29)	7	18.77 (54.64)
		8	24.14 (79.13)		8	21.24 (82.52)	8	25.27 (79.90)
4	25.50 (100.0)	9	11.86 (90.99)	9.43	9	9.38 (91.90)	9	11.79 (91.69)
		10	9.01 (100.0)		10	8.10 (100.0)	10	8.31 (100.0)

Note: Data from WVS wave 5 and ESS wave 3 (both 2006). Population weights applied. Countries included: France, Finland, Germany, Great Britain, The Netherlands, Norway, Poland, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Ukraine. WVS response options for happiness are labelled "Not at all happy" (=1), "Not very happy" (=2), "Rather happy" (=3), "Very happy" (=4). Extreme response options for happiness in ESS are labelled "Extremely unhappy" (=0) and "Extremely happy" (=10). Extreme response options for life satisfaction in ESS are labelled "Extremely dissatisfied" (=0) and "Extremely satisfied" (=10). Extreme response options for life satisfaction in WVS are labelled "Completely dissatisfied" (=0) and "Completely satisfied" (=10).

C Additional Tables

Table A3. OLS regressions of cumulative response share and of a just-sign reversing response scale using GSS data

	(1) s_{t1} (share in 1st response category)	(2) $s_{t1} + s_{t2}$ (share in 1st or 2nd response category)	(3) s_{t3} (share in 3rd response category)	(4) $\widetilde{hr}_t$ (just sign-reversing concave scale)
Log GDP per capita	-0.025	0.054**	-0.054**	0.000
	(0.017)	(0.019)	(0.019)	(0.029)
Constant	0.121***	0.679***	0.321***	2.400***
	(0.003)	(0.004)	(0.004)	(0.006)
Years	26	26	26	26

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Log GDP per capita is demeaned across years. Data are from the 1972–2006 waves of the GSS, as provided in the replication files of Stevenson & Wolfers (2008a). The just sign-reversing concave scale in column (4) has labels $l = (1, 2.37, 3)$. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Model titles denote the dependent variable used in each column.

Table A4. OLS regressions of continuous and discrete happiness on standard demographics using LISS data

	(1) Continuous happiness	(2) Discrete happiness
Log HH income	0.332*** (0.057)	0.299*** (0.047)
Working	0.242* (0.109)	0.203* (0.080)
Married	0.517*** (0.075)	0.403*** (0.057)
Has children	-0.347** (0.107)	-0.268** (0.085)
Has disability	-0.427*** (0.053)	-0.318*** (0.042)
Constant	5.527*** (0.473)	6.211*** (0.406)
Respondents	3,722	3,722

Note: Standard errors in parentheses (clustered by respondent). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Data are from the March and April 2011 waves of the LISS. Model titles denote the dependent variable used in each column.

Table A5. Full results corresponding to Figure 4, i.e. OLS regressions of $hd_{k,it}$ using SOEP data.

	(1) $hd_{1,it}$ i.e. $hr \leq 1$	(2) $hd_{2,it}$ i.e. $hr \leq 2$	(3) $hd_{3,it}$ i.e. $hr \leq 3$	(4) $hd_{4,it}$ i.e. $hr \leq 4$	(5) $hd_{5,it}$ i.e. $hr \leq 5$	(6) $hd_{6,it}$ i.e. $hr \leq 6$	(7) $hd_{7,it}$ i.e. $hr \leq 7$	(8) $hd_{8,it}$ i.e. $hr \leq 8$	(9) $hd_{9,it}$ i.e. $hr \leq 9$	(10) $hd_{10,it}$ i.e. $hr \leq 10$
					No Controls					
Log household income	-0.006***	-0.010***	-0.021***	-0.039***	-0.061***	-0.145***	-0.180***	-0.166***	-0.066***	**0.003****
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	**(0.001)**
Unemployed	0.015***	0.025***	0.052***	0.100***	0.150***	0.258***	0.295***	0.250***	0.102***	0.028***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)
Married	-0.002***	-0.004***	-0.009***	-0.016***	-0.023***	-0.035***	-0.040***	-0.043***	-0.013***	-0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Children	-0.001**	-0.002***	-0.004***	-0.009***	-0.014***	-0.032***	-0.042***	-0.039***	-0.029***	-0.003*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Disability	0.010***	0.019***	0.037***	0.066***	0.096***	0.176***	0.196***	0.166***	0.077***	0.014***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
					Full Controls					
Log household income	-0.003***	-0.006***	-0.014***	-0.027***	-0.045***	-0.110***	-0.143***	-0.143***	-0.068***	-0.009***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Unemployed	0.013***	0.022***	0.044***	0.083***	0.121***	0.190***	0.208***	0.160***	0.058***	0.017***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)
Married	-0.003***	-0.005***	-0.011***	-0.021***	-0.030***	-0.053***	-0.063***	-0.065***	-0.031***	-0.008***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Children	**0.000**	-0.001	-0.002**	-0.005***	-0.010***	-0.020***	-0.029***	-0.035***	-0.023***	-0.008***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Disability	0.009***	0.017***	0.034***	0.060***	0.088***	0.149***	0.164***	0.149***	0.071***	0.026***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
					Full controls & fixed effects					
Log household income	-0.002***	-0.005***	-0.010***	-0.019***	-0.030***	-0.054***	-0.067***	-0.065***	-0.033***	-0.009***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Unemployed	0.010***	0.016***	0.034***	0.064***	0.091***	0.129***	0.140***	0.103***	0.035***	0.015***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)
Married	-0.001*	-0.002***	-0.007***	-0.013***	-0.018***	-0.028***	-0.035***	-0.038***	-0.023***	-0.004**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)
Children	-0.000	-0.000	-0.001	-0.004**	-0.005***	-0.005**	-0.004+	**0.000**	**0.008*****	**0.003****
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	**(0.003)**	**(0.003)**	**(0.002)**
Disability	0.005***	0.009***	0.016***	0.028***	0.044***	0.070***	0.072***	0.053***	0.011***	**-0.001**
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	**(0.002)**
Observations	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999

Note: Cells in bold have opposite sign, implying possibility of reversal. Data are from the 1984-2015 waves of the SOEP. Standard errors in parentheses (clustered by respondent). * $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Model titles denote the dependent variable used in each column.

Table A6. OLS regressions of rank-order reported happiness and of $hd_{k,it}$ on individual-level variables (using GSS data)

	(1) Rank-order hr (no control)	(2) Rank-order hr (full controls)	(3) hd_1 (no controls)	(4) hd_1 (full controls)	(5) hd_2 (no controls)	(6) hd_2 (full controls)
Log eq. HH income	0.119***	0.076***	-0.057***	-0.039***	-0.061***	-0.038***
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)
Unemployed	-0.330***	-0.213***	0.180***	0.136***	0.150***	0.077***
	(0.023)	(0.024)	(0.014)	(0.014)	(0.013)	(0.013)
Married	0.289***	0.289***	-0.102***	-0.101***	-0.188***	-0.188***
	(0.011)	(0.010)	(0.004)	(0.004)	(0.008)	(0.007)
Has children	0.032**	-0.053***	**0.007**	0.027***	-0.038***	0.026***
	(0.009)	(0.007)	**(0.003)**	(0.004)	(0.007)	(0.005)
Waves	26	26	26	26	26	26
Observations	41,630	41,630	41,630	41,630	41,630	41,630

Note: Cells in bold have opposite sign, implying possibility of reversal. Data are from the 1972–2006 waves of the GSS, as provided in the replication files of Stevenson & Wolfers (2008a). Standard errors in parentheses (clustered by year). * $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Model titles denote the dependent variable used in each column.

Table A7. Results of heteroskedastic ordered probit models for hr using LISS data

	(1) HOP, variables entered separately	(2) HOP, full controls
$\boldsymbol{\mu_{it}}$		
Log HH income	0.698 (0.427)	0.532* (0.317)
	$c = 0.94$	$c = 0.88$
Working	-0.031 (0.084)	0.310 (0.253)
	$c = -0.04$	$c = 0.81$
Married	0.944* (0.497)	0.935* (0.568)
	$c = -61.56$	$c = -5.31$
Has children	-0.244 (0.161)	-0.576 (0.368)
	$c = -0.47$	$c = -3.45$
Has disability	-0.570* (0.308)	-0.649* (0.387)
	$c = 3.26$	$c = -5.27$
Constant		8.332** (4.233)
$\boldsymbol{\ln(\sigma_{it})}$		
Log HH income	-0.141*** (0.032)	-0.123*** (0.034)
Working	-0.137*** (0.026)	-0.077 (0.059)
Married	0.003 (0.027)	0.036 (0.043)
Has children	-0.107*** (0.027)	-0.034 (0.058)
Has disability	0.034 (0.027)	-0.025 (0.029)
Constant		0.800 (0.581)
Thresholds		
τ_0		-∞ (assumed)
τ_1		0.000 (assumed)
τ_2		1.000 (assumed)
τ_3		2.434*** (0.865)
τ_4		3.455** (1.433)
τ_5		4.100** (1.798)
τ_6		4.962** (2.289)
τ_7		6.147** (2.970)
τ_8		8.784* (4.494)
τ_9		11.934* (6.320)
τ_{10}		∞ (assumed)
Respondents	3,722	3,722

Note: Data are from the March and April 2011 waves of the LISS. Standard errors in parentheses (clustered by respondents). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Model titles indicate specifications used in each column.

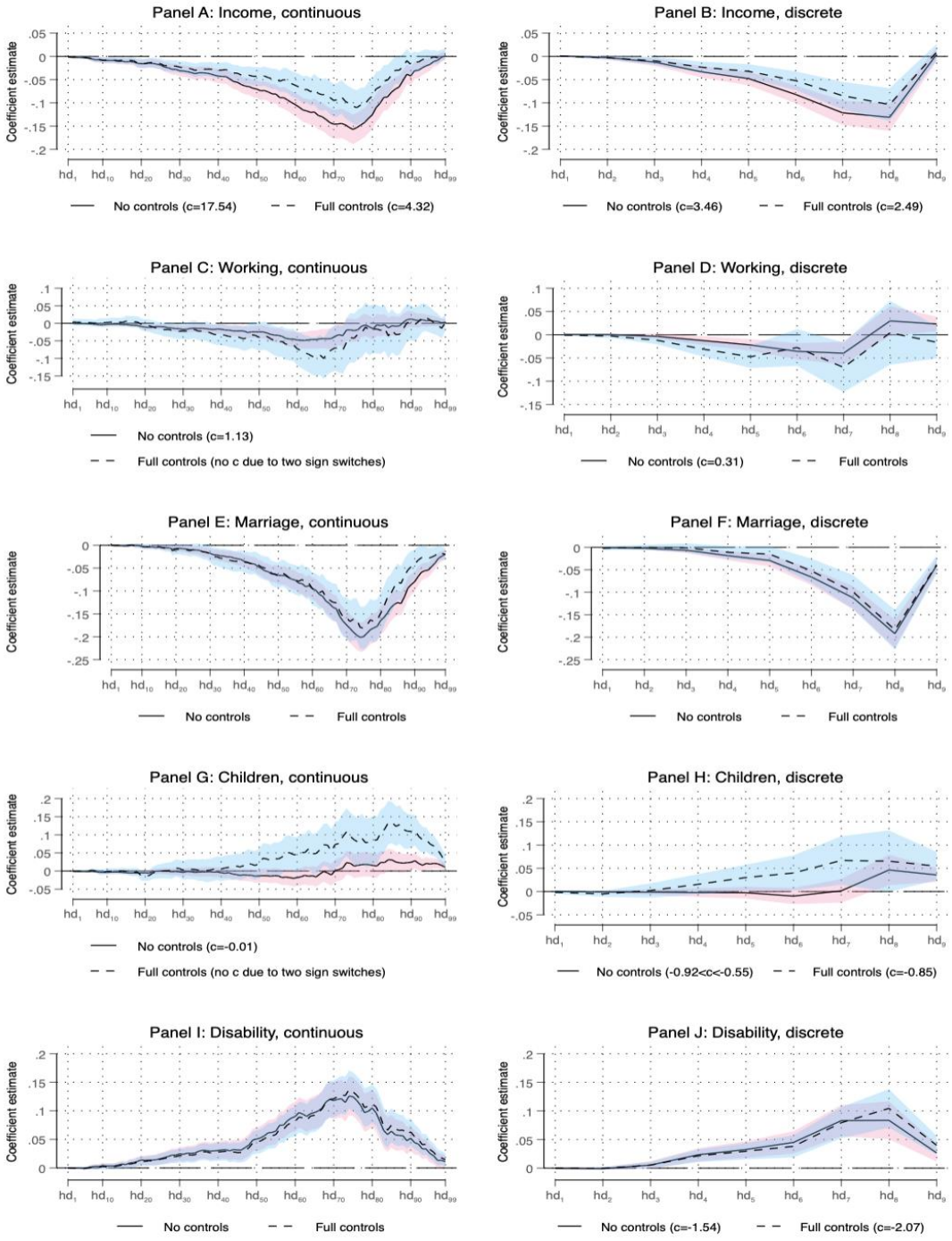
Table A8. Results of heteroskedastic ordered probit models for hr on individual-level socio-economic variables using GSS data

	(1) HOP, variables entered separately	(2) HOP, full controls
$\boldsymbol{\mu_{it}}$		
Log HH income	0.127*** (0.004)	0.082*** (0.004)
	$c = 1.60$	$c = 1.41$
Working	-0.359*** (0.026)	-0.233*** (0.027)
	$c = 7.74$	$c = 1.38$
Married	0.313*** (0.012)	0.310*** (0.011)
	$c = -45.30$	$c = -9.72$
Has children	0.040*** (0.010)	-0.055*** (0.008)
	$c - 1.31$	$c = 10.32$
Constant		-0.077 (0.050)
$\boldsymbol{\ln(\sigma_{it})}$		
Log HH income	-0.064*** (0.007)	-0.054*** (0.007)
Working	0.114*** (0.032)	0.136*** (0.029)
Married	0.020 (0.012)	0.029** (0.012)
Has children	0.085*** (0.011)	0.005 (0.014)
Constant		0.037 (0.081)
Thresholds		
τ_0	-∞ (assumed)	-∞ (assumed)
τ_1	0.000 (assumed)	0.000 (assumed)
τ_2	1.000 (assumed)	1.000 (assumed)
τ_3	∞ (assumed)	∞ (assumed)
Respondents	41,630	41,630

Note: Data are from the 1972–2006 waves of the GSS, as provided in the replication files of Stevenson & Wolfers (2008a). Standard errors in parentheses (clustered by year). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Model titles indicate the specification used in each column.

D Additional Figures

Figure A3. Coefficient estimates for each regression of $hd_{k,it}$ using LISS data.



Note: Continuous happiness is recorded with 100 unique values. Thus, each of the panels on the left shows 99 regressions of $hd_{k,it}$. Required c are shown in parentheses.