How much can we learn from happiness data?

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

May 2023

Abstract

Survey data on happiness, mental health, job satisfaction, and wellbeing are now widely used in economic research. They are also increasingly collected by government statistical agencies and bodies like the OECD. However, human feelings are measured on ordinal rather than cardinal scales. Hence, it has always been known that such data have to be handled with care. Some recent work has argued that, at least in principle, certain results in the research literature might be capable of being 'reversed' by recoding the scale – turning estimated positive effects into negatives, and vice versa. How important are such theoretical possibilities? We show that self-reported wellbeing data can in most relevant circumstances be used safely. To do so, we make three contributions. First, we derive a simple and general *non-reversal* condition for OLS regressions of ordinal data and obtain bounds for ratios of coefficients. Second, we demonstrate that respondents would have to answer numerical survey questions in a strongly non-linear fashion for reversals to actually appear. Yet, the evidence suggests approximately linear response behaviour. Third, using several large-scale datasets from the US, Germany, and the Netherlands, we show that effect reversals are rare or even impossible for a large number of core socio-economic variables, including people's incomes and employment status. We finish with a set of suggested recommendations for appropriate practice in empirical research and sketch avenues for future research.

JEL Codes: I31, C25

Keywords: ordinal reports, transformations of cardinal scales, subjective wellbeing, life satisfaction

---

* Department of Methodology and Statistics, Tilburg University; Institute for New Economic Thinking, Oxford University; Wellbeing Research Centre, Oxford University. 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

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, Nhat An Trinh, 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.

1 Introduction

In this paper we show that survey data on happiness and satisfaction are, broadly speaking, safe to use. We do so against a background of recent critiques of such data.

Scepticism against the use of 'subjective' data for economic research has a long-standing tradition (despite their enormous use in other scientific disciplines). Indeed, for a long time it seemed as though human feelings are simply unmeasurable, and that economics should purely focus on 'revealed' behaviour (e.g. Samuelson 1950; Stigler 1950). Yet, building on earlier pioneering work (e.g. Edgeworth 1881; Cantril 1965; Easterlin 1974), the idea that happiness and satisfaction can be reliably measured with survey data began to gain traction within economics in the 1990s (Van Praag 1991; Clark and Oswald 1994; 1996, 1999). About 30 years later, such work is now firmly in the social-scientific mainstream. A vast literature on the determinants and consequences life satisfaction, happiness, and mental health now exists. For overviews, see e.g. Di Tella and MacCulloch (2006), MacKerron (2012), Clark (2018), or Nikolova and Graham (2021). Crucially, national governments (e.g. HM Treasury, 2021) and international organisations (e.g. OECD, 2020) have begun to incorporate subjective measures of wellbeing into their decision-making.

Most of the academic 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.

Yet, this practice might not be robust. 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) made an even broader contribution and argued that discrete ordinal data are generally ill-suited for identifying the direction in which explanatory variables affect underlying wellbeing.

These papers raise a potentially important, although not new, issue. Their arguments could undermine the findings in the literature on subjective wellbeing. As these findings and data are increasingly used in policy-making and collected by national statistical agencies (ONS 2021), this is a critical concern. Since S&Y's and B&L's arguments apply to almost any kind of subjective data, their arguments also have broad implications beyond the literature on self-reported wellbeing. It is therefore valuable to give a comprehensive theoretical assessment of the reasons why and the conditions under which such reversals are possible. This paper provides such an analysis.

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 large at the top. We would then obtain a positive average effect of X. However, if we instead assumed that differences between response categories are 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, the sign of the average effect of X can be flipped by changing our assumptions about how respondents interpret the meaning of each response option.

Our first contribution is to derive a novel test of whether such reversals are possible in OLS regressions. This test is quite simple and only requires estimating regressions of dummies indicating each possible dichotomisation of the response scale. As it turns out, and contrary to the suggestion of B&L and S&Y, this test is satisfied in many empirical applications. This indicates that a significant part of the previous literature which almost universally relied on OLS regressions is robust to worries about sign reversals. However, ratios of OLS regression coefficients are not equally robust. The pricing of non-market goods with survey data depends on correctly estimating ratios of coefficients. Since this is a frequent and particularly important application of survey data on wellbeing (e.g. Levinson 2012; Danzer and Danzer 2016; Dolan et al. 2019; Frijters and Krekel 2021), we derive bounds on ratios of coefficients that hold for any positive monotonic transformation of how response options are labelled.

These analyses assume that mean wellbeing within each response category does not depend on X. Since this is a potentially strong assumption, we discuss how this assumption can be relaxed. In particular, an ordered probit model could be used in such cases. As shown by B&L, reversals are then always possible whenever estimated scale parameters depend on X. In large samples this condition will always be met. But this does not imply that reversals also easily occur in practice.

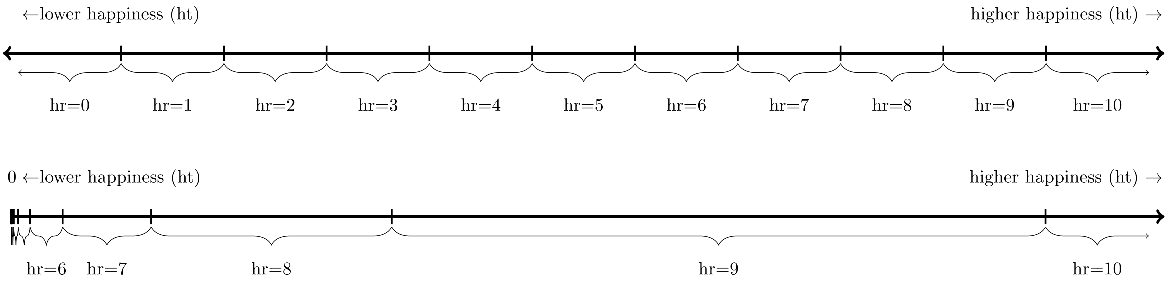
Our second contribution, therefore, is to analyse the conditions under which results based on ordered probit and OLS are actually reversed. This analysis shows that for reversals to occur, respondents must interpret response scales in a strongly non-linear manner. See Figure 1 for an illustration. Yet, our evidence shows that respondents tend to interpret scales with more than three response categories as approximately linear. Hence, results that rely on such scales are unlikely to be reversed. Moreover, by analysing data in which happiness is recorded on both an almost continuous and a discrete scale, we present evidence indicating that mean wellbeing within response categories only weakly depends on standard socio-economic characteristics, and that these associations do not enable reversals. This provides some justification for the OLS approach.

Our third contribution is to assess the practical plausibility of reversals in standard (eleven-point) life satisfaction data. In almost all cases we find that reversals of OLS regression coefficients are impossible or at least implausible. Similarly, the actual occurrence of reversals of ordered probit results often requires respondents to use response scales in a more strongly non-linear manner than what is supported by the evidence. Hence, when the goal is to establish the *direction* of a variable's effect on underlying wellbeing – or, indeed any other subjective feeling – self-reports are a strikingly robust measurement instrument.

That said, our analyses do suggest a number of 'best practices' that researchers may want to perform in the future when using such self-reports. For this purpose, our supplementary materials provide extensive Stata codes to implement the analyses showcased in this paper.[1]

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

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). As shown at the top, 'linear' scale use entails that differences in underlying wellbeing are constant across response categories. An example of 'non-linear' scale use is shown in the bottom. There, differences between boundaries of response categories ("thresholds") increase by a constant factor equal to 2.718 (i.e. e). In most of the cases we analyse, reversing a result requires that respondents use the scale in a way that is even more non-linear than is shown here. Appendix Figure D1 gives an alternative illustration.

Beyond our work, there are a few other complementary reactions to B&L's and S&Y's general arguments. Several years earlier, Ferrer-i-Carbonell & Frijters (2004) showed that OLS regressions and ordered probit models yield strikingly similar results. Their findings appeared to provide a justification for the wide-spread use of OLS in applied work on self-reported wellbeing. However, of course, this evidence alone cannot establish that survey data can be interpreted cardinally.[2] More recently, Bloem & Oswald (2021) and Chen et al. (2022) proposed 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 showed 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) showed how identification of rankings of mean happiness in two groups can be achieved with data on response times.

The rest of the paper proceeds as follows. In Section 2, we provide a general condition with which to test whether reversals of OLS regression coefficients are possible and derive bounds for ratios of coefficients. Section 3 explores the implications of allowing the mean of the latent dependent variable to vary within each response category and analyses reversals of marginal effects in ordered probit models. We show that although ordered probit reversals are almost always possible, their actual occurrence depends, like in the OLS case, on assuming that respondents interpret response scales in a (strongly) non-linear manner. In section 4, we present evidence to suggest that respondents interpret response scales in a roughly linear fashion and show that mean wellbeing within response categories does not strongly vary with standard socio-economic variables. Section 5 provides empirical evidence on the practical possibility and plausibility of reversals in self-reported data. We show that, when the focus is on the direction of a variable's influence on wellbeing, the use of survey data is 'safe' in most cases. We primarily use the German Socio-

[2] A separate concern is that the meaning of a given response category is not constant across people (Adler 2013; Diamond 2008; Viscusi 2020). Some papers evaluated the importance of this worry (Angelini et al. 2014; Kapteyn, Smith, and Van Soest 2010; Montgomery 2017; Odermatt and Stutzer 2019; Fabian 2021; Kaiser 2022). Although these studies find differences in scale use, these are typically too small to change substantive conclusions. To contain the length of our paper, we set this separate issue aside and assume that scale use is common across respondents.

Economic Panel (SOEP), but also utilise American GSS and Dutch LISS data in supplementary analyses. Proofs and further discussion are provided in an appendix.

2 Reversals of OLS regression coefficients

2.1 A *non-reversal condition* for OLS regressions

Most of the literature on wellbeing estimates OLS regressions of self-reported wellbeing on a variable of interest and a set of controls. Typically, the initial focus is then on the sign of the estimated coefficient. Given this standard practice in the field, we begin by presenting a general *non-reversal condition* for OLS regression coefficients. Although our discussion is framed in terms of 'happiness', our analyses apply to any ordinally reported construct. Indeed, the *non-reversal condition* we derive applies to regressions of any discrete dependent variable.

To motivate the analysis, suppose that the researcher's ultimate interest is in intensities of a person's true underlying happiness. Let this quantity be denoted by ht_i ('**t**rue **h**appiness'). Unfortunately, ht_i is only directly observable by respondents themselves. The analyst therefore has to resort to data on individuals' ordered self-reports about ht_i, which are recorded on a discrete scale with K response categories.

Let $hr_i \in \{1,2,\dots,K\}$ be respondent i's rank-order-coded **r**eported **h**appiness. We assign a 1 to the first category, a 2 to the second category, a 3 to the third category, etc. until the K^{th} category. This particular choice is not the only way of coding the data. Therefore, let $\widetilde{hr}_i = f(hr_i)$ be some transformation of hr_i and let $l = (l_1, \dots, l_k, \dots, l_K)$ denote a 'labelling scheme', which records how each of the K response categories are coded. Each labelling scheme l corresponds to a specific transformation f. That is, the elements in l are the image of f. Now consider the following individual-level regression equation:

$$\widetilde{hr}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i. \tag{1}$$

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 $\boldsymbol{\beta}$. Our primary interest is 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 test this, we first define a set of dummies $hd_{k,i} \equiv \mathbb{1}(hr_i \leq k)$, and consider the following regression equation:

$$hd_{k,i} = \boldsymbol{X}_i\boldsymbol{\beta}_k^{(d)} + \varepsilon_{k,i}^{(d)} \tag{2}$$

Let $\widehat{\boldsymbol{\beta}}_k^{(d)}$ be the OLS estimate of $\boldsymbol{\beta}_k^{(d)}$ in equation (2). Intuitively, means of the dummies $hd_{k,i}$ correspond to cumulative response shares of reported happiness until and including response category k.[3]

We assume that all permissible labelling schemes of $\widetilde{hr}_i$ have strictly increasing labels:

[3] See the working paper version of this paper (https://osf.io/gzt7a/) for an additional section on the intuition underlying the analysis in this section.

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 of rank-order coded hr_i 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. With this assumption in place, we can now state our *non-reversal condition*:

Proposition 1 (*non-reversal condition*). Given Assumption A1, all but the first element of $\hat{\boldsymbol{\beta}}$ 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 $\hat{\boldsymbol{\beta}}_k^{(d)}$ 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 weighted sum over $\hat{\beta}_{k,m}^{(d)}$. Specifically, it turns out that $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$.

This *non-reversal condition* has a direct and widely applicable use: It allows us to test whether the sign of an estimated OLS coefficient $\hat{\beta}_m$ of some explanatory variable X_m remains the same under every permissible labelling scheme, or equivalently, under every positive monotonic transformation of the dependent variable. 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, sign reversals of $\hat{\beta}_m$ are impossible.

Notably, when not all $\hat{\beta}_{k,m}^{(d)}$ share the same sign across k, the conditional association of $\widetilde{hr}_i$ with X_m is heterogeneous across response categories. Hence, sign-heterogeneities of variables' effects across the distribution of reported wellbeing are required to make reversals possible.[4]

For the special case where Equation (1) includes only one 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, they do not provide a non-reversal condition in settings with multiple explanatory variables. Our *non-reversal condition* fills this gap.

Finally, when just 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 $k = 1, \ldots, K-1$.[5] Here $s_{A,q}$ and $s_{B,q}$ denote observed shares in groups A and B in response category q. This condition is

[4] In comparison, 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 $\hat{\beta}_{k,m}^{(d)}$ is equivalent to a violation of first-order stochastic dominance (FOSD).

[5] This can be seen by considering individual-level regressions of $hd_{k,i}$ on a single dummy variable X to determine group membership (e.g. with $X_i = 1$ indicating membership in group A). Now note that $\hat{\beta}_k^{(d)}$ here yields the difference in the mean of $hd_{k,i}$ between groups A and B, and recall that means of $hd_{k,i}$ correspond to cumulative response shares. Hence, in that case, $\hat{\beta}_k^{(d)}$ is equivalent to the difference in cumulative response shares $\sum_{q=1}^{k} s_{A,q} - \sum_{q=1}^{k} s_{B,q}$.

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). The *non-reversal condition* simply restates the need for FOSD in a regression setting.

2.2 OLS-based reversals of effects on underlying happiness

The *non-reversal condition* of Proposition 1 is a general statement about the properties of OLS regression coefficients. 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.

Recall that ht_i denotes the unobservable cardinal quantity of underlying true 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 assumption is rarely made explicit. For one exception, see Layard et al. (2007). Substantively, Assumption A2 maintains that there is a 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 condition states that mean ht_i within each response category does not depend on $\boldsymbol{X}_i$. Section 4 provides a tentative test of this 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 this linear specification can be quite flexible in the sense of allowing for an arbitrary number of covariates, including higher-order terms and interactions, 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 shows that ordinal data on self-reported happiness can in some cases be safely used to study the sign of effects of explanatory variables on underlying happiness.

2.3 Bounds on ratios of OLS coefficients

When assessing the practical 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 absolute magnitude. For example, estimating shadow prices (Luechinger 2009; Levinson 2012; Danzer and Danzer 2016; Dolan et al. 2019) or equivalence scales (Rojas 2007; Biewen and Juhasz 2017; Borah et al. 2019) principally relies on ratios of coefficients.

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 $\hat{\boldsymbol{\beta}}$, 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 $k = 1, \ldots, K-1$.

As will become clear in section 5, the conditional of Proposition 3 is practically never satisfied. This means that ratios of coefficients are almost always affected by changing how reported happiness is labelled. Fortunately, 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 Appendices A3 and A4. As illustrated in section 5.2, Proposition 4 enables evaluating the relative impact of explanatory variables on happiness across all positive monotonic transformations of $\widetilde{hr}_i$.

2.4 OLS reversals using exponential transformations

The technique used to achieve reversals in our proof of the *non-reversal condition* makes use of highly irregular and generally non-linear labelling schemes (see Appendix A1). It would be surprising if respondents used response scales in the manner implied by such 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 $k = 1, \ldots, K-1$. Using the identity $\hat{\beta}_m = \sum_{k=1}^{K-1}(l_k - l_{k+1})\hat{\beta}_{k,m}^{(d)}$, we can then write $\hat{\beta}_m$ as a polynomial with coefficients $\hat{\beta}_{k,m}^{(d)}$:

$$\hat{\beta}_m = \sum_{k=1}^{K-1} -w^{k-1}\hat{\beta}_{k,m}^{(d)} . \tag{3}$$

For $w < 1$ differences between adjacent categories decline and for $w > 1$ differences increase. By Descartes' Rule of Signs, when $\hat{\beta}_{k,m}^{(d)}$ switches signs once across all k, equation (3) will be guaranteed to be zero for exactly one unique value of w.[6] Our empirical results shown in section 5 suggest that this is the prevalent case whenever any sign switches occur. For $w > 1$, a labelling scheme like that used in equation (3) 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 $w = e^{c}$.[7] In an ordered probit context, B&L use the same transformations, facilitating the comparison between the ordered probit and the OLS approach below.

Of course, these are not the only transformations that one might consider. For example, transformations of the form $\widetilde{hr} = hr^{c}$ would also yield positive monotonic transformations that are either concave (for $0 < c < 1$) or convex (for $c > 1$). However, empirically such transformations turn out to not be guaranteed to yield a reversal for some c. More complex transformations could also be considered. When there are at least two sign switches of $\hat{\beta}_{k,m}^{(d)}$ across k, transformations with an inflection point could lead to simpler reversals of coefficient signs than are feasible with an exponential transformation.[8] However, the results of section 5 show that this never occurs in our SOEP data. We therefore focus on exponential transformations.

3 Relaxing Assumption A2 and reversals based on ordered probit

OLS regressions of happiness data can be motivated on the basis of Assumption A2. But this assumption may not always be met. We therefore analyse the consequences of relaxing it. Section 3.1 discusses the general implications of doing so. Section 3.2 then focuses on the ordered probit model as one specific avenue for relaxing Assumption A2.

3.1 Relaxing Assumption A2

Suppose that Assumption A2 is replaced by the following 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 $\iota_0 < \iota_1 <\ldots< \iota_K$.

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.[9]

[6] 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 2021b). In these cases we search for roots numerically.

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

[8] With two sign switches across k, the simplest way to obtain a reversal is to either 'stretch' the middle part of the scale and to 'squeeze' the outer ends, or, *vice versa*, i.e. to 'squeeze' the middle part and to 'stretch' the outer ends. Transformations with one inflection point make this feasible.

[9] In comparison, 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).

Now consider two groups A and B, where group membership may be determined by combinations of values of the covariates in $\boldsymbol{X}$. Let $s_{j,q}$ denote the share of members of group j responding with response category q. We can then state:

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 $k = 2, \dots, K-1$.

Mutatis mutandis, analogous conditions hold for identifying whether $\mathrm{E}(ht_i|j = A) < \mathrm{E}(ht_i|j = B)$. A variant of this is also stated by B&L (cf. p. 1632). A proof of Proposition 5 in our notation is given in Appendix A5. In practice, the conditionals of Proposition 5 are much more demanding than the conditional of *our non-reversal condition* of Proposition 1. Consequently, the conditionals of Proposition 5 will almost never be satisfied for a sample of any reasonable size. See B&L, Table 1 for empirical examples. Hence, replacing Assumption A2 with Assumption A4 implies that we can almost never establish rankings of mean wellbeing between groups without restrictions on what may be considered as plausible scales.[10] The validity of Assumption A2 is thus key in considering how much we can learn from happiness data.

3.2 Reversals based on ordered probit models

In the case where we do not want to maintain Assumption A2 and want to estimate a parametric model, we may consider the use of an ordered probit model. Though less popular than the OLS approach, ordered probit models are also common in the literature on subjective wellbeing. This makes a comparison between the OLS and the ordered approach particularly useful.

B&L showed that differences in group means are almost never identified with an ordered probit model. We extend B&L's argument and derive a reversal condition for the sign of marginal effects of continuous explanatory variables. That analysis shows that reversals in ordered probit models depend, as was also true in the OLS case, on assuming non-linear scale use.

We require the following assumptions for the ordered probit model to be a viable:

Assumption A5.

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

$$hp_i = \boldsymbol{X}_i \boldsymbol{\beta}^{(\boldsymbol{p})} + \varepsilon p_i, \tag{4}$$

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 with mean zero and standard deviation σ_i, where the log of σ_i is given by:

$$\ln(\sigma_i) = \boldsymbol{X}_i \boldsymbol{\beta}^{(\boldsymbol{s})}. \tag{5}$$

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

[10] In Appendix B1 we discuss an interesting special case in which differences between thresholds $\iota_{k+1} - \iota_k$ are assumed to be constant for all k. In this case the conditions for identification of rankings of means are less demanding.

These assumptions are qualitatively different from those needed for the OLS approach (which relies on Assumptions A1-A3). A few elucidatory remarks about these may therefore be useful:

- On A5.II: 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. Note that εp_i could also be logistically distributed to yield an ordered logit model. Our arguments can be adapted to that case.
- On A5.III: It is typically assumed that $\boldsymbol{\beta}^{(s)} = \mathbf{0}$, i.e. that εp_i is homoscedastic. The heteroskedastic ordered probit model discussed here relaxes this assumption. The functional form in equation (5) is chosen for convenience, ensuring that σ_i is positive (Wooldridge 2010).
- On A5.IV: This assumption implies that the continuous variable hp is merely an *ordinal* proxy for underlying happiness ht. This emphasises that, as is widely understood, ordered probit only uses ordinal information. 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 (see Vendrik and Woltjer (2007), section 3.1, for a short discussion). Yet, most research implicitly assumes that g is linear. This is an arbitrary cardinalisation of the data.

The last of these remarks is most important in the current context. 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). Consequently, no sign reversals can occur when g is linear. 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 for certain classes of non-linear transformations.[11]

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_i and the latter function is concave in hp_i. We already introduced these functions when reversing OLS coefficients in section 2.4. When $c > 0$, the assumed model for ht_i is given by $ht_i = e^{c(\boldsymbol{X}_i\boldsymbol{\beta}^{(p)} + \varepsilon p_i)}$.[12] Here, ht_i will have a conditional distribution that is log-normal with mean (e.g. Weisstein, 2021b):

$$\mathrm{E}(ht_i|\boldsymbol{X_i}) = e^{c\mu_i + 0.5c^2\sigma_i{}^2}, \tag{6}$$

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:

$$\mathrm{E}(ht_i|\boldsymbol{X_i}) = -e^{c\mu_i + 0.5c^2\sigma_i{}^2}. \tag{7}$$

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 (6) 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

[11] This and the preceeding statement can be proven by a generalization of the derivation in the first two paragraphs of Appendix B2.

[12] 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 $\boldsymbol{X}_i$.

the case of (7), 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:[13]

> **Proposition 6:** Given Assumption 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 sign is given by $c = -\beta_m^{(p)}/e^{2X_i\beta^{(s)}}\beta_m^{(s)}$.

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

$$\frac{\partial E(ht_i|\boldsymbol{X_i})}{\partial X_{i,m}} = e^{c\mu_i + 0.5c^2\sigma_i{}^2}\left(c\beta_m^{(p)} + c^2 e^{2X_i\boldsymbol{\beta}^{(s)}}\beta_m^{(s)}\right). \tag{8}$$

Setting equation (8) to 0 and solving for c yields the condition in Proposition 6. The same expression is obtained when differentiating equation (7). 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 $\beta_m^{(s)} = 0$.

Importantly sign reversals of $\mathrm{E}(ht_i|\boldsymbol{X_i})$ in ordered probit models 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 directly produces estimates of how people use the response scale. The estimated thresholds $\tau_0, \tau_1, \dots, \tau_K$ seem to do just that. However, from Assumptions A4, A5.II, and A5.IV 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 response thresholds ι_k, the estimated thresholds τ_k must be transformed using g. Our beliefs about scale use therefore depend on the estimated probit thresholds and our choice of g. As stated above, g must be non-linear to yield reversals. Hence, in analogy to the OLS setting, reversals depend on assuming non-linear scale use.

Notably, in the special case of $ht_i = e^{chp_i}$, which we and B&L focus on, we obtain $(\iota_0, \iota_1, \dots, \iota_K) = (e^{c\tau_0}, \dots\dots, e^{c\tau_K}) = (0, 1, e^{\cdots}, \dots, \infty)$. 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. $(\iota_0 \dots, \dots, \iota_K)$ will increase (for $c > 0$) or decrease (for $c < 0$) by a factor $e^{c\Delta\tau}$.[14] 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. Finally, as we show in Appendix B2, the possibility of reversals in ordered probit is 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 are made possible by heterogeneities across the distribution of hr_i.

As will become apparent in section 5, a key difference between OLS and ordered probit is that OLS reversals are rarely possible, while ordered probit reversals are almost always possible.

---

[13] 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 $c = 2(\mu_A - \mu_B)/(\sigma_B^2 - \sigma_A^2)$.

[14] 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 $k = 1, 2, \dots, K-1$ (i.e. excluding the outer thresholds τ_0 and τ_K) we can write $(e^{c(\tau_{k-1}+\Delta\tau)} - e^{c(\tau_{k-2}+\Delta\tau)})/(e^{c\tau_{k-1}} - e^{c\tau_{k-2}}) = e^{c\Delta\tau}$. Analogous reasoning obtains the same factor for $c < 0$. Of course, differences between estimated thresholds are never exactly constant. However, deviations from this approximation imply even more strongly non-linear scale use.

Fundamentally, this is because 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 cumulative response shares for each level of hr_i (which occurs when all $\beta_{k,m}^{(d)}$ share the same sign). Nevertheless, the actual occurrence of sign reversals requires non-linear scale use in both cases.

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

4.1 General considerations

Thus far, the discussion established that reversals in both the OLS and ordered probit case are driven by assuming non-linear scale use. The exponential transformations we focus on 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). To generate some intuitions about these transformations, consider $c = 1$ 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. For $c = 1$, the difference in $\widetilde{hr}_i$ between subsequent response categories grows by a factor of $e^1 \approx 2.78$. Suppose now that there are 11 response categories (as in the often-used German SOEP survey). In that case, 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 for an illustration). The next section will survey the previous literature to assess whether such scale use is likely to be common.

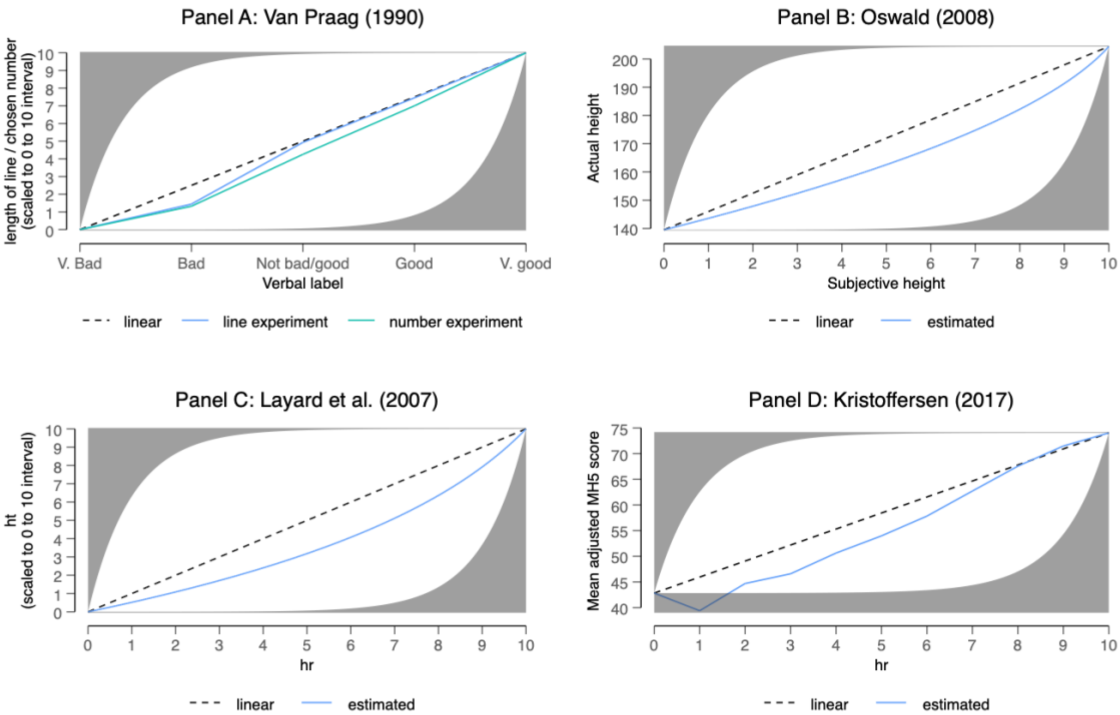
4.2 Evidence on scale use

Initial evidence is given by Van Praag (1991) who tested how individuals translate 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. This is most apparent when comparing his results with a multiplicative scale with $c = 1$, as also shown in Figure 2.

Another piece of general evidence is given by Oswald (2008), who asked respondents to report their height using 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 (measured in centimetres) and found a small but statistically significant negative coefficient on the squared term. In turn, when inverting his estimated regression equation, this implies convexity when transforming subjectively reported height into actual height. However, as shown in panel B, this estimate still is again close to linearity.

It is unclear whether this evidence generalises to other contexts. Being more directly concerned with scale use for wellbeing questions, Layard et al. (2007) pursued a rather different strategy. Using SOEP data, they estimated an OLS regression of rank-order-coded life satisfaction on a wide set of explanatory variables and assumed that the error of a model with actual life satisfaction (i.e.

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.

ht) as the dependent variable, is homoscedastic. As they show, any heteroskedasticity in their OLS regressions of rank-order coded hr then indicates non-linear scale use. They indeed find the residual variance to be larger for low predicted hr than for high hr, implying a convex response scale. Yet, as illustrated in panel C, the amount of convexity they infer is again rather small.

Finally, Kristoffersen (2017) was also specifically concerned with wellbeing data. She assumed that psychometrically adjusted scores from the MH5 index of mental health are a cardinal measure of ht. Using Australian HILDA data, she regressed 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, she also found a largely linear pattern.

Each of these papers made assumptions about the relationship between observable quantities and ht. Since ht is ultimately unobservable, none of these can be tested. Yet, despite each paper making different *sorts* of assumptions, their results all 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. We should therefore place less credence on strongly non-linear scales.

However, the evidence shown here only concerns questions with at least five categories. It is possible that scales with fewer response options, such as those in the American GSS are interpreted more non-linearly. In Appendix B3 we therefore compare response behaviour for ten and eleven-point scales with behaviour for three and four-point scales. We find that although three-point scales may be interpreted non-linearly, this does not appear to be the case for four-point scales.

4.3 A tentative test of Assumption A2

Approximately linear scale use looks to be a reasonable description of how respondents answer subjective survey questions. However, whether Assumption A2 holds – which would motivate the OLS approach – has not yet been investigated. To fill this gap, 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-point discrete scale. The other half received a continuous scale on a slider.[15] Although the scale appeared as continuous to respondents, the data was only recorded with a resolution of 100 distinct values (which we scale to also range from 1 to 10). In April, roles were reversed: The first half of respondents now answered on the continuous scale and the other half answered on the discrete scale. The data also contains socio-economic information, including respondents' household income, employment and marital status, numbers of children, and respondents' 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 ('hr_{cont}') does not vary across socio-economic characteristics within each of the ten response categories of *discrete* hr ('hr_{disc}'). More generally, if, across two groups, mean hr_{cont} is larger within a category of hr_{disc} for one of the two groups, we should also expect mean ht to be larger for that group within that category of hr_{disc}. We thus estimate a regression of hr_{cont} on dummies for each of the ten discrete response options of hr_{disc}. Each of these dummies are interacted with each of the aforementioned socio-economic characteristics. If these interactions are small and not statistically significantly different from zero, we interpret this as evidence in favour of Assumption A2.

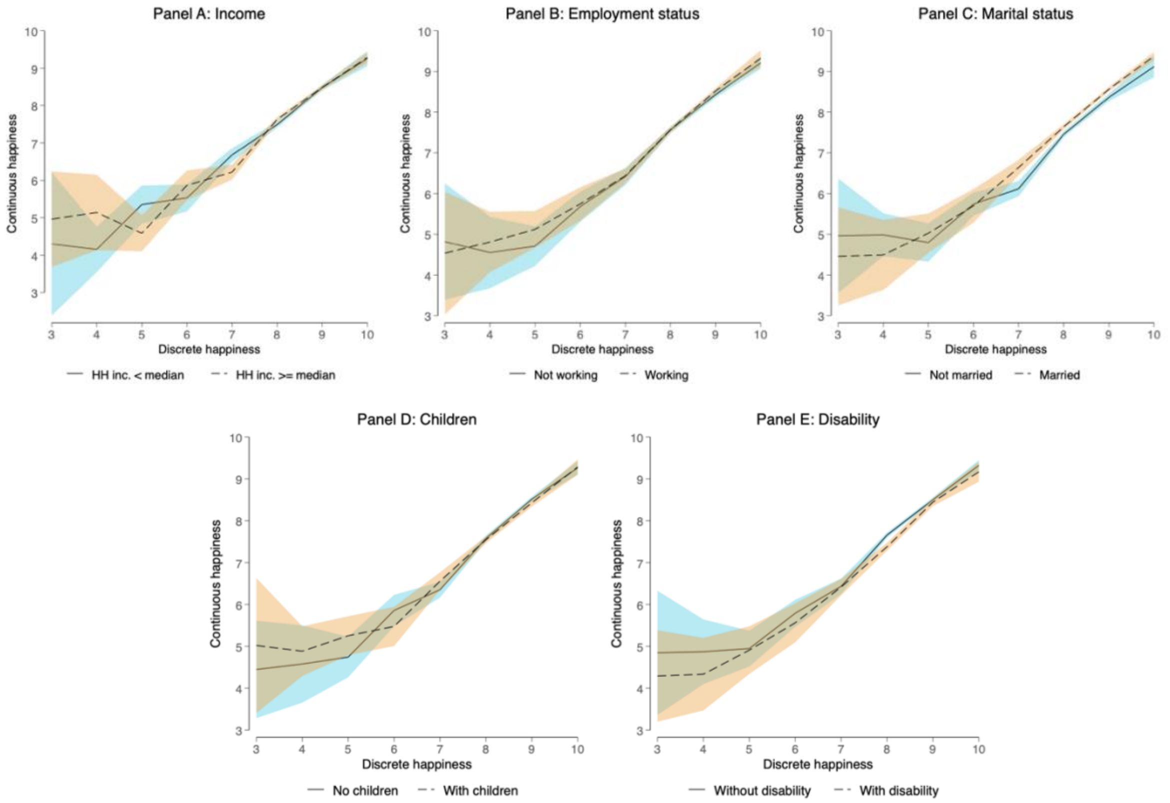
The results from this exercise are illustrated in Figure 3. Within each discrete response category, predicted levels of hr_{cont} are relatively homogenous across socio-economic characteristics. However, a formal hypothesis test of equality in predicted hr_{cont} across socio-economic groups cannot be rejected in all cases.[16] We therefore do see some violations of Assumption A2.

Can these violations cause the signs of OLS coefficients from a regression of hr_{disc} to be opposite to those from a regression of hr_{cont}? If so, Proposition 2 would be violated. However, for such sign switches to occur, Assumption A2 would have to fail in a rather specific manner. In particular, groups with higher mean hr_{disc} would have to have sufficiently lower hr_{cont} within at least some response categories of hr_{disc} (see footnote 20 in the working paper version for an intuition). In contrast to this requirement, Figure 3 shows a largely inconsistent pattern. At most, groups with

---

[15] The question reads *"Taking all things together, how happy would you say you are?"*, with extremes labelled *"completely unhappy"* and *"completely happy"*. This data was originally collected by Raphael Studer and Rainer Winkelmann (see Studer, 2012).
[16] The hypothesis 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 for disability. Moreover, for several subgroups no observations were available for the first and second response categories. This rendered a test of Assumption A2 impossible for these categories. Therefore, only categories 3 to 10 are shown in Figure 3. The fact that so few (or none) are observed in the bottom categories implies that the extent to which Assumption A2 is satisfied in these categories has less practical importance than for categories with many respondents. Furthermore, for the latter categories the interactions turn out to be small in size.

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).

higher hr_{disc} (see Appendix Table C2) tend to also have greater hr_{cont} within each discrete category. This is most visible in Panel E, where, within each category of hr_{disc}, non-disabled respondents report greater hr_{cont} than disabled respondents.

As can be seen in Appendix Table C2, where we report regressions of hr_{cont} and hr_{disc} side-by-side, such violations of Assumption A2 only bias OLS coefficients towards zero: Here, the coefficients in the regressions of hr_{cont} are larger than the coefficients in the regressions of hr_{disc}, but the corresponding coefficients always share the same signs. Hence, in our data, violations of Assumption A2 do not lead to sign switches. In that sense, our Proposition 2 is robust against mild violations of Assumption A2. Overall, these results and arguments give some justification for the practical use of OLS regressions and our *non-reversal condition*. Hence, in cases with a bounded scale of ht and many response categories, and if our non-reversal condition is satisfied, OLS is reasonably likely to lead to reliable estimates of the signs of the effects of explanatory variables on happiness. Nevertheless, we would welcome the collection of more data using a continuous slider to see if our observations can be replicated in other data, e.g. in large-scale surveys like the German SOEP.

Finally, Figure 3 also shows that for discrete response options larger than 4, hr_{cont} tends to increase approximately linearly in hr_{disc} within the various socio-economic groups. For response options lower than 5, we observe that mean continuous hr increases less steeply, but this non-linearity is not as pronounced as would be the case for a multiplicative scale with $c = 1$ (for comparison, see Appendix Figure D1). Hence, if we are willing to assume that the continuous scale allows respondents to report their ht cardinally (up to a linear transformation), this should be interpreted as additional evidence against strongly non-linear scale use in discrete scales. This idea is pursued further in Appendix B3, where we also investigate response scales with three or four response options.

5 Empirical Applications

We now assess the empirical relevance of the points of the preceding sections. Primarily, we do so by evaluating the possibility and plausibility of coefficient reversals in regressions of 11-point life satisfaction data. We use waves 1 (1984) to 32 (2015) of the German Socio-Economic Panel (SOEP). As shown in Appendices C and D, we can replicate our conclusions using Dutch LISS and American General Social Survey (GSS) data.[17]

Our primary explanatory variables include household income, unemployment, marriage, having children, and disability. Answers to the question *"How satisfied are you with your life, all things considered"* are used as our dependent variable. Answers are recorded with eleven response categories, labelled from 1 to 11. We use log net (post-tax) household incomes, deflated to 2005 prices and equivalized using the modified OECD scale. Regarding unemployment, we code a dummy that is 1 when a person reports to be unemployed, and 0 for all other employment statuses. We code similar dummies for being married, living with children in the household, and disability. In some specifications we add a set of additional control variables, including region and wave dummies, age(squared), tertiary education, home ownership, log(household size) and log(1+working hours).

5.1 OLS reversals using relabelling

Table 1 shows results for pooled and fixed-effects OLS regressions of hr_{it} on each explanatory variables of interest. 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 additional controls, are entered jointly. Column (3) adds individual fixed effects. Since the inclusion of individual fixed effects is equivalent to demeaning both the dependent and the explanatory variables, Propositions 1-4 remain applicable in the fixed-effects case.

In all specifications, household income, being married, and having children are associated with higher life satisfaction, while unemployment and 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 $k = 1,2, \ldots ,10$ when entering variables separately, when including controls, and when adding fixed effects. Figure 4 illustrates our results. For most variables and specifications, estimates

[17] That data is also used by B&L and is based on replication data by Stevenson & Wolfers (2008a).

Table 1. Application of the *non-reversal condition* for several socio-economic variables available in 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 $k = 1, 2, \dots, 10$ (see Figure 4 and Table C3 in Appendix C 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). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

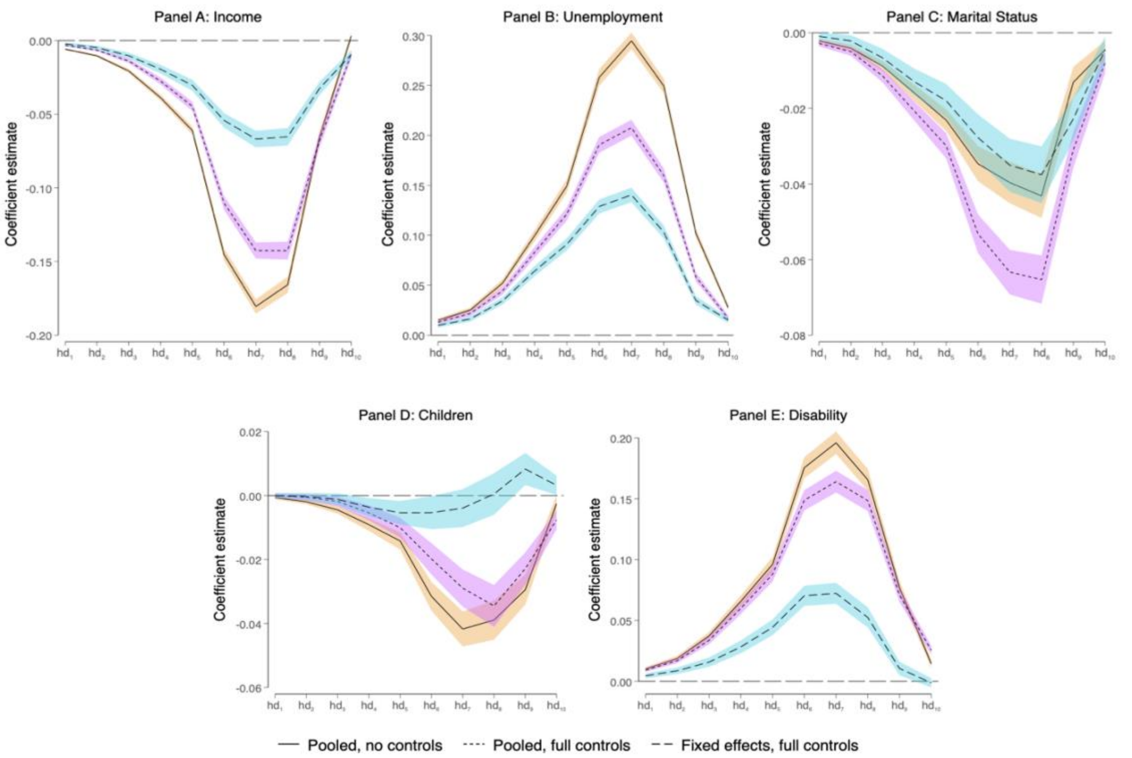
of $\hat{\beta}_{k,m}^{(d)}$ have the same sign across all k. Given our non-reversal condition of Proposition 1, reversals of $\hat{\beta}_m$ are therefore impossible in these cases.

However, there are four exceptions. First, when failing to include controls, the coefficient $\hat{\beta}_{10,income}^{(d)}$ for income has a positive sign, 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 yields that a multiplicative scale with $c > 3.18$ is required to achieve such a reversal. Second, the coefficient on children can be reversed in a pooled regression when including controls. Here the sign of $\hat{\beta}_{1,children}^{(d)}$ is positive while all other $\hat{\beta}_{k,children}^{(d)}$ are negative. A sufficiently concave transformation $\widetilde{hr}_{it} = -e^{chr_{it}}$ with $c < -2.83$ would yield a reversal here. Third, in a fixed-effects regression with full controls, the sign of $\hat{\beta}_{k,children}^{(d)}$ is negative for $k \leq 7$, but positive for $k > 7$. Here, a mild convex transformation with $c > 0.13$ is sufficient for a reversal. Fourth, the coefficient $\hat{\beta}_{10,disability}^{(d)}$ for disability in a regression with fixed effects is just negative, while all other coefficients are positive. Hence, a convex transformation with $c > 2.53$ would yield a reversal.

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

How likely are any of these theoretically possible reversals? In light of the arguments given in section 4, scales with c values in the order of at least 1 or at most -1 appear inconsistent with previous work and our evidence. Almost all the just-sign-reversing transformations found above are outside these values. The only exception is the effect of having children in a fixed-effects regression (where $c = 0.13$). We thus conclude that while reversals are possible for several

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).

variables in at least some specifications, the only clearly plausible reversal is that of the effect of having children in a fixed-effects regression.

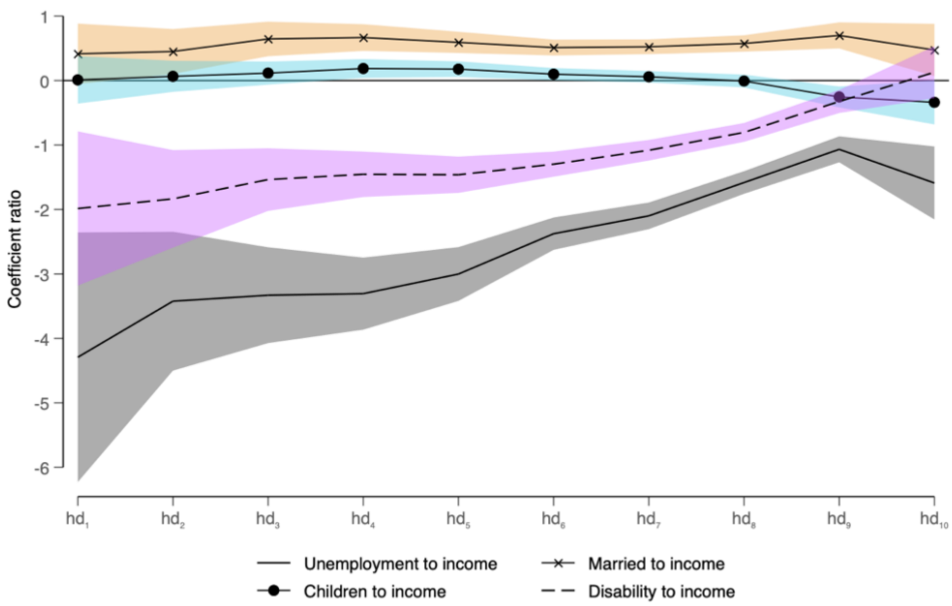
Appendix Figure D2 replicates the above analysis with LISS data. We there assess the possibility of reversals for both continuously reported happiness and discrete happiness. There, transformations required for continuously reported happiness tend to be more extreme than the corresponding transformations required for discrete happiness. Moreover, 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 ($c = -0.81$). Finally, we also performed similar tests with American General Social Survey data (see Appendix Table C4). Again, reversals are only feasible for the effect of having children.

5.2 Shadow prices and bounds on ratios of coefficients

We are often not only 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. 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}.

However, in our data these ratios differ substantially. Figure 5 plots the ratios of the coefficients for unemployment, being married, having children and disability to the coefficient for income in

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



Note: Displayed are ratios of coefficients for unemployment, being married, children, and disability 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 C3). Shaded regions represent 95% confidence intervals (obtained using the Delta method).

each of the fixed-effects regressions of $hd_{k,it}$ (corresponding to the bottom panel of Appendix Table C3). 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}.[18]

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 falls in a range determined by the largest and smallest ratio of coefficients obtained from regressions of $hd_{k,it}$.

Table 2 shows the results of this exercise. Estimated shadow prices for unemployment and disability cover particularly wide ranges. For unemployment, the estimated shadow price ranges

[18] This is because convex (concave) transformations give relatively more weight to higher (lower) levels of hr_{it}. Note that the ratios of the effects of unemployment and disability on hr_{it} to the effect of income on hr_{it} are positively-weighted averages of the ratios of the effects of unemployment and disability on $hd_{k,it}$ to the effect of income on $hd_{k,it}$ (see the proof in Appendix A4).

[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 2. Shadow prices for each explanatory variable based on OLS regressions using SOEP data.

Specification	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 ($\frac{0.023}{0.033}$)	-€3,249 ($\frac{0.004}{0.002}$)	-€2,296 ($\frac{0.001}{0.009}$)
Upper bound shadow price	€1,376,076 ($\frac{-0.010}{0.002}$)	-€6,478 ($\frac{0.001}{0.002}$)	€7,669 ($\frac{-0.003}{0.033}$)	€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 1, column (4), the lower panel of Table C3, 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 of these variables' effects (no such estimates are provided for having children or disability).

from €36,376 to €1,376,076 across all possible transformations of hr. Likewise, shadow prices for disability range from -€2,296 (implying *non-disabled* persons 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 response category. We thus evaluate how shadow prices vary for less extreme transformations $\widetilde{hr}_{it} = \pm e^{chr_{it}}$, with $c = 0.4$ and $c = -0.4$. These levels of c imply that differences in underlying life satisfaction between adjacent response categories increase or decrease by a factor of $e^{0.4} \approx 1.5$. Although we take these transformations to still be plausible, 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 3 shows our results. In column (1) we enter each variable in a separate 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 a lower mean. Analogously to the OLS fixed-effects specification, adding individual averages of all

Table 3. 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)
$\mathbf{ln(\sigma_{it})}$			
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		$-\infty$ (assumed)	$-\infty$ (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: Model titles denote the specification estimated in each column. Reversal conditions are evaluated at the sample means of all explanatory variables. 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. 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. Data are from the 1984-2015 waves of the SOEP. Standard errors in parentheses (clustered by respondents). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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 1. 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 1.

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 3, the level of c needed to reverse the sign of marginal effects is shown below each variable's estimated coefficients.

In most cases the required level of c is 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. But for income and marriage we sometimes find required levels of c that are smaller than 1. However, since 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) a transformation $ht = e^{chp}$ is typically more than twice as extreme for a given c than an analogous transformation of rank-order-coded hr. Hence, for a more reasonable comparison with the latter transformations, we should multiply the c values in Table 3 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 LISS and GSS data. See Appendix Tables C5 and C6. These results largely agree with those shown here.

6 Conclusions

This paper made three main contributions. First, our *non-reversal condition* provided a novel and general test of whether sign reversals of OLS regression coefficients are possible. This test provides applied researchers with a new tool to assess the robustness of results based on ordinal data. In this context, we showed that the possibility of reversals is caused by explanatory variables having heterogenous effects across the distribution of the dependent variable (such as 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 the actual occurrence of reversals of effects of variables requires analysts to assume that respondents in surveys use response scales in a strongly non-linear fashion. We then presented evidence to suggest that respondents instead use response scales with more than three response options in a roughly linear fashion. On that basis, we concluded that for such scales, reversals are a priori relatively unlikely to occur.

Third, we empirically investigated the practical possibility and plausibility of reversals of standard results in the literature. It turns out that reversals of OLS coefficients are impossible or unlikely in most of the cases we considered. Similarly, reversals using ordered probit require extreme non-linear transformations of the underlying scale. Although our main analyses relied 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.

Hence, it seems that many of the conclusions from earlier work on subjective wellbeing will hold up against the challenge put forward by Schröder & Yitzhaki (2017) and Bond & Lang (2019).

[20] As stated in section 3.2, after a transformation $ht = e^{chp}$, 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.

These are good news for this line of research. However, ratios of coefficients, which indicate the relative magnitudes of effect estimates and which are central to policy and cost-benefit analyses, are affected by mild transformations of the dependent variable. Practitioners and researchers should thus be careful when making policy recommendations on the basis of subjective variables.

To aid future research, we also suggest three new kinds of robustness tests. First, researchers may want to verify the sensitivity of their results against plausible transformations. 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, e.g., $c = 0.4$ and $c = -0.4$. Of course, these suggested values for c are tentative and depend on the specific application. Second, by ascertaining whether the *non-reversal condition* is satisfied, future research can now easily 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 able to defend Assumptions A1-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. Third, our Proposition 4 enabled evaluating how ratios of coefficients can change across all permissible labelling schemes. This is particularly useful when assessing the relative impact of explanatory variables.

There are a few gaps in our analysis that may be useful to fill in the future. First, our appraisal 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 aside that people may differ in how they use response scales. As argued in e.g. Angelini et al. (2014) or Kaiser (2022), such heterogeneities can bias estimates. Hence, future work should seek to jointly analyse issues relating to both (a) non-linear scale use and (b) interpersonal differences in scale use in an integrated fashion. Third, the extent to which response scales are interpreted as linear may depend on how questions are posed. Here we see untapped scope for improving the design of wellbeing-related questions. See Benjamin et al. (2021) for promising early work in this area.

References

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

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.

Benjamin, Daniel J, Jakina Debnam, Marc Fleurbaey, Ori Heffetz, and Miles Kimball. 2021. 'What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents'. *Working Paper*, 1–41.

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.

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. 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.

Cantril, Hadley. 1965. *The Pattern of Human Concerns*. New Brunswick: Rutgers University Press.

Chen, Le-Yu, Ekaterina Oparina, Nattavudh Powdthavee, and Sorawoot Srisuma. 2022. 'Robust Ranking of Happiness Outcomes: A Median Regression Perspective'. *Journal of Economic Behavior & Organization* 200 (August): 672–86. https://doi.org/10.1016/j.jebo.2022.06.010.

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

Clark, Andrew E., and Andrew J. Oswald. 1994. 'Unhappiness and Unemployment'. *The Economic Journal* 104 (424): 648–59.

———. 1996. 'Satisfaction and Comparison Income'. *Journal of Public Economics* 61 (3): 359–81. https://doi.org/10.1016/0047-2727(95)01564-7.

Danzer, Alexander M., and Natalia Danzer. 2016. 'The Long-Run Consequences of Chernobyl: Evidence on Subjective Well-Being, Mental Health and Welfare'. *Journal of Public Economics* 135 (March): 47–60. https://doi.org/10.1016/j.jpubeco.2016.01.001.

Di Tella, Rafael, and Robert MacCulloch. 2006. 'Some Uses of Happiness Data in Economics'. *Journal of Economic Perspectives* 20 (1): 25–46. https://doi.org/10.1257/089533006776526111.

Diamond, Peter. 2008. 'Behavioral Economics'. *Journal of Public Economics* 92 (8–9): 1858–62.

Dolan, Paul, Georgios Kavetsos, Christian Krekel, Dimitris Mavridis, Robert Metcalfe, Claudia Senik, Stefan Szymanski, and Nicolas R. Ziebarth. 2019. 'Quantifying the Intangible Impact of the Olympics Using Subjective Well-Being Data'. *Journal of Public Economics* 177 (September): 104043. https://doi.org/10.1016/j.jpubeco.2019.07.002.

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

Edgeworth, Francis Ysidro. 1881. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. Vol. 10. CK Paul.

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?' *Economic Journal* 114 (497): 641–59. https://doi.org/10.1111/j.1468-0297.2004.00235.x.

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.

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

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.

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.

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.

Kristoffersen, Ingebjørg. 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.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2008. 'The Marginal Utility of Income'. *Journal of Public Economics* 92 (8): 1846–57.

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.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *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.

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.

Nikolova, Milena, and Carol Graham. 2021. '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. 2020. *How's Life? 2020: Measuring Well-Being*. How's Life? OECD. https://www.oecd-ilibrary.org/economics/how-s-life/volume-/issue-_9870c393-en. https://doi.org/10.1787/9870c393-en.

ONS. 2021. 'Well-Being - Office for National Statistics'. 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing.

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.

Samuelson, Paul A. 1950. 'The Problem of Integrability in Utility Theory'. *Economica* 17 (68): 355–85.

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.

Stevenson, Betsey, and Justin Wolfers. 2008. 'Happiness Inequality in the United States'. *The Journal of Legal Studies* 37 (S2): S33–79.

Stigler, George J. 1950. 'The Development of Utility Theory. I'. *Journal of Political Economy* 58 (4): 307–27.

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.

Van Praag, Bernard M. S. 1991. 'Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept'. *Journal of Econometrics* 50 (1–2): 69–89.

———. 2015. 'A New View on Panel Econometrics: Is Probit Feasible After All?' *IZA Discussion Paper*, no. 9345.

Vendrik, Maarten C. M., and Geert B. Woltjer. 2007. 'Happiness and Loss Aversion: Is Utility Concave or Convex in Relative Income?' *Journal of Public Economics* 91 (7–8): 1423–48. https://doi.org/10.1016/j.jpubeco.2007.02.008.

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

———. 2021b. '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.

(Online) 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 (1) and (2) 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:

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

In moving from the second to the third line above, we used the fact that $\widehat{\boldsymbol{\beta}}_k^{(d)} = ((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{hd_k}$, and hence $((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')^{-1}\widehat{\boldsymbol{\beta}}_k^{(d)} = \boldsymbol{hd_k}$. The term $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'l_K\boldsymbol{I}$ 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 $k = 1, \dots, K-1$, 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 $k = 1, \dots, K-1$, $\hat{\beta}_m$ 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 $\mathrm{E}(\zeta_i|\boldsymbol{X}_i) = \sum_{k=1}^{K} s_k * \mathrm{E}(\zeta_i|hr_i = k; \boldsymbol{X}_i)$. 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 $\lim_{\iota_k - \iota_{k+1} \to 0 \text{ for } k \neq b} \hat{\beta}_m/\hat{\beta}_n \to \hat{\beta}_{b,m}^{(d)}/\hat{\beta}_{b,n}^{(d)}$. 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:

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

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

$$\sum_{k=1}^{K} \iota_{k-1} s_{A,k} - \iota_k s_{B,k} < \mathrm{E}[ht_A] - E[ht_B] < \sum_{k=1}^{K} \iota_k s_{A,k} - \iota_{k-1} s_{B,k}. \tag{A4}$$

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:

$$\begin{aligned} \mathrm{E}[ht_A] - E[ht_B] &> \sum_{k=1}^{K} \iota_{k-1} s_{A,k} - \iota_k s_{B,k} \\ &= (\iota_0 - \iota_1) s_{A,1} + (\iota_{K-1} - \iota_K) s_{B,K} + \sum_{k=2}^{K-1} (\iota_{k-1} - \iota_k) \left(\sum_{m=1}^{k} s_{A,m} - \sum_{m=1}^{k-1} s_{B,m} \right). \end{aligned} \tag{A5}$$

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):

$$\begin{aligned} &(\iota_0 - \iota_1) s_{A,1} + \sum_{k=2}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ &= \iota_0 s_{A,1} - \iota_1 s_{A,1} + \sum_{k=2}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ &= \iota_0 s_{A,1} + \iota_1 s_{A,2} - \iota_2 s_{A,1} - \iota_2 s_{A,2} + \sum_{k=3}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ &= \iota_0 s_{A,1} + \iota_1 s_{A,2} + \iota_2 s_{A,3} - \iota_3 s_{A,1} - \iota_3 s_{A,2} - \iota_3 s_{A,3} + \sum_{k=4}^{K-1} (\iota_{k-1} - \iota_k) \sum_{m=1}^{k} s_{A,m} \\ &= \cdots = \sum_{k=1}^{K-1} \iota_{k-1} s_{A,k} - \iota_{K-1} \sum_{k=1}^{K-1} s_{A,k} \\ &= \sum_{k=1}^{K-1} \iota_{k-1} s_{A,k} - \iota_{K-1} (1 - s_{A,K}) \\ &= -\iota_{K-1} + \sum_{k=1}^{K} \iota_{k-1} s_{A,k}. \end{aligned} \tag{A6}$$

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

$$(\iota_{K-1}-\iota_K)s_{B,K}-\sum_{k=2}^{K-1}(\iota_{k-1}-\iota_k)\sum_{m=1}^{k-1}s_{B,m}=\iota_{K-1}-\sum_{k=1}^{K}\iota_k s_{B,k}. \quad \text{(A7)}$$

Combining the results of (A6) and (A7) yields the expression in the first line of (A5).

The expression in (A5) 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 (A5) 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 (A5) 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 Implications of Proposition 5 for linear response scales

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 of the thresholds ι_k as a function of $hr_i=k$ is linear. The conditions for identification of rankings of means are less demanding in this case. They can be derived as follows.

In this case of a linear response scale 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 (A4). 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 (A5). 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 (A5) as:

$$E[ht_A]-E[ht_B]>\sum_{k=1}^{K}(\iota_k-d)s_{A,k}-\iota_k s_{B,k}=\sum_{k=1}^{K}\iota_k\left(s_{A,k}-s_{B,k}\right)-ds_{A,k}. \quad \text{(B1)}$$

Note that we can write d as $K/(\iota_K-\iota_0)$, 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 relation (B1) becomes:

$$E[ht_A]-E[ht_B]\geq\sum_{k=1}^{K}\iota_k\left(s_{A,k}-s_{B,k}\right)-\frac{1}{K}. \quad \text{(B2)}$$

The first term in this expression $\sum_{k=1}^{K} \iota_k (s_{A,k} - s_{B,k})$ 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 $\sum_{k=1}^{K} \frac{k}{K} (s_{A,k} - s_{B,k})$. 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 A2. 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.

Further 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 $E[ht_A] - E[ht_B] \geq \sum_{k=1}^{K} k(s_{A,k} - s_{B,k}) - 1$.Thus, we only require that the difference in mean rank-order-coded hr exceeds the value 1 between groups. Strikingly, this condition applies no matter the number of response options available.

However, a peculiarity of this case of linear response scales is that sign reversals cannot occur as a result of transformations of discrete hr scales as in OLS. This is because transformed scales are again linear, and hence imply identical signs of the differences in mean hr between groups. Sign reversals of differences in mean ht between groups can only occur as a result of intergroup differences in mean ht within each response category. Moreover, the latter differences have then to be oppositely signed to the differences in mean hr between the groups (see section 4.3 for more details). On top of that, when differences in mean rank-order hr between two groups are not much smaller than 1, sign reversals of differences in mean ht will require bunching up of ht near the borders of response categories (cf. the discussion in section 2.1 of B&L). Since linear response scales are bounded and for a high number of response categories K, both conditions seem a priori unlikely to occur in empirical practice. Finally, because of the boundedness of linear response scales, non-parametric identification of the rank order of mean happiness in two groups does not seem hard to achieve in most of these cases (again cf. section 2.1 of B&L).

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. For simplicity, we focus our analysis on the case of exponential functions. First write the normally distributed error εp_i in Equation (4) as $\sigma_i \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0,1)$ with density function $\varphi(\epsilon_i)$ 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 $\mathrm{E}(ht_i|\boldsymbol{X_i})$ as an integral over error ϵ_i. We then obtain:

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

Hence, the marginal effect of $X_{i,m}$ on mean ht_i (i.e. $\mathrm{E}\big(ht_{i_i}\big|\boldsymbol{X_i}\big)$) equals the mean effect of $X_{i,m}$ on individual ht_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 $ht_i = e^{chp_i}$ 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:

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

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 (4), the fact that $\varepsilon p_i = \sigma_i \epsilon_i$, and the relation $\sigma_i = e^{\ln(\sigma_i)} = e^{\boldsymbol{X_i \beta^{(s)}}}$, this local effect equals $\beta_m^{(p)} + \beta_m^{(s)} \sigma_i \epsilon_i$. Hence, we can write:

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

Thus, the marginal effect of $X_{i,m}$ on $\mathrm{E}(ht_i|\boldsymbol{X_i})$ 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 and ϵ_i. Hence, for given $c>0$ this weight is relatively higher for negative local effects beyond $\epsilon_i = -\frac{\beta_m^{(p)}}{\beta_m^{(s)} \sigma_i}$ in the right half of the standard normal distribution than for the positive local effects for lower ϵ_i. Nevertheless, for sufficiently low values of c the overall marginal effect of $X_{i,m}$ on $\mathrm{E}(ht_i|X_i)$ is positive.[21] However, 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 ht_i.[22] This is analogous to

[21] Note that for $c = 0$ the weights of the local effects are given by $\varphi(\epsilon_i)$. It is easily shown that these are the weights in an integral expression like (B5) for the overall effect of $X_{i,m}$ on the untransformed $\mathrm{E}(hp_i|X_i)$. This effect equals $\beta_m^{(p)}$, and hence is positive.

[22] This heterogeneity is equivalent to $\mathrm{E}(ht_i|\boldsymbol{X_i})$ being non-linear in X_i. 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.

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 $E(\mathrm{e}^{c\sigma_i\epsilon_i}) = e^{0.5c^2\sigma_i^2}$. 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:

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

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.

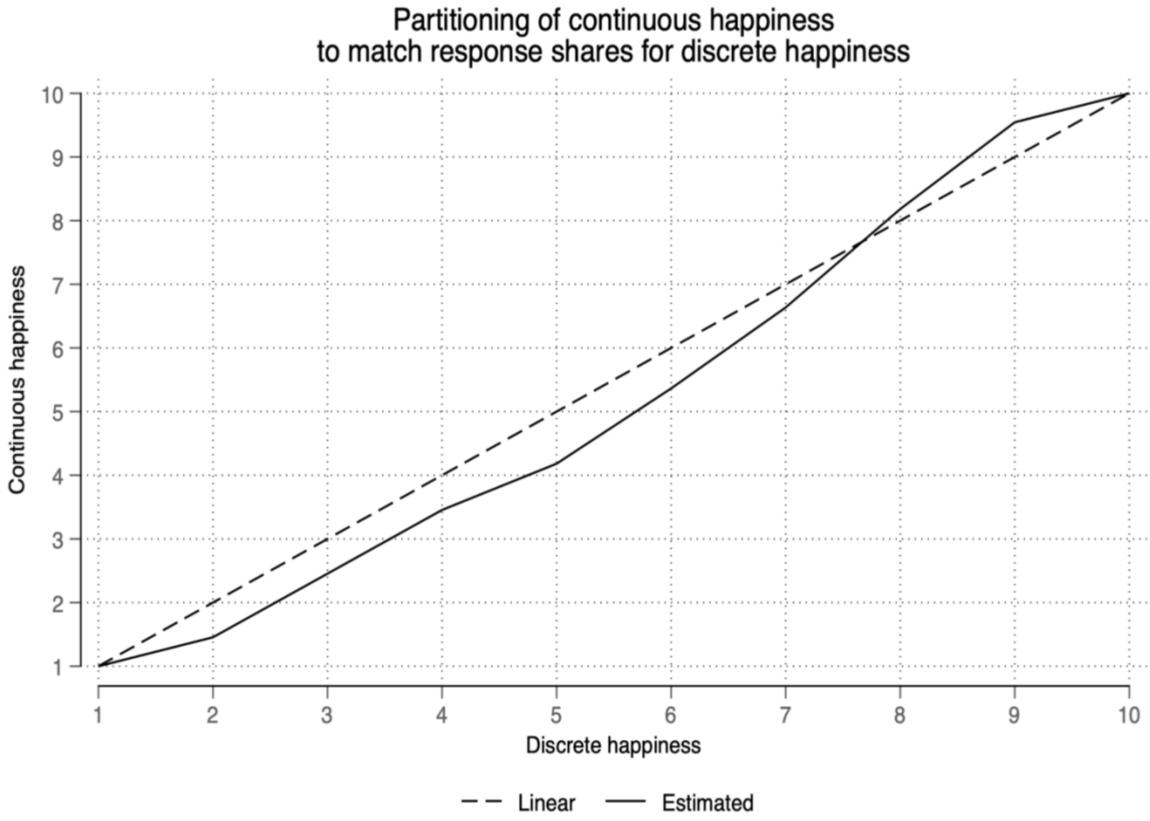
B3 Comparing scales with different numbers of response options.

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 $\tau_k, k = 0, \dots, 10,$ 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 B1. 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

Figure B1. Partitioning of continuous happiness to match response shares for discrete happiness (based on LISS data).



scale is a linear collapse of the continuous response scale. This is further evidence of approximately linear scale use for discrete response scales.

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-point scale of the 2006 GSS wave with responses to the ten-point 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-point or an eleven-point scale for a question on happiness in the United States. Nevertheless, both samples are representative of the same population, and we assume that the concepts of ht that are measured by the two questions are sufficiently strongly correlated to each other.

Table B2 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-point 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-point scale.

Assume now that the relative distribution of responses across the ten-point scale in the WVS sample (measuring life satisfaction) is a reasonable approximation of the distribution of responses

Table B2. 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.

we would observe had the GSS sample (measuring happiness) been given a ten-point scale. Further assuming that the ten-point 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 B2 for an illustration of this analysis, which shows a mildly concave pattern.

Furthermore, using WVS (four-point scale for happiness) and ESS (eleven-point scale for happiness) data, we also applied a similar procedure to a set of 14 European countries. As shown in Table B3 below, that exercise shows that differences between responses on the four-point WVS scale collapse in a roughly linear manner onto the eleven-point ESS scale. The figures for mean hr in the fifth column of Table B3 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

Table B3. Cumulative response shares for happiness and life satisfaction in ESS and WVS for European countries.

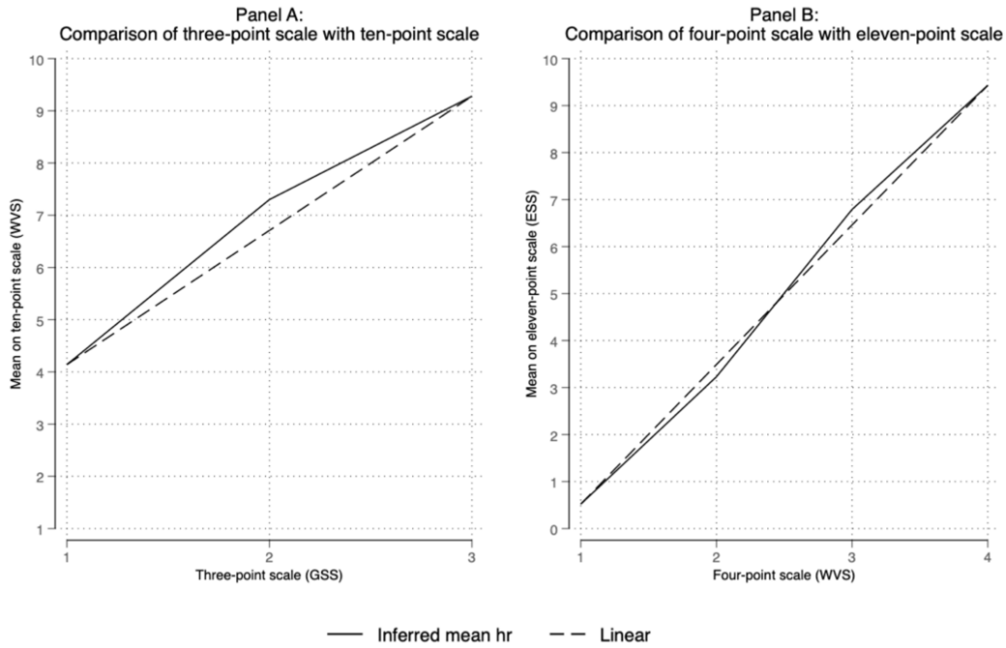
WVS (Happiness)		ESS (Happiness)			ESS (Life Satisfaction)		WVS (Life Satisfaction)	
hr	% share (cumulative)	hr	% share (cumulative)	$\overline{hr}$ after collapse	hr	% share (cumulative)	hr	% share (cumulative)
1	2.45 (2.45)	0	0.97 (0.97)	0.52	0	3.26 (3.26)	1	2.40 (2.40)
		1	1.04 (2.01)		1	2.15 (5.41)	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	18.77 (54.64)
		7	18.70 (55.00)		7	16.25 (61.29)		
		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).

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 B2 for an illustration of this analysis.

Taken together, the analysis of this appendix suggests that mildly convex/concave scales may be plausible for questions with three response options, but less so for questions with more response options.

Figure B2. Illustration of results of Tables A1 and A2.



C Additional Tables

Table C1. 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: Data are from the March and April 2011 waves of the LISS. Model titles denote the dependent variable used in each column. Standard errors in parentheses (clustered by respondent). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C2. 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. Model titles denote the dependent variable used in each column. Standard errors in parentheses (clustered by respondent). * $^{+}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C3. 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_2 (no controls)	(5) hd_1 (full controls)	(6) hd_2 (full controls)
Log eq. HH income	0.119***	0.076***	-0.057***	-0.061***	-0.039***	-0.038***
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)
Unemployed	-0.330***	-0.213***	0.180***	0.150***	0.136***	0.077***
	(0.023)	(0.024)	(0.014)	(0.013)	(0.014)	(0.013)
Married	0.289***	0.289***	-0.102***	-0.188***	-0.101***	-0.188***
	(0.011)	(0.010)	(0.004)	(0.008)	(0.004)	(0.007)
Has children	0.032**	-0.053***	**0.007**	-0.038***	0.027***	0.026***
	(0.009)	(0.007)	**(0.003)**	(0.007)	(0.004)	(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). Model titles denote the dependent variable used in each column. Standard errors in parentheses (clustered by year). * $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C4. 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)
$\mathbf{ln}(\boldsymbol{\sigma_{it}})$		
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. Model titles indicate specifications used in each column. Standard errors in parentheses (clustered by respondents). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

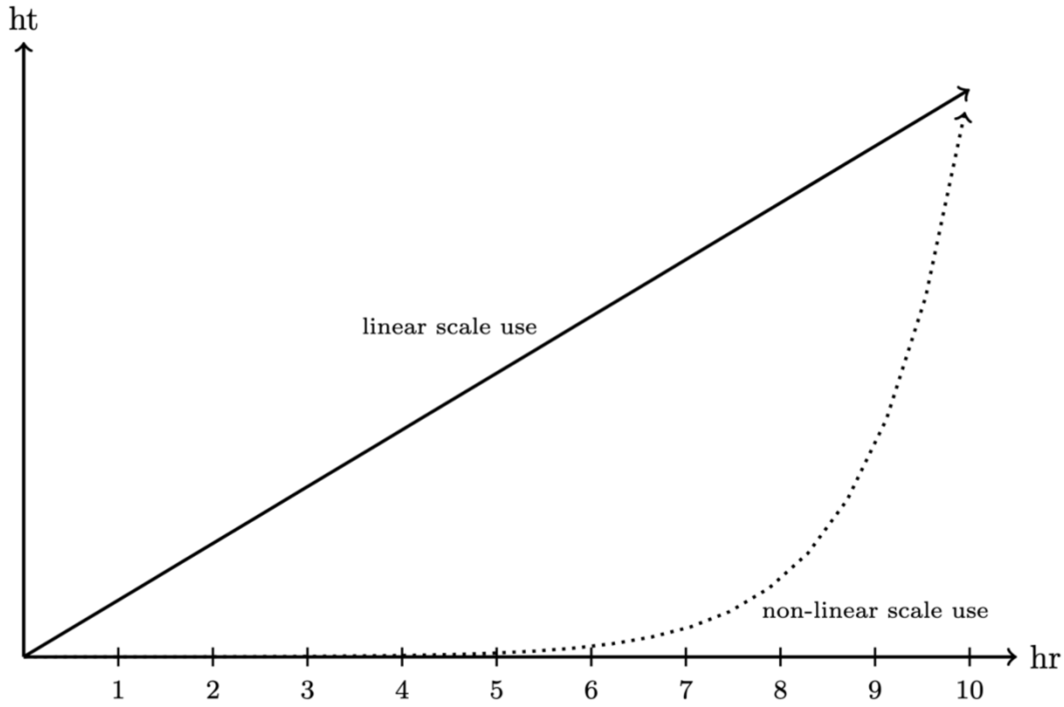
Table C5. Results of heteroskedastic ordered probit models for hr on individual-level socio-economic variables using GSS data.

	(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)
$\mathbf{ln}(\boldsymbol{\sigma_{it}})$		
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). Model titles indicate the specification used in each column. Standard errors in parentheses (clustered by year). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

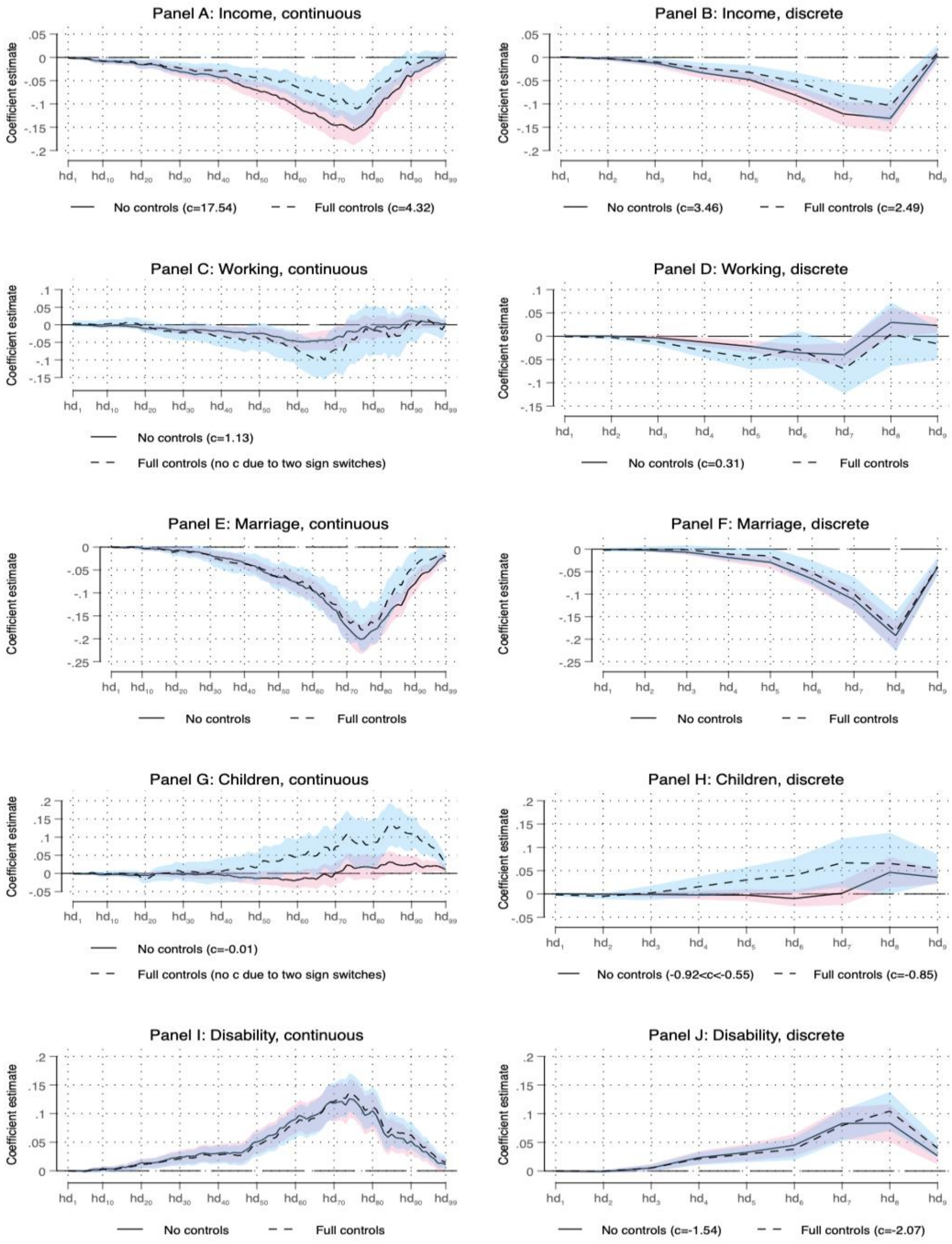
D Additional Figures

Figure D1. Two-dimensional illustration of linear and non-linear scale use.



Note: This is an alternative illustration of the two kinds of response scales shown in Figure 1.

Figure D2. Coefficient estimates for each regression of $hd_{k,it}$ using LISS data.



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