A Robustness Test for the Analysis of Human Feelings*

Jeffrey R. Bloem† and Andrew J. Oswald‡

July 13, 2020

Abstract

Governments, multinational companies, and researchers today collect previously unprecedented amounts of data on human 'feelings.' These new kinds of data provide information on citizens' happiness, levels of customer satisfaction, mental stress, societal trust, political sentiment, and many other important variables. Yet a key scientific difficulty tends to be downplayed, or even ignored, by many users of such information. Human feelings are not measured in objective cardinal units. This paper aims to address some of the ensuing empirical challenges. It suggests an analytical way to think about, and to approach, the scientific complications of ordinal data on human feelings. The paper describes a *dichotomous-around-the-median (DAM) test*, which, crucially, uses information only on direction within an ordering and deliberately discards the potentially unreliable statistical information in ordered data. Applying the proposed DAM approach, the paper demonstrates how it is possible to check and confirm some of the key conclusions of previous research—including earlier work on the effects upon human well-being of higher income, greater societal trust, and lower levels of corruption.

Keywords: Ordinal scales, Satisfaction, Subjective well-being, Happiness, Trust, Corruption, Robustness

JEL Codes: C18, C25, I31, I39

*The second author, AJO, wishes to record that the first author, JRB, wrote the first draft of this paper, contributed the key starting ideas, and should be assigned the great majority of any credit for this work. We thank Tim Bond for comments. All errors are our own.

†Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, Saint Paul, MN 55108. Email: bloem023@umn.edu.

‡Department of Economics, University of Warwick, CV4 7AL, UK. Email: andrew.oswald@warwick.ac.uk.

1 Introduction

"Feelings data" are all around us. International policy-makers have begun to incorporate measures of citizens' well-being into official statistics and the evaluation of public policies (Stiglitz et al. 2009; Durand 2015; Graham 2016, 2018; Centers for Disease Control 2019; OECD conference 2019). Businesses across the world now collect enormous amounts of information on customer satisfaction and employee engagement: such numbers long ago seemingly passed a key market test (to use Chicago-esque jargon). Organizations like Gallup (2019) provide regular surveys of people's feelings across the globe. For many decades, moreover, scientific researchers have included concepts such as life satisfaction, mental health, trust, and political corruption in formal regression analysis. Highly cited articles, published in journals such as the *Economic Journal*, the *Journal of Public Economics*, the *Journal of Development Economics*, the *American Economic Review*, the *Quarterly Journal of Economics*, *Review of Income and Wealth*, and the *Proceedings of the National Academy of Sciences of the USA*, empirically analyze such data (e.g., Frey and Stutzer 2000; Easterlin 2001; Alesina et al. 2004; Blanchflower and Oswald 2004; Ferrer-I-Carbonell and Frijters 2004; Booth and van Ours 2008; DiTella and MacCulloch 2008; Fafchamps and Shilpi 2008; Luechinger 2009; Clark and Senik 2010; Nunn and Wantchekon 2011; Verme 2011; Benjamin et al. 2012; Kahneman and Deaton 2010; Clark et al. 2016). Major portions of particular disciplines, such as psychology and psychiatry, rest on the analysis of this form of ordinal data. As described in recent reviews by Diener et al. (2017) and Clark (2018), there are now, as one example, tens of thousands of published articles on the empirical study of human well-being. Easterlin (1974) was a seminal paper.

The appropriate use of these kinds of data, however, presents scientific challenges. Those have typically been played down or ignored by government agencies, by commercial organizations, and often also in published academic research. This is the issue on which we focus. Our paper builds particularly upon the work of researchers such as Ferrer-I-Carbonell and Frijters (2004), Oswald (2008), Ravallion et al. (2016), Schröder and Yitzhaki (2017), Bond and Lang (2019), Chen et al. (2019), Kaiser and Vendrik (2019), and Bloem (2020).

By definition, an ordinal variable (measuring, for example, human happiness or customer satisfaction or feelings of trust) cannot be converted into objective cardinal units. Consider a question such as "All in all how satisfied are you with your life right now?" with several ordered answer categories: "very satisfied," "satisfied," "unsatisfied," and "very unsatisfied." The lack of information about the interval between response categories presents a conceptually deep empirical problem.

Since ordered-response categories only provide information about rank, and not the interval between categories, standard empirical methods—such as comparisons of means or linear regression analysis—can lead to invalid estimates.

These challenges have been known for many decades. Recent writings, however, help to clarify their nature. The work of Schröder and Yitzhaki (2017), for instance, explains the possible perils of assuming arbitrary and fixed intervals between each of the categories on an ordinal scale. The practice of comparing means or using a linear regression ignores the fact that, in principle, any transformation of the ordinal scale is theoretically permissible if it preserves the ordered rank of categories but changes the interval between categories. Additionally, the work of Bond and Lang (2019) argues that empirical results using ordinal response regression approaches (e.g., an ordered logit or probit regression) implicitly assume a specific distribution on the error term. Allowing for possible deviations from the assumed functional forms of the error term, Bond and Lang (2019) suggest that prominent results in the microeconomic literature on happiness might be uninformative. Since the true distribution of the error term is unknown, Bond and Lang (2019) are not able to establish that this literature is *definitively* incorrect. Previously published results may be either correct or incorrect, but careful analysis is necessary. In this paper we aim to provide some practical guidance for this analysis.

One response to these conceptual challenges would be to decline to use ordinal data. This extreme stance, however, has not been taken by modern governments and multinational companies. That is probably because its implied nihilism has a fundamental disadvantage. It means turning our backs, as a community, on a great deal of potentially valuable statistical information about the perceptions and attitudes that are central to people's lives. Despite the potential complications of using ordinal data, nihilism carries its own empirical dangers. Some balance, between conceptual purity and practical relevance, has to be struck, whatever the branch, in applied statistical science. One aim of the paper is to try to find a central ground that offers such balance.

As one example of the dangers, Ronald A. Fisher, then the most famous statistician in the world, famously refused to accept the early evidence that smoking caused lung cancer. He went to his grave assuring everyone that it was safe to smoke (Pearl and Mackanzie 2018). Fisher objected to the early researchers' cross-sectional statistical methods. He believed those methods were potentially flawed. "The curious associations with lung cancer found in relation to smoking habits do not [...] lend themselves easily to [...] simple conclusion [...]. Such results suggest that an error has been made of an old kind, in arguing from correlation to causation. There is nothing to stop those who greatly desire it from believing that lung cancer is caused by smoking cigarettes. [...] To believe this is, however, to run the risk of failing to recognize [...] more genuine causes" (Fisher 1958). Unfortunately for Fisher's standing in medical history, he turned out to be deeply wrong, in a substantive sense, even though his methodological concerns were in principle technically appropriate. If the world had listened to Fisher, large numbers of humans would have died prematurely. Of course, *ex ante*, we cannot know which statistical results are robust to future more rigorous analysis. Therefore, caution and pragmatism, of a scientifically constructive kind, are appropriate in the face of potential methodological concerns.

If ignoring all ordinal "feelings data" is unsatisfactory, how should researchers handle these ordinal variables in empirical analysis? Although some researchers object to the above critiques (see counter-arguments raised by Chen et al. 2019 and Kaiser and Vendrik 2019), the goal of the current paper is to be pragmatic and to attempt to be constructive. Given the foregoing, it seems imperative that empirical researchers use credible methods when using an ordinal dependent variable.[1]

This paper discusses a validity check on ordinal feelings variables. This check, which we call the dichotomous-around-the-median (DAM) test, relies only on the direction within ordinal data. This approach builds on the premise that one of the major challenges with the empirical analysis of ordinal variables is absent when analyzing dichotomous variables. To implement the method, we deliberately discard some information captured by the ordinal scale—namely, information about the full ordering of all the answers. This is done to ensure that our method does not rely on arbitrarily assumed interval information. In the DAM test, the ordinal dependent variable is redefined as a dichotomous variable for threshold points around the median value of the ordinal scale. Robustness to these DAM tests suggests robustness to rank-preserving transformations of the ordinal scale. This DAM method, it should be emphasized, consciously fails to use all the potential statistical information in an ordered data set. It does this precisely because some researchers object to, and doubt the reliability of, within-ranking information.

We then revisit a number of previous and familiar empirical results. First, we demonstrate the DAM test by re-analyzing the effect of unconditional cash transfers on psychological well-being in Kenya (Haushofer and Shapiro 2016) and the effect of the slave trade on trust in sub-Saharan

[1] In this article we focus on the use of ordinal *dependent* variables, as these tend to present applied researchers with a more difficult problem than compared to the use of ordinal explanatory variables. Indeed one common approach for using an ordinal explanatory variable is to separate the ordinal variable into distinct dichotomous variables.

Africa (Nunn and Wantchekon 2011). Second, we use the same approach to check the conclusions of prominent research articles on wellbeing, life satisfaction, and corruption. This bears on the findings in work such as Blanchflower and Oswald (2004), Fan et al. (2009), Triesman (2000), Kahneman and Deaton (2010), and Stone et al. (2010).

The rest of this paper is organized as follows. The next section highlights the the challenge of empirically analyzing ordinal variables. In section 3 we make the point that, despite the deep challenges, the use of ordinal variables is not going away anytime soon. Section 4 introduces and describes our DAM approach. Finally, we conclude in section 5.

2 An Intuitive Introduction to the Analytical Challenge

The challenge of empirically analyzing ordinal variables can be explained as follows. Think of a person who is about to fill in a questionnaire about her feelings. Call the "reporting function" the way the person looks inside herself and then writes answers about how she feels. Imagine that this individual actually has constant marginal utility of income. But now consider the possibility that, as she feels cheerier, she marks herself happier on a questionnaire scale in a way in which she is intrinsically reluctant to approach the highest level on the questionnaire form (the 5 on a 1-5 scale, for example). Then the reporting function itself is curved. In this case, we will have the illusion, when we study her pay increases over time and analyze the patterns in subjective well-being data, that true diminishing marginal utility of income has been established empirically. Yet it has not. The numbers written down by the person will be due, in part, to the sheer curvature of the reporting function.

The root of the problem here is not simply caused by the—perhaps inevitable—reality that subjective well-being is measured on an ordinal scale with discrete boxes for happiness of 5, 4, 3, etc. The problem would persist even if surveys got people to provide exact numerical answers anywhere on the real number line (Oswald 2008).

Consider a further example, as set out numerically in Table 1. A survey asks respondents to answer the following question: "All in all how satisfied are you with your life right now?" using the following ordered response categories: "very satisfied," "satisfied," "unsatisfied," and "very unsatisfied." To simplify this example even further, suppose there are two groups of people, each with only two members. Group A includes one person who is "Very Dissatisfied" and another who is "Very Satisfied." Group B includes one person who is "Dissatisfied" and another who is

"Satisfied." A seemingly simple question is: Which group is more satisfied? The answer to this question, however, is not simple. That is because the answer depends on the interval between the response categories. Perhaps a natural way to empirically answer this question is to assume a linear set of values for the response categories. In this case, "Very Dissatisfied" has a value of zero, "Dissatisfied" has a value of one, "Satisfied" is two, and "Very Satisfied" is three. The average satisfaction of the two groups is equal, with an average score of 1.5, and the groups are equally satisfied.

Similar to utility functions, however, ordinal scales only offer information about the relative rank of response categories and thus provide no information about the interval between categories. Therefore, a potentially valid alternative way to answer this question is to assume a concave set of values for the response categories. In this case, "Very Dissatisfied" again has a value of zero, "Dissatisfied" has a value of 1.75, "Satisfied" has a value of 2.5, and "Very Satisfied" again has a value of three. With this set of values, group B is more satisfied than group A. Finally, another valid alternative way to answer this question is to assume a convex set of values for the response categories. In this case, "Very Dissatisfied" again has a value of zero, "Dissatisfied" has a value of 0.5, "Satisfied" has a value of 1.25, and "Very Satisfied" again has a value of three. With this set of values, group A is more satisfied than group B. Table 1 illustrates this example and shows that the answer to the seemingly elementary question above depends on the assumed intervals between the response categories.

An additional issue is that of possible cross-sectional heterogeneity. That is, not only do we not have information about the intervals between response categories, but it is also possible—intuitively—that different individuals might interpret the distances between the response categories differently. Although this is an area for future research, we follow previous literature that analyzes an ordinal variable and assume no cross-sectional heterogeneity.

3 The Potential Use and the Potential Misuse of Ordinal Variables

Over past decades, researchers have increasingly studied variables measured on an ordinal scale. To date, for example, the single most-cited paper published in the history of the well-known *Journal of Public Economics* analyzes corruption using a zero through three ordinal scale as the dependent variable (Treisman 2000). While discussing the psychology literature on subjective well-being, Diener et al. (2017) note that in 2015 alone over 14,000 publications on Google Scholar mentioned subjective well-being. Additionally, while reviewing the economics literature on human happiness, Clark (2018) notes that four of the 20 most-cited articles ever published in the *Economic Journal* have the word "happiness" in their title—including Easterlin (2001)—and two of the three most-cited articles ever published in the *Journal of Public Economics* investigate subjective well-being.

The challenge is not limited to a huge research literature on "subjective well-being" (e.g., Easterlin 1974; Oswald 1997; Luttmer 2005; Graham 2005; Fafchamps and Shilpi 2008; Di Tella and MacCulloch 2008; Dedehouanou et al. 2013; Aghion et al. 2016; Boertien and Vignoli 2019; Perelli-Harris et al. 2019). It applies in contexts where any of the following variables are used as primary outcomes of interest: "satisfaction" (Frijters et al. 2004; Clark and Oswald 1994, 1996; Ritter and Anker 2002; Luechinger et al. 2010), "trust" (Nunn and Wantcheckon 2011; Putnam 2001), "corruption" (Treisman 2000; Fan et al. 2009; Martimort and Straub 2009), "work feelings" (Bryson and MacKerron 2017); "feelings of comparison" (Clark and Senik 2010; Budria and Ferrer-I-Carbonell 2019), "visibility" (Heffetz 2011), "hope" (Bloem et al. 2018; Glewwe et al. 2018), measures of mental well-being and personality traits (Borghans et al. 2008; Baird et al. 2013; Cornaglia et al. 2014), measures of "affect" (Krueger et al. 2009; Krueger 2017), measures of "quality" (Acemoglu et al. 2001), and standardized test scores (Bond and Lang 2013; Glewwe 1997; Jacob and Rothstein 2016; Lang 2010; Schröder and Yitzhaki 2016). Therefore, the matters discussed in this paper are in principle relevant to a variety of disciplines, including also: medicine, psychology, politics, marketing, biometrics, labor relations, psychiatry, and sociology.

Nor is the valid use of ordinal variables just an academic issue. Large corporations use ordinal scales to measure information about how their customers feel about their products and services. Delta and British Airways send customer-satisfaction surveys via email to customers after a flight. Amazon has a five-star ranking system to record customer satisfaction. Currently, Amazon reports an aggregate score that simply takes the average of the one-through-five ordinal scale, assuming a linear set of values associated with each interval. Additionally, the ride-share apps Uber and Lyft employ the same method to evaluate the quality of their drivers based on customer reports.[2]

[2]On their Community Guidelines page, Uber seems to acknowledges the ambiguity of ordinal scales but ultimately sets a minimum average rating that must be met for drivers to maintain access to their account. "There is a minimum average rating in each city. This is because there are cultural differences in the way people in different cities rate each other. We will alert you over time if your rating is approaching this limit, and you'll also get information about quality improvement courses that may help you improve. However, if your average rating still falls below the minimum after multiple notifications, you will lose access to your account. See https://www.uber.com/legal/community-guidelines/us-en/.

As explained earlier, recent work raises questions on the validity and robustness of some of the academic studies and corporate practices discussed above. In particular, Schröder and Yizhaki (2017) examine the reliability of the results reported by Ferrer-i-Carbonell and Frijters (2004). They find that the empirical conclusions are not robust to all the potential monotonic increasing transformations (i.e., transformations of the ordinal scale that maintain the relative rank of categories but change the interval between categories). Similarly, Bond and Lang (2019) attempt to re-evaluate several well-known results from the economics of happiness literature. These include: Easterlin's paradox (Easterlin 1974, 2001), U-shaped life-cycle happiness (Blanchflower and Oswald 2004), the unemployment-inflation trade-off (Di Tella et al. 2001), ranking countries based on happiness, results from the Moving to Opportunity project (Ludwig et al. 2012), the effect of marriage and children on happiness (Diener et al. 2000; Blanchflower and Oswald 2004), declining female happiness (Stevenson and Wolfers 2009), and adaption to disability (Kahneman 2011; Oswald and Powdthavee 2008). Ultimately, Bond and Lang (2019) argue that none of these pass their (admittedly rather extreme) theoretical requirements for the valid use of ordinal variables. See also the interesting issues raised in an important paper by Ravallion et al. (2016).

The core point of Schröder and Yitzhaki (2017) is that results using ordinal scales must be robust to monotonic increasing transformations of the *observed scale*. That is, empirical results must be robust to rank-preserving transformations of the ordinal scale that change the interval between the observed categories. A related but slightly different point of Bond and Lang (2019) is that results should ideally be robust to monotonic increasing transformations of the *latent variable* (say, for example, the unobserved distribution of happiness states). They show that non-parametric estimation techniques are likely not feasible in most empirical settings. There are two feasible approaches for parametric estimation. First, the researcher can assume that the ordinal scale is, in fact, an interval scale. Although this assumption is common, any rank-preserving cardinalization of the ordinal scale is, in principle, theoretically permissible. Second, the researcher can use an ordered-response regression and make the assumption that the distribution of the latent variable is normal (i.e., when using an ordered probit) or logistic (i.e., when using an ordered logit). Bond and Lang (2019) proceed to show that results using an ordered probit or ordered logit can in certain cases be reversed when assuming a different, and theoretically permissible, functional form of the distribution of the latent variable.

4 Dichotomous-Around-the-Median Tests

In introducing and discussing the following validity check, we attempt to take seriously these critiques and provide advice for how to calculate credible empirical results when using ordinal variables. This proposed test does not diminish the importance of acknowledging the underlying challenges. We hope, however, that the checks allow for the valuable information stored in ordinal variables to credibly inform empirical analysis.

One way to think of the complication of ordinal scales is that some researchers have come to put too much emphasis on the multiple levels of ordinal responses—for example the 4 levels of possible responses in Table 1. It is this multiple ordering, and the unknown intervals between categories, that generates the key problem.

Therefore, we begin in what may seem a counter-intuitive way—by deliberately discarding statistical information (about, in particular, some of the levels). First, and crucially, we exploit a property that can be relied upon, namely, the known qualitative direction within an ordinal response scale. No matter how many ostensible levels there are on an ordinal scale, the direction—which way is high and which way is low—is known and unambiguous. Second, the multiple-ordering aspect can then be avoided by compressing multiple scales into a single dichotomous scale. In the case of a customer-satisfaction scale in marketing science, for example, this would mean taking the multiple answers "Completely Satisfied," "Very Satisfied," [...], "Not Satisfied" and reclassifying them into only two categories.

The idea here is consciously to reduce an ordinal scale down into a dichotomous variable. Robustness of results to dichotomous tests suggest that core results are robust to rank-preserving transformations of the ordinal variable. Where exactly should the dichotomous split be inserted? We suggest that the use of a dichotomous test around the median of responses is particularly natural (although it should be emphasized that other choices about the position of the dichotomous split are feasible). This is because to divide the data at the median is one way to allow as much statistical information as possible into the upper and lower parts of the distribution.

We initially use two applications to illustrate our DAM method. In the first, we re-examine the effects of randomly distributed unconditional cash transfers on psychological well-being in Kenya (Haushofer and Shapiro 2016). Among other outcomes of interest, the authors of that study examine the treatment effect on happiness and life satisfaction measured using the questions included in the World Values Survey. Specifically, happiness is measured via the following question: "Taking all things together, would you say you are "very happy," "quite happy," "not very happy," or "not at all happy?" with the categories enumerated one through four. Similarly, life satisfaction is measured via the following question: "All things considered, how satisfied are you with your life as a whole these days on a scale of one to ten?" with one indicating "very dissatisfied" and ten indicating "very satisfied."

In the second application, we re-examine the effects of the trans-Atlantic slave trade on trust in sub-Saharan Africa (Nunn and Wantchekon 2011). Trust is measured via the Afrobarometer survey along five dimensions—trust of relatives, neighbors, the local council, intra-group trust, and inter-group trust—with the following categories: "not at all," "just a little," "somewhat," and "a lot." The authors code these categories from zero through three, with zero representing "not at all" and three representing "a lot."

In each of these data sources, the use of the default cardinalizations of the ordinal data implicitly makes the assumption that the intervals between each of the categories are known, represented by a unitary change of an integer value, and consistent across the entire scale. Yet, as noted above, any transformation of these numerical values that preserves the order of the response categories is theoretically permissible. To test the robustness of the core results to this assumption, we apply our DAM test.

Implementing the approach is fairly simple. For each application, we first calculate the median of the ordinal dependent variable. One complication is that many statistical distributions are asymmetric, so that the median can lie within an answer category. Thus, for completeness, we construct two dichotomous variables. The first includes the median value within the "upper" category and the second includes the median value within the "lower" category. To ensure comparability with results using the "original" ordinal scale, these variables are then standardized into z-scores.

Table 2 reports results from our dichotomous-around-the-median tests of the effect of unconditional cash transfers on psychological well-being in Kenya (Haushofer and Shapiro 2016). Panel A shows results on happiness, and Panel B shows results on life satisfaction. Column 1 replicates the original OLS effect estimates as reported by Haushofer and Shapiro (2016). The remaining columns report OLS results from our DAM tests around the median. In column 2 the dependent variable is a standardized dichotomous variable that equals one if the original scale value is greater than *or equal to* the median. In column 3 the dependent variable is a standardized dichotomous variable that equals one if the original scale value is strictly greater than the median.

Both columns 2 and 3 in Panel A of Table 2 show that the original estimate of the effect on happiness is robust to our DAM tests. Similarly, in Panel B, both columns 2 and 3 show that the original estimate of the effect on life satisfaction is robust to our DAM tests. The original effect-size estimate suggests that receiving the cash transfer increased happiness by 0.16 standard deviations and increased life satisfaction by 0.17 standard deviations. Our DAM tests report a slightly larger range of effect sizes on happiness, 0.12 through 0.16 standard deviations, but these estimates are qualitatively consistent. The DAM tests for the effects on life satisfaction have a relatively narrow range, 0.17 through 0.19 standard deviations. Taken together, the original estimates of the effect of unconditional cash transfers on psychological well-being seem to hold up encouragingly in the face of our DAM checks.

Table 3 reports results from our dichotomous-around-the-median tests of the effect of the slave trade on trust in sub-Saharan Africa (Nunn and Wantchekon 2011). Each Panel shows results on a different dimension of trust as measured via the Afrobarometer survey. Aside from the fact that the original scale is standardized to have a mean of zero and a standard deviation of one, column 1 replicates the original OLS effect estimates as reported by Nunn and Wantchekon (2011). Again, in column 2 the dependent variable is a standardized dichotomous variable that equals one if the original scale value is greater than *or equal to* the median. In column 3 the dependent variable is a standardized dichotomous variable that equals one if the original scale value is strictly greater than the median.

In Panel A of Table 3 the "Dichotomous lower" results are not required. This is because the median response for trust in relatives is the upper-most category (e.g., "a lot") and therefore it is impossible to generate results with a variable in which every category is coded as zero. In all other Panels the median response for interpersonal trust was in the "somewhat" category. In all Panels the qualitative result—that the slave trade decreased interpersonal trust in sub-Saharan Africa—persists in our dichotomous tests. In some cases, for example in Panel D for the effect on intra-group trust, the dichotomous tests suggest perhaps some additional uncertainty about the specific size of the effect. However, the effects remain both statistically significant and economically meaningful.

Tables 4, 5, and 6 provide additional applications of the DAM test. The tables are to be read by comparing the left-hand columns (which, in each case, give the original authors' approach) with the later dichotomous-estimate columns on the right-hand side of the tables.

In Tables 4 and 5 we examine a form of the well-being regression equation, as in Blanchflower and Oswald (2004) and Stone et al. (2020), using subjective well-being data from the United States collected by Gallup—in Table 4—and the Behavioral Risk Factor Surveillance System (BRFSS)—in Table 5. The Gallup data use "Cantril's Ladder of Life" which asks respondents to: "Imagine a ladder with steps numbered from zero at the bottom and ten at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life. On which step of the ladder would you say you personally feel you stand at this time?" The BRFSS data use a life satisfaction question which asks: "In general, how satisfied are you with your life?" Respondents choose between four possible categories: "Very satisfied," "Satisfied," "Dissatisfied," and "Very dissatisfied." By default, these categories are enumerated on a one through four scale, with one indicating "Very satisfied" and four indicating "Very dissatisfied."

In Table 6 we examine the effect of political decentralization and political corruption (Fan et al. 2009). To measure political corruption, these authors use several ordinal scales. In our illustration, we focus on their primary outcome which is based on the following survey questions included in the World Business Environment Survey (WBES): "Is it common for firms in your line of business to have to pay some irregular 'additional payments' to get things done?" Respondents had six response categories: "never," "seldom," "sometimes," "frequently," "usually," and "always." In the regression analysis, these response categories are enumerated by assuming a linear set of intervals, numbered one through six.

In Tables 4 and 5, we see evidence of a U-shaped relationship between age and subjective well-being (Stone et al. 2010), a positive relationship between marriage and subjective well-being (Diener et al. 2000; Blanchflower and Oswald 2004; Perelli-Harris et al. 2019), and a positive relationship between income and subjective well-being (Blanchflower and Oswald 2004; Stevenson and Wolfers 2013). These estimated relationships are robust to our DAM tests using both Gallup and BRFSS data.

Table 6 reports examines the effect of democratic decentralization on political corruption, originally reported by Fan et al. (2009). Specifically, we re-examine their core results reported in columns 1 and 10 in Table 5 of their paper. The dependent variable is an ordinal variable that measures political corruption, specifically the frequency of bribes. The coefficient of interest is the estimate on the $Tiers$ variable, which measures the level of democratic decentralization by describing the number of levels of government within a given country. Additional covariates aim to control for factors that may lead to biased estimates of the correlation between democratic decentralization and political corruption.[3] We find that these results are robust to our DAM approach.

[3]More detailed descriptions of these variables can be found in Fan et al. (2009).

These kinds of DAM checks seem to have several benefits. First, by using a dichotomous dependent variable—defined at the median value—the DAM tests directly address one kind of potential concern about robustness to monotonic increasing transformations of the *observed scale* and *latent variable*. Although this method relies on the observed median value, which is manipulable with monotonic increasing transformations of the ordinal scale, this value is only used when creating the dichotomous variables which remain unchanged for any monotonic increasing transformation of the ordinal scale. A brief mathematical sketch of this idea is included in the appendix.

Second, the technique is straightforward to execute. It only requires two additional regressions. One uses a dependent variable defined with the median value of the ordinal scale in the upper part and the other with the median value of the ordinal scale in the lower part of the dichotomous split. Focusing the dichotomous split on either side of a median simplifies calculations. Theoretically, any existing threshold point in an ordinal scale could be used to define the dichotomous split. Implementing this full approach with a large number of categories, however, can quickly become computationally intensive. Additionally, some dichotomous splits will suffer from very little variation. Therefore, we would advocate using a median value in the ordinal scale to guide choices about defining dichotomous dependent variables.

With that said, this approach is not without some limitations. Perhaps chief among these is that in order to generate dichotomous variables we are forced to discard potentially valuable information captured in multiple layers of an ordinal scale. How much this limitation matters will depend on the specific details of a given empirical setting and research question. Of course the reason, in the first place, to discard the extra information is because some critics will object to its use.

Another potential issue is that the DAM tests, as they are implemented here, rely on OLS estimates. This is deliberate. It is due to possible concerns about assuming a specific and potentially incorrect functional form of the error term when using more sophisticated regression techniques—such as logit or probit. Although the use of OLS is less sensitive to the distribution of the error term, the standard possible concerns associated with using OLS regression with a dichotomous dependent variable will persist. This, we argue, is less objectionable in empirical applications primarily concerned with estimating the causal effect of some variable X on an ordinal dependent variable Y—as is the case in both Haushofer and Shapiro (2016) and Nunn and Wantchekon (2011).

5 Conclusion

Feelings are central to the lives of human beings. The reason that we study income and wealth and the business cycle and optimal taxes and pollution and minimum wage laws (and so on) is because those variables are believed to affect human feelings. Feelings are an end. Nearly everything else is a means to that end.

This paper argues that, with sufficient care, data on people's reported feelings can provide useful empirical insights. The paper introduces a validity check approach, e.g., a dichotomous-around-the-median (DAM) test. The idea behind the paper's proposed DAM tests is a simple one. The test builds on the fact that the qualitative direction of an ordinal variable *is* reliably known—even though the size of intervals between multiple ranks is not.

By applying our DAM tests, which uses no information on interval sizes, we show that it is possible to re-analyze the substantive findings in a range of well-known journal articles that have relied, even if technically inappropriately, on cardinality. These re-analyzed results can be seen by comparing the estimates across each column in Tables 2 through 6.

It might be thought, considering again Table 1, that there is a possible intuitive objection to the paper's general argument. By cutting the data just below Very Satisfied, it might be argued, then Group B is better off. Yet cutting the data just above Satisfied, a critic might point out, makes Group A better off.

Such a concern is on the surface an appropriate one. Yet two points of rebuttal and clarification are relevant. First, this kind of potential objection illustrates the need for the paper's combination of the Upper and Lower approaches discussed in the earlier sections. An important feature of the current paper is that the DAM method involves a check on the robustness of the results from both 'upper' and 'lower' groupings of the data around the median. Second, although in principle any kind of dichotomous split in the data is theoretically feasible, some of those potential splits will intrinsically be under-powered to detect an effect, because of a lack of sufficient statistical information included within each response category. This motivates the current paper's proposed use of the dichotomous form of split around the median of the distribution.

In closing, it should be emphasized that the analysis in this paper is likely to be far from the last word on this complex set of issues. It is possible that the future will see the collection of ever-increasing amounts, and perhaps even new varieties of, feelings data. We would argue that a nihilistic rejection of all forms of such data (which are widely used by the world's organizations, and thus have passed the longstanding 'Chicago market test') would be inappropriate for economists and other social scientists. Nevertheless, the empirical challenges associated with ordinal information demand careful and constructive attention. Much remains to be understood about the optimal way to handle such data.

Table 1: Which Group is More Satisfied?

	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied	Most Satisfied?
A	1	0	0	1	
B	0	1	1	0	
Linear	0	1	2	3	Tie
Concave	0	1.75	2.5	3	B
Convex	0	0.5	1.25	3	A

Table 2: DAM Tests – Cash Transfers and Well-Being in Kenya (Haushofer and Shapiro 2016)

	Original ordinal scale	Dichotomous upper	Dichotomous lower
	(1)	(2)	(3)
Panel A: Happiness (World Values Survey)			
Treatment	0.165***	0.119**	0.160***
	(0.0503)	(0.0488)	(0.0534)
Village fixed effects	Yes	Yes	Yes
Observations	1,473	1,473	1,473
R-squared	0.079	0.082	0.081
Panel B: Life Satisfaction (World Values Survey)			
Treatment	0.170***	0.178***	0.185***
	(0.0480)	(0.0482)	(0.0511)
Village fixed effects	Yes	Yes	Yes
Observations	1,473	1,473	1,473
R-squared	0.135	0.180	0.133

Notes: The dependent variable in Panel A measures happiness and the dependent variable in Panel B measures life satisfaction. Both variables are measured using the survey question found on the World Values Survey. Following Haushofer and Shapiro (2016) each dependent variable are standardized so the control group has a mean of zero and a standard deviation of one. The original results are reported in Table 4 of Haushofer and Shapiro (2016). Standard errors clustered at the household level are shown in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: DAM Tests – Trust in Africa (Nunn and Wantchekon 2011)

	Original ordinal scale (1)	Dichotomous upper (2)	Dichotomous lower (3)
Panel A: Trust of relatives (Afrobarometer)			
ln(export area)	-0.139***	-0.112***	n.a.
	(0.0376)	(0.0327)	
Observations	20,062	20,062	n.a.
R-squared	0.133	0.126	n.a.
Panel B: Trust of neighbors (Afrobarometer)			
ln(export area)	-0.158***	-0.149***	-0.119***
	(0.0335)	(0.0347)	(0.0262)
Observations	20,027	20,027	20,027
R-squared	0.156	0.112	0.146
Panel C: Trust of local council (Afrobarometer)			
ln(export area)	-0.100***	-0.0909***	-0.0642**
	(0.0196)	(0.0194)	(0.0273)
Observations	19,733	19,733	19,733
R-squared	0.196	0.166	0.179
Panel D: Intra-group trust (Afrobarometer)			
ln(export area)	-0.143***	-0.141***	-0.0845***
	(0.0313)	(0.0334)	(0.0246)
Observations	19,952	19,952	19,952
R-squared	0.144	0.106	0.136
Panel E: Inter-group trust (Afrobarometer)			
ln(export area)	-0.0971***	-0.0879***	-0.100***
	(0.0279)	(0.0238)	(0.0273)
Observations	19,765	19,765	19,765
R-squared	0.112	0.057	0.089

Notes: The dependent variable in each Panel measures a different type of trust as measured by the Afrobarometer survey. The dependent variable is standardized to have a mean of zero and standard deviation of one in each column. The original OLS results are reported in Table 3 of Nunn and Wantchekon (2011). Standard errors clustered at the ethnicity and district levels are shown in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: DAM Tests — Cantril's Ladder of Life (Gallup)

	(1)	(2)	(3)	(4)	(5)	(6)
	Original ordinal scale		Dichotomous Upper		Dichotomous Lower	
Age	-0.052***	-0.051***	-0.044***	-0.043***	-0.044***	-0.043***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2	0.0005***	0.0005***	0.0005***	0.0004***	0.0005***	0.0005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.154***	0.145***	0.132***	0.123***	0.170***	0.162***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Log of household income	0.262***	0.263***	0.235***	0.236***	0.169***	0.171***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
MSA-level log of income		-0.0282		0.00735		-0.0146
		(0.029)		(0.029)		(0.030)
U-test results:						
Turning point (age)	48.19	48.07	48.64	48.53	44.35	44.17
Fieller 95% C.I.	[47.8; 48.7]	[47.6; 48.6]	[48.1; 49.2]	[48.0; 49.1]	[44.0; 44.7]	[43.8; 44.6]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000	0.000
Slope at Min	-0.032	-0.032	-0.028	-0.027	-0.026	-0.025
Slope at Max	0.013	0.013	0.010	0.010	0.015	0.015
Year and month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Other individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	No	Yes	No	Yes	No	Yes
Other MSA-level controls	No	Yes	No	Yes	No	Yes
Observations	556,300	461,054	556,300	461,054	556,300	461,054
R^2	0.117	0.120	0.104	0.106	0.062	0.065

Notes: The dependent variable in each column is the subjective well-being measure from Gallup. The first two columns show the original ordinal scale. The middle two columns show results using a dichotomous dependent variable that equals one for all values greater than or equal to the median. The last two columns show results using a dichotomous dependent variable that equals zero for all values less than or equal to the median. In each column the dependent variable is standardized to have a mean of zero and standard deviation of one. Individual-level controls include sex, race, and six education attainment categories. Additional MSA-level controls include the unemployment rate, the crime rate, and the population size. Standard errors clustered at the MSA level are shown in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: DAM Tests — Life Satisfaction (BRFSS)

	(1)	(2)	(3)	(4)	(5)	(6)
	Original ordinal scale		Dichotomous Upper		Dichotomous Lower	
Age	-0.045***	-0.044***	-0.032***	-0.031***	-0.038***	-0.038***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2	0.0005***	0.0005***	0.0003***	0.0003***	0.0004***	0.0004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.203***	0.190***	0.087***	0.076***	0.218***	0.207***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Log of household income	0.287***	0.295***	0.235***	0.241***	0.223***	0.230***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
MSA-level log of income		-0.011		0.001		-0.013
		(0.020)		(0.020)		(0.020)
U-test results:						
Turning point (age)	46.52	46.52	49.70	49.88	44.93	44.89
Fieller 95% C.I.	[46.2; 46.8]	[46.2; 46.9]	[49.1; 50.4]	[49.18; 50.67]	[44.6; 45.3]	[44.5; 45.3]
Sasabuchi p-value	0.000	0.000	0.000	0.000	0.000	0.000
Slope at Min	-0.027	-0.027	-0.020	-0.020	-0.023	-0.022
Slope at Max	0.013	0.013	0.007	0.006	0.013	0.013
Year and month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Other individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	No	Yes	No	Yes	No	Yes
Other MSA-level controls	No	Yes	No	Yes	No	Yes
Observations	846,269	738,770	846,269	738,770	846,269	738,770
R^2	0.126	0.130	0.065	0.067	0.093	0.097

Notes: The dependent variable in each column is the life satisfaction measure from BFRSS. The first two columns show the original ordinal scale. The middle two columns give results using a dichotomous dependent variable that equals one for all values greater than or equal to the median. The last two columns presents results using a dichotomous dependent variable that equals zero for all values less than or equal to the median. In each column the dependent variable is standardized to have a mean of zero and standard deviation of one. Individual-level controls include sex, race, and six education attainment categories. Additional MSA-level controls include the unemployment rate, the crime rate, and the population size. Standard errors clustered at the MSA level are shown in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: DAM Tests — Political Corruption (Fan et al. 2009)

	(1)	(2)	(3)	(4)	(5)	(6)
	Original ordinal scale		Dichotomous Upper		Dichotomous Lower	
Tiers	0.227***	0.197**	0.229***	0.192**	0.226**	0.197**
	(0.0827)	(0.0740)	(0.0674)	(0.0786)	(0.0900)	(0.0757)
Bottom unit size		-0.0260		-0.0164		-0.0342
		(0.0396)		(0.0431)		(0.0358)
Subnational revenues		-0.0294***		-0.0285***		-0.0166**
		(0.00841)		(0.00918)		(0.00762)
Total government revenues		-0.0165**		-0.0117		-0.0181**
		(0.00745)		(0.00737)		(0.00737)
Subnational government employment share		0.00602*		0.00551		0.00449
		(0.00330)		(0.00360)		(0.00289)
Total government employees		-0.0201		-0.0182		-0.0167
		(0.0218)		(0.0232)		(0.0206)
State ownership	-0.430***	-0.439***	-0.440***	-0.451***	-0.321***	-0.360***
	(0.0524)	(0.0494)	(0.0532)	(0.0482)	(0.0467)	(0.0497)
Foreign ownership	-0.0642	-0.112***	-0.0575	-0.0676**	-0.0606	-0.110***
	(0.0404)	(0.0279)	(0.0376)	(0.0326)	(0.0382)	(0.0337)
Exporter	0.0439	0.0516**	0.0555*	0.0628*	0.0485	0.0486
	(0.0326)	(0.0251)	(0.0319)	(0.0364)	(0.0340)	(0.0306)
Firm size	-0.00639	-0.00785	-0.00919	-0.00936	-0.00219	-0.00214
	(0.00665)	(0.00501)	(0.00592)	(0.00592)	(0.00697)	(0.00435)
GDP per capita	-0.247***	-0.126	-0.214***	-0.140	-0.191***	-0.0871
	(0.0571)	(0.0845)	(0.0497)	(0.0853)	(0.0574)	(0.0794)
Democratic	0.000351	0.157	0.0414	0.188	-0.0941	0.0582
	(0.125)	(0.161)	(0.117)	(0.160)	(0.116)	(0.150)
Fuel	0.000484	0.00207	0.000130	0.00349	0.000333	-0.00122
	(0.00196)	(0.00440)	(0.00142)	(0.00423)	(0.00191)	(0.00420)
Imports	0.000465	0.00234	0.00100	0.00281	-1.24e-05	0.00186
	(0.00268)	(0.00317)	(0.00230)	(0.00291)	(0.00287)	(0.00349)
Protestant	-0.591**	0.363	-0.630**	0.197	-0.378	0.316
	(0.267)	(0.367)	(0.262)	(0.385)	(0.268)	(0.355)
British colony	-0.0178	-0.225	-0.0130	-0.191	-0.00998	-0.251
	(0.0991)	(0.157)	(0.0943)	(0.166)	(0.0905)	(0.156)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,676	4,101	6,676	4,101	6,676	4,101
R^2	0.146	0.196	0.125	0.153	0.113	0.165

Notes: The dependent variable in each column is the frequency of political corruption measure from Fan et al. (2009). The original results are reported in columns (1) and (10) of Fan et al. (2009). The first two columns of this table show the original ordinal scale. The middle two columns show results using a dichotomous dependent variable that equals one for all values greater than or equal to the median. The last two columns show results using a dichotomous dependent variable that equals zero for all values less than or equal to the median. In each column the dependent variable is standardized to have a mean of zero and standard deviation of one. Standard errors clustered at the country level are shown in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

References

Acemoglu, D., Johnson, S., and Robinson, J. (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91 (5), pp. 1369-1401.

Aghion, P. Akcigit, U., Deaton, A., and Roulet, A. (2016) "Creative Destruction and Subjective Well-Being," *American Economic Review*, 106, 12, pp. 3869-3897.

Alesina, A., DiTella, R., and MacCulloch, R. (2004) "Inequality and Happiness: Are Europeans and Americans Different?," *Journal of Public Economics*, 88, 9-10, pp.2009-2042.

Baird, S., de Hoop, J., and Oxler, B. (2013) "Income Shocks and Adolescent Mental Health," *Journal of Human Resources*, 48 (2), pp. 370-403.

Benjamin, D.J., Heffetz, O., Kimball, M.S., and Rees-Jones, A. (2012) "What Do You Think Would Make You Happier? What Do You Think You Would Choose?", *American Economic Review*, 102 (5), pp. 2083-2110.

Blanchflower, D. and Oswald, A.J. (2004) "Well-Being over Time in Britain and the USA," *Journal of Public Economics*, 88, pp. 1359-1386.

Blanchflower, D. and Oswald, A.J. (2017) "Do Humans Suffer a Psychological Low in Midlife? Two Approaches (With and Without Controls) in Seven Data Sets," *NBER paper, No. 23724*.

Bloem, J., Boughton, D. Htoo, K., Hein, A., and Payongayong, E. (2018) "Measuring Hope: A Quantitative Approach with Validation in Rural Myanmar," *Journal of Development Studies*, 54 (11), pp. 2078-2094.

Bloem, J. (2020) "How Much Does the Cardinal Treatment of Ordinal Variables Matter? An Empirical Investigation," *Working Paper*.

Boertien, D. and Vignoli, D. (2019) "Legalizing Same-Sex Marriage Matters for the Subjective Well-being of Individuals in Same-Sex Unions," *Demography* vol. 56, pp. 2109-2121.

Bond, T. and Lang, K. (2013) "The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results," *Review of Economics and Statistics*, 95 (5), pp. 1468-1479.

Bond, T. and Lang, K. (2019) "The Sad Truth About Happiness Scales," *Journal of Political Economy*. (Available online as NBER Working Paper No. 19950).

Booth, A.L., and van Ours, J. (2008) "Job Satisfaction and Family Happiness: The Part-time Work Puzzle" *Economic Journal*, 118, pp. F77-F99.

Borghans, L., Duckworth, A.L., Heckman, J., and ter Weel, B. (2008) "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43 (4), pp. 972-1059.

Bryson, A., and MacKerron, G. (2017) “Are You Happy While You Work?” *Economic Journal*, 127, pp. 106-125.

Budria, S., and Ferrer-I-Carbonell, A. (2019) ”Life Satisfaction, Income Comparisons and Individual Traits”, *Review of Income and Wealth*, 65 (2), pp. 337-357.

Centers for Disease Control (2019) https://www.cdc.gov/brfss/

Chen, L., Oparina, E., Powdthavee, N., and Srisuma, S. (2019) “Have Econometric Analyses of Happiness Data Been Futile? A Simple Truth About Happiness Scales,” *IZA Working Paper, Bonn.*

Clark, A.E. (2018) “Four Decades of the Economics of Happiness: Where Next?,” *Review of Income and Wealth*, Series 64, 2, pp. 245-269.

Clark, A.E. and Oswald, A.J. (1994) “Unhappiness and Unemployment,” *Economic Journal* 104 (424), pp. 648-659.

Clark, A.E. and Oswald, A.J. (1996) “Satisfaction and Comparison Income,” *Journal of Public Economics* 61 (3), pp. 359-381.

Clark, A.E, Fleche, S., and Senik C. (2016) ”Economic Growth Evens Out Happiness: Evidence from Six Surveys”, *Review of Income and Wealth* 62, pp. 405-419.

Clark, A.E. and Senik, C. (2010) ”Who Compares to Whom? The Anatomy of Income Comparisons in Europe”, *Economic Journal* 120, pp. 573-594.

Cornaglia, F., Feldman, N.E., and Leigh, A. (2014) “Crime and Mental Well-Being,” *Journal of Human Resources*, 49 (1), pp. 110-140.

Dedehouanou, S. F. A., Swinnen, J., and Maertens, M. (2013) ”Does Contracting Make Farmers Happy? Evidence from Senegal”, *Review of Income and Wealth*, 59 (1), pp. S138-160.

Di Tella, R. and MacCulloch, R.J. (2008) “Gross National Happiness as an Answer to Easterlin’s Paradox?,” *Journal of Development Economics*, 86, pp. 22-42.

Di Tella, R. and MacCulloch, R.J., and Oswald, A.J. (2001) “Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness,” *American Economic Review*, 91, pp. 335-241.

Diener, E., Gohm, C.L., Suh, E., and Oishi, S. (2000) “Similarity of the Relations between Marital Status and Subjective Well-Being Across Cultures,” *Journal of Cross-Cultural Psychology*, 31, issue 4, pp. 419-436.

Diener, E., Heintzelman, S.J., Kushlev, K., Tay, L., Wirtz D., Lutes, L.D., and Oishi, S. (2017) “Findings All Psychologists Should Know From the New Science on Subjective Well-

Being," *Canadian Psychology/Psychologie canadienne*, 58, pp. 87-104.

Durand, M. (2105) "The OECD Better Life Initiative: How's Life? and the Measurement of Wellbeing". *Review of Income and Wealth*, 61, pp. 4-17.

Easterlin, R. (1974) "Does Economic Growth Improve the Human Lot? Some Empirical Evidence" David, P., and Reder, M. (Eds) *Nations and Households in Economic Growth*, pp. 89-125, New York: Academic Press.

Easterlin, R. (2001) "Income and Happiness: Towards a Unified Theory," *Economic Journal*, 111, 473, pp. 465-484.

Fafchamps, M. and Shilpi, F. (2008) "Subjective Welfare, Isolation, and Relative Consumption," *Journal of Development Economics*, 86, 1, pp. 43-60.

Fan, C.S., Lin, C., and Treisman, D. (2009) "Political Decentralization and Corruption: Evidence from Around the World," *Journal of Public Economics*, 93, issue 1-2, pp. 14-34.

Ferrer-i-Carbonell, A. and Frijters, P. (2004) "How Important is Methodology for the Estimates of the Determinants of Happiness?" *Economic Journal*, 114, pp. 641-659.

Fisher, R.A. (1958) "Smoking and Cancer," *Nature* 182, (1958 August 30), p. 596.

Frey, B.S., and Stutzer, A. (2000) "Happiness, Economy and Institutions" *Economic Journal*, 110, pp. 918-938.

Frijters, P., Haisken-DeNew, J.P., and Shields, M.A. (2004) "Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification" *American Economic Review*, 94 (3) pp. 730-740.

Gallup (2019) https://news.gallup.com/poll/101905/gallup-poll.aspx

Glewwe, P. (1997) "Estimating the Impact of Peer Group Effects on Socioeconomic Outcomes: Does the Distribution of Peer Group Characteristics Matter?" *Economics of Education Review*, 16 (1), pp. 39-43.

Glewwe, P. , Ross, P.H., and Wydick, B. (2018) "Developing Hope among Impoverished Children: Using Child Self-Portraits to Measure Poverty Program Impacts" *Journal of Human Resources*, 53 (2), pp. 330-355.

Graham, C. (2005) "Insights on Development from the Economics of Happiness" *World Bank Research Observer*, 20 (2), pp. 201-231.

Graham, C. (2016) "Subjective Well-Being in Economics," in *The Oxford Handbook of Well-Being and Public Policy*, Adler, M. and Fleurbaey, M. (eds.), New York: Oxford University Press, pp. 424-450.

Graham, C., Laffan, K. and Pinto, S. (2018) "Well-being in Metrics and Policy". *Science* 362, pp. 287-288.

Haushofer, J. and Shapiro, J. (2016) "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *Quarterly Journal of Economics*, 131, 4, pp. 1973-2042.

Heffetz, O. (2011) "A Test of Conspicuous Consumption: Visibility and Income Elasticities" *Review of Economics and Statistics*, 93, pp. 1101-1117. **Jacob, B. and Rothstein, J. (2016)** "The Measurement of Student Ability in Modern Assessment Systems" *Journal of Economic Perspectives*, 30 (3), pp. 85-108.

Kahneman, D. (2011) *Thinking Fast and Slow*, Penguin, London.

Kahneman, D. and Deaton, A. (2010) "High Income Improves Evaluation of Life but Not Emotional Well-being," *Proceedings of the National Academy of Sciences of the USA*, 107, 38, pp. 16489-16493.

Kaiser, C. and Vendrik, C.M. (2019) "How Threatening are Transformations of Reported Happiness to Subjective Wellbeing Research?," *Unpublished manuscript*, Maastricht University, Netherlands.

Krueger, A.B., Kahneman, D., Schkade, D., Schwarz, N., and Stone, A., (2009) "National Time Accounting: The Currency of Life" In *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*, (ed.) Krueger, A.B., University of Chicago Press.

Krueger, A.B. (2017) "Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate" *Brookings Papers on Economic Activity*, Fall 2017, pp. 1-87.

Lang, K. (2010) "Measurement Matters: Perspectives on Education Policy from an Economist and School Board Member" *Journal of Economic Perspectives*, 24 (3), pp. 167-181.

Luechinger, S. (2009) "Valuing Air Quality Using the Life Satisfaction Approach" *Economic Journal*, 536, pp. 482-515.

Luechinger, S., Meier, S., and Stutzer, A. (2010) "Why Does Unemployment Hurt the Employed? Evidence from the Life Satisfaction Gap Between the Public and the Private Sector" *Journal of Human Resources*, 45 (4), pp. 998-1045.

Ludwig, J., Duncan, G.J., Gennentian, L.A., Katz, L.A., Kessler, R.C., Kling, J.R., and Sanbonmatsu, L. (2012) "Neighborhood Effects on the Long-Term Well-Being of Low-

Income Adults," *Science*, 337, pp. 1505-1510.

Luttmer, E.F.P. (2005) "Neighbors as Negatives: Relative Income and Well-Being," *Quarterly Journal of Economics*, 120, no. 3, pp. 963-1002.

Manski, C. (2003) *Partial Identification of Probability Distributions,* Springer Series in Statistics, New York: Springer.

Martimort, D. and Straub, S. (2009) "Infrastructure Privatization and Changes in Corruption Patterns: The Roots of Public Discontent," *Journal of Development Economics*, 90, pp. 69-84.

Nunn, N. and Wantchekon, L. (2011) "The Slave Trade and the Origins of Mistrust in Africa" *American Economic Review* 101 (7), pp. 3221-3252.

OECD (2019) "Putting Well-being Metrics into Policy Action: Conference," *OECD Headquarters*, Paris, France.

Oswald, A.J. (1997) "Happiness and Economic Performance," *Economic Journal*, 107, pp. 1815-1831.

Oswald, A.J. (2008) "On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings," *Economics Letters* 100, pp. 369-372.

Oswald, A.J., and Powdthavee, N. (2008) "Does Happiness Adapt? A Longitudinal Study of Disability with Implications for Economists and Judges," *Journal of Public Economics*, 92, 5-6, pp. 1061-1077.

Pearl, J. and Mackanzie D. (2018) *The Book of Why: The New Science of Caus and Effect*, Basic Books: New York, NY.

Perelli-Harris, B., Hoherz, S., Lappegard, T., and Evans, A. (2019) "Mind the 'Happiness' Gap: The Relationship Between Cohabitation, Marriage, and Subjective Well-Being in the United Kingdom, Australia, Germany, and Norway," *Demography*, vol. 56, pp. 1219-1246.

Putnam, R. (2001) *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, 1st edition, New York: NY.

Ravallion, M., Himelein, K., and Beegle, K. (2016) "Can Subjective Questions on Economic Welfare Be Trusted?" *Economic Development and Cultural Change*, 64 (4), pp. 697-726.

Ritter, J. and Anker, R. (2002) "Good Jobs, Bad Jobs: Workers' Evaluations in Five Countries" *International Labor Review*, 141 (4), pp. 331-358.

Schröder, C., and Yitzhaki, S. (2016) "PISA Country Rankings and the Difficulty of Exams", Working Paper, Available at SSRN.

Schröder, C., and Yitzhaki, S. (2017) "Revisiting the Evidence for Cardinal Treatment of

Ordinal Variables" *European Economic Review*, 92, pp. 337-358.

Stevenson, B., and Wolfers, J. (2009) "The Paradox of Declining Female Happiness," *American Economic Journal: Economic Policy*, 1, 2, pp. 190-225.

Stevenson, B., and Wolfers, J. (2013) "Subjective Well-Being and Income: Is There Any Evidence of Satiation?," *American Economic Review: Papers & Proceedings*, 103, no. 3. pp. 598-604.

Stiglitz, J., Sen, A., and Fitoussi, J-P. (2009) Commission on the Measurement of Economic Performance and Social Progress. Paris: Final Report. Downloadable at www.stiglitz-sen-fitoussi.fr.

Stone, A.A., Schwartz, J.E., Broderick, J.E., and Deaton, A. (2010) "A Snapshot of the Age Distribution of Psychological Well-being in the United States," *Proceedings of the National Academy of Sciences of the USA*, 107, 22, pp. 9985-9990.

Tammer, E. (2010) "Partial Identification in Econometrics," *The Annual Review of Economics*, 2, pp. 167-195.

Treisman, D. (2000) "The Causes of Corruption: A Cross-Sectional Study," *Journal of Public Economics*, 76, 3, pp. 399-457

Verme, P. (2011) "Life Satisfaction and Income Inequality", *Review of Income and Wealth*, 57 (1), pp. 111-137.

Supplemental Appendix

5.1 Mathematical sketch of DAM tests

Let Y^* be the latent happiness variable, for example. Since, Y^* cannot be directly observed, let Y be the observed ordinal variable that measures Y^*. The observed ordinal variable Y is measured with various values of μ corresponding to threshold points on the ordinal scale:

$$Y = \begin{cases} 0 & \text{if} \quad Y^* \leq 0, \\ 1 & \text{if} \quad 0 < Y^* \leq \mu_1, \\ 2 & \text{if} \quad \mu_1 < Y^* \leq \mu_2 \\ \vdots & \\ N & \text{if} \quad \mu_{N-1} < Y^* \end{cases} \tag{1}$$

Next, we calculate the median of the observed ordinal scale and define both the upper and lower dichotomous variables as follows: Dichotomous Upper = 1 if $Y \geq Med(Y)$ and Dichotomous Lower = 1 if $Y > Med(Y)$, where $Med(Y)$ is the median of the observed scale Y.

Now, let $\tau(\cdot)$ be a monotonic increasing function. Transforming the latent variable Y^* with this monotonic increasing function, $\tau(Y^*)$, also transforms the observed ordinal, $\tau(Y)$, given the same set μ's as follows:

$$\tau(Y) = \begin{cases} 0 & \text{if} \quad \tau(Y^*) \leq 0, \\ 1 & \text{if} \quad 0 < \tau(Y^*) \leq \mu_1, \\ 2 & \text{if} \quad \mu_1 < \tau(Y^*) \leq \mu_2 \\ \vdots & \\ N & \text{if} \quad \mu_{N-1} < \tau(Y^*) \end{cases} \tag{2}$$

Note that $Med(\tau(Y)) = \tau(Med(Y))$. Therefore, the "new" dichotomous variables are defined as follows: Dichotomous Upper = 1 if $\tau(Y) \geq \tau(Med(Y))$ and Dichotomous Lower = 1 if $\tau(Y) > \tau(Med(Y))$, which is equivalent to Dichotomous Upper = 1 if $Y \geq Med(Y)$ and Dichotomous Lower = 1 if $Y > Med(Y)$, where $Med(Y)$ is the median of the observed scale Y. ■