

How Much Does the Cardinal Treatment of Ordinal Variables Matter? An Empirical Investigation

Jeffrey R. Bloem¹

¹United States Department of Agriculture, Economic Research Service, Email: jeffrey.bloem@usda.gov*

Abstract

Many researchers use an ordinal scale to quantitatively measure and analyze concepts. Theoretically valid empirical estimates are robust in sign to any monotonic increasing transformation of the ordinal scale. This presents challenges for the point-identification of important parameters of interest. I develop a partial identification method for testing the robustness of empirical estimates to a range of plausible monotonic increasing transformations of the ordinal scale. This method allows for the calculation of plausible bounds around effect estimates. I illustrate this method by re-visiting analysis by Nunn and Wantchekon (2011) on the slave trade and trust in sub-Saharan Africa. Supplemental illustrations examine results from (i) Aghion et al. (2016) on creative destruction and subjective well-being and (ii) Bond and Lang (2013) on the fragility of the black-white test score gap.

Keywords: Ordinal variables, Partial identification, Set estimation, Linear regression, Happiness, Subjective well-being, Trust, and Test scores.

1 Introduction

Concerns about the cardinal treatment of ordinal dependent variables are well-known. Consider an ordinal scale measuring "satisfaction," with the following four categories: "Very Dissatisfied," "Dissatisfied," "Satisfied," and "Very Satisfied." Suppose there are two groups of people, A and B, each consisting of two individuals. Group A has one person who is "Very Dissatisfied" and another who is "Very Satisfied." Group B has one person who is "Dissatisfied" and another who is "Satisfied." Which group is more satisfied?

The answer, in part, depends on the numerical values assigned to the response categories. It may seem reasonable to assign the integers zero, one, two, and three to each of the four categories. In this case, the groups are equally satisfied, with an average score of 1.5. Similar to utility functions, however, ordinal variables provide information about the rank of a specific concept, rather than representing a known or fixed interval. As such, any set of numerical values that preserve the ordering of the scale is also potentially valid. Therefore another reasonable set of numerical values could be: zero, 1.75, 2.5, and three. In this case, group B is more satisfied. Yet, another reasonable set of values could be: zero, 0.5, 1.25, and three. In this case group A is more satisfied.

Although ordinal response models (i.e., ordered logit or probit) are designed precisely for the situation where a researcher is interested in using an ordinal dependent variable,¹ there are numerous examples in applied empirical research where results using an ordinal response model are less

*I thank the editor and three anonymous reviewers for constructive feedback on this paper. The findings and conclusions of this paper are mine and should not be construed to represent any official USDA or US Government determination or policy. This research was conducted prior to my employment with the USDA. All remaining errors are my own.

1. Statistical models designed to appropriately handle an ordered dependent variable originated in the biometrics literature (Aitchison and Silvey 1957; Snell 1964). Use of such statistical models in the social sciences followed with McKelvey and Zavoina (1975).

Political Analysis (2020)

DOI: 10.1017/pan.xxxx.xx

Corresponding author
Jeffrey R. Bloem

Edited by
John Doe

© The Author(s) 2020. Published by Cambridge University Press on behalf of the Society for Political Methodology.

preferred compared to results from a more simple and straightforward linear regression (see, e.g., Nunn and Wantchekon 2011; Stevenson and Wolfers 2013; Aghion et al. 2016; Bryson and MacKerron 2017; Deaton 2018; as well as numerous other examples listed in the Online Supplement). Despite the well-known concerns, justification for using a linear regression with an ordinal dependent variable often include the incidental parameter problem (Neyman and Scott 1948; Heckman 1981; Lancaster 2000; Riedl and Geishecker 2014), debatable assumptions about the distribution of the error term (Bond and Lang 2019), or the use of a more sophisticated identification strategy—such as, for example, fixed effects (Ferrer-i-Carbonell and Frijters 2004).

Most fundamentally, concerns associated with the cardinal treatment of an ordinal dependent variable can be characterized as a missing information problem. That is, the researcher does not know the form of the function characterizing the relationship between the ordinal scale and the latent variable (Oswald 2008). Recent work demonstrates this point in related but distinct ways. Focusing on transformations to the observed values of the ordinal scale, Schröder and Yitzhaki (2017) demonstrate that valid empirical estimates must be robust to monotonic increasing transformations of the observed ordinal scale. Focusing on transformations to the unobserved latent variable, Bond and Lang (2019) point out that empirical results using ordered logit or probit regression analysis implicitly assume either a normal or logistic distribution on the error term, when alternative distributions of the error term are also theoretically permissible.

In this paper I develop a partial identification method for testing how much the cardinal treatment of ordinal variables matters for any empirical specification. It is generally not possible to perform credible statistical inference without any assumptions. Although some of these assumptions are plausible and based on known economic principles, inevitably some of these assumptions are arbitrary—and at times esoteric—but necessary to interpret empirical estimates (Tamer 2010). In the context of ordinal variables, most studies make an assumption about the functional form of the reporting function which cannot be supported by known economic principles.² These sorts of assumptions are necessary to point-identify effect estimates. The method I describe allows for estimation of a set of estimates based on a range of plausible monotonic increasing transformations of the ordinal dependent variable. I apply this method to re-examine empirical estimates of Nunn and Wantchekon (2011) on the slave trade and trust in sub-Saharan Africa.³

The method described in this paper first limits the universe of monotonic increasing transformations to be defined by a parameterized function representing a (relatively extreme) range of transformations. Next, based on this range of transformations, the researcher estimates a set of effect estimates. Finally, if effect estimates are not robust to this initial range of transformations and as demonstrated by Kaiser and Vendrik (2019), the researcher graphically assesses the plausibility of the transformation associated with specific effect estimates. This method, therefore, tests the sensitivity of the parameter of interest to a range of plausible monotonic increasing transformations.

The contribution of this paper is threefold. First, the method developed in this paper generalizes the work of Schröder and Yitzhaki (2017) to cases using econometric specifications with multiple covariates. In practice, the sufficient conditions developed by Schröder and Yitzhaki (2017) apply only to cases either comparing means between two groups or performing simple bivariate regression analysis.⁴ The empirical application shows that the inclusion of additional covariates in a given econometric specification influences robustness of results to monotonic increasing transforma-

2. Some work aims to identify the functional form of the reporting function in specific contexts using framed experiments (Oswald 2008; Van Praag 1991; Banks and Coleman 1981).

3. Supplemental analyses, shown in the Online Supplement, include analyses of Aghion et al. (2016) on creative destruction and subjective well-being and Bond and Lang (2013) on the black-white test score gap in kindergarten through third grade. Since Bond and Lang (2013) already establish the fragility of the black-white test score gap in kindergarten through grade three, the investigation of these results both illustrate and test the validity of the method developed in this paper.

4. Although note that Schröder and Yitzhaki (2017) do consider cases with multiple covariates, but frame their arguments based on theoretical results.

tions. That is, it is possible to fail the sufficient conditions developed by Schröder and Yitzhaki's (2017) and yet, when the full set of covariates are included in the empirical specification, the results can be robust to plausible monotonic increasing transformations.

Second, this method provides insight into the robustness of the sign, size, and statistical significance of effect estimates. Although the sufficient conditions of Schröder and Yitzhaki (2017) provide tests for the robustness of the sign of effect estimates, these conditions are silent about the size and statistical significance of these estimates.⁵ By calculating a set of effect estimates, this method allows researchers to fully assess the sensitivity of the sign, size, and statistical significance of effect estimates. This is important because the sign is not the only important piece of information derived from effect estimates. Both the size and precision (i.e., statistical significance) of effect estimates are necessary inputs into cost-benefit analysis, policy evaluation, and the estimation of important economic parameters.

Finally, the recent work on the valid statistical treatment of ordinal variables, by Schröder and Yitzhaki (2017) and Bond and Lang (2019), focus on ordinal variables measuring subjective well-being or happiness. These insights also apply to any variable that measures a latent concept using an ordinal scale. Therefore, concepts such as "satisfaction" (Frijters, et al. 2004; Clark and Oswald 1996), "trust" (Nunn and Wantchekon 2011; Putnam 2001), various 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"—of political institutions (Acemoglu et al. 2001), for example—and even standardized test scores (Bond and Lang 2013; Glewwe 1997; Jacob and Rothstein 2016; Lang 2010; Schröder and Yitzhaki 2016) are all measured with an ordinal scale. This paper extends existing theoretical insights to any empirical application using an ordinal dependent variable.

The next section briefly describes the theoretical framework motivating this research. In that section I outline the potential theoretical consequences of the cardinal treatment of ordinal variables and summarize the sufficient conditions developed by Schröder and Yitzhaki (2017). Section 3 introduces the methodology developed in this paper. Section 4 empirically illustrates this method. Finally, section 5 concludes.

2 Theoretical Framework

Although ordinal variables are used to measure a variety of concepts with no natural quantitative unit of measure, I proceed here by briefly discussing the subjective well-being literature specifically. As discussed below, the following also applies to other ordinal variables, such as happiness, satisfaction, trust, measures of quality, standardized test scores, and other concepts that require measurement via the use of an ordinal scale. Nevertheless, it is helpful to draw a connection to the familiar concept of utility theory and the relevant implications for econometric analysis (Greene 2012; Becker and Kennedy 1992). Suppose an individual's well-being is characterized by the following underlying relationship:

$$Y^* = X' \gamma \tag{1}$$

In this characterization, Y^* is the unobserved latent well-being of the individual. The vector X represents observable variables that define an individual's well-being and γ is a vector of regression coefficients. Since, Y^* cannot be directly observed, subjective well-being, Y , is measured via an

5. This contribution also applies to work specifically focusing on the sensitivity of test scores to transformations of the scale (see, e.g., Nielsen 2017).

ordinal variable with the various values of Y^* corresponding to threshold points on the ordinal scale:

$$Y = \begin{cases} \bullet & \text{if } Y^* \leq \bullet, \\ , & \text{if } \bullet < Y^* \leq ,, \\ f & \text{if } , < Y^* \leq f \\ \vdots & \\ N & \text{if } N-, < Y^* \end{cases} \quad (2)$$

A fundamental problem, when estimating equation (1) with OLS, is the values of Y^* are unknown.⁶ Using the observed ordinal scale of Y as the dependent variable implicitly assumes that the values of Y have known and fixed intervals. Thus, OLS assumes the ordinal variable measuring well-being is cardinal.

2.1 The Reporting Function

Oswald (2008) critiques the subjective well-being and happiness literature by arguing that this research has yet to establish the shape of the function relating reported subjective well-being to actual well-being. Oswald (2008) states:

As an example, imagine that there is constant marginal utility of income, but that people as they feel cheerier, mark themselves happier on a questionnaire scale in a way in which they are intrinsically reluctant to approach the upper possible level on the questionnaire form. Then the reporting function itself is curved. In this case, we will have the illusion [...] that true diminishing marginal utility of income has been established empirically.

A similar argument is easily applied to other ordinal variables, such as happiness, satisfaction, trust, and measures of quality. Standardized test scores perhaps require a brief explanation. As discussed in Bond and Lang (2013), standardized test scores may not have a known or fixed interval between values.⁷ Consider a simple case where a test score simply assigns values based on the number of questions answered correctly by each student. If some questions are more difficult than others, then assuming a fixed interval or cardinal scale may not be valid. Since test scores approximate student “learning,” answering difficult questions correctly may signal a larger marginal gain in learning than answering the easier questions correctly (see, e.g., Reardon 2008; Nielsen 2017). The reporting function for test scores, therefore, defines the relationship between actual student learning to performance on a test.

2.2 Sufficient Conditions for Robustness of Sign

Testing for the robustness of monotonic increasing transformations is complicated by the fact that there are an infinite number of ways to transform an ordinal variable. Precisely due to this reality, Schröder and Yitzhaki (2017) derive two theoretical conditions under which effect estimates will be robust in sign to monotonic increasing transformations. The first condition refers to the robustness of mean comparisons between groups and the second condition refers to the robustness of OLS regression estimates.⁸

6. Note that even if the values of Y^* are known, but the distributions of the latent variable within the discrete categories are unknown, then any empirical result will ultimately rely on strong assumptions.

7. A considerable amount of work by psychometricians aim at confronting the issue that test scores are measured on an ordinal, rather than an interval or cardinal, scale (see Stevens 1946; Thorndike 1966; Schröder and Yitzhaki 2016). Methods, such as item response theory (IRT), lead some to believe that test scores can be considered to be measured on a cardinal or interval scale (see Baker 2001). However, this is still an open area of research and the vast majority of researchers consider test scores to be measured on an ordinal scale.

8. I only summarize the details of these sufficient conditions to motivate the remainder of the paper, the interested reader should reference the work of Schröder and Yitzhaki (2017).

These conditions draw from the literature on stochastic dominance (Hadar and Russel 1969). In particular, the first condition states that the mean of one group is larger than that of another mutually exclusive group if and only if the former first-order stochastically dominates the latter. This condition implies that if the cumulative distribution functions of each group intersect, then it is possible to find a monotonic transformation of the ordinal scale that will change which group has a larger mean.

The second condition introduces the concept of the line of independence minus absolute concentration (LMA) curve. As the name suggests, the LMA curve takes the difference between two curves: the line of independence and the absolute concentration curve. The line of independence (LoI) is defined as the weighted mean of the dependent variable, Y , multiplied by the cumulative distribution, $F(X)$, of the explanatory variable, X .

$$LoI_i = \left[\frac{1}{N} \sum_{j=1}^N w_j Y_j \right] \times F(X_i) \quad (3)$$

The absolute concentration curve (ACC) is the cumulated product of the dependent variable, Y , and the frequency weight, w , divided by the sum of the frequency weights.

$$ACC_i = \frac{Y_i w_i}{\sum_{j=1}^N w_j}, ACC_{Cf} = \frac{Y_i w_i \cdot \dot{Y} Y_f w_f}{\sum_{j=1}^N w_j}, \dots, ACC_{CN} = \frac{Y_i w_i \cdot \dot{Y} \dots \dot{Y} Y_N w_N}{\sum_{j=1}^N w_j} \quad (4)$$

In both equations (3) and (4), $Y_1 \leq Y_f \leq \dots \leq Y_N$. Equation (4) can be interpreted as the generalized Lorenz curve. Finally, the LMA curve is defined as follows:⁹

$$LMA_i = LoI_i - ACC_i \quad [N] \quad (5)$$

The LMA curve is related to the concept of second-order stochastic dominance and the absolute Lorenz curve. Recall that second-order stochastic dominance states that if two Lorenz curves cross, then it is impossible to determine which of two mutually exclusive groups second-order stochastically dominate the other. Therefore the second condition, derived by Schröder and Yitzhaki (2017), states that if the LMA curve intersects the horizontal axis, then the absolute Lorenz curve intersect the line of independence, and there is some monotonic increasing transformation that will change the sign of the OLS regression coefficient.¹⁰ If the absolute Lorenz curves do not cross then there does not exist a monotonic increasing transformation that can change the sign.¹¹

Stated more formally, the logic of the second condition is as follows. Consider two simple bivariate OLS regression coefficients, β_1 and β_2 , from two separate specifications. One uses the raw ordinal variable, Y , and the other uses the transformed ordinal variable, $T(Y)$. If the LMA curve of Y , with respect to X , intersects the horizontal axis, it is possible to find a monotonic increasing transformation of the dependent variable, $T(Y)$, that can change the sign of the OLS regression coefficient. That is, if β_1 is positive (negative) then β_2 will be negative (positive). This implies:

$$\frac{\beta_1}{\beta_2} = \frac{\frac{Cov(X,Y)}{Var(X)}}{\frac{Cov(X,T(Y))}{Var(X)}} < \bullet \quad (6)$$

Although these sufficient conditions are instructive, important questions remain. First, suppose

9. A user-written STATA program is available which allows researchers to easily generate LMA curves (Schaffer 2015).

10. Formal proofs of this condition can be found in Yitzhaki and Schechtman (2012, 2013) and the Online Appendix of Schröder and Yitzhaki (2017).

11. Note the two Lorenz curves, in this explanation, refer to one representing the "raw" ordinal values and the second representing the "transformed" ordinal values. For the bulk of this paper, I will focus on the second condition derived by Schröder and Yitzhaki (2017).

there exists a transformation $T(Y)$ that allows equation (6) to hold, what is the shape of this transformation? Second, on the other hand, suppose there does not exist a transformation $T(Y)$ that allows equation (6) to hold, does the magnitude of the coefficient meaningfully change? Similarly, how is statistical significance affected by these transformations? Finally, equation (1) displayed an analytical example where there are multiple covariates and equations (3) through (6) only consider one X variable. This raises a final question. Since the existing tests of Schröder and Ytzhaki (2017) only focus on simple bivariate examples, how are researchers to test robustness of more complicated specifications to monotonic increasing transformations of the ordinal scale? These are the questions that this paper now aims to address.

3 A Partial Identification Method

The method consists of three steps. First, the researcher defines a parameterized function representing a (relatively extreme) range of transformations. In this section, I discuss two possible parameterized functions representing distinct classes of transformations: (i) globally concave and convex transformations and (ii) transformations with an inflection point. Second, the researcher estimates a set of effects associated with this range of transformations. Finally, within this set of transformations, and given the specific empirical setting, the researcher defines a (more narrow) range of *plausible* transformations. Consistent with the “law of decreasing credibility” (Manski 2003), as the strength of the assumptions used in this last step increases the credibility of the effect estimate decreases.

3.1 Globally Concave and Convex Transformations

In the spirit of Oswald (2008), the reporting function can be convex, concave, or linear in the raw ordinal rankings. The following parameterized function effectively allows the reporting function to be convex, concave, or linear depending on the value of the α parameter:

$$T(Y) = Y_{Max} \times \left(\frac{Y}{Y_{Max}} \right)^\alpha \quad [\alpha > 0] \quad (7)$$

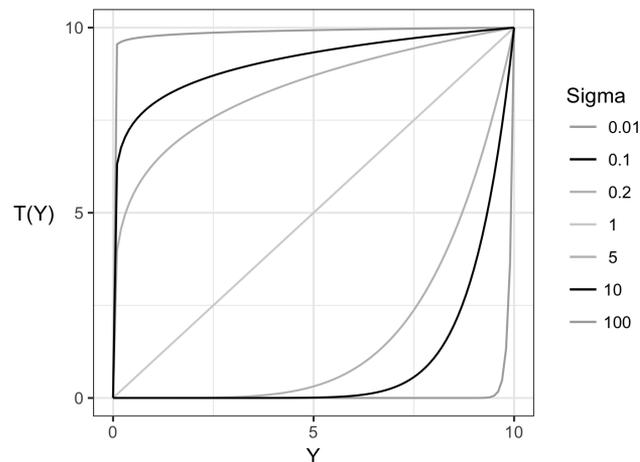
In this transformation Y is the linear ordinal scale ranging from zero through Y_{Max} —where Y_{Max} is the maximum value of the observed ordinal scale.¹² If $\alpha = 1$, then the scale remains in its linear form. If $0 < \alpha < 1$, then the scale will be concave to some degree, with the distances between relatively low levels being larger than the distances between relatively high levels. Finally, if $\alpha > 1$, then the ordinal scale will be convex to some degree, with the distances between relatively low levels being smaller than the distances between relatively high levels.

Theoretically, values of α could exist within the positive infinite interval $(0, \infty)$. If some restrictions to this domain are acceptable, however, equation (7) can help provide insight into the robustness of a particular effect estimate to plausible monotonic increasing transformations. Figure 1 shows equation (7), assuming—for illustrative purposes—a zero through ten ordinal scale, with several values of α . Plotting these functions allows researchers to make a choice about restrictions to the domain of transformations.

As argued by Kaiser and Vendrik (2019), graphically assessing plausibility may be the best and most transparent approach. Alternative ways for assessing plausibility could include benchmarking the shape of the distribution of the ordinal variable based on some other variable that measures a related concept. This is the approach used by Bond and Lang (2019) who argue that the distribution of subjective well-being can be just as skewed as the wealth distribution. Although this method of benchmarking can be potentially informative, it is still widely debated by empirical researchers.

12. This transformation ensures that the scale maintains its endpoints at zero and Y_{Max} , respectively. This point, however, is trivial as including a scaler in equation (7) can effectively shift the endpoints of the scale if necessary.

Figure 1. Specific Parameter Values of Globally Concave and Convex Transformations



Notes: This figure shows various transformation functions, given specific parameter values. The functions map the original variable, Y , into a transformed ordinal variable, $T(Y)$. In this figure the ordinal scale is assumed to run from zero through ten.

Kaiser and Vendrik (2019) demonstrate that the transformations used by Bond and Lang (2019) imply unrealistic assumptions about the way individuals reply to survey questions. Specifically, although the log-normal transformation used by Bond and Lang (2019) is theoretically permissible, it implies that small changes at low levels of true unobservable happiness are able to change responses on the observed ordinal scale by several categories and—at the same time—only large changes at high levels of true unobservable happiness are able to change responses on the observed ordinal scale. Moreover, Kaiser and Vendrik (2019) point out that it is unclear how knowing the predicted distribution of happiness is more or less skewed than the income distribution informs about the plausibility of a monotonic increasing transformation. If higher incomes are subject to decreasing marginal utility—as is assumed by many researchers and (arguably) empirically supported—then we should expect the distribution of true unobservable happiness to be less skewed than the income distribution.

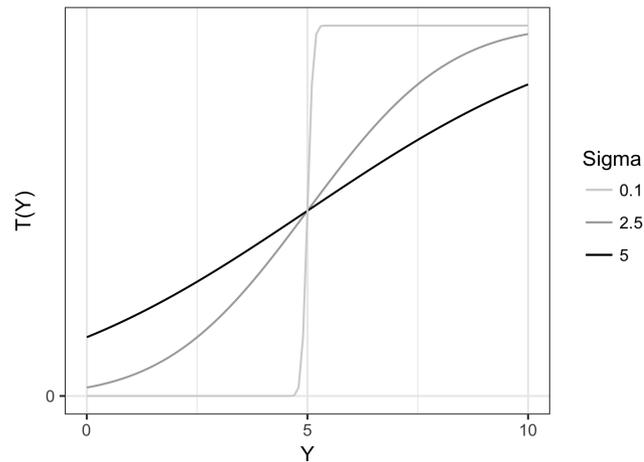
For the subsequent empirical illustrations, I assume that $\sigma \in [\cdot, \cdot, \cdot]$ define the range of transformations. In each of the empirical illustrations this range of transformations is relatively extreme. In the analysis of trust, shown in the main text, a transformation at each end of this range implies massive changes in latent trust associated with either only relatively low or relatively high reported trust levels. This range of transformations is purposefully extreme. It allows the researcher to compute a large set of effect estimates associated with this wide range of transformations. As I will discuss in section 3.3, the final step of this method is to assess a (more narrow) range of *plausible* transformations.

3.2 Transformations with an Inflection Point

An alternative class of transformations are those with an inflection point. Rather than being either concave or convex, this class of transformations are convex below and concave above an inflection point. The motivation for this class of transformations extends the intuition of Oswald (2008), where people are reluctant to report the highest values of some ordinal scale. If the “true” reporting function is one with an inflection point, this would imply that people are also reluctant to report the lowest values on the scale. Said differently, it takes a relatively large marginal gain or loss of some latent variable to move an individual off the mid-point of the observed ordinal scale.

One way to define such a class of reporting functions is to transform the observed ordinal scale

Figure 2. Specific Parameter Values of CDF Transformations



Notes: This figure shows various transformation functions, given specific parameter values. The functions map the original variable, Y , into a transformed ordinal variable, $T(Y)$. In this figure the scale is assumed to run from 0 - 10.

as follows:

$$T(Y) = Y_{Max} \times F\left(\frac{X - Y_{Mid}}{\sigma}\right) \quad (8)$$

In equation (8), $F(\cdot)$ is the cumulative distribution function (CDF) with a mean of Y_{Mid} —the middle point on the ordinal scale—and a standard deviation of σ . The domain of $F(\cdot)$ is determined by the range of the ordinal scale. In general, if the scale is zero through Y_{Max} and $Y_{Mid} = 5$, then $F(\cdot)$ will essentially be linear.¹³ If σ is relatively close to zero, then $F(\cdot)$ will look increasingly like a step function, with the step at Y_{Mid} . Therefore, limiting the domain of $F(\cdot)$ is straightforward when using these CDF transformations with an inflection point. Defining $\sigma \in [\cdot, Y_{Mid}]$ effectively characterizes all transformations from the nearly linear case to the relatively extreme step-function case. These details are visualized in Figure 2 for a zero through ten scale.

3.3 Assessing Plausibility

The final step of this method—which is only necessary if effect estimates are not robust to the initial range of transformations—is to limit the set of effect estimates to those associated with a range of *plausible* transformations. As the work of Kaiser and Vendrik (2019) demonstrates, this assessment is best and most transparent when implemented graphically. This is not to say that other ways for assessing plausibility are inappropriate. Alternative ways such as (i) finding the range of transformations that preserve empirical results that must be true based on theory or (ii) benchmarking the shape of the distribution of the ordinal variable based on some other variable that measures a related concept could be effective in the appropriate context.¹⁴ The present goal, however, is to provide a framework for assessing plausibility that could be reasonably applied in most empirical settings.

Graphically assessing plausibility first requires the researcher to examine the set of effect

13. Although this line will be roughly linear, it will not exactly replicate empirical results that assume a linear reporting function. In practice, CDF transformations do not preserve endpoints at zero and Y_{Max} , respectively. Additionally, it is not necessary to impose the inflection point at the mid-point of the scale. In practice, any category could serve as the inflection point.

14. Indeed Bond and Lang (2019) use a variation of this sort of validation exercise to benchmark plausibility. As the work of Kaiser and Vendrik (2019) shows, however, these validation exercises can be controversial. Examples of this type of validation can be found in Daly and Wilson (2009) and Di Tella et al. (2003) who compare subjective well-being data with revealed preference outcome-based measures of well-being in an effort to validate ordinal variables.

estimates associated with the relatively extreme range of transformations. If the estimated effects are qualitatively robust for the entire range of transformations, then the researcher can conclude that estimated effects are robust to extreme monotonic increasing transformations. Conversely, if the estimated effects are not robust in sign, statistical significance, and magnitude for the entire range of (relatively extreme) transformations, then the researcher must assess a range of plausible transformations given the empirical context. I demonstrate this sort of analysis in the subsequent empirical illustration.

This final step may seem esoteric. It is important to acknowledge, however, that it is generally not always possible to perform credible statistical inference without any untestable assumptions (see, e.g., Conley et al. 2012). As discussed by Tamer (2010), “the partial identification approach to econometrics views economic models as sets of assumptions, some of which are plausible and some of which are esoteric and are needed only to complete a model” (pp. 168). In this sense, partial identification allows for a way to test the sensitivity of effect estimates to these seemingly esoteric assumptions. Stronger assumptions (e.g., assuming a linear reporting function associated with an ordinal variable) will naturally allow for more information about a given effect estimate, however, as the strength of the necessary assumptions increase the credibility of the effect estimate decreases (Coombs 1965; Manski 2003).

Any assessment of plausible monotonic increasing transformations will critically rely on the specific elements of a given empirical application. Much like, for example, plausibly exogenous instrumental variables (Conley et al. 2012). In the empirical applications discussed in this paper, I rely on existing research that specifically assesses the plausibility of various transformations of subjective well-being and happiness scales (Banks and Coleman 1981; van Praag 1991; Oswald 2008; Kaiser and Vendrik 2019) and the cardinal properties of standardized test scores (Reardon 2008; Nielson 2017). Specifically, and as discussed in more detail in Section 4.2, if the reporting function of subjective well-being is curved at all it is likely to be concave and the relationship between test scores and student learning is likely convex. Combining the general analytical approach developed in this paper with specific assessments of plausible transformations within a given empirical context provides useful structure for testing robustness of the cardinal treatment of ordinal variables.

4 Empirical Illustration

The primary empirical illustration re-evaluates the effect of the slave trade on present-day trust in sub-Saharan Africa. Using OLS, Nunn and Wantchekon (2011) find a negative and statistically significant relationship between their preferred measure of slave trade activity and various measures of trust with the following specification:

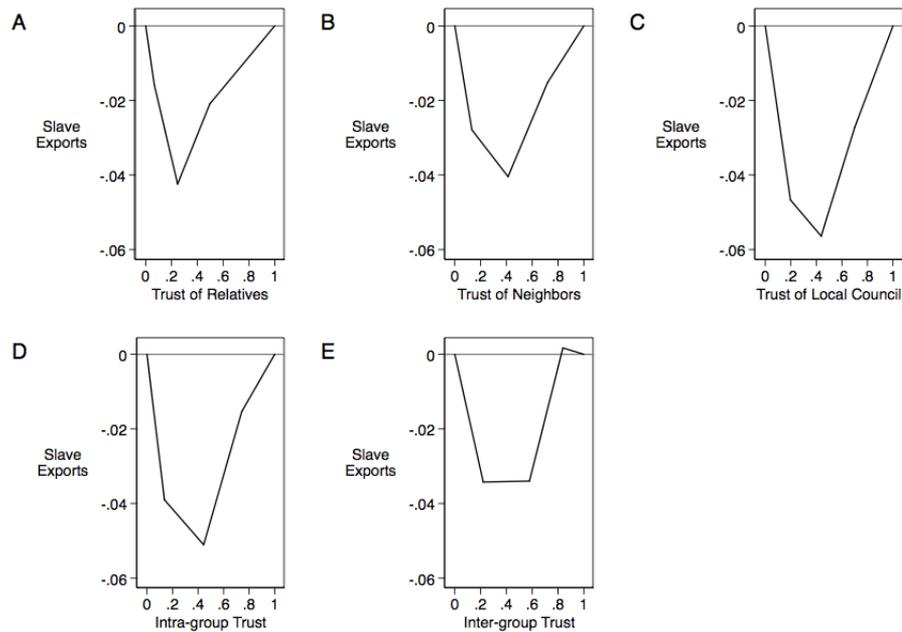
$$Trust_{iedc} = \alpha_c + \beta_e \ln Slave\ Exports_e + \gamma' X'_{iedc} + \delta' X'_{dc} + \epsilon' X'_e + \eta_{iedc} \quad (9)$$

In equation (9), i represents individuals, e ethnic groups, d districts, and c countries. The dependent variable $Trust_{iedc}$ represents each of the variables measuring trust of relatives, neighbors, the local council, intra-group, and inter-group measured on a zero through three ordinal scale—as measured in the Afrobarometer survey data. α_c captures country fixed effects, $Slave\ Exports_e$ indicates the number of slaves sold from a particular ethnic group e .¹⁵ The various X vectors are individual, district, and ethnic group level controls variables. Finally, η_{iedc} is the error term.

Concerned about the possibility of omitted variables biasing these results, the authors undertake several strategies to identify the causal relationship between the slave trade and trust. One of these strategies is instrumental variable analysis, where the distance of an individual’s ethnic group from an ocean coast instruments for slave trade activity. The instrument approximates an ethnic

15. Slave exports are natural log of one plus slave exports normalized by land area.

Figure 3. LMA Curves with Afrobarometer Measures of Trust and the Slave Trade



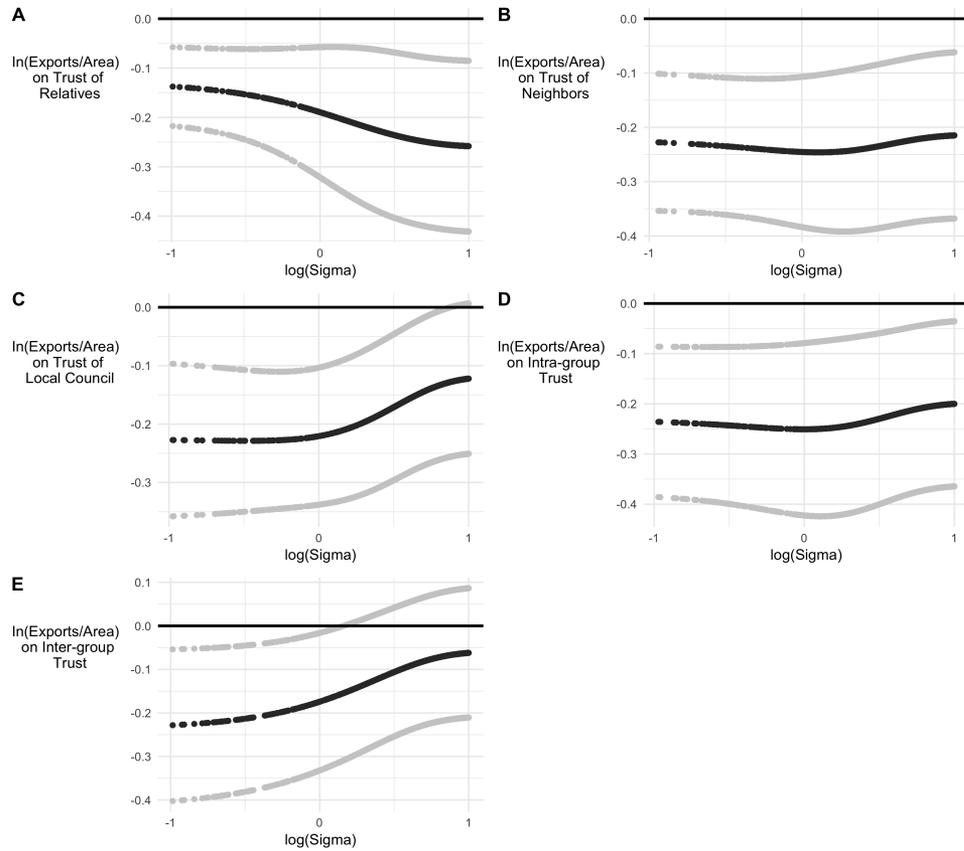
Notes: This figure shows LMA curves between the five measures of trust gathered via the Afrobarometer survey and the natural log of slave exports normalized by land area (Nunn and Wantchekon, 2011). The y-axis is fixed between all graphs.

group’s exposure to the slave trade and is unlikely to be correlated with factors that impact present day trust. In this illustration, I will examine the results from the instrumental variable estimation strategy. The simulation results from the simple OLS specification do not change the core findings from this analysis and are shown in the Online Supplement.

In this section I present three elements involved in performing this method to test for robustness to the cardinal treatment of ordinal variables. First, I comment on the results of the sufficient conditions derived by Schröder and Yitzhaki (2017). These results are illustrated as graphs of LMA curves and shown in the Online Supplement. Second, I graphically report the set of effect estimates. These results are shown by plotting the point estimate and the associated confidence interval for a (relatively extreme) range of monotonic increasing transformations. Finally, I comment on calculating plausible bounds on the originally-reported point estimates in the case that results are not robust to an extreme range of transformations.

Figure 3 shows the LMA curves for each of the five measures of trust and the variable of interest, the natural log of slave exports normalized by land area. In this figure all but one of the LMA curves do not cross the horizontal axis. The LMA curve that does cross the horizontal axis, inter-group trust, does so for relatively high values of trust. For all other measures of trust, the LMA curves suggest a negative covariance with the natural log of slave exports over land area. Taken together, these graphical results suggest that, except for perhaps the effect on inter-group trust, the empirical findings of Nunn and Wantchekon (2011) pass the second theoretical sufficient condition of Schröder and Yitzhaki (2017). Even so, the statistical significance of the findings or the overall magnitude of the results may be meaningfully affected by monotonic increasing transformations.

Figure 4. Effect Sets for Nunn and Wantchekon (2011)—Globally Concave and Convex Transformations



Notes: The dark lines represent the point estimates for a given specification with the corresponding value of $\log(\sigma)$. Logging the value of σ allows for equal share of the graph to represent concave and convex transformations. Lighter lines represent 95% confidence interval calculated with standard errors clustered by ethnicity. Each panel refers to a different specifications used in Table 5 of Nunn and Wantchekon (2011). Panel A refers to column (1) with the dependent variable trust of relatives. Panel B refers to column (2) with the dependent variable trust of neighbors. Panel C refers to column (3) with the dependent variable trust of local council. Panel D refers to column (4) with the dependent variable intra-group trust. Finally, panel E refers to column (5) with the dependent variable inter-group trust.

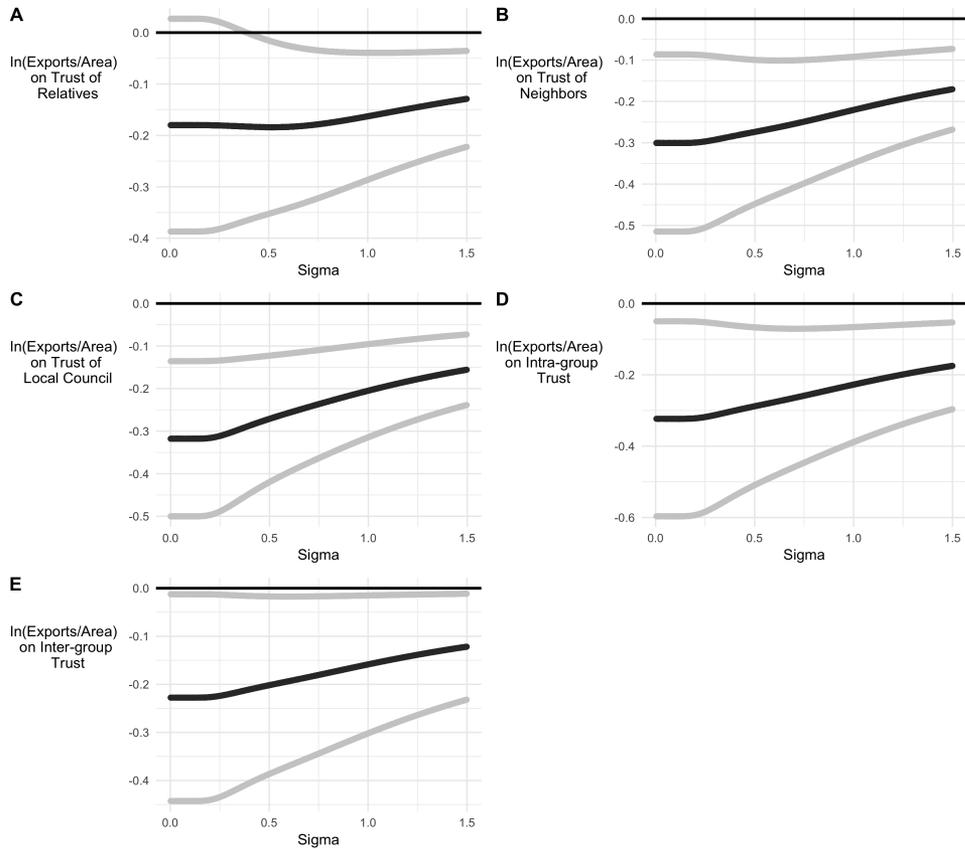
4.1 Graphical Results

Figures 4 and 5 show effect estimates for each of the two classes of transformations, for each of the five measures of trust in others, relating to the results reported in Table 5 of Nunn and Wantchekon (2011). The authors argue that the magnitudes of these estimated effects are economically meaningful. Specifically, a one standard deviation change in the intensity of the slave trade represents a -1.10 through -0.16 standard deviation change in each of the five different measures of trust in others. The authors also perform a variance decomposition to assess economic significance, as discussed below.

Globally Concave and Convex Transformations—Figure 4 shows estimated effect sets based on globally concave and convex transformations of specifications using each of the five measures of trust as dependent variables. The central finding of Nunn and Wantchekon (2011) is that individuals whose ancestors were heavily impacted by the slave trade are less trusting today. This qualitative result largely persists for the entire range of these (relatively extreme) transformations. There is no alternative transformation that changes the sign on any of the estimated coefficients.

Moreover, for three out of five specifications the effects remain statistically significant for the full range of transformations. The two exceptions are the effects of the slave trade on trust of the local council and inter-group trust, which both become statistically insignificant for convex

Figure 5. Effect Sets for Nunn and Wantchekon (2011)—Transformations with an Inflection Point



Notes: The dark lines represent the point estimates for a given specification with the corresponding sigma value. Lighter lines represent 95% confidence interval calculated with standard errors clustered by ethnicity. Each panel refers to a different specifications used in Table 5 of Nunn and Wantchekon (2011) presenting IV estimation results. Panel A refers to column (1) with the dependent variable trust of relatives. Panel B refers to column (2) with the dependent variable trust of neighbors. Panel C refers to column (3) with the dependent variable trust of local council. Panel D refers to column (4) with the dependent variable intra-group trust. Finally, panel E refers to column (5) with the dependent variable inter-group trust.

transformations. While discussing magnitude, the authors perform a variance decomposition and find that, along with the other covariates, slave exports explain 5.4% of the total variation of trust in neighbors. Additionally, of this 5.4%, about 16% is explained by slave exports. Results from the simulation analysis show that over all values of $\log(\sigma)$, along with the other covariates, slave exports explain between 4.2% and 5.4% of the total variation of trust in neighbors. Furthermore, of this 4.2 - 5.4%, roughly 16% is consistently explained by slave exports. Therefore, despite concerns about the invalid cardinal treatment of ordinal variables, the results of Nunn and Wantchekon (2011) are robust even relatively extreme transformations of the ordinal scale measuring trust.

Transformations with an Inflection Point—As discussed above, an alternative class of transformations are those with an inflection point. Figure 5 shows estimated effect sets based on transformations with an inflection point for the instrumental variable estimates from Nunn and Wantchekon (2011). Similar to the effect sets based on globally concave and convex transformations, the empirical findings of Nunn and Wantchekon (2011) are robust to a class of transformations with an inflection point. For all transformations, none of the coefficient estimates change sign. This is consistent with the theoretical conditions of Schröder and Yitzhaki (2017). Moreover, with the exception Panel A, referring to the effect on trust with neighbors, all coefficient estimates are statistically significant for all CDF transformations. In Panel A, the estimate becomes statistically insignificant for σ values close to zero, when the transformation resembles a step-function and is

essentially a binary indicator variable identifying if respondents trust their relatives or not. Finally, the magnitude of the coefficient estimates themselves are also relatively robust to these (relatively extreme) transformations.

4.2 Plausible Effect Bounds

So far the range of alternative transformations could be characterized as relatively extreme. In the case of the globally concave and convex transformations, the range is defined by any $\in [\cdot, \cdot, \cdot, \cdot]$. A transformation associated with $= \cdot, \cdot$, implies that moving from a reported trust score of one—on the zero through three ordinal scale—measures a massive relative change in latent trust while a transformation associated with $= \cdot, \cdot$ implies only moving from a reported trust score of three to a score of four measures any noticeable relative change in latent trust. In the case of transformations with an inflection point, the range extends from transformations suggesting a step function to transformations that are roughly linear. Step function transformations imply that the only useful information embedded within the trust score relating to latent trust is whether or not the respondent reports a trust score of greater than two. A roughly linear transformation, on the other hand, implies that each trust score represents an equal marginal gain in latent trust. In both of these cases the range of alternative monotonic increasing transformations likely represent transformations that are implausible in this specific empirical setting.

As noted above, however, the estimated effect sets reported in Figures 4 and 5 show that the results of Nunn and Wantchekon (2011) are robust to these relatively extreme and likely implausible transformations. This finding of robustness to extreme transformations provides evidence supporting the cardinal treatment of ordinal variables.¹⁶

This method requires one additional step in applications where estimated effects are *not* robust to a relatively extreme set of transformations—as is the case when re-analyzing the results from Aghion et al. (2016) and Bond and Lang (2013), shown in the Online Supplement. The task then is to determine a plausible range of transformations and therefore plausible bounds on the estimated effects.

What is a plausible transformation of the ordinal scale in these empirical contexts? In the case of Aghion et al. (2016), the work of Kaiser and Vendrik (2019), who focus on assessing the plausibility of transformations to subjective well-being and happiness scales, is instructive. Kaiser and Vendrik (2019) cite three studies that experimentally aim to identify the functional form of the reporting function defining the relationship between subjective feelings and objective reality. Empirical analysis reported by Oswald (2008) cannot reject that the reporting function is linear—if it is curved at all, it is slightly concave. Additionally, results found by van Praag (1991) and Banks and Coleman (1981) cannot rule out the finding that individuals in their study use a linear reporting function on average. Although these cited studies provide some suggestive evidence that assuming a linear reporting function may indeed be valid, this likely will be considered a relatively strong assumption. Assuming linearity of the reporting function leads to more precise effect estimates, but ultimately with less credibility.

Based on this assessment, Table A1 in the Online Supplement reports plausible bounds on the effect estimates for the results of Aghion et al. (2016) based on transformations associated with $\in [\cdot, \dots, f, \cdot]$. This range of plausible transformations allows for both concave and convex transformations, but only transformations such that response categories are only ten times larger on opposite ends of the scale.¹⁷ As shown in the Online Supplement, when limiting alternative

16. The evidence is weak in the sense that we may have simply chosen the wrong class of transformations, however the transformations used to estimate the effects in Figures 4 and 5 should represent a large share of plausible classes of transformations.

17. A conceptual way to characterize these transformations is that individuals who hold a reporting function defined by $= \cdot, \dots$ are pessimistic in the sense that only relatively high levels of latent well-being are sufficient to move them off

transformations to this more limited plausible range, I find that the set of effects extend from a small and statistically insignificant effect to an effect that is statistically significant and almost 50 percent larger than that originally reported. Thus, the results of Aghion et al. (2016) are not robust to even plausible transformations of the ordinal scale measuring subjective well-being.

In the case of Bond and Lang (2013), the work of Reardon (2008), and specifically insights from applying item response theory (IRT) results to the ECLS-K data, provide structure for an assessment of plausible transformations. In particular, Reardon (2008) estimates IRT parameters indicating how each question on the ECLS-K test predicts student learning.¹⁸ The author finds that roughly 40 percent of the questions in the ECLS-K have a relatively high likelihood of predicting no information about student learning. Most fundamentally, this suggests that the relationship between the observed test score and student learning is unlikely to be linear. It also suggests that at low (high) test score levels the marginal gain in student learning is relatively small (high). Taken together, this suggests that transformations with an inflection point are less plausible in the context of test scores, and implies that a plausible reporting function is convex to some degree. To what degree? Assuming that 40 percent of the questions could be answered correctly by guessing suggests that a value of 5 is relatively plausible.¹⁹ Therefore, a plausible range of transformations extends from $\in [, \uparrow]$. As shown in the Online Supplement, when limiting alternative transformations to this more limited plausible range, I find the set of estimates largely replicates the core findings of Bond and Lang (2013) and support the fragility of these results to transformations of the test score scale.

This discussion demonstrates how this method can be used to test for robustness of effect estimates to the cardinal treatment of ordinal variables. As the supplemental illustrations shown in the Online Supplement highlight, this method can apply to virtually any variable that cannot be directly observed and must be quantitatively measured on an ordinal scale.

5 Conclusion

This paper builds off recent contributions of Schröder and Yitzhaki (2017) and Bond and Lang (2019) on the appropriateness of the cardinal statistical treatment of ordinal variables. I develop a partial identification method for testing the robustness of empirical estimates to a set of monotonic increasing transformations of the ordinal scale. This approach allows for the calculation of a set of plausible effect estimates based on a range of assumptions about the cardinal properties of an ordinal variable. To illustrate this method, I re-examine the effect of the slave trade on trust in sub-Saharan Africa (Nunn and Wantchekon 2011). Supplemental illustrations include analysis on the effect of creative destruction on subjective well-being (Aghion et al. 2016) and the evolution of the black-white test score gap (Bond and Lang 2013).

The empirical applications clarify three empirical points that extend existing theoretical insights (Schröder and Yitzhaki 2017). First, failing existing theoretical tests for the valid cardinal treatment of ordinal variables may not be a serious problem in practice because it may be the case that only extreme monotonic increasing transformations substantially change empirical results. Second, passing existing theoretical tests does not necessarily imply that empirical results are robust because the size of estimated coefficients could change dramatically for monotonic increasing transformations. Third, passing existing theoretical tests does not imply that the statistical significance of results is robust to monotonic increasing transformations of the ordinal scale, even if the size of the coefficient estimates are relatively robust.

These findings carry implications for future empirical research which necessitates the use of

relatively low levels of the observed scale. Conversely, individuals who hold a reporting function defined by $= f. \uparrow$ are optimistic.

18. For transparency, the effectiveness of IRT is a historic and still active topic of debate in the psychometric literature. See, e.g., Lord (1975).

19. That is, only after a student answers over 40 percent of the questions correctly does the test score begin to take on a positive relationship with student learning. See Figure 1.

ordinal variables. Inevitably, as social science research extends itself into realms of society, the economy, and politics where factors cannot be quantitatively measured or directly observed, the need to use ordinal variables becomes increasing frequent. This situation presents a challenge to researchers regarding empirical methodology. The research builds on the recent work of Schröder and Yitzhaki (2017) and Bond and Lang (2019), who show that it is no longer valid to assume that the ordinality or cardinality of ordinal variables makes no qualitative difference. The cardinal treatment of ordinal variables can lead to incorrect empirical findings. Although some empirical findings may be robust to alternative monotonic increasing transformations, many will not be.

Supplementary Material

(This is dummy text) For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.xxxx.xx>.

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.
- Aitchison, J. and Silvey, S.D. (1957) "The Generalization of Probit Analysis to the Case of Multiple Responses" *Biometrika*, 44 pp. 131-140.
- Baker, F. (2001) *The Basics of Item Response Theory*, College Park: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.
- Banks, W.P., and Coleman, M.J. (1981) "Two subjective Scales of Number," *Perception & Psychophysics*, vol. 29, no. 2, pp. 95-105.
- Becker, W. and Kennedy, P. (1992) "A Graphical Exposition of the Ordered Probit" *Econometric Theory*, 8 (1) pp. 127-131.
- Baird, S., de Hoop, J., and Oxler, B. (2013) "Income Shocks and Adolescent Mental Health" *Journal of Human Resources*, 48 (2), pp. 370-403.
- Bloem, J. (2020) "Replication Data for: How Much Does the Cardinal Treatment of Ordinal Variables Matter? An Empirical Investigation," <https://doi.org/10.7910/DVN/BABCEN>, Harvard Dataverse.
- Bond, T. and Lang, K. (2013) "The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results" *The 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 June 10, 2019.
- 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.
- Box, G. and Cox, D. (1964) "An Analysis of Transformations" *Journal of the Royal Statistical Society* 26 (2) pp. 211-252.
- Bryson, A. and MacKerron, G. (2017) "Are You Happy While You Work?" *The Economic Journal* 127 (599) pp. 106-125.
- Clark, A. and Oswald, A. (1996) "Satisfaction and Comparison Income" *Journal of Public Economics* 61 (3) pp. 359-381.
- Clark, A. and Oswald, A. (1994) "Unhappiness and Unemployment" *The Economic Journal* 104 (424) pp. 648-659.
- Conley, T.G., Hansen, C.B., and Rossi, P.E. (2012) "Plausibly Exogenous," *The Review of Economics and Statistics*, vol. 94, issue 1, pp. 260-272.
- Coombs, C. (1965) *Theory of Data*, New York: Wiley & Sons.

- Cornaglia, F., Feldman, N.E., Leigh, A. (2014) "Crime and Mental Well-Being" *Journal of Human Resources*, 49 (1), pp. 110-140.
- Daly, M.C. and Wilson, D.J. (2009) "Happiness, Unhappiness, and Suicide: An Empirical Assessment," *Journal of the European Economic Association*, vol. 7, issue 2-3, pp. 539-549.
- Davis, S., Haltiwanger, J., and Schuh, S. (1996) "Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts" *Small Business Economics* 8 (4) pp. 297-315.
- Deaton, A. (2018) "What do self-reports of wellbeing say about life-cycle theory and policy" *Journal of Public Economics*, vol. 162, pp. 18-25.
- Deaton, A. (2008) "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll" *Journal of Economic Perspectives*, 22 (2), pp. 1-31.
- Di Tella, R., MacCulloch, R. and Oswald, A. (2003), "The Macroeconomics of Happiness," *The Review of Economics and Statistics*, vol. 85, no. 4, pp. 809-827.
- 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.
- 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.
- Fryer, R. and Levitt, S. (2004) "Understanding the Black-White Test Score Gap in the First Two Years of School" *The Review of Economics and Statistics*, 86 (2) pp. 447-464.
- Fryer, R. and Levitt, S. (2006) "The Black-White Test Score Gap Through Third Grad" *American Law and Economics Review*, 8 pp. 248-281.
- 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.
- Graham, C., Egger, A., and Sukhtanker, S. (2004) "Does happiness pay? An exploration based on panel data from Russia?" *Journal of Economic Behavior and Organization*, 55 (3) pp. 319-342.
- Greene, W. (2012) *Econometric Analysis* (Seventh ed.) Boston: Pearson Education.
- Hadar, J. and Russell, W.R. (1969) "Rules for Ordering Uncertain Prospects" *American Economic Review*, 49 (1) pp. 25-34
- Hawthorne, W. (2003) *Planting Rice and Harvesting Slaves: Transformations along the Guinea-Bissau Coast, 1400-1900*. Portsmouth, NH: Heinemann.
- Heckman, J.J. (1981) "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," In: Manski, C., McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA.
- Hopkins, D. and King, G. (2010) "Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability" *Public Opinion Quarterly*, pp. 1-22.
- Kaiser, C. and Vendrik C.M. (2019) "How threatening are transformations of reported happiness to subjective wellbeing research?" *SocArXiv Working Paper*.
- King, G. Murray, C. Salomon, J., Tandon, A. (2004) "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research" *American Political Science Review*, 98 (1) pp. 191-207.
- King, G. and Wand, J. (2007) "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes" *Political Analysis*, 15 pp. 46-66.
- Koelle, S.W. (1854) *Polyglotta Africana: Or a Comparative Vocabulary of Nearly Three Hundred Words and Phrases, in More than One Hundred Distinct African Languages*. London: Church Missionary House.
- 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.
- Jacob, B. and Rothstein, J. (2016) "The Measurement of Student Ability in Modern Assessment Systems" *Journal of Economic Perspectives*, 30 (3) pp. 85-108.
- Jencks, C. and Phillips, M. (1998) "The Black-White Test Score Gap: An Introduction" in Jencks, C. and Phillips, M. (eds.) *The Black-White Test Score Gap* Washington DC: Brookings Institution Press.
- Lancaster (2000) "The incidental parameter problem since 1948" *Journal of Econometrics*, 95 pp. 391-413.
- Lang, K. (2010) "Measurement Matters: Perspectives on Education Policy from an Economist and School Board Member" *The Journal of Economic Perspectives*, 24 (3) pp. 167-181.
- Lachowska, M. (2017) "The Effect of Income on Subjective Well-Being: Evidence from the 2008 Economic Stimulus Tax Rebates" *Journal of Human Resources*, 52 (2), pp. 374-417.
- Lord, F.M. (1975) "The 'Ability' Scale in Item Characteristic Curve Theory," *Psychometrika*, vol. 40. No. 2. pp. 205-217.
- Lorenz, M.O. (1905) "Methods of Measuring the Concentration of Wealth" *Journal of the American Statistical Association*, No. 70 pp. 209-219.
- Manski, C. (2003) *Partial Identification of Probability Distributions*, Springer Series in Statistics, New York: Springer.
- Marks, B., and Flemming, N. (1999) "Influences and consequences of well-being among Australian young people 1980-1995" *Social Indicators Research* 46 pp. 301-323.
- McKelvey, R. and Zavoina, W. (1975) "A Statistical Model for the Analysis of Ordered Level Dependent Variables" *Journal of Mathematical Sociology*, 4 pp. 103-120.
- Neyman, J. and Scott, E. (1948) "Consistent estimation from partially consistent observations" *Econometrica*, 16, pp. 1-32.
- Nielsen, E. (2017) "How Sensitive are Standard Statistics to the Choice of Scale?" *Unpublished Working Paper*.
- Nunn, N. and Wantchekon, L. (2011) "The Slave Trade and the Origins of Mistrust in Africa" *American Economic Review* 101 (7) pp. 3221-3252.
- Oswald, A.J. (2008) "On the curvature of the reporting function from objective reality to subjective feelings" *Economics Letters* 100 pp. 369-372.
- Piot, C. (1996) "Of Slaves and the Gift: Kabre Sale of Kin during the Era of the Slave Trade" *Journal of African History*, 37 (1): pp. 31-49.
- Putnam, R. (2001) *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, 1st edition, New York: NY.
- Reardon, S.F. (2008) "Thirteen Ways of Looking at the Black-White Test Score Gap," *Sanford University Working Paper*.
- Riedl, M. and Geishecker, I. (2014) "Keep it simple: Estimation Strategies for Ordered Response Models with Fixed Effects", *Journal of Applied Statistics*, vol. 41. (11), pp. 2358-2374.
- Schaffer, M.E. (2015) "GINIREG: Stata Module for Gini Regression" *Available online:* <https://EconPapers.repec.org/RePEc:boc:bocode:s457958>
- 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.
- Snell (1964) "A Scaling Procedure for Ordered Categorical Data" *Biometrics*, 20 pp. 592-607.
- Stevens, S. (1946) "On the Theory of Scales of Measurement" *Science*, vol. 10, pp. 667-680.
- Stevenson, B. and Wolfers, J. (2013) "Subjective Well-Being and Income: Is There Any Evidence of Satiation?" *American Economic Review: Papers and Proceedings*, 103 (3) pp. 598-604.

- Tamer, E. **(2010)** "Partial Identification in Econometrics" *The Annual Review of Economics*, 2 pp. 167-195.
- Thorndike, R. **(1966)** "Intellectual Status and Intellectual Growth" *Journal of Educational Psychology*, vol 57. pp. 121-127.
- Van Praag, B. **(1991)** "Ordinal and Cardinal Utility: An Integration of the Two Dimension of the Welfare Concept," *Journal of Econometrics*, vol. 50, no. 1-2, pp. 69-89.
- Wang, S. and Zhou, W. **(2018)** "Do Siblings Make Us Happy?" *Economic Development and Cultural Change*, vol. 66, no. 4. pp. 827-840.
- Wunder, C., Wiencierz, A., Schwarze, J. and Kuchenhoff, H. **(2013)** "Well-Being over the Life Span: Semiparametric Evidence from British and German Longitudinal Data" *Review of Economics and Statistics*, vol. 95 (1), pp. 154-167.
- Yitzhaki, S. and Schechteman, E. **(2012)** "Identifying monotonic and non-monotonic relationships" *Economics Letters*, 116 (1) pp. 23-25.
- Yitzhaki, S. and Schechteman, E. **(2013)** *The Gini Methodology: A Primer on a Statistical Methodology*. Springer. New York.