How Much Does the Cardinal Treatment of Ordinal Variables Matter?

An Empirical Investigation

Jeffrey R. Bloem, *Department of Applied Economics and Minnesota Population Center, University of Minnesota*

MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

A Simple Illustration

Consider an ordinal scale measuring "**satisfaction**", with the following four categories:

|----------|----------|----------|----------|

"Very Dissatisfied" "Dissatisfied" "Satisfied" "Very Satisfied"

Suppose there are two groups of people:

Group A:	Group B:
One person is "Very Dissatisfied"	One person is "Dissatisfied"
Another is "Very Satisfied"	Another is "Satisfied"

Question: Which group is more satisfied? The **answer** depends on the numerical values assigned to the response categories.

	Category				
	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied	**Which group is more satisfied?**
Numerical Values					
Linear	0	1	2	3	**Tie**
Concave	0	1.75	2.5	3	**Group B**
Convex	0	0.5	1.25	3	**Group A**

The Misuse of Ordinal Variables

Ordinal variables are used to measure any concept that does not have a direct quantitative unit of measure. Consider, for example, subjective well-being. Suppose an individual's well-being is characterized as follows:

$$Y^* = X'\beta + \epsilon$$

Here Y^* is unobserved actual well-being. The vector X represents observable variables and β is a vector of regression coefficients. Since, Y^* cannot be directly observed, subjective well-being, Y, is measured via an ordinal variable with various values μ representing threshold points on the ordinal scale.

$$Y = \begin{cases} 0 & \text{if} \quad Y^* \le \mu_1, \\ 1 & \text{if} \quad \mu_1 < Y^* \le \mu_2, \\ 2 & \text{if} \quad \mu_2 < Y^* \le \mu_3, \\ \vdots & \\ N & \text{if} \quad \mu_{N-1} < Y^* \end{cases}$$

The problem is the values of μ are unknown. Estimating the above regression specification using ordinary least squares (OLS) with the observed ordinal scale of Y as the dependent variable, assumes that the values of Y have known and fixed intervals.

The Method

Goal: To understand robustness of any empirical finding to a range of transformations.

1. Define a parameterized function that represents all "reasonable" transformations.

The main results use the following function:

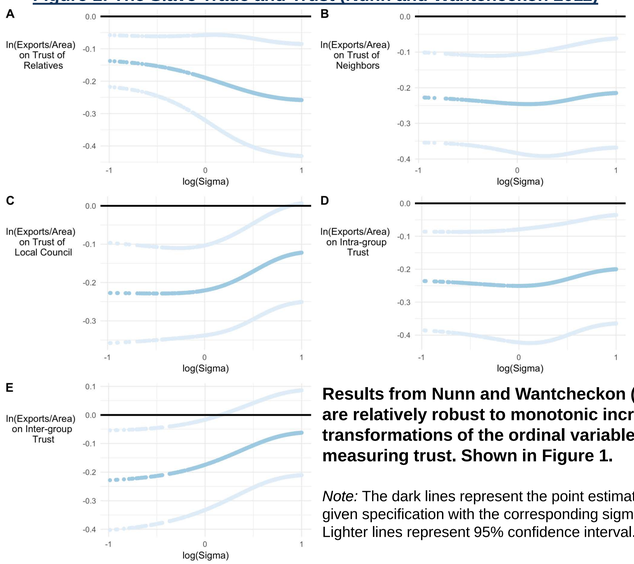
$$T(Y) = Y_{Max} \times \left(\frac{Y}{Y_{Max}}\right)^{\sigma}$$

In this function, Y is the linear ordinal scale ranging from zero through the maximum value of the scale. The σ parameter controls the convexity or concavity of the ordinal scale. If $\sigma = 1$, then the scale remains in its linear form. If $0 < \sigma < 1$, then the scale will be concave to some degree. If $\sigma > 1$, then the scale will be convex to some degree.

2. Run Monte Carlo simulations, randomly picking a parameter within the finite domain of σ.

Simulation Results

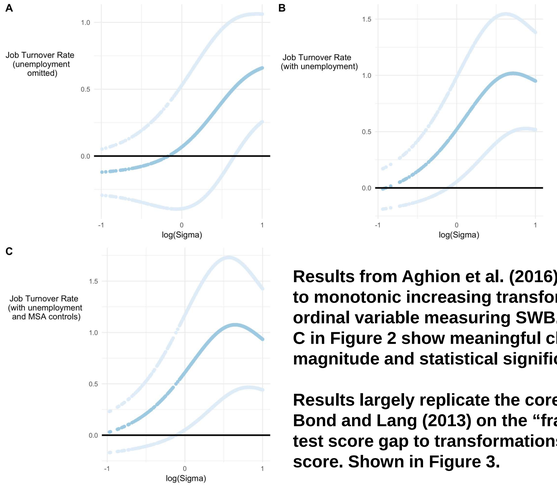
Figure 1: The Slave Trade and Trust (Nunn and Wantcheckon 2011)



Results from Nunn and Wantcheckon (2011) are relatively robust to monotonic increasing transformations of the ordinal variable measuring trust. Shown in Figure 1.

Note: The dark lines represent the point estimates for a given specification with the corresponding sigma value. Lighter lines represent 95% confidence interval.

Figure 2: Subjective Well-Being and Creative Destruction (Aghion et al. 2016)



Results from Aghion et al. (2016) are not robust to monotonic increasing transformations of the ordinal variable measuring SWB. Panels B and C in Figure 2 show meaningful change in magnitude and statistical significance.

Results largely replicate the core findings of Bond and Lang (2013) on the "fragility" of the test score gap to transformations of the test score. Shown in Figure 3.

Figure 3: K-3 Racial Test Score Gap (Bond and Lang 2013)