Economics 1425 Research Proposal

Dillon Smith

3/14/2019

Research Premise

I am interested in exploring the relationship between various measures of familial structure and relative moral values as discussed by Prof. Benjamin Enke during week 4's lecture "The Political Economy of Moral Universalism." I found the data to be incredibly interesting and see it having a large range of applicability to political, financial and behavioral economics, and thus I am interested in researching the effects of familial structure on moral values and whether or not the effects these familial structure characteristics have on an individual's income are merely proxies for effects coming from moral values.

Introduction

Much of the research in economics relating to moral values stems from Jonathan Haidt's Moral Foundations Theory, a theory based in pluralism suggesting that there are five foundations of morality: care, fairness/proportionality, loyalty/ingroup, authority/respect, and sanctity/purity.[1] These moral foundations have been shown to correlate with political and religious ideology, an idea strengthened by Benjamin Enke's articles *Moral Values and Voting*[2] and *Moral Universalism and the Structure of Ideology*[3], respectively.

Moving on to the second important part of the research, the link between familial structure and various measures of social outcomes has been established in many studies. Researchers have found links between it and educational achievments[4], educational attainment[5], likelihood of falling into poverty[6], and compulsive consumption[7].

Both moral values and familial structures are intricately linked to socioeconomic outcomes, and given this link I feel that they must be linked to each other as well. The goal of this paper will be to explore and shed light onto the effects familial structure has on moral values, and inderectly the effect it has on outcomes via values.

Data

The data for this research comes from two major sources: Raj Chetty's *Opportunity Insights*[8] project and Benjamin Enke's moral values data from *Moral Values and Voting*.

Raj Chetty and the Opportunity Insights team have made their data available on the project's website so that people may "analyze social mobility and a variety of other outcomes from life expectancy to patent rates by neighborhood, college, parental income level, and racial background." I began with data mapping all measured outcomes by race and gender at the county level and filtered this data to only include patrticularly pertinent variables—fraction of children who have a male claimer in the year they are linked to parents,

---

[1] Jesse Graham (2013)
[2] Enke (2018b)
[3] Enke (2018a)
[4] Astone and McLanahan (1991)
[5] Sandefurf and Wells (1999)
[6] McLanahan (1985)
[7] Aric Rindfleisch and Denton (1997)
[8] Chetty and Friedman (2019)

fraction of children who have a female claimer in the year they are linked to parents, fraction incarcerated on April 1st, 2010, fraction of women who grew up in the given tract who ever claimed a child who was born when they were between the ages of 13 and 19 as a dependent at any point, fraction of children claimed by two people in the year they are linked to parents, mean percentile rank (relative to other children born in the same year) in the national distribution of household income, and mean percentile rank (relative to other children born in the same year) in the national distribution of individual income.

For each variable listed above, I have obtained values for each race/gender combination as well as values for each race including both genders at the 1st, 50th, and 100th percentiles. This data is collected and cleaned by the Opportunity Insights team, so I did not perform any additional cleaning. Also from the Opportunity Insights page I downloaded data on various covariates to be used to discover underlying fixed effects/variables that may be contributing to omitted variable bias to be used in future regressions. Table 1[9] displays means for each of the income-related covariates in each of the seven regions of the U.S. (Northeast, "Southeast", "Midwest", "Southwest", "Rocky Mountains", "Pacific", and "Non-Contiguous"), and Table 2 does so for the education-related ones.

Table 1: Means by Region for Income-Related Covariates

Region	Mean H.H. Inc. 2000	Median H.H. Inc. 1990	Median H.H. Inc. 2016	Emp. Rate 2000
1	$77,298	$32,010	$60,277	60%
2	$65,705	$24,874	$47,004	56%
3	$69,130	$26,542	$52,164	62%
4	$64,705	$23,907	$49,659	55%
5	$71,382	$26,720	$53,963	62%
6	$75,385	$29,283	$56,126	56%
7	$90,403	$42,569	$73,141	61%

The Opporunity Insights Data contained information regarding 18 covariates to be used to account for state and area fixed effects as well as potential lurking variables causing omitted variable bias which may affect the accuracy of the regressions. For region, 1 corresponds to "Northeast" (containing CT, MA, ME, NH, NJ, NY, PA, RI, and VT), 2 to "Southeast" (containing AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, SC, TN, VA, and WV), 3 to "Midwest" (containing IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI), 4 to "Southwest" (containing AZ, NM, OK, and TX), 5 to "Rocky Mountains" (containing CO, ID, MT, NV, UT, and WY), 6 to "Pacific" (containing CA, OR, and WA), and 7 to "Non-Contiguous" (containing AK and HI). Mean H.H. Inc. 2000 is the mean household income in dollars calculated from the 2000 Decennial Census; Median H.H. Inc. 1990 is the median household income in dollars calculated from the 1990 Census; Median H.H. Inc. 2016 is the median household income in dollars calculated from the 2012-2016 American Community Survey; and Emp. Rate 2000 is the sum of employed females and employed males divided by the total population 16 years and over calculated from the 2000 Decennial Census.

This data was collected from a wide variety of government sources includng census and BLS data, suggesting that biases within the data are not an issue. There is, however, a small amount of noise added to the estimates by the Opportunity Insights team (<10% of the standard error) in orer to protect the privacy of respondents.

The data on moral values coming from Prof. Enke is also organized down to the county level. It is drawn from Jonathan Haidt's "Moral Foundations Questionnaire." It contains information on the relative importance of universal vs. communal values, absolute importance of universal values, and communal importance of universal values. Universal values are not context specific and reflect more consequentialist moral standpoints—they include considerations of rights, justice, impartial fairness, and avoidance of externalities—and communal values are values that can change dependent on situation and are relationship-/group-specific and reflect more deontological moral standpoints—they include considerations of loyalty, betrayal, obedience, and respect. Relative importance is the difference between a given county's absolute communal and absolute universal importances. Table 3 provides means for each of these variables for each of the aforementioned seven regions

---

[9]Tables created using R package stargazer, Hlavac (2018)

Table 2: Means by Region for Education-Related Covariates

Region	Ptg. College or More 2000	Ptg. College or More 2010	Mean Math Test Scores 2013
1	23%	26%	3.42
2	16%	19%	3.11
3	17%	20%	3.56
4	17%	19%	3.11
5	23%	26%	3.44
6	21%	24%	2.77
7	23%	26%	3.28

The Opporunity Insights Data contained information regarding 18 covariates to be used to account for state and area fixed effects as well as potential lurking variables causing omitted variable bias which may affect the accuracy of the regressions. For region, 1 corresponds to "Northeast" (containing CT, MA, ME, NH, NJ, NY, PA, RI, and VT), 2 to "Southeast" (containing AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, SC, TN, VA, and WV), 3 to "Midwest" (containing IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI), 4 to "Southwest" (containing AZ, NM, OK, and TX), 5 to "Rocky Mountains" (containing CO, ID, MT, NV, UT, and WY), 6 to "Pacific" (containing CA, OR, and WA), and 7 to "Non-Contiguous" (containing AK and HI). Ptg. College or More 2000 is the number of people aged 25 or older who have a bachelor's degree, master's degree, professional school degree, or doctorate degree, divided by the total number of people aged 25 or older in a tract calculated from the 2000 Census; Ptg. College or More 2010 is the same thing calculated using the 2006-2010 ACS; and Mean Math Test Scores 2013 is the mean 3rd grade math test scores in 2013 obtained from the Stanford Education Data Archive and measured at the district level.

of the U.S. in which we can see that regions are relatively evenly split on whether they place higher importance in universal or communal morals.

Table 3: Means by Region for Moral Values Variables

Region	Relative Morals	Universal Morals	Communal Morals
1	0.58	0.41	-0.48
2	-0.1	-0.05	0.11
3	-0.03	-0.03	0.01
4	-0.34	-0.06	0.37
5	-0.1	-0.31	-0.05
6	0.58	0.28	-0.55
7	0.46	-0.24	-0.7

The Moral Values Data contained information regarding 3 categories of moral values: relative importance of universal vs. communal values (calculated by subtracting importance of universal values from importance of communal values), absolute importance of universal values (implicit from the Haidt survey), and absolute importance of communal values (also implicit from the haidt survey). For region, 1 corresponds to "Northeast" (containing CT, MA, ME, NH, NJ, NY, PA, RI, and VT), 2 to "Southeast" (containing AL, AR, DC, DE, FL, GA, KY, LA, MD, MS, NC, SC, TN, VA, and WV), 3 to "Midwest" (containing IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, and WI), 4 to "Southwest" (containing AZ, NM, OK, and TX), 5 to "Rocky Mountains" (containing CO, ID, MT, NV, UT, and WY), 6 to "Pacific" (containing CA, OR, and WA), and 7 to "Non-Contiguous" (containing AK and HI).

Increasing the specificity from the region estimates, we can look at state-specific means. We can see a wide variety in the values for relative importance of universal vs. communal values across states, as seen in Figure 1. If we wanted to increase specificity even more, we would see that, as shown by Prof. Enke displays in his paper *Moral Values and Voting*, this variety exists across counties within states as well. These counties themselves vary in relative percentages of races, of children growing up under different family structures, and of a myriad of different factors which could be contributing to this variation in moral values.

Figure 1: Heatmap of Difference between Communal and Universal Values in the U.S.A.
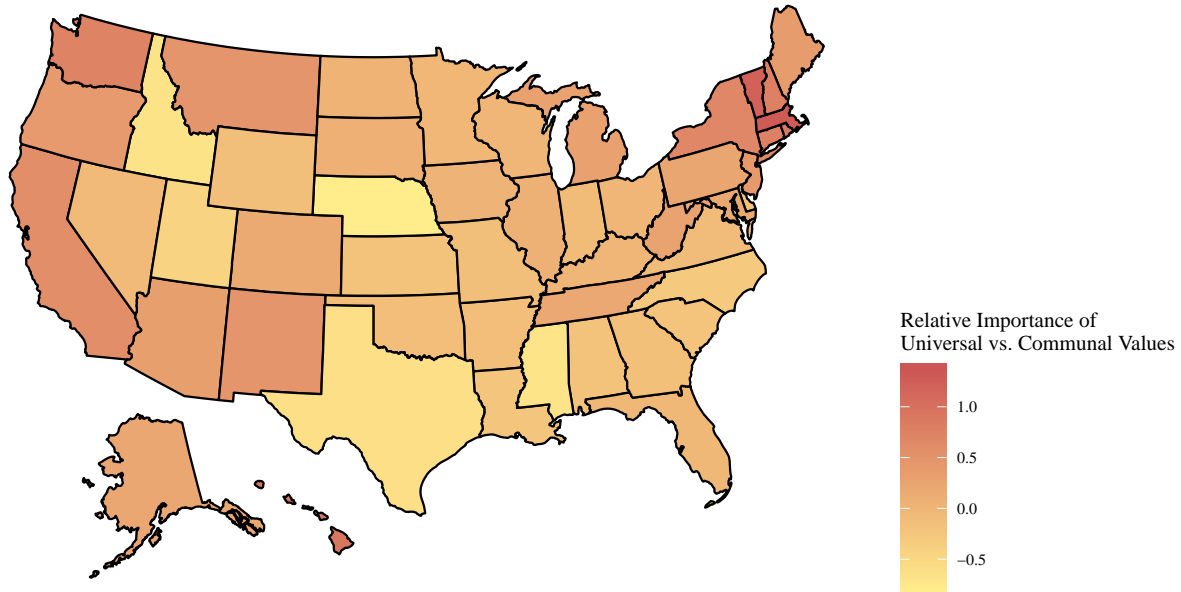
This data was collected by Jonathan Haidt from his "Moral Foundations Questionnaire"[10]: an online, elective quiz which reports individuals' scores on the five foundations of Moral Foundations Theory but also records demographic information and relative importance of universal and communal values to perform data analysis. Because of Haidt's relative popularity and fame, the dataset is very expansive, but the data is likely slightly skewed towards younger and more educated people.

Methods

I plan on exploring the relationship between the dependent and independent variables by expanding upon the regressions performed to obtain the preliminary results discussed in the next section. Much of the analysis I have done so far is on the effect the prevalence each independent variable has within each individual race has on relative importance of communal values over universal values.

Going forward, I will be creating multipe visualizations using both traditional x- and y-variable graphs as well as PCA analysis to look for relationships to further analyze in regression analysis. I will then use the inferences I draw from this exploratory analysis perform regressions using quadratic, cubic, and polynomial regressions as well as fixed effects regression techniques.

So far, each regression is a MLS regression following the formula:

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_n X_{ni} + \epsilon_i$$

where β_0 is the constant intercept, β_1 through β_n are the slope coefficients for the various independent variables, and ϵ_i is the error term for a given observation.

---

[10]http://yourmorals.org/haidtlab/mft/index.php?t=questionnaires

Table 4: Effects of Independent Variables on Relative Importance of Universal vs. Communal Values by Race

	Dependent variable:				
	Relative Importance of Universal vs. Communal Values				
	White	Black	Asian	Hispanic	Native American
	(1)	(2)	(3)	(4)	(5)
Has Dad	−0.586	−3.438***	−0.090	−0.490	3.052
	(6.704)	(0.993)	(4.298)	(1.255)	(3.210)
Has Mom	−2.115	−0.837	−0.336	0.404	3.205
	(5.605)	(1.000)	(4.184)	(1.193)	(3.170)
Incarceration	−9.072*	−0.245	1.079	−1.643	−2.302*
	(5.354)	(1.014)	(3.853)	(1.767)	(1.336)
Teen Birth	−5.360***	−0.994***	−1.780***	−2.019***	−0.511
	(0.583)	(0.292)	(0.452)	(0.269)	(0.352)
Two Parents	−1.083	0.455	−1.578	−1.960	−5.515*
	(6.246)	(0.979)	(4.092)	(1.224)	(3.146)
House. Inc.	−9.674***	3.613***	−2.694**	−2.993***	−1.685
	(1.707)	(1.070)	(1.101)	(0.874)	(1.175)
Indiv. Inc.	5.494***	−3.468***	3.480***	1.592*	−0.435
	(2.010)	(1.084)	(1.121)	(0.878)	(1.324)
Constant	6.684	3.434***	1.595	2.759**	−0.189
	(6.016)	(1.081)	(4.417)	(1.240)	(3.239)
Observations	2,262	1,408	884	1,639	642
R^2	0.100	0.094	0.045	0.074	0.098
Adjusted R^2	0.097	0.090	0.038	0.070	0.088

Note: *p<0.1; **p<0.05; ***p<0.01

The data for this regression comes from two sources: Raj Chetty's Opportunity Insights All Outcomes data and Benjamin Enke's Moral Values data. The dependent variable is calculated from the Moral Values data by subtracting the universal values score from the communal values score. The independent variables all come from the Opportunity Insights data. "Has Dad" represents the fraction of children at the 50th percentile for a given race with a male claimer in the year they are linked to a parents. "Has Mom" represents the fraction of children at the 50th percentile for a given race with a female claimer in the year they are linked to a parents. "Incarceration" represents the fraction of the population at the 50th percentile for a given race incarcerated on April 1st, 2010. "Teen Birth" represents the fraction of women at the 50th percentile for a given race who ever claimed a child who was born when they were between the ages of 13 and 19 as a dependent at any point. "Two Parents" represents the fraction of children at the 50th percentile for a given race who were claimed by two people in the year they are linked to parents. "House. Inc." represents the mean percentile rank at the 50th percentile for a given race in the national distribution of household income (i.e. own earnings and spouse's earnings) measured as mean earnings in 2014-2015 for the baseline sample. "Indiv. Inc." represents the mean percentile rank at the 50th percentile for a given race in the national distribution of individual income (i.e. just own earnings) measured as mean earnings in 2014-2015 for the baseline sample.

Preliminary Results

Looking at Table 4 and based on preliminary regressions, the strongest correlates with relative importance of universal vs. communal values are household income, individual income, and teen birth. There were comparatively high numbers for incarceration rates in white citizens, has dad for black citizens, and two parents for Native American citizens. Even though the coefficients for many of the variables were statistically significant, the R^2 and Adjusted R^2 for each of the regressions is extremely low. Further, without any controls, there is a large risk of omitted variable bias affecting the estimates returned by the regression.

Other Options

Given the breadth of the data available on the Opportunity Insights website, there are a variety of other variables for which I can explore its relationship with the moral values data. Some I am considering looking into if including controls eliminates the effects of the familial structure variables include kfr and kir—mean percentile rank (relative to other children born in the same year) in the national distribution of household incomeand mean percentile rank (relative to other children born in the same year) in the national distribution of individual income, respectively.

Bibliography

Aric Rindfleisch, James E. Burroughs, and Frank Denton. 1997. "Family Structure, Materialism, and Compulsive Consumption." *Journal of Consumer Research* 23 (4): 312–25. doi:10.1086/209486.

Astone, Nan Marie, and Sara S. McLanahan. 1991. "Family Structure, Parental Practices and High School Completion." *American Sociological Review* 56 (3): 309–20. http://www.jstor.org/stable/2096106.

Chetty, Raj, and John Friedman. 2019. "Opportunity Insights, Previously the Equality of Opportunity Project." Opportunity Insights, Harvard University. https://opportunityinsights.org/data/.

Enke, Benjamin. 2018a. "Moral Universalism and the Structure of Ideology." Cambridge, MA: Harvard University.

———. 2018b. "MORAL VALUES AND VOTING." Cambridge, MA: Harvard University. https://www-nber-org.ezp-prod1.hul.harvard.edu/papers/w24268.pdf.

Hlavac, Marek. 2018. "stargazer: Well-Formatted Regression and Summary Statistics Tables." R package version 5.2.2. https://CRAN.R-project.org/package=stargazer.

Jesse Graham, Sena Koleva, Jonathan Haidt. 2013. "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." *Advances in Experimental Social Psychology* 47: 55–130. doi:10.1016/B9780124072367000024.

McLanahan, Sara. 1985. "Family Structure and the Reproduction of Poverty." *American Journal of Sociology* 90 (4): 873–901. doi:10.1086/228148.

Sandefurf, Gary D., and Thomas Wells. 1999. "Does Family Structure Really Influence Educational Attainment?" *Social Science Research* 28 (4): 331–57. https://doi.org/10.1006/ssre.1999.0648.