

DEPARTMENT OF ECONOMICS

WORKING PAPER

2005



Department of Economics
Tufts University
Medford, MA 02155
(617) 627 – 3560
<http://ase.tufts.edu/econ>

Understanding Divergent Views on Redistribution Policy in the United States

by

Louise C. Keely and Chih Ming Tan[#]

This version: June 7, 2005

Abstract

Particular demographic groups are often associated with distinct points of view across various dimensions of redistribution policy. In this paper, we investigate which demographic groups account for heterogeneity in views on welfare policy and views on appropriate levels of overall redistribution. Using data from the General Social Survey and classification tools, we find evidence that classifications of the population by race, socioeconomic status, and age have some predictive power. However, much heterogeneity in views on redistribution policy persists even within these demographic groupings and remains unexplained. Our results suggest that identity-based explanations for variations in these views have to be interpreted with caution.

Keywords: Data mining, classification and regression trees, random forests, redistribution preferences, welfare, identity

JEL Classifications: C45, C49, H50, H53

[#] Keely: Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706. E-mail: lkeely@ssc.wisc.edu. Tan (corresponding author): Department of Economics, Tufts University, 8 Upper Campus Road, Medford, MA 02155. We thank Yannis Ioannides for comments on an earlier draft and Karl Scholz for discussion on the topic.

I. Introduction

A majority of Americans believe that too little is spent in the United States on assisting the poor. Further, a larger percentage of Americans think the government should redistribute income from the rich to the poor than the percentage who think the government should not redistribute. However, it is also the case that a majority (or near-majority) of Americans think too much is spent on welfare. Data supporting these claims from the General Social Survey (GSS) is summarized in Figures 1-3.

Opposition to welfare in the United States does not appear to be driven solely by ideological opposition to helping the poor. Rather, a substantial portion of Americans appears discontented with welfare's implementation as a policy intended to assist the poor. The term 'welfare' is not neutral in the United States. Its connotations go far beyond that of providing income assistance to the poor and unemployed. Welfare has been associated, justly or not, with programs that have distortionary effects such as eliminating the incentive to work and providing incentive to single mothers to have more children.

At the beginning of his term as president, Clinton promised to 'end welfare as we know it'. The legislatively-induced reform of welfare that followed was embodied in the act passed by Congress and signed by Clinton¹, the 'Personal Responsibility and Work Opportunity Reconciliation Act of 1996'. The reform reacted to but also appears to have temporarily reinforced negative sentiments regarding welfare in the years immediately preceding the act's passage; see Figures 1-3.

¹ For research on the national and state-level reforms see <http://www.ssc.wisc.edu/irp/welreform/home.htm>.

Views on assisting the poor generally, and welfare policy specifically, are often classified according to demographic variables by scholars and policy makers alike². Our study's purpose is to assess whether demographic variables are useful predictors of views on welfare policy in the United States over the past two decades and views on helping the poor more generally. The questions we pursue are as follows. Is opposition to welfare in the United States lined up with other views on redistribution, such as opposition to a public role in redistribution or ideological opposition to assisting the poor generally? Are there readily identifiable demographic groups who can be classified as having particular joint views³?

We investigate these questions in this paper by classifying these sets of views, and their interaction, across demographic groups, using thirteen demographic variables. The two classification tools we use, classification and regression trees (CARTTM) and random forests (RandomForestsTM), are described below. These statistical learning methods, which are widely used in other disciplines, provide insight into these views that linear regression could easily miss.

Understanding to what extent these views can be classified according to demographic groups is interesting not only as a matter of abstract intellectual inquiry, but also for policy-making. When programs for the poor, including welfare, are changed in scope or content, can we predict which readily identifiable demographic groups would be in favor of or opposed to such changes? If policy makers misclassify views according to demographic groups, and incorrectly predict the characteristics of those who will support or oppose a policy, they cannot determine overall levels of support for the policy nor appropriately design policy that would receive majority support.

² See, for instance, Edlund and Pande (2002) and Luttmer (2001), who focus on gender and race, respectively.

³ Our main concern in this study is with classification of joint responses to questions on income redistribution policy. We consider individual responses in Keely and Tan (2005).

Our results have some surprising implications. We find that salient demographic groupings in terms of responses to these questions correspond only to combinations of race, socioeconomic background, and age. Other variables considered, and described below, are not found to have important predictive power. Blacks are classified as being in favor of both increased spending on welfare and in assisting the poor generally, and in a public role in income redistribution. Whites of lower socioeconomic status are classified as in favor of a public role in income redistribution. They are classified particularly in support of increased welfare spending. They are, however, classified as opposed to increasing spending on assistance to the poor on the whole. Whites of higher socioeconomic status are classified as opposed to a public role in income redistribution. They are classified as supporting decreased spending both on welfare and assistance to the poor generally.

These classifications are not the end of the story, however. These classifications, and the population proportions associated with each, aggregate to the data presented in Figures 1-2. However we note that there remains substantial heterogeneity in views within each classification. The residual heterogeneity after classification suggests that these groupings provide only partial understanding of joint views on welfare, the public role in redistribution, and assistance to the poor generally. In public discussion and in scholarship, the typical classifications of the American constituency to predict their views on income redistribution include those according to race, gender, age, religious background, and socioeconomic status, and usually one at a time rather than in conjunction with each other. Our results suggest that these classifications – even when allowed to interact with each other in complex ways – do not go far in explaining the heterogeneity in these views.

In the next section the data and some exploratory summary statistics are discussed. In Section 3 we describe the CARTTM and RandomForestsTM methodologies. We present our results and interpretations in Section 4. Section 5 concludes.

II. The Data

The GSS contains the necessary data for our investigation. A variety of topics is covered in the survey, such as political activism, child-rearing practices, religious beliefs, and views on women's rights. Demographic variables such as the respondent's age, sex, income, and education are also collected. The samples are intended to be nationally representative of adults over 18.

The three questions on welfare and assistance to the poor that we use are:

1. NATFARE: *We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount... Welfare... are we spending too much, too little, or about the right amount on welfare? (1=too little, 2=about right, 3=too much)*

2. NATFAREY: *We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount... Assistance to the poor... are we spending too much, too little, or about the right amount on welfare? (1=too little, 2=about right, 3=too much)*

3. *EQWLTH: Some people think that the government in Washington ought to reduce the income differences between the rich and the poor, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor. Others think that the government should not concern itself with reducing the income difference between the rich and the poor. Here is a card with a scale from 1 to 7. Think of a score of 1 as meaning that the government ought to reduce the income differences between the rich and poor, and a score of 7 meaning that the government should not concern itself with reducing income differences. What score between 1 and 7 comes closest to the way you feel?*

Unfortunately, there is no overlap in the samples between respondents asked NATFARE and NATFAREY. There is, however, overlap in respondents asked each of those two questions and EQWLTH. We exploit these overlaps.

In order to clarify the correspondence in views on welfare and the poor, we use the following schematic representation. There are twenty-one possible pairs of responses to EQWLTH and one of the other two questions above, NATFARE or NATFAREY. We represent these respective 21 responses in Figures 4 and 5. We will refer to the joint EQWLTH-NATFARE variable as our “public redistribution and welfare” variable, and the joint EQWLTH-NATFAREY variable as our “public redistribution and assistance to the poor” variable. Our goal is to identify the most important predictive demographic characteristics of each group and contrast them with one another.

The relative overall sizes of these groups are represented using the relative frequency matrices in Figures 6 and 7 that contain unconditional sample proportions of each of the twenty-one groups described in Figures 4 and 5. These matrices are presented for the United States for all years of the sample⁴. In Figures 6 and 7 we also present the same information

⁴ The years include waves between 1978 and 2000.

but from only the year 1980 or 1984 and 1990 or 2000 in order to consider the constancy of the matrices over time. We find little variation.

These matrices reveal a complex and fascinating set of views among the American public. For instance, about 20% of Americans think that we spend too much on welfare, but that the government should take some role in redistributing income from the rich to the poor. Conditional on thinking the government should take some role in redistribution; about 40% of respondents think too much is spent on welfare.

About 35% of Americans think too little is spent on assisting the poor overall and that the government should take some role in redistributing income from the rich to the poor. Conditional on thinking the government should take some role in redistribution; about 75% of Americans think too little is spent on assisting the poor.

But about 15% of Americans think that little or no government action should be taken in redistributing income and that too little is spent on assisting the poor. Conditional on thinking that little or no government redistribution should be undertaken, about half of respondents also think that too little is spent on assisting the poor.

We consider a wide range of demographic characteristics, or identity markers, and let the data decide which dimensions are important for classifying views on public redistribution and welfare, and those on public redistribution and assistance to the poor, in ways we make precise below. We focus on characteristics that are exogenous; they are not choice variables of the respondent.⁵ This strategy avoids traditional endogeneity concerns and potential bias due to measurement error when using a subjective outcome as an independent variable (see Hamermesh (2004)).

⁵ We use the same variables as those in a related study, Keely and Tan (2005).

The identity markers are the respondent’s age in years (AGE); his gender (SEX); his self-reported race⁶ (RACE); the region of the US in which he was living at 16 (REG16); whether the respondent was born in the US (BORN); whether the respondent’s parents were born in the US (PARBORN); the respondent’s mother’s educational degree (MADEG; as a proxy for socioeconomic background); what religion in which the respondent was raised (RELIG16); and the respondent’s description of his religious upbringing as fundamentalist, moderate, or liberal (FUND16). A trend variable is also included (YEAR). These variables are detailed in the Appendix.

One endogenous variable we will also consider is the respondent’s real household income; we conduct our empirical examination with and without this variable. Our justification for including income is that it is such an obvious possible classifier of these views that we must check if it trumps the other variables, or how it interacts with other variables to classify such views. We do not, however, claim any causal inference from such analysis.

III. Classification and Regression Tree Methods

The main tool we use in the empirical analysis is classification and regression trees (CARTTM).

Formally, let $y \in Y$ denote an outcome variable of interest that takes on K categorical values $\{y_1, \dots, y_K\}$ and let $x \in X$ be a vector of M identity markers (which might be discrete or continuous variables or a mixture of both). We model the population of

⁶ This question asks the respondent to identify himself as white, black, or other. While we would have preferred a question with more ethnic detail, this is the best question that the GSS offers over many waves.

individuals as being classified by their identity markers into an unknown number b of subpopulations indexed by j . Within each subpopulation j , individuals are expected to return a response of y_j^* for the outcome variable of interest.

The classification of individuals into identity subgroups corresponds to the partitioning of the support of identity markers, X , into b partitions, $\Lambda = \{A_j\}_{j=1}^b$. The partitions A_j are mutually exclusive and their union is X . That is, $A_j \cap A_l = \emptyset$ and

$$\bigcup_{j=1}^b A_j = X.$$

For example, suppose y is a measure of views on (jointly) public redistribution and welfare, and $x = (\text{Race}, \text{Sex})$ where Race takes on values $\{B, W\}$ and Sex takes on values $\{M, F\}$. Then, a possible set of identity partitions, $\Lambda = \{A_1, A_2, A_3\}$, is $\{(BF, BM), (WM), (WF)\}$ with corresponding expected responses $\{y_B^*, y_{WM}^*, y_{WF}^*\}$. That is, in this example, if this were the set of identity groupings that we uncovered in the data, we would conclude that joint views on public redistribution and welfare differ systematically across subgroups in the population depending on whether respondents are black, white-male, or white-female. Our interest is in uncovering the identity partitions that characterize the data, as well as to estimate the predicted assignments of categorical outcome responses to each identity subgroup.

We now briefly describe the CARTTM algorithm that we employ to uncover the above. We refer the reader to Breiman, Friedman, Olsen, and Stone (1984)⁷ for further details on classification and regression tree methods.

⁷ We use the CARTTM software available from Salford Systems (<http://www.salfordsystems.com>).

CARTTM delivers a set of identity partitions by carrying out essentially two algorithms: (1) *recursive binary splitting* of the set of all observations, and (2) *cost complexity pruning* to address over-fitting. The recursive binary splitting algorithm starts with the set of all observations. It then classifies the observations into two subsequent sub-samples by exhaustively searching⁸ across the support points of all split variables (i.e., identity markers in our case) so as to find a split point that minimizes the joint node impurity across the two sub-samples. That is, the algorithm attempts to locate the split variable (i.e., identity marker) and associated split value (i.e., value for that identity marker) that produces the largest decrease in diversity in the outcome responses within each sub-sample.

Formally, for any partitioning, A_m , of the observations based on identity makers, let the proportion of y_k responses be given by $\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in A_m} I(y_i = y_k)$. Let Q_m be a measure of misclassification of responses (i.e., impurity) within this partition. For instance, the commonly used Gini index would be $Q_m^{GINI} = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk})$. The Gini index can be interpreted by noting that if we relabeled the responses as 1 for observations that yielded y_k and 0 otherwise, the variance in the partition A_m of this binary response is given by $\hat{p}_{mk} (1 - \hat{p}_{mk})$. Summing across all possible responses gives us the Gini index. That is, the Gini index is a variance estimate based on comparisons of all possible responses in a subgroup. An alternative impurity measure, the Twoing index (see Breiman et. al. (1984)) treats the k responses problem as if it were a binary response problem. It has been found

⁸ Loh and Shih (1997) point out that there may be variable selection bias towards identity markers which take on more values in CARTTM's exhaustive search algorithm. To get around this problem, we impose a penalty on high categorical variables in CARTTM. We calibrate the penalty to ensure that categorical variables have no inherent advantage in being selected for splitting over a continuous variable with unique values for each observation.

that Twoing tends to give considerably better prediction performance than Gini when the dependent variable is a higher-level categorical variable (i.e., with 10 or more categories). We therefore emphasize results which employ the Twoing index as the impurity measure in Section 4, but note that we find no substantive differences using the Gini index (unreported results).

CARTTM takes the set of all observations and partitions them into two sub-samples – the *Left* and *Right* nodes – by choosing an identity marker, j , and a corresponding value, s , in the support of j so as to minimize the joint impurity across the two sub-samples; i.e., $\min_{j,s}(\mathcal{Q}_L(j,s) + \mathcal{Q}_R(j,s))$. This process is then repeated iteratively on each of the subsequent sub-samples, and so on, until the number of observations in each sub-sample is too small for further splitting to occur.

The result of the recursive binary splitting algorithm is a full set of partitions of the original sample or “tree”. In order to avoid over-fitting, this tree is then “pruned”. Essentially, the pruning algorithm locates the (nested) subset of partitions within the full set of partitions that minimizes a generalized information criterion where the complexity penalty parameter is chosen by V-fold cross-validation⁹. The final set of partitions (the “pruned” tree) is then reported by CARTTM. To be clear, the end result of CARTTM is to deliver a set of homogeneous groupings of outcome responses and a pattern of identity partitions that characterizes these groupings, subject to not over-fitting the data.

In order to assess the validity of our CARTTM tree results (in terms of prediction error), we compare them with those obtained using Breiman’s (2001) RandomForestsTM (RF) algorithm. RF is an adaptive classification method which combines bootstrap aggregation (“bagging”) with pooling information from a multiplicity of randomly built trees to obtain

⁹ In our exercises, we set $V = 10$.

classifications of the outcome responses with lower mean prediction error compared to CARTTM. In fact, Breiman (2001) has shown that the prediction performance of RF is currently unmatched beating other leading adaptive learning methods like boosting. However, because RF pools information from a multiplicity of (randomly generated) trees, the results lack the sort of structural interpretability that CARTTM is able to offer in the form of a tree diagram. Because the uncovering of such structure is a main goal of this paper, we limit RF's role to two aspects. RF does offer guidance on which identity markers are salient in the classification of outcome responses into groups; we wish to compare the identity markers found to be important by RF with those in our CARTTM tree results. Also, we want to see how much better RF does in terms of reducing mean prediction error when compared to CARTTM in order to assess the validity of the latter's results.

We now briefly describe the RF algorithm and state key results. We refer the reader to Breiman (2001)¹⁰ for further details on random forests methods and implementation. RF generates a multiplicity of trees, and then pools information from these trees to obtain the best classification of responses in the following way. First, RF obtains L bootstrap samples (with replacement) from the data. Then, for each bootstrap sample, one third is left aside ("out-of-bag") while two thirds are used to generate a tree (fully grown without pruning) using CARTTM. To generate each tree, RF randomly selects a subset of identity markers of fixed size $m < M$ from the set of all identity markers to be used as split variables. Therefore, as a result, an outcome response assignment is obtained for each observation in about one-third of the trees.

Each tree now "votes" for the final outcome assignment for each observation. That is, at the end of the L iterations, take j to be the outcome response that was most

¹⁰ We use the RandomForestsTM software available from Salford Systems (<http://www.salfordsystems.com>).

frequently assigned to observation n when it was “out-of-bag”. This is then the RF predicted classification for that observation. In this way, each observation in the original sample is classified as corresponding to a particular outcome response depending on the modal classification accorded to it by the L trees. The “out-of-bag” misclassification estimate is then the proportion of times that j is not equal to the actual outcome response of observation n given by the data averaged over all observations. Breiman (2001) shows that this misclassification estimate is unbiased.

Finally, RF obtains a measure of variable importance for each identity marker by randomly permuting the values of each particular identity marker for the “out-of-bag” observations and then classifying these scrambled observations using the “in-bag” trees. RF defines the importance score for each identity marker as the average difference between the number of votes for the correct (i.e., observed) outcome response in the permuted “out-of-bag” data from the number of votes for the correct outcome response in the untouched “out of bag” data across the L trees. The idea is simple and compelling. If it is possible to substitute incorrect values for an identity marker and still obtain accurate predictions for outcome response classifications, then that identity marker cannot have been very important for classifying outcome responses in the first place.

IV. Results

The classification trees and random forest results are presented in Figures 9-16. The predicted classifications for both the EQWLTH-NATFARE variable (i.e., views on public redistribution and welfare) and the EQWLTH-NATFAREY variable (i.e., views on public redistribution and assistance to the poor) can range between 1 and 21 as defined in Figure 8.

Turning first to Figures 9-13, we see that the only key classification variable across all trees is race. Other key classification variables can, broadly speaking, be interpreted as indicative of socioeconomic status, and include age, household income, and mother's degree. As stated above, we conducted analysis with and without real household income. The following interpretation of the results incorporates both sets of trees and random forest results.

In general, the RF results affirm those of CARTTM. The error rates are similar, and the random forests' error rates do not decrease significantly with iteration. Moreover, the variables that RF identifies as most important generally reflect the splitting variables chosen by CARTTM. The error rates for random forests are 60% or higher, which is not surprising. This error rate should be compared to an error rate between predicted y and actual y in a linear regression context, which one would expect to be in the same sort of range. Nonetheless, given that the aim of the classification exercise is the identification of homogenous groupings, the remaining heterogeneity within groupings is important for our interpretation of the results.

Among the combinations of race, mother's degree, age, and income, the following groupings appear salient: non-whites¹¹, whites of low socioeconomic status, and whites of higher socioeconomic status.

Non-whites, particularly those self-defined as black, are overwhelmingly in favor of increasing assistance to the poor. They are more in favor of a government role in redistribution than whites, and also more strongly in favor of increasing spending on welfare.

¹¹ Depending on the tree, non-whites may include blacks only or those who self-classify as black or 'other'. Because the 'other' category is relatively small and ethnically heterogeneous, we place little emphasis on this classification in our interpretation. Instead, we focus on black-white differences that are robust. Further research into ethnic heterogeneity and views on income redistribution is necessary as data becomes available.

The joint views of whites are more complex. Overall, views on welfare change temporally. Across all white, the least support for increased welfare spending is evident at the end of the Carter administration and in the years immediately preceding the 1996 reform. Outside of those periods, younger whites are more pro-welfare than older whites. Views on a government role for redistribution are similar across whites, though those with the lowest socioeconomic status (MADEG=0) are more in favor than other whites (and not as much as blacks). Lowest socioeconomic status whites are also more in favor of increasing assistance to the poor, toward the levels of blacks. Among whites of other socioeconomic statuses (MADEG=1 through 4), older whites are less in favor of increasing assistance to the poor, and are indeed the least in favor of all demographic group classifications.

A broad summary of the predicted classifications for each group is provided in Figure 16. Non-whites are classified as in favor of a governmental role in redistribution as in favor of maintaining or increasing welfare spending, and in favor of increasing assistance to the poor. Whites of low socioeconomic status are classified as neutral to in favor of a governmental role in redistribution, neutral to in favor of increased welfare spending, and range in classifications on assistance to the poor. Whites of higher socioeconomic status are classified as being neutral to opposed to a governmental role in redistribution, as opposed to increased welfare spending, and as opposed to increasing assistance to the poor.

A summary statistic presented above is that a majority of respondents think that too much is spent on welfare. The results suggest that this majority is concentrated among whites with higher socioeconomic status, as they generally take a cool view toward governmental income distribution intervention. However, as one can see from Figure 10, there is enormous heterogeneity in the responses within terminal nodes. For instance, while blacks are classified

as believing that too little is spent on welfare, within the non-white terminal node 31% of responses correspond to a belief that too much is spent on welfare.

The set of demographic variables used to classify responses is quite wide-ranging and, within that set of variables, race stands out as an important classifying variable (see Figure 14). Even so, significant variation in individual responses is left unexplained by the classification exercise.

Another summary statistic described in Section 2 is that a majority of respondents think that too little is spent on assisting the poor. However, all whites are classified with a view that too much is spent. Again, the key to understanding this paradox is in the heterogeneity of responses (see Figure 13). The racial split, although relatively important as a classification of responses, does not provide a definitive split between views on assistance to the poor. Although this delineation is insufficient, this variable provides as much guidance as any other of the fourteen variables, including household income.

V. Conclusion

In the United States there is continuing debate over redistribution policies. One of the puzzling features of views on such policies is that, on average, Americans favor increased assistance to the poor generally but favor decreased spending on welfare specifically. The observation triggers several questions. Is welfare as a specific public policy viewed as ineffective or perverse in its incentives, or is a public role in redistribution the source of opposition? Do some demographic groups differ in the sources of opposition to welfare? Are there some identifiable demographic groups who actually tend to favor increased spending on welfare?

In considering these questions, we exploit data from the GSS that allow us to examine the *joint* responses to questions on views toward a public role in redistribution and welfare specifically, as well as views toward a public role in redistribution and assistance to the poor generally. We have sought to classify these joint responses along fourteen variables that are identifying features of individual respondents, including several variables that are often used to classify constituent views on redistribution policy.

A broad summary of our results is as follows. Non-whites tend to favor an increase in income redistribution, including that provided by the government and via direct transfers. Whites of higher socioeconomic status tend to favor a decrease in income redistribution via public and private channels. Whites of low socioeconomic status do not tend to favor an overall increase in spending, but do support a public role in income redistribution.

Still, much of the variation in these views is not attributable to broad demographic classifications, as is usually assumed. In other words, the preferences and information that underpin these responses is not readily reduced to any set of demographic variables, or their combinations, that are often taken to be salient. For example, there is no evidence of a monolithic 'black' perspective since substantial heterogeneity in views remains even in the groupings we do uncover. Finally, we do not find evidence for other commonly accepted explanations, such as religion, for differences in such views. We find no evidence for a 'fundamentalist Protestant' view towards income redistribution policies, for instance.

References

Breiman, L., J. H. Friedman, R. A. Olsen, and C. J. Stone (1984), Classification and Regression Trees, Wadsworth, Belmont.

Breiman, L. (2001), "Random Forests", *Machine Learning*, 45(1), p. 5-32.

Edlund, L. and R. Pande (2002), "Why Have Women Become Left Wing? The Political Gender Gap Decline in Marriage", *Quarterly Journal of Economics*, 117 (3), p. 917-961.

Hamermesh, D. S. (2004), "Subjective Outcomes in Economics", *Southern Economic Journal*, 71(1), p. 2-11.

Keely, L. C. and C. M. Tan (2005) "Understanding Preferences for Income Redistribution", Tufts University, Dept. of Economics Working Paper No. 2005-11.

Levine, K. and J. K. Scholz (2001), "The Evolution of Income Support Policy", in Understanding Poverty, Danziger and Havemen (Eds.), Harvard University Press and Russell Sage Foundation, 2001, p. 193-228.

Loh, W.-Y. and Shih, Y.-S. (1997), "Split Selection Methods for Classification Trees," *Statistica Sinica*, vol. 7, p. 815-840.

Luttmer, E. F. (2001), "Group Loyalty and the Taste for Redistribution", *Journal of Political Economy*, 109 (3), p. 500-528.

Appendix: Identity Markers

Here identity variables are detailed where their description in the text is incomplete. Those variables are SEX, RACE, REG16, BORN, PARBORN, MADEG, RELIG16, and FUND16.

1. **SEX** (1 = Male, 2 = Female)
2. **RACE:** *What race would you consider yourself?* (Recorded verbatim and coded) (1 = White, 2 = Black, 3 = Other)
3. **REG16:** *In what state or foreign country were you living when you were 16 years old?* (Coded by region) (1 = New England, 2 = Middle Atlantic, 3 = East North Central, 4 = West North Central, 5 = South Atlantic, 6 = East South Central, 7 = West South Central, 8 = Mountain, 9 = Pacific, 10 = Foreign)

New England = Maine, Vermont, New Hampshire, Connecticut, Rhode Island, Massachusetts

Middle Atlantic = New York, New Jersey, Pennsylvania

East North Central = Wisconsin, Indiana, Ohio, Illinois, Michigan

West North Central = Minnesota, Iowa, Missouri, North Dakota, South Dakota, Missouri, Kansas

South Atlantic = Delaware, Maryland, West Virginia, Virginia, North Carolina, South Carolina, Georgia, Florida, District of Columbia

East South Central = Kentucky, Tennessee, Alabama, Mississippi

West South Central = Arkansas, Oklahoma, Louisiana, Texas

Mountain = Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico

Pacific = Washington, Oregon, California, Alaska, Hawaii

4. **BORN:** *Were you born in this country?* (1 = Yes, 2 = No; don't know responses were treated as missing values)
5. **PARBORN:** *Were both of your parents born in this country?* (1 = Both born in the US, 2 = One born in the US, 3 = Neither born in the US; don't know responses were treated as missing values)
6. **MADEG:** Respondent's mother's education (Recorded by GSS from a set of questions regarding years of schooling and degrees attained) (0 = Less than high school, 1 = high school, 2 = Associate/junior college, 3 = Bachelor's, 4 = Graduate; don't know or NA responses treated as missing values)
7. **RELIG16:** *In what religion were you raised?* (1 = Protestant, 2 = Catholic, 3 = Jewish, 4 = None, 5 = Other)
8. **FUND16:** Fundamentalism/Liberalism of religion respondent raised in. (1 = Fundamentalist, 2 = Moderate, 3 = Liberal)

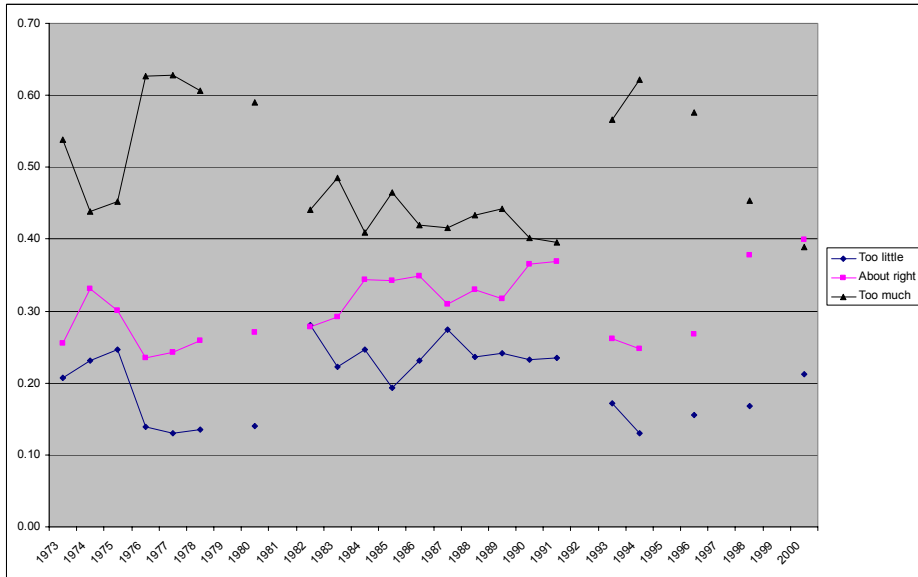


Figure 1: GSS NATFARE 1973-2000
Spending on welfare?

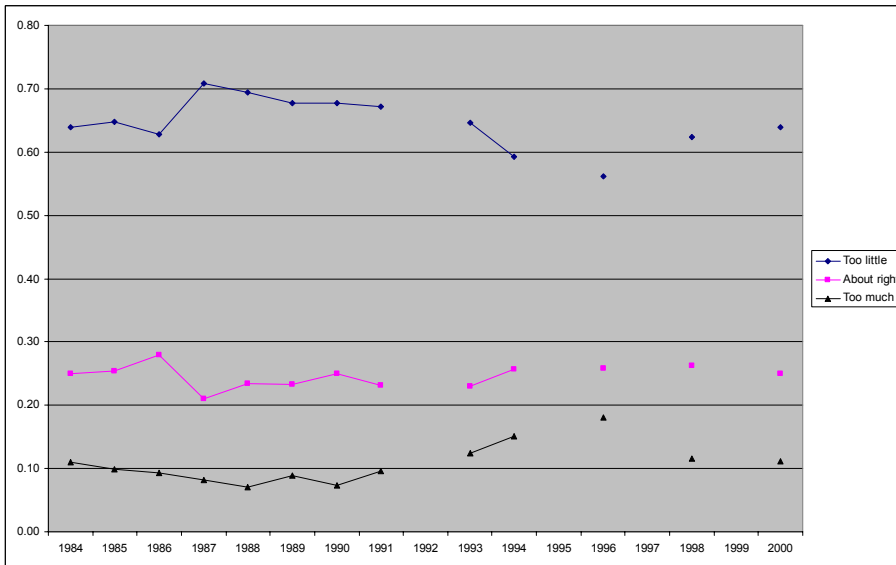


Figure 2: GSS NATFAREY 1984-2000
Spending on assisting the poor?

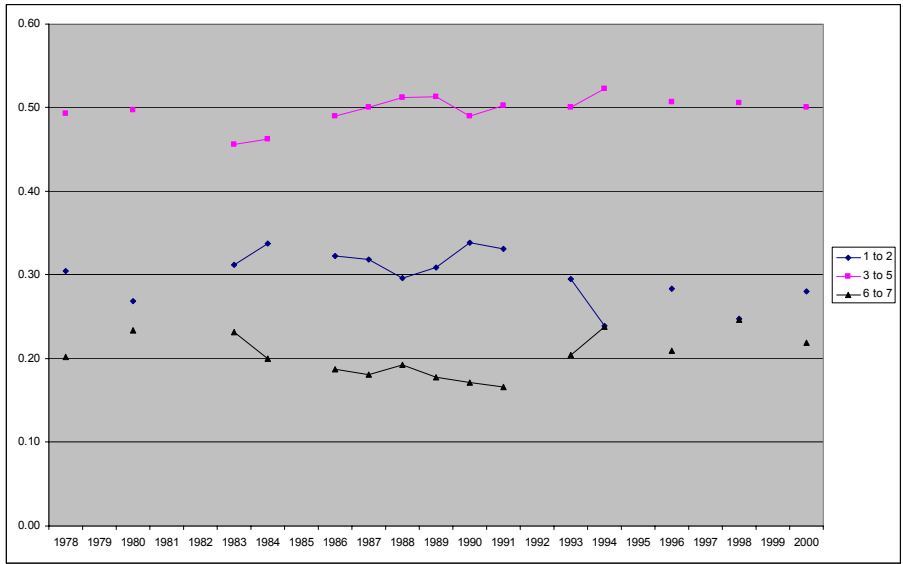


Figure 3: GSS EQWLTH 1978-2000
 Government redistribution from rich to poor? (1= Take action, 7= No action)

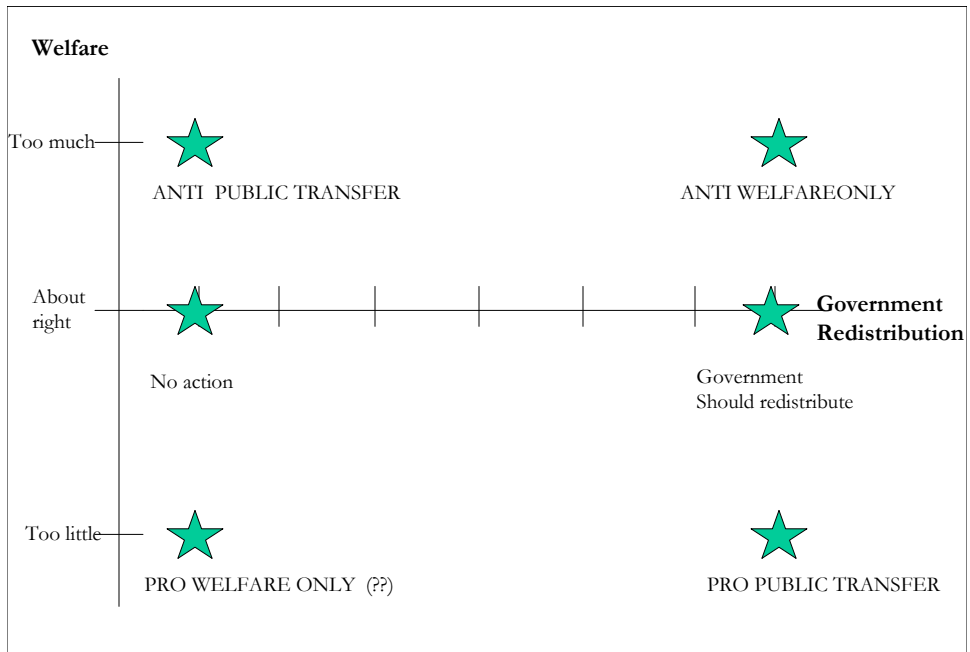


Figure 4: Views on Welfare v. Views on Government Redistribution from Rich to Poor

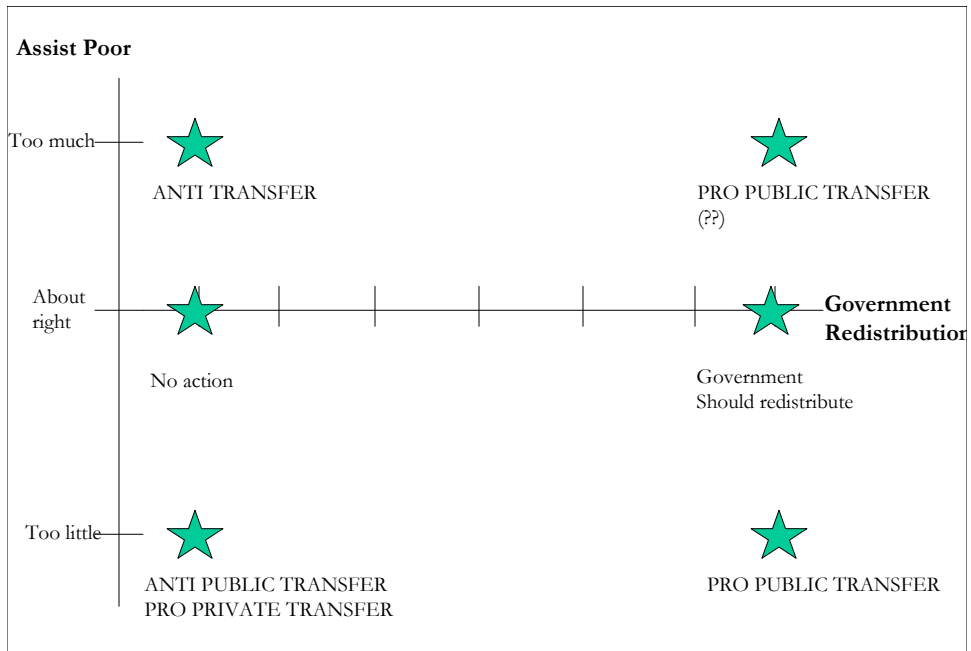


Figure 5: Views on Assisting the Poor v. Views on Government Redistribution from Rich to Poor

ALL YEARS				
EQWLTH	NATFARE			Total
	too little	about right	too much	
Gov't reduce diff	6%	5%	7%	18%
2	3%	3%	5%	11%
3	3%	6%	7%	17%
4	3%	7%	10%	20%
5	2%	4%	7%	12%
6	1%	2%	5%	8%
No Gov't action	1%	3%	10%	14%
Total	19%	31%	50%	

Year 1980				
EQWLTH	NATFARE			Total
	too little	about right	too much	
Gov't reduce diff	4%	4%	9%	17%
2	1%	3%	5%	10%
3	2%	6%	9%	17%
4	3%	6%	11%	20%
5	1%	3%	8%	13%
6	0%	2%	5%	8%
No Gov't action	1%	3%	12%	16%
Total	14%	27%	59%	

Year 2000				
EQWLTH	NATFARE			Total
	too little	about right	too much	
Gov't reduce diff	5%	5%	4%	14%
2	3%	5%	4%	12%
3	4%	7%	5%	16%
4	4%	10%	7%	21%
5	3%	6%	5%	14%
6	1%	3%	4%	8%
No Gov't action	1%	5%	8%	15%
Total	21%	41%	38%	

Year 1990				
EQWLTH	NATFARE			Total
	too little	about right	too much	
Gov't reduce diff	6%	7%	7%	20%
2	4%	4%	7%	14%
3	5%	7%	6%	17%
4	5%	10%	7%	22%
5	2%	2%	4%	8%
6	2%	1%	3%	6%
No Gov't action	2%	3%	8%	12%
Total	25%	34%	41%	

Figure 6 Relative frequency matrices: EQWLTH and NATFARE

All years				
EQWLTH	NATFAREY			Total
	too little	about right	too much	
Gov't reduce diff	15%	3%	1%	19%
2	8%	2%	1%	11%
3	12%	4%	1%	17%
4	13%	6%	2%	21%
5	7%	4%	1%	13%
6	3%	3%	1%	7%
No Gov't action	5%	4%	3%	13%
Total	63%	26%	11%	

Year 1984				
EQWLTH	NATFAREY			Total
	too little	about right	too much	
Gov't reduce diff	12%	6%	1%	19%
2	7%	2%	1%	10%
3	8%	4%	2%	14%
4	11%	9%	2%	22%
5	8%	5%	1%	14%
6	3%	3%	1%	7%
No Gov't action	4%	6%	4%	14%
Total	53%	35%	12%	

Year 2000				
EQWLTH	NATFAREY			Total
	too little	about right	too much	
Gov't reduce diff	14%	3%	1%	18%
2	9%	2%	1%	12%
3	12%	3%	1%	16%
4	11%	5%	3%	19%
5	7%	6%	1%	14%
6	4%	3%	1%	9%
No Gov't action	3%	5%	3%	11%
Total	62%	27%	11%	

Year 1990				
EQWLTH	NATFAREY			Total
	too little	about right	too much	
Gov't reduce diff	20%	2%	1%	23%
2	10%	3%	0%	13%
3	14%	3%	1%	18%
4	13%	5%	2%	20%
5	6%	4%	0%	10%
6	4%	3%	0%	7%
No Gov't action	4%	3%	1%	9%
Total	70%	24%	6%	

Figure 7 Relative frequency matrices: EQWLTH and NATFAREY

		<i>NATFARE(Y)</i>			
		<i>Too little</i>	<i>About right</i>	<i>Too much</i>	
<i>EQWLTH</i>			<i>1</i>	<i>2</i>	<i>3</i>
<i>Government should have role</i>	<i>1</i>	1	8	15	
	<i>2</i>	2	9	16	
	<i>3</i>	3	10	17	
	<i>4</i>	4	11	18	
	<i>5</i>	5	12	19	
	<i>6</i>	6	13	20	
<i>Government shouldn't have role</i>	<i>7</i>	7	14	21	

Figure 8: Coding of intersection variable between EQWLTH and NATFARE or NATFAREY

Classification Variable	A	B	C	D	E
Race	2,3	1	1	1	1
Year	All	78,80,93,94,96	83,84,86,87,88,89, 90,91,98,00	83,84,86,87,88,89, 90,91,98,00	83,84,86,87,88,89, 90,91,98,00
Age	All	All	>45	<46	<46
Reg16	All	All	All	2,7	0,1,3,6,8,9
Predicated Class	1	19	14	6	3

Figure 9: Regression tree terminal nodes of EQWLTH and NATFARE tree

Terminal Node	A	Classification:	1
	NATFARE		
EQWLTH	too little	about right	too much
Gov't reduce diff	0.14	0.09	0.08
2	0.05	0.04	0.04
3	0.05	0.06	0.05
4	0.07	0.07	0.07
5	0.04	0.02	0.03
6	0.01	0.01	0.02
No Gov't action	0.02	0.02	0.03
Total	0.37	0.31	0.32

Terminal Node	B	Classification:	19
	NATFARE		
EQWLTH	too little	about right	too much
Gov't reduce diff	0.03	0.03	0.08
2	0.02	0.03	0.05
3	0.02	0.05	0.09
4	0.02	0.06	0.13
5	0.01	0.04	0.09
6	0.01	0.02	0.06
No Gov't action	0.01	0.03	0.13
Total	0.11	0.26	0.64

Terminal Node	C	Classification:	14
	NATFARE		
EQWLTH	too little	about right	too much
Gov't reduce diff	0.04	0.06	0.07
2	0.02	0.04	0.03
3	0.02	0.07	0.06
4	0.03	0.10	0.10
5	0.01	0.05	0.06
6	0.01	0.03	0.05
No Gov't action	0.01	0.05	0.10
Total	0.16	0.38	0.46

Terminal Node	D	Classification:	6
	NATFARE		
EQWLTH	too little	about right	too much
Gov't reduce diff	0.04	0.04	0.07
2	0.03	0.03	0.03
3	0.03	0.08	0.08
4	0.05	0.08	0.09
5	0.01	0.04	0.06
6	0.02	0.04	0.05
No Gov't action	0.01	0.02	0.10
Total	0.18	0.33	0.49

Terminal Node	E	Classification:	3
	NATFARE		
EQWLTH	too little	about right	too much
Gov't reduce diff	0.05	0.04	0.06
2	0.03	0.03	0.05
3	0.05	0.08	0.07
4	0.04	0.06	0.07
5	0.03	0.05	0.06
6	0.01	0.02	0.06
No Gov't action	0.02	0.03	0.09
Total	0.23	0.31	0.46

Figure 10: Relative frequencies of responses within tree terminal nodes of EQWLTH and NATFARE tree

Classification Variable											
Race	2,3	1	1	1	1	1	1	1	1	1	1
Realinc	All	<19780	<19780	>19779	>19779, <49906	>19779, <49906	>19779, <49906	>19779, <49906	>19779, <49906	>19779, <49906	>49905
Age	All	<60	>59	All	>45	<46	<46	<46	<46	<46	All
Year	All	All	All	78,80,94, 96	83,84,86, 87,88,89, 90,91,93	83,84,86, 87,88,89, 90,91,93	83,84,87, ,89,91	86,88,90, ,93	83,84,86, ,87,88,8 9,90,91, 93	83,84,86, ,87,88,8 9,90,91, 93	83,84,86, ,87,88,8 9,90,91, 93
Reg16	All	All	All	All	All	1,7	2,3,5,9	2,3,5,9	0,4,6,8	0,2,3,4,5 ,6,8,9	All
Madeg	All	All	All	All	All	All	1,2,3	1,2,3	1,2,3	0,4	All
Predicted Class	1	2	14	20	14	6	13	6	12	10	20

Figure 11: Regression tree terminal nodes of EQWLTH and NATFARE tree with income as explanatory variable

Without REALINC as explanatory variable

Classification Variable	A	B	C	D	E
Race	2	1,3	1,3	1,3	1,3
Madeg	All	0	1-4	1-4	1-4
Year	All	All	86-91, 93,98,00	84,94,96	84,94,96
Age	All	All	All	<40	>39
Predicted Class	1	15	13	19	20

With REALINC as explanatory variable

Classification Variable		
Realinc	<19780	>19780
Predicted Class	15	20

Figure 12: Regression trees terminal nodes of EQWLTH and NATFAREY tree

WORKING PAPER SERIES 2005

<http://ase.tufts.edu/econ/papers/papers.html>

- 2005-01** EGGLESTON, Karen, Keqin RAO and Jian WANG; “From Plan to Market in the Health Sector? China's Experience.”
- 2005-02** SHIMSHACK Jay; “Are Mercury Advisories Effective? Information, Education, and Fish Consumption.”
- 2005-03** KIM, Henry and Jinill KIM; “Welfare Effects of Tax Policy in Open Economies: Stabilization and Cooperation.”
- 2005-04** KIM, Henry, Jinill KIM and Robert KOLLMANN; “Applying Perturbation Methods to Incomplete Market Models with Exogenous Borrowing Constraints.”
- 2005-05** KIM, Henry, Jinill KIM, Ernst SCHAUMBURG and Christopher A. SIMS; “Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models.”
- 2005-06** KIM, Henry, Soyoung KIM and Yunjong WANG; “International Capital Flows and Boom-Bust Cycles in the Asia Pacific Region.”
- 2005-07** KIM, Henry, Soyoung KIM and Yunjong WANG; “Fear of Floating in East Asia.”
- 2005-08** SCHMIDHEINY, Kurt; “How Fiscal Decentralization Flattens Progressive Taxes.”
- 2005-09** SCHMIDHEINY, Kurt; “Segregation from Local Income Taxation When Households Differ in Both Preferences and Incomes.”

- 2005-10** DURLAUF, Steven N., Andros KOURTELLOS, and Chih Ming TAN; “How Robust Are the Linkages between Religiosity and Economic Growth?”
- 2005-11** KEELY, Louise C. and Chih Ming TAN; “Understanding Preferences For Income Redistribution.”
- 2005-12** TAN, Chih Ming; “No One True Path: Uncovering the Interplay between Geography, Institutions, and Fractionalization in Economic Development.”
- 2005-13** IOANNIDES, Yannis and Esteban ROSSI-HANSBERG; “Urban Growth.”
- 2005-14** PATERSON, Robert W. and Jeffrey E. ZABEL; “The Effects of Critical Habitat Designation on Housing Supply: An Analysis of California Housing Construction Activity.”
- 2005-15** KEELY, Louise C. and Chih Ming TAN; “Understanding Divergent Views on Redistribution Policy in the United States.”
- 2005-16** DOWNES, Tom and Shane GREENSTEIN; “Understanding Why Universal Service Obligations May Be Unnecessary: The Private Development of Local Internet Access Markets.”
- 2005-17** CALVO-ARMENGOL, Antoni and Yannis M. IOANNIDES; “Social Networks in Labor Markets.”
- 2005-18** IOANNIDES, Yannis M.; “Random Graphs and Social Networks: An Economics Perspective.”
- 2005-19** METCALF, Gilbert E.; “Tax Reform and Environmental Taxation.”
- 2005-20** DURLAUF, Steven N., Andros KOURTELLOS, and Chih Ming TAN; “Empirics of Growth and Development.”
- 2005-21** IOANNIDES, Yannis M. and Adriaan R. SOETEVENT; “Social Networking and Individual Outcomes Beyond the Mean Field Case.”

- 2005-22** CHISHOLM, Darlene and George NORMAN; “When to Exit a Product: Evidence from the U.S. Motion-Pictures Exhibition Market.”
- 2005-23** CHISHOLM, Darlene C., Margaret S. McMILLAN and George NORMAN; “Product Differentiation and Film Programming Choice: Do First-Run Movie Theatres Show the Same Films?”
- 2005-24** METCALF, Gilbert E. and Jongsang PARK; “A Comment on the Role of Prices for Excludable Public Goods.”