ELSEVIER

Journal of Public Economics 92 (2008) 944–961

JOURNAL OF PUBLIC ECONOMICS

www.elsevier.com/locate/econbase

Understanding preferences for income redistribution ☆

Louise C. Keely [a], Chih Ming Tan [b,*]

[a] *Department of Economics, University of Wisconsin, United States*
[b] *Department of Economics, Tufts University, 8 Upper Campus Road, Medford, MA 02155, United States*

Received 11 July 2005; received in revised form 21 June 2007; accepted 8 November 2007
Available online 15 December 2007

Abstract

Recent research suggests that income redistribution preferences vary across identity groups. We employ statistical learning methods that emphasize pattern recognition; classification and regression trees (CART™) and random forests (RandomForests™), to uncover what these groups are. Using data from the General Social Survey, we find that, out of a large set of identity markers, only race, gender, age, and socioeconomic class are important classifiers for income redistribution preferences. Further, the uncovered identity groupings are characterized by complex patterns of interaction amongst these salient classifiers. We explore the extent to which existing theories of income redistribution can explain our results, but conclude that current approaches do not fully explain the findings.

JEL classification: C45; C49; H50; H53
Keywords: Data mining; Classification and regression trees; Random forests; Redistribution preferences; Identity

1. Introduction

The idea that an individual's identity potentially plays a key role in determining preferences that affect economic decisions and outcomes has gained increasing acceptance in the economics literature since the seminal work by Akerlof and Kranton (2000). This emerging literature should rightly be seen as an extension of the large body of work in sociology examining how people make sense of their world and how identity plays into their views of themselves and others (see, for instance, Lamont (2000)). In this paper, we focus our attention on one such preference — an individual's preferred level of income redistribution.

☆ We thank Jim Andreoni, Buz Brock, Steven Durlauf, Carol Graham, Yannis Ioannides, Wei-Yin Loh, Larry Samuelson and seminar participants at the London School of Economics, University of North Carolina-Chapel Hill, and the University of Wisconsin Institute for Research on Poverty for comments. We also thank two anonymous referees for their valuable suggestions and insights. We are grateful for funding under the Robock Award in Empirical Economics from the University of Wisconsin. Keely thanks WARF and the Brookings Institution for generous research support, and Brookings for their hospitality. Tan thanks the generous research support provided by the Program of Fellowships for Junior Scholars, MacArthur Research Network on Social Interactions and Economic Inequality. We greatly appreciate the excellence and diligence of our research assistant, Zhiguo Xiao.
* Corresponding author.
E-mail address: chihming.tan@tufts.edu (C.M. Tan).

0047-2727/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2007.11.006

Existing theoretical treatments of how an individual's identity determines her income redistribution preferences highlight various mechanisms. In *preference-based* theories, identity matters because people care, in an exogenous fashion, about the actions or outcomes of others in the same or across identity groups. The relevance of identity to economic decision-making is modeled via modifications to the preference structure. For instance, Alesina et al. (1999) and Alesina et al. (2001) both propose simple models which capture an individual's utility being dependent on the utilities of members of other ethnic groups. They conclude that this awareness of ethnic heterogeneity, or "racism" (Alesina et al. (2001)), could be responsible for the divergence in views on redistribution across groups.

In *information-based* theories, identity provides information about an individual's future economic circumstances in an environment with uncertainty. Identity groupings may correspond to a set of initial conditions that have persistent implications for income mobility (Benabou (1996) and Benabou and Ok (2001)) or for shaping mobility beliefs (Loury (1998) and Piketty (1995)) that in turn determine redistribution preferences. Loury (1998), for instance, argues that people are 'socially located' — they are part of social and cultural networks that exert strong influence on behavior. Behavior may be ex-post rational in that it is self-fulfilling and persistent. As a result, initial differences across groups can have long-run effects on outcomes such as income or preferences for income redistribution.

The aim of this paper is twofold. The first and primary objective of this paper is to propose a new way to explore the empirical implications of information-based theories using data from the General Social Survey (GSS). To do so, we exploit the fact that these theories imply a mapping between identity variables and various outcome variables that are in turn related to income redistribution preferences. These mappings imply restrictions on the data that we propose to investigate. We make clear the exact restrictions we are investigating and how we do so in Section 2.

The second objective of this paper is to collect a (broad) set of stylized facts regarding redistribution preferences. For instance, as noted above, preference-based theories tend to focus on ethnicity as the important identity marker for determining redistribution preferences. It is of interest, however, to ascertain whether there are other prominent dimensions of identity that matter. We also extend our analysis to investigate the relationship between American's views on helping the poor more generally and their views on welfare policy in the United States over the past two decades.

To address the above two objectives, and to uncover the role of identity in driving differences in redistribution preferences, we employ statistical learning methodologies that emphasize pattern recognition; i.e., classification and regression trees (CART™) and random forests (RandomForests™). These statistical learning methods, which are widely used in other disciplines, provide insight into these views that linear regression could easily miss.

Existing empirical work in this area typically employ the following strategy: (i) a linear relationship between redistribution preferences and other covariates is assumed, (ii) investigations of heterogeneity in redistribution preferences are carried out using pre-specified identity groups; for instance, the existing empirical literature tends to emphasize the (a priori) importance of race and gender, and (iii) typically, only a small number of alternative specifications (such as interactions between covariates, and polynomial terms) are explored before settling on a particular specification that is then reported.

For instance, in a work that is closest to ours in spirit, Alesina and La Ferrara (2005) employ a linear ordered probit model to study the determinants of redistribution preferences. They consider a large number of such determinants, including respondents' age, race, gender, socioeconomic class, etc., but do not report any results for possible interaction effects for these variables. On the other hand, Fong (2001), which explores support for redistribution, does consider the interaction effect between race and gender. However, her choice of interaction effect is made a priori; there is no systematic attempt to explore other possibilities.

An important consequence of the above empirical strategy is that, because nonlinearity and heterogeneity are not systematically investigated, this work in effect makes strong prior claims about the correct (econometric) model for preferences. Researchers essentially focus their attention on a small number of models out of the very large set of possible models that could have been generated if one considered nonlinearity as well as the full range of possible interactions between the covariates. However, as pointed out by Brock and Durlauf (2001) in the economic growth context, there are many instances in economics where theories about particular outcomes are "open-ended". By open-ended, Brock and Durlauf are referring to the possibility that in these instances, the fact that one theory (e.g., gender) may be important to redistribution preferences, does not automatically exclude some other theory (such as race or parental income or any of the many other possible alternative explanations for variations in such preference) from also being important. It also does not exclude the possibility that their interaction may be important.

The important point is that theory open-endedness implies model uncertainty, and therefore, any assessment of the likely effect of an explanatory variable on redistribution preferences should be made with the full universe of possible