Re-estimation and Extension of Barnes et al. (2018)

Information, Knowledge and Attitudes: The Taxpayer Receipt

Matthew Draper

Department of Political Science, University of California San Diego

Contact Information:
Department of Political Science
University of California San Diego
9500 Gilman Dr. La Jolla 92093

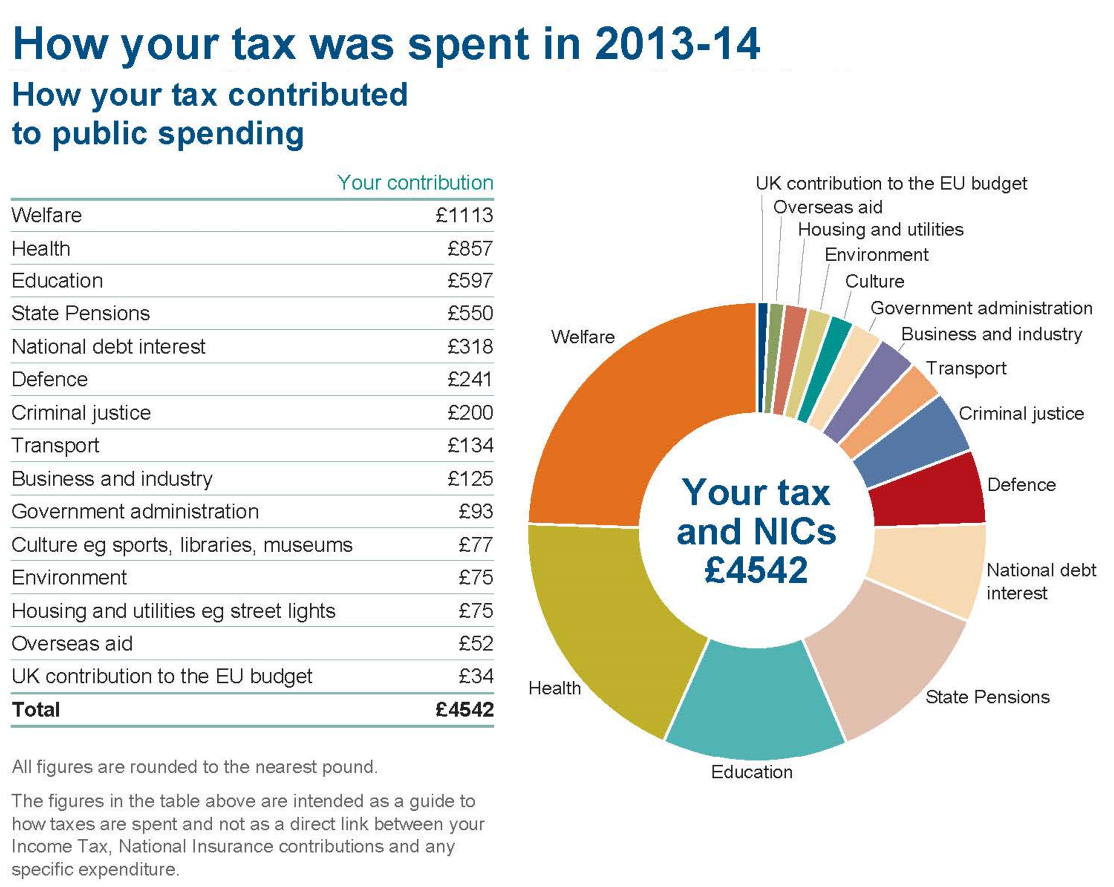
Phone: +1 (434) 327-3382
Email: mdraper@ucsd.edu

Abstract

This article re-estimates and extends published work on the impact of government-issued taxpayer receipts on political knowledge and political attitudes. Previous work had found that tax receipts can increase knowledge but have no effect on attitudes or preferences (Barnes et al. (*JoP* 2018)). After reproducing the authors' findings using the original survey data, I fit a cumulative logistic regression model in place of the authors' ordered logit, and use this cumulative logistic regression to test the parallel regressions assumption on which the authors' use of an ordered logit relied. Finding that this assumption is not satisfied, I fit a multinomial logistic regression in place of the authors' ordered logit. I find strong evidence to suggest that a multinomial logistic regression is a more appropriate model for the data-generating process studied in Barnes et al. (2018).

Introduction

Barnes et al. (*JoP* 2018) investigates whether the dissemination of government-issued 'taxpayer receipts' affects political knowledge and attitudes. They found that these receipts increased political knowledge, but had no effect on political attitudes or preferences. Indeed, '[c]itizens can learn, but we find no evidence that they change their minds as a result' (p.701).



How your tax was spent in 2013-14

How your tax contributed to public spending

	Your contribution
Welfare	£1113
Health	£857
Education	£597
State Pensions	£550
National debt interest	£318
Defence	£241
Criminal justice	£200
Transport	£134
Business and industry	£125
Government administration	£93
Culture eg sports, libraries, museums	£77
Environment	£75
Housing and utilities eg street lights	£75
Overseas aid	£52
UK contribution to the EU budget	£34
Total	**£4542**

All figures are rounded to the nearest pound.

The figures in the table above are intended as a guide to how taxes are spent and not as a direct link between your Income Tax, National Insurance contributions and any specific expenditure.

Figure 1: An example taxpayer receipt (2014)

1 Re-Estimation

Barnes et al. estimate treatment effects on knowledge acquisition via the following model (Model 2):

$$W_2K_i = \alpha + \tau T_i + \beta_0 W_1K_i + \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i \qquad (1)$$

where i indexes respondents, W_2K_i represents political knowledge at Wave 2, T_i is a dummy variable indicating assignment to treatment, $x_{ik}$is the kth covariate for individual i and W_1K_i is a control for Wave 1 knowledge level.

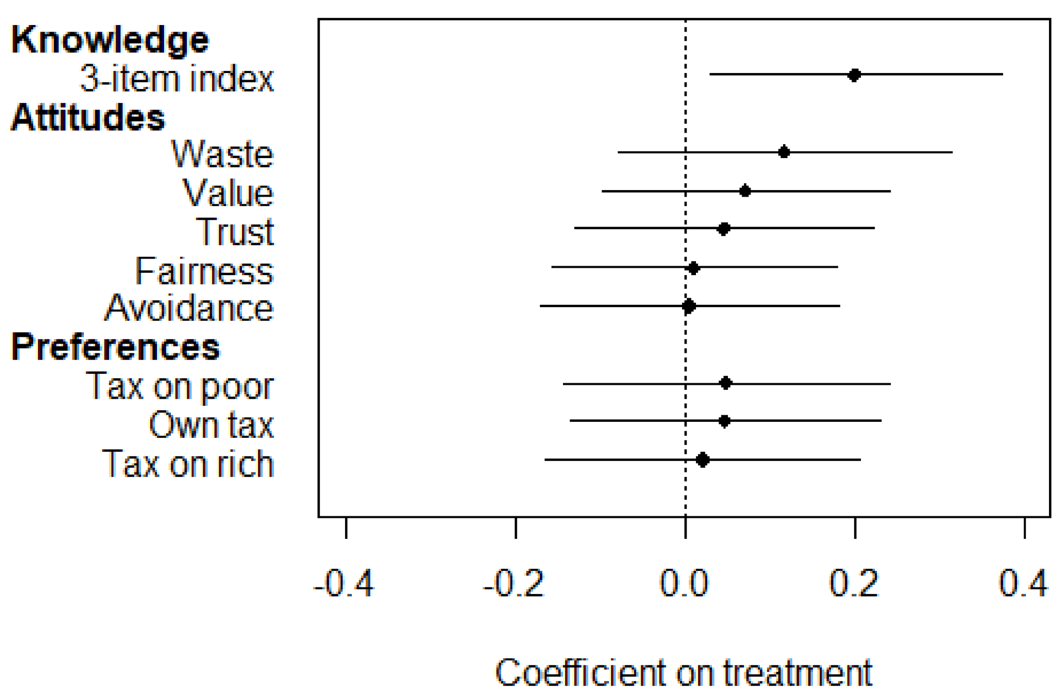


Figure 2: Replication of ordered logit models (Figure 3) in Barnes et al. (2018)

TABLE 2. Treatment effects on Wave 2 budget knowledge, ordered logit models.

	3-item knowledge index
Treatment	0.201**
	(0.087)
Age	0.037
	(0.038)
Female	−0.437***
	(0.091)
White	0.279**
	(0.132)
Conservative	0.159
	(0.118)
Labour	0.018
	(0.108)
Liberal Democrat	−0.013
	(0.173)
Working full time	0.053
	(0.130)
Education scale	0.915***
	(0.200)
Wave 1 knowledge	−0.356*
	(0.182)
Observations	2,072
Log Likelihood	-2,180
AIC	4,360

*p < .1; **p < .05; ***p < .01

Figure 3: Replication of the knowledge model (Table 5) in Barnes et al. (2018)

2 Extension

The ordered logit model constrains slope parameters β_m to be identical across covariates. Instead of conceptualizing the logit as a single model, we can instead see it as the constrained estimation of a system of models, because we can reexpress the dependent (ordered categorical) variable Y_i as a series of binary variables, $\tilde{Y}_{im}$, such that $\tilde{y}_{im} = 1 \Leftrightarrow y_i \leq m$ for some category, m. I fit a logit model for each of these $\tilde{Y}_m$, together comprising a cumulative logit model.

$$Pr(Y_i \leq M-1) = logit^{-1}(\tau_{M-1} + \mathbf{x}_\mathbf{i}^\mathbf{T} \beta_{M-1}) \qquad (2)$$

This assumption of common slope parameters across levels of the response variable is known as the parallel regressions assumption (PRA), which is easier to satisfy with fewer covariates but becomes more demanding as covariates increase in number. I use the unconstrained cumulative logit developed above to test the PRA in Model 2 (30 covariates). Observing deviations from linearity on multiple covariates, I find that the PRA is likely to be violated.

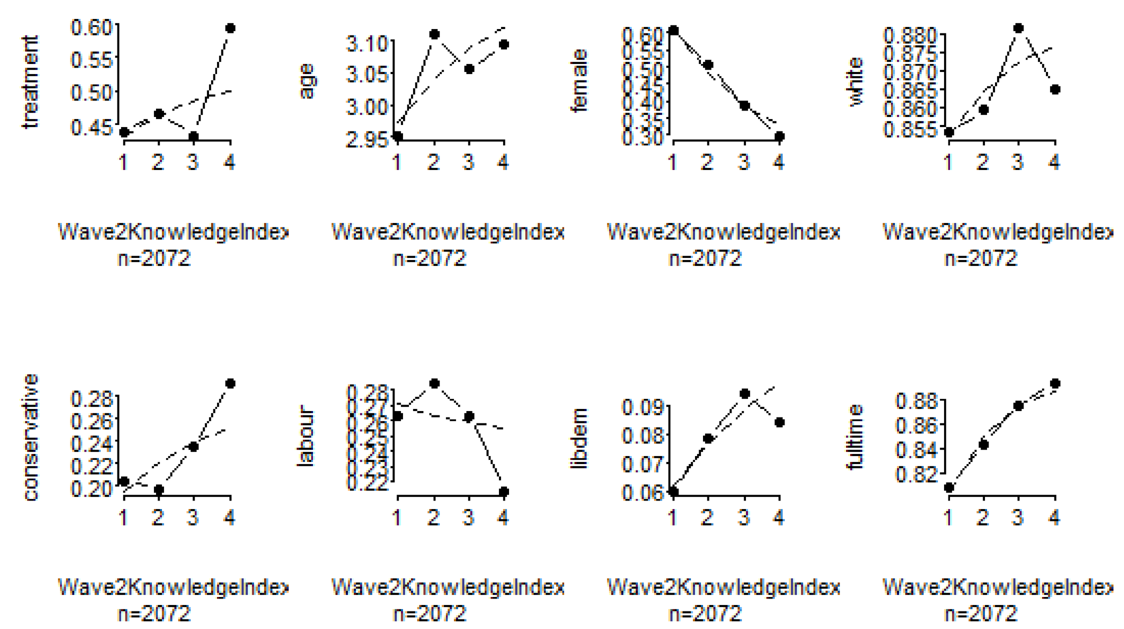


Figure 4: Plot of the conditional means of non-binary regressors at different levels of the response variable 'Wave 2 Knowledge Index'

2.1 Multinomial Logit

Since the parallel regressions assumption appears to be violated, I fit a multinomial logit in place of the ordered logit used in Barnes et al. (2018). The multinomial logit model is a generalization of the binomial distribution involving $M-1$ binary logits estimated simultaneously, with the probability constrained to sum to one. The influence of each independent variable will differ by outcome category. To make sure that probabilities will sum to 1 across the outcome categories, we must divide by the sum across all M categories, as shown here:

$$Pr(Y_i = m \mid \mathbf{x_i}) = \frac{exp(\mathbf{x}_\mathbf{i}^\mathbf{T} \beta_m)}{1 + \sum_{j=2}^{M} exp(\mathbf{x}_\mathbf{i}^\mathbf{T} \beta_j)} \qquad (3)$$

Results

The multinomial logit performs best at the highest and lowest values of the knowledge index. This indicates that the model is doing a good job of predicting the extremes but a poorer job of predicting middle categories, where performance is comparable to the ordered logit. A confusion matrix (not pictured) confirms these findings.

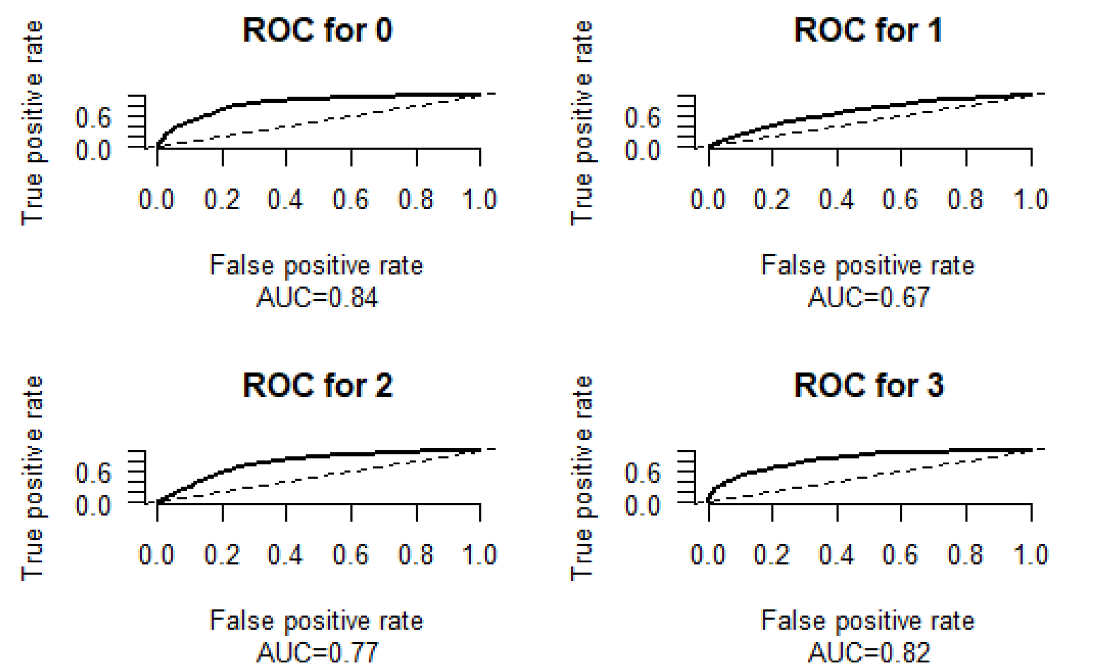


Figure 5: 'One vs. all' ROC curve diagnostics for the multinomial logistic regression

Conclusions

- I find moderately strong indications that the PRA is violated in Model 2. Ordered logit may be an inappropriate model because of variation in covariate slope parameters.
- The proposed multinomial logit model makes superior predictions to the ordered logit model at the extreme categories (0, 3) and performs comparably to the ordered logit on the intermediate categories (1, 2).
- I find evidence suggesting that a multinomial logistic regression is a more appropriate model for the data-generating process studied in Barnes et al. (2018).