doi:xx.xxxx/xxxxx

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

MATTHEW DRAPER *University of California San Diego*

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 (2018), JoP). 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 evidence to suggest that a multinomial logistic regression is a better model of the data-generating process studied in Barnes et al. (2018).

Word Count: 4,921

INTRODUCTION

Barnes et al. (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). My interest in this finding concerns the null effect of new information on existing attitudes. While it is unsurprising that exposure to more information about politics results in more political knowledge, it is quite surprising that more political knowledge does not lead to changed political attitudes or preferences. Converse (1964) famously found that most people know very little about their political system, and this finding has been repeatedly confirmed (Lewis-Beck et al. 2008). Democratic theorists have asserted that it

Matthew Draper is a PhD student in the Department of Political Science at the University of California San Diego (mdraper@ucsd.edu).

The author thanks Zhengyu Yang and John Ahlquist for detailed comments and feedback.

This is a manuscript submitted for review.

would be an (instrumentally) good thing for people to increase their political knowledge (Delli Carpini and Keeter 1996, Pettit 1996, Mackie 2003, Achen and Bartels 2016). However, if an increase in knowledge were to lead to no corresponding change in attitude or preference, it is unclear whether increased political knowledge would actually be desirable.

This article will proceed as follows. First, I will re-estimate the OLS and ordered logit knowledge models used in Barnes et al. (2018), including Figure 2 and Figure 3 in the main paper and Table 4, Table 5 and Table 7 in the Appendix. Next, I will fit a cumulative logit model in place of the ordinal logit approach used by the authors. Using this cumulative logit model to evaluate the parallel regressions assumption, I will then fit a multinomial model to the same data. Finally, I will re-examine the authors' assumptions regarding missing data, particularly the MAR (missing at random) assumption.

RE-ESTIMATION

Barnes et al. estimate treatment effects on knowledge acquisition via the following model (Model 1):

$$W_2K_i = \quad + \quad T_i + \sum_{k=1} {}_k x_{ik} + {}_i \tag{1}$$

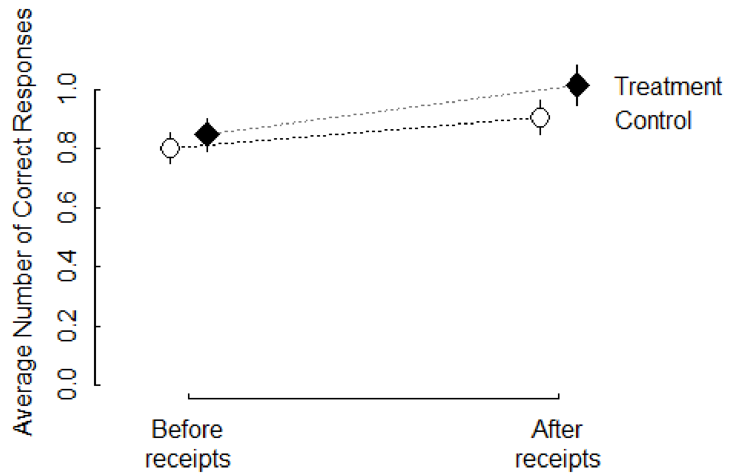
where i indexes respondents, W_2K_i represents respondents' political knowledge at Wave 2, T_i is a dummy variable indicating the subject is assigned to treatment, and x_{ik} is the kth covariate for individual i.

The paper also includes a model that controls for political knowledge levels as measured in Wave 1 (Model 2):

$$W_2K_i = \quad + \quad T_i + {}_0 W_1K_i + \sum_{k=1} {}_k x_{ik} + {}_i \tag{2}$$

where i indexes respondents, W_2K_i represents respondents' political knowledge at Wave 2, T_i is a dummy variable indicating the subject is assigned to treatment, x_{ik} is the kth covariate for individual i and W_1K_i is subject i's Wave 1 knowledge level. The authors estimate the impact of encouragement on knowledge acquisition using OLS and ordered logit models.

FIGURE 1. Comparison of treatment and control means before and after receipt distribution



2.png

Note: Replication of Figure 2 in Barnes et al. (2018).

Treatment Effects - OLS Models

I re-estimate the authors' OLS knowledge models in Table 1. Initially, I observed minor discrepancies between these re-estimation results and Table 4 in Barnes et al. (2018). While most values were identical or very close to the authors' results, the values for several covariates diverged dramatically. After investigation, it seems that the authors' directional recoding of their variables may have been implemented subsequent to the generation of published tables. The OLS models use versions of the variables without ordinal recoding, which I have replicated here and in this paper's replication file. Sample sizes (n) are the same in both cases, indicating that our statistical software package has not inadvertently omitted observations.

By taking simple averages of the Wave 1 and Wave 2 knowledge indices it is straightforward to estimate treatment effects with confidence intervals for measurement before and after treatment. I plot the results here (Figure 1), replicating Figure 2 in Barnes et al. (2018). The treatment effects given in the Results section of Barnes et al. (2018, p.702) are incorrect, and diverge from the results given in their appendix (Table 4). Model 1 actually yields a treatment effect of 0.097 correct answers ($p = 0.11$, SE = .042), while Model 2 (controlling for Wave 1 knowledge) actually yields a treatment effect of 0.074 correct answers ($p = 0.36$, SE = .035).

The knowledge indices that the authors use as their dependent variable in these models are ordinal data (integers from 0-3), and the regressors are either binary or ordinal variables. Without information on the distances between adjacent categories, treating ordered categories as continuous variables requires strong assumptions (Ward and Ahlquist 2018, p.142). I will test these OLS models by examining their predicted probabilities.

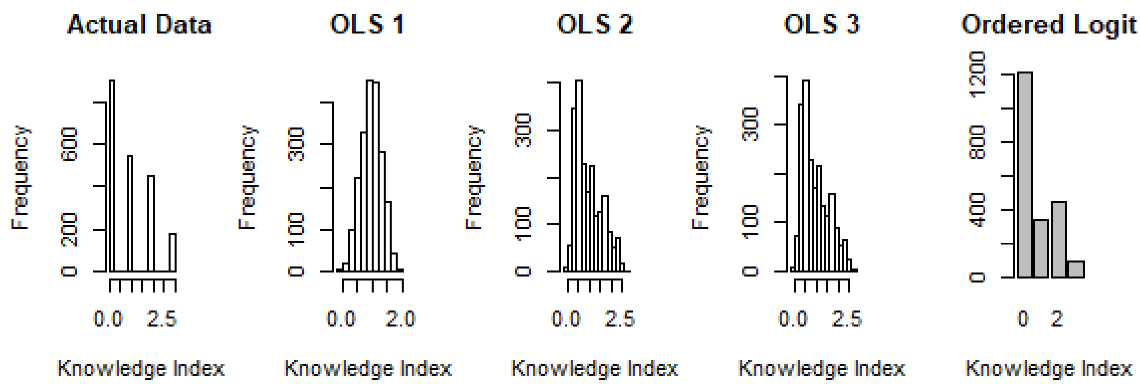
TABLE 1. Treatment effects on Wave 2 knowledge index, OLS models.

	Wave 2 knowledge index		
	(1)	(2)	(3)
Treatment	0.097	0.074	0.080
	(0.042)	(0.035)	(0.037)
Age	0.033	0.007	0.007
	(0.018)	(0.015)	(0.015)
Female	0.394	0.180	0.182
	(0.043)	(0.037)	(0.037)
White	0.112	0.124	0.151
	(0.062)	(0.053)	(0.052)
Conservative	0.084	0.070	0.062
	(0.056)	(0.047)	(0.047)
Labour	0.014	0.008	0.042
	(0.052)	(0.044)	(0.044)
Lib Dem	0.139	0.013	0.030
	(0.083)	(0.071)	(0.069)
Work full time	0.004	0.013	0.004
	(0.061)	(0.052)	(0.049)
Education	0.131	0.057	0.061
	(0.016)	(0.014)	(0.014)
Wave 1 Know.		0.547	0.551
		(0.019)	(0.019)
Constant	0.168	0.091	0.067
	(0.137)	(0.116)	(0.118)
N	2,072	2,072	2,072
R^2	0.120	0.367	0.372
Adjusted R^2	0.108	0.358	0.362
Resid. Std. Er.	0.941 (df = 2042)	0.798 (df = 2041)	2.050 (df = 2041)
F Statistic	9.618 (df = 29; 2042)	39.447 (df = 30; 2041)	40.239 (df = 30; 2041)

$p < .1$; $p < .05$; $p < .01$
Replication of Table 4 in Barnes et al. (2018)

All three OLS models predict minimum values below 0, and their maximum predicted values are 1.95, 2.63 and 2.72, respectively. Because they assume a continuous data-generating process, the

FIGURE 2. Predicted probabilities (knowledge index) of OLS and ordered logit models



Note: All OLS models predict outside the [0,3] interval, and OLS 1 predicts no values higher than 2.

OLS models predict non-integer values. These shortcomings of OLS are well-known, and the authors' decision to fit an ordered logit is a sound first step toward a better model.[1]

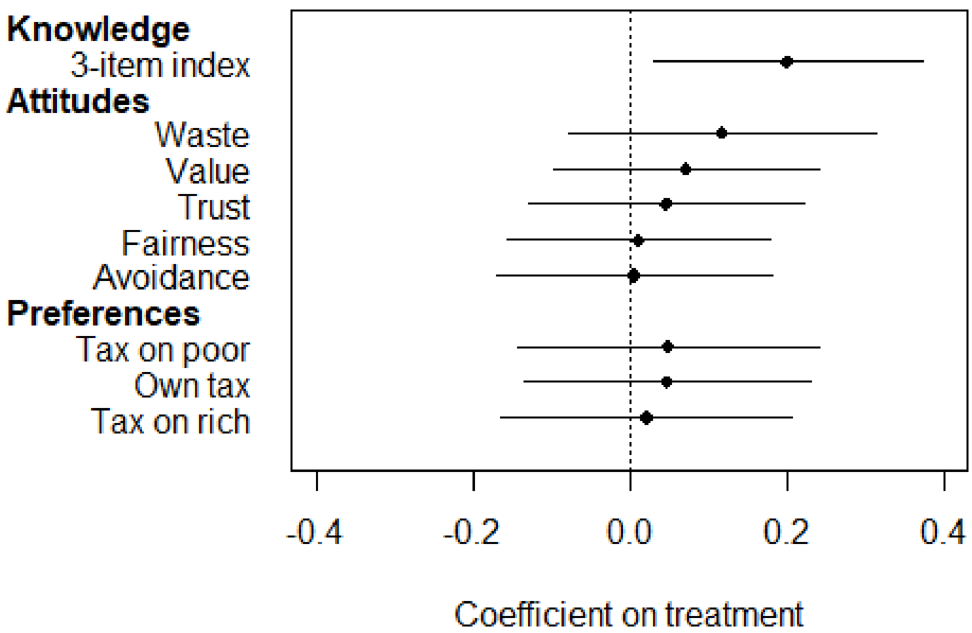
Treatment Effects - Logistic Models

Barnes et al. (2018) fit an ordinal logistic regression model of Wave 2 budget knowledge, controlling for demographic characteristics and Wave 1 knowledge. I reproduce their results in Table 4 (in the multinomial analysis below). Barnes et al. (2018) also disaggregate the knowledge index by subject area, modeling responses to each of the four knowledge categories separately using ordered logit. I replicate this disaggregation in Table 2. Sample sizes (n) are the same in all cases, indicating that our statistical software package has not inadvertently omitted observations.

We observe immediately that the ordered logit model makes better predictions than the OLS models. Predicted values are integers, and the strong bias in the data toward 0 is captured more fully than in the OLS models. However, the ordered logit model predicts more values of 2 than of 1, which is not borne out in the data. As mentioned, the authors found that the encouragement treatment affected knowledge but not attitudes or preferences. I replicate a graphical summary of their findings in Figure 3, showing significance for treatment effects on knowledge but none for attitudes or preferences. I have

[1]We might also consider an ordered probit for these data. However, the logistic and normal distributions are nearly indistinguishable except in the extrema, making the choice between them irrelevant for our purposes.

FIGURE 3. Treatment effects on Wave 2 knowledge, attitudes and preferences, ordered logit models



3.png

Note: Replication of Figure 3 in Barnes et al. (2018).

now re-estimated the core knowledge models used in Barnes et al. (2018). In the next section, I will interrogate the assumptions behind the use of an ordered logit and suggest an alternative model of the data-generating process. I will then address the problem of missing data in Wave 2 of the survey.

TABLE 2. Disaggregated knowledge index, ordered logit models

	Overseas aid		Defense		Health	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.258	0.159	0.089	0.021	0.168	0.163
	(0.115)	(0.158)	(0.102)	(0.119)	(0.112)	(0.123)
N	1,597	1,374	2,529	1,864	1,590	1,403
Log Likelihood	922.745	534.588	1,208.901	886.280	958.609	808.558
AIC	1,905.491	1,131.176	2,477.803	1,834.559	1,977.217	1,679.116

$p < .1$; $p < .05$; $p < .01$
Replication of Table 7 in Barnes et al. (2018)

EXTENSION

Ordered Logit Models as Cumulative Logit

Mapping variables onto coarse ordinal categories inevitably causes information loss, because binning data into categories fails to preserve the metric content of the original data. The ordered logit model is intended to circumvent this difficulty by positing a latent but unobserved continuous variable that is only reported in particular bins if its values fall between particular cutpoints (Ward and Ahlquist p.142). The ordered logit takes the form:

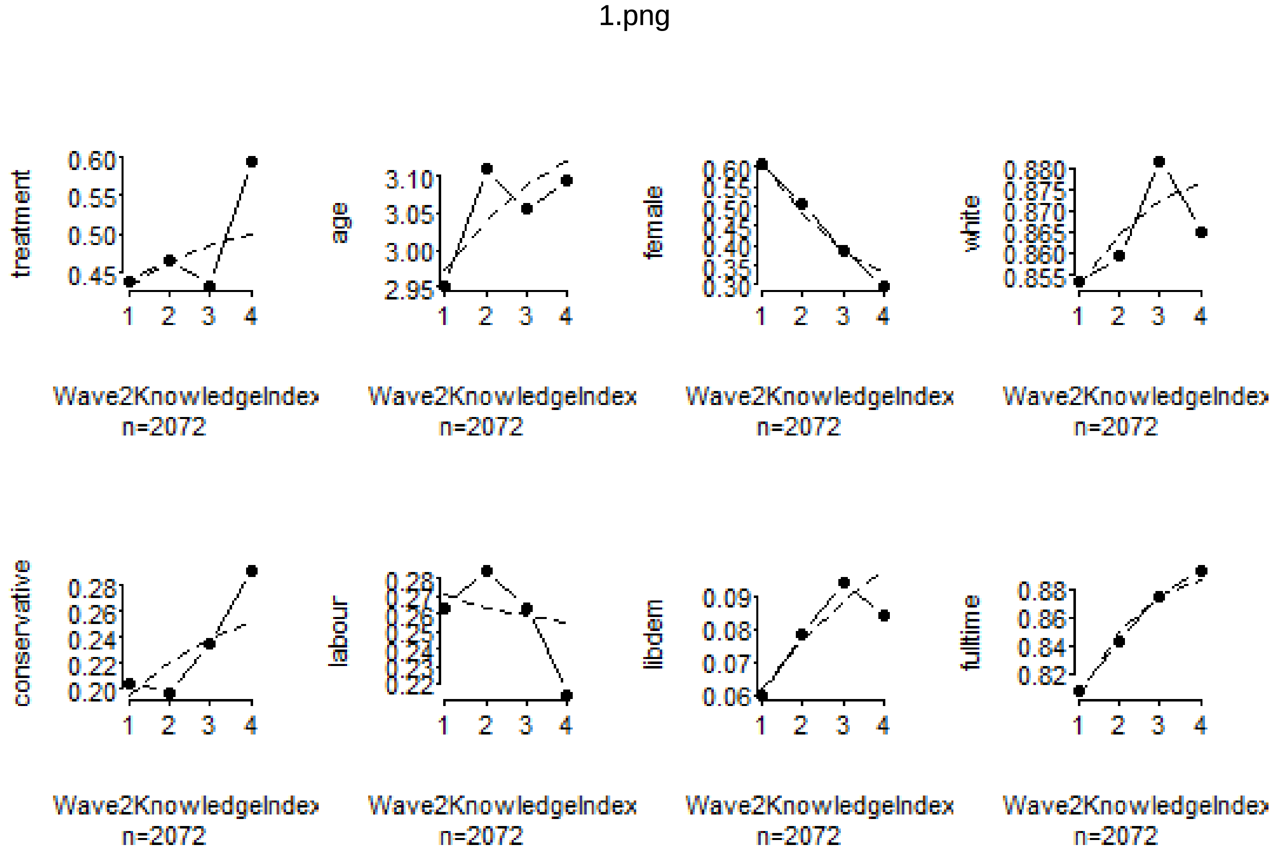
$$L = \prod_{i=1}^{n} \prod_{M=1}^{M} \left[\Lambda(\tau_m - x_i^T \beta) - \Lambda(\tau_{m-1} - x_i^T \beta)\right]^{1_{im}} \tag{3}$$

While this approach is certainly a defensible choice for the data-generating process studied in Barnes et al. (2018), the ordered logit model constrains slope parameters β_m to be identical across covariates. In general, continuous regression models are more desirable as the number of categories increases, particularly when data are roughly evenly distributed among outcome categories. Instead of conceptualizing the logit as a single model, we can instead imagine it as the constrained estimation of a system of models. This is because we can reexpress the dependent (ordered categorical) variable Y_i as a series of binary variables, $\tilde{Y}_{im}$, such that $\tilde{y}_{im} = 1 \iff y_i \leq m$ for some category, m. We thus need to fit a logit model for each of these $\tilde{Y}_m$, together comprising a cumulative logit model.

$$Pr(Y_i \leq M - 1) = logit^{-1}(\alpha_{M-1} + \mathbf{x_i^T} \beta_{M-1}) \tag{4}$$

In this general form, the cumulative logit model allows each equation to have its own set of slope parameters, β_m. Contrast this with ordered logit, where $\beta_m = \beta \; \forall m$. To address this aspect of the data-generating process, I fit a cumulative logistic link model to the date used in Barnes et al. (2018). It is important to emphasize that the effect of a covariate on the probability of being in the extreme categories is monotonic, but a covariate's effect on the probability of being in intermediate categories is not. I will use the results of this cumulative logistic regression to evaluate the initial choice of an ordered logit.

FIGURE 4. Plot of the conditional means of selected regressors at different levels of the response variable "Wave 2 Knowledge Index".

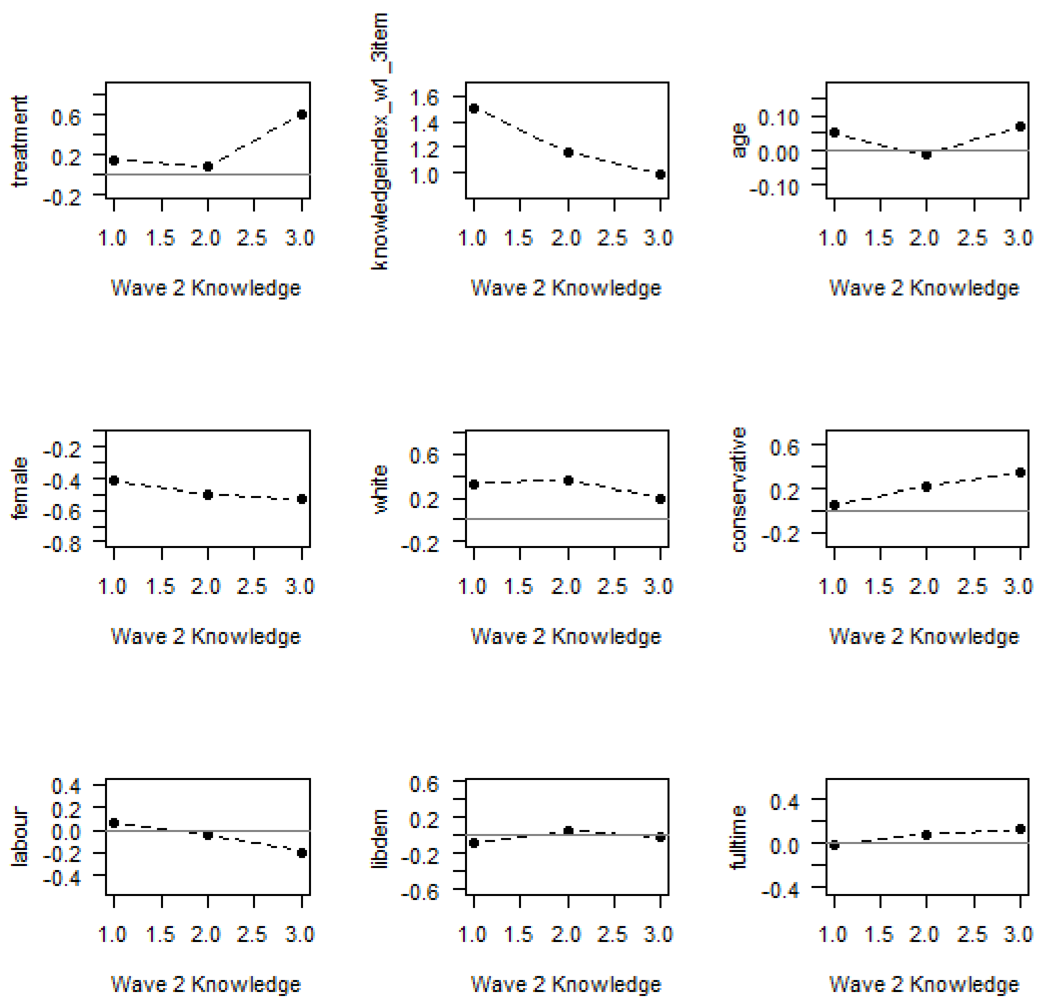


Parallel Regressions Diagnostics

In general, use of the ordered logit model requires an assumption of common slope parameters across levels of the response variable, also known as the "parallel regressions assumption", or "proportional odds". This assumption is easier to satisfy with few covariates but becomes more demanding as covariates increase in number. I will use the unconstrained cumulative logit developed above to test the parallel regressions assumption.

Barnes et al. (2018) employ no fewer than 30 covariates in each of their model. Approximately twenty of these are binary (dummy) variables coding for income and region. I will investigate to see whether the inclusion of so many covariates in their ordered logit models violates the parallel regressions assumption by examining the means of the regressors at different levels of the response variable. Covariates appearing as significant should show linear relationships with the dependent variable. I plot the results in Figure 4. We observe non-linear relationships in several covariates,

FIGURE 5. Plot of the estimated regression coefficients from M 1 regressions on $\tilde{Y}_m$



including the informational treatment itself. This non-linearity is an indication that the parallel regressions assumption may be violated.

We can also test the parallel regressions assumption by examining the stability of our cumulative logit coefficients across all levels of the response variable. Since our response variable has four categories, this implies three equations using the cumulative logit conceptualization. I plot the results in Figure 5. We see immediate signs of nonlinearity, particularly on the treatment variable. We also observe instability in the estimated coefficients across different levels of $\tilde{Y}$. These are indications that the parallel regressions assumption may be violated in this model.

Multinomial Logit

Since the parallel regressions assumption appears to be violated, we may wish to consider a multinomial logit instead 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:

$$Pr(Y_i = m \mid \mathbf{x_i}) = \frac{exp(\mathbf{x_i^T} \beta_m)}{1 + \sum_{j=2}^{M} exp(\mathbf{x_i^T} \beta_j)} \tag{5}$$

$$L = \prod_{h=1} \left(\frac{\sum_{h=1}^{M} exp(\mathbf{x_i^T} \beta_h)}{\sum_{l=1}^{M} exp(\mathbf{x_i^T} \beta_l)} \right)^{\mathbf{1}_{ih}} \tag{6}$$

So far, this is similar to the cumulative logit which we estimated above. However, the β_m need not be constant across categories because we will estimate $M - 1$ sets of regression parameters, each of which will give us the log odds of being in category m versus the reference category (Ward and Ahlquist pp.163-4). Results of the multinomial logit appear in Table 4. We see that the multinomial logit performs best at higher values of the knowledge index, and indeed the treatment covariate is only significant at the highest category of the response variable. This indicates that the model is doing a good job of predicting the extremes but a poorer job of predicting middle categories. The multinomial predicts fewer values of 3 (62) than the ordered logit model (92), but both underpredict actual values of 3, as shown in Table 3.

TABLE 3. Actual and predicted values of the knowledge index (n = 2072)

	0	1	2	3
actual	899	547	448	178
ordered.logit	1,206	349	455	62
multinomial	1,202	339	439	92

A core assumption of the standard multinomial model is that the ratio of the probabilities of choosing among any two outcomes must be invariant with respect to other alternatives. This is known as the independence of irrelevant alternatives assumption, or IIA, which must be met if we are to use a multinomial model. To determine whether the multinomial model is appropriate for these data, I will conduct tests of the IIA assumption. First, I will administer a Wald test, a generalized version of the standard t-test (^{2}), based on the variance-covariance matrix.

$$W = \frac{1\,\hat{}\qquad h^{o2}}{I^{1\,\hat{}\,o}} \tag{7}$$

This test describes how the likelihood changes as we impose restrictions on our models, relying on standard regularity assumptions. I give the results in Table 5.

TABLE 4. Treatment effects on Wave 2 knowledge index, ordered logit and multinomial logit.

	ordered logit	multinomial (1)	multinomial (2)	multinomial (3)
Treatment	0.192	0.146	0.0001	0.644
	(0.087)	(0.120)	(0.140)	(0.195)
Age	0.031	0.079	0.009	0.065
	(0.037)	(0.051)	(0.059)	(0.084)
Female	0.434	0.250	0.570	0.831
	(0.091)	(0.124)	(0.144)	(0.208)
White	0.304	0.241	0.521	0.511
	(0.132)	(0.180)	(0.217)	(0.296)
Conservative	0.161	0.063	0.121	0.409
	(0.118)	(0.164)	(0.186)	(0.248)
Labour	0.018	0.086	0.041	0.148
	(0.108)	(0.146)	(0.173)	(0.253)
Liberal Democrat	0.025	0.123	0.025	0.132
	(0.172)	(0.249)	(0.277)	(0.380)
Working full time	0.058	0.055	0.032	0.153
	(0.130)	(0.172)	(0.209)	(0.313)
Education scale	0.160	0.209	0.159	0.278
	(0.035)	(0.047)	(0.055)	(0.081)
Wave 1 knowledge	1.246	1.192	1.780	2.137
	(0.053)	(0.089)	(0.095)	(0.120)
Constant		2.745	3.259	5.735
		(0.402)	(0.480)	(0.724)
N	2,072	2,072	2,072	2,072
AIC	4,359.70	4,397.67	4,397.67	4,397.67

$p < .1$; $p < .05$; $p < .01$

Note: the multinomial logit baseline category is a knowledge index score of 0.

Next, I will compare actual and predicted values ($\hat{y}_i$ and y_i) from our multinomial logistic regression

TABLE 5. Wald test of variables in the multinomial logistic regression model in Table 6.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
LR Chisq	30	27.545	115.912	0.339	1.260	8.489	639.407
Df	30	3.400	2.191	3	3	3	15
Pr(>Chisq)	30	0.391	0.348	0.000	0.037	0.739	0.952

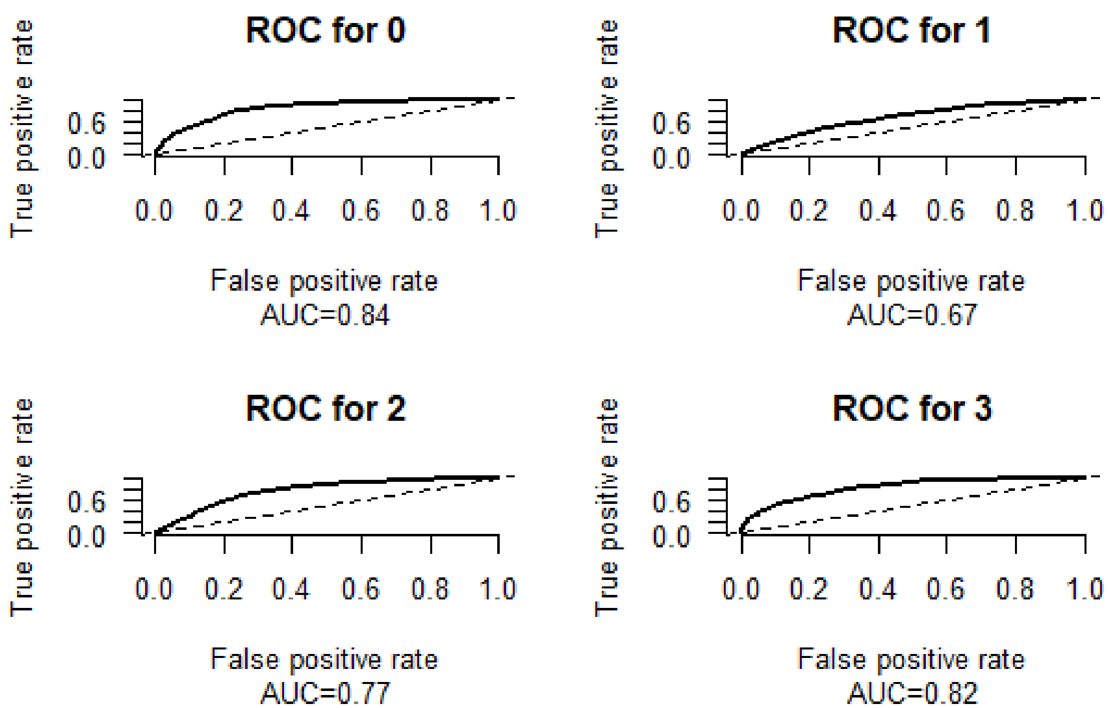
in a confusion matrix to assess outcome prediction. This will allow us to determine the correct classification rate and the implied error rate (these two values sum to 1). I show the results in Table 6. We see that the model is doing a good job of predicting 0 vs. 1-3, but is not performing much better than random assignment on the remaining categories.

TABLE 6. Confusion matrix of the classifications produced by the multinomial logistic regression in Table 6. Horizontal axis: predicted values; vertical axis: actual values.

	0	1	2	3
0	781	66	50	2
1	258	161	122	6
2	119	105	201	23
3	39	27	80	32

Finally, I generate "One-v.-all" ROC curves considering the false positive rate against the true positive rate for each value of the outcome variable. I plot the results in Figure 6. As space between the 45-degree line and the ROC curve increases, prediction performance improves. We observe the lowest rates of false positives in the baseline category (0) and at the highest category (3). While no test of the IIA assumption is dispositive, these diagnostics give us no particular reason to worry that IIA is violated. Of greater concern, the multinomial model is not a significant improvement on the ordered logit. AIC is comparable in both models. Because multinomial models rapidly consume degrees of freedom, we have a weak reason to prefer a multinomial to an ordered logit at the same AIC, but we do not see significantly better prediction in the multinomial model. However, the survey data used in Barnes et al. (2018) contains significant amounts of missing data which were simply dropped (complete case analysis). In the next section, I will explore whether estimating values for these missing data enables us to say more about model fit.

FIGURE 6. "One vs. all" ROC curve diagnostics for the multinomial logistic regression given in Table 6.



Missing Data Analysis

Barnes et al. (2018) explicitly rely on the Missing Completely at Random (MCAR) assumption when they choose to analyze only complete cases. That is, they stipulate that "missingness is ignorable given covariates, treatment assignment, and responses in Wave 1" (p.A12). I produce a "missingness" map of their data in Figure 7. We immediately notice an odd feature, which is that approximately 500 consecutive observations appear to lack answers to the Wave 2 survey questions. Without speculating on the cause of the error, this lacuna introduces bias into the estimates produced. These data were gathered regionally, and the dramatic gap likely occurred in one or a small number of regions, biasing the sample.

I now generate plots of the proportion of observations missing for a selection of variables, and the frequency of combinations for missing and non-missing variables (Figure 8). There is clearly moderate cause for concern that these missing data are biasing Barnes et al.'s results. The authors' decision to employ complete case analysis is defensible, but alternative approaches would assuage worries about

FIGURE 7. Missing data in Barnes et al. (2018)

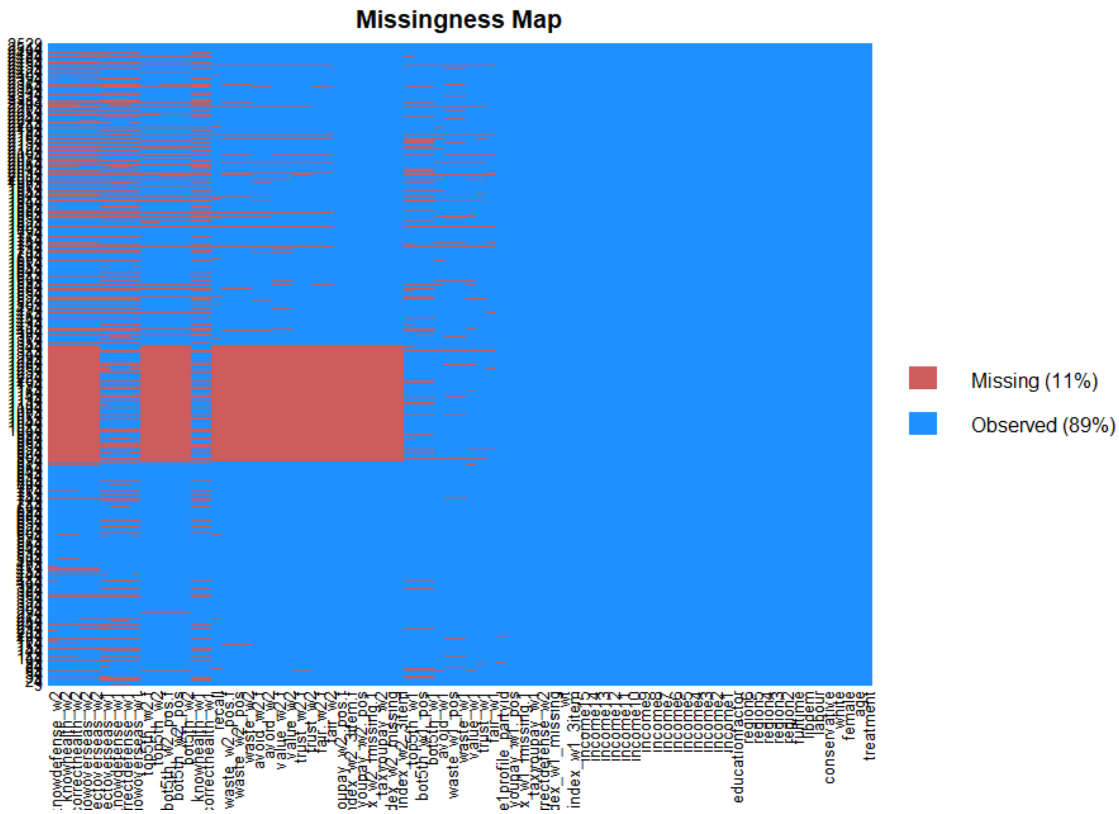
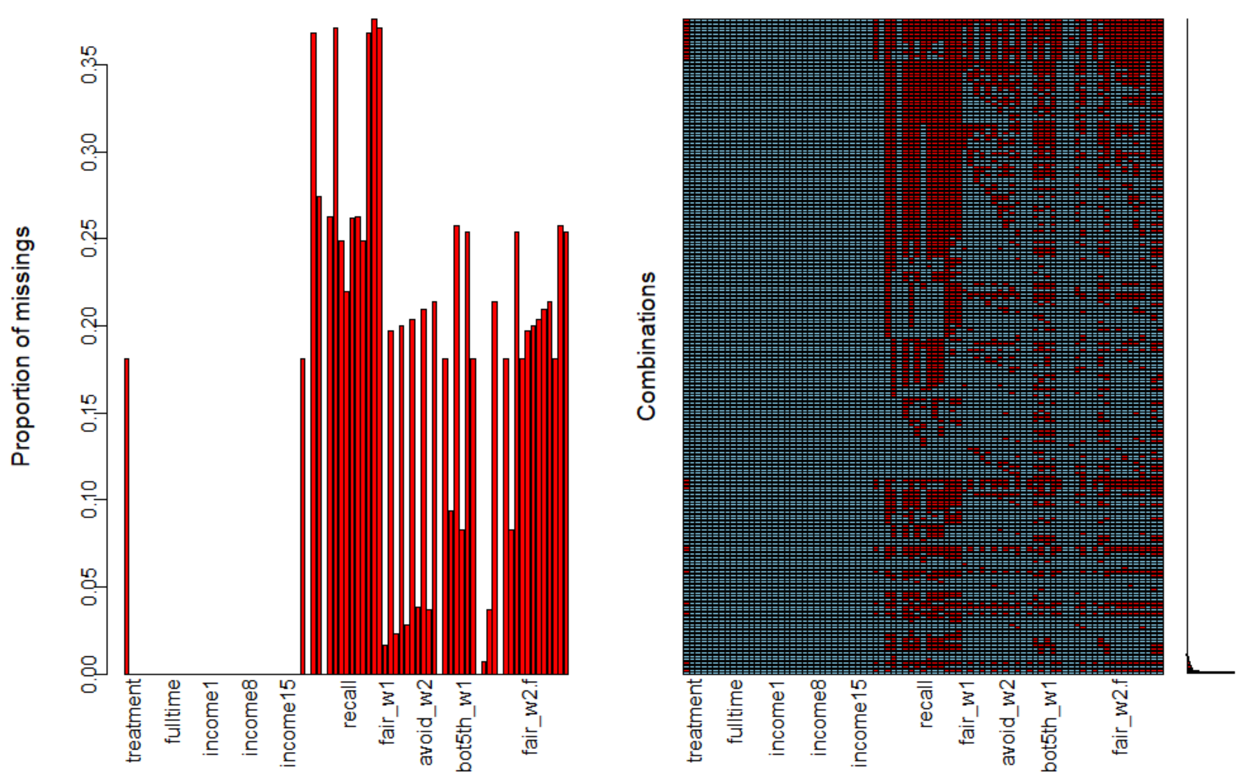


FIGURE 8. Missing data in Barnes et al. (2018)



bias. We might consider available case analysis or conditional mean imputation, but these techniques can still lead to bias (even under MCAR), and they fail to account for estimation uncertainty in the regression model (Ward and Ahlquist p.258). I attempted to use multiple imputation to generate multiple values for each missing value. However, the inherently multicollinear nature of the survey data made multiple imputation impossible. I will close by noting that the missingness present in this survey appears to be non-random and consequently biases the results obtained in Barnes et al. (2018).

CONCLUSION

In this article, I have re-estimated and extended published work on the impact of government-issued taxpayer receipts on political knowledge and political attitudes. After replicating the authors' findings using the original survey data (and finding errors in the published paper), I re-estimated Barnes et al.'s ordered logit knowledge models by fitting a cumulative logit model and using it to test the parallel regressions assumption. After finding that this assumption may have been violated, I fit a multinomial logit model and compared it to the ordered logit. The multinomial model gives better classification results at the extreme values, and no worse classification at the intermediate values. Finally, I attempted to provide a more rigorous treatment of the missing data problem by applying multiple imputation. In summary, I find evidence to suggest that a multinomial logistic regression is a more appropriate model for the data-generating process studied in Barnes et al. (2018).

REFERENCES

Achen, Christopher and Bartels, Larry. "Democracy for Realists" Princeton University Press (2016).

Carpini, M., Keeter, S. "What Americans Know about Politics and Why It Matters." New Haven; London: Yale University Press. (1996). http://www.jstor.org/stable/j.ctt1cc2kv1

Kulikinski, James et al. "Misinformation and the Currency of Democratic Citizenship". Journal of Politics Vol. 62, Issue No. 3 (23 January 2003). https://doi.org/10.1111/0022-3816.00033

Lewis-Beck, M., Jacoby, W., Norpoth, H., Weisberg, H., Converse, P. (2008). "The American Voter Revisited." University of Michigan Press. http://www.jstor.org/stable/10.3998/mpub.92266

Lucy Barnes, Avi Feller, Jake Haselswerdt, and Ethan Porter, "Information, Knowledge, and Attitudes: An Evaluation of the Taxpayer Receipt," The Journal of Politics 80, no. 2 (April 2018): 701-706. https://doi.org/10.1086/695672

Mackie, Gerald. "Democracy Defended". Cambridge University Press (2003).

Pettit, Philip. "Republicanism", Oxford University Press (1997).

Philip E. Converse. "The nature of belief systems in mass publics" (1964), Critical Review, 18:1-3, 1-74, (2006). https://doi.org/10.1080/08913810608443650

Ward, Michael Don and Ahlquist, John. "Maximum Likelihood for Social Science". Cambridge University Press (2018).