

PLSC 597
Causal Inference
Homework 4
Due: April 25th 11:59pm

We are going to work with the following paper,

Robertson, Claire E., et al. "Negativity drives online news consumption." *Nature Human Behaviour* (2023): 1-11.

Read the entire paper and get a sense of what they're analyzing, how and why. [Here is more context.](#)

Here are the links to the [paper](#) and [all replication materials](#). Using these materials, carry out the following:

1. Scope conditions: This is a dataset of RCTs, so we are less concerned about confounding, but there is still the question of scope conditions. Describe

- Substantive/theoretical quantity of interest
- The estimand defined by the specific comparisons (eg "treatment" vs "control") in the current paper
- The statistical estimator

2. Replication: Replicate Figure 2. Discuss the choice of the addition of control variables in panel a. Discuss the decision to marginal effects in panels b and c with a linear model.

3. Hypothesis generation: think about another hypothesis that might be tested with the Upworthy Research Archive data. You will need to do some text analysis in order to implement it, so don't worry about that for now – instead, just describe hypothetically how you would answer the three bullet points in question 1 in your hypothetical study.

4. In the manuscript, the authors conduct a random effects meta-analysis of the RCTs, using platform age ("relative_date") as a contextual variable for each of the RCTs. They find that

the control for platform age is negative ($\beta = -0.309$, $SE = 0.005$, $z = -56.917$, $P < 0.001$, 99% $CI = (-0.323, -0.295)$). Hence, stories published later in Upworthy's career had lower CTRs than stories published at the beginning of Upworthy's career, implying that Upworthy headlines were most successful when its editorial practices were novel to online users.

Given that they are primarily interested in the “theoretically relevant” variables related to the sentiment etc, they also fit random effects models where the slopes of the relationships of the CTR and sentiment are allowed to vary across experiments.

What if we are concerned with the temporal validity findings? Re-run the analysis you conducted above, in two ways:

- Divide the dataset into four equal time periods and simply replicate the analysis on each of them. How much do the coefficients in Figure 2a change over time?
- Re-run the analysis in Table 3, removing the varying slopes for “liwc_posemo” and “liwc_negemo” and replacing them with varying slopes for the platform age at which the analysis was run. What do you conclude about the temporal validity of these findings?