Matthew-David Ilomuanya

PSAM UN3707

Eunji Kim and Chris Wiggins

03/28/25

Task 2

1) Introduction and IV Strategy

Understanding whether social media engagement causally influences political behavior is a major challenge in both political science and data science. In particular, people who consume political content on platforms like Snapface are likely to already have strong opinions or be more politically engaged, making it difficult to disentangle correlation from causation. To address this issue of endogeneity, we use an instrumental variable (IV) strategy based on a naturally occurring shock: Hurricane Blandy's blackout in Manhattan.

In October, Hurricane Blandy caused widespread power outages south of 39th Street, unexpectedly disrupting internet access and thus limiting Snapface usage in affected areas. We use a binary indicator for blackout exposure as an instrument for Snapface usage. This approach allows us to isolate exogenous variation in Snapface exposure, enabling a causal estimate of its impact on the probability of voting for the Purple party.

2) Key IV Assumptions: Relevance and Exogeneity

Relevance: The instrument must be strongly correlated with the treatment. In our first-stage regression, we found that experiencing a blackout reduced Snapface usage by an average of 84.64 seconds ($p < 0.0001$). The model has an R-squared of 0.518 and an F-statistic

of 5365, far above the conventional threshold of 10 for a strong instrument. This provides robust evidence that the relevance assumption is satisfied.

Exogeneity: The instrument must only affect the outcome (voting behavior) through the treatment (Snapface usage). While the blackout was a natural, unanticipated event, threats to this assumption remain. For example, if neighborhoods more likely to lose power also differ politically in ways unrelated to Snapface use, or if the blackout itself affected voting behavior (e.g., through anger at local government), then the exclusion restriction would be violated. While we control for income to mitigate this risk, we acknowledge that exogeneity is ultimately untestable and must be justified based on context and theory.

3) Data Patterns and Anomalies

Snapface usage is right-skewed, with most users spending moderate amounts of time but some spending considerably more. Income varies widely, with a few high-income outliers that may influence the analysis. About 45.4% of respondents voted for the Purple party, and 40% experienced the blackout, providing good variation across key variables. No severe anomalies were found, though the high condition number (~6.8e+05) in the first-stage regression suggests the possibility of multicollinearity, likely due to the wide range of income values.

4) Interpretation of First-Stage Results

In the first-stage regression, we modeled Snapface usage as a function of blackout exposure and income. The coefficient on blackout was -84.64 seconds and highly significant ($p < 0.0001$). This confirms that the instrument meaningfully affects the treatment variable. The model's F-statistic (5365) indicates a very strong instrument, satisfying the relevance

assumption. Additionally, income is positively and significantly related to Snapface usage, suggesting higher-income individuals use the platform slightly more.

5) Interpretation of Second-Stage Results

In the second-stage regression, we regressed voting for the Purple party on the predicted values of Snapface usage from the first stage, controlling for income. The coefficient on predicted Snapface usage is -0.0003, and it is statistically significant ($p = 0.0285$). This means that, on average, an increase of one second in Snapface exposure reduces the probability of voting for the Purple party by 0.03 percentage points. While the effect size is small, the direction is notable: Snapface usage appears to causally decrease support for the Purple party. This could be due to the nature of content promoted by the platform (e.g., polarizing or anti-Purple messaging).

6) Validity of the IV Approach

The evidence supports the validity of the IV strategy in terms of instrument strength. The blackout is clearly a powerful and significant predictor of Snapface usage. The more difficult question is whether the blackout affects voting only through Snapface usage. We control for income, and the event itself was exogenous, supporting the plausibility of the exogeneity assumption. However, it remains possible that other effects of the blackout (e.g., reduced access to news, stress, changes in mobility) influenced political behavior. Overall, the IV strategy is credible but not immune to criticism, and future studies should explore robustness through alternative specifications or placebo tests.

7) Conclusion and Implications

This analysis suggests that Snapface usage has a causal, negative effect on support for the Purple party. Even though the magnitude is small, the statistical significance and direction of the effect point to an important political consequence of social media exposure. For policymakers, this underscores the urgency of regulating misinformation and content moderation during elections. For researchers, it illustrates the value of natural experiments and instrumental variable techniques in isolating causal effects in politically sensitive contexts.

8) GenAI Tool Usage

ChatGPT:

I used some sample code from another class that I am taking (Big Data with Python), and after I took some of the code and input the data from this class, I asked ChatGPT if it could add notes to the code explaining what each part of the code does, instead of manually adding notes to the code myself. To ensure quality control I looked over the notes that were made to make sure that they were correct and then manually copied them into Colab. To reflect on how this affected my final work, it makes it easier for others to keep track of what my code does, and as I am someone who generally retroactively adds notes to code instead of adding them as I code, it makes the workload easier for me, and allows me to move onto other parts of assignments faster. Going back over the notes also helps me to remember the syntax moving forward as it forces me to review what I have written