Cells Distract Capstone

By: David Fix, Scott Miketa, Tanner Olson
University of Michigan, School of Information

Introduction

The ubiquity of cell phone use has introduced a new and pervasive challenge for anyone in the field of education. The physical act of using the phone during class might be the most immediate problem presented, which can lead to discussions about the discipline policy, compliance, insubordination and classroom management. However, deeper problems exist that might be less obvious, and more detrimental to learning. These issues must also be researched and weighed when crafting educational policies. This analysis focuses on a question, purely academic in nature: Do cell phone bans in public school districts in Michigan increase student academic performance?

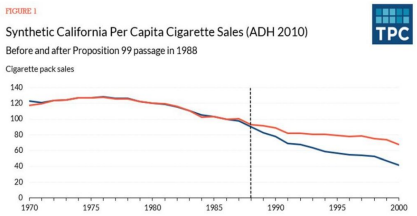
Data

Michigan school data was obtained from the State of Michigan website, where we were able to aggregate test scores as a measure of academic progress. Census data was obtained from the censusdata API, and we were able to join this data to the Michigan school data. We kept key attributes about demographic information constant so that our analysis could be more robust, and more resistant to variables outside of the phone ban having a significant effect.

Common data cleaning procedures were necessary to clean text strings such as removing whitespace from school district names, as well as using delimiters to separate text so that the appropriate joins would function properly. Some missing or ambiguous numerical data was replaced by mean scores in aggregate.

Synthetic Control Analysis

Synthetic control is a growing approach in causal inference which creates a synthetic copy of the control units that closely mirrors the pre-treatment patterns of the treated units. This approach is based on the theory that the similarity between the synthetic copy and the treatment group would continue in the absence of a treatment. Synthetic control became widely popularized in 2010 when three researchers, Abadie, Diamond, and Hainmueller, utilized synthetic control to evaluate the impact of a cigarette tax in California on cigarette sales(figure below taken from the Tax Policy Center).



In practice, synthetic control finds the weighted average of the control group that most closely resembles the treated units. Covariates used in the model are normalized to ensure that the covariates' scale and assigned measurement do not overly influence the model. An example of this would be a covariate that is assigned a binary would be weighted far less than a covariate that is measured in thousands or even billions. From an implementation standpoint, researchers have historically leveraged Lasso regressions on control data to create a synthetic control. However, in 2020, a synthetic controls package was created and launched in python to help simplify the implementation of synthetic control. This has been widely adopted and has further increased the use of this method for causal inference. We have leveraged the SyntheticControlMethods package for our analysis.

Methodology

We concluded that 14 school districts in Michigan had policies that constituted a district-wide cell phone ban. We chose a prominent school district, Forest Hill Public Schools, as our treatment group and subtracted the other 13 school districts from our data frame in order to maintain the integrity of our control group. The outcome we used to assess academic progress was the 8th grade math test. We chose this test because it was universally assigned and provided a measure of consistency between districts

For this analysis, we evaluate the output of two synthetic control models that were run with a dataset representing the top five relevant features found to impact 8th grade math scores (outcome variable) and with a dataset that is representing the top ten relevant features. These features were identified by: 1) leveraging a PCA analysis to understand most heavily weighted covariates that impact school quality and 2) leveraging teacher experience to further select covariates based on experience. When running the data through the models, the synthetic control is made of only Ann Arbor Public Schools for the analysis with five features and is made up of Troy School District and West Ottawa Public School District for the analysis with ten features.

Results

To examine whether or not we see an impact on 8th grade math scores when school districts ban cell phones, we apply the weights as described above to create a synthetic control.. When running the model and visualizing its outcome on the 5-feature dataset, we get the following output (Figure below). The top line chart in the cluster shows the outcome of the synthetic control plotted against the actual outcome of the treatment. The middle line chart in the cluster adjusts the synthetic control to baseline at 0 and adjusts the treatment district accordingly to help show the change in actual percentage points. The bottom line chart in the cluster shows the summed treatment effect. Through the "Cumulative Effects" chart in the cluster, we can observe that the treatment had a 5 percentage point positive impact the year after the cell phone ban was initiated and close to a 12.5 percentage point positive impact the second year after the cell phone ban was initiated.