# The Impact of Natural Disasters on Education: Evidence from Standardized Testing

Gregor Steiner

May 3, 2022

**Abstract**

# 1 Introduction

# 2 Data

## 2.1 Natural Disaster Data

Natural disasters are declared as such by the president, usually upon request by the affected state's governor. Once a disaster is federally declared, states or local governments can receive federal assistance. The Federal Emergency Management Agency (FEMA) provides data on all federally declared natural disasters, beginning in 1953. The data is easily accessible via their API (Turner, 2022).

Figure 1 shows the number of declared disasters between 2009 and 2018 across the US. It seems that the variation in the number of declared disasters may be driven by the governor's proactiveness in requesting a declaration. Thus, it could be interesting to compare counties on different sides of state borders, whose actual disaster exposure is likely very similar in order to analyze the effect of a declaration.

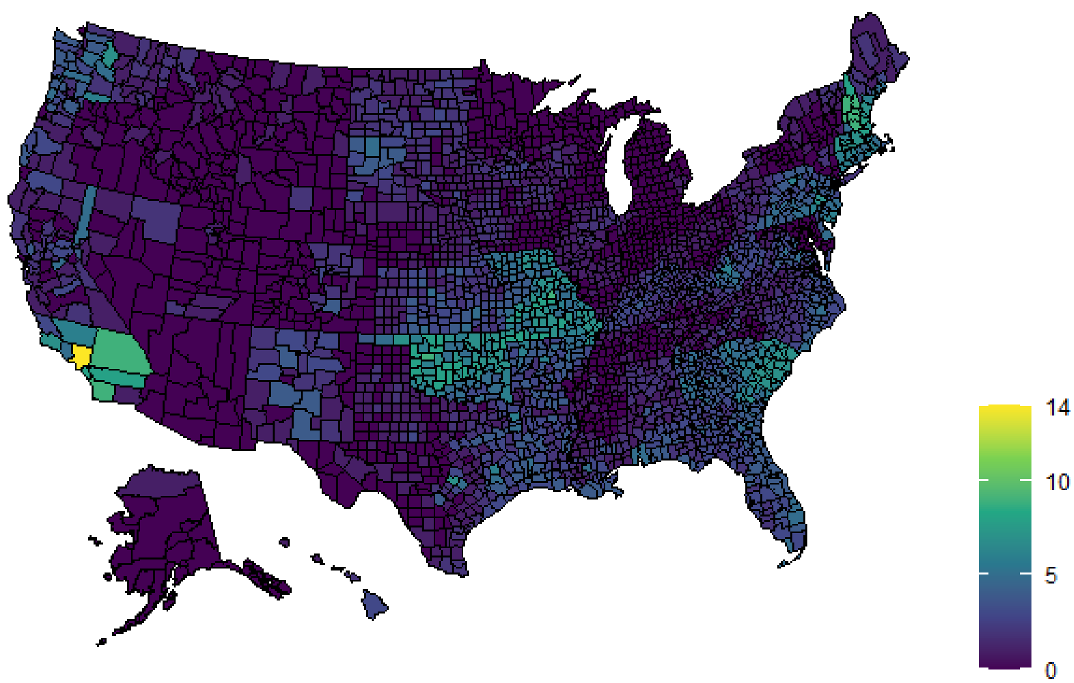


Figure 1: Number of declared natural disasters from 2009 to 2018

FEMA also provides a dataset on their Public Assistance Applicants Program Deliveries. This contains information on applicants and their recovery priorities, including the amount of damage caused and amount of federal disaster assistance granted. Unfortunately, this data is only available since October 2016. Figure 2 shows the total federal assistance awarded to counties.

Figure 3 and figure 4 show boxplots and kernel density estimates by county application status. Counties that did apply for federal disaster assistance tend to have lower median income, higher poverty rates, and higher shares of single motherhood. Thus, it seems that counties that had to apply for federal disaster assistance were more socially vulnerable in the first place. Visually, the distribution of democratic votes in the 2016 election does not seem to be different in the two groups.

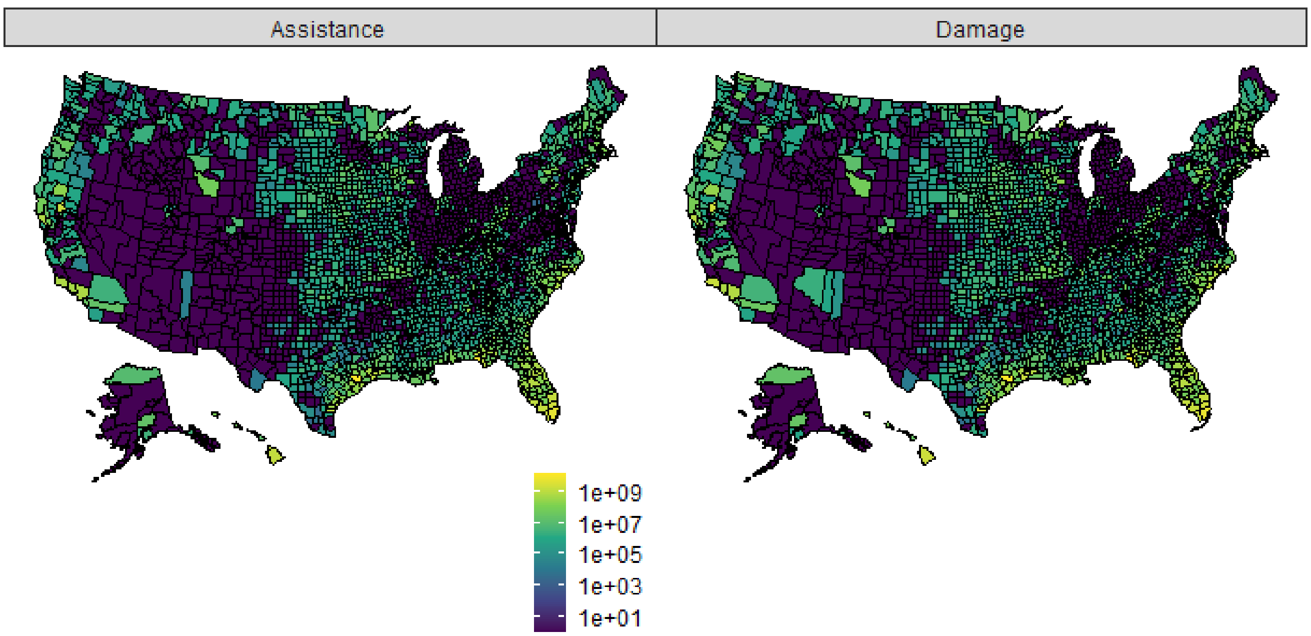


Figure 2: Amount of disaster damage and federal disaster assistance (in USD) awarded to counties since October 2016

### 2.2 Standardized Testing Data

Data on academic achievement is available from the Stanford Education Data Archive (Reardon et al., 2021). They provide mean test results from standardized tests by county, year, grade and subject among all students and various subgroups (including race, gender, and economically disadvantaged). The most recent version 4.1 covers grades 3 through 8 in mathematics and Reading Language Arts (RLA) over the 2008-09 through 2017-18 school years.

Test scores are cohort-standardized, meaning they can be interpreted relatively to an average national reference cohort in the same grade. For instance, a county mean of 0.5 indicates that the average student in the county scored approximately one half of a standard deviation higher than the average national student in the same grade.

In addition to mean test scores, the data includes estimates of gap estimates for various sub-

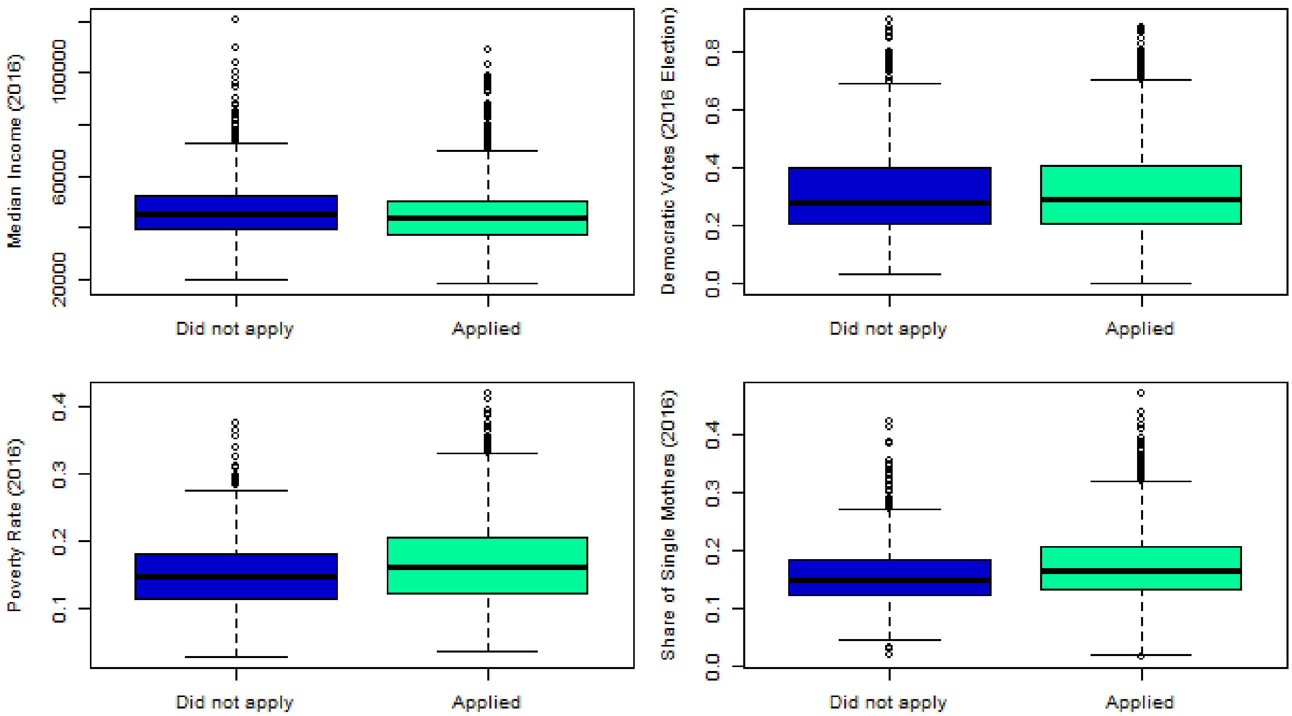


Figure 3: Boxplots by application status

groups, e.g. mean difference in test scores among white and black students. These are only reported if the subgroups' sample sizes are large enough. Thus, the number of observations for some of the gap statistics is substantially smaller.

Furthermore, the Stanford Education Data Archive maintains a large set of covariates for each county and year. They include variables like the county's median income, unemployment and poverty rate.

### 2.3 Combining disaster and testing data

Natural disasters should only have an effect on test scores if they occur before the test. Standardized tests are generally administered during spring. We will use March 1st as a cut-off point. Thus, any disaster happening within the same school year before the 1st of March will be considered. School years tend to start in late August or early September with some variation across states. We will use September 1st, meaning any disaster happening between September 1st and March 1st will be counted for a given school year.

Each disaster is assigned to a school year as described above. Then, disaster and test score data can be merged by school year and county. This yields a panel data set with six grades and two subjects for each county-year combination. Table 1 shows summary statistics for all relevant variables

Table 1: Summary Statistics

| Variable | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 75 | Max |
|---|---|---|---|---|---|---|---|
| Disasters | 336415 | 0.221 | 0.566 | 0 | 0 | 0 | 6 |
| Disaster Dummy | 336415 | | | | | | |
| ... 0 | 282581 | 84% | | | | | |
| ... 1 | 53834 | 16% | | | | | |
| Cumulative Disasters | 336415 | 1.245 | 1.631 | 0 | 0 | 2 | 14 |
| Grade | 338349 | | | | | | |
| ... 3 | 58603 | 17.3% | | | | | |
| ... 4 | 58501 | 17.3% | | | | | |
| ... 5 | 57265 | 16.9% | | | | | |
| ... 6 | 56992 | 16.8% | | | | | |
| ... 7 | 54392 | 16.1% | | | | | |
| ... 8 | 52596 | 15.5% | | | | | |
| Subject | 338349 | | | | | | |
| ... Mathematics | 165352 | 48.9% | | | | | |
| ... RLA | 172997 | 51.1% | | | | | |
| Mean test score | 327358 | -0.042 | 0.294 | -3.196 | -0.214 | 0.152 | 1.669 |
| White-Black gap | 130595 | 0.618 | 0.257 | -0.754 | 0.454 | 0.771 | 2.358 |
| White-Hispanic gap | 140326 | 0.451 | 0.258 | -1.713 | 0.282 | 0.611 | 2.213 |
| Male-Female gap | 306055 | -0.131 | 0.199 | -1.612 | -0.258 | 0.001 | 1.248 |
| Disadvantaged gap | 282423 | 0.543 | 0.211 | -0.995 | 0.413 | 0.669 | 2.052 |
| Total Damage | 110901 | 1076549.238 | 21067263.821 | 0 | 0 | 0 | 1868575345.74 |
| Federal Assistance | 110901 | 817694.776 | 17101611.639 | 0 | 0 | 0 | 1628231852.72 |
| Log Income | 338088 | 10.693 | 0.253 | 9.305 | 10.546 | 10.834 | 11.727 |

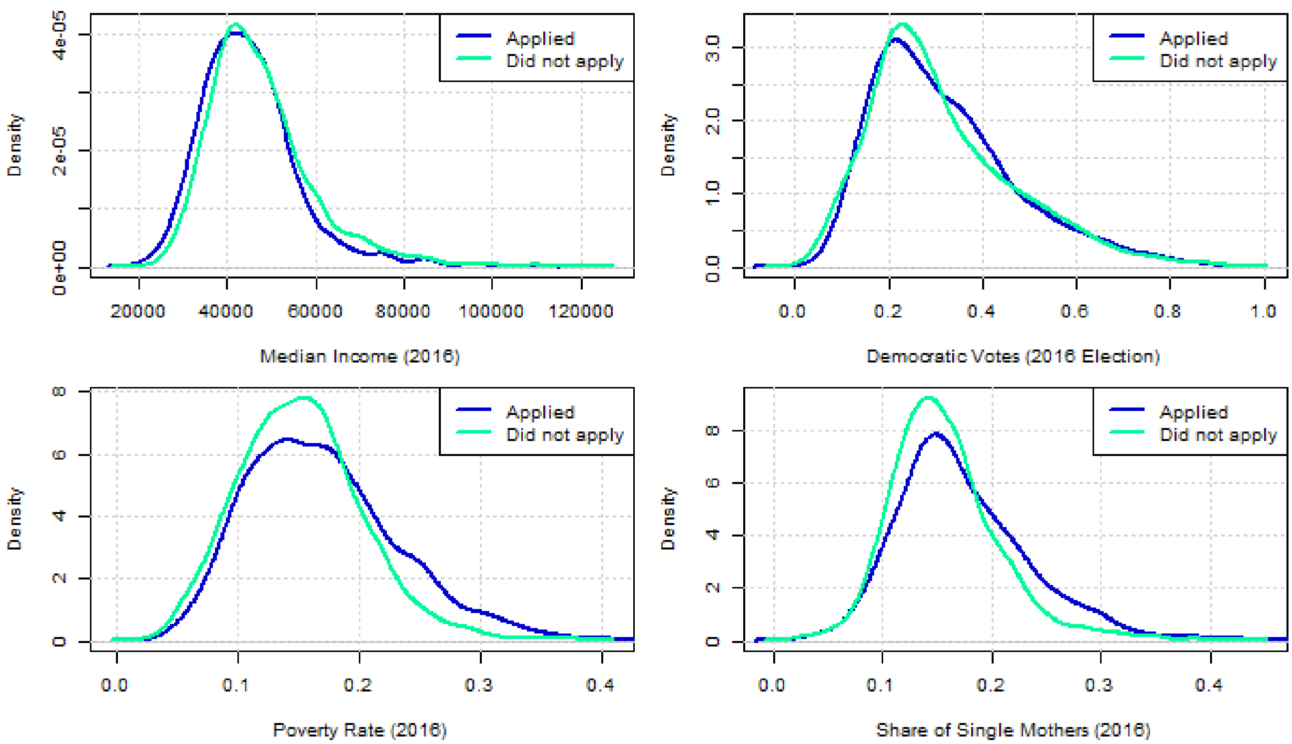


Figure 4: Kernel density estimates by application status

## 3 Empirical Strategy

### 3.1 Setting

We employ an event study design to measure the effect of natural disasters on standardized test outcomes. An event study design is a staggered adoption design where units are first-treated at different points in time, and there may or may not be never-treated units (Sun and Abraham, 2021).

In order to identify a causal effect, unobservable determinants of a county's test scores must be unrelated to natural disasters conditional on observable characteristics of that county. The occurrence of natural disasters is plausibly random conditional on location. Furthermore conditioning on the year should account for an increasing trend in natural disasters due to climate change. Thus, independence of mean test scores and natural disasters is plausible conditional on county and year fixed effects.

Consequently, the baseline specification is

$$y_{i,t,g} = \sum_{\tau=-9,\tau\neq-1}^{9} \beta_\tau D_{i,t-\tau} + \alpha_i + \lambda_t + \zeta_g + X_{i,t}\gamma + \varepsilon_{i,t,g} \ , \tag{1}$$

where $y_{i,t,g}$ is the outcome of interest for county $i$, year $t$, and grade $g$. County, year, and grade fixed-effects are given by $\alpha_i$, $\lambda_t$, and $\zeta_g$ respectively and $X_{i,t}$ is a row vector of additional control variables. $D_{i,t-\tau}$ is a treatment indicator for county $i$ in year $t-\tau$. That is, $D_{i,t-\tau} = 1$ if the county had already experienced a disaster $\tau$ years ago at time $t$. Since we consider the time period 2009-2018, $-9 \leq \tau \leq 9$.

Note that treatment must be absorbing, meaning the sequence $(D_{i,t})_{t=1}^{T}$ must be a non-decreasing sequence of 0s and 1s. In other words, after being treated for the first time a county stays treated. In the present application this means treatment refers to having experienced a disaster rather than experiencing a disaster in that year. This is common practice and does not cause bias due to the conditionally random nature of natural disasters (Deryugina, 2017). Thus, the emphasis lies on cumulative long-term effects rather than instantaneous short-term effects.

It is implausible that the treatment effects are constant in our setting. The extent of disasters varies substantially, and also the level of preparation for such disasters likely displays high variance across counties.

### 3.2 Interaction-weighted estimator

We utilize the interaction-weighted (IW) estimator proposed by Sun and Abraham (2021) that is robust to treatment effects heterogeneity. The main interest lies on the cohort average treatment effect on the treated (CATT),

$$CATT_{e,\tau} := \mathbb{E}\left[Y_{i,t+\tau} - Y_{i,t+\tau}^{\infty} | E_i = e\right] ,$$

where $Y_{i,t+\tau}^{\infty}$ is the counterfactual of being never treated and $E_i$ denotes the first treatment period. Thus, $CATT_{e,\tau}$ is the average treatment effect on the treated $\tau$ years after being treated for the first time for the cohort that was first treated in year $e$.

The estimation procedure consists of three main steps:

1. Estimate $CATT_{e,\tau}$ using a linear fixed effects specification with interactions between relative period indicators and cohort indicators:

$$y_{i,t,g} = \sum_{e\notin C} \sum_{\tau\neq-1} \delta_{e,\tau} (\mathbb{1}\{E_i = e\} D_{i,t-\tau}) + \alpha_i + \lambda_t + \zeta_g + \varepsilon_{i,t,g} \ , \tag{2}$$

where $C$ is the set of comparison cohorts. In our case $C$ is the never treated cohort, i.e. $C = \infty$. If there is a cohort that is always treated, i.e. that already receives treatment in the first period, then we need to exclude this cohort. The coeffiecient estimator $\widehat{\delta}_{e,\tau}$ that we obtain from (2) estimates $CATT_{e,\tau}$.

2. Weight the estimators by the share of the respective cohort in the sample in that period. Let $\hat{W}^{\tau}$ be a weight matrix with element $(t, e)$

$$[\widehat{W^{\tau}}]_{t,e} := \frac{\mathbb{1}\{t - e = \tau\} \sum_{i=1}^{N} \mathbb{1}\{E_i = e\}}{\sum_{e \in h^{\tau}} \sum_{i=1}^{N} \mathbb{1}\{E_i = e\}},$$

   where $h^{\tau} := \{e : 1 - \tau \leq e \leq \max(E_i) - 1 - \tau\}$ is the set of cohorts that experience at least $\tau$ periods of treatment.

3. Take the average over all $CATT_{e,\tau}$ estimates weighted by the cohort shares in the weight matrices. Let $vec(A)$ be the vectorize operator that vectorizes matrix $A$ by stacking its columns and let $\widehat{\delta}$ be the vector that collects $\widehat{\delta}_{e,\tau}$ for all $e$ and $\tau$. Then, the IW estimator $\widehat{v}_g$ for bin $g$ can be written as

$$\widehat{v}_g := \frac{1}{|g|} \sum_{\tau \in g} [vec(\widehat{W^{\tau}})]^{\intercal} \widehat{\delta}. \tag{3}$$

   For a singleton bin $g = \{\tau\}$, this simplifies to

$$\widehat{v}_g := [vec(\widehat{W^{\tau}})]^{\intercal} \widehat{\delta}.$$

Under some standard assumptions, $\widehat{v}_g$ is asymptotically normal (for a proof and a detailed description of said assumptions see Sun and Abraham, 2021, Appendix C). Under the additional assumptions of parallel trends and no anticipatory behavior, $\widehat{v}_g$ is consistent, that is it converges in probability to

$$\widehat{v}_g \xrightarrow{p} [vec(W^{\tau})]^{\intercal} \delta = \sum_{e \in h^{\tau}} \mathbb{P}(E_i = e | E_i \in h^{\tau}) CATT_{e,\tau} \ ,$$

where $W^{\tau}$ is the probability limit of the weight matrix $\widehat{W}^{\tau}$.

Under quasi-random treatment assignment, the parallel trends assumption is likely justified (Roth et al., 2022).

## 4 Results

Figure 5 shows the dynamic treatment effects.

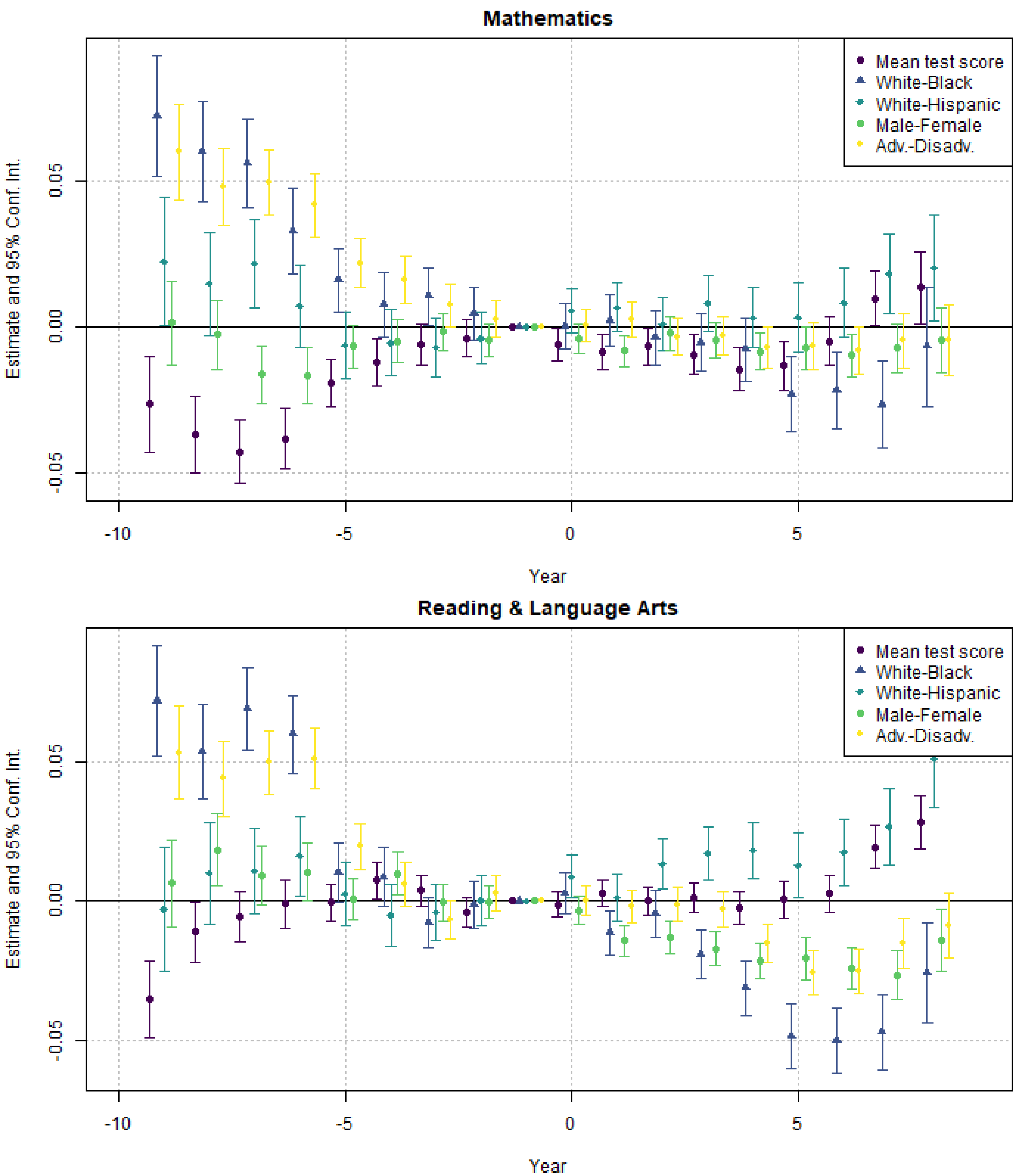


Figure 5: Dynamic Treatment effects by dependent variable and subject

## 5 Conclusion

## References


Deryugina, T. (2017). The fiscal cost of hurricanes: Disaster aid versus social insurance. *American Economic Journal: Economic Policy*, 9(3):168–98.

Reardon, S., Kalogrides, D., Ho, A., Shear, B., Fahle, E., Jang, H., and Chavez, B. (2021). Stanford education data archive (version 4.1).

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. (2022). What's trending in difference-in-differences? a synthesis of the recent econometrics literature.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199. Themed Issue: Treatment Effect 1.

Turner, D. (2022). rfema: Access the openfema api. *rOpenSci*.