TV Identities: TITLE*

Andrew Kao†

January 2020

Abstract

Here's an abstract

*We would like to thank XYZ.

†University of Chicago, andrewkao@uchicago.edu.

1 Introduction

Mass media lets us know what the outside world thinks, and this shapes the way that we think.

- Media plays a large role in shaping our lives
- Latino consumption of broadcast TV remains relevant
- Relevant subquestion: how identity is affected

Three domains

- Education
- Firms
- Politics

The high level research question is to look at the effect of reinforcing identity within Hispanic populations on their schooling outcomes. Specifically, I'll be using the influence of Spanish language television as the channel by which identity is reinforced, and look at how it affects everything from graduation rates to disciplinary action taken to math abilities and English proficiency for Hispanic students in public schools. In short, if I have access to more programming from my home country, does this make me less engaged in school (perhaps because there are more distractions or because it socially ostracizes me etc.), or does this make me perform better (perhaps because l have more role models or because I have something to talk with peers about in school, and hence motivation to attend/perform)?

There's good reason to believe that identity, as reinforced through mass media, has a large effect on the lives people lead. Oberholzer-Gee, Waldfogel (AER 2009) demonstrate that the presence of Spanish language local news increases Hispanic voter turnout, while Yanigazawa-Drott (QJE 2014) shows that radio broadcasts in Rwanda contributed to the violence and genocide that took place in the 90s. It would be reasonable to think then, that there could be a meaningful effect of Spanish language TV on education.

2 Data

Overall data and brief explanation of sources

Data for the instrument comes from both the FCC and TMS (a telecommunications company that was kind enough to let me use their API for free), and the instrument is fully constructed. The relevant data here is essentially just the coverage contour spatial data and the broadcast language of the station.

The data on public schools comes from the US government's CRDC (Civil Rights Data Collection) dataset. It's a very large dataset with a ton of outcome/control variables, but importantly, it breaks down all major variables of interest by ethnicity. These variables includes graduation rates, chronic absenteeism, suspensions, expulsions, arrests, bullying, AP test results, English proficiency,

math class performances, gifted program enrolment etc., so I can look at effects on both the top and bottom end of the distribution, and examine potential mechanisms driving outcomes. These are all at the school level, and so I can run this through ArcGIS to get physical locations of these schools as well.

I get additional controls for things like population, income, density of Hispanic population etc. at the county level from IPUMS.

2.1 Broadcast TV

2.2 Outcomes

Table 1: School-District Level Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Distance to Boundary	17,280	136.855	146.751	0.000	15.786	217.567	806.543
SLTV Coverage Dummy	17,280	0.292	0.455	0.000	0.000	1.000	1.000
% County Hispanic	17,280	7.051	11.950	0.000	0.668	6.974	97.216
Log(Population)	17,280	11.618	1.840	5.869	10.242	13.110	15.997
Log(Income)	17,280	9.428	0.257	7.976	9.257	9.593	10.245

Note: Distance to SLTV Boundary measured in KM

Table 2: School Level Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Total Students	96,349	524.859	449.354	2.000	254.000	662.000	14,164.000
# Hispanic Students	91,019	143.195	243.873	2.000	13.000	166.000	7,675.000
Contains Grade 1	96,350	0.538	0.499	0	0	1	1
Contains Grade 6	96,350	0.364	0.481	0	0	1	1
Contains Grade 9	96,350	0.253	0.435	0	0	1	1
Hispanic Suspension Dummy	94,535	0.382	0.486	0.000	0.000	1.000	1.000
Hispanic Chronic Absentees	94,540	22.920	57.838	0.000	0.000	22.000	2,131.000
# Teachers	93,934	35.219	33.892	1.000	19.000	44.000	6,031.000

Note: Dummies indicate whether event occurred in the school over the past year

3 Empirical Strategy

To isolate the causal effect of Spanish language television, I adopt the technique used in [**?**] Newman, Velez (AJPS 2019) and generalize it from three counties to the entirety of the US. Newman and Velez

exploit a FCC (Federal Communications Commission) regulation which determines the distance from a TV station in which the station's broadcast signal is protected from interference. This creates a natural regression discontinuity, where the decaying strength of a signal over distance is combined with this cutoff in broadcast protection to create a split among people just inside and outside these coverage 'contours' that are presumably comparable save for their access to broadcast TV.

In the case of Spanish language TV in particular, this should allow me to examine its causal effect on Hispanic populations for spatially located outcomes, such as public schooling results. It's worth noting that these contours are purely determined by an algorithm that looks at things like local elevation and antennae strength, so that the cutoffs are located in more or less random locations, and that coverage is large enough that these contours tend to cut across towns and suburbs, rather than cities. Finally, regressions using US census data indicate that Hispanic people do not migrate across counties in response to these contours.

A standard regression thus looks like restricting the universe of schools to only those within a small radius of the contour boundary, where the key independent variable of interest is an indicator for the school being inside or outside the boundary.

3.1 Main Specification

A standard regression thus looks like restricting the universe of schools to only those within a small radius of the contour boundary, where the key independent variable of interest is an indicator for the school being inside or outside the boundary, interacted with the distance to the boundary:

$$Y_i^{j,k} = \beta_0 + \beta \mathbb{I}[InsideContour_i] \times Distance_i + \gamma X_i + \delta Z^j + \epsilon_i^k \quad \epsilon \overset{iid}{\sim} N(0, \sigma_i^{k^2})$$

where Y_i is an outcome for school i in county j and school district k, X is a vector of school-level controls, and Z is a vector of county-level controls. Errors are often clustered by school district, meaning that $Corr(\sigma_i^k, \sigma_{i'}^k) \neq 0$ is permissible.

By limiting the analysis to a small distance from the contour boundary (100 KM/63 miles by default), we also minimize the potential concerns of omitted variable bias etc., as these schools must now be at least fairly close to one another, meaning that they probably share many overarching characteristics.

3.2 Migration

4 Public Schools

4.1 Data

4.2 Results

4.3 Discussion

Evidence of Mechanism Targeting based on identity

5 Firms

5.1 Data

5.2 Results

5.3 Discussion

6 Campaign Contributions

6.1 Data

6.2 Results

Wave 1: Intervention Before the Election.

6.3 Discussion

7 Conclusion

abc

Figures and Tables

Figure 1: The Coverage Contours of Spanish Language TV stations

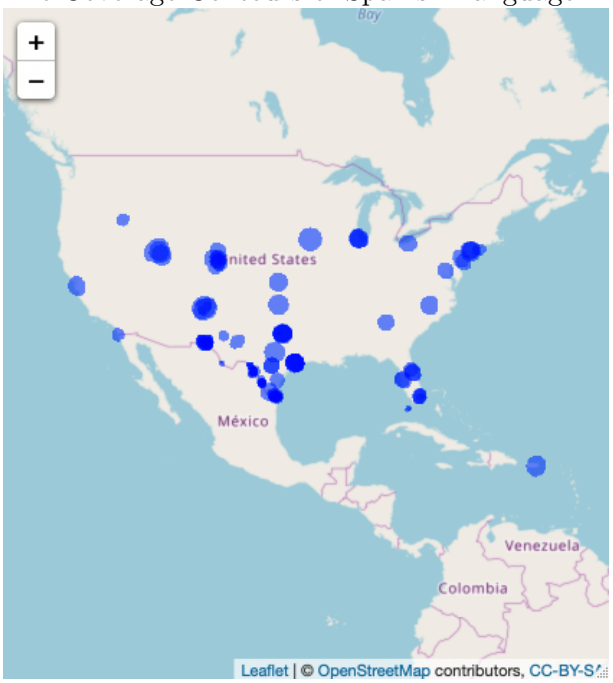

Figure 2: Map of School Districts in the US