

Seeing is Believing: Inequality and the Impact of Television on Hispanics*

Andrew Kao[†]

February 2020

Abstract

This paper investigates the diverse impacts of Spanish Language Television (SLTV) on Hispanic communities. By exploiting a FCC regulation concerning broadcast signal protection, I estimate the impact of SLTV on closing inequality in businesses and schools via a spatial regression discontinuity. The results show that SLTV creates more Hispanic owned firms and Hispanic named firms, and induces higher educational performance and fewer disciplinary incidents among Hispanic students. I provide evidence that suggests a strengthening of identity drives these results.

*Thanks to Victor Lima and Kotaro Yoshida for overseeing the honors thesis workshop, to Leonardo Bursztyn for advice and guidance, and to Trip from SatelliteGuys for technical advice.

[†]University of Chicago, andrewkao@uchicago.edu.

1 Introduction

[Television] has altered every phase of the American vision and identity.

- Marshall McLuhan, *War and Peace in the Global Village*

We spend our waking hours inundated by mass media. The internet, newspapers, radio, and television stations broadcast to us a constant stream of facts, beliefs, entertainment, and ideas—ideas that percolate in our heads, and ideas that, I argue, shape our own identities and the way we live our lives. In this paper, I examine the impact of Spanish Language Television (SLTV) on Hispanic communities, looking at how it might reduce inequality in the domains of entrepreneurship and education. In each case, I find that SLTV induces changes in behavior can be attributed, at least in part, to a stronger sense of identity; causality is established via a spatial regression discontinuity introduced by a regulation on television signal protection.

These two major results are highlighted in Figures 1, 2, corresponding to the two domains of analysis. Figure 1 shows that the number of Hispanic owned firms with Hispanic names sharply increases under the presence of television (distances above 0), while Figure 2 shows the decline in Hispanic students suspended when in communities with SLTV access.

There's good reason to believe that mass media has a large effect on the lives people lead. Olken (2009) finds that that radio and television decrease 'social capital' in Indonesia, in line with Putnam (2001)'s argument that mass media is socially isolating. Yanagizawa-Drott (2014) shows that radio broadcasts in Rwanda contributed to the violence and genocide that took place in the 90s. DellaVigna and Kaplan (2007) find that the introduction of Fox News induced gains in Republican voteshares, while other work establishes the link between media and gender norms, media and anti-Americanism, and media and fertility.¹

We focus on Hispanic viewership of television for several reasons. First, Hispanics face substantial inequality as both entrepreneurs and students. They open 33% fewer businesses than would be expected from their proportion of the population and command the fewest business loans from financial institutions (they secure 35% less than white-owned businesses according to Orozco, Oyer and Porras (2017)). They have the lowest high school graduation rates (they are 23% less likely to complete high school than whites) and the lowest college completion rates too (they are 60% less likely to complete college than whites according to Tienda (2009)).

Second, Hispanics consume substantial amounts of television—out of the 115 million households with television in the United States, there are 14 million Hispanic ones, proportional to the overall fraction of households. Half of these Hispanic television households get their television content via satellite or broadcast television, substantially larger than the 30% national average (FCC (2016a),

¹ Papers that investigate the means by which media affects identity and action include Jensen and Oster (2009), Gentzkow and Shapiro (2004), Ferrara, Chong and Duryea (2012), and Kearney and Levine (2015). For an overview, see DellaVigna and La Ferrara (2015).

De La Merced and Gelles (2014))—important because the instrument I use only affects access of satellite and broadcast TV.

Spanish Language Television, in particular, allows us to take a closer look at Hispanic communities and examine its ties to identity. 78% of Spanish-dominant households watch SLTV, while 50% of multi-language Spanish-speaking homes do. Notably, 85% of SLTV viewership occurs over broadcast television; in 2010, the top 10 broadcast shows in the Hispanic demographic category were all Spanish language programs (Pardo and Dreas, 2011). Broadcast television, and Spanish Language television in particular, remains an important institution in Hispanic households.

Compared to other viewers of television, Hispanics are uniquely likely to watch television in a social context rather than watching alone—this is partially driven by the fact that non-Hispanic households have 40% more TV sets per person than Hispanic ones (Coghill and McGinnis, 2018). This social aspect, wherein SLTV is watched with family/friends (or people that speak Spanish), may be one way in which identity is reinforced through television.

SLTV programming is also more likely to contain content that is directly salient to a Hispanic person’s identity. This occurs not only because of the language of the broadcast, but also its content: roughly 20% of programming on SLTVs are telenovelas produced in foreign (Latin American) countries, with a similar proportion of programming dedicated to non-locally produced news and paid programming, which may come from abroad as well.²

When explicitly looking at the effect of television on Hispanic communities, Oberholzer-Gee and Waldfogel (2009) demonstrate that the presence of Spanish language local news increases Hispanic voter turnout, whereas Velez and Newman (2019), who first developed the instrument used in this paper, show that SLTV leads to lower voter turnout rates. Trujillo and Paluck (2012) run an experiment testing trust in the government and the census based on a soap opera scene. I extend on this literature by expanding the scope of research to a national-level analysis while attaining more granularity with spatial microdata, and also by exploring outcomes beyond the political realm.

The central instrument utilized in this paper is the spatial regression discontinuity in television coverage contours introduced via a FCC regulation guaranteeing TV station signal protection only within a certain distance of a station’s main antenna. Households just inside the coverage contour are able to receive broadcast and/or satellite TV coverage, whereas those just outside oftentimes cannot, and so the regression discontinuity keeps those just inside and outside the coverage boundary (usually defined as those within 100 KM of the boundary) for comparability between observations. This, combined with the fact that television signals grow weaker with greater distance from a station, lends credibility to the discontinuity actually demarcating a difference between those with and without access to SLTV.

I argue that this allows us to identify the causal effect of SLTV, given several factors: (1) con-

² The statistics come from FCC (2016a), but unfortunately, the dataset they use do not allow them to precisely determine from whence the programming originated.

tours are mechanically decided by a formula involving geographical features and antenna strength, (2) contours are large and tend to cut across suburban and ‘small town’ areas, rather than dense urban areas which corporations might try to include for profitability reasons (urban centers fall squarely within contours), (3) SLTV stations were often built before the regulation was imposed, (4) demographic and other controls across the regression discontinuity look similar, making it plausible that there are not external factors driving the differences observed, and (5) Hispanics do not appear to migrate across contours in either direction, minimizing the probability of effects being driven by selection.

The first major outcome I examine looks at Hispanic firm ownership and Hispanic named firms. A small literature has investigated the impact of television on financial outcomes. Bjorvatn et al. (2019) find that television imparts better financial knowledge, while Berg and Zia (2017) find that television does not foster more or better entrepreneurship, but does lead to more students dropping out of schools. While these studies rely on the randomization of individual television episodes or programs to study their effects, I extend the analysis to the overall effect and presence of television, and find that SLTV actually leads to higher rates of Hispanic firm ownership and firms named in ways that draw on a Hispanic identity.

Another segment of the entrepreneurship literature looks at the overall ways in which entrepreneurship might be better fostered. The results tend to be bleak: Karlan and Valdivia (2011) find that explicit training in entrepreneurship has minimal impact on the overall success of entrepreneurs, while Gin and Mansuri (2014) show that entrepreneurs tend not to be cash constrained, but rather idea constrained. I thus provide some initial evidence on how television might be tied to business formation, and conclude that approximately 10% more Hispanic-owned businesses are formed under the influence of television.

On the education side, there is substantial controversy surrounding the impact of television on students, with the mainstream line of reasoning contending that television serves as a distraction which ‘rots’ the mind (Zavodny, 2006).³ On the other hand, there has been some pushback claiming that individual shows can lead to fewer teenage births (Kearney and Levine, 2015), or that specific audiences can attain greater educational outcomes (Gentzkow and Shapiro, 2008*b*). We find results that broadly align with the latter, showing that Hispanic enrolment in gifted programs and AP classes increases in response to the presence of SLTV, while instances of out of school suspension and chronic absenteeism decrease. However, some of our findings do cut against this narrative: the number of Hispanic students placed into Limited English Proficiency programs and the number of Hispanic students bullied or harassed on the basis of their ethnicity both increase—we are inclined to tie attribute this to a stronger sense of ethnic identity.

There has not been much work done by economists looking at the performance of Hispanic students in schools in particular, though Cascio and Lewis (2012) provides a look at these outcomes

³ Winn (2002) and Gentile (2004) make similar arguments.

in relation to schooling and social segregation in California. Through a comprehensive look at the outcome of Hispanic students in the US public school system, I provide additional color to this area of study at a national level, showing that Hispanic students are roughly 5% more likely to take and pass more advanced coursework and 10% less likely to earn disciplinary infractions under the influence of SLTV.

There is a growing literature that looks at how identity can be a mechanism driving behavior; this has been studied in both lab environments⁴ and more organic settings too (Bursztyn et al., 2015). However, the underlying factors that strengthen identity in the first place (rather than simply triggering them via priming or other short-term interventions) is less well understood. Bisin et al. (2010), Atkin, Colson-Sihra and Shayo (2019), and Bazzi et al. (2019) encompass some reason studies on this topic, and all come to the conclusion that intergroup tensions or differences lead to a strengthening of identity. Earlier work includes ?, who take the very long view and show how gender norms can be traced back to early agricultural practices. With our work on Hispanic firm names and harassment based on Hispanic identity, I provide an alternate, media-based look at how identity may be strengthened.

Thus, to summarize the key contributions of the paper: I open the door to studying the effect of media on Hispanic communities in ways other than the political, and in tackling these questions, use a national natural experiment that is larger in scope than most of the extant literature, while simultaneously making use of geocoded microdata to provide a more precise look at the underlying effects. I also provide an additional bridge into the existing literature on identity, showing how media might bolster and strengthen it.

The rest of the paper is structured as follows: Section 2 presents the data used across later sections, Section 3 addresses the empirical strategy. The following two sections, Sections 4 – 5, present data, results, and discussion on the results for our analysis on firms and schools respectively. Finally, Section 6 concludes.

2 Data

2.1 Broadcast TV and Geography

The central instrument in this paper is the discontinuity in coverage contours of SLTV stations introduced via FCC regulation.

Coverage Contours To build the coverage contours of SLTV stations (and thus find the boundaries across which people just receive/do not receive SLTV) in the US, we collected a list of the callsigns for all SLTV stations via the TMS API (TMS is a large provider of data on TV, movies,

⁴ See Benjamin, Choi and Strickland (2007) or Benjamin, Choi and Fisher (2010).

and other media).⁵ There are 100 of these stations located across the United States. These callsigns were then matched against data from the FCC’s OET Bulletin 69 and the FCC’s CDBS Database to directly obtain the relevant coverage contour boundaries as prescribed and regulated by the FCC.⁶ A map of all these contours can be seen in Figure 4.

Geocoding Location data for all outcomes was collected in the form of addresses written in text. To transform this into proper spatial data (coordinates with latitude & longitude), two geocoding tools were used: (1) ArcGIS, which has its own proprietary database of locations. Over 99% of addresses were successfully matched to one location and geocoded. This was used to geocode the schooling data, as well as portions of the campaign contribution data. (2) The US Census Geocoder, which contains the census database of locations. Over 80% of addresses were successfully matched to one location and geocoded.⁷ This was used to geocode the business data, as well as portions of the campaign contribution data. It is unlikely for non-geocoded addresses to be correlated with the instrument, given the relatively narrow band around the contour retained for the spatial regression discontinuity.

For data that take the form of spatial points (such as the location of a school), determining its distance to the boundary and whether the datapoint falls within the coverage boundary is a straightforward process. For data aggregated into grid points (typically a grid composed of 2×2 KM² chunks), we treat each any degree of intersection as the grid point falling within coverage, given its relatively small size. In locations covered by multiple SLTV stations, the distance to the boundary is taken as the distance to the closest boundary.

2.2 Controls and Other Non-Outcome Data

Controls at the county level are sourced from IPUMS and consist of basic relevant demographic information: population, income, percent of county that is Hispanic etc. County level data is mapped to its relevant location using census data as well.

Data on migration comes from the 2011-2015 American Community Survey (ACS), which reports the number of people moving from each origin county to destination county (aggregated over the four years).⁸ This sample also contains migration flows by Hispanic origin, allowing us to determine whether they move based on geographic boundaries.

Finally, data attached to specific outcomes are discussed under their relevant section.

⁵ A TV station is defined to be SLTV if at least one of the primary broadcasts languages are Spanish.

⁶ 2015 coverage contour data is used due to the ‘FCC Spectrum Repack’ that began in 2018, which relocates a number of signals, affecting the reception and coverage for a substantial number of stations (Fletcher, Heald and Hildreth, 2018).

⁷ The US Census geocoder, unlike the ArcGIS geocoder, is free. However, due to the higher precision of the ArcGIS geocoder, data constructed from it is used wherever possible.

⁸ Historically, approximately 15% of the ACS migration data has been allocated, or imputed based on salient characteristics (United States Census Bureau (2020a)).

3 Empirical Strategy

To isolate the causal effect of Spanish language television, I adopt the technique used in Velez and Newman (2019) and generalize it from two counties to the entirety of the US.⁹ Velez and Newman 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.

Digital and satellite TV stations operate by broadcasting signals from a central antenna, and the field strength at a given point resulting from this antenna is a mechanical product of several factors: The antenna’s ERP (Effective Radiated Power, which is the amount of input power given to the antenna adjusted for idiosyncrasies in the antenna that may boost or attenuate the effective power), the antenna’s HAAT (High Above Average Terrain), and the distance from the point to the antenna.

This signal declines in strength as one grows more distant from the station, making it subject to interference. The FCC regulation OET Bulletin No. 69 (FCC, (2004a)) protects signals for commercial TV stations from interference in a contour area for which service holds at 50% of locations 90% of the time.¹⁰ An example of this coverage contour can be seen in Figure 3; note that they tend to be sizable enough to fully cover major metropolitan areas, with contours boundaries ending substantially beyond them.

This creates a natural spatial regression discontinuity, where the decaying strength of a signal due to distance from a station 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. This minimizes the potential concern of omitted variable bias, as the groups we are comparing across this border should share many overarching characteristics.

In the case of Spanish Language TV in particular, this should allow us to examine its causal effect on Hispanic communities for spatially located outcomes. As mentioned, these contours are purely determined by an algorithm and only dependent on physical variables like local elevation and antennae strength. Thus, the precise regulatory boundaries are located in more or less random locations, and coverage is large enough that these contours tend to cut across towns and suburbs, rather than large cities — television networks are not constructing their antennas to be just large enough to only cover the most dense and populous areas. This implies that network executives, if they are aiming to maximize profit, ratings, or audiences, would not have these boundaries at the forefront of their calculus.

In order for the causal effect of SLTV to be identified, the actual coverage of the contours must be uncorrelated with any of the other determinants for the outcome variables with which

⁹ The paper was retracted in 2019, but this was due to usage of unauthorized data, and unrelated to the efficacy of the underlying identification strategy.

¹⁰ These contour lines are termed (50,90) lines. There is a small adjustment made for different channel numbers, which have varying noise-limited coverage.

we are interested. One reassurance is that the interference protection regulation, OET Bulletin 69, was only codified in 1977 — in contrast, Univision, the largest owner of SLTV stations, was founded in 1955, and had built a substantial number of their television stations and antennas by 1977.¹¹ Furthermore, the most recent Longley-Rice methodology used to determine TV service coverage was only adopted in 1997, making it even less likely that stations were built or adapted in response to the policy.¹² Nonetheless, one may be concerned that SLTV stations target areas with more Hispanic people, or wealthier communities, or more populous areas, all of which are factors that could affect the areas of interest. Hence, I include explicit controls for these variables in the regression.

The instrument therefore consists of two variables interacted: First, a dummy for whether the outcome data falls within a SLTV station’s coverage contour boundaries, and second, the distance from the outcome of interest to the closest coverage boundary. To guarantee similarity between the people inside and outside the boundaries, only data points located within a distance of 100 KM of the boundary are kept.¹³

Several concerns that potentially remain:

- *Can we guarantee that it is Hispanic people who watch SLTV?* If it were the case that non-Hispanic people were frequent viewers of SLTV, the interpretation of the main effects would potentially be different: we would be looking at the effect of SLTV on all people. Thus, though outcomes restrict the analysis to how the lives of Hispanic people change, this could be driven by, for instance, white people treating Hispanic people differently due to having viewed SLTV. However, only 4% of total SLTV station programming watched can be attributed to non-Hispanic people, a number that is only as high as it is because some SLTV stations also broadcast in English (FCC (2016a)). Similarly, < 1% of all programming watched by non-Hispanics is in the Spanish language.
- *How do we account for the possibility of selection?* It is theoretically possible that Hispanic people move in response to these television coverage contour boundaries, and that the effects seen are therefore a result of Hispanic people self-sorting. If this were true, it would be a fairly remarkable result—people moving in significant quantities for access to better television in a way that influences life outcomes ranging from education to business to politics. However, as the subsection on migration beneath demonstrates, the selection story does not appear to be borne out by the data.

¹¹ Though Telemundo, the second largest owner of SLTV, was technically founded in 1984, the stations it initially acquired were built in 1954. It also primarily expanded through the acquisition of existing stations, rather than building out its own new ones.

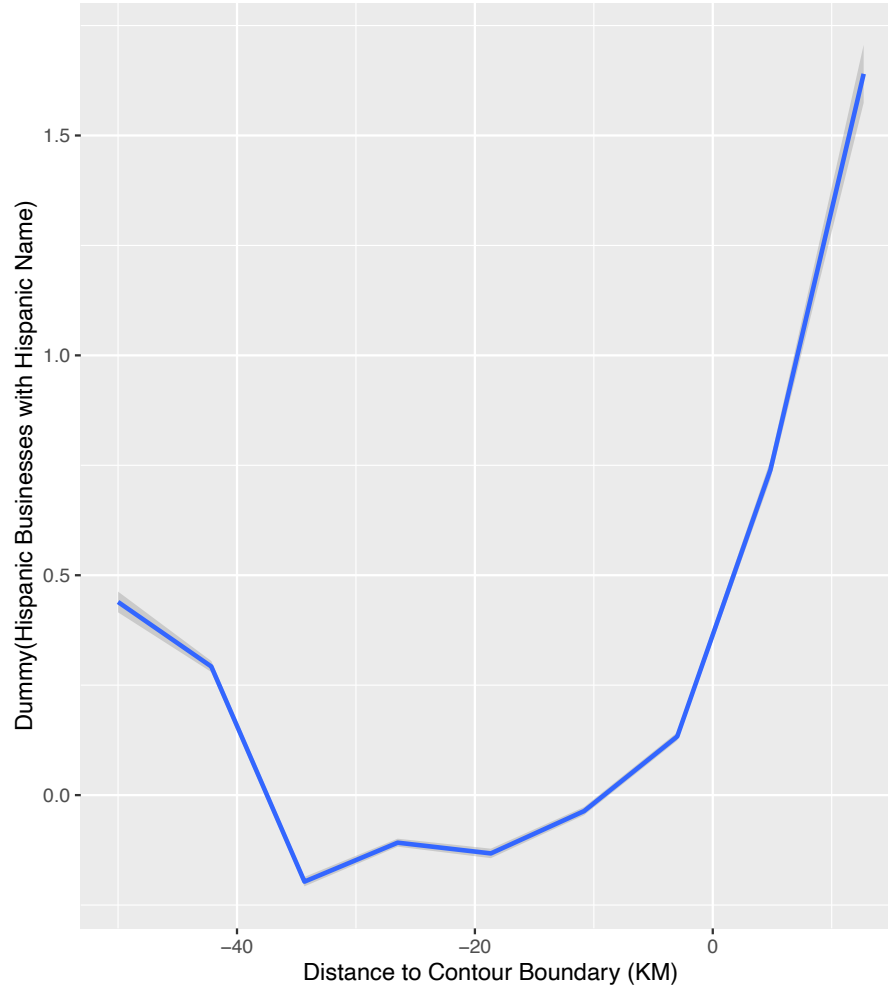
¹² See FCC (2004a) for details.

¹³ Using a round number in kilometers rather than miles makes the cutoff less likely to be correlated with some real-world phenomena.

Figures and Tables

6.1 Figures

Figure 1: Dummy for Hispanic Owned Business with Hispanic Name by Distance to Contour Boundary



Notes: The figure presents data at the firm level, where a smoothed average of a residualized dummy for Hispanic businesses with Hispanic-indicating names is plotted against the distance of the school to the closest Spanish Language Television station contour boundary. Positive distances denote schools that are located within the boundary, while negative distances denote schools outside of them. Controls at the county level include log population, income, and percentage population Hispanic.

Table 5: Influence of Spanish Language Television on Hispanic Business Ownership

	<i>IHS(# Hispanic Owned Businesses)</i>			
	(1)	(2)	(3)	(4)
TV Dummy	0.261*** (0.014)	0.122*** (0.014)	0.112*** (0.014)	0.132*** (0.015)
TV Dummy \times Distance to Boundary	0.010*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
Distance to Boundary (meters)	0.006*** (0.001)	0.009*** (0.001)	0.010*** (0.001)	0.011*** (0.001)
Log(Population)		0.412*** (0.011)	0.388*** (0.012)	0.342*** (0.014)
County % Hispanic			1.261*** (0.133)	1.414*** (0.136)
Log(Income)				0.391*** (0.070)
Observations	23,853	23,853	23,853	23,853

Notes: The table presents coefficient estimates from regressions at aggregated into grids of size approximately 4 KM², only keeping grid points within 100 KM of a contour boundary. The dependent variable is the inverse hyperbolic sine transformed counts of Hispanic owned firms within the grid. The key independent variable of interest is the TV Dummy, which tracks whether the school is within a coverage contour boundary for a Spanish language television station. This is interacted with the distance to the boundary. Controls are at the county level. Standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Influence of Spanish Language Television on Businesses with Hispanic Names

	<i>Hispanic Named Business Dummy</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
TV Dummy	0.839*** (0.052)	0.638*** (0.066)	0.637*** (0.066)	0.769*** (0.071)	0.849*** (0.077)	0.775*** (0.071)
TV Dummy \times Distance to Boundary	0.008*** (0.002)	0.002 (0.002)	0.002 (0.002)	0.0002 (0.002)	-0.0002 (0.002)	0.0002 (0.002)
Distance to Boundary (meters)	0.010** (0.004)	0.021*** (0.004)	0.021*** (0.005)	0.031*** (0.005)	0.035*** (0.005)	0.031*** (0.005)
Log(Population)		0.957*** (0.052)	0.979*** (0.070)	0.702*** (0.074)	0.761*** (0.081)	0.701*** (0.074)
County % Hispanic			-0.151 (0.312)	1.428*** (0.367)	1.514*** (0.388)	1.434*** (0.368)
Log(Income)				2.350*** (0.319)	2.534*** (0.344)	2.356*** (0.320)
Observations	23,853	23,853	23,853	23,853	23,853	23,853
Only Hispanic Owners	No	No	No	No	Yes	No
Only Non-Hispanic Owners	No	No	No	No	No	Yes

Notes: The table presents coefficient estimates from logit regressions at aggregated into grids of size approximately 4 KM², only keeping grid points within 100 KM of a contour boundary. The dependent variable is a dummy for whether there is a firm with a Hispanic name within the grid. The key independent variable of interest is the TV Dummy, which tracks whether the school is within a coverage contour boundary for a Spanish language television station. This is interacted with the distance to the boundary. Controls are at the county level. Standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.