Effect of High Information Environments on Voter Turnout

Marc Trussler
Political Science Department
Vanderbilt University
marc.j.trussler@vanderbilt.edu
V 4.0 Comments Welcome

October 31, 2016

Abstract: Political information becoming cheaper and more readily accessible has been heralded as a boon for democracy. Many have hoped that these changes will lead to higher political knowledge and, in turn, more participation in the political process. However, changes in the information environment have also increased the amount of entertainment options available, reducing the amount of political information actually consumed. Estimating the effect of a changing information environment is difficult due to the bias in individual self-reports of media usage related to over-reporting by those who are better educated and have higher political interest. I circumvent these problems by focusing instead on access to information, using within-county variation in the roll-out of broadband in the early 2000s. This conditionally exogenous variation in broadband indicates that increases in the availability of information has an average negative effect on turnout measured at both the aggregate and individual levels. This average effect hides substantial heterogeneity, where more motivated individuals receive substantially more positive effects from increases in the availability of information.

Paper prepared for the Toronto Political Behavior Workshop: November 11-12, 2016. Thanks to Larry Bartels, Josh Clinton, and Cindy Kam for helpful advice. Also to the RIPS group and graduate-student Journal Club at Vanderbilt for suffering through early presentations of these ideas.

Political knowledge and participation are closely related. In their work on the role of political knowledge in American democracy Delli Carpini and Keeter (1997: 244) state: "'Political knowledge boosts participation because it promotes an understanding of why politics is relevant." It has never been less costly for citizens to gain the information they need to feel confident turning out to vote; as such, it is tempting to link some of the gains in participation witnessed in the early 2000s to the lower costs of obtaining information. However, changes in the information environment have also increased the opportunity cost of accessing political information by offering an expanded choice-set of entertainment options. If so, turnout *decreases* when information becomes more readily accessible. I examine this question of the effect of high information environments on voter turnout.

Individual level analyses of the relationship between information access and political participation have found positive effects on average (for a meta-analysis see: Boulianne 2009), but there are reasons to be wary. Individual self-reports of media use are biased by over-reporting, particularly by those who are better educated and have higher political interest (Bartels 1993; Dilliplane et al. 2013; Prior 2009a, b). I avoid this difficulty in measuring media exposure by focusing instead on access to information. Using FCC data on the roll-out of high-speed broadband internet, I estimate within-county estimates of the effect of high-information environments on voting rates at both the aggregate and individual levels. By focusing only on within-county variation while controlling for a flexible time trend, changes in county income, and changes in county population, I identify conditionally-exogenous variation in the information environment that is (1) predictive of home internet use in a county, and (2) orthogonal to potentially endogenous variables like county education, racial composition, county average age, and previous voter turnout.

I use this conditionally exogenous variation to predict voting rates at both the aggregate and individual level, and I show that high-information environments have, on average, a demobilizing effect on voters. At the aggregate level, the demobilizing effect of an additional broadband provider on %VAP turnout ranges from $-.25\%$ (for presidential elections) to

−1.86% (for senate elections). At the individual level, the demobilizing effect of an additional broadband provider on the probability an individual votes ranges from −2.48% for presidential election years to −5.15% in midterm elections.

Moreover, the negative impact of high-information environment is significantly moderated by voter motivation. More motivated voters – those voting in presidential elections, those individuals in electoral swing-states, and those with an a priori higher likelihood of voting – receive a significantly more positive impact from access to high-information environments. The role of information appears to be blessing and a curse for increased participation, creating virtuous circles for those motivated to use it, and vicious circles for those who are not.

The Promise and Reality of a High-Information Environment

The reasons why individuals decide to vote vary, but being reasonably informed and knowledgeable about politics is often thought a necessary condition(Delli Carpini and Keeter 1997; Rosenstone and Hansen 1993; Verba et al. 1995; Wolfinger and Rosenstone 1980). Delli Carpini and Keeter (1997: 218) for example find that a one standard deviation change in political knowledge leads to a 9% increase in the probability of voting. Similarly, Verba et al. (1995: 352) find political information to be a significant predictor of political participation in their "Civic Voluntarism" model. Why does political information matter for turnout? For Delli Carpini and Keeter (1997), "Knowledge is an instrumental good that helps enlighten one's self interest and to translate it into effective political action." This perspective has its roots in Downs's (1957) *An Economic Theory of Democracy*, which discusses at length the necessary role of political information in reducing uncertainty about politics, where political action is more likely for individuals with lower levels of uncertainty. For both Downs (1957) and Delli Carpini and Keeter (1997), information is instrumental: it is a tool used by individuals to participate in politics. Delli Carpini and Keeter (1997: 224) discuss that political information influences participation directly by "promoting an understanding of why politics is relevant", as well as indirectly by promoting civic attitudes and behaviors.

This Downsian rational perspective leads to a clear prediction: that political participation is directly related to the cost and availability of political information (Bimber 2003). Individuals desire to reduce uncertainty when it comes to politics, but their ability to do so is constrained by the cost of collecting information. If the cost of that political information decreases (say, due to an increase in the availability of broadband internet), they can collect information at a lower cost, and thus be more likely to participate. There is a great deal of research associated with this perspective (Barber 2001; Delli Carpini 2000; Krueger 2002; Ward et al. 2003; Weber et al. 2003). This research posits that those individuals whose knowledge deficiencies generated disengagement can now be reached at a lower cost than before, helping to close the divide in participation between, for example, those with high and low socioeconomic status (Delli Carpini and Keeter 1997). The "cost" of political information can also be considered to be related to the degree to which it is presented in a style that meshes with the viewer. New, more exciting, presentation styles could entice those individuals who find traditional media boring (like young people). Additionally, the integration of social-networks and political news could allow for a more exciting group-based politics on-line, creating a news environment akin to the highly-motivating yellow journalism of the 19th century(Schudson 2000).

There are certainly historical instances where political information has become more readily available and participation has increased as a result. Both Schudson (2000) and Bimber (2003) assign a role to the increased availability of political information in the explosion of popular participation in politics in the early to mid 19th century, for example. That being said, the straightforward story does not contend as well with recent history. The previous 50 years has seen a dramatic decline in the cost of political information through new media like cable television and the internet, but according to Bimber (2001: 57): "Neither telephones, radio, nor television exerted a net positive effect on participation, despite the fact that they apparently reduced information costs and improved citizens access to information".

The reason may be due to the nature of information changes in this time period. The increase in the availability of political information has ridden a wave of technological change that has made *all* information cheaper. An increase in the availability of political information, in other words, comes with a poison pill of more entertainment options. The net effect may be that individuals consume less political information then they otherwise would have when political information was more costly but made up a higher proportion of information options. This seems to have been the case for the introduction of television. Exploiting exogenous variation in the roll-out of television in the 1940s and 50s, Gentzkow (2006) found a negative effect of TV on turnout. A similar pattern has been found in regards to the introduction cable television in the late 20th Century. Baum and Kernell (1999) demonstrate that the audience size for presidential addresses declined from the 1970s through the 1990s despite more and more houses having access to television. The decline, according to the authors, was caused not by voter disenchantment, but by the ability that cable news gives individuals to "tune-out".[1] Baum and Kernell point out that – while technological change has lowered the cost of collecting information – the increased choice environment means that choosing any particular television station or website imposes an increasingly high opportunity cost on the participant.

A more nuanced perspective moves away from average effects and takes into account heterogeneity in the way that individuals react to an increase in the information environment. An exogenous increase in the information environment presents an *opportunity* for individuals to become more politically informed. The uses and gratifications approach (see e.g. Chaffee and Kanihan 1997) takes the perspective that those individuals who already have a high level of knowledge and sophistication are more likely to find the collection of political-information more gratifying. Research on informational "schemas" likewise finds that knowledgeable people are more likely to learn from an increase in the availability of information (Bimber 2003). The clearest distillation of the idea that more choice makes the informationally-rich

[1] Others who are pessimistic about the impact of lower-cost information media focus on how the availability of entertainment will impact important *offline* determinants of political participation like civic and social activities. This line of thinking has its roots in the work of Putnam (2001), who found that much of the decline in civic life has its roots in the roll-out of television. That is, it became easier to use media-consumption time for entertainment rather than news gathering, and as such overall levels of political information dropped.

richer comes in Prior's (2007) *Post-Broadcast Democracy* which lays out a litany of evidence that significant heterogeneity exists in media-effects, whereby the most highly motivated individuals are the most likely to gain from a higher choice media environment. Similarly, Norris (2001) predicts that an increase in the choice and availability of information will create a "virtuous" circle of political information and knowledge for those motivated to take advantage of new technologies, and a "vicious" circle of political ignorance for those who are not (See also: Bimber 2001; Bonfadelli 2002; DiMaggio et al. 2004; Vettehen et al. 2004; Krueger 2002; Polat 2005; Weber et al. 2003).

The previous research suggests several hypotheses regarding the effects of an exogenous increase in the availability of information on turnout. There are reasons to believe the average effect may be either positive or negative. A positive effect would indicate that individuals are using less costly information to reduce their electoral uncertainty in a Downsian fashion, where a negative effect would indicate that the high-choice environment is causing people to tune-out from political information. Further, this effect is expected to be heterogeneous. Those individuals with a higher level of motivation are more likely to receive a positive impact from a high-information environment, while individuals with a lower level of motivation are more likely to receive a negative impact.

Data and Modeling Strategy

To relate changes in the media environment to electoral turnout what is needed is a measure of the availability of information that is exogenous to confounding factors while also being a valid predictor of access to information. Traditionally, individual level data have been used to test this relationship. These data usually rely on self-reports of media use which have been repeatedly found to be unreliable and systematically biased, often conflating media use with political interest (Bartels 1993; Dilliplane et al. 2013; Prior 2009a, b). In this paper I instead rely on FCC data that tracks the roll-out of broadband over time to counties. This allows me to relate within-county changes in broadband to within-county changes in turnout, effectively controlling for all between-county reasons that broadband adoption may be endogenous to turnout. Additionally, I find that controlling for within-county changes in income and population, as well as controlling for a flexible time trend, results in a model of broadband adoption that isolates exogenous variation while also strongly predicting access to the internet.

The Federal Communications Commission collects bi-annual data from all broadband providers operating in the United States on where they have deployed service. From 1999-2008 the FCC tabulated these results into data listing the number of broadband providers[2] in each zip-code[3] in the United States. This allowed the construction of a dataset which records the number of broadband providers in a county in each of the election years from 2000-2008.

The number of broadband providers is an indirect measure of the degree to which people in an area have access to low-cost information. These data do not measure the number of broadband providers individuals have access to in actuality. If a county has eight broadband providers operating within it, some individuals may have access to one provider, others may have access to 3 or 4, while some may have access to none. At the same time, for a given county an increase in the number of broadband providers can only mean one of two things. First, it may be that individuals who already have a broadband provider now have more choice. Previous research has found that a competitive market for broadband decreases the cost of broadband and increases the speed. Wallsten and Mallahan (2010) analyzed a similar dataset and found that those areas with multiple broadband providers had on-average lower costs and higher internet speeds. If instead the new broadband provider is deploying to a

[2] Broadband providers in this case are identified as the over-arching company which owns the service provider. That is, if two internet service providers have different names but are owned by the same corporation, they are treated as one "provider" in this dataset.

[3] Because I am interested in relating the number of broadband providers to election results, I transformed these zip-code tabulations to the county level. The Census provides Zip Code to County crossover tables that list the percentage of each county that lives within the zip codes that overlap with that county. Using this data, I was able to calculate the number of broadband providers in a given county for a given year by taking the average of the number of providers in each of the zip-codes within the county, weighted by the percent of the county population living in that zip code. All zip codes with between one and three broadband providers are coded the same in the FCC data, so these cases were coded as having 2 providers. Any zip-codes not included in the data were coded as having no broadband providers.

new area without broadband, than those individuals are getting access to faster information at a lower cost. Lelkes et al. (2015) show that the number of providers and number of subscribers correlate quite strongly at the cross-sectional county level.

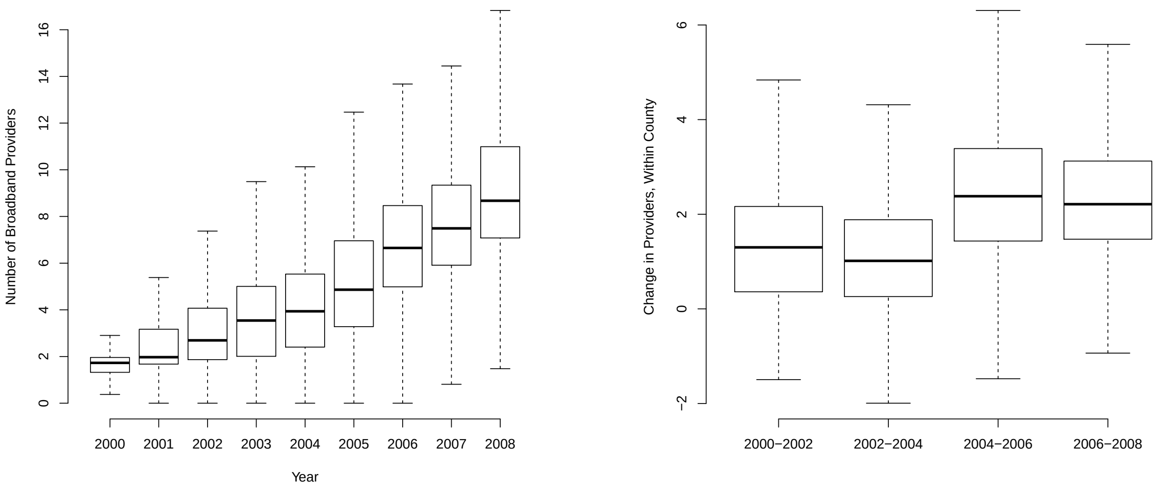
Another criticism of the use of this data is that I make no attempt to measure the actual content of the internet. However, this is not a problem in so much as the content of the internet does not change when an area receives more broadband providers. That is, content cannot bias the estimate of the effect of the internet here, because the content of the internet doesn't change for areas that suddenly gain additional providers.[4] There is interesting work to be done on the complex interaction between supply and demand of different forms of content, but it is not the focus of this work.

The goal of this study is to use these data to estimate the effect of an increase in the availability of information on electoral outcomes. In a perfect world, this would be estimated by randomly assigning one group of counties to have more access to information, with the other group acting as a control. As this cannot be done, the goal instead is to isolate in the FCC data exogenous variation. The vast majority of the variation in the broadband data is not random. Broadband providers are more likely to expand to areas that will be profitable and these areas may also vote at higher rates. This endogeneity means that a simple regression of turnout on number of providers would result in a positively biased estimate. Instead, this paper predicts turnout using a within-county fixed effects model with controls for changes in income and population at the county level, as well as controlling for a flexible time trend.

The panel structure of this dataset (repeated observation of the same county over time) provides a powerful solution to this problem of endogeneity. Including county fixed-effects (i.e. an intercept for each county) forces a model of turnout on broadband to look only at *within*-county variation. Using county fixed-effects makes it so that each county is it's

[4]This would be a problem in estimating the effect of newspapers on turnout. For example, a small newspaper who gained access to a new market would likely change their product to accommodate that market. The same cannot be said for the internet.

own "starting point". This estimation strategy effectively controls for any variable – whether observed or unobserved – that varies between counties and could cause a spurious correlation. Instead, counties serve as their own "control".



(a) Levels (b) Changes

Figure 1: Broadband Descriptives

Descriptive statistics on the levels and changes in broadband are shown in the two panels of Figure 1. The left panel uses a box-and-whisker plot to show the distribution of broadband providers across counties in each year. The substantial over-time growth in broadband providers is clear from the figure, as is the increasing between-county heterogeneity in access to the internet.The within-county variation in the number of broadband providers is shown in the right panel. The figure shows that, for example, the median county added 1.3 broadband providers in the interval of 2000-2002. In each of the 4 intervals the median change in broadband providers was positive, with the vast majority of counties seeing positive growth in each period.

The within-county design does not remove all sources of endogeneity. The first additional source is that within-county levels of broadband have variation because of secular growth in the availability of broadband. That is, to isolate random variation, the model must control for the fact that rates of broadband adoption were consistently rising in all counties across this time period. To do so, I include in the turnout models a 2nd order polynomial time trend which will control for this secular growth.

The second problem is endogeneity in changes. While county fixed effects remove all bias that results from between county confounding variables, there is a possibility that variation in broadband could be due to endogenous changes. A county may, for example, become richer over time, and therefore attract additional internet service providers. As such, the turnout model must control for variables at the county level that have variation over the time period which may cause endogenous changes to where broadband is provided.

To assess how these different controls partition the variance in the number of broadband providers, Table 1 presents a simple within-county regression predicting the number of logged broadband providers in a given county-year [5]. The purpose of this exercise is to examine how these particular variables partition the variance of broadband in order to show that when these variables are included in a model with the logged number of broadband providers the remaining variation is random.

The first model has only county fixed-effects. The R^2 of this model signifies that a full 79% of the variation is explained by simply giving an intercept to each county – in other words the vast majority of variation in the level of broadband is *between* counties. Model 2 adds to this a 2nd order polynomial time trend. The coefficients on $Year$ and $Year^2$ indicate that within-county the number of broadband providers increased at a decreasing rate over time. Adding this secular time trend explains a further 16% of the variation in broadband.

The remaining 5% of variation is a mixture of endogenous changes in broadband due to changes in county characteristics, as well as the exogenous variation that this paper looks

[5] Throughout, I use logged broadband providers for both statistical and theoretical reasons. Statistically, the number of broadband providers is positively skewed, with many counties having a small number of broadband providers, and few counties have a large number of providers. Theoretically, it makes good sense in terms of internet cost and speed for the end user to treat each successive broadband provider as having less impact. Going from having a monopoly of one broadband provider to two is a large change that will seriously affect the probability individuals in a county have accessible high-speed internet. We would not expect the same impact going from seven to eight broadband providers.

Table 1: Determinants of Broadband Adoption

	$ln(Providers_{jt})$		
	(1)	(2)	(3)
$Year_t$		0.28***	0.29***
		(0.004)	(0.004)
$Year_t^2$		−0.01***	−0.01***
		(0.0004)	(0.0004)
$ln(Med.Income_{jt})$			0.63***
			(0.08)
$ln(Pop_{jt})$			−0.69***
			(0.07)
County F.E.	Yes	Yes	Yes
N	28,293	28,293	28,293
R^2	0.82	0.95	0.95
F Statistic	37.73*** (df = 3137; 25156)	153.84*** (df = 3139; 25154)	156.08*** (df = 3141; 25152)

*p < .1; **p < .05; ***p < .01
OLS, County-clustered standard errors in parentheses

to exploit (and, of course, some degree of variation is due to measurement error). The third model takes into account features of the county that may cause endogenous changes. (Recall that, because of the county fixed effects, any feature of the county that is stable cannot cause endogenous changes). The variables included are (logged) median income, and (logged) population.[6] On their book on the economics of broadband expansion, Nuechterlein and Weiser (2005) identify the major drivers of new communications infrastructure to be these two features. Income, in particular, should be an important driver of demand for internet services. While other features of counties might be *correlated* with low levels of broadband (i.e. race, education, unemployment), they are ultimately caused or reflected by the median income of the county. Including these other features would serve only to decrease the precision of the estimates due to the collinearity between themselves and county median income. The coefficients in the model indicate that within county changes in median income are associated with a higher number of broadband providers, as expected. However, within-county changes in population are associated with *lower* levels of broadband, which is counter to the theory posited by Nuechterlein and Weiser (2005). While this variable does not run in the direction anticipated, I nevertheless continue to include it as a summary measure of

[6]Data from the 2000 US Census and the 2005-2008 ACS Surveys. Data for 2001-2004 linearly interpolated.

important over-time demographic changes. The population flows in and out of a county during the time period studied could be indicative of other important changes (for example, housing prices) that are correlated with the turnout of a county.

The residuals from this fully specified model represent the remaining variation that will be exploited as an exogenous predictor of an increase in the availability of information. The first important question is whether this remaining variation is truly random. To test this, I run a simple regression of the residuals from the fully specified model on education levels, race, the median age, and 1996 presidential election turnout. None of these other county features are significantly related to the residuals from the fully specified model (%Bachelor $p = .75$; %White $p = .30$; %Black $p = .48$; Median Age $p = .87$; Turnout 1996 $p = .75$), nor are they jointly significant ($p = .89$).[7]

While this identification strategy isolates plausibly exogenous variation in broadband, it is useless if that exogenous variation is too weak to generate actual differences in the availability of internet for individuals living in these counties. In other words, it's not clear whether within-county variations have validity as an inducement for individuals to select into the "treatment" of signing up for internet at home. To test this I use data from the Current Population Survey Internet Supplement[8] for the years 2001,2003, and 2007. This survey simply asked individuals whether or not they had access to the internet at home. I run the following linear probability model with standard errors clustered at the county level, where individuals are indexed by i and counties by j.

$$\begin{aligned} p(InternetHome_i) &= \alpha_j + \beta ln(providers_{jt}) + \beta ln(Med.Income_{jt}) + \\ &\quad \beta ln(Population_{jt}) + \beta Time_t + \beta Time_t^2 + \epsilon_j \end{aligned}$$

The fixed-effects for county makes it so individuals in a county-year serve as a control for

[7] The full results of this regression can be found in the appendix.

[8] Data collected from the Integrated Public Use Microdata Series: Version 6.0

other individuals in that same county in different years. Adding controls for changes in the counties median income and population, as well as a flexible time trend, isolates exogenous variation in the (logged) number of broadband providers in the county. Estimating this with a linear probability model indicates that a one unit increase in $ln(providers)$ within-county is associated with a 5.2% ($p < .01$) increase in the probability an individual has the internet at home.[9] As such, I conclude that this plausibly exogenous variation in broadband is a reasonably strong predictor of whether individuals have internet at home.

Aggregate Level Methodology

At the aggregate level, within-county changes in the number of broadband providers will be related to within-county changes in turnout in both presidential and senate elections.[10]. The purpose of testing the relationship on both presidential and senate elections is to naturally vary the motivation of the voters in question. Senate elections – even in mid-term years – generate a far lower degree of excitement compared to presidential elections, and as such, have lower turnout. The voting data was collected for election years between 2000-2008 from the CQ election database. The CQ elections database lists the raw number of votes cast in each county in each election. I use the Voting Age Population (VAP) from the US Census and Americas Community Survey as a denominator to determine the percentage of the voting age population that voted in each county. For presidential election turnout each county has an observation in each of the three election years (2000,2004,2008). For senate elections, counties only have an observation for a given year if a Senate seat was voted on. So for example, a county in California would have observations in 2000,2004 and 2006; while a county in Tennessee would have observations for senate voting in 2000, 2002, 2006 and 2008.

The fully specified model is as follows:

[9] The full results of this regression can be found in the appendix.

[10] Unfortunately, results at the county level are not available for House races

$$\begin{aligned} Vote\%_{jt} &= \alpha_j + \beta ln(providers_{jt}) + \beta ln(Med.Income_{jt}) + \\ &\quad \beta ln(Population_{jt}) + \beta Time_t + \beta Time_t^2 + \epsilon_j \end{aligned}$$

$Vote\%_{jt}$ – the %VAP turnout in county j in time period t – is modeled as a function of a county fixed effect, a 2nd order polynomial time trend, (logged) number of broadband providers, (logged)median income and (logged) population. As the specification controls for endogenous variation in the number of broadband providers in a county, the value of β_3 provides an estimate of the impact of exogenous variation in the availability of information on the percent of the VAP turning out in an election. As is indicated by ϵ_{county}, the standard errors of each model are adjusted for clustering at the county level. This model is run separately for turnout in presidential and senate elections.

There is also an expectation of heterogeneity based on motivation. I operationalize motivation in two ways at the aggregate level. The first way is simply by examining the effect for both presidential and senate elections. Within a given year motivation is higher for the presidential elections, which means that individuals are on average more likely to gain an informational advantage from a high-information source. Second, I test for whether the impact of broadband changes in closer elections. I calculate for each state-year the closeness of the senatorial and presidential elections by taking the absolute value of the Democratic percent of the vote minus the Republican percent of the vote. Lower numbers on this scale indicate that voting in that state took place in a more competitive environment. Voters in closer elections tend to be more motivated and turn out to vote more often as they perceive their vote to "matter" more when the margin of victory is thin. Cox and Munger (1989) find that closeness has an effect on turnout even after considering higher campaign spending in states with close margins. I expect that counties in close electoral states will have a significantly more positive impact from broadband than counties in states with lower levels electoral competition.

Aggregate Level Results

Table 2 presents the results main aggregate level results. The goal here is to test whether exogenous increases in availability of information have a positive or negative effect on turnout. A positive impact would suggest that more information allows individuals to reduce political uncertainty and thus be more likely to turn out. A negative impact would indicate that an expansion of the information environment serves more as a distraction, thus lowering turnout. Models 1 and 3 show the average treatment effect of an additional logged broadband provider on the percentage of individuals in a county that turn out to vote. An additional logged broadband provider depresses turnout by -.17% in presidential elections and -1.24% in senate elections. The average effect of an exogenous increase in the availability of information on electoral turnout is negative and statistically significant.

Table 2

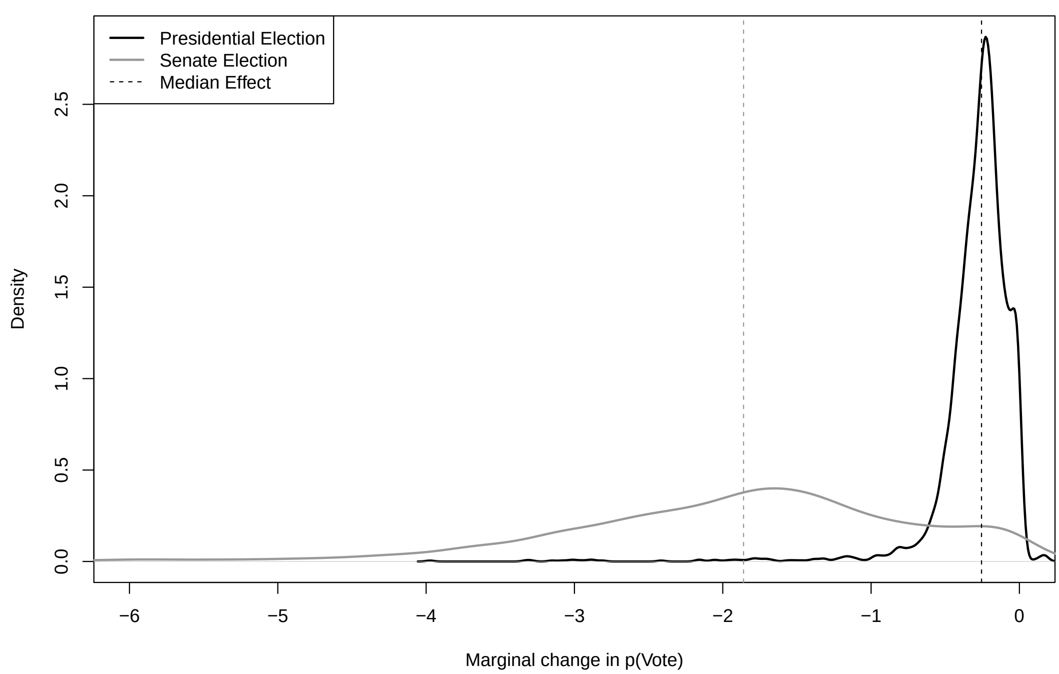
	Presidential Elections		Senate Elections	
	(1)	(2)	(3)	(4)
$ln(Providers_{jt})$	-0.17^{*}	0.88^{***}	-1.24^{***}	-1.05^{***}
	(0.09)	(0.12)	(0.20)	(0.22)
$ln(Med.Income_{jt})$	-12.33^{***}	-10.55^{***}	-8.03^{***}	-10.67^{***}
	(0.79)	(0.80)	(1.59)	(1.54)
$ln(Population_{jt})$	-8.13^{***}	-9.10^{***}	-21.52^{***}	-22.87^{***}
	(0.79)	(0.77)	(1.69)	(1.62)
$Year$	1.76^{***}	1.76^{***}	-2.59^{***}	-2.20^{***}
	(0.04)	(0.04)	(0.10)	(0.10)
$Year^2$	-0.15^{***}	-0.15^{***}	0.40^{***}	0.36^{***}
	(0.004)	(0.004)	(0.01)	(0.01)
$Elect.Closeness_{jt}$		0.05^{***}		-0.12^{***}
		(0.01)		(0.01)
$ln(Providers_{jt})$:$Elect.Closeness_{jt}$		-0.06^{***}		-0.02^{***}
		(0.005)		(0.004)
$County.FE$	Yes	Yes	Yes	Yes
N	9,341	9,341	10,403	10,403
R^2	0.997	0.997	0.978	0.98
F Statistic	756.37^{***} (df = 3116; 6225)	775.63^{***} (df = 3118; 6223)	102.76^{***} (df = 3116; 7287)	116.08^{***} (df = 3118; 7285)

*p < .1; **p < .05; ***p < .01

To make sense of these coefficients, it is in instructive to take the derivative of the regression with respect to the (non-logged) number of providers which yields: $\frac{\partial \%Vote}{\partial providers} = \beta_1 * \frac{1}{providers}$. That is, the marginal effect of a one unit change in providers depends on the "starting point" number of providers. The cleanest interpretation is to think about an impact starting from a county only having 1 provider, as then $\frac{\partial \%Vote}{\partial providers} = -.17\%$ for a one-unit increase in providers for presidential elections, and -1.25% for a one unit increase for senate elections.

To better understand the magnitude of this effect, I calculate the predicted effect of broadband on turnout between the 2000 and 2004 elections for each of the counties in the dataset. To do this, I extract the number of broadband providers in each county in the year 2000 (mean=1.7, sd=1.03) and the change in broadband providers between 2000 and 2004 for each of the counties (mean=2.64, sd=1.78). With these values I am able to calculate the predicted impact of broadband between these two elections for both presidential and senate voting. For example, if a county had one provider in the year 2000 and three providers in the year 2008, the predicted impact of broadband would be $2 * \frac{-.17}{1} = -0.34\%$. The distribution of effects are presented in Figure 2. The median effect for presidential elections is -.25%, and for senate elections -1.86%.

Figure 2: Predicted Impact of Broadband: 2000-2004



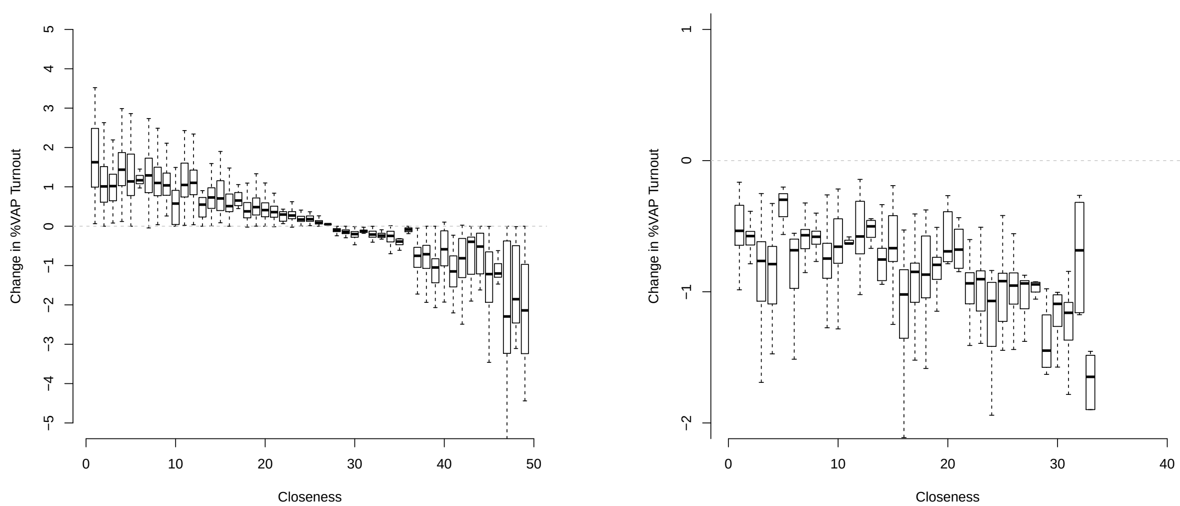
To put these effect sizes in context Gentzkow (2006) used a similar design to assess the impact of television on Congressional voting rates, finding an effect size of -.8% over a four year period. The effects found here indicate that the Internet had a stronger effect on Congressional elections compared to the effect of television. The effect on presidential election turnout is much smaller, though this is in line with the theory that more motivated individuals will be less negatively affected by an increase in the availability of information.

Models 2 and 4 in Table 1 further investigate whether the main effect is moderated by the closeness of the electoral margin in the state. Individuals in states with a smaller electoral margin should be more electorally motivated, and therefore receive a less negative effect from an increase in information. The results for presidential elections show the expected pattern. The coefficient on $ln(Providers)$ represents the effect of an additional logged broadband provider in counties where the Democratic and Republican candidates received the same share of the vote, and we see that this effect is *positive* and significant. However, the coefficient on the interaction term indicates that as the margin between the two candidates increases in the county, the effect of logged broadband providers becomes more negative, which is evidenced by the negative interaction term. The relationship is the same in Model 4 for the Senate, though in this case, the effect of broadband is *always* negative, and becomes more so as the election becomes less close.

As with the main effect, presenting the marginal effects of the number of logged broadband providers is somewhat misleading as all counties have a different "starting point" for number of providers. To better assess the marginal effect across levels of electoral closeness, we can take the derivative of each moderation equation with respect to the (non-logged) number of broadband providers. If we consider the number of broadband providers as variable x and the electoral closeness as variable z:

$$\frac{\partial \% Vote}{\partial x} = \frac{\beta_{ln(x)}}{x} + \frac{\beta_{ln(x):z} \cdot z}{x}$$

The marginal effect of the number of broadband providers is dependent on both the starting level of providers as well as the level of the moderating variable. Instead of using arbitrary values to evaluate these, I examine the predicted marginal effect of a one unit increase on the real year 2000 values in the data. This step is important as the marginal effect will be smaller for counties that start with a larger number of providers. If the number of providers is not randomly distributed across levels of electoral closeness then this could affect the interpretation of the results.



(a) Presidential Elections (b) Senate Elections

Figure 3: Marginal Effect of Additional Broadband Provider on Year 2000 Counties

Figure 5 displays the marginal effect of one additional broadband providers on year 2000 counties across levels of electoral closeness. Each state is represented by a box and whisker plot. The within-state variance in year 2000 broadband rates is what generates dispersion in each predicted marginal effect, while between state variance in the number of providers makes it so the slope of the marginal effects is not strictly linear. That is, states with more broadband providers in the year 2000 will have an attenuated marginal effect. Despite this variance, A clear negative pattern is evident in the figure for both presidential and senate elections. In presidential swing states an additional broadband provider produces a positive effect on voter turnout. That is, when voters have more reason to be motivated, an exogenous increase in the availability of information helps the voters and generates turnout. In states where the election is not as close, and therefore voters are less motivated, an exogenous increase in the availability of information serves to depress turnout.

Individual Level Methodology

Research in social science that use aggregate data will always be plagued with questions of the ecological fallacy – the phenomenon whereby a relationship measured at the aggregate level is found to be different then the same relationship when measured at the individual level. Considering that the meta-analysis cited above (Boulianne 2009) found that the average effect found for the usage of the internet on participation to be *positive*, it is prudent to be skeptical of the aggregate results.

I investigate whether the negative relationship between increased access to information and voting rates is stable at the individual level by pairing the exogenous variation in broadband providers exploited above to samples of voters for each election year, 2000-2008, from the Current Population Survey Voter Supplement[11]. While the CPS does not ask voters for whom they voted, its substantial benefit over an academic study like the *National Election Study* is the size of the sample. The 2000-2008 cross section includes over 150,000 respondents. This generates sufficient statistical power for what is essentially a natural field experiment. The outcome variable of interest in the dataset is whether an individual voted in the election or not. The CPS does not ask separately whether an individual voted for each race in a given year, so at the individual level I cannot distinguish between voting for President and lower level candidates, other than to look at the relationship separately for

[11]Data collected from the Integrated Public Use Microdata Series: Version 6.0

president and midterm elections.[12]

The estimation strategy is to run a within-county fixed effect linear probability model predicting whether an individual i situated in county j in time t reported voting (1) or not (0) using the same specification above that isolated the exogenous variation in $ln(Providers)$.

$$\begin{aligned} P(Vote_i) &= \alpha_j + \beta ln(Providers_{jt}) + \beta ln(Med.Income_{jt}) \\ &\quad + \beta ln(Pop_{jt}) + \beta Time_t + \beta Time_t^2 + \epsilon_j \end{aligned}$$

In the aggregate analysis the observations were counties and the fixed effects made it so that counties served as their own control. The logic is the same here, but instead of counties taken in aggregate, groups of respondents in a county act as the control group for other groups of respondents in that same county in different years. As above, the controls for county income, population, and a flexible time trend make it so that the number of broadband providers varies exogenously. In effect, this design approximates a field experiment where broadband is being randomly assigned to groups of respondents.

It should be noted, however, that a county being assigned an exogenously higher number of broadband providers is not the "treatment" in this case. Rather, an individual taking advantage of the increased ease of access to information is the treatment, which I cannot directly measure. Some individuals, who Gerber and Green (2012) call "compliers" will respond to the treatment by increasing the amount of information accessed. Others, the "never-takers" will not respond to the treatment. Because the proportion of "compliers" is unknown, I cannot estimate the Complier Average Causal Effect (CACE), which would be a better estimate of the effect of the treatment as it only focuses on those who received it. Instead, I rely here on the Intent to Treat Effect (ITT), which averages across those who

[12]The question text reads: "In any election, some people are not able to vote because they are sick or busy or have some other reason, and others do not want to vote. Did (you/name) vote in the election held on Tuesday, November X, XXXX?"

respond to treatment and those who don't. As such, the estimates below are likely smaller than what would be found if I was able to calculate the CACE (Gerber and Green 2012).

As with the aggregate effect above, I examine how this effect is heterogeneous in different election contests by examining the effect across different levels of electoral competitiveness, as well as in both midterm and presidential election years.

Additionally, I examine whether an individual's propensity to vote affects the magnitude of the treatment effect. I determine each individual's a priori propensity to vote by running a demographic model on all individuals in the CPS voter supplement from 1996 to 2014. I estimate the following model using a probit model, where Φ is the CDF of the standard normal distribution and $\mathbf{K}$ is a vector of demographic variables (age, gender, education, race, state) factored out into dummy variables:

$$\begin{aligned} P(Vote_i) &= \Phi(\alpha + \beta\mathbf{K}_{it} + \beta Year_t + \beta Year_t^2 \\ &\quad + \beta Year_t * \mathbf{K}_{it} + \beta Year_t^2 * \mathbf{K}_{it}) \end{aligned}$$

Each individuals probability of voting is modeled as a function of the year they are in and their demographic profile, where the coefficients for each demographic category can vary with time. I use the estimates from this model to predict, for each individual in the dataset matched to the broadband provider data, the probability that they will vote based on their demographic characteristics. I then separate respondents into three distinct datasets based on terciles of the probability of voting. The model is then run separately on those three datasets and the coefficient on $ln(Providers_{jt})$ is compared. One potential issue with running such an analysis is related to the discussion of compliance above. If the rate of compliance is different among the three groups, then differences in the estimates could result from either different response to treatment, or different probabilities of compliance. However, I show in the appendix using the CPS internet supplement that a priori probability of voting based on demographics does not increase or decrease the strength of the encouragement (increases in $ln(Providers_{jt})$) on the probability individuals have internet at home. In other words, there is no evidence that compliance varies across levels of the probability of voting.

Individual Level Results

Models 1 and 3 of Table 4 present the individual level results separated by presidential and midterm election years. The effect of an exogenous increase in the logged number of broadband providers in both types of elections has a negative impact on the probability that individuals vote. As would be expected by the motivation hypothesis, the effect of broadband is about twice as strong for midterm election years where voters are less motivated to use a high information source to seek out political information.

Table 3

	$P(Vote_i)$			
	Presidential Years		Midterm Years	
	(1)	(2)	(3)	(4)
$ln(Providers_{jt})$	−0.02**	−0.01	−0.04**	−0.01
	(0.01)	(0.01)	(0.02)	(0.05)
$ln(Med.Income_{jt})$	−0.18***	−0.12**	−0.12	0.36
	(0.05)	(0.06)	(0.10)	(0.24)
$ln(Population_{jt})$	0.16***	0.16***	−0.15*	−0.19
	(0.04)	(0.04)	(0.08)	(0.22)
$Year_t$	0.02***	0.02***	0.01***	0.02**
	(0.003)	(0.003)	(0.002)	(0.01)
$Year_t^2$	−0.001***	−0.002***		
	(0.0002)	(0.0002)		
$Elect.Closeness_{jt}$		0.003***		−0.003
		(0.001)		(0.003)
$ln(Providers_{jt}):Elect.Closeness_{jt}$		−0.001***		0.001
		(0.0004)		(0.001)
$County.FE$	Yes	Yes	Yes	Yes
N	87,719	87,719	64,088	37,159
R^2	0.74	0.74	0.56	0.57
F Statistic	730.50*** (df = 335; 87384)	726.31*** (df = 337; 87382)	242.37*** (df = 333; 63755)	178.57*** (df = 277; 36882)

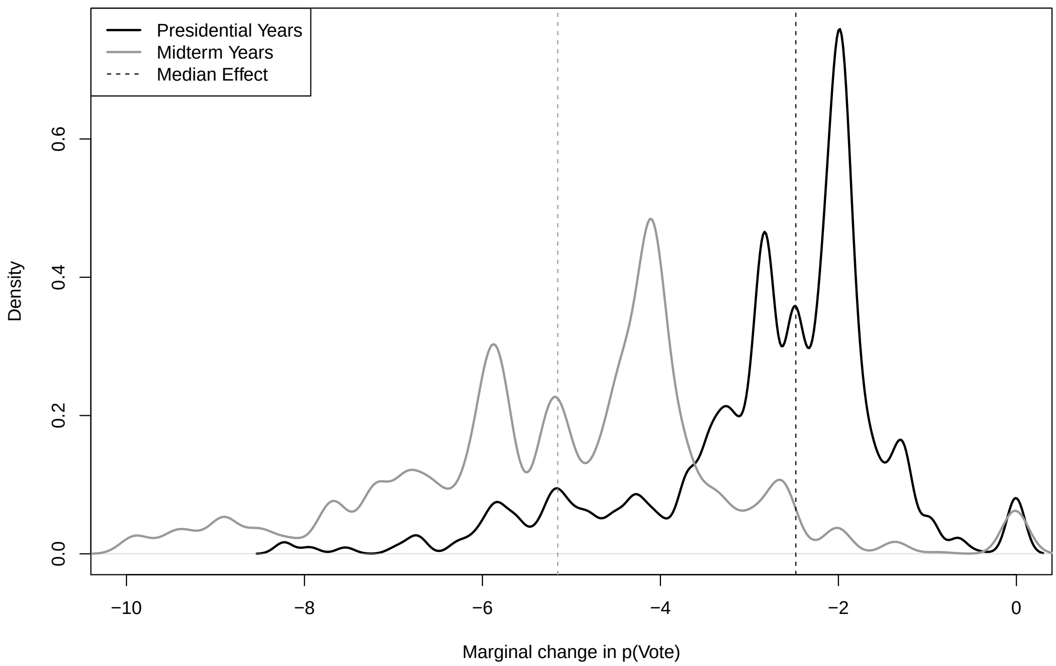
*p < .1; **p < .05; ***p < .01

To determine the magnitude of these effects, I calculate the estimated between-election impact of changes in broadband for the sample of individuals from the year 2000.[13] I extract for each individual the number of providers in their county in the year they were interviewed, and the number of providers in their county 4 years later. From these number I calculate the

[13] I evaluate the effect size for a common sample despite the fact that year 2000 voters are not those voting in midterms. Doing so gives a clearer picture of the relative size of effects.

expected impact of increases in broadband for that individual. The distribution of predicted effects is presented in figure 5.

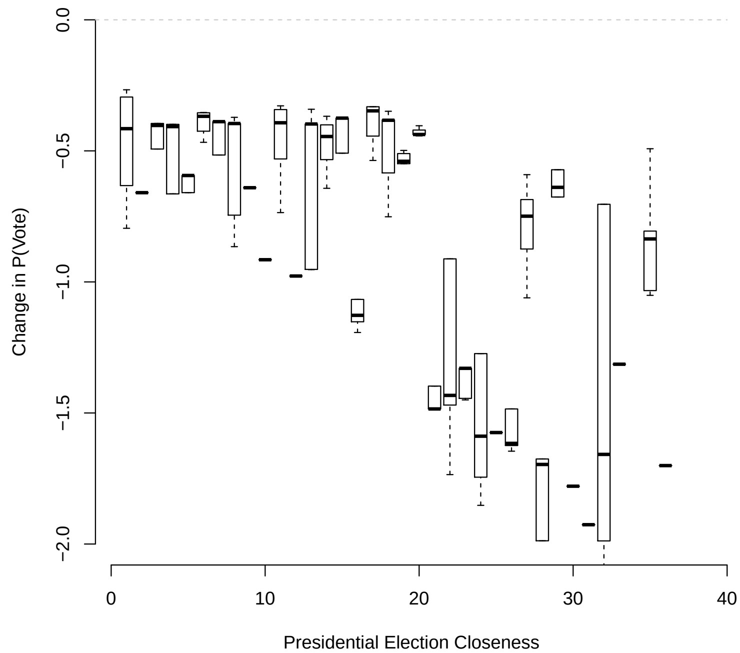
Figure 4: Predicted Impact of Broadband on Individuals, 2000-2004



For presidential election years, the median predicted effect of changing broadband between for a 4 year period is a -2.48% reduction in the probability an individual will turn out to vote. For midterm election years, the median predicted effect of changing broadband for a 4 year period is a 5.15% reduction in the probability an individual will turn out to vote.

To put these individual level results in context, we can look to other Intent-to-Treat effect field experiments on turnout. Arceneaux and Nickerson (2009) present a meta-analysis of 11 Get-out-the-Vote field experiments that took place in different contexts and for different elections. The ITT effects on the probability of turnout for the 11 studies in that analysis ranged from a high of 10.9% to a low of 1.4%. This puts the effects found here squarely in the range of those found in GOTV field experiments, though obviously running in the opposite direction.

Figure 5: Impact of One Additional Broadband Provider on Year 2000 CPS Respondents



Models 2 and 4 investigate whether the effect of high information environments are moderated by electoral closeness. The coefficients on the interaction terms indicate that the closeness of the presidential election does impact the effect of high information environments in presidential election years, though closeness of senate elections does not have an impact in midterm years. Figure 6 visually displays this interaction effect. The figure shows, for year 2000 respondents, the marginal impact of an additional broadband provider given both the number of providers each individual has, as well as the closeness of the election in that individual's state. A clear negative pattern emerges, though this effect is attenuated at the low end of the electoral closeness scale, as these individuals happen to live in counties with a relatively higher base number of broadband providers. Those individuals who are voting in closer elections have a substantially less negative impact from decreases in the cost of information than those individuals living in less electorally competitive states, which further suggests that the impact of information on voting rates is contingent on the electoral environment.

Table 4: Individual Results by a Priori Propensity of Voting

	$P(Vote_i)$		
	Low	Med	High
$ln(Providers_{jt})$	-0.04^{***}	-0.04^{***}	-0.003
	(0.01)	(0.01)	(0.01)
$ln(Med.Income_{jt})$	-0.23^{**}	-0.06	-0.05
	(0.10)	(0.08)	(0.06)
$ln(Pop_{jt})$	0.01	0.05	0.05
	(0.06)	(0.06)	(0.05)
$Year_t$	-0.01^{**}	0.02^{***}	0.002
	(0.004)	(0.004)	(0.003)
$Year_t^2$	0.003^{***}	-0.001^{***}	0.0002
	(0.0004)	(0.0003)	(0.0002)
County F.E.	Yes	Yes	Yes
N	50,350	49,833	49,199
R^2	0.41	0.69	0.87
F Statistic	105.42^{***} (df = 335; 50015)	332.33^{***} (df = 335; 49498)	970.93^{***} (df = 335; 48864)

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

The final test is to see whether individual differences in a priori propensity to vote affect the impact of high-information environments on whether an individual turns out or not. Table 4 displays three separate regressions for low ($p(Vote) < 58\%$), medium ($58\% < p(Vote) < 78\%$), and high propensity voters($p(Vote) > 78\%$). For simplicity I've pooled together presidential and mid-term election years for this analysis. The results indicate that the negative effect of high-information environments are only found in individuals with low and medium propensities to vote. Individuals who, based on their demographics, are likely to turn out to vote are not affected by variation in the number of broadband providers in their counties, which suggests that the negative impact of broadband is mitigated for individuals who are likely to gather political information in similar rates regardless of the information context. In other words, the distracting nature of the internet only works on those who can be distracted from politics.

Conclusion

This paper set out to determine the effects of increases in information on voter turnout. To do so, I used a unique within-county dataset that isolated conditionally-exogenous variation in the roll-out of broadband internet. There were good reasons to believe that this effect could be either positive or negative, and that it would be significantly moderated by electoral motivation. The results indicate that the average effect of turnout is negative, but as electoral motivation increases, so does the positive impact of information. This relationship holds at both the aggregate and individual levels, and is of a magnitude roughly equivalent to those found in other studies.

The goal of this study was to study the first generation question of the direction and magnitude of the effect. Determining the causal pathways of that effect is something that I do not take on here. At the individual level determining how different information environments affect political knowledge is an important step. Delli Carpini and Keeter (1997) split their scale of political knowledge into three components: rules of the game, people and parties, and the substance of politics. Research can ask: which of these banks of knowledge are most affected by changes in the information environment, and how do deficiencies in these categories affect turnout?

Another pathway left unexplored here has to do with motivation. I've discussed motivation as an individual phenomenon, whereby those individuals in more competitive electoral environment and those with higher a priori likelihoods of voting get a substantially more positive impact from high-information environments. In addition to this voter-behavior story, there is likely a parallel story from the campaign side. More competitive elections would spur campaigns to put more resources into an area: resources that may more easily reach citizens in a high-information environment. Research by Cox and Munger (1989) on the motivating nature of electoral closeness suggest both are likely to contribute to the effects found here, in that electoral closeness does spur additional campaign spending, but voters seem to have motivation to turnout above and beyond what would be expected based solely on these investments by campaigns.

Progressive-era thinkers like John Dewey saw information as the key to building a more representative and equitable society. There has been an assumption – for example in work examining the so-called digital divide – that if there is a problem with information, then it is that there is too little of it. I hope that this work here helps us to examine that assumption. New technologies like the internet are nothing more than what we make of them. On average, citizens in the United States have not used this new opportunity in a virtuous manner. However, when citizens are motivated to participate in politics these new technologies are there to support them. As such, the solution to this problem is not creating more or better political content on-line, but rather finding ways to boost motivation to participate in politics both on-line and off.

References

Arceneaux, K. and Nickerson, D. W. (2009). Who is mobilized to vote? a re-analysis of 11 field experiments. *American Journal of Political Science*, 53(1):1–16.

Barber, B. R. (2001). The uncertainty of digital politics. *Harvard International Review*, 23(1):42.

Bartels, L. M. (1993). Messages received: the political impact of media exposure. *American Political Science Review*, 87(02):267–285.

Baum, M. A. and Kernell, S. (1999). Has cable ended the golden age of presidential television? *American Political Science Review*, 93(01):99–114.

Bimber, B. (2001). Information and political engagement in america: The search for effects of information technology at the individual level. *Political Research Quarterly*, 54(1):53–67.

Bimber, B. (2003). *Information and American democracy: Technology in the evolution of political power*. Cambridge University Press.

Bonfadelli, H. (2002). The internet and knowledge gaps a theoretical and empirical investigation. *European Journal of communication*, 17(1):65–84.

Boulianne, S. (2009). Does internet use affect engagement? a meta-analysis of research. *Political communication*, 26(2):193–211.

Chaffee, S. H. and Kanihan, S. F. (1997). Learning about politics from the mass media. *Political communication*, 14(4):421–430.

Cox, G. W. and Munger, M. C. (1989). Closeness, expenditures, and turnout in the 1982 us house elections. *American Political Science Review*, 83(01):217–231.

Delli Carpini, M. X. (2000). Gen. com: Youth, civic engagement, and the new information environment. *Political communication*, 17(4):341–349.

Delli Carpini, M. X. and Keeter, S. (1997). *What Americans know about politics and why it matters*. Yale University Press.

Dilliplane, S., Goldman, S. K., and Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, 57(1):236–248.

DiMaggio, P., Hargittai, E., Celeste, C., and Shafer, S. (2004). From unequal access to differentiated use: A literature review and agenda for research on digital inequality. *Social inequality*, pages 355–400.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper and Brothers, New York.

Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, pages 931–972.

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Krueger, B. S. (2002). Assessing the potential of internet political participation in the united states a resource approach. *American Politics Research*, 30(5):476–498.

Lelkes, Y., Sood, G., and Iyengar, S. (2015). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.

Nuechterlein, J. E. and Weiser, P. J. (2005). *Digital crossroads: American telecommunications policy in the Internet age*. MIT press Cambridge, MA.

Polat, R. K. (2005). The internet and political participation exploring the explanatory links. *European journal of communication*, 20(4):435–459.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Prior, M. (2009a). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, page nfp002.

Prior, M. (2009b). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, 71(03):893–908.

Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Rosenstone, S. and Hansen, J. (1993). *Mobilization, participation, and American democracy*. McMillan, New York.

Schudson, M. (2000). The good citizen: A history of american civic life.

Verba, S., Schlozman, K. L., and Brady, H. E. (1995). *Voice and equality*. Cambridge, MA: Harvard University Press.

Vettehen, P. H., Hagemann, C. P., and Van Snippenburg, L. B. (2004). Political knowledge and media use in the netherlands. *European Sociological Review*, 20(5):415–424.

Wallsten, S. and Mallahan, C. (2010). Residential broadband competition in the united states. *FCC Omnibus Broadband Initiative*.

Ward, S., Gibson, R., and Lusoli, W. (2003). Online participation and mobilisation in britain: Hype, hope and reality. *Parliamentary affairs*, 56(4):652–668.

Weber, L. M., Loumakis, A., and Bergman, J. (2003). Who participates and why? an analysis of citizens on the internet and the mass public. *Social Science Computer Review*, 21(1):26–42.

Wolfinger, R. E. and Rosenstone, S. J. (1980). *Who votes?*, volume 22. Yale University Press.

Appendix

Regression Results Predicting Residuals From Broadband Roll-Out Model

Table 5

	Residuals Table 2 Model 3
%Bachelor.Degrees	−0.0001
	(0.0004)
%Black	0.0002
	(0.0003)
%White	0.0003
	(0.0003)
Median.Age	0.0001
	(0.001)
*Turnout*96	−0.0001
	(0.0003)
Constant	−0.02
	(0.03)
N	28,012
R^2	0.0001
F Statistic	0.34 (df = 5; 28006)

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Regression Results Predicting Home Broadband Use

Table 6

	$P(Internet.Home_i)$	
	(1)	(2)
$ln(providers_{jt})$	0.05***	0.07***
	(0.01)	(0.02)
$ln(Med.Income_{jt})$	0.19***	0.02
	(0.05)	(0.08)
$ln(Pop_{jt})$	−0.08*	−0.17***
	(0.04)	(0.06)
$Year$	−0.02***	−0.02***
	(0.003)	(0.005)
$Year^2$	0.01***	0.01***
	(0.0004)	(0.001)
$p(Vote_i)$		0.05
		(0.04)
$ln(providers_{jt}) : P(Vote_i)$		−0.02
		(0.02)
County F.E.	Yes	Yes
N	110,995	59,024
R^2	0.73	0.72
F Statistic	1,074.65*** (df = 277; 110718)	554.98*** (df = 273; 58751)

*p < .1; **p < .05; ***p < .01

Model 1 tests to see whether the within-county specification used in the turnout models is predictive of individuals having internet at home. In field-experiment terms, this is determining whether the inducement (cheaper access to broadband) results in individuals complying with treatment (getting internet at home). The results show that every additional logged broadband provider increases the probability an individual has the internet at home by 5%. In other words, small, conditionally exogenous, variation in Broadband does have significant effects on individuals actually having access to less costly information.

Mode 2 relates individual level heterogeneous effects. Because this study requires compliance, there is a potential issue in measuring heterogeneous effects where differences across individuals may be due to the treatment having a different effect or individuals having different rates of compliance. To test this, I use the demographic turnout model discussed above to assign each individual in the CPS internet supplement an a priori probability of voting (which is the variable for which I am interested in heterogeneous effects). I test to see whether levels of this variable influence the degree that individuals comply with treatment. A positive coefficient on the interaction term would indicate that as an individuals probability of voting increases, so does their propensity to get internet at home when the number of broadband providers increases in their area. This is not the case. The coefficient on the interaction term is negative, with a standard error as large as the effect size. This result suggests that the heterogeneous results based on a priori voting probability are picking up actual differences in effects, not differences in rates of compliance.