Car Accidents and 3G Coverage: New Evidence Using Cell Phone Tower Construction

Bree Lang*

Matthew Lang†

Jonathan Hersh‡

ABSTRACT

We examine the relationship between the growth of 3G cellular phone coverage and traffic accidents in California between 2001 and 2013. Cellular coverage maps are only available from the FCC in 2015 and 2016, but not in the mid-2000s when 3G coverage was introduced to the public. We link cellular coverage along a highway in 2016 to the location of antenna towers and then apply machine learning techniques to predict coverage between 2001 and 2013. Fixed-effect Poisson regressions find that car accident rates increase 1.1 percent when 3G cell phone coverage is introduced to an area, controlling for traffic volume. The types of accidents most responsive to 3G coverage are non-severe crashes that take place in highly trafficked areas. Accidents caused by drivers over 65 do not change in response to 3G coverage. In contrast to much of the previous research, we find a persistent increase in traffic accidents when access to cellular coverage increases. Our empirical findings suggest that 3G coverage is responsible for approximately 3,305 accidents per year in California.

Keywords: cellular phone coverage, car accidents, machine learning, random forest

JEL Codes: R41, C52, O33

*University of California, Riverside, Department of Economics. Email: bree.lang@ucr.edu.

†University of California, Riverside, Department of Economics. Email: matthew.lang@ucr.edu.

‡Chapman University, Argyros School of Business. Email: hersh@chapman.edu.

I. Introduction

Between 2000 and 2015, the number of mobile cellular subscriptions in the United States increased from 109 million to 382 million (International Telecommunications Union, 2017). The rise in mobile phones corresponds with an increase in cell phone functionality. In addition to being able to make a phone call, hand-held cellular devices are able to text message, browse the internet, video chat and analyze data anywhere there is an adequate wireless connection.[1] A potential consequence of the increased use of wireless devices is that users may become distracted by their phone while engaging in common activities, such as driving.

The potential costs of using a cell phone while driving has caught the attention of public health advocates, cellular providers, policy makers and researchers. The National Highway Traffic Safety Administration (NHTSA) frequently promotes distracted driving awareness campaigns[2] and AT&T has released controversial advertisements depicting fatal car crashes caused by cell phone use as part of their "It Can Wait" campaign.[3] There is strong public support for cellular bans while driving, with support for texting bans ranging from 86 to 98 percent.[4] All but two states have enacted a cellular phone ban for drivers (National Conference of State Legislatures, 2018) and the state-level bans are often used to proxy for cell phone use by drivers. Research relating cellular bans to traffic accidents are sensitive to the geographical setting, time period and severity of accidents being analyzed.[5] The inconclusive relationship between aggregate cell phone use and accident rates are in contrast to individual-level research, which finds that cell phone use reduces the quality of driving.[6]

In this paper, we predict 3G coverage along all road segments in California from 2001 to 2013 and examine how the frequency of accidents change before and after 3G coverage

[1] Wireless 3G coverage is reportedly available on the summit of Mt. Everest (Oberhaus, 2016).

[2] The NHTSA has turned to Twitter in order to personally tell drivers to stay off their phone (Matyszczyk, 2016)

[3] https://www.itcanwait.com/videos

[4] These numbers come from polls conducted by the New York Times (Connelly, 2009), the American Automobile Association (American Automobile Association, 2013) and the Insurance Institute for Highway Safety (Williams et al., 2011).

[5] Kolko (2009), Abouk and Adams (2013), Nikolaev et al. (2010), Sampaio (2014) and Burger et al. (2014) all examine the relationship between cell phone laws and accident rates.

[6] See Redelmeier and Tibshirani (1997), Abdel-Aty (2003), Strayer et al. (2003) and Törnros and Bolling (2005) for evidence showing that cell phones reduce the quality of driving and increase the likelihood of a crash.

becomes available. Poisson fixed-effect regression results show that car accident rates increase significantly by 1.1 percent when a road segment gains 3G coverage. In event-study analyses, there is no change in accident rates prior to a road gaining 3G coverage, but a significant and persistent increase in accidents after an area gains 3G coverage. The types of accidents most responsive to 3G coverage are those that do not involve bodily harm, take place in high traffic areas and where the driver at fault is younger than 65 years old. While the magnitude of our results are smaller than related research studying the determinants of fatal accidents, we find that 3G coverage is responsible for approximately 3,305 accidents per year in California.[7]

In order to carry out our empirical analysis, we construct a novel dataset that estimates annual 3G coverage for a quarter-mile stretch of road (road segment). Each road segment in California is assigned 3G coverage in 2016 using data from the Federal Communications Commission (FCC). Because accurate coverage data is not available prior to 2016, information on licensed antenna towers are linked to road segments in 2016. Machine learning techniques are used to determine which attributes of towers best predict 3G coverage. Tower information is available annually and the machine learning results are applied to road segments between 2001 and 2013, the time period in which 3G cellular towers were built. The resulting dataset consists of the annual number of reported traffic accidents along a road segment in California, annual traffic counts for each road segment and an estimate of whether 3G coverage is available.

The reliability of our results hinge on a number of assumptions. The first assumption is that the predicted coverage variable we create accurately measures the growth in 3G coverage. It is not possible to confirm the existence of 3G coverage before 2015, but we have evidence that our coverage estimates are accurate. The machine learning results are created using a random sample of 80% of data. The remaining 20% of the data is used as test data and our results accurately predict coverage in over 98.5% of road segments in the test data. Additionally, predicted growth in the fraction of road segments covered by 3G are strongly correlated with the growth in mobile broadband subscriptions over the time period of our analysis.

[7] The fatal accident rate increases by 15 percent after an individual turns 21 (Carpenter and Dobkin, 2009) and the accident rate in Tippecanoe County, Indiana increased by 47 percent in the months following the introduction of Pokémon Go (Faccio and McConnell, 2018). Adams et al. (2012) find that a 10 percent increase in the minimum wage is associated with a 5 to 10 percent increase in fatal accidents involving 16-20 year olds.

Another assumption we make is that the construction of new towers explains the growth in 3G coverage, but tower construction is not related to unobservable characteristics that increase accidents. Our empirical analysis controls for the annual average daily traffic (AADT) along a road segment, mitigating concerns that the results are biased by an increase in overall traffic. Additional analyses of the machine learning prediction method show the 3G coverage prediction is unlikely to be capturing unobservables that increase the likelihood of an accident. We are reasonably confident that the increase in accidents we observe along a road segment when 3G coverage is introduced is caused by drivers changing their cell phone related behavior.

The existing data is not able to identify what drivers may be doing on their phone when they enter a road segment that has 3G coverage. In areas that already had basic cellular coverage prior to gaining 3G coverage, our results are measuring the effect of moving from a 2G to 3G network. The primary difference between 2G and 3G coverage is that data can be transferred hundreds of times faster on a 3G network allowing users to quickly access data from the internet and use web-based applications (Qualcomm, 2014). If texting behavior does not change when drivers enter an area with 3G coverage, the results suggest that checking email and using phone applications while driving increase the likelihood of an accident.

It is also possible that the positive relationship between car accidents and 3G coverage is caused by an increase in text messaging. Although texting was available throughout much of California prior to 2001, the growth in text messaging coincided with the proliferation of 3G coverage and an increase in the functionality of cellular devices.[8] If drivers with 3G compatible phones are more likely to operate their phone when they have 3G coverage, and the phone is used primarily for texting, it is possible that text messaging is the primary cause of the increase in traffic accidents when 3G coverage becomes available. Because we do not know what exactly drivers are doing on their phones when they get in an accident, our results capture the overall impact of drivers becoming distracted as they enter an area with 3G coverage.

One contribution of our paper is that we are able to better inform policy makers about the

[8] In December, 2000, there were 14.4 million text messages sent in the US. By December, 2011, there were 193.1 billion text messages sent. These numbers are generated by the Cellular Telecommunications Industry Association (CTIA) and the Wayback Machine (https://archive.org/web/). The CTIA's Wireless Quick Facts (www.ctia.org/consumer_info/service/index.cfm/AID/10323) from April 20, 2010 and April 22, 2012 yield the monthly texts sent in the United States.

consequences of using a cell phone while driving. We improve upon many related studies that use state-level data and fatal accidents to estimate the relationship between cell phones and accidents. We analyze accidents along a one-quarter mile stretch of road and road segment fixed-effects capture time-invariant unobserved determinants of accidents, such as the inherent dangers along a road. These unobservable factors cannot be captured when data are aggregated to the county, state or national level. A policy implication of our paper is that existing cell phone laws do not appear to be effective, but developing laws that substantially alter how drivers use their cell phone may lead to a reduction in car accidents.

In addition to providing new and important insight into a relevant policy discussion, the current paper adds to the growing use of machine learning in economics (Athey, 2017; Mullainathan and Spiess, 2017).[9] Cicala (2017) builds random forest models to non-parametrically estimate operating rules for counterfactual electricity markets, which are then embedded in a difference-in-difference framework. Bajari et al. (2015) use machine learning based ensemble methods to estimate consumer demand, showing machine learning methods often outperform the standard demand estimation methods. Other researchers have used machine learning for predicting relevant information when surveys may be too expensive or of insufficient spatial or temporal resolution. Glaeser et al. (2017) use data on changes in businesses and restaurants on the website Yelp to forecast change in establishments and restaurants at small geographies in official data. Engstrom et al. (2017) and Babenko et al. (2017) generate poverty headcount rates for Sri Lanka and Mexico respectively, using machine learning models developed on features derived from high resolution satellite images. The novel application here is embedding a machine learning prediction into a causal model.

In the next section, we review the growth of mobile and broadband coverage. Section 3 introduces the data and develops coverage maps using machine learning techniques. Empirical results are presented in section 4. We discuss the implications of our findings in section 5 before concluding in section 6.

[9] The range of applications has led some researchers to suggest that machine learning methods should be taught as part of core empirical methods in graduate programs (Hersh and Harding, 2018).

II. Background

The first commercial mobile phone service in the United States was developed in 1983 by AT&T Bell Labs. Ten antenna towers ranging from 150- to 550- feet high covered 2,100 square miles around Chicago. This early analog system technology was costly, used sparingly and eventually replaced by a digital, second generation (2G) wireless system in the early 1990s. The 2G mobile system allowed for better voice quality, more efficient use of frequency bandwidths and improved security as simple encryption became possible. The 2G networks facilitated the creation of the Short Messaging Service (SMS), also known as text messaging and initially transferred data at a speed of around 10 kilobytes per second. Improvements in the structure of 2G networks and the quality of antenna towers increased the maximum downloading speed to 120 kilobytes by the late-1990s, but this speed was difficult for the typical user to obtain. By the early 2000s, technology for 3G networks had been developed, although the adoption of 3G coverage did not begin until the mid-2000s.[10] Devices on 3G networks could initially download data at a rate of 144 kilobytes per second when driving and 2000 kilobytes per second in a building environment. By 2005, 3G coverage data speeds was over 500 kilobytes per second for the average user.[11]

Improving wireless download speeds spurred growth in related industries. When Ameritech launched the first 1G network phone in 1983, the 2-pound DynaTAC 8000x retailed at $3,995 and could be used for 35 minutes before needing ten hours of charging (NBC, 2005). The growth in 2G networks led to the creation of smaller mobile phones that had longer battery lives and improved functionality. Browsing the internet on a 2G network was possible, but limited as a result of the slow data transmission speed.[12]

The growth of the faster 3G networks led to further improvements in the functionality of phones. When 3G service in the United States was first introduced by Verizon in 2002, two devices were capable of accessing the 3G network as long as the device was connected to a computer or had a PC card (CNN, 2002). In June 2007, Apple released the first iPhone (Apple,

[10]Prior to 2005, less than one percent of the United States population had a mobile broadband subscription, which increased to 52 percent by 2010 (TekCarta, 2018).

[11]See Ghosh et al. (2010) and Sauter (2013) for more information on the history and technology of cellular networks.

[12]By the end of the 1990s, mobile phones in Europe were being used to purchase goods from vending machines (Peña-López et al., 1999) and Japan released full internet service on mobile phones in 1999 (Ishii, 2004).

2007). The processing power of the 4.8-ounce first generation iPhone was faster than the 1980s supercomputer CRAY-1 and comparable to laptops in the year 2000 (Experts Exchange, 2018). The HTC Dream, with its QWERTY keyboard, was launched in 2008 and became the first smartphone to use Google's Android operating system (T-Mobile, 2008).

The influence of modern smartphones on individual lifestyles and behaviors is hard to understate. Increased smartphone functionality gave rise to a market of smartphone applications. By 2009, over 1 billion "apps" had been downloaded in Apple's "App Store" (Dormehl, 2018) and by 2012, both the Google Play and the App Store had over 500,000 unique apps available for download (Dogtiev, 2018). Smartphones are now used for many activities, including watching TV, getting directions, trading stocks and video calls. In 2017, Deloitte found that 89 percent of individuals check their phone within an hour of waking up and 81 percent look at their phone less than an hour before going to bed (Deloitte, 2017). Teens are estimated to spend nine hours per day on social media (Common Sense Media, 2015) and approximately 20% of adults are estimated to own a smartphone but do not have landline broadband internet at home (Pew Research Center, 2018).

Cell phone use is often compared to substance abuse and gambling disorders (De-Sola Gutiérrez et al., 2016) and cell phone addiction has been linked to depression, relationship issues and anxiety (Babadi-Akashe et al., 2014; Andreassen, 2015). Smartphone overuse can lead to physical problems such as eye strain, "text neck" and male infertility (Rosenfield, 2016; Lee et al., 2015; Deepinder et al., 2007). Researchers have found a negative correlation between cell phone use and grades in science courses (Douglas et al., 2012) and productivity at work (Thornton et al., 2014). Nasar and Troyer (2013) show that mobile phone related injuries among pedestrians that resulted in emergency room visits increased significantly between 2004 and 2010. Palsson (2017) makes a significant stride in measuring high-speed wireless coverage in the United States and shows that when a hospital becomes part of AT&T's 3G network, injuries for young children increase and the results are arguably driven by parents being distracted.

Considering that parents are less attentive to their children as a result of 3G coverage, it is reasonable to believe that drivers would also be less attentive when using their cell phone. Redelmeier and Tibshirani (1997) examine cell phone records from 700 accidents and find that the likelihood of a car crash increases significantly when a driver uses a cell phone. Studies from driving simulations find that drivers on cell phones have reduced peripheral detection (Törnros and Bolling, 2005), increased braking time (Strayer et al., 2003) and are more likely to be involved in a crash (Abdel-Aty, 2003). Reviews of laboratory and simulator studies by Caird et al. (2008) and Caird et al. (2014) suggest that texting while driving leads to increased reaction time, more collisions and worse lane positioning.

Despite the evidence that cell phone use reduces the effectiveness of individual drivers, there is less evidence showing that cell phone use increases the overall accident rate. The lack of evidence can be partially attributed to the difficulty in obtaining data that reports traffic accidents and cell phone accurately over a relevant period of time. Many studies use the adoption of cell phone and texting bans to proxy for changes in the use of cell phones, but the results of the studies vary based on the empirical approach used and time period examined (Nikolaev et al., 2010; Sampaio, 2014; Burger et al., 2014; Abouk and Adams, 2013; Kolko, 2009). Cheng (2015) does show that bans are associated with reductions in cell phone use at specific intersections, but it is possible that drivers respond to cell phone bans by increasing their use of a hands-free cellular device (Carpenter and Nguyen, 2015). Bhargava and Pathania (2013) exploit discontinuities in the "free nights and weekends" cellular plans from the mid-2000s and show that reducing the marginal cost of cellular calls at 9pm is not associated with a change in the rate of car accidents at the state level.

A recent survey of 3 million individuals over 3 months by Zendrive found that drivers use their phones on 88 percent of trips and are on their phone 3.5 minutes per hour of driving (Zendrive.com, 2018). Although many drivers use their cell phone, and cell phone use is related to poorer driving, researchers have been unable to consistently show that cell phone use leads to an increase in the overall accident rate. The current paper improves upon existing research by using a unique dataset and a new identification strategy to explore if gaining access to 3G cellular coverage increases the rate of traffic accidents. Our analysis provides important insight into the the relationship between cell phone use and traffic accidents that has been hard to observe in previous research.

III. Data Description and Predicting 3G Coverage

The first step in the construction of our data is defining a unit of observation. California defines their highways with a postmile system. A postmile value reports the mileage along a route, within a county. The postmile value starts at zero whenever a route crosses over a county border or when it originates.

In the analysis below, the postmiles of interest represent the road segment east or north until the next postmile on the route. This is done because postmile values increase to the west and south. Imagine a route runs west to east, the start of the route is postmile zero, one mile east is postmile one and two miles east is postmile two. The first road segment is defined between postmile zero (which lies on the west side of the road segment) and postmile one. The second road segment is defined between postmile one and postmile two. The same process occurs when the route runs from south to north.

California reports two different postmiles for each location on a highway, one for each direction. We combine these two points into a single postmile observation, which yields a total of 63,733 postmiles along California highways. The average distance between postmiles is approximately 0.25 miles, although the distance can be larger in rural areas. However, only 5% of the postmile observations represent road segments greater than one mile and only 1% are greater than 2.25 miles. Each postmile can be thought of as representing a road segment along a California highway. The postmile definitions do not change from year-to-year, so we use them as our unit of observation.[13].

A. 3G Coverage Data

The Federal Communications Commission's (FCC) Mobile Deployment Form 477 shapefiles provide information on mobile broadband coverage, by provider, on December 31, 2014 and December 31, 2015. Because the coverage maps are so similar between the two years, we focus on the shapefile that was published on December 31, 2015, but refer to it as the broadband coverage in 2016. The 477 files define and report broadband coverage in categories based on

[13]We only include postmiles that are in existence over the entire time period, so we do not account for road construction or destruction. For more information about postmiles, see `https://postmile.dot.ca.gov/`

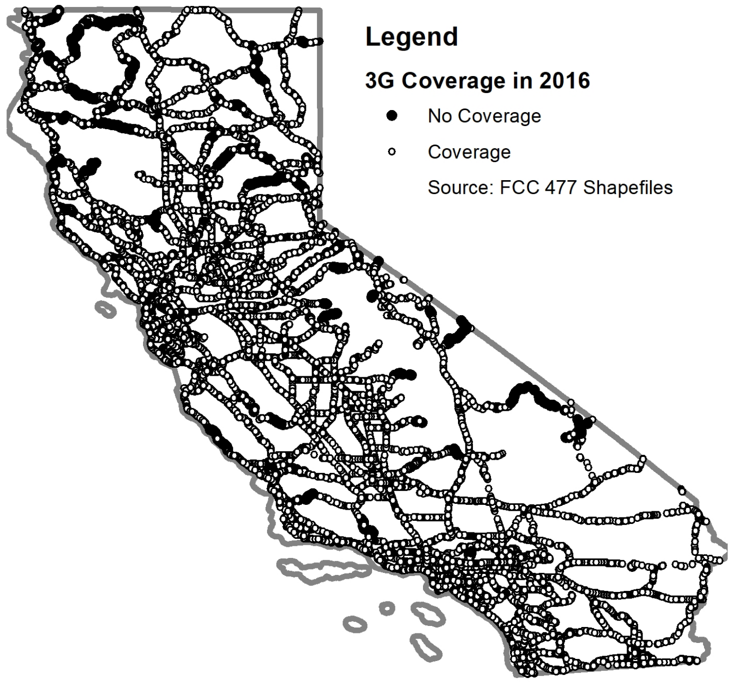


Figure 1: Postmile 3G Coverage in 2016

speed.[14] We build our coverage data using the shapefiles for coverage defined as 3G or faster. The data only includes the coverage maps for the four largest mobile service providers, measured by market share in 2016 (in parentheses): Verizon (34.9%), AT&T (32.3%), T-Mobile (16.8%) and Sprint (14.4%).[15]

Using ArcMap, software for Geographical Information System (GIS) analysis, we overlay the 3G coverage maps with the postmiles. Then we define a postmile as having 3G coverage in 2016 if the postmile intersects with the shapefile from the FCC. If the shapefile overlaps with the postmile (a point on the map), the entire road segment the postmile represents is assumed to have 3G coverage. This assumption is not strong, given the average distance between postmiles is only 0.25 miles.

We define 61,074 out of the 63,733 postmiles as having 3G coverage in 2016, or 97%. Figure 1 is a map of the postmiles, showing whether or not they are reported as having 3G coverage

[14] The codes for 2G coverage are 85 and 86. The codes for 3G coverage are 80 and 82. The codes for 4G coverage are 81 and 84. The code for 4G LTE coverage is 83.

[15] Market share statistics are from `https://www.androidheadlines.com/2016/11/top-7-us-wireless-carriers-q3-2016.html/`.

in 2016.

While the FCC shapefiles are useful for establishing coverage in 2016, there is no information reported between 2001 and 2013. Because this is the time period in which 3G coverage was expanding, we devise a strategy to predict coverage during that time period. We first assume that no road segments had 3G coverage prior to 2005, which is consistent with the fact that less than one percent of the population had mobile broadband service before 2005. A second 3G coverage data point is available by combining the information on 3G coverage in 2016 with information on the location and attributes of cellular towers. With 3G coverage and tower attributes at two points in time for each road segment, we can predict 3G coverage for the time period that accident data is available in California, 2001 to 2013.

B. Cellular Tower Data

The FCC requires all antenna structures over 200 feet high (towers) to be registered through the Antenna Structure Registration (ASR) system and information on tower location, elevation, year of construction and height is available publicly. Even though all structures over 200 feet high are required to be registered, roughly half of the towers in the registry are less than 200 feet.

The location of towers are geocoded using Geographical Information System (GIS) software and are laid on the map consisting of California postmiles. We restrict the sample of towers to those that were built after 1990, as older towers are less likely to be used for cellular coverage. In 2001, there were 3,101 existing towers within California's borders, which increased to 6,255 towers in 2013. In 2016, there were 6,556 towers. We create variables that measure the spatial relationship between postmiles and towers, based on information at the beginning of each year. Table I reports those statistics.

The first variable we create is the number of towers that exist within a 20-mile radius of each postmile. The range of coverage from a tower is unknown and varies. A tower's range depends on the type of cell tower built, the strength of the transmitter, the relative height, the surrounding landscape and the number of users in the area. In densely populated areas, such as business parks and campuses, picocell towers are common and have a range of 250 yards (Harris, 2011). Macrocell towers in rural areas can provide coverage more than 20 miles away (Heimerl et al., 2013). Based on estimates of maximum tower ranges, we determine that a 20-mile radius is a sufficiently large range to capture the relevant towers for each postmile. The average number of towers within a 20-mile radius in 2001 was 82.4 and increased to 170.4 in 2013.

Table I: Relationship between Postmiles and Towers: Descriptive Statistics

		Attributes of Tower Nearest to Postmile		
Year	Towers Within 20 Miles	Distance	Construction Year	Max Tower Elevation
2001	82.4	3.66	1996.2	455.9
2002	92.1	3.42	1996.8	452.4
2003	97.3	3.21	1997.1	449.7
2004	103.2	3.11	1997.5	447.4
2005	111.9	3.03	1997.9	447.0
2006	122.1	2.98	1998.2	446.7
2007	131.9	2.86	1998.7	448.8
2008	138.2	2.82	1999.0	447.5
2009	144.3	2.78	1999.3	449.4
2010	149.3	2.77	1999.5	448.3
2011	158.5	2.72	2000.0	446.5
2012	165.6	2.68	2000.4	448.8
2013	170.4	2.65	2000.7	445.6
2016	179.0	2.60	2001.4	445.7

Describes 63,733 postmiles in California. Distance is reported in miles. Elevation is reported in feet. The average elevation of the postmiles is 498.3 feet and the standard deviation of the elevation within 20 miles is 36.3 feet.

The remaining variables we create are based on the attributes of the tower that is closest to the postmile in a given year. The average distance of the closest tower decreased from 3.66 miles in 2001 to 2.60 miles in 2016; the average year of construction for the closest tower increased from 1996.1 to 2001.4 between 2001 and 2016; and the average elevation of the top of the nearest towers decreased from 455.9 to 445.7 between 2001 and 2016. The shortening of towers over time is consistent with an increase in demand for ”stealth towers” by citizens. Stealth towers are built to look like the surrounding area (Wikle, 2002). Assuming that new and closer towers are indicative of an area gaining 3G coverage, the tower data is consistent with the expansion of 3G coverage throughout California over this time period.

C. Predicting 3G Coverage with Machine Learning

Using information on whether a postmile has 3G coverage in 2016, we use machine learning methods to predict whether a postmile has 3G coverage during the years 2001-2013, during which we only observe detailed cellular tower characteristics and not 3G coverage. Specifically we fit a random forest model (Breiman, 2001) to predict 3G coverage based on characteristics of cellular towers close to the postmile.

We estimate a model using 2016 data to estimate if postmile i has 3G coverage y_i where

$$y_i = \begin{cases} 1 & \text{if postmile has 3G coverage} \\ 0 & \text{if postmile does not have 3G coverage} \end{cases}$$

The desired output is a model of conditional probability $p_i \equiv Pr(y_i = 1|\mathbf{X})$ where $\mathbf{X}$ is a vector of characteristics about cellular towers located near the postmile. Table II shows the full list of variables available to model 3G coverage. These include the number of towers within 20 miles of the postmile, the distance to the nearest cellular tower, elevation and standard deviation of elevation of the areas near postmiles, year of construction of the closest tower, among others. In total we use fourteen continuous variables to model the probability a postmile has 3G coverage.

The standard approach is to fit a parametric logistic regression of y_i on $\mathbf{X}$. There are several reasons why this approach is inadvisable for this prediction task. The relationship between y_i and $\mathbf{X}$ may be sufficiently non-linear such that we cannot assume to know the functional form mapping the two. Indeed, estimating a logit of y_i results in a very poor fit of the data. Without assuming the functional form of y_i we estimate a model of the form

$$p_i = (Y = 1|\mathbf{X}) = f(\mathbf{X}) + \epsilon$$

where $f(\mathbf{X})$ is estimated using a random forest. Random forests are a popular method for classification and regression prediction that produce reliable out-of-sample prediction. A random forest model is an ensemble method, meaning it combines many different models, each of which is a simple decision tree. Decision trees partition the target data (postmile 3G coverage) by finding a variable (e.g. height of tower) and a proposed split of that variable (e.g. less than 30 feet) that best partitions the target into groups that reduce within-group variance.[16] A series of variable selection and splits are performed until some stopping criterion is reached.

Random forests are comprised of many of these decision trees, however there is a catch in estimating each regression tree. Each time a decision tree is estimated, we bootstrap a subsample of data on which we estimate (also know as train) the decision tree. Further, at each decision node (or variable split), we sample which variables the decision tree can use to partition the response data. This may seem as if we are needlessly handicapping the model by reducing the variables the model can choose. However, in practice this approach works well for fully exploring the parameter space and preventing overfitting, or fitting noise rather than signal. Prediction in a random forest is performed by combining (or ensembling) all of the decision trees. The resulting p_i for any postmile i is the fraction of decision trees that predict the postmile with characteristics X_i have 3G coverage.

Our random forest model is estimated using 1,000 decision trees. We further partition our estimation dataset into a training and test set, of sizes 80% ($N_{train} = 254,912$) and 20% ($N_{test} = 63,728$) of the full estimation dataset respectively.[17] Our model is estimated against the training set and model diagnostics are performed on the test data – which was not used to fit the model and thus functions as a better approximation of out-of-sample performance. The remaining parameter to select is how many variables to randomly sample at each decision split. Appendix figure 10 shows cross-validation model accuracy across a number of values of the parameter. We achieve best cross-validated accuracy by randomly sampling five variables at every decision split.

A remaining assumption to make is the date of 3G cellular introduction. We set 3G coverage to be equal to zero from 2001 - 2004, giving 2005 as the first date of introduction of 3G service. This assumption best matches rates of mobile broadband in the US (TekCarta, 2018), which

[16] Classification problems such as this amounts to separating postmiles into those that have and don't have 3G coverage.

[17] Summary stats for the estimation and prediction stages are shown in table II. Summary statistics for the observations used for prediction are shown in panel B of the same table.

Table II: Summary statistics, 3G coverage estimation data

	mean	sd	min	max
Panel A: Summary statistics, data used to estimate postmile 3G coverage, 2001-05; 2016				
Towers within 20 mi. of postmile	111	138	0	843
Towers within 20 mi. squared	31,279	76,678	0	710,649
Distance to nearest tower from postmile	3.2	4.3	.0018	46
Distance to nearest tower squared	29	86	3.2e-06	2,130
Std. dev. of elevation of land within 20 mi.	36	11	7.2	64
Elevation at centroid of postmile	498	514	-21	2,665
Year of construction of nearest tower	1,998	4.4	1,990	2,015
Height in feet of nearest tower	30	22	0	584
Height squared	1,405	3,773	0	340,706
Height above sea-level of nearest tower minus height	-48	275	-1,373	1,850
Std. dev. elevation X Towers in 20 mi.	1102698	2987691	0	2.9e+07
Year of construction X height nearest tower	60,518	44,099	0	1176156
Year of construction X std. dev elevation	2888102	1709135	103,996	8222200
Towers within 20 miles X height nearest	2,906	4,767	0	158,246
Observations	318640			
Panel B: Summary statistics, data used to predict 3G coverage, 2006 - 2013				
Towers within 20 mi. of postmile	132	160	0	843
Towers within 20 mi. squared	42,846	90,502	0	710,649
Distance to nearest tower from postmile	2.9	4	.0018	46
Distance to nearest tower squared	25	76	3.2e-06	2,130
Std. dev. of elevation of land within 20 mi.	36	11	7.2	64
Elevation at centroid of postmile	498	514	-21	2,665
Year of construction of nearest tower	1,999	4.7	1,990	2,015
Height in feet of nearest tower	30	22	0	584
Height squared	1,369	3,027	0	340,706
Height above sea-level of nearest tower minus height	-50	271	-1,373	1,850
Std. dev. elevation X Towers in 20 mi.	1521921	3519389	0	2.9e+07
Year of construction X height nearest tower	60,134	43,046	0	1176156
Year of construction X std. dev elevation	2889420	1710047	103,996	8222200
Towers within 20 miles X height nearest	3,435	5,535	0	158,246
Observations	892237			

show broadband subscriptions per 100 persons in 2004 of 0.41. In 2005 the US achieved 1.03 mobile broadband subscriptions per 100 persons.[18]

The Receiver Operating Characteristic (ROC) plot from using the fitted random forest model to predict 3G coverage for observations in the test set is shown in figure 11. The ROC plot shows an exceptional model fit, with an Area Under Curve (AUC) of 0.9966. This is an impressive result given that these are diagnostics from predicted observations into the test set, which were not used to estimate the random forest model. The prediction task is all the more challenging given that the majority of postmiles have 3G coverage in 2016, and none had 3G coverage in '01 - '04.

Threats to identification embedded in the prediction model

One threat to identification is that the random forest model learns and predicts characteristics that increase the probability of traffic accidents that are correlated with changes in predicted 3G coverage but are not related to 3G coverage itself. An example is that changes in predicted 3G coverage may be correlated with road disrepair, which may increase the likelihood of traffic accidents in absence of 3G coverage.

One is naturally tempted to ask the random forest model how it calculates its predictions. While inference with these models is cumbersome, we can infer which variables are most useful for the estimation by constructing a variable importance plot, as shown in figure 2, according to the method described by (Breiman, 2001). The variable importance plot shows that the year of construction of the closest tower is the most important variable used to predict 3G coverage – over four times as important as the next variable. Other important predictor variables are elevation, number of cellular towers and the standard deviation of elevation near the postmile. The random forest appears to have "learned" that postmiles located close to few, old cellular towers, and near rocky terrain, do not have 3G coverage. We have no reason to believe why any of these are correlated with unobservables that positively impact the probability of traffic accidents, thus we are reasonably certain the prediction model is measuring changes in 3G

[18] Appendix figure 14 shows model diagnostics varying the assumption of 3G introduction from 2003, 2004, and 2005. Prediction performance is comparable across models varying this date of introduction.

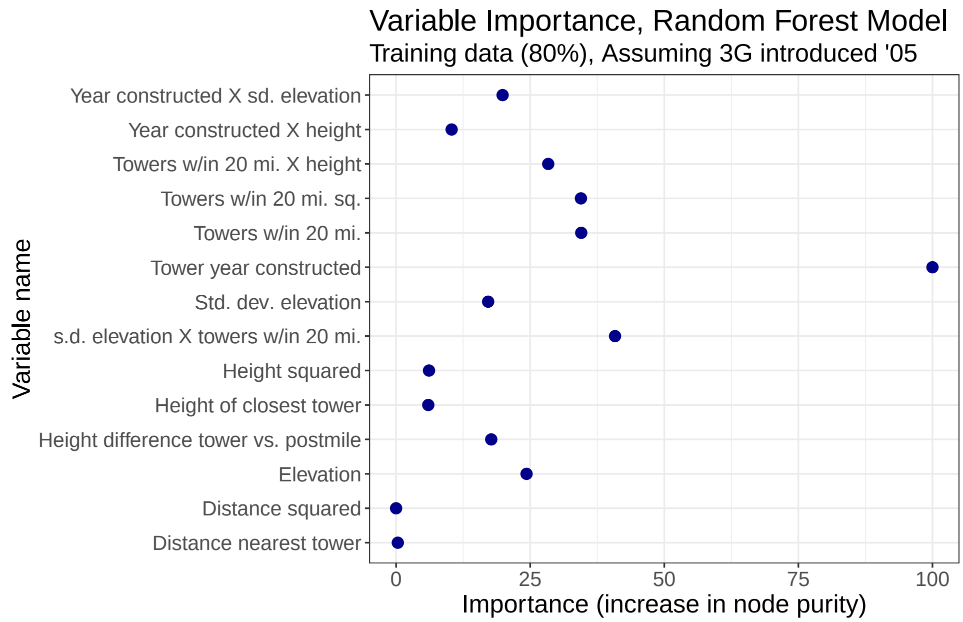


Figure 2: Random forest model variable importance plot

Variable importance plots are constructed according to (Breiman, 2001). The importance of variable j is calculated by taking each decision tree in which the variable appears and re-calculating the out-of-bag error. Importance is calculated by averaging over all of the decision trees the difference in out-of-bag error with and without this permutation. Scores are normalized by dividing by the importance score of the variable with the highest importance scores, thus the variable importance scores measure contribution relative to the highest variable.

coverage and not unobservables that are correlated with 3G coverage.

From continuous probabilities to binary 3G coverage prediction

To generate a binary prediction from the continuous probabilities of postmile 3G coverage p_i, we must define some threshold c such that

$$\hat{y}_i = \begin{cases} 1, & \text{if } p_i \geq c \\ 0, & \text{if } p_i < c \end{cases}$$

This is to say that we must make a threshold decision such that postmiles with probabilities below this cutoff will be classified as not having 3G coverage, and those above this threshold will be classified as having 3G coverage. The ROC curve, which presents the true and false positives from any threshold decision, presents guidelines for that decision process.

Our preferred threshold cutoff value is 0.8. This choice is motivated by the fact that when using this cutoff, observed mobile broadband subscribers most closely matches predicted 3G postmile coverage, as shown in figure 3. The correlation between observed mobile broadband coverage and predicted 3G postmile coverage is highest at this threshold, compared to discretized cutoff values at 10% increments. The consequence of various thresholds in terms of type I and type II error rates is shown in table III. We experiment with a variety of threshold values between 0 and 1. The accuracy of our predictions using 0.8 as a threshold is 98.63%. Choosing an alternative cutoff near 0.8 does not qualitatively impact the results, although the magnitude and significance of the impact are generally increasing with the threshold used.

Table III: Binary prediction accuracy varying cutoff threshold

Prediction Threshold	Accuracy	False negative rate (type 1 error)	False positive rate (type II error)	Correlation Coefficient
1	88.42%	60.52%	0.00%	0.9852
0.9	97.78%	11.43%	0.04%	0.9671
0.8	98.63%	6.89%	0.07%	0.9729
0.7	99.02%	5.55%	0.09%	0.9728
0.6	99.16%	3.78%	0.14%	0.9682
0.5	99.27%	3.01%	0.19%	0.9590
0.4	99.29%	2.52%	0.28%	0.9433
0.3	99.22%	2.13%	0.46%	0.9204
0.2	99.22%	1.78%	0.81%	0.8865
0.1	98.23%	1.51%	1.83%	0.8306

D. Car Accident and Traffic Data

The previous sections combine recent 3G coverage data along road segments with tower information to estimate the annual growth in 3G coverage between 2001 and 2013. A potential consequence of the increase in 3G coverage is that automobile drivers will be more distracted by their phones and the accident rate will rise. To explore this possibility, we merge 3G coverage information to the number of accidents that took place along a postmile each year. Accident data is available continuously in California between 2001 and 2013 from the California Highway Patrol (CHP) Statewide Integrated Traffic Records System. The CHP reports the location of a collision based on the county, route number and exact postmile. In most incidents, the age of the drivers in the accident are reported, but whether the cause of the accident is inattention and

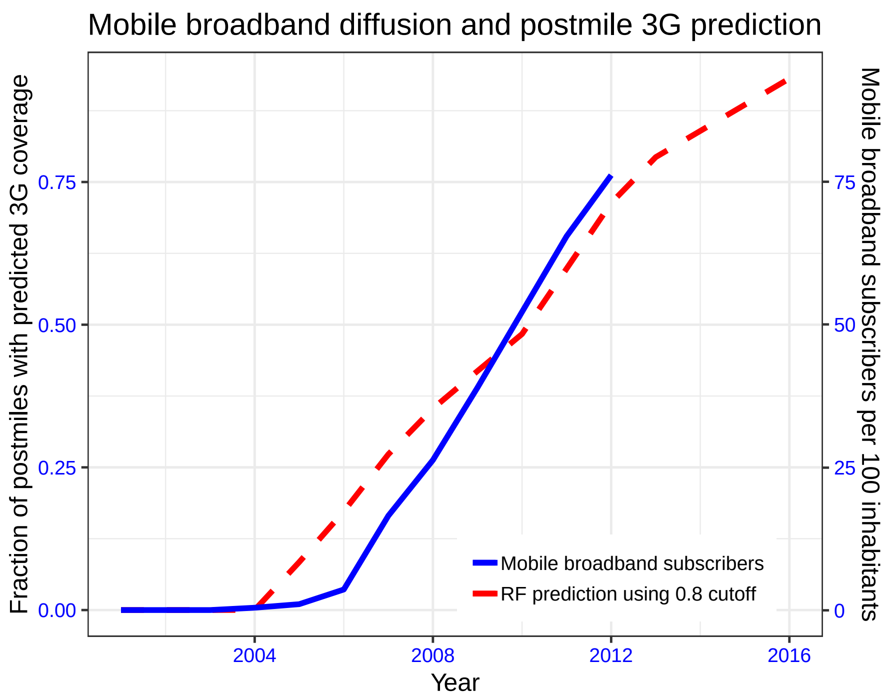


Figure 3: Mobile broadband diffusion and postmile 3G prediction over time

alcohol involvement is less reliable, as detailed recording likely differs across police jurisdictions.

The dependent variable in our analysis is constructed by first aggregating the number of accidents along a road segment each year. Each accident along a road segment is assigned to the nearest postmile to the west or south of the accident. We also utilize information on the the volume of traffic along a road segment, which allows us to construct an accident rate. Traffic volume is reported annually by the California Department of Transportation (CalTrans) through their Traffic Census Program.[19] Their specific metric is the Annual Average Daily Traffic (AADT), which is the total number of vehicles that travel along a road segment in a year, divided by 365. The AADT provides a measure of how busy the road is on an average day in a particular year.

Each year, Caltrans publishes traffic volume in a spreadsheet for approximately 4,000 locations in both directions. The locations are defined by county, route number and postmile. This information is available for every year between 2001 and 2013 except 2009, in which there was an error in the administrative data. We extrapolate values for 2009 using 2008 and 2010. We assign AADT values to the nearest postmile west or south of the reporting location and

[19] See `http://www.dot.ca.gov/trafficops/census/` for more information.

take the average of the two directions.

Using this process yields precise traffic volume data for 3,709 postmiles in our dataset. This is only a small subset of postmiles in our sample, so we use this data to construct average AADT measures for routes within counties. For example, imagine we only have traffic information for three postmiles of a route with 10 postmiles in a particular county. We take the average of the AADT on the three postmiles and apply that average to all 10 postmiles on the route in that county. Using this process, we are able to construct traffic value data for 61,804 of the 63,733 postmiles in our sample.

Because the sample of postmiles with traffic is a subset of all available postmiles, it is important to note that the accident counts are similar, regardless of the sample. Appendix figure 9 shows that the trend in accidents is the same for all postmiles, postmiles with traffic data, all postmiles that have at least one accident between 2001 and 2013 and postmiles with at least one accident between 2001 and 2013 that has traffic data. Accidents in all four postmile subsets are increasing until 2005, decrease from 2007 to 2010, then remain relatively stable until 2013.[20]

In the previous subsection, the most accurate measure of 3G coverage along a road segment occurs when the predicted probability of coverage is greater than 0.8. This distinct cutoff allows us to evaluate how traffic accident rates change as a result of 3G coverage. Table IV presents accident rate averages for all postmiles, postmiles that did not have coverage in 2005 but gained coverage by 2013 ("Changers"), postmiles that gained 3G coverage in 2005 ("Always Covered") and postmiles that did not gain 3G coverage by 2013. The first column of Table IV shows that the final dataset includes 51,727 postmiles and 70 percent of those postmiles gained coverage between 2006 and 2013.

The second column of Table IV reports the Annual Average Daily Traffic average for each group and shows that there is a noticeable difference in the level of traffic across the three subsets of postmiles. The postmiles that are always covered have a relatively high AADT average (100,192), compared to postmiles that gained 3G coverage between 2006 and 2013

[20] In the empirical analysis, we employ a fixed-effect count model, which omits the 16 percent of postmiles that do not experience at least one car accident over the time period of interest.

Table IV: Descriptive Statistics

	Postmiles	AADT All Years	Accident Rate All Years	2001	2005	2013	%Δ Accident Rate 2005-2013
All	51,727	61,784	0.94	1.07	0.99	0.77	-22%
Changers	36,305	61,493	0.92	1.05	0.97	0.75	-23%
Always Covered	3,904	100,192	0.87	0.91	0.90	0.64	-28%
Never Covered	11,518	49,680	1.06	1.18	1.12	0.89	-20%

Predicted Coverage Threshold=0.8

(61,493). Areas that do not gain 3G coverage have a lower AADT average (49,680) than both postmiles that were always covered and the postmiles that gained coverage in our analysis.

The accident rate for all years, 2001, 2005 and 2013 are reported for each subset of postmiles, as well as the universe of all postmiles in the data. The average accident rate on a road segment in all years is 0.94 per 10,000 AADT. This means that road segments that average 10,000 vehicles per day average 0.94 accidents per year. The AADT average for all years of 61,784 implies that the average road segment has 5.807 accidents per year.

Although the subsets of postmiles may be inherently different regarding their location and traffic patterns, the relative change in accident rates across the postmile subsets suggest that 3G coverage could have had an observable impact on accidents. Specifically, in 2005, the subset of postmiles that gained 3G coverage had an accident rate of 0.97 per 10,000 AADT, which decreased to 0.75 in 2013, a 23 percent decrease. Postmiles that had 3G coverage starting in 2005 observed a 28 percent reduction in their accident rate (0.90 to 0.64).

Table IV only provides information on three specific years making the comparison of accident rates across postmiles over time limited. Nonetheless, a first glance at the data suggests that areas that gained 3G coverage between 2006 and 2013 had a lower reduction in accidents than postmiles that had 3G coverage beginning in 2005. To more completely understand how accident rates change in response to 3G coverage, the next section explores the relationship in a fixed-effects regression framework.

IV. Empirical Analysis

A. Empirical Specifications

The distribution of car accidents is skewed heavily to the right with 41 percent of postmile-year observations having no accidents and 92 percent reporting fewer than 10 accidents. Because of the nature of car accident data, we propose a Poisson fixed-effects model to examine the relationship between car accidents and 3G coverage within a postmile. Assuming that the conditional mean assumption holds, the estimates are consistent with robust standard errors clustered on postmile (Cameron and Trivedi, 2005; Gourieroux et al., 1984).[21]

The exposure variable is average annual daily traffic (AADT) along the postmile route for a given year, so it enters the specification with its coefficient constrained to equal one.[22] The resulting Poisson specification is the following:

$$E[Accident_{it}|\cdot] = exp(\alpha_1 PredictedCoverage_{it} + ln(AADT_{it}) + \gamma_i + \tau_t + \upsilon_{it}). \qquad (1)$$

The variable $Accident_{it}$ represents the number of accidents that occur at postmile i in year t. $PredictedCoverage_{it}$ can represent one of various measures of predicted 3G coverage. The analysis will initially use a continuous measure of predicted coverage between zero and one and then an indicator variable defined by a threshold between zero and one. $AADT_{it}$ is the estimated annual average traffic volume for the postmile, γ_i is a postmile fixed effect and τ_t is a time fixed effect.

In specification 1, the estimated coefficient α_1 is interpreted as the percentage increase in the accident rate when a postmile gains 3G coverage. When using a strict cutoff to define 3G coverage, α_1 is the percentage difference in the accident rate between all the years in which $PredictedCoverage_{it}$ equals zero and all the years in which $PredictedCoverage_{it}$ equals one. The identifying assumption in equation 1 is that the change in 3G coverage in a postmile, caused

[21]Although not provided, we also estimate these models with a fixed-effects negative binomial, fixed-effects Poisson with bootstrapped standard errors and the quasi-maximum likelihood estimator with robust standard errors suggested by Wooldridge (1999) and Simcoe (2008). All of the results from these estimation strategies yield results that are similar in magnitude and statistical significance.

[22]If the coefficient is not constrained to one, the estimated coefficient of interest is not meaningfully altered.

by the construction of new cellular towers, is not related to unobservable characteristics that could influence the accident rate. While the regression does not have a large number of controls, the analysis is able to control for changes in the AADT along postmiles, reducing concerns that the results are capturing a change in traffic patterns. Postmile fixed-effects control for time-invariant road segment characteristics associated with traffic accidents, further strengthening our identification strategy.

Although specification 1 is able to capture many elements that explain traffic accidents, the regression is unable to show the dynamic change in accident rates before and after a postmile gains 3G coverage. To examine the annual effects of 3G coverage, we construct an event study specification with a four year lead and lag, which illustrates how the number of accidents changes when a postmile gains 3G coverage. The omitted category is the first full year of cell phone coverage.[23] The general equation is:

$$E[Accident_{it}|\cdot] = exp(\sum_{k=-4+}^{-1} \theta_k S_{it+k} + \sum_{j=1}^{4+} \theta_j S_{it+j} + ln(AADT_{it}) + \gamma_i + \tau_t + \upsilon_{it}). \quad (2)$$

The term S_{it} equals one when the postmile has its first full year of 3G coverage in time t and zero otherwise. The event study can only facilitate the use of binary treatment variables. The previous section suggests that the most accurate cutoff occurs when a road segment is defined as having 3G coverage when the predicted probability is above 0.8. However, we show results when defining the coverage threshold to be 0.4, 0.6, 0.8 and 1.

B. Results

Table V shows the results from the fixed-effects Poisson specification (Equation 1). The first two columns report results using continuous predicted 3G coverage. In column (1), the AADT or traffic volume, enters the regression as a control. According to the coefficient of interest, accidents in a postmile increase by 0.65 percent when predicted coverage increases by one standard deviation in a postmile.[24] When traffic volume is the exposure variable in column (2),

[23]A postmile is said to have coverage for the full year if the tower was built before January 1st of the given year.

[24]The standard deviation of predicted coverage is 0.41.

the strength of the results are unchanged and increasing predicted coverage by one standard deviation is associated with a 0.70 percent increase in the accident rate.

Columns (3) through (6) of table V report the change in the accident rate when 3G coverage is defined as a binary variable. In column (3), all postmile-year observations are defined as covered if the continuous predicted 3G coverage measure is 0.4 or above. The coefficient of interest in column (3) suggests that when predicted coverage in a postmile crosses the threshold of 0.4, the accident rate increases insignificantly by 0.58 percent. When the threshold is changed to 0.6 in column (4), the results show that accident rates increase significantly by 0.83 percent when a postmile gains 3G coverage.

Table V: Fixed Effect Poisson Regressions: Accidents and 3G Coverage

	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Coverage	0.0158***	0.0170***				
	(0.00613)	(0.00613)				
ln(Traffic Volume)	0.571***					
	(0.0357)					
I(Coverage≥0.4)			0.00581			
			(0.00408)			
I(Coverage≥0.6)				0.00834**		
				(0.00410)		
I(Coverage≥0.8)					0.0111**	
					(0.00493)	
I(Coverage≥0.99)						0.01686***
						(0.00484)

p<0.01, ** p<0.05, * p<0.10. Dependent variable is annual number of accidents at a postmile. All regressions include 672,451 observations, which represent 51,727 postmiles over 13 years. The exposure variable in columns (2) through (6) is traffic volume.

Our preferred definition of 3G coverage is when predicted 3G coverage in a postmile is 0.8 or above. Column (5) of table V suggests that when a postmile gains 3G coverage the accident rate increases significantly by 1.1 percent. In column (6), postmiles are not considered covered by 3G unless predicted coverage is greater than 0.99. Even when this extreme definition of coverage is used, the relationship between coverage and traffic accidents continues to be significant at conventional levels.

In the preferred specification, a postmile is considered to have 3G coverage if the predicted

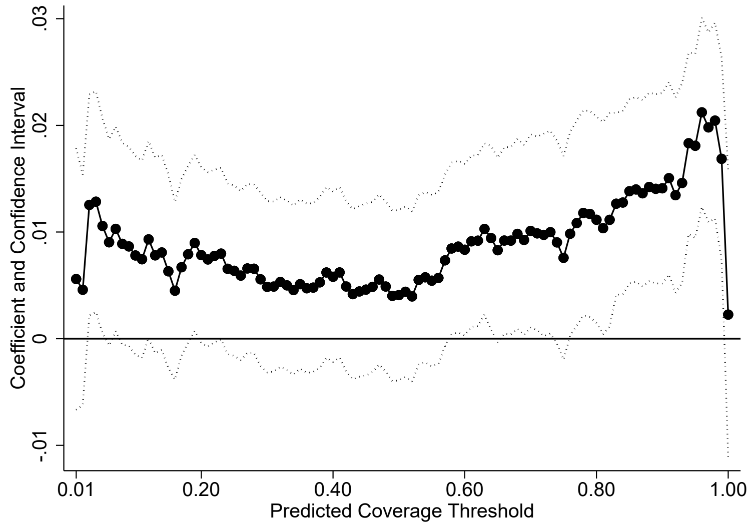


Figure 4: Coefficient Estimates and Confidence Intervals for Panel Regression by Predicted Coverage Threshold

Notes: This figure shows the coefficient estimate in a panel regression framework on an indicator variable equal to one if the predicted 3G coverage for a postmile is above a defined threshold.

probability of coverage is 0.8 or greater. Using this threshold aligns with subscription rates for mobile broadband coverage, but the results in table V are not sensitive to the threshold used over a relatively large range of cutoffs. Figure 4 shows the coefficient estimates and 95 percent confidence intervals when the coverage threshold is defined by every value from 0.01 to 1.00, in increments of 0.01. For nearly every threshold above 0.57, the coefficient of interest from equation 1 is significant at the 95% level. Of the thresholds between 0.58 and 0.99, only the cutoffs 0.65 and 0.75 are insignificant. The 0.8 cutoff is our preferred threshold, but the cutoff choice over a wide range of values does not meaningfully alter our findings.

Table V and figure 4 provide evidence that the introduction of 3G coverage is associated with a significant increase in the traffic accident rate. However, the results are not able to provide insight as to whether or not our results are picking up a larger trend in accident rates that could lead to a spurious correlation between 3G coverage and accidents. In order to better understand the dynamic changes in accident rates when 3G coverage is introduced, table VI reports the results from event study regressions, equation 2.

The four columns in table VI define 3G coverage using a different threshold. Consistent with the previous table, there is not a significant relationship between 3G coverage and traffic

Table VI: Event Study Coefficients

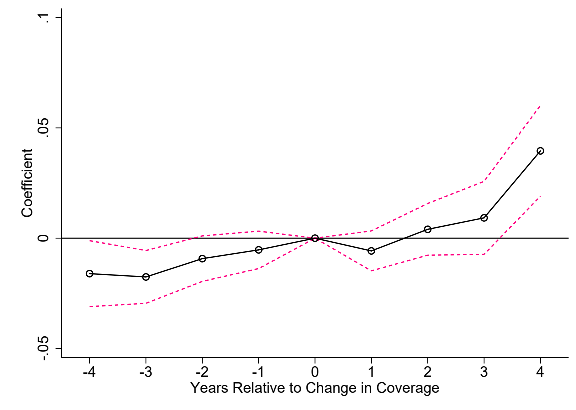
	(1) I(Coverage≥0.40)	(2) I(Coverage≥0.60)	(3) I(Coverage≥0.80)	(4) I(Coverage≥0.99)
4+ years before	-0.016**	-0.020***	-0.015**	-0.0093
	(0.0075)	(0.0073)	(0.0076)	(0.0096)
3 years before	-0.018***	-0.014**	-0.0058	0.0027**
	(0.0060)	(0.0059)	(0.0062)	(0.0081)
2 years before	-0.0093*	-0.00035	-0.0028	-0.0186
	(0.0052)	(0.0052)	(0.0054)	(0.0073)
1 year before	-0.0053	-0.0066	-0.0062	-0.0071
	(0.0042)	(0.0043)	(0.0046)	(0.0062)
1 year after	-0.0058	0.0052	0.0070	0.0097
	(0.0045)	(0.0046)	(0.0057)	(0.0064)
2 years after	0.0040	0.013**	0.021***	0.0227**
	(0.0059)	(0.0058)	(0.0071)	(0.0099)
3 years after	0.0092	0.022***	0.030***	0.033***
	(0.0083)	(0.0082)	(0.0086)	(0.0123)
4+ years after	0.040***	0.061***	0.076***	0.0490***
	(0.010)	(0.010)	(0.0091)	(0.0163)
Observations	632,216	597,246	531,401	315,627
Postmiles	48,632	45,942	40,877	24,279

p<0.01, ** p<0.05, * p<0.10. Dependent variable is annual number of accidents at a postmile. Traffic volume is the exposure variable in every column. Observations vary based on the number of postmiles that cross the defined predicted coverage threshold between 2001 to 2013.

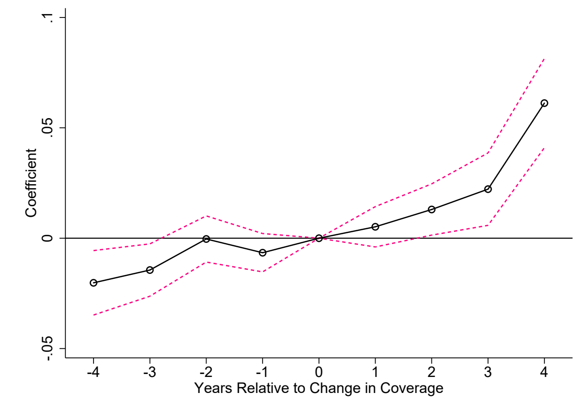
accidents when coverage is defined at a threshold of 0.4 or 0.99. Columns (2), (3) and (4) use the thresholds of 0.6, 0.8 and 0.99, respectively. The columns all show that 2 years after a postmile gains coverage, there is a significant increase in the accident rate. The increase in the accident rate persists 3 and 4 or more years after gaining coverage. In the years immediately preceding 3G coverage, there is not a significant difference in accident rates.

The results in table VI are conveyed visually in figure 5. Figure 5 shows the plots of the coefficients from table VI, along with the 95 percent confidence intervals. In panel (a), the threshold of 0.40 is used to define 3G coverage. There is a slight upward trend in the accident rate prior to the year 3G coverage is gained (year 0), an insignificant increase in the accident rate in years 2 and 3 and a significant increase 4 or more years after coverage is gained.

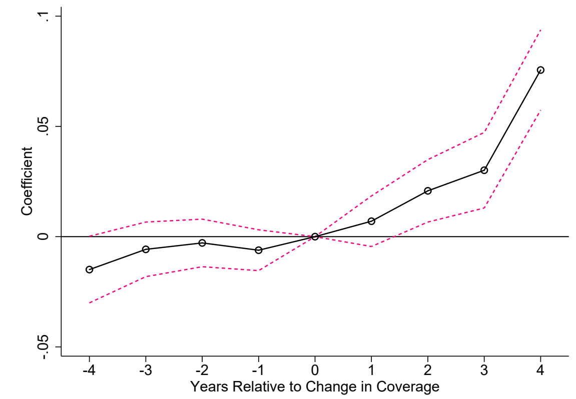
Using the threshold of 0.6 to define coverage in panel (b) generates a similar graph as panel



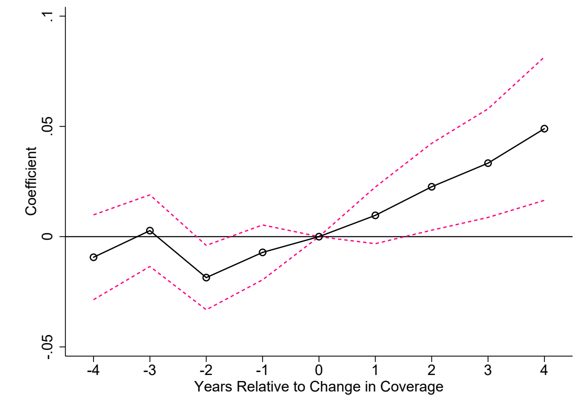
(a) ≥0.40 Threshold, N=48,632



(b) ≥0.60 Threshold, N=45,942



(c) ≥0.80 Threshold, N=40,877



(d) Threshold=0.99, N=24,279

Figure 5: Accident Event Studies Over Different Predicted Coverage Thresholds

(a). There are two noticeable differences between panels (a) and (b). First, in panel (b), there is essentially no change in the accident rate in the two years leading up to the change in coverage. Second, within two years after gaining coverage, the accident rate in panel (b) is significantly greater than in the year 3G coverage was acquired.

Panel (c) uses our preferred threshold of 0.8. The trend in the accident rate prior to the change in coverage is relatively flat. There is a noticeable rise in the accident rate after a postmile is defined as having coverage. When the threshold of 0.99 is used to define coverage in panel (d) a similar, but more pronounced pattern exists. It is worth noting the event study regressions only include postmiles that crossed the defined threshold. This causes each event study result to have a different number of postmiles, which are reported in the figure.

The results in tables V and VI, along with figures 4 and 5 provide evidence that there is a significant relationship between gaining 3G coverage and the traffic accident rate. Although event study results are only reported for four unique thresholds, any threshold between 0.58 and 0.99 yields similar figures, further strengthening the robustness of the results.

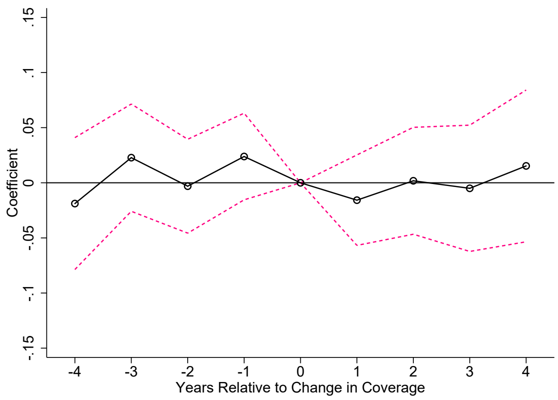
C. Heterogeneous Effects

Using the universe of highway postmiles in California shows that introducing 3G coverage is associated with a significant increase in the accident rate. However, the effect of using a cell phone while driving may depend on many factors, such as the traffic patterns along a postmile, the age of the driver at fault and the severity of the accident.

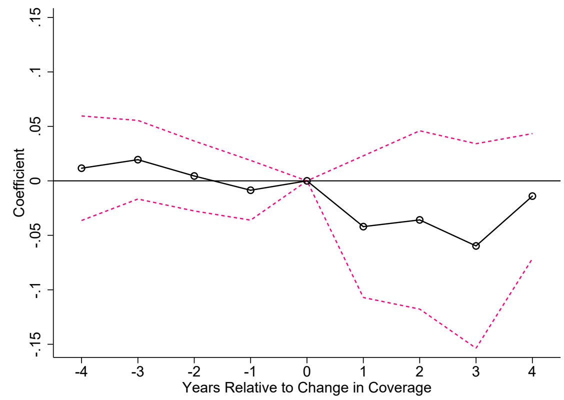
Postmile Traffic Volume

Postmile fixed-effects do control for many unobserved, fixed characteristics associated with a road segment, but it is possible that areas with less traffic are impacted by 3G coverage differently than areas with more traffic. Figure 6 shows event study results using our preferred threshold of 0.8 after creating subsets of postmiles based on the average traffic volume in the postmile between 2001 and 2013. In panel (a) the results using postmiles with an average AADT below the first quartile are reported. No pattern emerges and none of the coefficients are significant at conventional levels. Similar conclusions are drawn when examining the second and third quartiles of traffic. The postmiles with the most traffic are reported in panel (d). One year following the onset of 3G coverage, there is a significant increase in the traffic accident rate and there is no evidence of a trend prior to coverage.

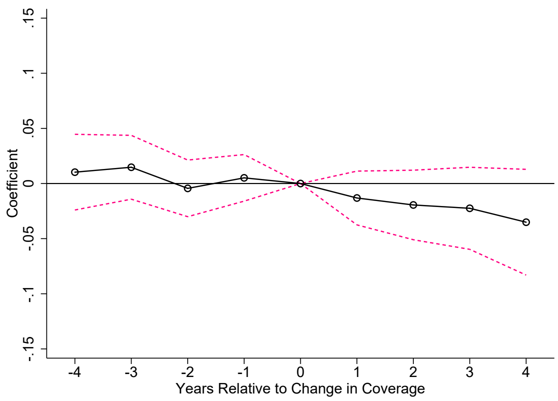
The results from figure 6 provide evidence that postmiles with high traffic volume are responsible for the main results above. Postmiles with less traffic do not have significantly different accident rates before and after 3G coverage becomes available. Although we cannot weigh in on differences in driving behaviors across the quartiles of traffic, 3G coverage appears to be more detrimental when traffic volume is high.



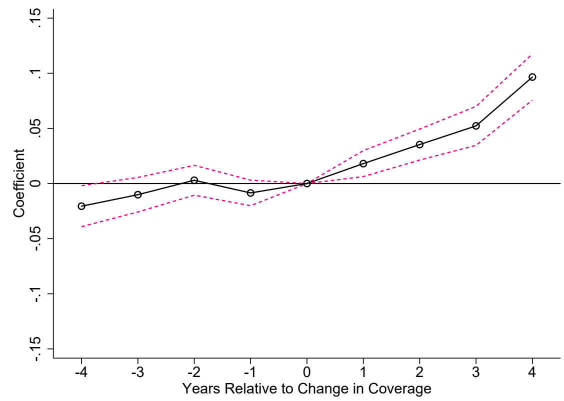
(a) First Traffic Quartile, N=9,301



(b) Second Traffic Quartile, N=10,207



(c) Third Traffic Quartile, N=10,131



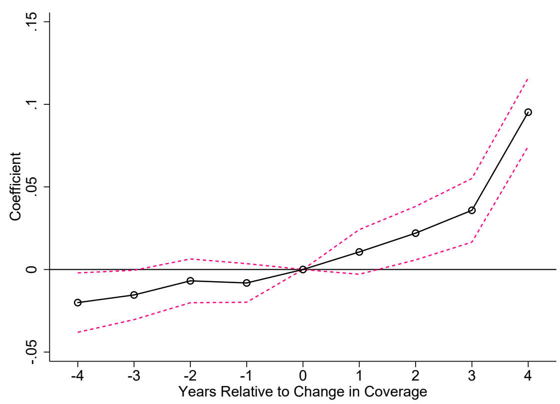
(d) Fourth Traffic Quartile, N=11,238

Figure 6: Heterogeneous Effects: Traffic Volume, Predicted Coverage Threshold=0.80

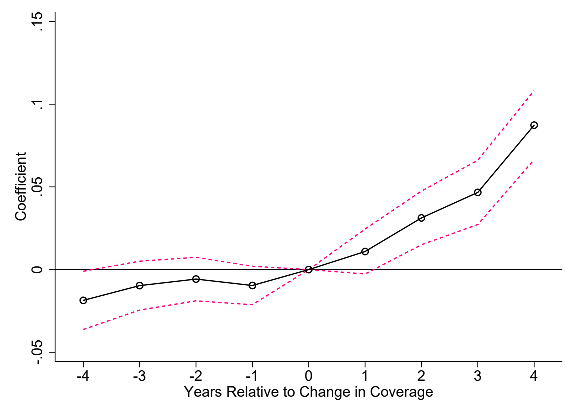
Age of the Driver At Fault

Although many details of the traffic accident data are not consistently reported within and across police jurisdictions, we are reasonably confident that the age of the driver at fault in an accident is accurately reported. According to multiple studies discussed above, younger individuals are more likely to use and own a smartphone, compared to older individuals. This would suggest that younger drivers would respond more strongly to 3G coverage than older drivers. However, it is possible that younger drivers are better at using a smartphone than older drivers and an older driver with a smartphone may be more likely to get in an accident than a younger driver with a smartphone.

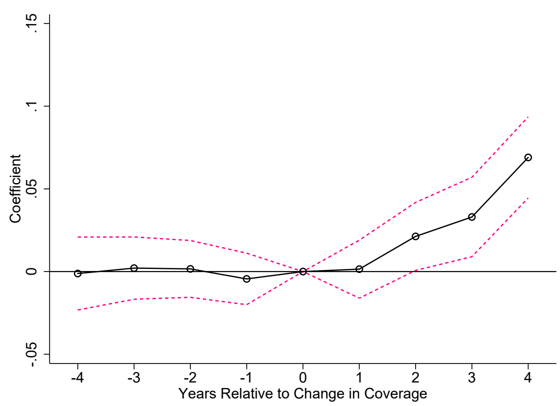
Figure 7 shows the results of event study regressions, again using 0.8 as the threshold for coverage, and stratifying the sample based on the age of the driver deemed at fault in the accident. In panel (a), the effect of 3G coverage on accident rates where the driver at fault is younger than 29 years old is reported. A familiar pattern is seen. There is not a significant difference in accident rates prior to 3G coverage, but accident rates increase significantly two years after a postmile obtains 3G coverage. Panels (b) and (c) report the results for accidents where the driver at fault is between 30 and 49 years old, and 50 and 64 years old, respectively. The same pattern is seen between all three age groups in panels (a), (b) and (c). In panel (d), 3G coverage does not change the accident rate in crashes where the driver at fault is 65 years old or older.



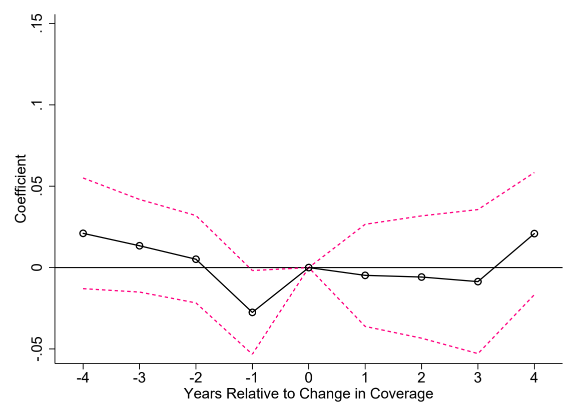
(a) 29 Years and Younger, N=36,880



(b) Between 30 and 49 Years, N=37,377



(c) Between 50 and 64 Years, N=33,857



(d) 65 Years and Older, N=27,494

Figure 7: Heterogeneous Effects: Age, Predicted Coverage Threshold=0.80

The age-specific event study results in figure 7 show that accidents caused by older individuals do not change when a postmile gains 3G coverage. Accidents caused by individuals between 16 and 64 increase in a postmile when 3G coverage is gained. The results suggest that

3G coverage has an impact on the driving behavior across a wide age range. However, we are unable to decipher whether this is because there is an equal increase in cell phone use across the majority of ages while driving when 3G coverage is available or because cell phone use impacts drivers differently, depending on their age.

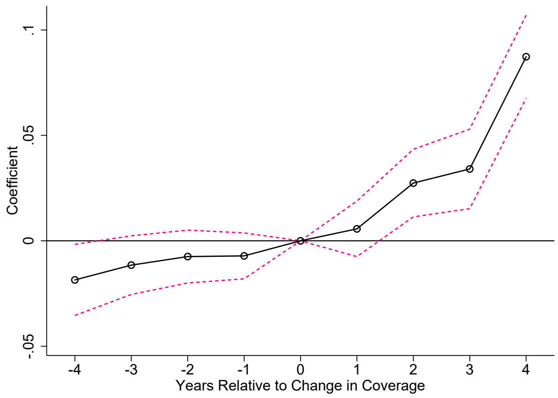
Severity of the Accident

Each accident reported in the data is assigned to one of five general severity categories. Accidents that only damage vehicles are labeled as "property damage only". The category of "no visible bodily harm" captures accidents where the most severe outcome is that an individual says they feel pain, but it is not visible. Whiplash falls into this category. Accidents categorized as "Non-Severe, Visible Bodily Harm" have individuals with cuts or bruises, but the wounds are not severe. "Severe Bodily Harm" implies that an individual requires immediate assistance and has a critical injury. Accidents with a fatality are graded as the most severe type of accident.

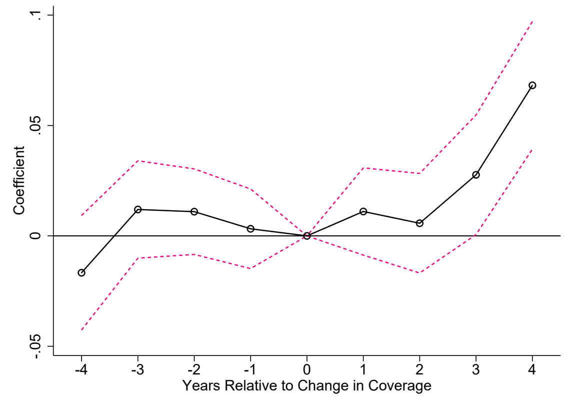
Figure 8 shows the effect of 3G coverage on accidents, by severity. Event study results from "property damage only" accidents are presented in panel (a) and mimic event study results above. After a postmile gains 3G coverage, less severe accidents increase and there is not evidence of a strong trend prior to gaining coverage.

Results for accidents where an individual has a non-visible injury are reported in panel (b). There are similarities between the pattern found in panel (b) and the least severe accidents in panel (a), but the estimates using accidents with non-visible injuries are relatively noisy. Panels (c), (d) and (e) report accidents that are increasingly severe. There is not a strong relationship between 3G coverage and accidents that involve visible injuries, severe bodily harm or fatalities.

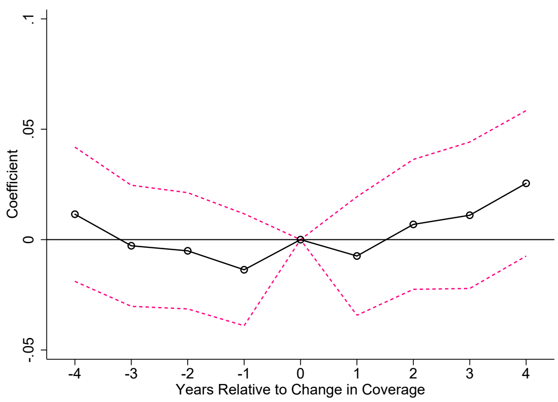
The results in figure 8 suggest that minor traffic accidents are driving the main results in the previous section. The relationship between accidents and 3G coverage becomes weaker as the severity of the accident increases. It is noteworthy that the number of postmiles in the regressions decrease significantly as the severity of the accident being examined rises. This is because there are relatively few postmiles that have a fatal accident in the data and those postmiles without a fatal accident between 2001 and 2013 are automatically dropped from the regression. Nonetheless, the insignificant relationship between severe accidents and 3G coverage



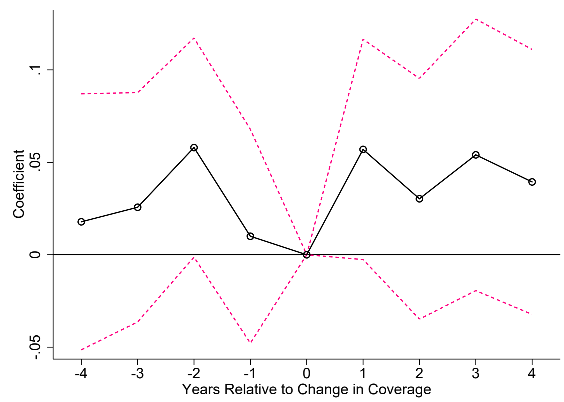
(a) Property Damage Only, N=37,765



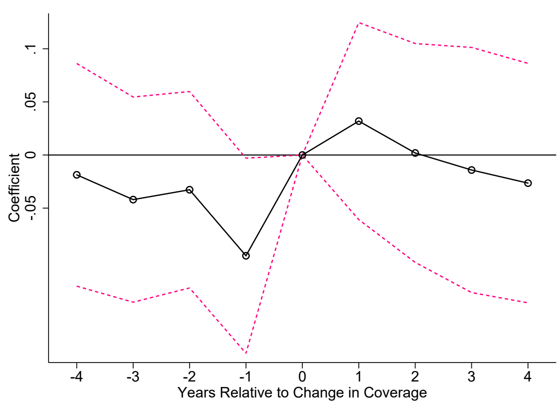
(b) No Visible Bodily Harm, N=30,601



(c) Non-Severe Visible Bodily Harm, N=28,617



(d) Severe Bodily Harm, N=12,254



(e) Fatal, N=8,367

Figure 8: Heterogeneous Effects: Severity of Accident, Predicted Coverage Threshold=0.80

is consistent with many of the previous studies that find a limited relationship between cell phone use and fatal accidents.

Unreported Regressions

In addition to examining accidents by severity, age of the driver at fault and the traffic volume of the postmile, we investigate the potential effect of 3G coverage along a number of other dimensions that are worth mentioning [25]. In the time period we examine, California enacted two cell phone laws. A hand-held cellular device ban while driving was enacted on July 1, 2008. Using monthly accident data in 2008, we do not find a significant change in the accident rate along postmiles with or without 3G coverage, after the ban went into effect. The insignificant finding is consistent with Burger et al. (2014), who explore California's hand-held ban in detail.

California enacted a specific texting ban six months later on January 1, 2009. We do find that there is a reduction in monthly accident rates between December, 2008 and January, 2009, but the reduction is not dependent on 3G coverage. This finding is not surprising as nearly all postmiles in California had 2G coverage by the early 2000s. Consequently, there is not a strong counter factual and the reduction in accidents after the ban must be interpreted with extreme caution.

Each observation in our data has information on the time-of-day that an accident took place. We use this information to explore whether 3G coverage impacts drivers more during the day or at night. Daytime accidents are broadly defined as accidents that occurred between 6am and 6pm and nighttime accidents are defined as accidents that occurred between 6pm and 6am. Daytime and nighttime accident rates increase in a similar way when a postmile gains 3G coverage. It is possible that assigning a specific sunlight index to each accident could yield a different conclusion. However, when daylight is defined as 5am to 7pm or 7am to 7pm, the effect of 3G coverage is the same across daytime and nighttime accidents. Our findings suggests that the effect of 3G coverage is not dependent on the time-of-day.

Unreported regressions that stratify the data into different subsets based on the attributes of accidents do not alter the findings. We also show that the results are not dependent on the

[25] All results discussed in this subsection are available upon request.

inclusion of a particular year. When any single year is dropped from the sample, there is not a meaningful change in the results. Regressions that isolate the effect of gaining coverage by year show that gaining 3G coverage between 2004 and 2007 are associated with the strongest increase in accidents. However, gaining coverage during the Great Recession (2008 and 2009) was associated with a decrease in the accident rate. We also find negative effects associated with gaining coverage during the last two years of the sample (2012 and 2013). The estimates in these later years are derived from a relatively small number of observations, which may partially explain the relationship we find in 2012 and 2013.

When a probit model and linear input variables are used to predict coverage, our conclusions are unchanged. The machine learning techniques employed in section III increase the precision of our coverage estimates, but our main findings are not altered by the use of machine learning.

V. Discussion

Many studies exist that examine the relationship between cell phone use and driving behavior. Individual driving behavior is impaired when a driver is using a phone, but data limitations have made it difficult for researchers to find a link between cell phone use and traffic accident rates. We overcome many of the limitations in previous research by exploring how all traffic accidents respond to changes in 3G coverage.

Our analysis first attaches cellular tower information to postmiles for each year of data. The FCC reports 3G coverage data in 2016 and we assume that 3G coverage was not available prior to 2005. Machine learning techniques allow us to estimate how the construction of cellular towers change the predicted level of 3G coverage along a postmile for every year that accident data is available.

Poisson fixed-effects regressions compare postmile accident rates before and after the introduction of 3G coverage. We find that accident rates increase in a postmile after a tower is constructed and a postmile gains 3G coverage. The identifying assumption in the empirical analysis is that unobserved characteristics that influence traffic accidents are unrelated to the construction of cellular towers. Although we cannot directly test whether or not our identifying assumption holds, we are confident in our results. Regressions incorporate changes in traffic volume, reducing concerns that the effect of 3G coverage is being confounded by an increase in traffic. Postmile fixed-effects capture many important unobservable characteristics of a road segment that are related to accidents, such as road quality, line-of-sight and unchanging traffic patterns. Event study results mitigate concerns that the increase in accidents is part of a larger trend that began prior to a postmile gaining 3G coverge. Any remaining endogeneity in our analysis comes from a systematic change in driving behavior that is unrelated to cell phone use and happens to occur when a cellular tower is constructed.

In summarizing the results above, a plausible story emerges. Cell phone tower construction increases the likelihood that a road segment has 3G coverage. Postmile accident rates in the years preceding 3G coverage are statistically similar to the year coverage is acquired. In the year after gaining coverage, there is a subtle increase in the accident rate. Two, three and four or more years after a postmile gains coverage, the accident rate increases significantly. Because changes in 3G coverage results are determined largely by the construction of cellular towers, the delayed accident rate response is not surprising. Historically, cellular towers have not been constructed by cellular carriers such as Verizon or AT&T. Instead, firms that specialize in cellular tower construction build the towers and then lease the tower to cellular providers. This suggests that it is possible for a tower to be constructed near a postmile, leading us to assign the postmile 3G coverage, but a lag between tower construction and service suggests that users may not change their driving behavior until the following year.[26]

Gaining 3G coverage does not influence all road segments in the same way. Road segments that have relatively low traffic volume do not see a significant change in traffic accidents after 3G coverage becomes available. Accidents where the driver at fault is over 65 years old are not impacted by 3G coverage. The trend in severe accidents where bodily harm is visible or there is a fatality does not change when 3G coverage becomes available. This suggests that 3G coverage increases the likelihood of a non-severe accident occurring in areas of high traffic and for drivers under the age of 65. In other words, when a road segment gains 3G coverage, fender

[26] In 2017, Verizon and AT&T signed a contract with the cellular tower construction company, Tillman Infrastructure, where Tillman builds towers specifically for Verizon and AT&T, reducing the time between tower completion and usage (Reuters, 2017).

benders in high traffic areas where the driver at fault is relatively young are more likely.

The magnitude of the results are comparable to previous research using traffic accidents as an outcome. Our point estimates suggest that the accident rate in a postmile is 1.1 percent higher when there is 3G coverage, compared to there being no 3G coverage. The average postmile in our data over all years has an accident rate of 0.94 per 10,000 AADT and an average AADT of 61,784. If all 51,727 postmiles in the data gained 3G coverage, a 1.1 percent increase would translate into 3,305 more accidents per year in California.

A 1.1 percent increase in the accident rate is comparable to work by Adams et al. (2012), but noticeably less than Carpenter and Dobkin (2009) and Faccio and McConnell (2018). Adams et al. (2012) use changes in state-level minimum wages to show that fatal accidents involving 16 to 20 years increase 5 to 10 percent when there is a 10 percent increase in the minimum wage. Carpenter and Dobkin (2009) show that motor vehicle fatalities increase by 15 percent after an individual becomes legally allowed to drink at the age of 21. Faccio and McConnell (2018) conclude that Pokémon Go was responsible for 134 crashes, a 47 percent increase, in Tippecanoe County between July 6, 2016 and November 30, 2016, the 5-month time period in which the game Pokémon Go became popular. Being legally allowed to drink alcohol and addictive mobile games are strongly related to traffic accidents, but the effects found in these studies are temporary by nature. Both alcohol consumption and motor vehicle fatalities decrease as individuals age (Carpenter and Dobkin, 2009). Pokémon Go users peaked at approximately 25 million users per day in July, 2016 then quickly fell to less than 7 million users by January, 2017 (Windels, 2017). General cell phone use grew throughout the 2000s and continues to grow today (Deloitte, 2017).

The majority of previous studies do not find a strong relationship between cell phones and accident rates, but our analysis differs considerably from those studies. Arguably the most important difference lies in our measurement of cell phone use. Previous papers using laws to proxy for changes in cell phone use rely to some extent on self-enforcement. Cell phone use by drivers may lead to more traffic accidents, but if a cellular ban does not lead to a persistent change in driver behavior, observing a change in accidents is difficult to detect.

Using the introduction of 3G coverage to capture cell phone use is beneficial because in areas without 3G coverage, the benefits of a smartphone are limited; only calling and texting are available in areas with 2G coverage. This does highlight a limitation of using 3G coverage to identify effects. If an area that gains 3G coverage previously had 2G coverage, we are capturing the effect of a postmile moving from 2G to 3G coverage. Assuming that a driver is texting when there is 2G coverage and using other smartphone functions when there is 3G coverage, our positive findings then suggest that the functionality of the smartphone is more distracting than texting. It is possible that a driver is less likely to use a smartphone to text when there is not 3G coverage and entering 3G coverage increases the use of all smartphone functions. We cannot observe what drivers are doing when they enter an area with 3G coverage, and are unable to provide more insight into the exact mechanism causing the increase in accidents when 3G coverage becomes available. Despite this limitation, the use of 3G coverage as a measure of cell phone use improves upon previously used proxies.

In many of the studies, fatal traffic accidents are used (Abouk and Adams, 2013; Kolko, 2009). In figure 8, less severe accidents increase when 3G coverage becomes available, but more severe accidents and fatal accidents are not impacted by 3G coverage. Our results suggest previous studies may not find a persistent relationship between cell phone use and accidents because fatal accidents are used as the dependent variable. Additionally, the current paper is able to examine traffic accidents at the postmile level as opposed to the state level, which allows us to capture important unobservable characteristics that may influence accident rates. Using machine learning techniques to estimate the relationship between traffic accidents and 3G coverage at a fine geographic level yields an important finding that researchers have been unable to identify: cell phone use leads to a persistent increase in traffic accidents.

VI. Conclusion

This paper uses the construction of cellular towers between 2001 and 2013 in California to show that accident rates increase when a road segment gains 3G coverage. Estimates of 3G coverage are derived from historical cellular tower information, current 3G coverage and machine learning techniques. The main results show that traffic accidents in a postmile increase after gaining 3G coverage. Event study results show there is not a trend in accidents within a postmile prior to gaining coverage and the rise in accidents is most noticeable two, three and four or more years after gaining coverage. The findings are not sensitive to the threshold used to define coverage, the time-of-day that the accident took place or using a probit model to predict coverage instead of machine learning.

We contribute to the literature on cellular phone use and accidents. Unlike much of the previous work, we show that there is a significant and persistent increase in accidents following the introduction of 3G coverage. Although we cannot identify changes in specific driver behavior, heterogeneous results show that accidents that increase after 3G coverage is introduced do not involve severe injuries, occur in highly trafficked areas and are caused by drivers younger than 65.

These detailed results speak directly to policy makers. Banning cell phone use by drivers has not effectively reduced traffic fatalities, but laws requiring cars to have certain safety features may prove to be more influential. Blind spot monitoring, lane drift alerts and collision avoidance systems are becoming increasingly common. Effective in May, 2018, back-up cameras are required on all new vehicles less than 10,000 pounds (Bomey, 2018). Advancing car safety features may be able to overcome the increase in distraction caused by increasing cell phone usage by drivers.

VII. Appendix

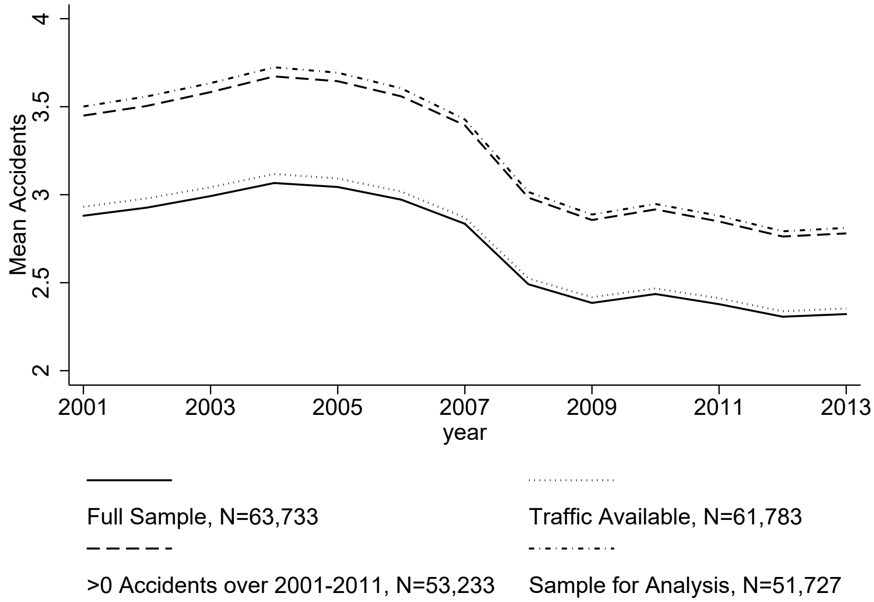


Figure 9: Average Accidents Over Time by Sample of Postmiles (N)

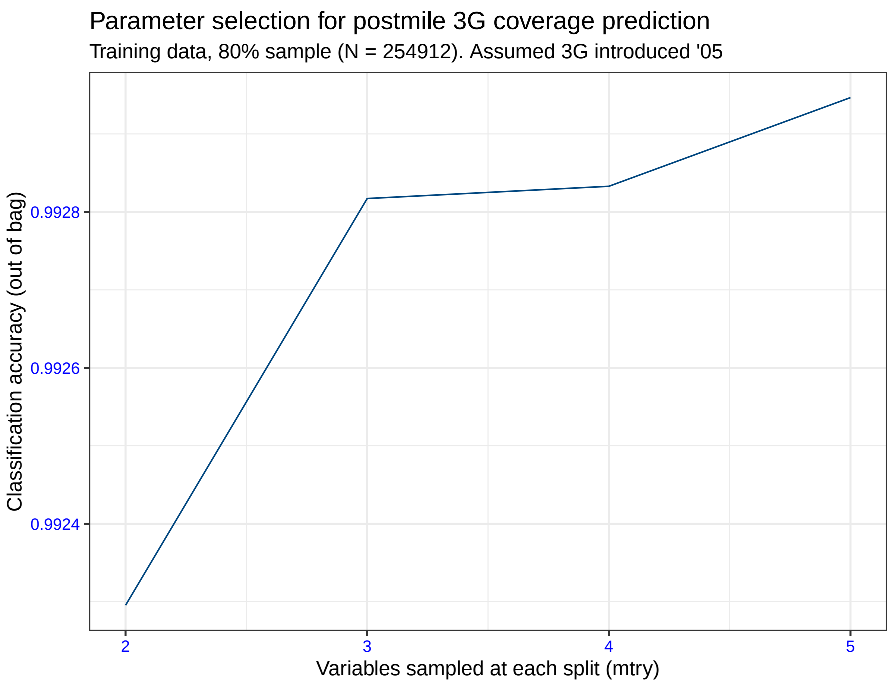


Figure 10: Random forest parameter selection

This plot shows cross-validated performance varying the number of variables sampled at each split. This is implemented by cross-validating over the parameter *mtry* in the *caret* package.

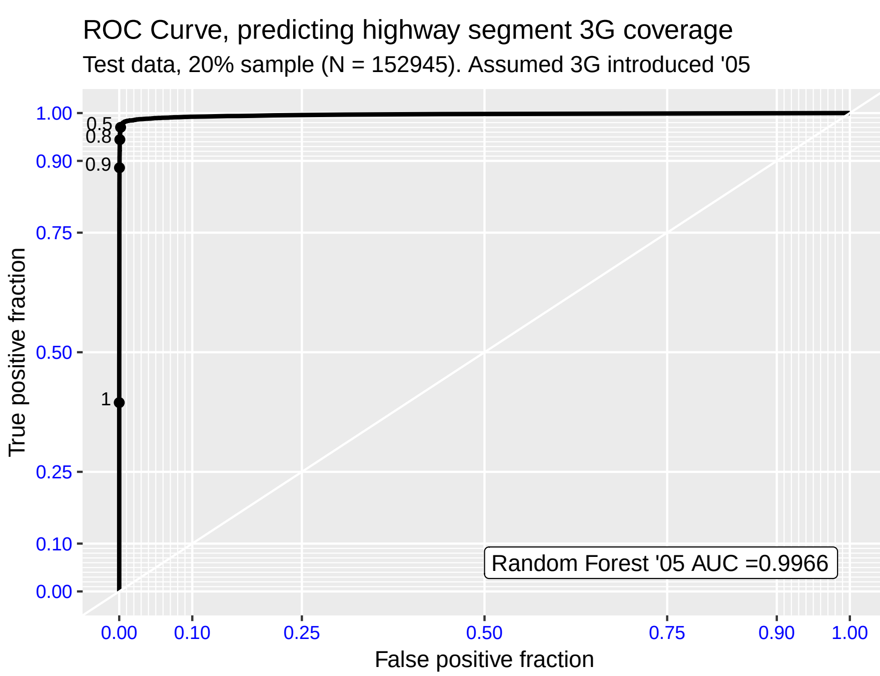


Figure 11: ROC Curve

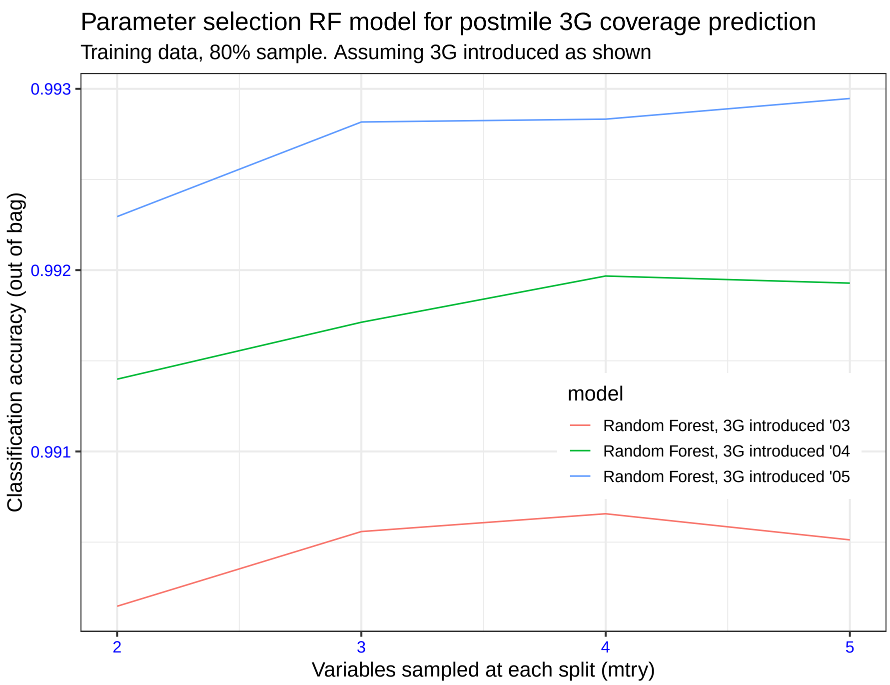


Figure 12: Parameter selection for models varying assumption of 3G introduction
This plot shows cross-validated performance varying the assumed year of 3G introduction, and varying the number of variables sampled at every decision node.

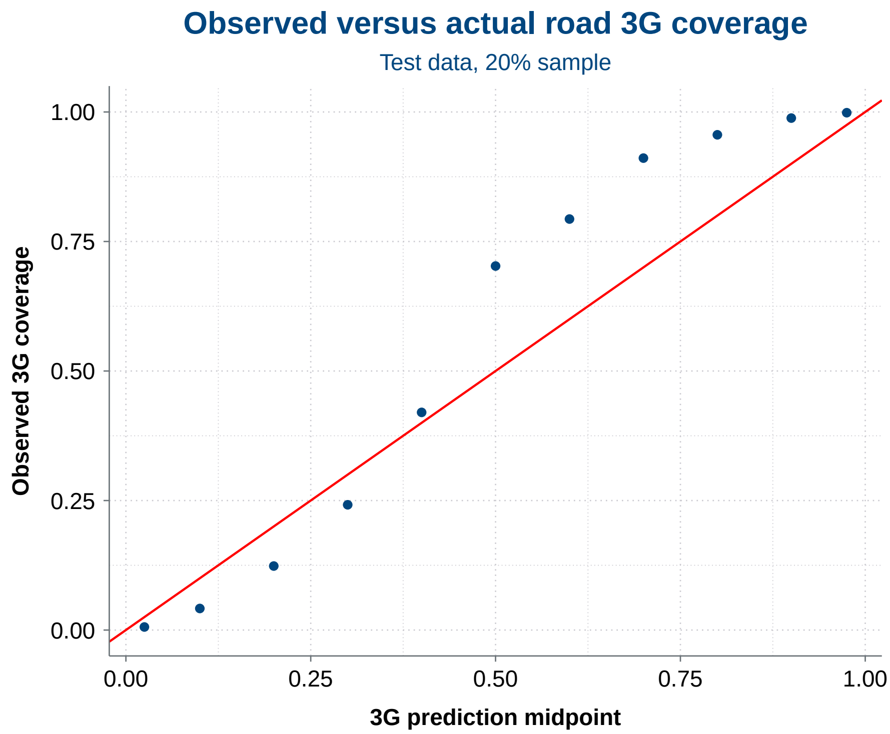


Figure 13: Calibration plot

This plot shows prediction calibration by binning the 3G predictions into eleven bins, and for each bin calculating the observed 3G coverage. According to (Dawid, 1982), forecasts are "well calibrated if, for example, the long-run proportion of forecast 75 percent credible intervals that succeed in covering the actual value of the predicted quantity turns out to be 75 percent." This corresponds to observed event percentages on the y-axis lining up with the bin midpoints on the x-axis, shown in the red line.

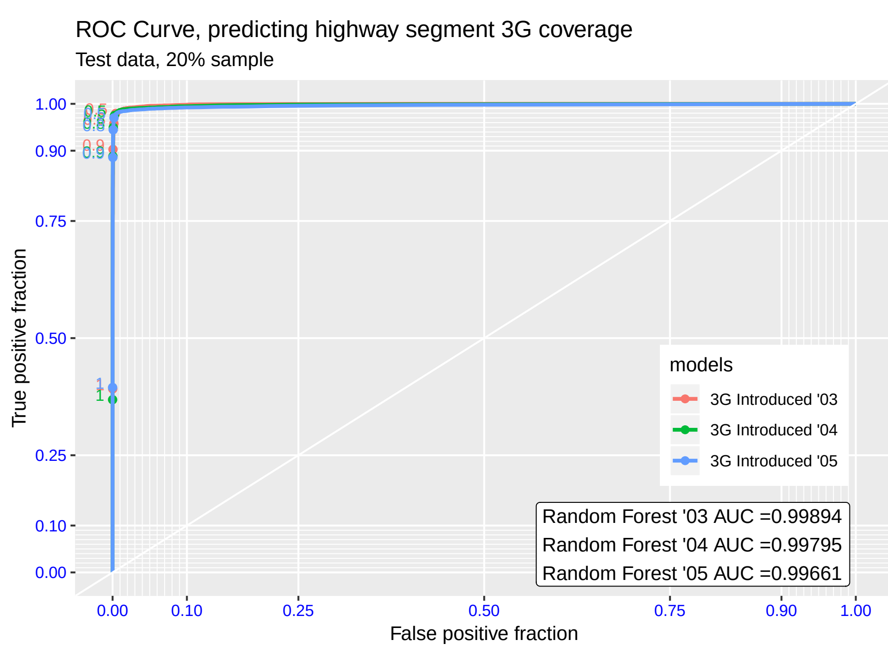


Figure 14: ROC plot varying date of 3G introduction.

This plot shows three ROC curves, one for each model where the date of 3G introduction is varied.

References

Abdel-Aty, M. (2003). Investigating the relationship between cellular phone use and traffic safety. *Institute of Transportation Engineers. ITE Journal 73*(10), 38.

Abouk, R. and S. Adams (2013). Texting bans and fatal accidents on roadways: do they work? Or do drivers just react to announcements of bans? *American Economic Journal: Applied Economics 5*(2), 179–99.

Adams, S., M. L. Blackburn, and C. D. Cotti (2012). Minimum wages and alcohol-related traffic fatalities among teens. *Review of Economics and Statistics 94*(3), 828–840.

American Automobile Association (2013). Aaa foundation for traffic safety 2012, traffic saftey culture index.

Andreassen, C. S. (2015). Online social network site addiction: A comprehensive review. *Current Addiction Reports 2*(2), 175–184.

Apple (2007). iPhone Premieres This Friday Night at Apple Retail Stores. `https://www.apple.com/newsroom/2007/06/28iPhone-Premieres-This-Friday-Night-at-Apple-Retail-Stores/`. Last accessed August 20, 2018.

Athey, S. (2017). The impact of machine learning on economics. *Economics of Artificial Intelligence*.

Babadi-Akashe, Z., B. E. Zamani, Y. Abedini, H. Akbari, and N. Hedayati (2014). The relationship between mental health and addiction to mobile phones among university students of shahrekord, iran. *Addiction & Health 6*(3-4), 93.

Babenko, B., J. Hersh, D. Newhouse, A. Ramakrishnan, and T. Swartz (2017). Poverty mapping using convolutional neural networks trained on high and medium resolution satellite images, with an application in mexico. *arXiv preprint arXiv 1711.06323*.

Bajari, P., D. Nekipelov, S. P. Ryan, and M. Yang (2015). Demand estimation with machine learning and model combination. *National Bureau of Economic Research. No. w20955.*

Bhargava, S. and V. S. Pathania (2013). Driving under the (cellular) influence. *American Economic Journal: Economic Policy 5*(3), 92–125.

Bomey, N. (2018). Backup cameras now required in new cars in the U.S. `https://www.usatoday.com/story/money/cars/2018/05/02/backup-cameras/572079002/`. Last accessed October 11, 2018.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Burger, N. E., D. T. Kaffine, and B. Yu (2014). Did Californias hand-held cell phone ban reduce accidents? *Transportation research part A: policy and practice 66*, 162–172.

Caird, J. K., K. A. Johnston, C. R. Willness, M. Asbridge, and P. Steel (2014). A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention 71*, 311–318.

Caird, J. K., C. R. Willness, P. Steel, and C. Scialfa (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention 40*(4), 1282–1293.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Carpenter, C. and C. Dobkin (2009). The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics 1*(1), 164–82.

Carpenter, C. S. and H. V. Nguyen (2015). Effects of a Driver Cellphone Ban on Overall, Handheld, and Hands-Free Cellphone Use While Driving: New Evidence from Canada. *Health economics 24*(11), 1452–1467.

Cheng, C. (2015). Do cell phone bans change driver behavior? *Economic Inquiry 53*(3), 1420–1436.

Cicala, S. (2017). Imperfect markets versus imperfect regulation in us electricity generation. *National Bureau of Economic Research. No. w23053*.

CNN (2002). Verizon launches first U.S. 3G Network, 2002. `http://www.cnn.com/2002/TECH/ptech/01/28/verizon.3g/`. Last accessed August 20, 2018.

Common Sense Media (2015). The Common Sense Census: Media Use by Tweens and Teens. `https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens`. Last accessed August 20, 2018.

Connelly, M. (2009). Many in U.S. Want Texting at the Wheel to Be Illegal. *New York Times*. `https://www.nytimes.com/2009/11/02/technology/02textingside.html`. Last accessed August 20, 2018.

Dawid, A. P. (1982). The Well-Calibrated Bayesian. *Journal of the American Statistical Association 77*(379), 605–610.

De-Sola Gutiérrez, J., F. Rodríguez de Fonseca, and G. Rubio (2016). Cell-phone addiction: a review. *Frontiers in psychiatry 7*, 175.

Deepinder, F., K. Makker, and A. Agarwal (2007). Cell phones and male infertility: dissecting the relationship. *Reproductive Biomedicine Online 15*(3), 266–270.

Deloitte (2017). 2017 Global Mobile Consumer Survey: US Edition. `https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html`. Last accessed August 20, 2018.

Dogtiev, A. (2018). App Download and Usage Statistics. `http://www.businessofapps.com/data/app-statistics/`. Last accessed August 20, 2018.

Dormehl, L. (2018). Today in Apple History: App Store hits 1 billion downloads. `https://www.cultofmac.com/543017/app-store-1-billion-downloads/`. Last accessed August 20, 2018.

Douglas, D., H. Angel, and W. Bethany (2012). Digital devices, distraction, and student performance: Does in-class cell phone use reduce learning? *Astronomy education review*.

Engstrom, R., J. Hersh, and D. Newhouse (2017). Poverty from space: Using high-resolution satellite imagery for estimating economic well-being. *World Bank Policy Research Working Paper No. 8284*.

Experts Exchange (2018). Processing Power Compared. `https://pages.experts-exchange.com/processing-power-compared`. Last accessed August 20, 2018.

Faccio, M. and J. J. McConnell (2018). Death by Pokémon GO: The Economic and Human Cost of Using Apps While Driving. Technical report, National Bureau of Economic Research.

Ghosh, A., J. Zhang, J. G. Andrews, and R. Muhamed (2010). *Fundamentals of LTE*. Pearson Education.

Glaeser, E. L., H. Kim, and M. Luca (2017). Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity. *National Bureau of Economic Research. No. w24010*.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, 681–700.

Harris, B. (2011). How Cell Towers Work.

Heimerl, K., K. Ali, J. E. Blumenstock, B. Gawalt, and E. A. Brewer (2013). Expanding rural cellular networks with virtual coverage. In *NSDI*, pp. 283–296.

Hersh, J. and M. Harding (2018). Big data in economics. *IZA World of Labor*.

International Telecommunications Union (2017). Itu statistics. `https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx`. Last accessed August 20, 2018.

Ishii, K. (2004). Internet use via mobile phone in japan. *Telecommunications policy 28*(1), 43–58.

Kolko, J. D. (2009). The effects of mobile phones and hands-free laws on traffic fatalities. *The BE Journal of Economic Analysis & Policy 9*(1).

Lee, S., H. Kang, and G. Shin (2015). Head flexion angle while using a smartphone. *Ergonomics 58*(2), 220–226.

Matyszczyk, C. (2016). US Government Agency Shames Texting Drivers on Twitter. `https://www.cnet.com/news/government-agency-shames-texting-drivers-on-twitter/`. Last accessed August 20, 2018.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

Nasar, J. L. and D. Troyer (2013). Pedestrian injuries due to mobile phone use in public places. *Accident Analysis & Prevention 57*, 91–95.

National Conference of State Legislatures (2018). Cellular Phone Use and Texting While Driving Laws. http://www.ncsl.org/research/transportation/cellular-phone-use-and-texting-while-driving-laws.aspx. Last accessed August 20, 2018.

NBC (2005). First cell phone a true 'brick'. http://www.nbcnews.com/id/7432915/ns/technology_and_science-wireless/t/first-cell-phone-true-brick/#.W3yItegzqUk. Last accessed August 20, 2018.

Nikolaev, A. G., M. J. Robbins, and S. H. Jacobson (2010). Evaluating the impact of legislation prohibiting hand-held cell phone use while driving. *Transportation research part A: policy and practice 44*(3), 182–193.

Oberhaus, D. (2016). How to Use the Internet on the Summit of Everest. https://motherboard.vice.com/en_us/article/4xa4zp/when-the-internet-came-to-everest. Last accessed August 20, 2018.

Palsson, C. (2017). Smartphones and child injuries. *Journal of Public Economics 156*, 200–213.

Peña-López, I. et al. (1999). World telecommunication development report: Mobile cellular 1999.

Pew Research Center (2018). Mobile Fact Sheet. http://www.pewinternet.org/fact-sheet/mobile/. Last accessed August 20, 2018.

Qualcomm (2014). The Evolution of Mobile Technologies: 1G ¿ 2G ¿ 3G ¿ 4G ¿ LTE. https://www.qualcomm.com/media/documents/files/the-evolution-of-mobile-technologies-1g-to-2g-to-3g-to-4g-lte.pdf. Last accessed August 20, 2018.

Redelmeier, D. A. and R. J. Tibshirani (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine 336*(7), 453–458.

Reuters (2017). AT&T, Verizon striker tower agreement in effort to diversify vendors. https://www.reuters.com/article/us-at-t-verizon-mobilephone/att-verizon-strike-tower-agreement-in-effort-to-diversify-vendors-idUSKBN1DD2G2. Last accessed October 11, 2018.

Rosenfield, M. (2016). Computer vision syndrome (aka digital eye strain). *Optometry in Practice 17*(1), 1–10.

Sampaio, B. (2014). Identifying peer states for transportation policy analysis with an application to new york's handheld cell phone ban. *Transportmetrica A: Transport Science 10*(1), 1–14.

Sauter, M. (2013). *3G, 4G and beyond: Bringing networks, devices and the web together.* John Wiley & Sons.

Simcoe, T. (2008). Xtpqml: Stata module to estimate fixed-effects poisson (quasi-ml) regression with robust standard errors.

Strayer, D. L., F. A. Drews, and W. A. Johnston (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of experimental psychology: Applied 9*(1), 23.

T-Mobile (2008). T-Mobile Unveils the T-Mobile G1 - the First Phone Powered by Android. `https://www.t-mobile.com/news/t-mobile-unveils-the-t-mobile-g1-the-first-phone-powered-by`. Last accessed August 20, 2018.

TekCarta (2018). Processing Power Compared. `https://www.nakono.com/tekcarta/databank/mobile-broadband-subscriptions-per-100-inhabitants/`. Last accessed August 20, 2018.

Thornton, B., A. Faires, M. Robbins, and E. Rollins (2014). The mere presence of a cell phone may be distracting: Implications for attention and task performance. *Social Psychology 45*(6), 479.

Törnros, J. E. and A. K. Bolling (2005). Mobile phone useeffects of handheld and handsfree phones on driving performance. *Accident Analysis & Prevention 37*(5), 902–909.

Wikle, T. A. (2002). Cellular tower proliferation in the united states. *Geographical Review 92*(1), 45–62.

Williams, A. F., K. A. Braitman, and A. T. McCartt (2011). Views of parents of teenagers about licensing policies: a national survey. *Traffic Injury Prevention 12*(1), 1–8.

Windels, J. (2017). Pokémon go figure - a data analysis of the most popular game of all time. `https://www.wandera.com/pokemon-go-data-analysis-popular-game/`. Last accessed October 11, 2018.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics 90*(1), 77–97.

Zendrive.com (2018). Largest Distracted Driving Behavior Study. `http://blog.zendrive.com/distracted-driving/`. Last accessed August 20, 2018.