# Quantifying the Premium Externality of the Uninsured

Sun, Stephen Teng and Yannelis, Constantine

Stanford University, Stanford University

20 May 2013

# Quantifying the Premium Externality of the Uninsured *

Stephen Teng Sun† Constantine Yannelis‡

January 2013

**Abstract**

In some insurance markets, the uninsured can generate a negative externality on the insured, leading insurance companies to pass on costs as higher premia. Using a novel panel data set and a staggered policy change that exogenously varied the rate of uninsured drivers at the county level in California, we quantitatively investigate the effect of uninsured motorists on automobile insurance premia. Consistent with predictions of theory, we find uninsured drivers lead to higher insurance premia. Specifically, a 1 percentage point increase in the rate of uninsured drivers raises insurance premia by approximately 1%. We also derive corrective Pigouvian taxes and determine that the optimal fine is much closer to those seen in European countries as opposed to the United States.

*We especially wish to thank Caroline Hoxby for guidance and helpful comments. We also thank Nick Bloom, Tim Bresnahan, Liran Einav, Han Hong, Xing Li, Florian Scheuer, Stephen Terry, and seminar participants at Stanford, École des Mines and the University of Illinois at Urbana-Champaign for helpful comments. We thank the California Department of Insurance and in particular Luciano Gobbo for providing us with data and regulatory information which was crucial to the development of this project. Constantine Yannelis gratefully acknowledges the financial support of the Alexander S. Onassis Foundation.

†Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072. doubles@stanford.edu.

‡Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072. yannelis@stanford.edu.

# 1 Introduction

The uninsured can generate a negative externality on the insured, leading insurance companies to pass on costs as higher premia. Following the passage and subsequent controversy over the Patient Protection and Affordable Care Act, insurance externalities have received substantial media coverage and public attention in the United States. The externality of the uninsured is present in the automobile insurance market, and the potential magnitude of this externality could be quite large given the size of this market and the large number of uninsured drivers. The National Association of Insurance Commissioners estimated that Americans spent $186 billion on automobile insurance premia in 2009, and roughly 15% of American drivers lack automobile insurance. The externality of the uninsured has been a large part of the impetus behind ten states passing "No Pay, No Play" legislation in the past decades, which restricts the ability of uninsured drivers to sue for damages following a collision as well as the large fines for driving without insurance seen in many European countries.

The aim of this paper is to estimate the size of the externality caused by uninsured drivers in the automobile insurance market and discuss the optimal policy response.[1] In this market, when a collision occurs and an uninsured individual is at fault, the insured individual will typically be compensated by his own policy.[2] When the uninsured driver has insufficient resources to cover the cost of the damage they can declare bankruptcy, passing the costs of the accident on to the insurance company and finally onto insured drivers via higher premia. Despite the theoretical and policy interests behind this externality,[3] there is relatively little empirical support in this area. We find clearly identified empirical evidence that this externality is present, and that a 1 percentage point increase in the rate of uninsured drivers raises premia by roughly 1%.

Estimating the size of the effect of the uninsured on premia unavoidably poses a substantial empirical challenge. The most significant concern is the endogeneity of the rate of the uninsured with respect to insurance premia, which will bias regression coefficients. If insurance premia are high for reasons other than there being a high fraction of uninsured individuals, fewer people will buy insurance, generating reverse causality that could lead the researcher to misstate the causal

[1] There are numerous other externalizes arriving from driving, for example the vehicle size externality studied by Anderson (2008) and Anderson and Auffhammer (2014).

[2] Most insurance policies sold in the US come with uninsured motorist coverage, which is the primary, but not the sole, channel through which the uninsured affects the premia of the insured. Department of Insurance data indicates that in 2008 in California 84.38% of policies came with uninsured motorist coverage. A majority of states also mandate uninsured motorist coverage.

[3] For example, see Smith and Wright (1992) and Keeton and Kwerel (1984). The policy relevance of this effect is clearly exemplified by the United Kingdom Motor Insurers' Bureau, which compensates damage done by uninsured motorists explicitly by adding a surcharge to insurance premia. In the United States various state and federal laws mandate insurance coverage under penalties of a fine or tax, which are presumably designed to internalize insurance externalities.

effect of the uninsured on premia. This makes it difficult for the researcher to identify the true effect of the uninsured on premia. Although the literatures on insurance and health are large, empirical research on the effect of the uninsured on premia has been lacking due to the aforementioned problem. Our paper attempts to fill this gap for the case of the automobile insurance market. Using a novel panel data set and a plausibly exogenous policy change varied at the county level in California, we quantify the extent of this negative externality. Our findings have substantial implications for policymaking in this area.

We exploit variation in the rate of uninsured drivers resulting from an exogenous policy change to identify the effect of uninsured drivers on insurance premia. Between 1999 and 2007 the California Low Cost Automobile Insurance (CLCA) Program was introduced in the state of California and rolled out sequentially on a county-by-county basis. The introduction of the CLCA program, together with the accompanied media campaign in areas in which the program was in effect, resulted in an approximate 1 percentage point decrease in the rate of uninsured drivers. The sequential rollout of the program makes it possible to obtain credible identification of the causal effect of uninsured drivers on premia.

In order to accomplish this, we compiled novel panel data for 58 counties in California between 2003 to 2007.[4] Our main data set consists of insurance premium quotes collected by the California Department of Insurance from licensed insurers based on several hypothetical risks including demographic and driving characteristics, policy limits, location, and coverage availability. Each observation in our sample represents an offer price for one of two typical insurance plans, for hypothetical consumers with specific observable demographics from an insurance company operating in a particular zip code. The main variation of interest to us is the geographic variation – at the zip code level – in insurance premia. Automobile insurance companies collect zip codes from clients and vary prices accordingly.[5] Controlling for year and zip code fixed effects can absorb many environmental factors since auto insurance companies typically price at the zip code level. Furthermore, we exploit the within-geographical variation induced by the CLCA program to obtain estimates for the average effect of uninsured drivers on insurance premia.

The use of policy-driven variation in the rate of uninsured motorists along with new administrative data on insurance premia leads us to conclude that uninsured drivers raise premia for other drivers, as predicted by theory. We find that a 1 percentage point increase in the share of drivers who are uninsured leads to a 1 percent increase in premia. To illustrate, this implies that consumers could save approximately $500 annually if the county with the highest uninsured drivers rate, 29%

[4]The data used was not collected statewide in 2004 and 2008, and there are significant delays in the construction of data on uninsured motorists. At the time of writing, California data on uninsured motorists at the zip code level beyond 2008 did not exist.

[5]California has been attempting to ban auto insurance pricing based on zip codes since 2005. However, the change did not officially come into effect until late 2008, and there is substantial evidence that the majority of insurance companies did not comply with the ban in 2008.

in San Joaquin, sees its uninsured drivers rate fall to that of the county with lowest uninsured drivers rate, 9% in Mono. We rule out a number of potential alternative explanations such as an increased competition or unobserved selection channel. [6]

We discuss the optimal corrective Pigouvian tax on uninsured drivers using a simple model informed by our empirical estimates. Given that uninsured individuals increase premia paid by insured individuals, the government can levy a fine or tax on the uninsured to try to capture the effect of the externality. We find that the optimal tax is $2,240, which forces uninsured drivers to fully pay for the externality. Given that enforcement is stochastic, this is substantially higher than current fines in the US, although in line with fines in some European countries such as France. Such a high fine, if enforced rigorously, would effectively eliminate uninsured drivers as purchasing insurance on the private market would be cheaper than paying the fine.

We also conduct a battery of robustness checks and examine alternative explanations. Our results are robust under weak instruments consideration. We find that our results are robust to dropping any wave of the CLCA program and controlling for a county-specific time trend. We also vary the definitions of our instrumental variables and obtain consistent results. Utilizing the eligibility requirement of the CLCA program, we discuss and reject alternative explanations such as unobserved selection on accident risk and moral hazard.

The paper is organized as follows. Section 2 presents a concise motivating model based on prior literature. Section 3 describes the data, which to our knowledge has not been used in the economics literature. Section 4 discusses our estimation strategy, explaining how we use a policy change to overcome the endogeneity problem. Section 5 presents our main empirical results, in which we find evidence of a significant externality arising from uninsured drivers. The section then discusses Pigouvian taxation. Section 6 presents robustness checks and rules out alternative explanations for our results. Section 7 concludes and offers suggestions for future research.

# 2 Theory

In this section we discuss the theory behind the externality caused by uninsured drivers on auto insurance premia, and we illustrate the endogenous relationship between premia and uninsured

[6] One concern is that the CLCA program, being an insurance plan itself, could affect the insurance premia in the commercial market through an increased competition channel. As well as lowering the rate of uninsured drivers, introducing the CLCA program also offered another low-cost plan to consumers which may have forced insurance providers to react by lowering premia. Thus it is possible that our results are partially or entirely driven by increased competition rather than the effect of the CLCA program on uninsured drivers. We exploit an eligibility requirement of the program, the maximum allowed value of the insured vehicle, to obtain results that are free from this potential confound. Our estimates restricted to samples with car values well-above the eligibility requirement do not significantly differ from our main results, suggesting they are indeed mainly driven by the decrease in uninsured drivers. We can similarly rule out other potential competing explanations such as unobserved selection.

drivers which creates difficulties in estimating the effect of uninsured drivers on premia. In Section 5 we use the model as a framework to discuss the optimal policy response to uninsured drivers. The basic intuition behind the theory is straightforward. Typically when a driver is found at fault in an accident, the at-fault driver's insurance covers the cost of damages. However, when an uninsured or underinsured driver causes an accident the driver may not have sufficient resources to cover damages.[7] In this case the damaged party will be forced either to pay expenses out of pocket or collect payment from his own insurance plan. Thus in an area with a higher proportion of uninsured drivers, insurance companies will charge higher premia to obtain a given rate of return. The ability of an uninsured driver to declare bankruptcy is a crucial part of the burden shifting from the uninsured to the insured.

The model draws heavily from Smith and Wright (1992) and Keeton and Kwerel (1984). Define an individual $i$ with wealth $w_i$ and probability of being involved in an accident $\pi_i$. The individual purchases liability insurance from firm $j$ with uninsured motorist coverage that costs $p_{ij}$. The liability insurance, which is the minimum insurance coverage required by law in most US states, pays for damage incurred by the holder of the policy to other individuals.[8] The individual $i$ who purchases insurance has a payoff of $w_i - p_{ij}$ if he is not involved in an accident or if he is involved in an accident with another driver and found not to be at fault. For simplicity and without loss of generality,[9] we assume that an individual has an equal probability of being found at fault or not at fault in an accident and an accident always involves two cars. If an individual is involved in an accident and is found at fault, the individual must either pay for the damage incurred to his vehicle or declare bankruptcy, hence the individual's payoff is $max\{w_i - p_{ij} - L_i^s, 0\}$ where $L_i^s$ is the stochastic cost of damage incurred by either party equally from the accident. In this case, the insurance company covers the losses $L_i^s$ of the other driver who is not at fault[10]. This event occurs with probability $\frac{\pi_i}{2}$. Thus an insured driver has expected utility, assuming a utility function $U(.)$ with standard properties:

$$V_{ins}(p_{ij}, w_i) = U(w_i - p_{ij})(1 - \pi_i + \frac{\pi_i}{2}) + \mathbb{E}[U(max\{w_i - p_{ij} - L_i^s, 0)\}]\frac{\pi_i}{2}$$

[7]There are also other concerns, for example an uninsured driver may be more likely to flee the scene of an accident.

[8]In the above discussion we restrict the externality to passing through the uninsured motorist coverage channel, this is not the only mechanism through which the externality is present. Auto insurance policies list uninsured motorist premia separately, but this does not capture the full extent of the uninsured driver externality due to several factors. First, while an entire package may be actuarially fair, the individual components may not. Second, uninsured drivers may still incur costs on insured drivers through property and collision coverage as the uninsured motorist coverage in California covers only bodily injury. See Smith and Wright (1992) for a further discussion.

[9]With the notable exception of moral hazard. We discuss the literature on moral hazard in Section 6, which has mixed results. Allowing for any other arbitrary probability of being at fault will not change the basic intuition and prediction of our model.

[10]We note that since he holds a liability only policy which pays for the damage done to the other individual's car, the insured driver must still pay for the damage to his own vehicle, $L_i^s$.

Let $\lambda$ be the fraction of uninsured motorists in a market, and note that $\lambda$ is a function of premia, since when premia are high few drivers will purchase insurance. For an uninsured driver, if no accident occurs, or if an accident occurs with an insured driver and the uninsured driver is not found at fault, the driver obtains payoff $w_i$. The probability of not being involved in an accident is $1 - \pi_i$ and the probability of being involved in an accident with an insured driver and not being found at fault is $\frac{\pi_i}{2}(1 - \lambda)$. The expected utility for an uninsured driver if involved in an accident and found at fault is similar to that of a driver with liability insurance, with the exception of never having paid a premium to an insurance company, and that the driver must pay for the other driver's losses, rather than the insurance company paying: $max\{w_i - 2L_i^s, 0\}$. Finally, if an uninsured driver is involved in an accident with another uninsured driver who is at fault, the driver receives a payoff $min\{w_i - L_i^s + R_i, w_i\}$, which occurs with probability $\lambda\frac{\pi_i}{2}$. We let $R_i$ refer to the amount the driver recovers from the uninsured individual who caused the accident, which is random. Assuming a continuous, increasing and concave utility function $U(.)$, the total expected utility $V_{unins}(w_i)$ for the uninsured driver becomes:

$$V_{unins}(w_i) = \mathbb{E}[U(w_i)(1-\pi_i+\frac{\pi_i}{2}(1-\lambda))]+\mathbb{E}[U(max\{w_i-2L_i^s,0\})]\frac{\pi_i}{2}+\mathbb{E}[U(min\{w_i-L_i^s+R_i,w_i\})]\lambda\frac{\pi_i}{2}$$

A driver will choose to insure if $V_{ins}(p_{ij}, w_i) \geq V_{unins}(w_i)$. As we would expect, a driver is less likely to choose to insure when his premium is higher. Thus $\lambda$, the rate of uninsured drivers, is increasing in the premium $p_{ij}$. This property leads to simultaneity bias which, as we will see, presents significant empirical challenges to estimating the effect of uninsured drivers on insurance premia.

We assume a representative risk-neutral firm in a competitive insurance market and we compute the actuarially fair premium by equating revenues with expected indemnities, which amount to the expected liability loss from an insured driver as well as the expected loss from being involved in an accident with an uninsured driver who declares bankruptcy. We thus have

$$p_{ij} = \mathbb{E}[(max\{L_i^s - R_i, 0\}\lambda + L_i^s)\frac{\pi_i}{2}].$$

Assuming that accident rates of the policy holder are a function of observable demographics $X_i$ we have $\frac{\pi_i}{2}\mathbb{E}[L_i^s] = X_i'\gamma$. We can then define $\beta_i = \mathbb{E}[max\{\frac{L_i^s - R_i}{\mathbb{E}[L_i^s]}, 0\}X_i'\gamma] \geq 0$ and we have the following equation for the premium that individual $i$ pays to firm $j$

$$p_{ij} = \beta_i\lambda + X_i'\gamma.$$

The premia charged by the insurance company are weakly increasing in $\lambda$, the rate of uninsured drivers. Hence, *ceteris paribus* we would expect an area with a higher rate of uninsured drivers to have higher insurance premia. At the same time $\lambda$ is increasing in $p_{ij}$ as higher premia will cause fewer drivers to insure. Thus an area with high premia for reasons totally unrelated to the rate of uninsured drivers could also have a high rate of uninsured drivers. This endogeneity problem makes it difficult to estimate the true effect of $\lambda$ on $p_{ij}$, since $\lambda$ will be significantly correlated with the error term in any regression.[11] Separating the effect of uninsured drivers on insurance premia from drivers choosing not to insure due to otherwise high premia presents a challenge. In the next section, we discuss how we overcome the endogeneity problem and estimate the true effect of uninsured drivers on insurance premia.

# 3 Data

## 3.1 Main Dataset

Our main dataset, which to our knowledge has not been used in the economics literature, comes from the California Department of Insurance. Following January 1, 1990, California law[12] required that the California Department of Insurance collect data on insurance rates in the state. The Department of Insurance ran the Automobile Premium Survey (APS) which collected data on automobile insurance premia from insurers licensed to provide automobile insurance in California based on several hypothetical risks including demographic and driving characteristics, policy limits, location and coverage availability. Each observation represents an offer price for consumers with particular observable demographics from a firm operating in a particular zip code. Auto insurance pricing is heavily regulated in California, and insurers must charge customers prices based on formulae registered with the state Department of Insurance. The survey oversampled hypothetical drivers with speeding tickets and at fault accidents, leading to a higher average premium in comparison to the general populace. The premium survey data is available from 2003 to 2010 and our final sample is from 2003 to 2007, matching the available uninsured drivers rate data.[13] We view non-compliance or false information as unlikely to be a major concern in the survey data since both false information and non-response are punishable by large fines according to the California Insurance Code.[14] There is a surprising degree of price dispersion in the data, with different firms charging higher or lower premia for drivers in the same zip code with identical characteristics. This is consistent with prior studies of automobile insurance, such as Dahlby and West (1986).[15]

The database consists of several million observations, the main variable of interest being the annual premium for an automobile insurance plan. The observations are indexed by zip codes, al-

[11] We tend to think that this correlation should be positive, as higher premia will cause fewer drivers to insure, biasing our results upwards. However, other biases such as measurement error will bias the coefficient towards zero.

[12] Specifically, the California Insurance Code Section 12959.

[13] In 2008 the APS survey was not conducted for administrative reasons, and in 2004 the survey was not conducted statewide.

[14] We drop premium quotes above \$20,000, however our results are robust to varying this threshold and not dropping observations. See Section 6 for more information on robustness.

[15] Dahlby and West (1986) offer costly consumer search as a possible explanation for this phenomenon.

lowing the researcher to match the database to county-level data. The database also contained data on National Association of Insurance Commissioner (NAIC) codes of insurers, which allows the researcher to identify the number of firms offering plans in a county and to match insurance company characteristics to each surveyed premium. The APS database also contains data on vehicle make and year, which we matched to vehicle value using pricing information.[16] The APS collected data on two types of plans from licensed insurers in zip codes, a basic plan and a standard coverage plan for different demographics. The basic plan represents a plan just above the minimum required threshold for coverage in California, while the standard plan was deemed by the Department of Insurance to be the most common automobile insurance plan in California. Table 2 summarizes the two types of private plans and the basic CLCA plan.

One potential concern is that our results could be driven by compositional changes in the survey data. It is worth noting that our premium data comes from an administrative survey, which uses a host of hypothetical risk profiles of drivers. A priori, there is no reason to believe that the government surveyed insurance premia for different groups of drivers after the CLCA program took effect. In Table 3, we demonstrate this is indeed the case. Since the insurance companies set prices based on several individual-specific characteristics, we directly examine the characteristics of the drivers surveyed before and after CLCA program to make sure that we compare prices for the same group of people. We compare the mean of major risk factors used in the main analysis for insurance pricing in the period before and after the CLCA program has been active for at least four months. These factors include sex, age, plan type, accident rate, daily miles driven, whether the driver has incurred an at-fault accident as well as whether the driver has a recent history of speeding tickets. Our F-test can not reject at 5% level the hypothesis that these characteristics ever changed after the CLCA program took effect. We reject that the 10% level that the accident rate is the same, which is consistent with moral hazard, insured drivers being less cautious and being involved in more accidents. We discuss this issue, which will not bias our results as we control for accident rates, further in Section 6. Another potential concern regards the CLCA program attracting some particular group of drivers whose behaviors could affect the insurance premium independent of the uninsured drivers' externality effect. This concern is also dealt with in Section 6, as we restrict the sample only to individual who would have been ineligible for the CLCA program.

The raw APS survey data was matched with demographic, driving, policy and vehicle characteristics using the annual APS Hypothetical Risk Codebooks which were provided to us by the Department of Insurance. This allowed us to match each observation to create variables for age, gender, the number of years an individual possessed a license, the number of miles an individual drives to work daily, the number of miles an individual drives in a year, the number of persons covered under a plan, the types of vehicles covered under the plan, the number of speeding tickets a hypothetical individual received in the three years prior to the survey date, and the number of

[16]The website Auto Loan Daily was used as the source for vehicle values.

at-fault automobile accidents in which an individual was involved in the three years prior to the survey. Since premia were unadjusted for inflation, we collected data on the Consumer Price Index from the Bureau of Labor Statistics using the BLS December CPI of each year in our adjustments.

## 3.2 Matched Data

The main APS survey data was matched to three other data sources which were obtained from the California Department of Insurance, the California Highway Patrol Integrated Traffic Records System, and the US Census Small Area Estimates Branch. Whether or not the CLCA was in effect in various counties as well as premium rates in effect was obtained from the California Department of Insurance 2011 Report to the Legislature.

We used zip codes to match data from our sample premium database to zip code level data from California using various sources. Zip code level data on uninsured bodily injury claims and bodily injury claims was also obtained from the California Department of Insurance between 2002 and 2007. We used this data to construct a measure of uninsured drivers following Smith and Wright (1992) and Cohen and Dehejia (2004)[17]. For each zip code, we use the average rate of uninsured motorists in zip codes within a 25 mile (40km) radius of the zip code area[18]. Using claims data to estimate the rate of uninsured motorists introduces another form of bias to our results– measurement error. Since we do not observe the rate of uninsured motorists directly, but rather claims from accidents which are stochastic, OLS estimates will be biased towards zero.

To construct our measure of accident rates, county level data on injuries and fatalities resulting from automobile collisions was obtained from the California Highway Patrol. Since 2002, the California Statewide Integrated Traffic Records System has provided a database of information on monthly traffic collisions in California counties. The system provides data on all reported fatal and injury collisions occurring on public roads in California. The data is compiled from local police and sheriff jurisdictions and California Highway Patrol field offices. We can use this data, and data on the total number of exposures and percentage of uninsured motorists from the Department of Insurance, to compute the injury collision and fatality collision rates in various California counties by taking the number of injury accidents over the number of registered vehicles.

[17]See Appendix A for more on estimating the rate of uninsured drivers. Our measure is the number of Uninsured Motorist Bodily Injury claims divided by the number of Bodily Injury claims in a given zip code. This measure will be identical to the rate of uninsured motorists given two very plausible assumptions, one, we must assume that the probability of being involved in an accident is the same for both insured and uninsured motorists and two, in accidents between insured and uninsured motorists each party is equally likely to be found at fault.

[18]According to the Bureau of Transportation Statistics (2006), this is roughly the number of kilometers that the average Californian drives per day. The main results are robust to varying the uninsured motorist zip code region. We use a standard equirectangular approximation to compute distance.

# 4 Empirical Strategy

## 4.1 The CLCA Program

Despite great policy interest in the topic, credible estimates of the effect of uninsured drivers on premia are lacking. Any simple estimates that do not directly address the issue of reverse causality would be plagued by the obvious endogeneity problem noted above. Given that when premia are higher, drivers are less likely to buy automobile insurance, the rate of uninsured drivers is endogenous in a usual hedonic regression. Estimating the causal effect on insurance premia requires the use of instrumental variables for the rate of uninsured drivers. In order to obtain a valid instrument, we must find a variable that is (1) correlated with the rate of uninsured drivers and, (2) uncorrelated with any other unobservable determinants of the dependent variables. In practice, finding such an instrument has proven to be quite difficult since most factors that would affect the rate of uninsured drivers would also have direct effects on premia through channels other than the rate of uninsured drivers. We find a set of credibly valid instruments using a policy change that generated variation in the share of drivers who were uninsured in each California zip code. Starting in 1999, California introduced a program that subsidized automobile insurance premia for uninsured drivers who fit certain eligibility criteria. This program was not introduced in every county at the same time, but was rolled-out to different counties at different times. We demonstrate that the sequence of the roll-out is not correlated with other factors that might affect insurance premia. We can therefore exploit both variation over time within a county and variation among counties at a point in time.

California mandates, as do all US states with the exception of New Hampshire, that drivers purchase basic liability automobile insurance. In California the basic liability insurance required by law consists of \$15,000 of bodily injury insurance per individual, \$30,000 of total bodily injury insurance per accident, and \$5,000 of property damage insurance per accident. Despite the mandates, many drivers remain uninsured. For instance, in 1998, the Department of Insurance estimated 16.38 percent of California drivers were uninsured. To reduce the share of drivers who are uninsured, California introduced the California Low Cost Automobile Insurance program (CLCA) in 1999, starting with two pilot counties. CLCA offers basic liability insurance to eligible low-income individuals who live in California counties where the program is active. Rates under the CLCA program are set annually at the county level by the California Automobile Assigned Risk Plan (CAARP) commissioner. They are set well below rates for plans available in the market.[19] The rates set by CAARP are intended to cover the administrative costs of the program but not to allow insurance companies to make a profit. Premia are not directly subsidized by the government, and policyholders are assigned to insurance firms based on their share of the voluntary auto

[19]CLCA coverage is lower than the minimum required insurance coverage for holders of normal private automobile insurance plans.

insurance market in each county. When setting rates, the CAARP commissioner is allowed only to consider insurance firms' loss in the previous year in each county. The commissioner is also constrained to set rates 25 percent higher for eligible, unmarried male drivers between the ages of 19 and 24.

The CLCA program was instituted in two pilot counties in 1999, and then expanded across the state in five different waves between April 2006 and December 2007. The introduction of the CLCA program was coupled with intense media campaigns in counties that were thought to be underserved or having a high proportion of uninsured drivers by the Department of Insurance. Advertisements were put out via print, radio, cable television, community organizations and government agencies. This media campaign about the legal requirement for carrying insurance would likely have had a second effect in decreasing the rate of uninsured drivers, as well as the primary effect of decreasing uninsured drivers via insurance plans under the CLCA program.[20] Figure 1 illustrates the expansion of the CLCA program via waves between 1999 and 2007.

After the initial pilot program in San Francisco and Los Angeles counties was deemed successful, the California State Senate voted to expand the program in 2005 to the six counties with the highest volume of inquiries received by the CAARP. In 2006 and beyond, the commissioner was allowed to introduce the CLCA program based on determination of need, which was interpreted as the number of uninsured drivers in a county between 1998 and 2007.[21] The number of uninsured drivers depends largely on the size of counties rather than the rate of uninsured drivers. County borders are somewhat arbitrary, and the population size of California counties varies drastically while the rate of uninsured drivers, which is the driving force behind the externality, does not vary as much, ranging from 9% to 29%. Effectively, this means that CLCA program waves were assigned by the population of counties. Figure 2 illustrates the means of certain key variables of counties across county waves. There is a clear declining trend in population across the five waves, while other variables such as accident rates, rates of uninsured drivers, and premia are close to being identical. The exception to this rule is in the final wave, where the results are affected by several small counties in the Sierra Nevada mountains which have a very high measured accident rate:[22] Alpine, Placer, Nevada, El Dorado and Sierra. The results are robust to excluding these counties, and our results are robust to omitting both the final wave and the pilot counties.

Eligibility for the program was determined by two main factors, income and a vehicle value threshold.[23] We do not observe income, as it is illegal in California for automobile insurers to price

[20]See Schultz and Yarber (2011).

[21]For more details on the implementation of the CLCA program consult Schultz and Yarber (2006).

[22]The sharp spike in accident rates likely represents the way in which we measure the accident rate. Our measure of accidents is the number of injury accidents over the total number of vehicles in a county, and this measure reports implausible accident rates several times higher than those of other counties. The Lake Tahoe region is a popular tourist destination, and it is very likely that the high measured accident rates simply reflect tourists getting into accident in counties with very low numbers of registered vehicles.

[23]See Appendix A.3 for a further discussion on eligibility and the CLCA program in general.

on income, however we do observe vehicle value. This eligibility criteria is extremely valuable, as it allows us to test and reject competing explanations for our observed effects. If premium prices drop following the introduction of the CLCA program, this could be due to insurers competing with the new CLCA plan, or due to riskier individuals selecting into the CLCA plan. However, we can restrict the sample to vehicles above the vehicle value threshold, which are ineligible for the CLCA program and hence would not be affected by competition or selection.

## 4.2 Empirical Specifications

As mentioned earlier, the rate of uninsured drivers is endogenous to premia; if premia are higher fewer drivers are likely to insure. This makes any OLS estimation results for the effect of uninsured drivers on premia inconsistent for the true effect, and essentially meaningless to the researcher. As well as the endogeneity bias caused by reverse causality, we face another bias in the form of measurement error. The rate of uninsured drivers is estimated by the ratio of uninsured bodily injury claims over the insured bodily injury claims. Endogeneity should bias these estimates upwards, while measurement error will bias the coefficients towards zero. These two effects moving in opposite directions make any OLS results uninformative in regards to the true causal effect of the rate of uninsured drivers on insurance premia. In order to overcome these difficulties we employ an instrumental variables strategy exploiting the introduction of the CLCA program to various California counties, which was plausibly exogenous.

The first assumption is that the instrumental variables are correlated with the rate of uninsured drivers, which is supported in results presented in Section 5.1. The second assumption is that the instrumental variables are orthogonal to unobserved determinants of insurance premia. Thus the identifying assumption for our empirical strategy is that, had it not been for the introduction of the CLCA program, there would have been no differential conditional changes in the insurance premia across California counties in different waves over our sample period. It is important to note given that we control for year and zip-code fixed effects, any confounding factor must be a systematic time-varying zip-code-specific change that coincides with our observed trend in insurance premia.

While our identifying assumption cannot be tested directly, Figure 3 provides further support that there was no significant pre-existing trend in the insurance premia across the different CLCA program waves. Figure 3 shows wave-by-year fixed effects from regressing premia on controls for individual, geographic, temporal and vehicle controls. None of the fixed effects are significant at the 5 percent level, and there do not appear to be significant differences in the waves conditional on observables. The figure also provides graphical evidence for our hypothesis that the CLCA program reduced the rate of uninsured drivers, thereby reducing automobile insurance premia. In 2006, when the CLCA program begins, we see a sharp drop in premia for the first two waves, where the CLCA program took effect. We also plot the average annual rate of uninsured drivers by the number of years before and after the introduction of the CLCA program in Figure 4. This visual illustration makes the case that there is no clear declining pretrend in the rate of uninsured drivers before the implementation of the CLCA program. We conclude that examining the dynamic variation of both the insurance premia and the rate of uninsured drivers combined with the specific timing of CLCA waves provide strong support for our identifying assumption.

Our instruments from the CLCA program include the following: the average number of months during the year in which the CLCA program was active in a zip code cluster; the square of the previous term; and a dummy of whether the CLCA program was active for more than four months in a zip code cluster.[24] We use the number of months during the year in which the CLCA program was active since typically the CLCA program was introduced in the middle of a year, and we avoid any arbitrary cutoffs associated with an indicator variable of whether or not the CLCA program was in effect. The results are robust if instead we use an indicator of whether or not the CLCA program was in effect for the entire year, or an indicator of whether or not the CLCA program was in effect for any part of the year.[25]

The CLCA program being in effect is associated with a drop in the rate of uninsured drivers due to both the direct effect of uninsured drivers entering the program and through the media campaign associated with the introduction of the program. It is also highly plausible that the introduction of the CLCA program was exogenous to insurance premia in a county.[26] Furthermore, the rate of uninsured drivers varies much more within zip code clusters in counties as opposed to across counties. Since insurance companies price at the zip code level, including zip code fixed effects absorbs geographic factors in pricing. The inclusion of zip code fixed effects greatly strengthens our identification strategy – even if certain counties have higher average premia, our analysis at the zip code level will estimate the average effect of an increase in the rate of uninsured drivers. While the exogeneity of the introduction of the CLCA program is highly plausible, it is impossible to fully test the exclusion restriction, which is necessary for the validity of an instrument.

We also include as an instrument the number of months the CLCA program is in effect squared. If the average number of months that the CLCA program is in effect is a valid instrument, the square of the instrument will always mechanically be a valid instrument. However, there is also an intuitive reason to include the square of the CLCA program as an instrument– we expect the effect of the CLCA program to be greater in geographic areas where the program has been in effect for

[24]We have varied the threshold for the number of months the CLCA program has been in effect, and the results are robust to changing the dummy to an indicator or whether or not the program has been in effect for more than one month, or more than six months.

[25]Appendix table A4 also shows results using only one instrument, the average number of months that the CLCA program was in effect. The results are less precise than our preferred estimates.

[26]California government documents regarding the introduction and expansion of the CLCA program do not make any mention of premia being used as a determinant of where the CLCA program was introduced, and from Figure 2 it appears that the California government simply rolled out the program in counties with a larger population first. We also find that population is not a significant determinant of premia when we control for population, and our results are robust to including population in the specification.

more time. Thus including a square term would put more weight on zip code clusters where the CLCA program has been active for more than several months.

Given our set of instruments we can exploit variation orthogonal to premia, conditional on zip code and year fixed effects, to address both the problem of reverse causality between premia and the rate of uninsured drivers and the issue of measurement error using a standard approach. To implement the IV estimator, we first run the following regression (first stage):

$$\lambda_{gt} = \alpha_g + \alpha_j + \alpha_t + \alpha_v + X_{it}'b_1 + CLCA_{gt}'b_2 + e_{gijt}, \tag{1}$$

where $\lambda_{gt}$ is the rate of uninsured drivers in geographic area $g$ in which firm $j$ offers an insurance premium to individual $i$ at time $t$, $CLCA_{gt}'$ is a vector consisting of our CLCA instruments, $X_{it}$ is a vector of control variables and $\alpha_g, \alpha_j, \alpha_t$ and $\alpha_v$ are zip code, firm, year and vehicle fixed effects. Since automobile insurance companies typically price at the zip code level, including zip code fixed effects absorbs all environmental factors that do not vary over time within a zip code, for example certain zip codes may have worse road conditions or higher speed limits leading to frequent accidents and higher premia. We then estimate the second stage:

$$premium_{gijt} = \alpha_g + \alpha_j + \alpha_t + \alpha_v + X_{it}'\gamma + \beta\hat{\lambda}_{gt} + \varepsilon_{gijt}, \tag{2}$$

where $premium_{gijt}$ is the real (inflation-adjusted) premium offered in geographic area $g$ by firm $j$ to individual $i$ at time $t$ and $\hat{\lambda}_{gt}$ are predicted values of the rate of uninsured drivers from our first stage, (1). We use year fixed effects to control for any time-specific macro effects that shift the premium of automobile insurance in California. In our context, such macro effects could involve technological progress in automobiles that reduced loss in accidents or changes in the degree of competitiveness in automobile insurance markets that affect areas across California. We use zip code fixed effects to capture any unobserved zip code characteristics that are fixed over time, such as population characteristics, general weather conditions, traffic conditions, and any other bias associated with geographic characteristics. These zip code fixed effects are important for mitigating potential bias associated with the likely endogeneity of the rate of uninsured drivers. For example, the bias can arise from the fact that wealthier zip code areas have fewer uninsured drivers and tend to have higher insurance premia for reasons like price discrimination, which is difficult for the researcher to control directly. We also use company fixed effects to control for any time-invariant company-specific effects. For example, some firms may be more competitive and focus on thrift consumers while some firms charge higher premia for superior quality of service and brand capital. The vehicle fixed effects control for vehicle specific pricing factors, for example, more expensive vehicles may be more expensive to insure. We define the vehicle fixed effects by brand and model, and all results are robust to specifying the vehicle fixed effects by brand, model and year. Our coefficient of interest is $\beta$, which we interpret as the average effect of a 1 percentage point increase in the rate of uninsured drivers on the average premium. It is important to mention the caveat that our estimates are local. It is quite likely that there are nonconstant average effects in how uninsured drivers affect insurance premia. The average rate of uninsured drivers in California during our time period is 20.6%, with a standard deviation of 4%.

It is worth noting that while the different waves of the CLCA program were introduced by county, for our purpose, the minimal source of variation in the rate of uninsured drivers affected by this program is actually at what we call zip code cluster level, since a driver living in a border zip code could drive to other nearby zip codes of a different county. We define a zip code cluster as zip codes within a 25 mile (40 km) radius of the zip code under consideration. We chose 25 miles as this is the average number of miles the average Californian drove daily in 2007. Our rates of uninsured drivers as well as the instruments are all averages of the raw variables within the zip code cluster. This particular feature also justifies why we are clustering at zip code cluster level instead of county level.[27]

# 5 Main Results

## 5.1 First Stage

Before showing our main results, we first examine how the introduction of the CLCA program affected the rate of uninsured drivers in California. In the first three columns of Table 4, we regress the rate of uninsured drivers on each of our three instrumental variables and find the rate of uninsured motorists decreased. In terms of economic magnitude, the introduction of the CLCA program led to a roughly one percentage point drop in the rate of uninsured drivers. We also run a first-stage regression by including all our instrumental variables and find a F-statistic for the hypothesis that all instruments jointly have no effects to be 14.83, which is above the standard threshold for weak instruments. Results in column 4 allow us to further investigate how the reduction in the number of uninsured drivers corresponds with the number of drivers actually enrolled in the CLCA program. More specifically, we have administrative data on the number of drivers enrolled in the CLCA program at the county level in 2006 and 2007. We impute the number of uninsured drivers from the estimate for the rate of uninsured drivers and the number of registered drivers. Results from a simple OLS regression shows that the enrolled drivers can account for 40% of the total reduction in the total number of uninsured in a county. That this number is less than 100% is not surprising, given that there is a large accompanied media campaign on driving with insurance taking place concurrently with the introduction of the CLCA program, which could have made more drivers buy insurance. The CLCA Report to the Legislature noted that many uninsured motorists called insurance companies to ask about the CLCA program, and ended up purchasing

[27] Our results are, in fact, robust to clustering at county level and varying the radius of the zip code cluster.

more comprehensive policies. Overall, our results demonstrate that the CLCA program indeed reached the desired goal of reducing the number of uninsured drivers, which provides the variation essential to our empirical strategy.

## 5.2 Estimates of the Externality

Table 5 presents a set of linear regressions of the insurance premium on the rate of uninsured drivers and other controls, where we add more controls gradually. In the first two columns we are treating the rate of the uninsured as exogenous and do not control for zip code fixed effects in the OLS regression. In both specifications, the coefficient on the rate of uninsured drivers is negative and significant at 0.10 level, indicating the rather nonsensical result that more uninsured drivers reduce insurance premia. This is not surprising given that in these specifications our main source of variation is the cross sectional difference in the rate of uninsured drivers and we do not control for any fixed effects. Geographic factors such as wealth differences, leading to price discrimination, or low vehicle values leading to lower accident costs may result in a negative correlation between premia and the rate of uninsured drivers. These factors necessitate controlling for zip code and other fixed effects. Indeed, when we control for zip code and year fixed effects in Table 5, columns (3)-(4), the coefficient on the rate of the uninsured changes its sign and becomes positive and statistically significant. The inclusion of zip code fixed effects corrects only part of the endogeneity problem that arises from cross-sectional differences across zip codes. The simultaneity bias illustrated in our simple model in Section 2 will lead the coefficient to be biased upwards in OLS regression even after controlling for fixed effects. At the same time, we face another potential source of bias, measurement error in the rate of uninsured drivers. We use a widely used measure for the rate of uninsured drivers, the uninsured motorist bodily injury claims over the insured motorist bodily injury claims. Since this measure is not a direct observation of the rate of uninsured motorists, but rather an estimate based on accident data, we expect this to be a rather noisy measure of the true rate of uninsured motorists. This measurement error effect will bias the coefficient towards zero.[28] In fact this bias appears to be quite significant in our data, which is not surprising given the inherent noisiness of using accident claims data to measure the rate of uninsured motorists. In Appendix A.3, we further gauge the magnitude of the measurement error in the data and reconcile the seemingly large contrast in the magnitude of our OLS fixed effects estimates and the IV estimates. These competing effects of simultaneity bias and measurement error make the OLS fixed effects estimates uninformative in regards to the true causal effect of the

[28] If instead of observing a variable $x_i$, we observe a noisy measure $x_i^* = x_i + \eta_i$ where $\eta_i \perp x_i$, $E[\eta_i|x_i] = 0$ and $Var[\eta_i|x_i] = \sigma_\eta^2$ and $Var[x_i] = \sigma^2$ the coefficients $\hat{\beta}$ of the regression $y_i = x_i^*\beta + \epsilon_i$, under standard assumptions, will be consistent for $\frac{\sigma^2}{\sigma_\eta^2+\sigma^2}\beta$. When we follow Cohen and Dehejia (2004) and estimate our main specification in logs, which is more robust to measurement error, we find that the difference between the fixed effects and instrumental variables estimates is smaller supporting our hypothesis that measurement error accounts for much of the bias.

rate of uninsured drivers on insurance premia, other than providing us with evidence for the rather weak assertion that the effect is nonnegative.

Fortunately, we can solve the above problems by instrumenting for the rate of uninsured drivers using the staggered introduction of the CLCA program that changes the rate of uninsured drivers. As reported in Table 5, columns (5)-(6), once instrumented for, the coefficient for the rate of uninsured drivers becomes higher in absolute value, with a positive value of \$29.5, or roughly 1-2% of the total value of a typical insurance contract in our data, showing a much larger effect of the uninsured on the insured than methods not controlling for the endogeneity and measurement error problem. Our empirical findings are consistent with theoretical predictions of Smith and Wright (1992) and Keeton and Kwerel (1984) in the auto insurance industry. The magnitude of our results does not change much when we add various demographic and driving record controls, providing an additional test that our instrument is uncorrelated with these controls. The (untabulated) $R^2$ is quite high when we include all controls, at .722, suggesting that our controls explain a great deal of the variation in automobile insurance premia. This is not surprising given that we control for most factors on which firms are legally allowed to price in California, and that we include zip code fixed effects.

As an additional test of the validity of our instrumental variables strategy, the final row of table 5 presents the results of a Hausman test for over identification. The Hausman test is based on the assumption that the average number of months that the CLCA program is in effect is a valid instrument. The null hypothesis is that the estimator is consistent for the same value when additional instruments are added. The test statistic is distributed $\chi^2_2$, and the small values indicate that we fail to reject the null hypothesis that the additional instruments are valid.

Insurance premia are also increasing with the accident rate in a county, which is again consistent with Smith and Wright (1992). If we drop the accident rate from the specification, the coefficient on the rate of uninsured drivers does not change substantially, which suggests that moral hazard does not play a large part in explaining our results.[29] The sign and magnitude of other coefficients in the results presented in Table 5 are also consistent with riskier drivers paying higher premia. Premia are also lower for women and middle aged drivers, which is likely to reflect lower accident rates for women and higher accident rates for inexperienced drivers. The latter point is supported by adding in the number of years licensed to the specifications as controls. Insurance premia are also increasing in the number of miles an individual drives to work daily as well as in speeding tickets and at-fault accidents, both of which are likely to be correlated with an increased risk of being involved in an accident. While our main variable of interest is the rate of uninsured drivers, the other coefficients in the regression also the basic theoretical underpinnings of Smith and Wright (1992), Keeton and Kwerel (1984) and Arrow (1963), namely that premia will be increasing in accident rates and the inherent riskiness of a driver.

[29]See Section 6 for a discussion of moral hazard.

It is illustrative of the challenges in estimating the effect of uninsured drivers on premia to contrast the results of IV estimates in Table 5 with the OLS estimates presented in Table 5. In contrast to the IV estimates, the OLS estimates are not in line with theoretical predictions. The coefficients on the rate of uninsured drivers are negative and significant, which would seem to contradict standard economic theory. The inconsistency between the OLS and the IV estimates is not unexpected, and is likely due to a number of biases. First, we have geographic, time, firm and vehicle biases which probably bias the results in different directions. The negative coefficient in the OLS specification without fixed effects is likely to reflect geographic and firm specific factors such as firms price discriminating by charging customers more in wealthier zip code areas, where we would tend to see fewer uninsured drivers, higher premia and the fact that cars are likely to be cheaper, and thus expected insurer losses are smaller, in poorer areas with a higher rate of uninsured drivers. When we include time and zip code fixed effects in the OLS specification to deal with temporal and geographic biases, the coefficient on the rate of uninsured drivers becomes positive but is still quite small. This coefficient is still uninformative due to a number of biases. First, we have strong endogeneity of the rate of uninsured drivers and insurance premia, $Cov[\lambda_{gt}, \varepsilon_{gijt}] \neq 0$, which should bias the coefficient upwards. Second, we have measurement error bias from our measure of uninsured drivers – uninsured bodily injury claims over total bodily injury claims. This effect would substantially bias our coefficient towards zero, particularly in a panel structure where the true uninsured driver's rate is highly serially correlated and the measurement error is serially uncorrelated white noise. We demonstrate in Appendix A.3 that our estimates and data are indeed consistent with measurement error driving a large portion of the bias. Third, omitted variables bias may also be present which could bias our coefficient in any direction. Due to these biases, the fixed effects OLS results only tell us that the effect is nonnegative, and the magnitude of the effect seems small given prior theoretical work. However, once these biases are dealt with using aspects of the CLCA program as instruments, we see a significant effect of the rate of uninsured drivers on premia, which is consistent with theory.

While the instrumental variables results above provide some evidence for the presence of an externality, there are two potentially serious confounds. First, riskier drivers may sort into the CLCA plan, lowering premia through changing the risk composition of drivers. Second, the CLCA plan is an insurance plan itself, and the presence of the plan in the market may directly cause insurers to lower premia. The following section shows that the potential effects of both concerns are at most very small quantitatively.

## 5.3 Results Excluding Increased Competition and Unobserved Selection

A potential concern to our empirical strategy and results is that the CLCA program, being an insurance plan itself, affects the insurance premium in the commercial market through an increased competition channel. As well as lowering the rate of uninsured drivers, introducing the CLCA program also offered another low-cost plan to consumers which may have forced insurance providers to react by lowering premia. Thus it is possible that our results are partially or entirely driven by increased competition rather than the effect of the CLCA program on uninsured drivers. While we have no data on income to determine eligibility for the CLCA program,[30] we exploit another eligibility requirement of the CLCA program to produce results that are free from this potential confound. In years prior to 2005, only vehicles worth less than $12,000 could be insured under the CLCA program, and this cap was raised to $20,000 in 2006 and following years. We can thus restrict our sample to only those surveyed insurance plans covering vehicles of higher value as to be ineligible for the CLCA program.[31]. Importantly, this restriction also rules out unobserved selection driving our results as these drivers are ineligible for the CLCA program, and hence the risk pool of drivers in this sample did not change before and after the introduction of the program.

Our findings restricting vehicles to be above certain threshold values and ineligible for the CLCA program are reported in Table 6. In column 1 and 2, we restrict the sample to vehicles above their survey year's maximum allowed car value for the CLCA program, while we restrict to vehicles above $20,000, the maximum allowed car value throughout the years in the CLCA program in column 3 and 4. To the extent that one might be concerned about potential spillover from lower car-value plan to higher car-value plan or a coarse pricing strategy by insurance companies, we conduct a "stress test" by restricting the sample to vehicles to be above $25,000 in column 5 and 6. If increased competition due to a new plan being offered could explain the bulk of our findings, we would expect the coefficient on the rate of uninsured drivers to drop substantially. Results from our three different sub-samples show this is not the case: while the point estimates of these regressions are slightly lower than that of the regression in Table 5, they are in fact statistically indistinguishable from our main result, given the magnitude of standard errors. Our results are again statistically significant at 1% level. This demonstrates that increased competition cannot explain our findings, and that the effect of the CLCA program on premia comes almost entirely from decreasing the rate of uninsured drivers.

The magnitude of the IV estimate are consistent with a stylized model of automobile insurance pricing in which the externality is fully passed onto consumers. An insurance company will be required to pay for damages in two scenarios, one, if the driver is involved in an accident and found to be at fault, and two, if the driver is involved in an accident with an uninsured driver. We denote a rate $\lambda$ of uninsured drivers and assume that (1) the probability $\theta = \frac{\pi_i}{2}$ of being at fault in an accident for insured or uninsured driver is the same and (2) insured and uninsured drivers in expectation cause the same amount of loss, $L$. Under competitive pricing, the insurance premium is $P = L(\theta + \lambda\theta)$. Thus, for a 1 percentage point decline in the rate of uninsured drivers, the percentage decrease in premium will be $\frac{dP}{P} = \frac{0.01L\lambda\theta}{L(\theta+\lambda\theta)} = \frac{0.01}{1+\lambda}$. In California the rate of uninsured

[30]In fact it is illegal for insurers in California to price on factors such as income or race.

[31]It is important to note that we are also likely throwing out many individuals who were not eligible for the CLCA program, as vehicle value was not the only criterion for eligibility.

drivers is roughly 20%, so a 1 percentage point drop in the rate of uninsured drivers will reduce the premium by 0.83%. Given that the average premium in our data is roughly $2,356,[32] and we estimate that a 1% decrease in the rate of uninsured drivers reduces premia by $24.8 (column 6 of Table 6), the aforementioned logic is very much in line with our results.

When aggregated over all insured drivers in California the social costs of the externality[33] are substantial. Based on our main specification, and uninsured motorists rates in California in 2007 as well as rates of uninsured motorist coverage,[34] the total cost of the externality to California is about $6 billion. If the magnitude of the effect in other US states is similar in size to California on a per-person basis, the size of the externality would be quite large, which we calculated to be at $27 billion nation-wide using NAIC estimates of average premia.[35] If the magnitude of the effect is similar in the United Kingdom, we would estimate the size of the externality to be roughly £1.6 billion. This is substantially smaller than in the United States, given that the rate of uninsured motorists in the United Kingdom is only 3.5%. The Motor Insurers' Bureau levies a £33 surcharge on automobile insurance premia to fund damage arising from uninsured motorists. We note that this is quite close to our estimates in California– we would predict that uninsured motorists would raise premia by $100 (£50) if the rate of uninsured motorists is 3.5%.

## 5.4 Pigouvian Taxation

The presence of externalities can be corrected by pricing the damage caused by uninsured drivers to other drivers. One way to accomplish this task is by levying a Pigouvian tax, or equivalent fine on uninsured drivers. Individuals would then only fail to purchase insurance if their private benefit exceeds the external social cost of being uninsured. This is in effect the system already in place in most of the United States directly or indirectly,[36] as well as many other countries. While ostensibly it is illegal for motorists to drive without insurance in most US states, the current system closely mimics a Pigouvian tax. In most US states drivers who are caught without insurance are forced to pay a citation, which is essentially equivalent to a stochastic Pigouvian tax on driving uninsured. In theory authorities could set fines large enough so that very few drivers drive without insurance,[37]

[32]The average premium in our data is larger than the typical premium paid in California since the survey data oversamples drivers with at fault accidents and speeding tickets.

[33]There are, of course, other externalities associated with automobile use. See Parry et al. (2007) for a survey of externalities associated with automobile use and Edlin and Karaca-Mandic (2006) for a discussion of the general externality caused from miles driven.

[34]In 2007, Department of Insurance data indicate that 17.83% of motorists were uninsured, and there were 19,280,329 vehicles with uninsured motorist coverage in the state of California.

[35]We caution that our estimates are local.

[36]Most US states levy substantial fines for driving without insurance. Virginia directly allows individuals to pay a $500 fine to opt out of auto insurance.

[37]This is the case in some European countries, for example, in France in 2012 if one is caught driving without insurance the fine is €3,750 accompanied with a three-year license suspension. Given these exceptionally high fines,

but intuitively the welfare effects of forcing uninsured motorists to buy insurance without a subsidy are ambiguous. The fine would disproportionately affect low income households, where most uninsured drivers tend to be located.[38]

There exists a long tradition since Pigou (1920) of economists advocating corrective taxes on externalities.[39] However, despite the optimality of Pigouvian taxation in the presence of externalities, determining what corrective taxes should be levied is often difficult in practice. Typically, the most daunting challenge is measuring the size of the externality, which we have accomplished in the previous section of this paper. To accomplish our objective, we can levy a Pigouvian tax on uninsured drivers in a fashion similar to how most US states currently fine uninsured motorists. Authorities force uninsured drivers to pay a tax $\tau$ if they are uninsured and redistribute a subsidy $s$ to all drivers. Given the framework outlined in the theory section and under some weak assumptions, we can compute the optimal fine which only depends on observables. Implicitly, the probability of being caught uninsured must be factored into the tax, as currently drivers will only pay the tax if they are stopped by law enforcement officials. The tax will reduce the size of the externality by discouraging drivers from driving uninsured, while at the same time directly lowering premia by subsidizing insured drivers. Essentially the government can use a tax to correct the externality, fining uninsured drivers and redistributing the proceeds to all drivers. Given three possible states, no accident, an accident with an insured driver, and an accident with an uninsured driver, consumers choose optimal amounts of insurance to purchase much along the lines presented in Section 2. After consumers have made optimal insurance choices, the government solves for a representative consumer with insurance choice determined by consumers' optimization, $max_\tau\ V(s,\tau)$ for given tax $\tau$ and subsidy $s$, subject to the government budget being balanced, $s = \lambda(\tau)\tau$. Solving the government's problem and applying the envelope theorem, after some algebra we obtain that the optimal corrective tax depends only on $\beta$, and $\lambda(\tau)$ given by

$$\tau^* = \beta(1 - \lambda(\tau)).$$

See Appendix B for a detailed derivation of the formula, which follows Chetty (2006) in spirit. The optimal tax formula is simple and intuitive, depending on $\beta$, the amount of premia increase from uninsured drivers and $\lambda(\tau)$, the rate of uninsured drivers. The result indicates that uninsured individuals should fully bear the cost of the externality, which is similar to the Pigouvian tax found in Edlin and Karaca-Mandic (2006). The fine is unambiguously increasing in $\beta$ , which is

it is no surprise that the rate of uninsured motorists in France is quite low, at .1% of registered vehicles compared to 14% in the US. Many European countries also have rates of uninsured motorists substantially lower than the US, as well as higher penalties for driving without insurance.

[38]See Hunstad (1997) for a discussion of the characteristics of uninsured motorists in California. See Zimolo (2010) for more information.

[39]For the sake of brevity, we do not offer a full treatment of Pigouvian taxation. See Sandmo (1978) for a classic treatment of the problem or Mankiw (2009) for a more recent discussion of Pigouvian taxes.

the externality that the Pigouvian tax is designed to correct. A larger effect stemming from this externality would mean a larger corrective fine. As we would expect, the fine is zero if there is no externality. We note that the optimal tax is always positive and thus will be a fine on the uninsured and a subsidy for the insured.

The results indicate that any redistributive fines for driving without insurance should be $2,240. This value is substantially higher than current fines in California, where individuals pay between $100-200 for the first offense and $500 for the second. This difference becomes even clearer when we note that enforcement is stochastic.[40] It is thus quite possible that, if relatively few drivers are caught driving without insurance, current fines are substantially below the optimum. It is difficult to determine the expected fine that California residents would pay for driving uninsured, as statewide data does not exist on tickets for driving uninsured. Many states and several European countries levy fines for driving without insurance that are substantially greater than those of California, and it is quite possible that those are in line with the optimal rate. If the optimal fine of $2,240 were enforced rigorously, this would effectively eliminate the uninsured driver problem as it would be cheaper for nearly all individuals to purchase a basic insurance plan rather than pay a heavy fine.

# 6 Robustness

Table 7 presents several robustness checks which indicate that our basic result holds controlling for several potential confounds. All specifications include fixed effects as well as proxies for the three mandatory auto insurance pricing factors: years licensed, driving record and miles driven per day and other controls. In all cases the coefficients remain significant at the .01 level and are similar in magnitude to the main results. In the main dataset, we restrict the sample to only observations where there is one driver on the insurance plan. Our main results are robust to including multi-driver policies as well. Concerns with the data and our measure of uninsured drivers are addressed in Appendix A. We go over potential concerns about our results one by one in the following sections.

## 6.1 Weak Instruments

A typical concern regarding the use of the instrumental variables approach is the strength of the instruments used in the study. While the F-statistic of 14.83 in our first-stage regression exceeds the standard rule-of-thumb threshold of 10 for weak instruments (Staiger and Stock (1997)), one might still be concerned about whether instruments with this degree of power were able to produce stable estimates. As recommended by Stock et al. (2002), we reestimate the models in column (5) and (6)

[40]See Polinsky and Shavell (1979) for a discussion of the tradeoff between the probability and magnitude of fines.

of Table 5 using limited information maximum likelihood (LIML) methods. The implementation follows the approach in Moreira and Poi (2003). The corrected confidence intervals are constructed from tests of coefficients based on the conditional distributions of nonpivotal statistics. We find in column 1 and 2 of Table 7 that our results from LIML are extremely close to the IV results, which indicates that weak instruments should not be a serious concern in our context.

## 6.2 County Waves

As we show in Figure 2, counties in the last wave of the CLCA program have higher accident rates and are smaller and more sparsely populated. These counties also tend to have slightly lower premia than other counties. To guard against the potential confound that the results are driven by some particular counties that have different characteristics than others, we exclude the last wave counties. The results are presented in column 3 of Table 7. In this specification the coefficient on the rate of uninsured drivers turns out to be close to our main results. We conclude that our results are not driven by the counties in the final wave being different from other counties. Our results are robust to excluding counties in any wave, for example, Los Angeles and San Francisco in the pilot wave, and we conclude that our results are not driven by any single wave.

## 6.3 Unobserved Selection

A potential concern is that unobserved selection on accident risk could play a major role in determining premia. For example, drivers switching to the CLCA program could be unobservably riskier than those remaining in traditional insurance plans. This effect could lead insurance premia to fall for those remaining in traditional insurance plans. We view unobserved selection as unlikely given the regulation of automobile insurance pricing in California. First, following Proposition 103, automobile insurers are only allowed to price on certain factors, the vast majority of which are in our dataset. It is not clear why unobservably risky individual would prefer the CLCA plan to traditional insurance plans with higher coverage limits. At the same time, we refer to our results in Table 6 where we have restricted to several subsamples with car-values above the eligible value for the CLCA program. Drivers purchasing these types of insurance plans would not be affected by any unobserved selection into the CLCA program as they are ineligible for the program. We see nearly identical effects for individuals who were ineligible to enter the CLCA program due to high vehicle values, and for this group unobserved selection into the CLCA program cannot explain the price effects. Thus we conclude that unobserved selection is not a major driving force of our results.

## 6.4 County Specific Time Trend

It is possible that the introduction of the CLCA program at county level was simply correlated with other factors that reduced premia. We test for this possibility by including county-specific linear time trend in our main specification and the result is in column 4 of Table 7. The source of variation is at the zip code cluster level, given that a zip code could border other zip codes within its 25 miles radius in a different county that had the CLCA program introduced in a different wave. This feature will make sure that including a county-specific linear time trend will not fully absorb our source of variation for identification. Indeed we still obtain a significant result of the effect of uninsured drivers on premia at the 1% level. The point estimate changes from 29 to 34, but given the relatively large standard error, we can not reject that it is different from the results in our main regression in Table 5 at 5% level of significance. We also provide results for a full range of specifications including county-specific linear time trend in the Appendix Table A1.

## 6.5 Choice of Instruments

Our instruments in the main regression include an indicator of whether the CLCA program was active for over 4 months in a zip code cluster. We vary the definition of this instrument and our results remain robust. Column 5 of Table 7 reports results when we replace it with an indicator of whether or not the CLCA program was active at all during the year. We use an indicator of whether or not the CLCA program was active for over 6 months during the year instead in column 6 and dropped this indicator in column 7. In either of the above cases, altering the definition of our instrument creates almost no change to our main result. Our results are also robust to dropping any individual instrument. We conclude that our results are robust to changes in instrument specification.

## 6.6 Omitted Variables

Another potential concern is that coefficients in our specifications are subject to omitted variables bias. We do not think that this is a significant source of bias given the richness of our data and the regulatory framework in California. Automobile insurance is highly regulated in California, and we have all factors on which insurers are required to price, as well as, in the authors' view, the more important optional pricing factors. Proposition 103, passed in 1988, modified the California Insurance Code[41] to mandate that automobile insurers in California could only price on driving record, miles driven annually, and the number of years licensed. In addition, insurers were also allowed to price on secondary factors permitted by the insurance commissioner. For the period in which the authors have data (2003-2007), insurance companies were permitted to price on loca-

[41] Section 1861.02 (a)

tion (zip code), vehicle type and performance, number of vehicles owned by the household, the use of vehicles, gender, marital status, age, demographic characteristics of secondary drivers, persistency, the academic standing of any student in the household, completion of a driver training course, smoking, bundling of products with the same company and claims frequency and severity. Automobile insurers were not allowed to price on any other characteristics, and firms were required to report rate changes in their pricing formulae to the Department of Insurance. The mandatory pricing factors were also required to have a larger weight in the pricing formula than the optional pricing factors.[42] Given that our data includes information on all mandatory pricing factors, as well as the major optional pricing factors for automobile insurance pricing in California, we think it is unlikely that our results are significantly biased by omitted variables.

## 6.7 Underlying Change in Variables that Drive CLCA Wave Timing

In Section 4.1, we show that the specific timing of waves in the CLCA program is driven by the level of county population size. While the county population should be a stationary variable that we do not expect to be changing rapidly, we guard against the effect of changing population on the demand of automobile insurance in each county by including county population interacted with year dummies in column 8 in Table 7. We find the effect of uninsured drivers on premia declines slightly, but still we can not reject that it is identical to what we find in our main results at 5% level of significance.

## 6.8 Moral Hazard

One potential concern is that our results may slightly overestimate the effect of uninsured drivers, as the CLCA program also introduced moral hazard. In theory, increased insurance coverage should increase the risk of an accident.[43] By covering previously uninsured individuals, the program may have given some drivers an incentive to drive in a less safe manner. Chiaporri and Salanié (2000) find no evidence of asymmetric information in the automobile insurance market using the positive correlation test. Dionne et al. (2013) find evidence of moral hazard among people with less driving experience in France from 1995-97. They find no information problem for people with more than 15 years of experience. Abbring et al. (2003) also find no evidence of moral hazard using dynamic insurance data and a test similar to Abbring et al. (2003). However Cohen (2005) notes that the results of Chiaporri and Salanié (2000)[44] are also consistent with asymmetric

[42] As stipulated under California Insurance Code Section 1861.02(a).

[43] See Shavell (1979) or Arrow (1971) for early discussions of this effect.

[44] Chiaporri and Salanié (2000) use a French dataset which focuses on younger drivers. Cohen (2005) finds a positive correlation between a lower deductible and accidents for drivers with three or more years of experience, but not for drivers with less than three years of experience. Dionne et al. (2013) also find no information problem for experienced drivers using a French dataset. This relationship is consistent with classic adverse selection theory,

information and learning. Cohen and Dehejia (2004) estimate the effect of automobile insurance on traffic fatalities and find significant effects of moral hazard in the automobile insurance market. Furthermore Table 3 indicates that there is a small but marginally significant (at the 10% level) increase in accident rates following the roll out of the CLCA program. This is consistent with the moral hazard hypothesis. However, we control for moral hazard effects of the CLCA program by including the local accident rate in our regression. Thus our estimates may overcontrol for moral hazard as any increase in traffic accidents due to moral hazard will be reflected in the coefficient on the accident rate. As a test for whether or not moral hazard significantly affects automobile insurance premia, we can drop the county level accident rate from the specification. When we run this alternative specification, the coefficient on the rate of uninsured drivers changes only slightly, and the difference is not significantly different from zero. We take this as evidence that moral hazard does not play a significant part in our results.

### 6.9 Results Collapsing to the Level of Variation

We provide an additional robustness check by collapsing our raw insurance premia data to the zip code by year level. We take the average of the insurance premia from the raw data for each zip code and year. This way our sample is reduced from around four million observations to roughly one thousand. We use this collapsed sample to examine the effect of the rate of uninsured drivers on insurance premia, adopting the same instrumental variables approach as in our main results section.[45] The results are reported in Appendix Table A2 when we gradually add controls for the fixed effects in columns 1 to 3. The results indicate the coefficient for uninsured driver to be 23.13, significant at 1% level. Given the standard error of the estimates, this is very close to our main results.

## 7 Concluding Remarks

This paper makes two contributions. First, we empirically gauge the magnitude of the negative externality generated by uninsured parties in insurance markets, and second, we discuss the optimal corrective Pigouvian tax for this externality based on our empirical analysis. This paper uses a novel panel data set on auto insurance premia in California to quantify the negative externality generated by uninsured drivers on the insured. We overcome the endogeneity challenge inherent

however the study cannot disentangle between adverse selection and moral hazard. Cohen and Einav (2003) also find that seat belt use does not increase reckless driving, providing further evidence against another type of moral hazard effects in automobile accidents. See Cohen and Siegelman (2011) for a review of the empirical literature on adverse selection in insurance markets and Einav et al. (2010) for a treatment of welfare effects.

[45] We modify the months instrument accordingly to be the number of months that the CLCA program is in effect in a particular zip code.

in the relationship between insurance premia and the rate of the uninsured, utilizing exogenous variations from the staggered introduction of a policy that lowers the rate of uninsured drivers.

Our data set and empirical strategy enable us to directly estimate the effect of uninsured on premia. Consistent with predictions of the theory, our study suggests that higher rates of uninsured drivers has a significant effect on the auto insurance premium. We estimate that a 1-percentage-point increase in the rate of uninsured drivers leads to a roughly \$28 increase in automobile insurance premia, which is 1% of the total value of the insurance contracts in our data. These estimates imply that each driver could save almost \$500 if every motorist became insured in the state of California, which would reduce automobile insurance costs by roughly a third.

This study also develops a new formula for computing the optimal corrective tax or fine on uninsured individuals. This formula is parsimonious, relying only on the size of the externality and the rate of uninsured drivers. We compute that the optimal fine should be \$2,240, which is substantially higher than current fines in most US states, although similar to fines in some European countries like France.

A fruitful avenue for further research would be to estimate the effect, if any, of the uninsured on health insurance premia. Theory work has noted that there may be a similar effect in the health insurance market resulting from the regulatory requirement that hospitals cross-subsidize the uninsured, and this effect has of late become an important policy issue in the United States due to the passage of the Patient Protection and Affordable Care Act in 2010.[46] However, as of yet there is no direct empirical evidence that this regulatory externality raises premia. While the direct effect of the uninsured not paying medical bills is similar to the effect of uninsured motorists not paying for collision damages after accidents in which they are at fault, there are also a host of significant moral hazard risks associated with medical care as well as externalities from communicable diseases and other effects which could make the true effect of the uninsured on premia substantially different. While our quantitative results concern only the automobile insurance market, estimating the effect, if any exists, of the uninsured on health insurance premia would serve both to test the predictions of economic theory and better inform the policy debate about health care.

[46] See Gruber (2008) for a survey of the literature on the uninsured in the health care market.

# References


Abbring, J., P.-A. Chiappora, and J. Pinquet (2003). Moral hazard and dynamic insurance data. *Journal of the European Economic Association 1*(4), 767–820.

Abbring, J., J. Heckman, P.-A. Chiappora, and J. Pinquet (2003). Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish. *Journal of the European Economic Association 1*(2-3), 512–521.

Anderson, M. (2008). Safety for whom? The effects of light trucks on traffic fatalities. *Journal of Health Economics 27*(4), 973–89.

Anderson, M. and M. Auffhammer (2014). Pounds that kill: The external costs of vehicle weight. *The Review of Economic Studies*.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review 53*(5), 941–973.

Arrow, K. (1971). Insurance, risk and resource allocation. *Essays in the Theory of Risk Bearing Chicago: Markham Publishing Company.*

Bureau of Transportation Statistics (2006). Highway vehicle miles traveled. *Federal Highway Administration, Washington, DC*.

Chetty, R. (2006). A general formula for the optimal level of social insurance. *Journal of Public Economics 90*(10), 2351–2356.

Chiaporri, P.-A. and B. Salanié (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy 108*(1), 56–78.

Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. *Review of Economics and Statistics 87*(2), 197–207.

Cohen, A. and R. Dehejia (2004). The effect of automobile insurance and accident liability laws on traffic fatalities. *Journal of Law and Economics 47*(2), 357–93.

Cohen, A. and L. Einav (2003). The effect of mandatory seat beat laws of driving behavior and traffic fatalities. *Review of Economics and Statistics 87*(4), 823–843.

Cohen, A. and P. Siegelman (2011). Testing for adverse selection in insurance markets. *The Journal of Risk and Insurance 77*(1), 39–84.

Dahlby, B. and D. S. West (1986). Price dispersion in an automobile insurance market. *The Journal of Political Economy 94*(2), 418–438.

Dionne, G., P.-C. Michaud, and M. Dahchour (2013). Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from France. *Journal of the European Economic Association 11*(14), 897–917.

Edlin, A. S. and P. Karaca-Mandic (2006). The accident externality from driving. *Journal of Political Economy 114*(5), 931–955.

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics 124*(3), 877–921.

Gruber, J. (2008). Covering the uninsured in the United States. *Journal of Economic Literature 46*(3), 571–606.

Hunstad, L. (1997). Characteristics of uninsured motorist. *California Department of Insurance* (February).

Hunstad, L. (1999). Estimating the uninsured vehicle rate from the uninsured motorist/bodily injury ratio. *NAIC Research Quarterly 5*, 1–7.

Insurance Research Council (1999). Uninsured motorists. *Insurance Research Council Wheaton, Illinois, October*, 1–7.

Keeton, W. R. and E. Kwerel (1984). Externalities in automobile insurance and the uninsured driver problem. *Journal of Law and Economics 27*, 149–79.

Khazzoom, D. J. (1999). What we know about uninsured motorists and how well we know what we know. *Journal of Insurance Regulation 18*(1), 82.

Levitt, S. and J. Porter (2001). How dangerous are drinking drivers? *Journal of Political Economy 109*(6), 1198–1237.

Mankiw, N. G. (2009). Smart taxes: An open invitation to join the Pigou club. *Eastern Economic Journal 35*, 14–23.

Mas-Colell, A., M. Whinston, and J. Green (1995). Microeconomic theory. *Oxford University Press, Oxford*, 1–7.

Moreira, M. J. and B. P. Poi (2003). Implementing tests with correct size in the simultaneous equations model. *Stata Journal 3*(1), 57–70.

Parry, I., M. Walls, and W. Harrington (2007). Automobile externalities and policies. *Journal of Economic Literature 45*(2), 373–399.

Pigou, A. (1920). The economics of welfare. *London: MacMillan*.

Polinsky, M. and S. Shavell (1979). The optimal tradeoff between the probability and magnitude of fines. *American Economic Review 69*(5), 880–891.

Sandmo, A. (1978). Optimal taxation in the presence of externalities. *The Swedish Journal of Economics 77*(1), 86–98.

Schultz, C. and L. Yarber (2006). Report to the legislature consumer education and outreach plan. *California Department of Insurance*.

Schultz, C. and L. Yarber (2011). Report to the legislature consumer education and outreach plan.

*California Department of Insurance*.

Shavell, S. (1979). On moral hazard and insurance. *The Quarterly Journal of Economics 93*(4), 541–562.

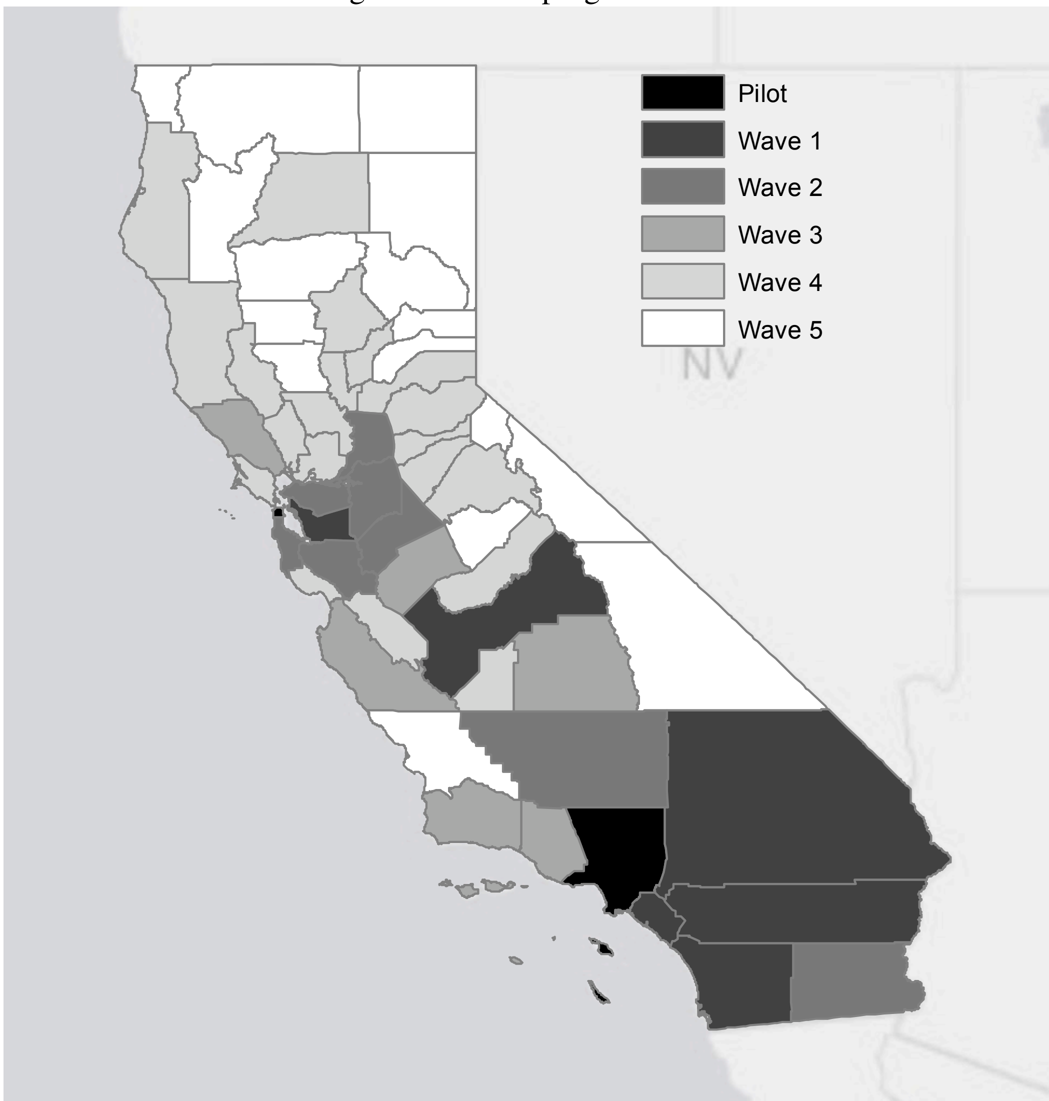
Smith, E. F. and R. Wright (1992). Why is automobile insurance in Philadelphia so damn expensive. *American Economic Review 82*(4), 756–72.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002). Testing for weak instruments in linear iv regression. *Journal of Business and Economic Statistics 20*(4), 518–529.

Zimolo, A. (2010). Normative and management characteristics of motor third party liability insurance in the world. *Motor Insurance Working Group, Paris, May 17-20*.

Figure 1: CLCA program Waves




Pilot Counties (1999)- Los Angeles and San Francisco.

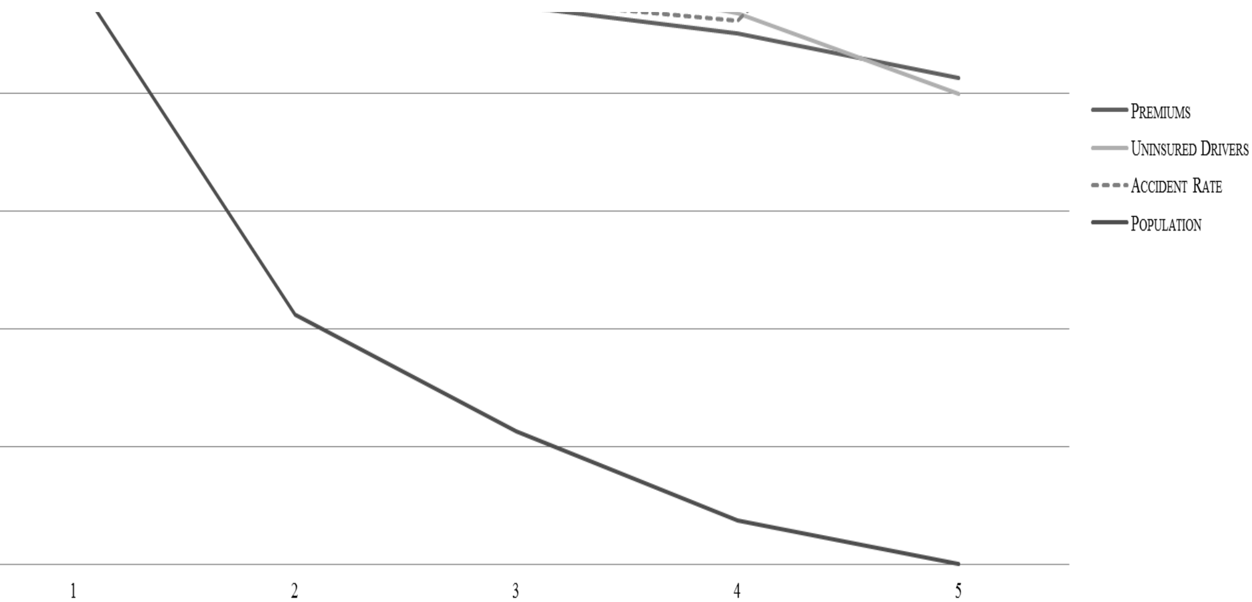
Wave 1 (April 1, 2006)-Alameda, Fresno, Orange, Riverside, San Bernardino, San Diego.

Wave 2 (June 1, 2006)-Contra Costa, Imperial, Kern, Sacramento, San Joaquin, San Mateo, Santa Clara, Stanislaus.

Wave 3 (March 30, 2007)- Merced, Monterey, Santa Barbara, Sonoma, Tulare, Ventura.

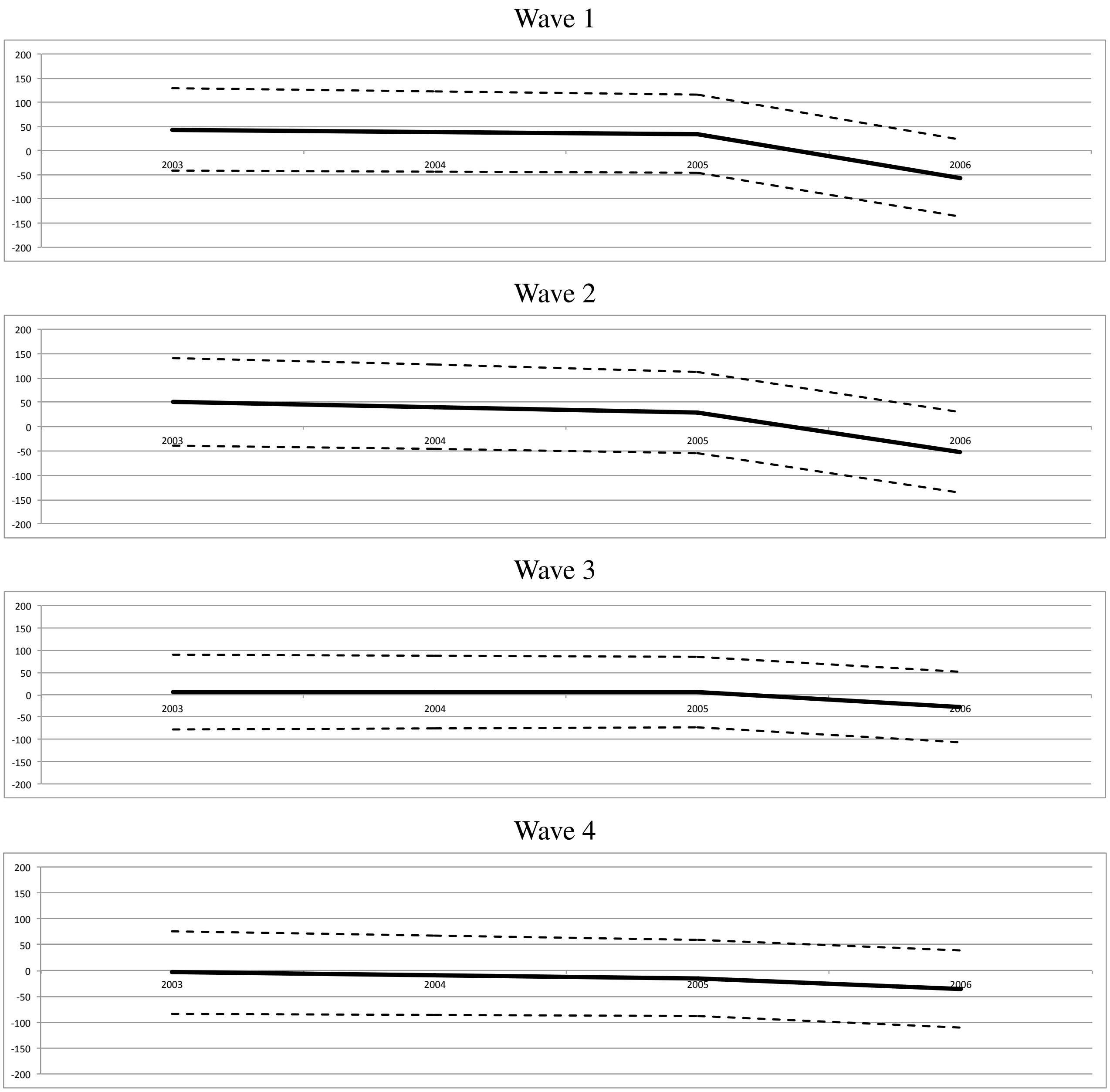
Wave 4 (October 1, 2007)-Amador, Butte, Calaveras, El Dorado, Humboldt, Kings, Lake, Madera, Marin, Mendocino, Napa, Placer, San Benito, Santa Cruz, Shasta, Solano, Sutter, Tuolumne, Yolo, Yuba.

Wave 5 (December 10, 2007)-Alpine, Colusa, Del Norte, Glenn, Inyo, Lassen, Mariposa, Modoc, Mono, Nevada, Plumas, San Luis Obispo, Sierra, Siskiyou, Tehama, Trinity

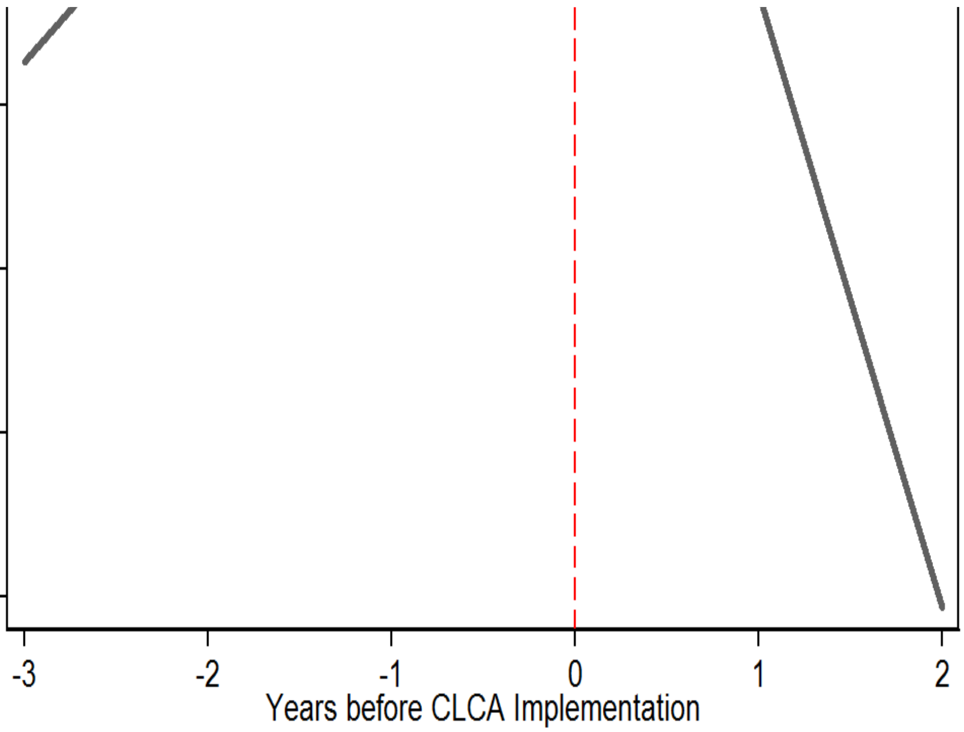



s figure plots average inflation adjusted premia, the rate of uninsured drivers, accident rates and population across each e implementation of the CLCA program. The spike in accident rates in the fifth wave is driven by counties in the Lake on. Including these counties does not change the results significantly. The CLCA program was effectively assigned ion size so we see a clear decreasing trend in population size across CLCA waves, while we do not see significant in other variables.

Figure 3: No Significant Pre-Trend Across Waves




Notes: This figure plots the estimated difference (wave by year fixed effects, where the fifth wave is omitted to avoid multicollinearity) from a regression of premia on individual, geographic, temporal and vehicle controls. Confidence bands at the 95% level are included matching each line style. Note that in the first two waves, the CLCA program went into effect in 2006.




s figure plots average the rate of uninsured drivers leading up to and immediately after the implementation of program. See figure 1 for the dates of implementation of the CLCA program in specific counties. The dashed s the year in which the CLCA program was implemented in the county.

Table 1: Summary Statistics

| | Mean of Analysis Variable | | | | | |
|---|---|---|---|---|---|---|
| **Wave** | Pilot | 1 | 2 | 3 | 4 | 5 |
| Premium | 3167.957 | 2224.571 | 2224.995 | 2092.118 | 1992.882 | 1821.962 |
| Uninsured Rate | 20.66453 | 20.78925 | 22.13009 | 22.09136 | 23.52509 | 18.61633 |
| Accident Rate | 1.209594 | .8731114 | .8489298 | .8195649 | .7927318 | .9557761 |
| At Fault | .4724665 | .4727527 | .4741157 | .4737676 | .4729685 | .4727535 |
| Standard | .7673367 | .752907 | .7473121 | .7522681 | .7477136 | .7486086 |
| Age | 30.02012 | 30.00553 | 30.00228 | 30.00606 | 30.00942 | 30.00897 |
| Mile Driven per Day | 12.44483 | 12.44016 | 12.43941 | 12.43821 | 12.43904 | 12.43824 |
| Speeding | .4724166 | .4728509 | .472685 | .4727655 | .4721361 | .4720728 |
| Female | .4654293 | .4314502 | .4203862 | .4321344 | .420959 | .4220819 |
| $N$ | 1,744,022 | 1,405,293 | 518,616 | 335,310 | 496,198 | 224,781 |

Notes: The table provides the means of the analysis variables used in the study. The data source for uninsured motorists is the California Department of Insurance. The data source for the number of accidents is the California Highway Patrol. The data source for all other variables is the California Department of Insurance Automobile Premium Survey.

Table 2: Automobile Insurance Plan Coverage

| | Basic Coverage | Standard Coverage | CLCA Plan |
|---|---|---|---|
| Bodily Injury | \$15,000/\$30,000 | \$100,000/\$300,000 | \$15,000/\$20,000 |
| Property Damage | \$5,000 | \$50,000 | \$3,000 |
| Medical Payments | \$2,000 | \$5,000 | – |
| Uninsured Motorist Bodily Injury | \$15,000/\$30,000 | \$30,000/\$60,000 | – |
| Comprehensive Deductible | – | \$250 | – |

Notes: Bodily Injury (BI) claims are the maximum that an insurance company will pay per person and the maximum an insurance company will pay for injuries from a specific accident. Uninsured Motorist Bodily Injury (UMBI) claims are the maximum that an insurance company will pay per person and the maximum an insurance company will pay for injuries from a specific accident where an uninsured motorist is at fault. California law mandates BI and property damage coverage according to the basic liability-only policy.

Table 3: Survey Sample Characteristics

| | No CLCA Mean | CLCA Mean | Difference | p-value | Observations |
|---|---|---|---|---|---|
| Female | .454 | .427 | .027 | .236 | 4,723,816 |
| | (.006) | (.020) | | | |
| Age | 30.002 | 29.974 | .055 | .139 | 4,723,816 |
| | (.004) | (.017) | | | |
| Standard | .761 | .748 | .013 | .219 | 4,723,816 |
| | (.003) | (.009) | | | |
| Accident Rate | .872 | 1.065 | -.193 | .054 | 4,723,816 |
| | (.029) | (.095) | | | |
| Daily Miles Drive | 12.441 | 12.436 | .005 | .243 | 4,723,816 |
| | (.001) | (.004) | | | |
| At Fault Accident | .483 | .484 | -.001 | .659 | 4,723,816 |
| | (.001) | (.003) | | | |
| Speeding Ticket | .483 | .484 | -.001 | .616 | 4,723,816 |
| | (.001) | (.002) | | | |

Notes: The first column presents the mean of the variable in the row before the CLCA program has been active for at least four months. The second column presents the mean of the variable in the row after the CLCA program has been active for at least four months. The third column presents the difference. The fourth column presents the p value from an F test that the hypotheses are the same. The final column presents the number of observations. Standard errors are clustered at the county level.

Table 4: Effect of CLCA program on the Rate of Uninsured Drivers

| | (1) CLCA Program in Effect | (2) Months CLCA in Effect | (3) Months CLCA in Effect$^2$ | (4) CLCA Enrollment on Uninsured |
|---|---|---|---|---|
| | -1.090*** | -0.151*** | -0.0154*** | -0.395*** |
| | (0.3145) | (0.0437) | (0.0038) | (0.14304) |
| Observations | 4,724,220 | 4,724,220 | 4,724,220 | 71 |

Notes: * $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the rate of uninsured drivers in a county, measured by UMBI/BI. The independent variable in specification (1) is an indicator of whether or not the CLCA program was in effect in the zip code cluster for more than half the year. The independent variable in specification (2) is the average number of months the CLCA program is active in a 25 mile radius around the zip code where the premium quote is located. The independent variable in specification (3) is the average number of months the CLCA program is active in a 25 mile radius around the zip code where the premium quote is located squared. The dependent variable in specification (4) is the numbered of uninsured drivers in a county, while the independent variable is the number of individuals enrolled in the CLCA program in 2006 and 2007. Standard errors are in parentheses and are clustered at the zip code cluster level.

Table 5: Main Results for Effects of the Uninsured on Premia

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | OLS | OLS | FE | FE | IV | IV |
| Uninsured Drivers | -11.19* | -13.01* | 3.205** | 3.169** | 25.17*** | 29.51*** |
| | (6.613) | (6.708) | (1.553) | (1.565) | (4.875) | (5.853) |
| Accident Rate | 1337.6*** | 1352.8*** | 262.6*** | 257.9*** | 258.1 | 286.2* |
| | (359.8) | (363.1) | (75.49) | (75.45) | (167.5) | (171.5) |
| At Fault Accident | 770.5*** | 804.8*** | 739.4*** | 780.5*** | 726.1*** | 789.6*** |
| | (13.18) | (13.79) | (13.64) | (14.43) | (10.95) | (13.02) |
| Standard | 1669.2*** | 1695.0*** | 1020.9*** | 1784.0*** | 1590.4*** | 1750.3*** |
| | (24.69) | (25.63) | (14.72) | (29.94) | (19.68) | (152.9) |
| Age | | -43.37*** | | -46.47*** | | -43.69*** |
| | | (0.928) | | (0.967) | | (4.179) |
| Daily Miles | | 38.59*** | | 39.50*** | | 43.24*** |
| | | (0.546) | | (0.565) | | (3.190) |
| Speeding Ticket | | 593.1*** | | 556.8*** | | 558.2*** |
| | | (9.984) | | (10.59) | | (8.233) |
| Female | | -148.2*** | | -169.3*** | | -148.4*** |
| | | (8.947) | | (4.046) | | (6.884) |
| Hausman | | | | | .124 | .126 |
| Observations | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 |

Notes:* $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. The rate of uninsured drivers is measured between 0 and 100. In the IV estimates the rate of uninsured drivers is instrumented using (i) the average number of months during which the CLCA program was in effect in a zip code cluster (ii) the average number of months during which the CLCA program was in effect in a zip code cluster squared and (iii) an indicator of whether or not the program has been in effect for more than four months. The rate of uninsured drivers is measured by UMBI/BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. Columns 3,4,5 and 6 include zip code, year, firm and vehicle fixed effects. Standard errors are in parentheses and are clustered at the level of the zip code cluster.

Table 6: Results Using Various Vehicle Value Thresholds Above Eligibility

| | (1) Ineligible | (2) Ineligible | (3) >$20,000 | (4) >$20,000 | (5) >$25,000 | (6) >$25,000 |
|---|---|---|---|---|---|---|
| Uninsured Drivers | 23.78*** | 26.37*** | 21.44*** | 27.06*** | 20.20*** | 24.82*** |
| | (4.628) | (5.391) | (4.239) | (5.343) | (5.749) | (5.562) |
| Accident Rate | 255.1 | 271.2* | 222.0 | 254.3 | 268.9* | 302.1* |
| | (155.9) | (154.4) | (147.1) | (154.8) | (157.5) | (168.4) |
| At Fault Accident | 693.6*** | 756.2*** | 643.8*** | 700.0*** | 759.3*** | 856.2*** |
| | (11.38) | (14.09) | (9.340) | (11.03) | (15.03) | (12.97) |
| Standard | 1959.0*** | 1721.5*** | 1905.8*** | 1886.9*** | 1661.7*** | 1319.1*** |
| | (19.32) | (63.79) | (25.73) | (41.63) | (132.8) | (50.51) |
| Age | | -40.61*** | | -36.62*** | | -37.84*** |
| | | (3.009) | | (1.376) | | (5.234) |
| Daily Miles | | 38.63*** | | 38.26*** | | 45.56*** |
| | | (3.336) | | (1.966) | | (1.821) |
| Speeding Ticket | | 536.1*** | | 493.4*** | | 596.0*** |
| | | (11.24) | | (12.55) | | (42.66) |
| Female | | -121.4*** | | -81.05*** | | -129.7*** |
| | | (4.334) | | (3.737) | | (38.18) |
| Observations | 3,802,252 | 3,802,252 | 3,230,538 | 3,230,538 | 1,699,610 | 1,699,610 |

Notes:* $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. The rate of uninsured drivers is measured between 0 and 100. Columns 1 to 2 restrict the sample to vehicles ineligible for the CLCA program in the current year, while columns 3 to 4 restrict the sample to vehicles ineligible for the CLCA program during the entire sample period. We restrict the sample to vehicles with values above $25,000 in columns 5 to 6. In all estimates the rate of uninsured drivers is instrumented using (i) the average number of months during which the CLCA program was in effect in a zip code cluster (ii) the average number of months during which the CLCA program was in effect in a zip code cluster squared and (iii) an indicator of whether or not the program has been in effect for more than four months. The rate of uninsured drivers is measured by UMBI/BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. All columns include zip code, firm and vehicle fixed effects as well as a county specific time trend. Standard errors are in parentheses and are clustered at the level of the zip code cluster.

| | Instrument | Instrument | Wave | Year Trend | Ever Active | Active 6M | Inst. | X Year |
|---|---|---|---|---|---|---|---|---|
| insured Drivers | 27.82*** | 27.22*** | 30.05*** | 34.08*** | 29.25*** | 29.34*** | 29.34*** | 26.32*** |
| | (1.864) | (1.612) | (7.497) | (5.284) | (5.834) | (5.821) | (5.822) | (5.256) |
| ntrols | | √ | √ | √ | √ | √ | √ | √ |
| o Code Fixed Effects | √ | √ | √ | √ | √ | √ | √ | √ |
| ar Fixed Effects | √ | √ | √ | √ | √ | √ | √ | √ |
| nicle Fixed Effects | √ | √ | √ | √ | √ | √ | √ | √ |
| m Fixed Effects | √ | √ | √ | √ | √ | √ | √ | √ |
| servations | 4,724,220 | 4,724,220 | 4,499,035 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 |

$< .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. We use Limited on Maximum Likelihood to estimate results in column 1 to 2. We drop the final wave counties in column 3. In column 4, county-specific ls are used as extra controls. The rate of uninsured drivers, which is measured between 0 and 100, is instrumented using (i) the average f months during which the CLCA program was in effect in a zip code cluster (ii) the average number of months during which the CLCA was in effect in a zip code cluster squared and (iii) an indicator of whether or not the program has been in effect for more than four all columns except 5 to 7. We replace the instrument (iii) with an indicator of whether or not the program was active at all in column the indicator of whether or not the program was active for over six months instead in column 6. We dropped the instrument (iii) in . We include controls for the county population interacted with year fixed effects in column 8. The rate of uninsured drivers is measured /BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. Each ion includes the accident rate, driving history variables, age, plan type and gender. All specifications include zip code, year, firm and xed effects. Standard errors are in parentheses and are clustered at the level of the zip code cluster.

# A Estimating the Rate of Uninsured Drivers

## A.1 Methodology

From 1996 to 2004 the California Department of Insurance collected data on the number of registered vehicles in California by county, as well as the rate of uninsured motorists. The numbers were based on DMV Currently Registered Vehicles by zip codes up until July 1st for a given year. However, due to budgetary constraints the data collection was discontinued after 2004. Despite this inconvenience, other data was collected by the Department of Insurance which allows us to estimate the rate of uninsured motorists by county and zip code using standard methods. Our discussion draws heavily from Khazzoom (1999). This method is described by the Insurance Research Council (1999) and used for estimates of uninsured motorists in Smith and Wright (1992) and Cohen and Dehejia (2004).

The California Department of Insurance collects data on the number of exposures for bodily injury in car accidents (BI), as well as the number of claims for bodily injuries by uninsured motorists (UMBI). The data is reported based on where the car is garaged, giving us a method to estimate the rate of uninsured motorists per zip code area. We also have corresponding data by county. In order to construct estimates of the rate of uninsured motorists by zip code, we must make two implicit assumptions. First, we must assume that the probability of being involved in an accident is the same for both insured and uninsured motorists. Two, in a collision between an insured and uninsured motorist, both parties are equally likely to be found at fault. Hunstad (1999) provides a further discussion of using BI and UMBI data to estimate the rate of uninsured motorists. Formally, we can write

**A1** Let $p_I$ be the accident rate of insured motorists, and let $p_U$ be the accident rate of uninsured motorists. Then $p_I = p_U$.

**A2** In accidents between an insured and an uninsured motorist, each party is equally likely to be at fault.

The proportion of uninsured motorists is estimated in two steps:

1. We assume that both insured and uninsured motorists have an at-fault injury accident rate of $p$. Then $pX = BI$ accidents involving bodily injury are caused by the group of insured motorists, and consequently if there were $p.Y$ total accidents, $pY - pX$ accidents were caused by uninsured motorists.

2. Given that uninsured and insured motorists are equally likely to be at fault in an accident and the same propensity to cause an accident, we have that $\frac{X}{Y}$ of the $pY - pX$ accidents

caused by uninsured motorists are with insured motorists. Thus there will be $\frac{X}{Y}(pY - pX) = UMBI$ uninsured motorist claims filed by the insured motorists.

The total number of motorists is the number of insured combined with the uninsured, so $Y = X + U$ where $U$ is the number of uninsured motorists. Thus after some algebra we have that $\frac{UMBI}{BI} = \frac{U}{Y}$, so UMBI/BI is the same as the proportion of uninsured motorists in the population of all motorists. It is then straightforward to compute the number of uninsured motorists given the total number of registered motorists for each county which we do using data from the California Department of Insurance.

In fact, we can also derive that UMBI/BI is a good theoretical proxy for the rate of uninsured drivers under the aforementioned two assumptions and the two-car accident framework of Levitt and Porter (2001). We denote $N_I$ and $N_U$ to be the number of insured and uninsured drivers respectively. As in equations 5 to 7 of Levitt and Porter (2001), we define the probability of accident between two insured drivers to be $P_{II} = \frac{PN_IN_I}{K}$, the probability of accident between an insured and an uninsured to be $P_{UI} = \frac{2PN_UN_I}{K}$ and the probability of accident between two uninsured drivers to be $P_{UU} = \frac{PN_UN_U}{K}$, where the $K$ is such that the three probabilities add up to one. For the total number of two-car accidents $T$, we have the number of bodily injury (BI) claims for crashes where the insured drivers are at fault to be $BI = T(P_{II} + 0.5P_{UI}) = T(\frac{PN_IN_I}{K} + \frac{PN_UN_I}{K})$. At the same time, the uninsured BI claims submitted by insured drivers where the uninsured drivers are at fault is $UMBI = T(0.5P_{UI} = T(\frac{PN_UN_I}{K})$. Therefore, in the Levitt and Porter (2001) framework, we also arrive at $\frac{UMBI}{BI} = \frac{N_U}{N_I+N_U}$.

Finally, given that we have individual zip codes, we geocode each location and compute for each zip code the average rate of uninsured motorists for zip codes within a 25 mile (40km) radius. Since the distances involved are relatively small, we use a standard equirectangular approximation to compute distance $distance = R\sqrt{(lon_2 - lon_1)^2 Cos(\frac{lat_1+lat_2}{2})^2 + (lat_2 - lat_1)^2}$, where $R$ is the Earth's radius, $lon$ is the longitude and $lat$ is the latitude. We record the CLCA program as not being in effect if the average number of months is below a third of the year.

## A.2 Points of Concern

It is important to note several points of concern in relation to this methodology. First, the model rules out multiple ownership of vehicles in which some of the drivers are insured and others are uninsured. Furthermore, if our assumption that the rate of accidents is the same between insured and uninsured drivers is violated, and if in fact uninsured drivers have a higher or lower accident rate, then our estimates of uninsured drivers will be biased upwards or downwards respectively. In this case our estimates can be viewed respectively as lower or upper bounds for the true effect of uninsured motorists on insurance premia. However, the IRC found no evidence that uninsured drivers have higher accident rates than insured drivers. There are also concerns that the measure of uninsured motorists is biased upwards as UMBI claims will include injuries caused by drivers of stolen vehicles, as well as injuries caused by hit and run accidents. Cohen and Dehejia (2004) find evidence for moral hazard in their study using state laws on compulsory insurance to instrument for the endogeneity of uninsured motorists. Their estimates, while marginally significant, are small. These effects may also bias the results downwards, as uninsured drivers may have a lower accident rate due to concerns about payments for damage. These issues are dealt with in section 6 of the paper. Regardless, even if our results are biased due to uninsured motorists have higher or lower accident rates, UMBI/BI should still be an excellent proxy for the rate of uninsured motorists in a county. Furthermore using an instrumental variables approach addresses the problem of measurement error, alleviating many of the concerns regarding the use of UMBI/BI as a measure of uninsured motorists.

One concern that is often voiced regarding using collision data to estimate rates of uninsured motorists is that uninsured drivers may have higher accident rates than other drivers as many uninsured drivers in California may be illegal immigrants and hence unfamiliar with driving in the United States. This concern is largely unwarranted. While California does have a quarter of the nation's illegal immigrants, with the Department of Homeland Security estimating slightly less than 3 million illegals living in California in 2006, they account for less than 10% of state's population. In 2006, there were roughly 4.6 million uninsured motorists in California, so if all illegal immigrants drove and all were uninsured, illegal immigrants could potentially make up a high fraction of uninsured motorists. There are several reasons to believe that concerns are exaggerated and largely irrelevant. First, illegal immigrants have been able to obtain auto insurance in California since 2003. Thus for the entire sample period over which we have data, illegal immigrants have been able to obtain auto insurance. Second, given very real concerns regarding the threat of deportation after encounters with police that lead to a revelation of undocumented status, one could easily claim that illegal immigrants would be more likely to purchase auto insurance than the general population. Third, the CLCA advertising programs included advertising in several languages other than English, including Spanish. Thus we would expect similar effects from the CLCA program on illegal immigrants. Fourth, illegal immigrants being generally lower income than the general population, they are likely to have low rates of vehicle ownership. Finally, a 2008 study by Utah's Office of the Legislative Auditor General found that rates of insurance are nearly the same between illegal immigrants and the general populace, being 76% and 82% respectively. The Utah Office of the Legislative Auditor General sampled a group of 3,461 holders of driving privilege cards and matched them to vehicle insurance policies, and then sampled a similar number of driving licenses. Only 1.7% of holders of driving privilege cards in Utah had a legal presence in the US. The Utah evidence leads us to believe that the presence of illegal immigrants in California, and potential differences in accident rates in this group, would not significantly bias our results.

## A.3 Magnitude of the Measurement Error

The difference in the coefficient estimates between FE and IV specifications may appear large. Directionally, measurement error in the uninsured driver's rate is able to attenuate our FE estimates and bias the linear least squares result towards zero. For a simpler case of uninvariate panel regression with uninsured driver's rate being the only independent variable, the ratio of the estimated-to-true coefficient is proportional to $\frac{\sigma^2_{\triangle u}}{\sigma^2_{\triangle u}+\sigma^2_{\triangle \mu}}$ where $\sigma^2_{\triangle u}$ is the variance of the *first-differenced* true uninsured rate and $\sigma^2_{\triangle \mu}$ is the variance of the *first-differenced* measurement error. To explain the observed 8-to-1 difference between our FE and IV estimates, the ratio of $\sigma^2_{\triangle u}$ to $\sigma^2_{\triangle \mu}$ should be 1-to-7. However, this condition is different from requiring the ratio of the variance of the true uninsured driver's rate $\sigma^2_u$ over the variance of the measurement error $\sigma^2_\mu$ to be 1-to-7. While on first sight this number looks implausibly large, it is consistent with relatively high serial correlation of the true uninsured driver's rate and the measurement error being serially uncorrelated white noise. Intuitively, high serial correlation of the true uninsured driver rate will make $\sigma^2$ small while serially uncorrelated measurement error will make $\sigma^2_\mu$ two times the variance of the measurement error.

In fact, we can gauge the magnitude of the serial correlation of the true measurement error and the ratio of $\sigma^2_\mu$ over $\sigma^2$ by assuming the measurement error is a classical measurement error: ie, $\mu_t \perp\!\!\!\perp u_s$ for any $s$ and serially uncorrelated white noise: ie, $\mu_t \perp\!\!\!\perp \mu_{t-1}$. We have,

$$\frac{\sigma^2_{\triangle \mu}}{\sigma^2_{\triangle u}} = \frac{2\sigma^2_\mu}{2\sigma^2_u - 2cov(u_t, u_{t-1})} = 7$$

we also calculate in our sample the correlation coefficient of our measured uninsured driver's rate with its one-period lag to be about 0.5, that is,

$$\frac{cov(u_t + \mu_t, u_{t-1} + \mu_{t-1})}{var(u_t + \mu_t)} = \frac{cov(u_t, u_{t-1})}{\sigma^2_\mu + \sigma^2_u} = \frac{1}{2}$$

From these two equations, we can solve that, $\sigma^2_\mu = \frac{9}{7}\sigma^2_u$ and $cov(u_t, u_{t-1}) = \frac{8}{9}\sigma^2_u$ , which means the correlation coefficient of the true uninsured driver's rate and its one-period lag is very high, at 0.89. The implied ratio of the variance of the true uninsured rate and the variance of the measurement error is indeed quite plausible.

# B Pigouvian Taxation

Let agents be endowed with a standard utility function $U(.)$, which is concave, continuous, and increasing. Let $\lambda$ be the proportion of uninsured motorists and all other parameters are as defined in section 2. Agent $i \in I$ follows a distribution of types $F(.)$ which determines wealth $w_i$, accident probability $\pi_i$ and losses $L^s_i$ are stochastic. The government levies a fine $\tau$ on the uninsured and returns a subsidy $s$ to all individuals. This setup largely mimics current policies in most US states. Agents who purchase insurance a portion $q_i \in [0,1]$ of their wealth from a representative firm at price $p_i$ consume $c^i_{na} = q_i[w_i - p_i + s] + (1-q_i)[w_i - \tau + s]$ in the event that they do not cause an accident with an insured motorist, and $c^i_{ai} = q_i[w_i - L^s_i - p_i + s] + (1-q_i)max\{w_i - 2L^s_i - \tau + s, -\tau + s\}$ in the event of an at fault accident, and $c^i_{au} = q_i[w_i - p_i + s] + (1-q_i)min\{w_i - L^s_i + R_i - \tau + s, w_i - \tau + s\}$ in the event of an accident with an uninsured motorist who is at fault. Agents optimize and decide how much insurance $q_i$ to purchase, and the sum of these decisions determines the rate of uninsured motorists, in other words $1 - \lambda(\tau) = \frac{\sum_{i\in I} q_i}{N}$. The government can levy a fine or tax on uninsured motorists, which we denote $\tau$. Higher fines discourage motorists from driving without insurance, so we can write $\lambda(\tau)$ with $\lambda'(\tau) < 0$. The agent's problem at time 0 is to choose the optimal level of insurance $q_i$ such that

$$\max_{0\le q_i\le 1} (1 - \frac{\pi_i}{2} - \frac{\pi_i\lambda}{2})\mathbb{E}[U(c^i_{na})] + \frac{\pi_i}{2}\mathbb{E}[U(c^i_{ai})] + \frac{\pi_i\lambda}{2}\mathbb{E}[U(c_{au})]$$

$$s.t. \qquad c^i_{na} \le q_i[w_i - p_i + s] + (1-q_i)[w_i - \tau + s]$$

$$c^i_{ai} \le q_i[w_i - L^s_i - p_i + s] - (1-q_i)[\tau - s] + (1-q_i)max\{w_i - 2L^s_i, 0\}$$

$$c^i_{au} \le q_i[w_i - p_i + s] - (1-q_i)[\tau - s] + (1-q_i)min\{w_i - L^s_i + R_i, w_i\}$$

The externality is captured entirely in the $p_i$ term. For reasons discussed in the main text of the paper, uninsured drivers cause premia to be raised for other drivers. We can denote the solution to the optimization problem for a given subsidy $s$ and tax $\tau$ by $V(s,\tau)$. We assume a benevolent government which levies a fine (or tax) $\tau$ on uninsured drivers and redistributes a subsidy $s$ to all individuals. We assume that the government redistributes from the uninsured to the general populace, and uses revenues from the fine to finance the subsidy. Thus the government's budget constraint is $s = \lambda(\tau)\tau$. The government solves the following problem, maximizing utility for a representative consumer who purchases insurance at the rate $q = 1 - \lambda(\tau)$ determined by the consumer's optimization problem. We assume that the wealth effects of the tax are insignificant.

$$max_\tau \, V(s,\tau)$$

$$s.t. \; s = \lambda(\tau)\tau$$

At an interior optimum, the optimal tax must satisfy

$$\frac{dV(s,\tau^*)}{d\tau(\tau^*)} = 0.$$

We note that we can write $V(s,\tau)$ as

$$V(s,\tau) = \max_{q,c_{na},c_{ai},c_{au},\mu_1,\mu_2,\mu_3} (1-\frac{\pi}{2}-\frac{\pi\lambda}{2})\mathbb{E}[U(c_{na})] + \frac{\pi}{2}\mathbb{E}[U(c_{ai})] + \frac{\pi\lambda}{2}\mathbb{E}[U(c_{au})]$$

$$\mu_1[w-(1-q)(\tau-s)-(p-s)q-c_{na}]+\mu_2[q(w-L^s-p+s)-(1-q)(\tau-s)+(1-q)max\{w-2L^s,0\}-c_{ai}]+$$

$$\mu_3[q(w-p+s)-(1-q)(\tau-s)+(1-q)min\{w-L^s+R,w\}-c_{au}]$$

Where $\mu_i$ is the Lagrange multiplier, the marginal value of relaxing the budget constraint. Since we have already optimized the function over $\{q, c_{na}, c_{ai}, c_{au}, \mu_1, \mu_2, \mu_3\}$, changes in these variables do not have any first order effects on $V(s,\tau)$, by the envelope theorem (see the mathematical appendix of Mas-Colell et al. (1995) for a proof). We note that $1-\lambda(\tau) = \frac{\sum_{i\in I} q_i}{N} = q$. We thus have the following necessary condition:

$$\frac{dV(s,\tau^*)}{d\tau(\tau^*)} = (\mu_1+\mu_2+\mu_3)q(1-\frac{dp}{d\tau}) + (\mu_1+\mu_2+\mu_3)(\frac{ds}{d\tau}-1).$$

Given the first order condition we have the following optimality condition

$$1-\frac{dp}{d\tau} = (1-\frac{ds}{d\tau})\frac{1}{1-\lambda(\tau)}.$$

Recall that $p = \beta\lambda(\tau)+k$ and rearrange the government's budget constraint to obtain $s = \lambda(\tau)\tau$. We thus have

$$\frac{ds}{d\tau} = \lambda'(\tau)\tau + \lambda(\tau).$$

Substituting above we have

$$1-\beta\lambda'(\tau) = (1-\lambda'(\tau)\tau-\lambda(\tau))\frac{1}{1-\lambda(\tau)}.$$

We can thus rearrange the preceding equation to have

$$1-\beta\lambda'(\tau) = 1-\frac{\lambda'(\tau)\tau}{1-\lambda(\tau)}$$

$$\tau^* = \beta(1-\lambda(\tau)).$$

It is evident that the optimal tax is increasing in $\beta$, the size of the externality, and decreasing in $\lambda$, the rate of uninsured drivers.

# C Institutional Background and Data

## C.1 The CLCA Program

California law requires that all drivers in the state maintain a liability only insurance policy that covers up to $15,000 in damage. Violating this mandate can lead to fine of between $100-500 depending on the number of offenses. However, the fine is only enforced if individuals are cited and a police officer asks to see proof of insurance. Given the low probability of being caught driving without insurance, roughly 20% of California drivers choose not to purchase liability insurance. As explained in the main text, this decision can lead to insurance premia rising for other drivers.

The California Low Cost Automobile Insurance (CLCA) Program was created by the state legislature in 1999 under California Insurance Code Section 11629.7 to provide low cost liability automobile insurance for low income persons at affordable rates. The program initially began as a pilot program in Los Angeles and San Francisco counties, and gradually by December 2007 expanded to all 58 counties in California. The program was initially set to sunset on January 1, 2011. The California state assembly passed a bill to extend the program, however Governor Schwarzenegger vetoed the bill due to perceived costs. Eventually the program was extended. The program's expansion was based on the determination of need, which was defined as the absolute number of uninsured drivers in a county. Effectively this meant that the roll out of the program was determined by population size of counties. See the main text for additional information regarding this point.

The roll out of the program in a county was accompanied by a media campaign financed by a special purpose assessment on each vehicle insured in the state. The media campaign was targeted at lower income and minority individuals. Brochures and advertisements were created in several languages. The media outreach program resulted in 151,924,800 impressions in newspapers, the internet and on television by 2011. In 2011, 66,375 individuals were assigned to the CLCA program.

The CLCA program provides insured individuals with coverage of $10,000 for bodily injury for an individual person in an accident, and up to $20,000 for all individuals injured in an accident. In addition $3,000 is provided for property damage as a result of an accident. Roughly 60% of those enrolled in the program were previously uninsured.

The incidence of the program primarily is designed to fall on insurance companies. Individuals apply to the CLCA program, and are assigned to insurance companies at the county level based on the insurance company's market share in a particular county. Individuals are charged rates that are determined by the California department of insurance. Individuals are able to apply online or use a paper application. There is anecdotal evidence from the department of insurance that insurance companies took advantage of the media campaign to enroll individuals in their own low cost automobile insurance policies.

The rates of the CLCA program varied by county each year. The rates are administered by the California Automobile Assigned Risk Plan (CAARP) commissioner. By law, the rates are required to cover losses incurred and expenses, costs of administration, underwriting, taxes, commissions, and claims adjusting in each county. Moreover the rates are required to be set so that there is no projected subsidy between counties. The rates are set so that insurance companies break even in each county, with the goal of the CLCA program not being a disincentive to insurance companies operating in lower income counties. In 2007 rates ranged from $222 per year in Tulare county to $354 per year in Stanislaus county. Young men under the age of 25 must pay an additional 25% surcharge. Rates typically do not vary much from year to year, however the stipulation that rates only cover losses leads to sharp spikes or drops in the premium of up to 25% percent was enacted in several years for some counties. These likely reflect stochastic shocks which increased accident rates and drove up premia.

Eligibility for the program was determined by several criteria. First, individuals had to have an annual household income below 250% of the federal poverty line. Second, vehicle values had to be below a threshold, which was $20,000 in 2007. Third, an individual must have fewer than two at fault accidents or moving violations. Fourth, an individual must have had a license for at least three years, be at least 19 years old and be a resident of the State of California.

## C.2 The Automobile Insurance Premium Survey

The main data source employed in this study is the California Department of Insurance Automobile Insurance Premium Survey. Following 1990, the California Department of Insurance was required to collect data on automobile insurance premia. The California Department of Insurance surveys licensed insurers in accordance with the California Insurance Code sections 12959, 10234.6 and 10192.2. The goal of this annual survey is to inform consumers about the premia charged by different companies. The results of the survey are published online and consumers are able to access premium data via an online tool. We obtained full data from 2003 to 2010, excluding 2008 when the survey was not conducted for administrative reasons. We also have partial survey results prior to 2003. In the years in our sample, the demographic hypothetical risk profiles do not change significantly across zip codes and over time.

By law, licensed automobile insurers are required to submit their pricing formulas to the California Department of Insurance. Insurance companies must base the premiums they charge on the rates filed with the Department of Insurance. The pricing formula is heavily regulated in the State of California. Following Proposition 103 in 1988, automobile insurers are required to price on three factors: driving record, annual miles driven, and years licensed. In addition, companies may price from a list of optional factors that can vary from year to year. The optional factors typically include vehicle type and performance, number of vehicles, use of vehicle, gender, marital status, age, characteristics of secondary drivers, persistency, location (zip code), academic standing of students in the household, completion of driver training course, smoking, bundling products with same company, as well as claims frequency and severity.

More than 50 companies are surveyed in each zip code each year, and there are over 1 million premium quotes in the data each year. The companies must give price quotes for each hypothetical risk profile in the APS. Each observation in the data represents an offer price for consumers with particular observable demographics from a firm operating in a particular zip code. The survey oversampled individuals with speeding tickets and at fault accidents, leading to a higher average premium in comparison to the general populace. If insurers do not respond to the survey, they are fined heavily. The database also contained data on National Association of Insurance Commissioner (NAIC) codes of insurers as well as data on vehicle make and year, which we matched to vehicle value using pricing information. The survey collected data on two types of plans in zip codes, a basic plan and a standard coverage plan for different demographics. The basic plan represents a plan just above the minimum required threshold for coverage in California, while the standard plan was deemed by the Department of Insurance to be the most common automobile insurance plan in California.

Table A1: Main Results: Collapsed to Zip Code

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Uninsured Drivers | 19.83*** | 19.23*** | 23.13*** | 23.10*** | 23.19*** | 23.19*** |
| | (4.453) | (4.568) | (4.444) | (4.445) | (4.441) | (4.442) |
| Accident Rate | 291.0* | 291.0* | 228.3 | 228.5 | 228.0 | 228.0 |
| | (160.1) | (160.1) | (149.0) | (149.0) | (148.9) | (148.9) |
| Year Effects | NO | TREND | YES | YES | YES | YES |
| Zip Code Effects | YES | YES | YES | YES | YES | YES |
| Instrument Dropped | | | | Indicator | Months$^2$ | Months |
| Observations | 995 | 995 | 995 | 995 | 995 | 995 |

Notes:* $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. The rate of uninsured drivers is measured between 0 and 100. In column 1 to 3, the rate of uninsured drivers is instrumented using (i) the average number of months during which the CLCA program was in effect in a zip code (ii) the average number of months during which the CLCA program was in effect in a zip code squared and (iii) an indicator of whether or not the program has been in effect for more than four months. The rate of uninsured drivers is measured by UMBI/BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. Zip code fixed effects, year fixed effects, or a linear time trend are noted below the point estimates. In columns (4)-(6) one of the instruments in dropped. The dropped instrument is noted below the fixed effects. Standard errors are in parentheses and are clustered at the level of the zip code.

Table A2: Main Results: IV Full

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Premium | Premium | Premium | Premium | Premium | Premium |
| Uninsured Drivers | 25.22*** | 25.33*** | 25.50*** | 25.58*** | 40.54*** | 29.30*** |
| | (4.872) | (4.884) | (4.903) | (4.909) | (7.689) | (5.824) |
| Accident Rate | 257.7 | 259.3 | 259.9 | 260.5 | 360.3* | 283.7* |
| | (167.3) | (168.1) | (168.4) | (168.8) | (208.6) | (170.4) |
| Standard | 1799.4*** | 1789.7*** | 1766.8*** | 1776.6*** | 1665.3*** | 1617.8*** |
| | (24.48) | (22.27) | (23.85) | (24.30) | (67.04) | (49.75) |
| At Fault Accident | | 734.5*** | 734.0*** | 777.4*** | 796.9*** | 790.0*** |
| | | (11.10) | (11.11) | (11.74) | (13.48) | (13.52) |
| Female | | | -151.9*** | -155.6*** | -139.2*** | -148.8*** |
| | | | (18.85) | (18.91) | (6.719) | (6.648) |
| Speeding Ticket | | | | 557.2*** | 567.0*** | 558.9*** |
| | | | | (11.40) | (9.021) | (8.682) |
| Age | | | | | -39.81*** | -43.14*** |
| | | | | | (3.538) | (3.493) |
| Daily Miles | | | | | | 43.02*** |
| | | | | | | (2.785) |
| Observations | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 |

Notes:* $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. The rate of uninsured drivers is measured between 0 and 100. In all estimates the rate of uninsured drivers is instrumented using (i) the average number of months during which the CLCA program was in effect in a zip code cluster (ii) the average number of months during which the CLCA program was in effect in a zip code cluster squared and (iii) an indicator of whether or not the program has been in effect for more than four months. The rate of uninsured drivers is measured by UMBI/BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. All columns include zip code, firm,year and vehicle fixed effects. Standard errors are in parentheses and are clustered at the level of the zip code cluster.

Table A3: Main Results with County Year Trend

| | (1) Premium | (2) Premium | (3) Premium | (4) Premium | (5) Premium | (6) Premium |
|---|---|---|---|---|---|---|
| Uninsured Drivers | 29.42*** | 18.33*** | 20.96*** | 28.66*** | 47.17*** | 34.08*** |
| | (4.870) | (3.326) | (3.688) | (4.854) | (7.361) | (5.284) |
| Accident Rate | 84.97 | 19.35 | 39.07 | 102.6 | 214.7 | 134.8 |
| | (149.63) | (120.7) | (127.1) | (148.4) | (205.3) | (151.5) |
| Standard | 1641.2*** | 1671.4*** | 1695.0*** | 1550.1*** | 1602.1*** | 1559.5*** |
| | (20.68) | (19.84) | (25.16) | (70.66) | (70.16) | (77.63) |
| At Fault Accident | | 738.1*** | 737.3*** | 777.0*** | 789.7*** | 786.6*** |
| | | (14.43) | (14.31) | (17.89) | (17.86) | (17.92) |
| Female | | | -166.2*** | –161.9*** | -150.3*** | -158.2*** |
| | | | (18.86) | (23.93) | (22.91) | (22.10) |
| Speeding Ticket | | | | 563.2*** | 568.9*** | 564.6*** |
| | | | | (10.94) | (10.99) | (12.18) |
| Age | | | | | -41.76*** | -43.71*** |
| | | | | | (1.137) | (1.151) |
| Daily Miles | | | | | | 41.67*** |
| | | | | | | (5.203) |
| Observations | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 | 4,724,220 |

Notes:* $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. The rate of uninsured drivers is measured between 0 and 100. In all estimates the rate of uninsured drivers is instrumented using (i) the average number of months during which the CLCA program was in effect in a zip code cluster (ii) the average number of months during which the CLCA program was in effect in a zip code cluster squared and (iii) an indicator of whether or not the program has been in effect for more than four months. The rate of uninsured drivers is measured by UMBI/BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. All columns include zip code, firm and vehicle fixed effects as well as a county specific time trend. Standard errors are in parentheses and are clustered at the level of the zip code cluster.

Table A4: Main Results Using Only One Instrument

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Uninsured Drivers | 38.88* | 38.88* | 39.85* | 39.93* | 39.62* | 39.71* |
| | (23.61) | (23.61) | (23.52) | (23.48) | (23.61) | (23.61) |
| Accident Rate | 176.2** | 176.2** | 173.0** | 173.0** | 173.8** | 173.7** |
| | (87.16) | (87.16) | (86.92) | (86.72) | (87.23) | (87.21) |
| At Fault Accident | 736.4*** | 736.4*** | 737.0*** | 777.0*** | 774.5*** | 774.5*** |
| | (6.04) | (6.04) | (6.04) | (6.22) | (6.19) | (6.19) |
| Standard | | 1021.4*** | 1679.7*** | 1890.8*** | 1804.5*** | 1800.2*** |
| | | (22.37) | (22.59) | (22.90) | (22.90) | (22.89) |
| Age | | | -45.60*** | -45.99*** | -276.7*** | -286.7*** |
| | | | (0.42) | (0.42) | (3.61) | (3.58) |
| Speeding Ticket | | | | 553.6*** | 551.9*** | 551.8*** |
| | | | | (4.39) | (4.38) | (4.37) |
| Age$^2$ | | | | | 3.53*** | 3.67*** |
| | | | | | (0.05) | (0.05) |
| Daily Miles | | | | | | 42.86*** |
| | | | | | | (0.36) |
| Observations | 4,723,816 | 4,723,816 | 4,723,816 | 4,723,816 | 4,723,816 | 4,723,816 |

Notes:* $p < .1$, ** $p < .05$, *** $p < .01$. The dependent variable in all columns is the real premium quote offered by a firm. The rate of uninsured drivers is measured between 0 and 100. In all estimates the rate of uninsured drivers is instrumented using the average number of months during which the CLCA program was in effect in a zip code cluster . The rate of uninsured drivers is measured by UMBI/BI. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. All columns include zip code, firm and vehicle fixed effects as well as a county specific time trend. Standard errors are in parentheses and are clustered at the level of the zip code cluster.