The Real Effects of the Uninsured on Premia

Yannelis, Constantine and Sun, Stephen Teng

Stanford University, Stanford University

20 May 2013

The Real Effects of the Uninsured on Premia *

Stephen Teng Sun[†] Constantine Yannelis[‡]

May 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 between 1-2%. We also discuss corrective Pigouvian taxes.

*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 and the Midwest Macro conference at 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 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. 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.[1] 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 interest behind this externality, for example see Smith and Wright (1992) and Keeton and Kwerel (1984), 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 increases premia by roughly 1 percent.

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 there exist various state and federal laws mandating insurance coverage under penalties of a fine or tax, which are presumably designed to internalize insurance externalities. Unfortunately, estimating the size of the effect of the uninsured on premia 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 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 in either health or automobile insurance markets has been lacking due to the aforementioned problem. Our

[1]Most insurance policies sold in the US come with an uninsured motorist coverage. Department of Insurance data indicates that in 2008 in California 84.38% of policies came with uninsured motorist coverage.

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 a 1-2 percentage point decrease in the rate of uninsured drivers. The sequential rollout of the program makes it possible to obtain a credible identification of the causal effects of the uninsured motorist rate. We argue that the CLCA program can generate valid instrumental variables for the rate of uninsured drivers.

In order to accomplish this, we compiled a novel panel data for the 58 counties in the state of California for years from 2003 to 2007.[2] Our main data set consists of insurance premium quotes collected by the California Department of Insurance from most 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 consumers with particular observable demographics from a firm 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.[3] Controlling for year and zip code fixed effects can absorb many environmental factors since auto insurance companies typically price at the zip code level. We exploit this geographic variation to obtain estimates for the average effect of uninsured drivers on insurance premia.

The use of policy-driven variation in the prevalence of uninsurance along with new administrative data on insurance premia leads us to conclude that uninsured drivers raise premia for other drivers, as predicted by theory. Specifically, we find that a 1 percentage point increase in the share of drivers who are uninsured leads to a 1-2 percent rise in premia. To illustrate, this implies that consumers could save about $500 annually if the county with the highest uninsured drivers rate, 29% in San Joaquin, sees its uninsured drivers rate fall to that of the county with lowest uninsured drivers rate, 9% in Mono.

[2]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.

[3]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.

We also discuss the optimal corrective Pigouvian tax on uninsured drivers. 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.

Alternative explanations for our results are examined and rejected. Other phenomena, such as the introduction of the CLCA program inducing insurance companies to lower prices to compete with subsidized plans, or unobserved selection on accident risk could potentially explain our results. The structure of the CLCA programs allows us to rule out such alternative explanations. We are able to test these alternative hypotheses by restricting our sample to individuals ineligible for the CLCA program, and reject these explanations for the observed effects following the introduction of the CLCA program.

The paper is organized as follows. Section 2 presents a concise motivating model based on prior literature. Section 3 discusses and motivates our estimation strategy, explaining how we use a policy change to overcome the endogeneity problem. Section 4 describes the data, which to our knowledge has not been used in the economics literature. 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 various robustness checks and rules out alternative explanations for our results such as competition and selection. 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 drivers. It is precisely this endogeneity which creates difficulties in estimating the effect of uninsured drivers on premia. We present a concise model of how insurers determine automobile insurance premia which draws heavily from Smith and Wright (1992) and Keeton and Kwerel (1984). 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.[4] 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.

More formally, we can define an individual i with wealth w_i and probability of being involved in an accident π_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. 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,[5] we assume that an individual has an equal probability of being found at fault or not at fault in an accident. 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[6]. This event occurs with probability $\frac{\pi_i}{2}$. Thus an insured driver has expected utility, assuming a utility function $U(.)$ with standard properties:

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

Let λ be the fraction of uninsured motorists in a market, and note that λ 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

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

[5] With the notable exception of moral hazard. We discuss the literature on moral hazard in section 6, which has mixed results.

[6] 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.

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 λ, 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 can assume a representative risk-neutral firm in a competitive insurance market and we can 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 thus weakly increasing in λ, 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 λ 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 λ on p_{ij}, since λ will be significantly correlated with the error term in any regression.[7] Separating the effect of uninsured drivers on insurance premia from drivers choosing not to insure due to

[7]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.

otherwise high premia presents a challenge to the researcher. In the next section, we discuss how we can overcome the endogeneity problem and estimate the true effect of uninsured drivers on insurance premia.

3 Empirical Strategy

3.1 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 zip code in California. 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.[8] 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 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 areas of the relevant 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.[9] 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.[10] 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

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

[9] See Schultz and Yarber (2011).

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

accident rate:[11] 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.[12] We do not observe income, as it is illegal in California for automobile insurers to price 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 insures 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.

3.2 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. Column 1 of Table 2 indicates that the introduction of the CLCA program was associated with a roughly two percentage point drop in the rate of uninsured drivers. Column 6 of table 2 also indicates that uninsured motorist claims fell by almost 10% following the introduction of the CLCA program. Both of these effects are significant at the .01 level. The second assumption is that the instrumental variables are orthogonal to unobserved

[11] 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.

[12] See appendix D for a further discussion on eligibility and the CLCA program in general.

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 should be 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 obtain several instruments from the CLCA program. First, we use the average number of months during the year in which the CLCA program was active. Second, we use the rates set by the commissioner at which participants in the program can purchase liability coverage. Finally we also include variants of our first instrument, the average number of months that the CLCA program is in effect squared and an interaction between the CLCA program being in effect and being a high uninsured zip code. 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 wanted to 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. 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.[13] 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.

[13]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.

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 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. In the same spirit, we can also exploit the heterogenous effects of the CLCA program. We would expect the CLCA program to be more effective in areas with large numbers of uninsured drivers, so we can include an interaction between the average number of months the CLCA program is in effect and being a high uninsured zip code cluster.[14]

Our other instrument is the CLCA rates set by the commissioner at the county level, which vary by county from year to year depending on the previous year's loss experience, dropping by as much as 25% in some years. The CLCA program is essentially a burden on insurance companies, with firms being assigned low-income participants based on their market share in the voluntary market. The commissioner then varies the rate, by law, only based on the previous year's loss experience, which is stochastic. While the CLCA rates generally remain stable, in some years the rates jump or drop substantially, likely reflecting abnormal loss experience in specific counties due to shocks such as pile up accidents or beneficial and adverse weather conditions. These random events would shock loss experience and would translate into higher CLCA rates in the following year. The rates are valid instruments if they affect the share of drivers who are uninsured (an effect we can verify) and are orthogonal to market premia. The latter condition is plausible, however this condition is not as obvious in the case of the CLCA rate instrument. Therefore, we employ Hausman-type econometric tests later in the paper to demonstrate that our CLCA rate and interaction instruments are valid, under the assumption that our first instrument is valid, the number of months that the CLCA program was in effect.

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):

[14] More than 25% rate of uninsured drivers in the entire sample period.

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

where λ_{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 α_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 β, 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%.

4 Data

4.1 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[15] required that the California Department of Insurance collect data on insurance rates in the state. Following 1990, 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. The survey oversampled hypothetical drivers with speeding tickets and at fault accidents, leading to a higher average premium in comparison to the general populace. We obtained data from 2003 to 2010 excluding the year 2008.[16] 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.[17] 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).[18]

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, allowing 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.[19] 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,

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

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

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

[18] Dahlby and West (1986) offer costly consumer search in the sense of Stigler (1961) as a possible explanation for this phenomenon, testing predictions from the search model of Carlson and McAfee (1983).

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

while the standard plan was deemed by the Department of Insurance to be the most common automobile insurance plan in California. Table 1 summarizes the two types of private plans and the basic CLCA plan.[20]

One potential concern is that our results could be driven by compositional changes in the survey data. It is to note 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 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 at-fault accident as well as whether the driver has 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 cautions 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 has 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 at-fault automobile accidents in which an individual was involved in the three years prior to the survey.

[20]Unfortunately the plans do not vary deductible choice, otherwise we would be able estimate risk preference as in Cohen and Einav (2007). Also, the plans do not decompose specific parts of the premium. Thus we are unable separately examine the premium for uninsured motorist coverage and collision coverage, which should be the parts of the premium affected by the uninsured motorist problem.

4.2 Matched Data

The main APS survey data was matched to three other data sources, 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)[21]. 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[22]. Since premia were unadjusted for inflation, we collected data on the Consumer Price Index from the Bureau of Labor Statistics. We used the BLS December CPI of each year in our adjustments.

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.

[21]See Appendix B for more on estimating the rate of uninsured drivers. Our measure used 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.

[22]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.

5 Main Results

5.1 Estimates of the Externality

Table 4 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 0.05 level, indicating the rather nonsensical result more uninsured drivers reduce insurance premia. This is not surprising given that in these specifications 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 make controlling for zip code and other fixed effects critical. Indeed, when we control for zip code and year fixed effects in Table 4, columns (3)-(4), the coefficient on the rate of the uninsured changes its sign and becomes positive and statistically significant. However, 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.[23] 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. 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 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

[23] 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 $\hat{\beta}$ 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.

uninsured drivers. As reported in Table 4, columns (5)-(6), once instrumented for, the coefficient for the rate of uninsured drivers becomes higher in absolute value, with a positive value of \$28, 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 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 when we add various demographic and driving record controls, providing an additional test that our instrument is uncorrelated with these controls. Our 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.

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.[24] The sign and magnitude of other coefficients in the results presented in Table 4 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 also supported by adding in the number of years licensed to the specifications as controls. However, once we add a quadratic term to the regression specification, the coefficient on age squared is significant and positive suggesting that elderly drivers pay higher insurance premia. 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 support the basic theoretical underpinnings of Smith and Wright (1992), Keeton and Kwerel (1984) and Arrow (1963), namely that premia will also be increasing in accident rates and the inherent riskiness of a driver.

The final row in Table 4 shows results from a Hausman test. The validity of our instruments rests on two assumptions. The first assumption is that the instrumental variables, (1) the number of months that the CLCA program was in effect in a zip code cluster and a square and interaction term, and (2) the rate that an eligible individual had to pay for coverage under the CLCA program, are correlated with the rate of uninsured drivers. Table 2 presents evidence that this is indeed the case, and that the introduction of the CLCA program was associated with a roughly two percentage point decrease in the rate of unin-

[24]See section 6 for a discussion of moral hazard.

sured drivers. The second assumption is that the instrumental variables are exogenous to insurance premia. This second assumption cannot be directly tested, however we can provide some partial tests for the exogeneity of the instrumental variables following Wooldridge (2010). Our test, which is based on the Hausman (1983) test, assumes the exogeneity of the introduction of the CLCA program, and tests the hypothesis that the CLCA rates and interaction are exogenous. The test statistic is distributed χ^2 with two degrees of freedom. The results indicate that we cannot reject the null hypothesis that the CLCA rates and interaction are exogenous to insurance premia and thus valid instruments.

It is illustrative of the challenges in estimating the effect of uninsured drivers on premia to contrast the results of IV estimates in Table 4 with the OLS estimates presented in Table 4. 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 bias our coefficient towards zero. 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.

The magnitude of the results is not surprising if we consider how automobile insurance companies price and assess risk. An insurance company will be forced to pay damages in two scenarios, one, if the driver is involved in an accident and found at fault, and two, if the driver is involved in an accident with an uninsured driver. We have a rate λ of uninsured drivers and furthermore we can assume that (1) a driver is equally probable to be at fault or not at fault in an accident and (2) insured and uninsured drivers in expectation cause the same amount of loss. In California the rate of uninsured drivers is roughly 20%, so a 1 percentage point increase in the rate of uninsured drivers should increase the payouts that an insurance company faces by approximately 1%. Given that the average premium in our data is roughly $2,356,[25] and we estimate that a 1% increase in the rate of uninsured drivers increases premia by $28, the aforementioned logic is very much in line with our results. This suggests that insurance companies entirely pass on the damage caused by uninsured drivers to insurance premia, and perhaps that insurance companies recover very little in damages from uninsured drivers.

When aggregated over all insured drivers in California the social costs of the externality[26] are substantial. Based on our main specification, and uninsured motorists rates in California in 2007 as well as rates of uninsured motorist coverage,[27] the total cost of the externality to California is about $6 billion, which is substantial. 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.[28] 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 rate premia by $80 (£50) if the rate of uninsured motorists is 3.5%.

5.2 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,[29] as well as

[25] 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.

[26] 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.

[27] 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.

[28] We caution that our estimates are local.

[29] 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.

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,[30] 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[31].

There exists a long tradition since Pigou (1920) of economists advocating corrective taxes on externalities.[32] 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 τ if they are uninsured and redistribute a subsidy s to all drivers. However, 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 τ 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 can obtain the following that the optimal corrective tax depends only on β, and

[30] 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, 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.

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

[32] 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.

$\lambda(\tau)$ and is given by

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

See Appendix C for a detailed derivation of the formula, which follows Chetty (2006) in spirit. The optimal tax formula is simple and intuitive, depending on β, the amount 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 β , which is 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.[33] 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 5 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 driver per day and other controls. In all cases save one we cannot reject at the 5% level that the coefficients are the same as in our main specification, although in one specification the effect is significantly larger than in the main specification. In the main dataset, we restrict the sample to only observations where there is one driver on the insurance plan. Our main specifications are robust to dropping observations above a threshold at various values

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

between \$10,000 and \$15,000, not dropping any values, as well as to including multi-driver policies. Concerns with the data and our measure of uninsured drivers are addressed in Appendix B. We go over potential concerns about our results one by one in the following sections.

6.1 County Waves

One potential confound is that the results are driven by the pilot group of counties, Los Angeles and San Francisco, which have different characteristics than other counties. Both are urban areas, and Los Angeles County has the highest average premia of any county in the state of California. A similar concern applies to the last wave of counties, which are typically smaller and more sparsely populated. Figure 2 indicates that the counties in wave 5 tend to have slightly lower premia than other counties, and this may bias our coefficients downwards.

Column 1 of Table 5 presents results when the pilot wave is excluded. The coefficient on the rate of uninsured drivers increases slightly, and the standard errors on the coefficient decrease, but the basic result remains unchanged. We are unable to reject at the 5 percent level that the coefficients are identical in magnitude to the main specification. We conclude that differing characteristics of the first wave of CLCA counties are not driving our results. Column 2 of Table 5 provides results when the final wave of counties is excluded. In this specification the coefficient on the rate of uninsured drivers does not change significantly. We again conclude that our results are not driven by the counties in the final wave being different from other counties. Our results are also robust to dropping any other individual wave and dropping the first and the last wave together. We thus conclude that no single wave is driving our results.

6.2 Competition

Another potential confound is that introducing the CLCA program lowers premia through more than one 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 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,[34] the structure of the CLCA program allows us to test this possibility. 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

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

our sample to only those vehicles which were ineligible for the CLCA program by throwing out all premium quotes for vehicles that were eligible for the CLCA program.[35].

Column 3 of Table 5 reports our findings restricted to only those vehicles ineligible for the CLCA program. 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. However, the coefficient on the rate of uninsured drivers remains largely similar and remains significant at the 5% level. Again the coefficient is statistically indistinguishable from the coefficients in our main result. This is not surprising, given our theory that uninsured drivers will generate a negative externality to insured drivers. This suggests 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.

6.3 Unobserved Selection

A major 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. Second, the specification in column 3 of Table 5 provides further evidence that unobserved selection is not driving our results. This column restricts to the sample to vehicles which were above the $20,000 threshold, and hence ineligible for the CLCA program. This group of drivers would not be affected by any unobserved selection into the CLCA 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 Choice of Instruments

Our main instrument is the number of months during which the CLCA program was active in a zip code cluster. We vary the definition of this instrument and our results remain robust.

[35] 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.

Column 4 of Table 5 reports results when we replace the number of months during which the CLCA program was active in a given zip code cluster with an indicator of whether or not the CLCA program was active at all during the year. The coefficient on the rate of uninsured drivers increases slightly while our basic result remains qualitatively unchanged. If instead we drop the CLCA rates instrument, the results of this specification being reported in Column 5 of Table 5, the measured effect of uninsured drivers remains largely unchanged. We wish to draw specific attention to this robustness check, as this suggests that our results are not driven by the rates instrument. One potential concern is reverse causality in the first stage– since the rates are set by losses, more uninsured drivers may lead to more losses, in turn leading to higher rates making the program's subsidy smaller. While we cannot rule this possibility out, our results are robust to excluding the rate instrument so we do not view this phenomenon as being able to drive our results. In both cases, altering the definition of our instrument slightly does not alter our main result significantly. Finally, if we use only the average number of months that the CLCA program is in effect as an instrument, our coefficient increases, but is less precisely measured. Again we cannot reject that the coefficient is $28 as estimated in Table 4. Our results are thus robust to changes in instrument specification.

6.5 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[36] 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 location (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

[36]Section 1861.02 (a)

formula than the optional pricing factors. [37] 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.

6.6 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.[38] 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. However Cohen (2005) notes that the results of Chiaporri and Salanié (2000)[39] are also consistent with asymmetric information and learning. Cohen and Dehejia (2004) also 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 very 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.

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

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

[39] 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. This relationship is consistent with classic adverse selection theory, 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.

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 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 formally modelled by Smith and Wright (1995). 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 between 1-2% 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.

Another 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.[40] 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.

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

References

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.*

Carlson, J. A. and R. P. McAfee (1983). Discrete equilibrium price dispersion. *The Journal of Political Economy 91*(3), 480–493.

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 L. Einav (2007). Estimating risk preference from deductible choice. *American Economic Review 97*(3), 745–788.

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.

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.

Hausman, J. (1983). Specification and estimation of simultaneous equation models. *Handbook of Econometrics, 1*(1), 391–448.

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 ration. *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.

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.

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.

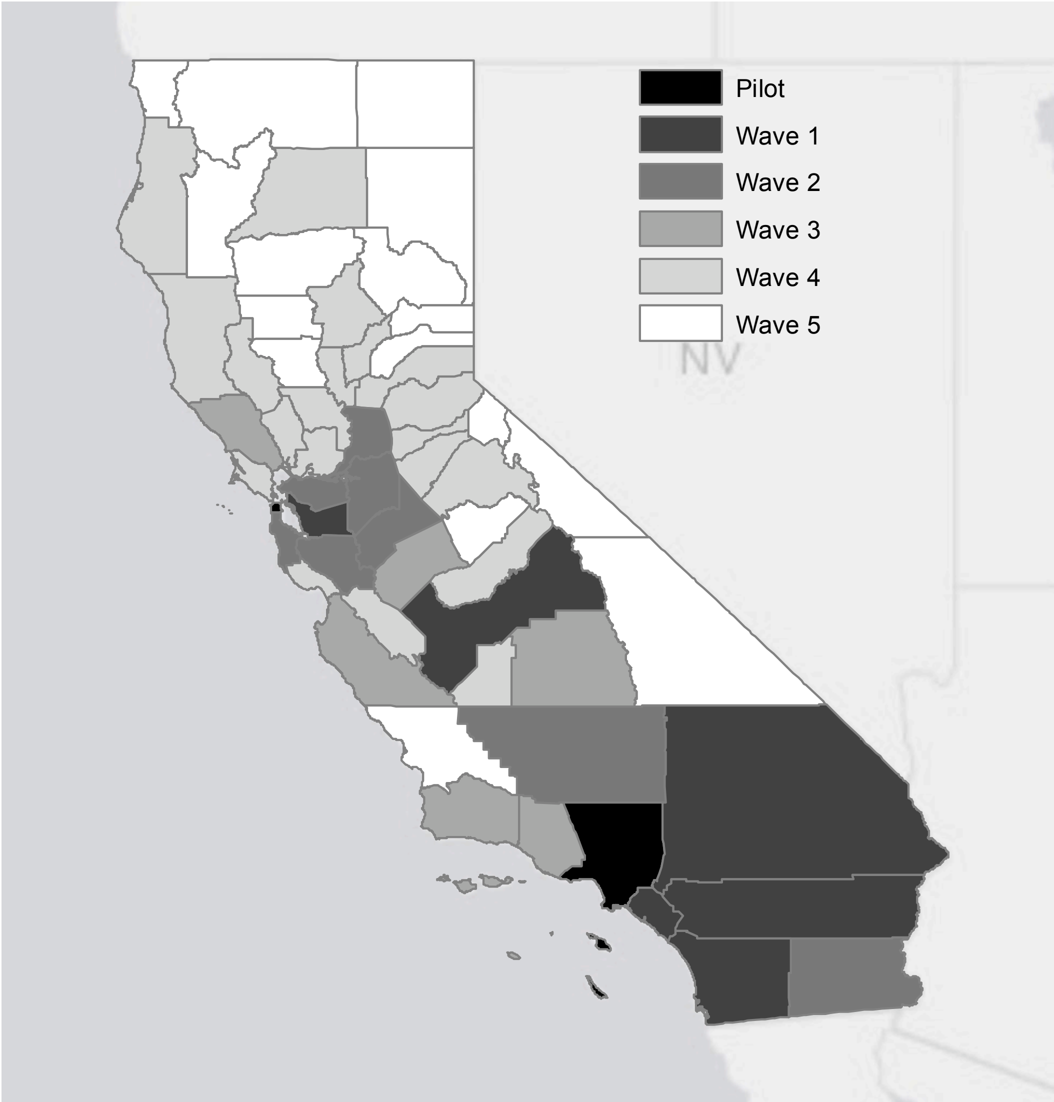
Smith, E. F. and R. Wright (1992). Why is automobile insurance in philadelphia so damn expensive. *American Economic Review 82*(4), 756–72.

Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy 69*(3), 213–225.

Wooldridge, J. (2010). Econometric analysis of cross section and panel data. *MIT Press.*

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.

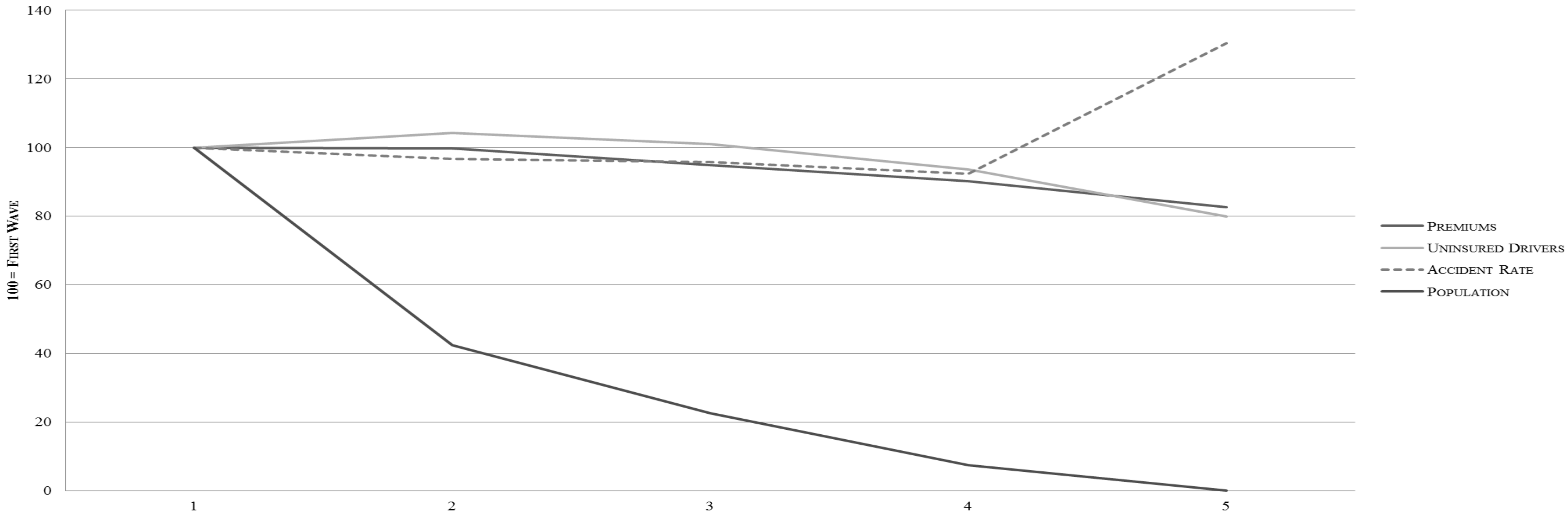
Wave 2 (June 1, 2006)-Contra Costa, Imperial, Kern, Sacramento, San Joaquin, San Mateo, Santa Clara, Stanislaus.

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

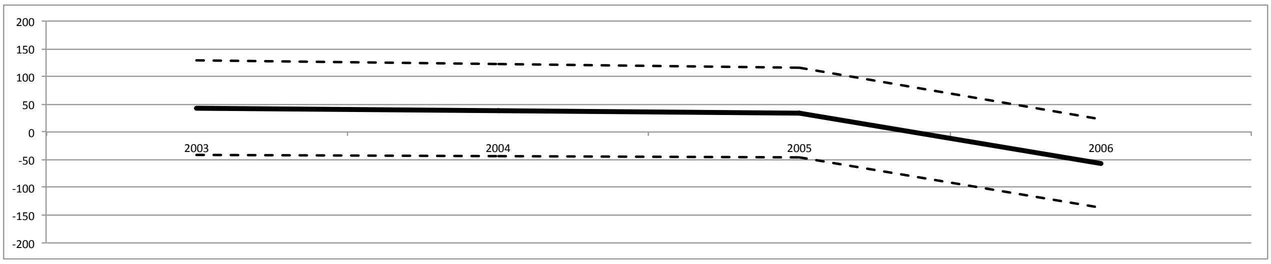
Figure 2: CLCA Waves Assigned by Population



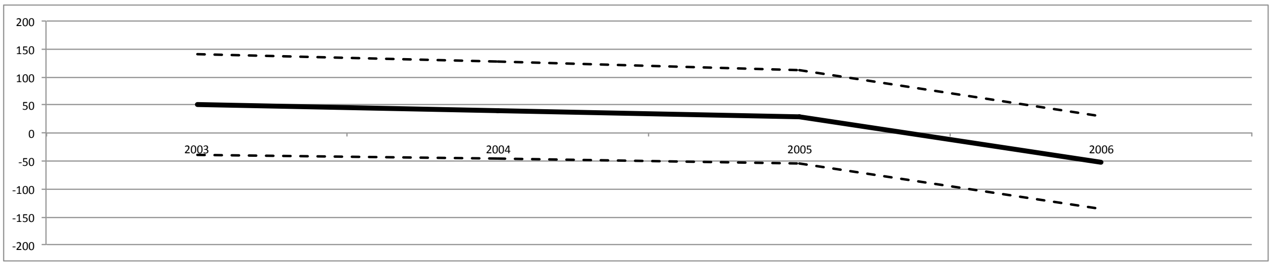
Notes: This figure plots average inflation adjusted premia, the rate of uninsured drivers, accident rates and population across each wave of the implementation of the CLCA program. The spike in accident rates in the fifth wave is driven by counties in the Lake Tahoe region. Including these counties does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

Figure 3: No Significant Pre-Trend Across Waves

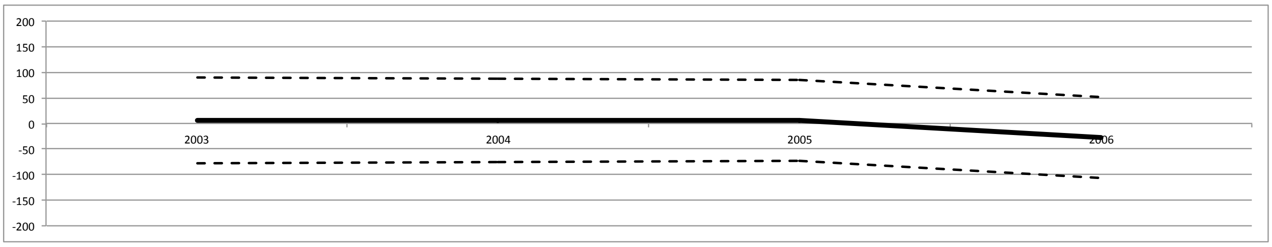
Wave 1



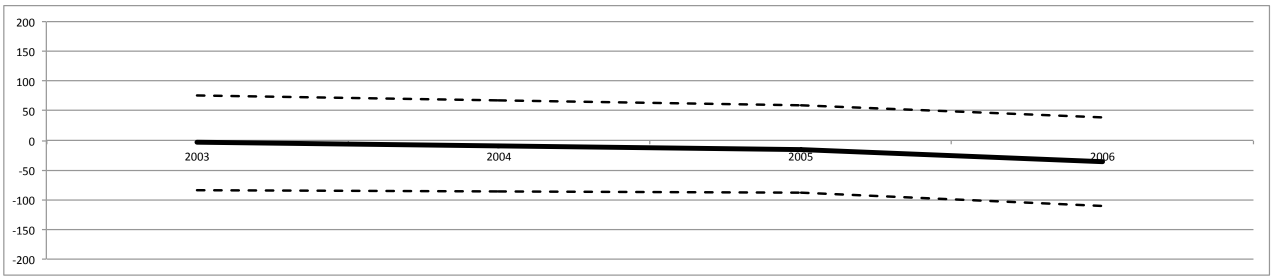
Wave 2



Wave 3



Wave 4



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.

Table 1: 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 2: 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 in Effect High Uninsured Zip	(5) CLCA Rates in Effect	(6) Effect on UI Claims
	-1.526***	-0.148***	-0.015***	0.894***	-0.004***	-0.326***
	(0.0034)	(0.0047)	(0.0005)	(0.0155)	(0.0001)	(0.0178)
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 the first five 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 independent variable in specification (4) is an interaction between the average number of months the CLCA program is in effect in the zip code cluster and an indicator of whether the rate of uninsured drivers is greater than 25% for the entire sample period. The independent variable in specification (5) consists of the premium rates in each county that were set annually for participants in the program. The dependent variable in (6) is the uninsured motorist claims in a zip code. Each specification includes demographic controls analogous to the specification in Table 5, with the exception of (6), which excludes individual level controls. The independent variable in (6) is the average number of months that the CLCA program is in effect. The rate of uninsured drivers is measured between 0 and 100. The accident rate is measured by the number of injury exposures over the total number of registered vehicles in a county. Standard errors are in parentheses and are clustered at the zip code by year by company level.

Table 3: Survey Sample Does Not Change Significantly

	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: Main Results

	(1) OLS	(2) OLS	(3) FE	(4) FE	(5) IV	(6) IV
Uninsured Drivers	-11.15***	-12.95***	3.23***	3.19***	27.84***	28.32***
	(1.31)	(1.27)	(1.14)	(1.14)	(8.88)	(8.93)
Accident Rate	1332.7***	1348.5***	264.3***	259.7***	203.5***	197.6***
	(44.88)	(44.87)	(58.14)	(58.39)	(61.11)	(61.43)
At Fault Accident	767.2***	801.0***	736.4***	774.9***	736.4***	774.9***
	(6.67)	(6.73)	(6.04)	(6.19)	(6.04)	(6.19)
Standard	1667.2***	1729.6***	1021.3***	1695.4***	1021.3***	1695.5***
	(14.90)	(14.67)	(22.39)	(22.90)	(22.36)	(22.86)
Age		-273.9***		-285.4***		-285.4***
		(5.72)		(3.58)		(3.58)
Age^2		3.53***		3.66***		3.66***
		(0.08)		(0.05)		(0.05)
Daily Miles		41.81***		42.82***		42.82***
		(0.37)		(0.36)		(0.36)
Speeding Ticket		591.3***		553.4***		553.4***
		(5.48)		(4.39)		(4.39)
Female		-146.6***		-164.7***		-164.6***
		(3.25)		(2.64)		(2.64)
R^2	.12	.45	.64	.72	.64	.72
Hausman	..	..	..	..	.75	.77
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 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 rates set by the CAARP for participants of the CLCA program in the county in which the premium quote is located (iii) the average number of months during which the CLCA program was in effect in a zip code cluster squared and (iv) an interaction between the number of months the CLCA program in in effect and the zip code cluster having a high rate of uninsured drivers. 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 by year for each firm and vehicle.

Table 5: Robustness Checks

	(1) No First Wave	(2) No Final Wave	(3) CLCA Inelig.	(4) Indicator Inst.	(5) No Rate Inst.	(6) Only Month Inst.
Uninsured Drivers	41.75***	27.00***	27.64**	31.60***	23.58***	38.93*
	(7.04)	(8.35)	(10.47)	(9.05)	(8.66)	(22.90)
Controls	✓	✓	✓	✓	✓	✓
Zip Code Fixed Effects	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓
Vehicle Fixed Effects	✓	✓	✓	✓	✓	✓
Firm Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	2,980,193	4,499,035	3,943,086	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, which is measured between 0 and 100, is instrumented using (i) the average number of months during which the CLCA program was in effect in a zip code cluster (ii) the rates set by the CAARP for participants of the CLCA program in the county in which the premium quote is located (iii) the average number of months during which the CLCA program was in effect in a zip code cluster squared and (iv) an interaction between the number of months the CLCA program in in effect and the zip code cluster having a high rate of uninsured drivers. 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. Each specification includes the accident rate, driving history variables, age, plan type and gender. All specifications include zip code, year, firm and vehicle fixed effects. Standard errors are in parentheses and are clustered at the level of the zip code by year for each firm and vehicle.

Table 6: Instrument Only Using CLCA Months in Effect Instruments

	(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, which is measured between 0 and 100, 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. 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 specifications include zip code, year, firm and vehicle fixed effects. Standard errors are in parentheses and are clustered at the level of the zip code by year for each firm and vehicle.

A Appendix Tables

Table A1: Summary Statistics

COUNTY	Premium	% Uninsured	Accidents	Age	Miles/Day	Speeding	License	Standard	Population	Wave
ALAMEDA	2349.55	22.37	.91	29.98	12.43	.49	12.01	.75	1,466,761	2
sd	1531.91	2.00	.09	8.14	2.50	.50	8.10	.44	6094.71	..
ALPINE	1786.93	29.81	5.28	29.97	12.43	.49	11.9956	.75	1208	6
sd	1082.35	0	0	8.14	2.50	.50	8.11	.44	0	..
AMADOR	1762.57	24.95	.90	29.98	12.44	.49	12.01	.75	37,667.69	5
sd	1092.46	4.28	.048	8.14	2.50	.50	8.11	.43	490.16	..
BUTTE	1766.13	18.25	.61	29.98	12.44	.49	12.01	.75	214,382.80	5
sd	1069.90	3.62	.08	8.14	2.50	.50	8.1	.43	2337.12	..
CALAVERAS	1816.12	25.71	.76	29.98	12.44	.49	12.01	.75	44,780.52	5
sd	1091.12	3.24	.09	8.14	2.50	.50	8.11	.43	899.90	..
COLUSA	1797.52	21.61	1.10	29.98	12.44	.49	12.00	.75	20,594.31	6
sd	1108.30	1.55	.13	8.14	2.50	.50	8.11	.43	520.66	..
CONTRA COSTA	2138.26	23.45	.59	29.98	12.44	.49	12.00	.75	1,006,720	3
sd	1309.79	1.88	.05	8.13	2.50	.50	8.10	.43	10884.05	..
DEL NORTE	1668.57	17.03	.95	29.99	12.44	.49	12.01	.75	28,246.55	6
sd	1041.26	4.70	.14	8.15	2.50	.50	8.11	.43	213.30	..
EL DORADO	1837.04	21.74	.68	29.98	12.44	.49	12.00	.75	173,500.30	5
sd	1114.74	4.58	.08	8.14	2.50	.50	8.11	.43	3400.57	..
FRESNO	2232.66	32.08	1.00	29.97	12.44	.48	12.00	.74	872,522.50	2
sd	1279.98	3.07	.16	8.14	2.50	.50	8.10	.43	20740.67	..
HUMBOLDT	1737.33	24.42	.86	29.99	12.44	.49	12.01	.75	131,833.30	5

sd	1113.70	4.88	.05	8.14	2.50	.50	8.11	.43	892.34	..
IMPERIAL	1750.15	26.63	.83	29.98	12.44	.48	12.01	.75	158,225	3
sd	1051.23	3.76	.13	8.14	2.50	.50	8.11	.43	6162.08	..
KERN	1753.13	21.88	.88	29.98	12.44	.49	12.00	.75	764,739.70	3
sd	1021.19	3.70	.068	8.14	2.50	.50	8.10	.43	35444.22	..
KINGS	1697.53	25.16	1.05	29.99	12.43	.48	12.01	.75	144,288.40	5
sd	1032.22	2.72	.09	8.13	2.50	.50	8.10	.43	4528.68	..
LAKE	1684.33	23.11	.67	29.97	12.44	.49	12.00	.75	63,256.41	5
sd	1049.98	1.55	.07	8.15	2.50	.50	8.12	.44	756.10	..
LASSEN	1680.33	17.96	.67	29.99	12.43	.49	12.01	.75	34,849.42	6
sd	1050.12	3.31	.07	8.15	2.50	.50	8.11	.44	608.61	..
LOS ANGELES	2981.83	20.56	1.21	30.00	12.44	.48	12.02	.76	9,805,209	1
sd	1814.89	1.82	.10	8.13	2.50	.50	8.10	.43	18516.37	..
MADERA	1865.02	31.08	1.09	29.97	12.43	.48	12.00	.74	139,841.10	5
sd	1117.16	2.92	.17	8.14	2.50	.50	8.10	.44	5416.02	..
MARIN	1996.72	23.13	.64	29.97	12.44	.49	12.00	.75	247,729.80	5
sd	1158.69	2.10	.06	8.13	2.50	.50	8.10	.43	876.01	..
MARIPOSA	1780.69	20.14	.67	29.98	12.43	.49	12.00	.74	18,120.16	6
sd	1101.84	1.81	.10	8.14	2.50	.50	8.11	.44	220.48	..
MENDOCINO	1682.99	21.45	.74	29.99	12.43	.49	12.01	.75	87,809.14	5
sd	1045.10	3.50	.10	8.15	2.50	.50	8.12	.44	202.80	..
MERCED	1853.60	24.18	1.05	29.97	12.43	.49	12.00	.75	241,588.50	4
sd	1127.41	2.51	.19	8.14	2.50	.50	8.10	.46	7301.07	..
MODOC	1651.03	13.80	.82	29.99	12.43	.49	12.01	.75	9,585.33	6
sd	1046.81	5.50	.08	8.15	2.50	.50	8.11	.44	44.89	..
MONO	1622.74	11.85	1.14	29.97	12.43	.48	11.99	.74	13,896.98	6

sd	1028.85	.93	.18	8.14	2.50	.50	8.10	.44	267.42	..
MONTEREY	1744.56	22.24	.74	29.96	12.43	.48	11.99	.74	408,377.80	4
sd	1039.29	2.66	.10	8.13	2.50	.50	8.10	.44	2204.47	..
NAPA	1910.95	24.25	.99	29.99	12.43	.49	12.01	.75	131,353.30	5
sd	1130.38	2.00	.10	8.15	2.50	.50	8.12	.44	1340.38	..
NEVADA	1814.99	22.35	.65	29.98	12.44	.49	12.00	.75	97,708.64	6
sd	1091.57	2.10	.06	8.14	2.50	.50	8.11	.43	953.73	..
ORANGE	2152.89	20.05	.81	29.97	12.44	.49	12.00	.75	2,954,766	2
sd	1219.76	2.36	.06	8.14	2.50	.50	8.10	.43	9146.53	..
PLACER	1927.67	24.43	.67	29.98	12.44	.49	12.00	.75	314,529.70	5
sd	1116.55	4.06	.07	8.14	2.50	.50	8.11	.43	14039.38	..
PLUMAS	1687.41	19.16	.83	29.99	12.43	.49	12.01	.75	20,730.07	6
sd	1058.35	2.87	.09	8.15	2.50	.50	8.12	.44	94.40	..
RIVERSIDE	2000.98	21.58	.91	29.98	12.44	.49	12.00	.75	1,933,874	2
sd	1140.62	2.83	.09	8.13	2.50	.50	8.10	.43	124406.70	..
SACRAMENTO	2276.41	23.55	1.19	29.98	12.44	.49	12.00	.75	1,363,683	3
sd	1320.29	3.48	.14	8.14	2.50	.50	8.11	.44	22052.15	..
SAN BERNARDINO	2144.13	21.17	.96	29.97	12.44	.49	12.00	.75	1,938,083	2
sd	1243.25	3.70	.10	8.14	2.50	.50	8.11	.43	55900.64	..
SAN DIEGO	1890.45	18.76	.82	29.98	12.44	.48	12.01	.75	2,973,480	2
sd	1118.99	2.08	.07	8.14	2.50	.50	8.11	.43	23122.56	..
SAN FRANCISCO	2807.61	22.62	1.18	29.97	12.44	.49	12.00	.75	783,820.50	1
sd	1737.91	1.68	.11	8.13	2.50	.50	8.10	.43	4383.95	..
SAN JOAQUIN	2132.06	23.07	1.16	29.97	12.43	.49	11.99	.74	657,030.80	3
sd	1239.72	2.17	.13	8.15	2.50	.50	8.11	.44	13020.83	..
SAN LUIS OBISPO	1522.18	22.40	.66	29.98	12.44	.49	12.01	.76	260,423	6

sd	908.02	3.31	.06	8.13	2.50	.50	8.10	.43	3437.35	..
SAN MATEO	2110.60	20.92	.62	29.97	12.44	.49	12.00	.75	701,331.70	3
sd	1307.71	1.80	.05	8.13	2.50	.50	8.10	.43	2208.75	..
SANTA BARBARA	1585.20	21.04	.80	29.98	12.44	.48	12.01	.75	412,171.20	4
sd	961.01	1.66	.05	8.13	2.50	.50	8.10	.43	2572.64	..
SANTA CLARA	1933.24	19.77	.66	29.97	12.43	.49	12.00	.74	1,712,505	3
sd	1174.65	2.06	.05	8.14	2.50	.50	8.10	.44	15745.14	..
SANTA CRUZ	1881.42	20.76	.66	29.98	12.44	.49	12.01	.75	255,706.10	5
sd	1175.06	1.94	.03	8.14	2.50	.50	8.10	.43	1098.14	..
SHASTA	1753.84	23.33	.89	29.99	12.43	.49	12.01	.75	174,405.80	5
sd	1082.21	4.91	.11	8.15	2.50	.50	8.11	.44	1411.67	..
SIERRA	1671.28	17.41	.75	29.95	12.43	.48	12.01	.75	3344	6
sd	1048.77	0	0	8.15	2.50	.50	8.12	.44	0	..
SISKIYOU	1628.00	15.56	.63	30.00	12.43	.49	12.02	.75	44,833.20	6
sd	1040.03	5.01	.08	8.15	2.50	.50	8.12	.44	129.50	..
SOLANO	1987.17	23.84	.69	29.97	12.43	.49	12.00	.74	411,217.50	5
sd	1181.62	2.00	.08	8.14	2.50	.50	8.11	.44	1103.07	..
SONOMA	1901.88	23.55	.75	29.98	12.43	.49	12.01	.74	470,355.30	4
sd	1125.56	2.23	.09	8.15	2.50	.50	8.11	.44	1657.37	..
STANISLAUS	2148.44	22.12	1.12	29.98	12.43	.49	12.00	.74	499,943	3
sd	1251.35	2.57	.147	8.15	2.50	.50	8.11	.44	8271.63	..
SUTTER										
mean	1875.38	21.10	.91	29.99	12.44	.49	12.01	.75	89,648.47	5
sd	1117.68	1.98	.08	8.15	2.50	.50	8.11	.43	3049.33	..
TEHAMA	1736.32	19.97	1.01	29.98	12.44	.49	12.00	.75	60,678.09	6
sd	1082.09	5.37	.17	8.14	2.50	.50	8.11	.43	1330.10	..

TRINITY	1666.16	24.46	.99	29.99	12.43	.49	12.01	.75	13,709.18	6
sd	1068.06	4.61	.07	8.14	2.50	.50	8.11	.44	171.73	..
TULARE	1832.37	27.31	1.01	29.97	12.43	.49	11.99	.75	412,794.50	4
sd	1087.60	5.58	.14	8.14	2.50	.50	8.11	.44	10770.54	..
TUOLUMNE	1802.03	23.62	.85	29.99	12.44	.49	12.01	.75	56,402.36	5
sd	1080.80	2.92	.07	8.15	2.50	.50	8.12	.43	157.91	..
VENTURA	2317.30	19.83	.78	29.99	12.44	.48	12.01	.76	797,853.40	4
sd	1428.92	2.06	.07	8.13	2.50	.50	8.10	.43	5410.74	..
YOLO	1783.13	22.75	.75	30.00	12.44	.49	12.02	.75	188,983.40	5
sd	1085.61	2.61	.09	8.15	2.50	.50	8.12	.4345037	4424.111	..
YUBA	1899.98	21.81	.88	29.99	12.44	.49	12.01	.75	67,898.20	5
sd	1144.79	1.49	.28	8.15	2.50	.50	8.12	.43	2273.73	..
CALIFORNIA AVERAGE										
mean	2355.88	21.13	.98	29.99	12.44	.48	12.01	.75	4,383,212	..
sd	1539.10	3.47	.26	8.14	2.50	.50	8.10	.43	4,106,721	..
Observations	4,937,664									

B Estimating the Rates of Uninsured Drivers

B.1 Methodology

From 1996 to 2005 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_1 be the accident rate of insured motorists, and let p_2 be the accident rate of uninsured motorists. Then $p_1 = p_2$.

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 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 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. 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 using longitude (lon) and latitude (lat), $distance = R\sqrt{(lon_2 - lon_1)^2 Cos(\frac{lat_1+lat_2}{2})^2 + (lat_2 - lat_1)^2}$, where R is the Earth's radius. We record the CLCA program as not being in effect if the average number of months is below a third of the year.

B.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.

C Pigouvian Taxation

Let agents be endowed with a standard utility function $U(.)$, which is concave, continuous, and increasing. Let λ 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 π_i and losses L_i^s are stochastic. The government levies a fine τ 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 $c_{na}^i = q_i[w_i - p_i + s] + (1 -$

$q_i)[w_i - \tau + s]$ 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 τ. 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 who have uninsured motorist coverage. We can denote the solution to the optimization problem for a given subsidy s and tax τ by $V(s, \tau)$. We assume a benevolent government which levies a fine (or tax) τ 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 μ_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

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

We recall that $p = \beta\lambda(\tau)+k$ and we can 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 β, the size of the externality, and decreasing in λ, the rate of uninsured drivers.

D Institutional Background and Data

D.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 had 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 on 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 over 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.

D.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.