Selection and Unraveling in Automobile Insurance: The Effect of Per-Mile Insurance*

Steven Berry, Yale University
Kenneth Gillingham, Yale University
James Levinsohn, Yale University

PRELIMINARY DRAFT - PLEASE DO NOT CITE
April 28, 2019

Abstract

Economists have long argued that automobile insurance does not appropriately price the external costs the driver imposes on others. With the introduction of inexpensive GPS equipment, technology now allows for pricing automobile insurance per mile driven, more closely reflecting the true costs, which scale with miles driven. Using data from a per-mile insurance start-up, this paper examines the effect of the introduction of per-mile insurance, exploring both the selection of drivers into per-mile insurance and the behavioral response of these drivers. We then use these results to calibrate a simple model of the evolution of the automobile insurance market to show how per-mile insurance could lead to an unraveling of the current insurance market. However, our results provide little support for reductions in driving and emissions.

Keywords: automobile insurance, Pigouvian pricing, pay-as-you-drive insurance.
JEL classification codes: D12, H25, L11, L62, L71, Q4

*Berry: Yale University and NBER, Department of Economics, P.O. Box 208264, New Haven, CT 06520, phone: 203-432-3556, e-mail: steven.berry@yale.edu; Gillingham: Yale University and NBER, 195 Prospect Street, New Haven, CT 06511, United States, phone: 203-436-5465, e-mail: kenneth.gillingham@yale.edu; Levinsohn: Yale University and NBER, School of Management, Jackson Institute for Global Affairs, P.O. Box 208206, New Haven, CT 06520.

The authors are grateful for excellent research assistance from Wade Davis. We also appreciate the comments from seminar participants at the NBER pre-conference workshop.

1 Introduction

...the manner in which [auto insurance] premiums are computed and paid fails miserably to bring home to the automobile user the costs he imposes in a manner that will appropriately influence his decisions - William Vickrey (1968) p.464

Economists have long been aware of suboptimal pricing in automobile insurance markets. A fundamental issue is that accidents are proportional to the amount of driving, while conventional automobile insurance contracts exhibit only a weak relationship between the quantity of driving and the insurance premium. In this paper, we explore the potential for a per-mile insurance offering to upend the existing automobile insurance market by providing a contract that tightens the relationship between the amount driven and the premium. As in any insurance market, adverse selection and behavioral responses are of crucial importance in the automobile insurance market. Per-mile insurance might be expected to both select low-mileage drivers who have lower absolute risk and at the same time discourage driving due to the higher marginal cost per mile for adopters. The selection effect could leave conventional fixed-rate insurance providers with a pool of increasingly higher mileage–and thus increasingly risky–customers. And the behavioral response could further exacerbate the effect. Just as health insurance markets can unravel due to adverse selection (Rothschild and Stiglitz 1976; Cutler and Reber 1998; Einav et al. 2010), one could imagine per-mile insurance leading to "death spiral" of the automobile insurance market, whereby conventional fixed-rate insurance offerings are entirely replaced by per-mile insurance.

Such an unraveling in the automobile insurance market could have important implications for the market structure of automobile insurance, leading to a completely different set of insurance offerings dominating the market, and potentially providing headroom for new entrants to disrupt the market. It also may have important implications for greenhouse gas emissions, local air pollutant emissions, oil consumption, and congestion. Economic evidence indicates that these further externalities of driving are also underpriced in the United States (Parry et al. 2007), leading to overuse of our vehicles. By discouraging driving, per-mile insurance could have the added benefit of reducing these impacts of driving, bringing the quantity of miles driven closer to the socially-optimal level. Indeed, policymakers have even discussed per-mile insurance as a potential policy tool to reduce greenhouse gas emissions.[1]

This paper examines the potential for unraveling in the automobile insurance market using unique data from one of the first per-mile insurance providers in the United States. Until recently, per-mile insurance was not technologically feasible, but vehicle-level GPS equipment has become inexpensive, allowing for a large increase in per-mile insurance over the past several years. We observe the vehicles and trips for a large sample of customers in four states, and compare the amount driven and the characteristics of the vehicles to the full population of drivers in three of these states, using vehicle inspection data. We further find the accident probability using data from a database of police-reported accidents in several states. These data provide rich insight for calibrating a simple model of the evolution of the per-mile and conventional automobile insurance markets, allowing us to uncover the conditions under which unraveling of the dominance of conventional insurance could occur.

We find substantial selection into per-mile insurance based on the amount driven. The average per-mile insurance customer from our provider is in the 27th percentile of the overall distribution of driving. Surprisingly, we find little evidence of a behavioral response in our sample. Per-mile insurance customers who were previously enrolled as testers of the GPS technology did not appear to drive less after adopting per-mile insurance. We posit that this result may be due to either our sample of very early adopters or to a behavioral response that occurred upon enrollment as a tester. Thus, for our calculations of the effects of per-mile insurance on driving and emissions, we explore the sensitivity of our results to different behavioral responses, including values from standard cost per mile of driving elasticities in the literature.

[1] For example, California considered adding per-mile insurance in its Draft Plan to lower the state's greenhouse gas emissions to meet its 2020 limit under Assembly Bill 32 (Nichols and Kockelman 2015)

We develop a calibrated simulation model to explore unraveling in the insurance market, and find that such unraveling can occur under realistic conditions. Indeed, we find that there are large groups of consumers that would save money under per-mile insurance and that these consumers tend to be relatively lower risk consumers. Yet, per-mile insurance is relatively unknown, even in the markets where it exists, and there is considerable consumer inertia in consumer insurance choices. We show that if consumers come to view per-mile insurance with the same perceived quality as conventional insurance, the unraveling becomes plausible.

Our work contributes to the literature on the economics of insurance markets. Much of the work in this area has focused on the economics of health care insurance markets, going back to classic papers on adverse selection (Rothschild and Stiglitz 1976; Cutler and Reber 1998). More recently, there has been work on behavioral responses in these markets as well (Einav et al. 2010; Aron-Dine et al. 2015). There is also literature on automobile insurance, exploring such aspects as risk preferences (Cohen and Einav 2007), asymmetric information and learning (Cohen 2005), the moral hazard effect of mandatory automobile insurance on traffic fatalities (Cohen and Dehejia 2004) and the efficiency implications of restrictions on rate classifications (Harrington and Doerpinghaus 1993).

There is limited previous work on per-mile automobile insurance. Early research argued that per-mile insurance is a nice idea that is simply infeasible because of high monitoring costs (Rea 1992). However, recent market developments have shown that technology can overcome this hurdle. Edlin (2003) develops a simple theoretical model of accidents and costs under fixed-price and per-mile insurance and then simulates the effects of switching entirely to per-mile insurance on crash counts and costs. Parry (2005) built on this framework to add more detailed estimates of driving responsiveness and a more complete exposition of reductions in externality costs that would be possible from a complete switch to per-mile insurance. However, Edlin (2003) and Parry (2005) did not observe heterogeneity in driving behavior and did not model the dynamics of a possible unraveling of the current insurance market.[2]

[2]A series of other papers outside of the economics literature have also pointed to welfare benefits from

Our analysis of how per-mile insurance could influence consumer behavior and emissions also contributes to the large literature on pricing externalities from driving. Economists have long known that driving is imperfectly priced, with environmental, energy security, congestion, and accident external costs that remain only partly internalized by gasoline taxes, tolls, and conventional automobile insurance (Parry et al. 2007; Coady et al. 2018). There is a line of research quantifying the spatial distribution of the external costs of driving (e.g., Newbury 1990; Safirova et al. 2007; Holland et al. 2016). Studies have also begun exploring the optimal mileage tax for alternative fuel vehicles Davis and Sallee (2019). A key take-away from this literature is that in many settings around the world there is substantial room for changes in policies to improve economic efficiency.

Indeed, our analysis has clear policy relevance. Only a small number of states in the United States have a regulatory structure that permits per-mile insurance offerings. Our findings suggest sizable emissions reductions and welfare improvements from an unraveling of conventional insurance offerings that leads to a switch to per-mile insurance. This suggests that state insurance regulators may improve economic efficiency by encouraging per-mile insurance. While such a regulatory change would still not perfectly internalize all externalities from driving, it would come closer to internalizing the accident externalities of driving an additional mile.

The remainder of this paper is organized as follows. The next section describes the per-mile insurance offering that we are examining. Section 3 presents our unique data set. Section 4 provides evidence on selection and potential behavioral responses from a switch to per-mile insurance and section 5 develops a model of unraveling in the automobile insurance market. Section 6 concludes.

a switch to per-mile insurance. For example, Bordoff and Noel (2008) use insurance claims data to find a strong positive relationship between miles driven and use this to argue for per-mile insurance. Ferreira and Minikel (2010) and Litman (2011) perform similar analyses for Massachusetts and Vancouver respectively. Nichols and Kockelman (2015) uses survey data from the entire United States in an analysis that parallels the others. Remarkably, there is very little mention of adverse selection in the transition to a per-mile insurance market in any of these papers.

2 Per-Mile Insurance Offerings

The market for per-mile insurance in the United States is nascent. We focus our study on a start-up founded in 2011 and based in San Francisco called MetroMile. MetroMile began underwriting policies in September 2016. MetroMile provides a per-mile insurance offering in several states around the country by using a small GPS unit that plugs into each subscriber's on-board vehicle computer (see Appendix A for a photo of the "Pulse" unit that plugs into the OBD-II port, which is the same port used by mechanics to diagnose a "check engine" light).[3] MetroMile charges a low fixed monthly rate and a per-mile rate for every mile driven. These values vary by state and driver, but a typical monthly base rate is $29 and per mile cost is $0.06 for all miles traveled under 250 miles in a single day (150 miles in some states). Above 250/150 miles per day, there is no further charge. The mile per day limit is intended to avoid deterring potential customers who do not use their vehicles often, but occasionally make longer trips. MetroMile also offers services to help optimize trips, get street sweeping alerts, find your parked car, and provide diagnosis of your cars running condition from a mobile phone app. Customers can also disable the GPS function of the device through an online dashboard, and when the GPS is disabled, the unit records the miles driven using the vehicles on-board computer.

MetroMile began in California and is currently offered in seven other states: Arizona, Illinois, New Jersey, Oregon, Pennsylvania, Virginia, and Washington. With a $90 million influx of capital in 2018, MetroMile is planning on scaling up to all 50 states.[4] Our study focuses on MetroMile, which is the first large-scale marketing of a per-mile insurance offering. Recently other conventional insurance carriers have begun creating similar per-mile offerings, including Progressive, Allstate (Esurance Pay Per Mile and Milewise), Liberty Mutual, and National General. Some of these other programs by incumbent insurance providers go beyond pricing per mile and also adjust prices based on metrics of

[3]The vehicle must be of model year 1996 or newer for the dongle to work.

[4]https://www.insurancebusinessmag.com/us/news/breaking-news/paypermile-insurer-plans-nationwide-expansion-after-funding-from-industry-titans-107173.aspx

safety of driving, such as the number of hard stops and number of fast accelerations per day. Per-mile insurance offerings also exist in several other countries, including Japan, Australia, and several countries in the European Union. However, while growing, the market share of per-mile insurance remains small in all of the markets where it exists.

3 Data

Our analysis is based on unique anonymized data from MetroMile. These data include all customers from MetroMile as of April 7, 2015. By this date, MetroMile had expanded to California, Illinois, Oregon, and Washington. The data also include a set of 'testers' who were given the MetroMile unit to plug into their vehicle's computer port (so their driving could be tracked), but they were not MetroMile customers. These testers began as friends and family of MetroMile employees, but expanded to people who responded to marketing expressing interest in trying out the MetroMile Pulse unit. MetroMile continued marketing to these testers and many of the testers subsequently became MetroMile customers after several months. These testers are a selected sample of interested parties in per-mile insurance, but they provide unique insight into the behavior of customers before and after becoming per-mile insurance customers. As per-mile insurance was not available before MetroMile, all testers can safely be assumed to have been enrolled in conventional insurance, rather than an alternative per-mile insurance offering.

For each anonymized identifer denoting an individual customer, we observe the vehicle identification number (VIN) prefix (first 10 digits), zip code of the customer, the premium contract terms, the coverage, and detailed information on the start time, end time, average speed, number of miles driven, and number of gallons for each trip taken by a tester or customer.[5] We also observe the dates at which the household becomes a tester and a customer. We use a VIN decoder from DataOne to decode all of the VIN prefixes, providing detailed vehicle characteristics for each vehicle in the sample.

These data provide unparalleled insight into some of the first adopters of per-mile

[5] For privacy reasons, we do not observe where the trips occurred.

insurance. To understand how these drivers compare to other drivers, we acquire vehicle inspection data on odometer readings from most registered vehicles in California, Illinois, and Oregon. The vehicle inspection programs in each state cover nearly all of each of the states, but miss some of the more remote areas.[6] Fortunately, there are very few vehicles in these more remote areas. As we were unable to acquire data from Washington State, we either exclude Washington from the analysis or use data from Oregon as a proxy for the distribution of driving. We observe daily average vehicle miles traveled (VMT) from the vehicle inspections, by calculating the difference in odometer readings between two inspections and dividing this by the number of days between the inspections. The vehicle inspection data also include the VIN prefix, allowing us to decode these large samples as well.

Panel A of Table 1 details the number of observations in our final dataset (see the Appendix B for details on data cleaning). It shows that we have 6,970 vehicles, of which 1,676 are customers. 51% of the vehicles are from California, and each of the other states has roughly 15% of the sample. Panel B of Table 1 shows the mean daily VMT for test drivers and customers, along with information on the average daily VMT that is free among the test drivers and customers. We immediately see that testers both have more daily trips and longer trips than customers, and that this is also true for testers who became customers. We observe that only a small percentage of the miles driven are free miles, indicating that MetroMile customers are almost always facing a per-mile insurance charge.

Finally, we acquire data on the insurance market from the California Department of Insurance, the National Association of Insurance Commissioners, `insure.com`, and `ValuePenguin.com`. We use these data to calibrate our simulation model.

[6]More specifically, the inspection programs cover areas that are or have recently been non-attainment zones under the National Ambient Air Quality Standards, which are most areas in these three states.

4 Evidence on Selection and Behavioral Responses

In this section, we provide descriptive evidence on selection into MetroMile insurance and behavioral responses to per-mile insurance. The behavioral response we are focusing on here is the number of miles driven. The insights from this section will be used in the next section to calibrate our model.

4.1 Selection

Recall that we are interested in the extent to which the customers who select into per-mile insurance are different than other customers in ways that are correlated with the absolute probability of a claim. If per-mile insurance can skim the "cream-of-the-crop" drivers who are low risk, this raises the possibility of an unraveling in the insurance market. There are many factors that contribute to the probability of a claim. While factors such as driving speed and overall risky driving behavior are unobservable, we can observe both the vehicle being driven and the number of miles driven.

As discussed in Jacobsen (2016), the vehicle consumers buy is highly correlated with unobserved driving behavior. We examined the most common vehicles driven by MetroMile customers. The most common vehicle is the Honda Civic, followed by the Ford Focus, Honda Accord, and Honda Civic. These are all small or midsize cars. As shown in Jacobsen (2016), these vehicle classes have relatively lower traffic fatality rates than other classes, such as sport utility vehicles or pickups. Of course, nearly all common vehicle models are represented in the data. Looking across the entire dataset, just over 8% of the sample are 4-door hatchbacks and nearly 39% of the sample are 4-door sedans. Fewer than 10% of the sample are four-wheel-drive vehicles, a low percentage relative to the United States or even California itself. Combined, these simple statistics suggest that lower risk customers may be selecting into MetroMile.

The risk of a claim obviously also scales with the number of miles driven, for exposure to accidents is directly proportional to the number of miles driven. The summary statistics in Table 1 already show that MetroMile customers drive less than testers. But it turns out that they drive much less than drivers on average. For this analysis, we use data from California. Figure **??** compares the distribution of the mean daily VMT by vehicle from the California inspection data to the distribution for California MetroMile testers and customers. The figure very clearly shows that MetroMile customers drive less than MetroMile testers, and much less than all drivers in California. The mean daily VMT for the population of drivers who receive inspections in California is 27.7 miles (10,110 miles per year), while the mean daily VMT for MetroMile drivers is 25.1 miles (9,162 miles per year). The mean daily VMT for MetroMile drivers is only in the 27th percentile of the distribution of drivers in California. We also see that the testers drive slightly less than the overall population of drivers in California, but not very much less, suggesting that the testers are only a somewhat selected sample.

We can also some further gain insight into selection by more carefully comparing the testers to customers. Testers have revealed themselves to be interested in per-mile insurance, so it is instructive to observe which of the testers choose to become customers. Table **??** illustrates the distribution of VMT among test drivers and test drivers who became customers during the time when they were still testers. We observe that across the entire distribution, the test drivers who became customers have a lower VMT (the K-S statistic of XX shows that they are statistically significant). This mirrors the difference in the distributions for California that we saw in Figure **??**. One caveat about these summary statistics is that they are comparing all testers over the full time period with testers who eventually became customers and thus are dropped from the sample when they become customers. In Appendix C, we present a simple difference-in-differences specification showing that the selection continues to hold even after controlling for day-of-week and month-of-year fixed effects.

These results indicate that both those who are interested in per-mile insurance (the testers) as well as those who sign up to be per-mile insurance customers are a strongly selected sample with much lower driving than the overall population. This may be expected, but it has important implications for the automobile insurance market. In future versions of this paper, we will make the comparison to the data from Oregon and Illinois as well.

4.2 Behavioral Responses

Per-mile insurance increases the effective cost per mile of driving. Thus, one might expect a behavioral response in the number of miles driven.[7] We consider behavioral responses in two ways. First, we consider estimates in the literature of the cost-per-mile price elasticity of driving, as these should provide insight into how consumers would be expected to change their driving upon an increase in the cost per mile of driving. Second, we explore whether testers change their behavior upon becoming customers in our MetroMile data.

There is a large and growing literature on the how consumers respond to changes in the cost per mile of driving. These studies tend to report the elasticity of driving with respect to the cost per mile of driving. Studies estimating this elasticity are largely identified based on how changes in fuel prices influence driving, so the setting is not exactly analogous in timing to consumers facing a per-mile insurance bill each month, but it can provide useful insight. We will focus here on some of the recent studies estimating a cost per mile elasticity from the United States. Knittel and Sandler (2016) find a medium-run (two-year) elasticity of VMT with respect to the cost per mile of driving of -0.15 in California. Wenzel and Fujita (2018) find a similar elasticity of -0.16 in Texas. Langer et al. (2017) find a short-run VMT price elasticity of -0.11 in Ohio. Hymel and Small (2015) find a long-run elasticity ranging from -0.04 to -0.18 for the United States. These estimates suggest that consumers are quite inelastic in their short- or medium-run driving behavior, but that there is a clear response to changes in the cost per mile of driving. For the sake of illustration, we perform our calculations using an estimated cost-per-mile elasticity of

[7] It is also possible that there is a behavioral response in the way people drive because they know that their insurance company can monitor their driving. For example, they may be more cautious drivers. However, we are unable to verify this.

driving of -0.10.

Using an elasticity from other studies provides guidance on the behavioral responses on the margin for the larger population of drivers, but several studies have suggested substantial heterogeneity in the response to changing gasoline prices (Langer et al. 2017; Gillingham 2014, 2015). Thus, it is possible that the marginal customer who selects into per-mile insurance when it is still a niche offering would respond differently than drivers on average. To explore this, we use a simple difference-in-difference strategy with the MetroMile testers, customers, and testers who become customers. This empirical strategy is facilitated by the fact that testers become customers at different times, allowing us to control for time-varying unobservables. The primary identification assumption is that the trends by other testers and customers that do not switch can serve as a reasonable control for the treated households. For a causal interpretation, we also must assume that the timing of when testers become customers is exogenous.

$$\log(VMT)_{it} = \beta 1(\text{customer})_{it} + \gamma_i + \eta_t + \delta_t + \varepsilon_{it} \tag{1}$$

where VMT_{it} is the log of mileage driven on day t by vehicle i.[8] The term $1(\text{customer})_{it}$ is an indicator for whether or not the driver is a MetroMile customer. The terms γ_i, η_t, and δ_t account for vehicle, day-of-week, and month-of-year fixed effects, respectively, and ε_{it} is the error term.

Table 3 presents the results by first decomposing the changes in miles driven per day into the miles per trip and number of trips per day. Columns (1) and (2) replace the dependent variable in equation (1) with the log of the number of miles per trip (plus one to account for zeros), while columns (3) and (4) replace the dependent variable with the number of trips per day. Columns (5) and (6) show the results of estimating (1) directly. For each dependent variable, we present the results using the complete final sample and a subsample where 871 vehicles that switched from being testers to customers are also excluded because they were either not observed more than one day during the tester

[8] To be precise, VMT_{it} is log of daily mileage plus one to address days with zero mileage.

period or the customer period. This second subsample is included to explore different sources of variation; the slightly smaller subsample improves the balance of the panel and drops vehicles where the pre-period or post-period does not have much information. Our preferred estimates are those on this smaller subsample.

A reasonable hypothesis for what one might expect to see in Table 3 is that households that switch to per-mile insurance would reduce their driving due to the increase in the cost per mile of driving. Indeed, in column (2) we see that the miles per trip decreases when testers switch to being customers, as expected. The coefficient indicates that those who switch reduce driving by 3%. However, in column (4) we observe that the total number of trips *increases* upon becoming a per-mile insurance customer, with roughly 0.18 additional trips per day (the average number of trips per day for drivers who became customers is 1.8).

There are several possible explanations for this finding. One possibility is that MetroMile insurance is so much less expensive than conventional insurance that there is an income effect: the switch to per-mile insurance relaxes the budget constraint so much that consumers take more trips.[9] Another possibility is that when households switch to per-mile insurance, they try to reduce their driving by taking shorter trips, but they unintentionally end up taking more trips. An example of this could be if households avoid trip chaining and going to a big box store further away on regular trips, but then realize later on that they need something from that store anyway. A third possible explanation is that when drivers become testers, they actively try to decrease their driving, but the effect wears off over time and they return to their normal driving habits after becoming customers. A final potential explanation is that testers chose to become customers when they expect to be taking more trips. In other words, they do not bother to change their insurance until they realize that they will be taking more trips in the near future, and this

[9]There is scant recent evidence from well-identified studies of income elasticities of driving. Glaister and Graham (2002) review the literature up to 2000 and find estimates in the range of 1.1 to 1.3. Under plausible assumptions, MetroMile insurance would increase the cost per mile of driving by 30%, which translates into a 36% increase in driving using a 1.2 elasticity. For reference, the decline in insurance cost on average would be from about \$1,100 per year (`https://newsroom.aaa.com/tag/driving-cost-per-mile/`) to \$360 (base rate) + 14.3 miles/day $\times$ 365 days $\times$ \$0.06 per mile = \$673 per year.

spurs them to make the switch.

The net result in column (6) shows that when we combine the negative effect on the miles per trip and the positive effect on the number of trips per day, we find that the total number of miles per trip actually *increases* by just over 7% when testers become customers. This is a surprising result that may not hold when the broader population adopters per-mile insurance, but it is a fascinating result for the population of testers. Thus, we perform our calculations of the implications of per-mile insurance both assuming that there is an increase in driving of 7% and a decrease in driving following the elasticities in the literature.

5 Unraveling in the Automobile Insurance Market

Now that we have evidence on the extent of selection and behavioral responses to per-mile insurance, we turn our attention to developing a stylized model of the automobile insurance market to explore how unraveling could occur. We begin by developing a random utility model for automobile insurance demand based on automobile insurance premiums for different driver VMT 'types.' Based on this random utility framework, we make a set of reasonable assumptions that allows us to model the evolution of the insurance market as consumer knowledge and acceptance of the per-mile offerings increase. We then discuss the implications of such an unraveling.

5.1 Random Utility Model of Insurance Choice

We begin by modeling the demand for automobile insurance offering j by driver i in a classic random utility framework:

$$u_{ij} = \delta_j - \alpha_i p_j(Z_i) + Z_i \gamma_j + \varepsilon_{ij} \tag{2}$$

where δ_j is a vector of dummy variables for each insurance offering and p_j is the insurance premium for offering j. p_j is modeled as a function of a vector of driver or vehicle characteristics Z_i, which can include miles driven, vehicle age, vehicle class, and location. α_i can be interpreted as the marginal utility of money, and is given an i subscript to allow for a possible random coefficient. ε_{ij} captures idiosyncratic taste shocks for insurance.

This model thus far is quite general and can easily be expanded along with data availability. For reasons of simplicity and data availability, we focus on the choice between per-mile insurance (MetroMile) and conventional (legacy) insurance. ε_{ij} can be assumed to be an i.i.d Type I extreme value error term, making the estimation a logit estimation (random coefficients logit if we have α_i rather than α). Alternatively, we could use a more flexible distributional assumption, as in Souza-Rodrigues (2019). For our analysis here, we estimate a logit model.

The coefficients from estimating equation (2) have a causal interpretation if the variation in price and characteristics is exogenous. While it may be possible to argue this, for the purposes of our stylized simulation, it is not strictly necessary. We are interested in the relationship between insurance premia and insurance choice to model how insurance choice will change along with changes in premia. Thus our context only requires that the effect of any omitted variables in our equation is stable over time, which is a more relaxed assumption. This assumption is also common in the literature on the economics of insurance markets (Hendel and Lizzeri 2003; Einav et al. 2010).

In this version of the draft, we use data on insurance premia from ValuePenguin for a male driver with a clean record who drives 12,000 miles a year, and use State Farm's conventional insurance offering as a proxy for all other conventional insurance. We collect data from ValuePenguin for a typical make and model for each vehicle class and for model year 2015 and model year 2000. We then bin all vehicles in California into classes and model year category bins. To perform this estimation we also need to match MetroMile customers to customers in the vehicle inspection data, as we do not observe the actual VIN for the MetroMile customers. We do this by finding vehicles with the same VIN prefix, driven roughly the same amount (within plus or minus 1 mile per day) over the same time period, and that are registered in the same zip code.[10]

[10] See Appendix D for further details of this matching.

The results of this illustrative logit estimation are given in Table 4. Column (1) performs the matching using the most refined 5-digit zip code, column (2) brings in more MetroMile customers by performing the matching using the 4-digit zip code, and column (3) brings in all MetroMile customers by performing the matching using the 3-digit zip code. We do not interpret these coefficients directly, since in a logit model they are not the marginal effects. However, the signs of the coefficients and relative magnitudes are useful to interpret. All of the coefficients are also statistically significant. The results show a large (relative to other coefficients) negative coefficient on the per-mile MetroMile insurance dummy. This is not surprising, as MetroMile is a start-up that most consumers do not know, so many consumers could save money and yet still have not yet adopted MetroMile. This coefficient is also likely capturing inertia in consumer choices in the insurance market. We also observe a negative coefficient on VMT, indicating that those who drive more are less likely to choose MetroMile insurance (the selection effect). Further, we see a positive coefficient on the monthly premium for conventional automobile insurance and a negative coefficient on the monthly premium for MetroMile insurance.

5.2 The Automobile Insurance Market

With the demand estimates in hand, we can make further headway on a model of the automobile insurance market. We intentionally keep the model simple to fix ideas. Specifically, we make a key assumption that is common in the economic literature on health insurance:

Assumption 1. A zero-profit condition holds for each insurance offering.

This assumption also implies no markups by insurers and 100% passthrough of insurer costs to customers. The assumption greatly simplifies the analysis and allow us to focus on the economics underlying the evolution of the insurance market. It disallows insurers to cross-subsidize insurance plans, which may occur in the short run. Thus, it is best seen as an assumption valid in the longer-run.

We next divide up the sample of drivers into bins $Z \in \mathcal{Z}$, where each bin characterizes drivers over a set of attributes (e.g., the simplest set of bins would be low, medium, and high VMT drivers). Under the zero-profit condition, revenues must equal costs:

$$\underbrace{\sum_{\mathcal{Z}} p(Z)M(Z)N(Z)}_{\text{Revenues}} - \underbrace{\sum_{\mathcal{Z}} C(Z)M(Z)N(Z)}_{\text{Claims Paid}} - F = 0 \tag{3}$$

where $p(Z)$ is the insurance premium per mile, $M(Z)$ is miles driven on average in bin Z, $N(Z)$ is number of vehicles in bin Z, $C(Z)$ are average claims paid per mile, and F are any fixed costs of running an insurance offering. For convenience, we assume $F = 0$ in this illustrative analysis.

The next step is to model how the average claims paid per mile are affected by the miles driven and characteristics. In our current analysis, we focus on the miles driven, although this could be extended to examine other characteristics. Specifically, we create three bins for drivers with low, medium, and high VMT per day. We have data from the California Department of Insurance on claims by average VMT drivers and low VMT drivers (we have a pending request for more complete data and this is an area we plan to explore further in the next iteration of the paper). Thus to make progress here, we apply the claims for average VMT drivers as the claims for the medium and high VMT per day bins. We apply the claims by low VMT drivers to the low VMT bin. We further assume that the claims per mile in each VMT bin will remain constant over time in our simulation–only the composition of vehicles in each insurance offering will change over time, changing the average claims per mile for the insurance offering.

A first observation from this stylized model is that with the relatively highly negative $\delta_{PerMile}$ in Table 4, per-mile insurance will remain a niche offering as it has a large negative utility shock to overcome. This motivates our approach to our simulation. We consider the case where $\delta_{PerMile}$ slowly converges to zero, putting per-mile insurance on a level playing field with conventional insurance so that they compete solely on the basis of the insurance premiums. For illustrative purposes, we ramp the $\delta_{PerMile}$ coefficient from -

11.86 to zero over 10 years using a simple linear function (i.e., the coefficient increases by 1.186 each year). Such a change could come about due to a major marketing effort. It could also come about due to government policy intervention to encourage per-mile insurance.

Next we use our random utility model to calculate the choice probabilities for a typical household in each VMT bin for per-mile and conventional insurance using the standard logit expression:

$$\mathrm{Pr(PerMile)} = \frac{\exp(\delta_{PerMile} + \alpha_{PerMile} p_{PerMile} + \alpha_{Conv} p_{Conv} + \gamma VMT)}{1 + \exp(\delta_{PerMile} + \alpha_{PerMile} p_{PerMile} + \alpha_{Conv} p_{Conv} + \gamma VMT)}$$

The probability of conventional insurance is just equal to $1 - \mathrm{Pr(PerMile)}$.

Finally, as $\delta_{PerMile}$ changes and $\mathrm{Pr(PerMile)}$ increase, we invert equation 3 to calculate what the premium for both per-mile insurance and conventional insurance must be under the zero-profit condition. This allows the exogenous change in $\delta_{PerMile}$ to change the composition of the number of drivers of each VMT bin that choose each insurance offering. Specifically, more low-VMT drivers will switch from conventional insurance to per-mile insurance, thus increasing the claims for conventional insurance. The zero-profit condition then leads conventional insurance to raise the insurance premium, thus leading to further defections from conventional insurance to per-mile insurance. Appendix XX provides further details on the simulation, including the exact numbers used.

Figure 2 shows an illustrative unraveling of the conventional insurance market. It starts slow, and then rapidly picks up steam as the claims for conventional insurance steadily deteriorate. By the last year, nearly 80% of the insurance market has switched over to per-mile insurance. This would imply much more business for MetroMile and likely per-mile insurance offerings from many of the legacy insurers. The key condition in our analysis for making this happen is that per-mile insurance faces a more modest utility detriment as it becomes better known and understood.

5.3 Implications of Insurance Market Unraveling

With insurance market unraveling, there would be subsequent effects on driving, emissions, and accidents. This section is inspired in part by Edlin (2003), but is extended to also model emissions. These calculations should be viewed as illustrative, given the nature of our simulation. We begin with the change in driving that might occur with broader adoption of per-mile insurance. The widespread adoption of per-mile insurance would lower the cost per mile of driving and thus would be expected to reduce driving, following standard elasticities in the literature. However, our behavioral response results in section 4 indicate that the per-mile customers on the margin do not appear to reduce overall driving, and may even slightly increase overall driving. Thus, we also present results showing what would happen with the increase in driving suggested by our results.

We begin by calculating the average cost per mile of driving over time in our unraveling simulation, and then we apply an elasticity of driving with respect to the cost-per-mile of -0.10, following the literature. For these illustrative calculations, we have to make a set of assumptions about the other costs per mile of driving. We assume a gasoline price of $2.50 per gallon, which was the average gasoline price in the United States on March 11, 2019.[11]. With a fuel economy of 25 miles per gallon,[12] this translates into a gasoline cost of 10 cents per mile. The American Automobile Association estimates that maintenance costs are 5 cents per mile. Tolls may add a further variable cost, but these vary by location and thus are impossible to estimate, so our calculations exclude them. Thus, the total variable cost per mile of driving under the 6 cents per mile MetroMile rate would be 21 cents per mile on average. Conventional insurance may add some cost with additional driving, but this is difficult to quantify. For the purposes of our illustrative calculation, we assume that conventional insurance is a fixed cost that does not vary with the number

[11] https://www.cnbc.com/2019/03/11/average-us-price-of-gas-jumps-6-cents-per-gallon-to-2point50.html

[12] The average fuel economy of the full U.S. vehicle fleet in 2018 was 24.7 miles per gallon according to the U.S. Environmental Protection Agency (https://www.reuters.com/article/us-autos-emissions/u-s-vehicle-fuel-economy-rises-to-record-24-7-mpg-epa-idUSKBN1F02BX).

of miles driven per year. Under this assumption, the total variable cost per mile of driving under conventional insurance would be 15 cents per mile on average.

Using the cost per mile of driving increase of 29% along with the elasticity of -0.1 implies that under a complete unraveling, driving will decline by 2.9% on average. This change in driving would first affect lower-VMT households, so the impact would be small in the early years after the introduction of per-mile insurance. It would increase rapidly and be more noticeable just prior to the unraveling. And a decline in driving of about 3% would translate into decreased congestion and traffic fatalities, although it of course depends on where the reduced miles occur. Lower-VMT households are known to be in cities (Gillingham 2014), where congestion is the greatest. Our selection results indicate that we would see low-VMT households being the first to switch to per-mile insurance, so there may be disproportionately larger improvements in congestion and accidents in the early years before complete unraveling. By the time of complete unraveling, the decrease in congestion and traffic fatalities would be expected to proportionally scale with changes in VMT.

However, our behavioral response results suggested an *increase* in driving from the switch to per-mile insurance. If this increase is due to an income effect, such an effect would lessen as the insurance market unravels, for per-mile insurance will begin drawing in higher-VMT households, and thus the reduction in overall insurance premia would be small. By the time the insurance market is fully unraveled, the zero-profit condition would suggest that there would be no net income effect on average across all drivers, and thus the behavioral response would be more likely to revert to a value similar to the -0.1 elasticity. If the increase in driving we observe is due to other behavioral explanations, it is more difficult to hypothesize what the behavioral response would be. We plan to continue exploring this further in future iterations of the draft.

A decline in VMT would imply a decline in greenhouse gas and local air pollutant emissions. Fuel economy changes depending on the speed of driving and driver behavior (e.g., acceleration from lights). For illustrative calculations, we assume that these factors are held roughly constant, although we recognize that less congestion, the average speed may increase. Higher speeds improve fuel economy and reduce emissions up to around 55 miles per hour, so it is possible that the change in speeds increases or reduces emissions. Holding fuel economy constant, carbon dioxide emissions from driving scale linearly with VMT. A roughly 3% decline in VMT would also imply a roughly 3% decline in carbon dioxide emissions. Local air pollutant emissions, such as sulfur dioxide, nitrogen oxides, carbon monoxide, and particulate matter are more complicated, as vehicles have air pollutant control systems that (for most cars) largely break the link between air pollutants and driving. However, local air pollutants would likely decrease at least somewhat due to per-mile insurance. An increase in speeds from reduced congestion would reduce carbon monoxide and particulate matter that occur in stop-start traffic (emissions control systems typically do not work as efficiently at low speeds). As well, local air pollutant emissions are roughly proportional to miles driven for older vehicles and vehicles for which the pollutant control systems are not working (Merel et al. 2014). Thus, a 3% decline in driving would correspond to a roughly 3% decline in local air pollutant emissions for many of these vehicles.

6 Conclusions

This paper explores the possibility of unraveling in the insurance market due to the introduction of per-mile insurance capturing some of the lowest-risk customers. Our evidence from a major per-mile startup suggests that customers who select into per mile insurance drive much less than other customers. We might also expect a reduction in driving due to the increased the cost per mile of driving from per-mile insurance, but the evidence from testers who choose to become customers suggests that upon becoming customers, the length of trips decreases, but the number of trips increases, so that overall driving actually increases. One plausible explanation for this finding is an income effect due to the reduction in the overall expenditure on automobile insurance from switching from conventional insurance to per-mile insurance. If this is the explanation, consumers would be better off from per-mile insurance, potentially countering the externalities from the additional driving.

We developed and calibrated a simple model to demonstrate how unraveling in the insurance market could occur as per-mile insurance becomes more widely known. We demonstrate that if per-mile insurance is viewed by consumers similarly to conventional insurance, so that decisions are based solely on insurance premia, then unraveling in the automobile insurance market can occur. Using estimates from the literature on the behavioral response to changes in the cost-per-mile of driving, we show that a change to per-mile insurance would reduce emissions, accidents, and congestion. As each of these externalities from driving are imperfectly priced, policymakers may consider encouraging per-mile insurance in the absence of first-best policies to internalize these externalities. For example, Edlin (2003) suggests mandatory provision of information about the premium per mile on all automobile insurance contracts, which could influence households that do not drive very much to switch to per-mile insurance. This perhaps could facilitate an unraveling of the automobile insurance market, just as we modeled.

There are several extensions of this paper still underway. We are in the process of extending the model to the other three states that MetroMile is working in. Further, we have a request in for data on claims from different insurers from the California Department of Insurance. This request should provide data on insurance premiums that will allow us to estimate a richer choice model for insurance demand. We also are interested in updating our MetroMile data to the present day. Finally, we plan on refining our current illustrative calculations on the implications of per-mile insurance with updated estimates and extending these calculations to examine the overall welfare effects of an unraveling of the insurance market.

References

Aron-Dine, A., Einav, L., Finkelstein, A. and Cullen, M. (2015), 'Moral hazard in health insurance: Do dynamic incentives matter?', *Review of Economics and Statistics* **97(4)**, 725–741.

Bordoff, J. and Noel, P. (2008), 'The impact of pay-as-you-drive auto insurance in california', *The Hamilton Project Working Paper* pp. 1–19.

Coady, D., Parry, I. and Shang, B. (2018), 'Energy price reform: Lessons for policymakers', *Review of Environmental Economics and Policy* **12(2)**, 197–219.

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

Cohen, A. and Dehejia, R. (2004), 'The effects of automobile insurance and accident liability laws on traffic fatalities', *Journal of Law and Economics* **47(2)**, 357–393.

Cohen, A. and Einav, L. (2007), 'Estimating risk preferences from deductible choice', *American Economic Review* **97(3)**, 745–788.

Cutler, D. and Reber, S. (1998), 'Paying for health insurance: The trade-off between competition and adverse selection', *Quarterly Journal of Economics* **113(2)**, 433–466.

Davis, L. and Sallee, J. (2019), 'Should electric vehicle drivers pay a mileage tax?', *Energy Institute at Haas Working Paper 301* .

Edlin, A. (2003), Per-mile premiums for auto insurance, *in* R. Arnott, B. Greenwald, R. Kanbur and B. Nalebuff, eds, 'Essays in Honor of Joseph E. Stiglitz', MIT Press, Cambridge, Massachusetts, chapter 5, pp. 53–82.

Einav, L., Finkelstein, A. and Cullen, M. (2010), 'Estimating welfare in insurance markets using variation in prices', *Quarterly Journal of Economics* **125(3)**, 877–921.

Ferreira, J. and Minikel, E. (2010), ‘Pay-as-you-drive auto insurance in massachusetts’, *Conservation Law Foundation Working Paper* pp. 1–64.

Gillingham, K. (2014), ‘Identifying the elasticity of driving: Evidence from a gasoline price shock in california’, *Regional Science and Urban Economics* **47**, 13–24.

Gillingham, K. (2015), ‘Heterogeneity in the response to gasoline prices: Evidence from pennsylvania and implications for the rebound effect’, *Energy Economics* **52(S1)**, S41–S52.

Glaister, S. and Graham, D. (2002), ‘The demand for automobile fuel: A survey of elasticities’, *Journal of Transportation Economics and Policy* **36(1)**, 1–25.

Harrington, S. and Doerpinghaus, H. (1993), ‘The economics and politics of automobile insurance rate classification’, *Journal of Risk and Insurance* **60(1)**, 59–84.

Hendel, I. and Lizzeri, A. (2003), ‘The role of commitment in dynamic contracts: Evidence from life insurance’, *Quarterly Journal of Economics* **11(1)**, 299–327.

Holland, S., Mansur, E., Muller, N. and Yates, A. (2016), ‘Are there environmental benefits from driving electric vehicles? the importance of local factors’, *American Economic Review* **106(12)**, 3700–3729.

Hymel, K. and Small, K. (2015), ‘The rebound effect for automobile travel: Asymmetric response to price changes and novel features of the 2000s’, *Energy Economics* **49**, 93–103.

Jacobsen, M. (2016), ‘Fuel economy and safety: The influences of vehicle class and driver behavior’, *American Economic Journal: Economic Policy* **5(3)**, 1–26.

Knittel, C. and Sandler, R. (2016), ‘The welfare impact of second-best uniform pigouvian taxation: Evidence from transportation’, *American Economic Journal: Economic Policy* **10(4)**, 211–242.

Langer, A., Maheshri, V. and Winston, C. (2017), 'From gallons to miles: A disaggregate analysis of automobile travel and externality taxes', *Journal of Public Economics* **152**, 34–46.

Litman, T. (2011), 'Distance-based vehicle insurance feasibility, costs, and benefits', *Victoria Transport Policy Institute Working Paper* pp. 1–90.

Merel, P., Smith, A., Williams, J. and Wimberger, E. (2014), 'Cars on crutches: How much abatement do smog check repairs actually provide?', *Journal of Environmental Economics and Management* **67(3)**, 371–395.

Newbury, D. (1990), 'Pricing and congestion: Economic principles relevant to pricing roads', *Oxford Review of Economic Policy* **6(2)**, 22–38.

Nichols, B. and Kockelman, K. (2015), 'Pay-as-you-drive insurance: Its impacts on household driving and welfare', *Transportation Research Record* **2450**, 76–82.

Parry, I. (2005), 'Is pay-as-you-drive insurance a better way to reduce gasoline than gasoline taxes?', *American Economic Review* **95(2)**, 288–293.

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

Rea, S. (1992), Insurance classifications and social welfare, *in* G. Dionne, ed., 'Contributions to Insurance Economics', Kluwer, Boston, Massachusetts.

Rothschild, M. and Stiglitz, J. (1976), 'Equilibrium in competitive insurance markets: An essay on the economics of imperfect information', *Quarterly Journal of Economics* **90(4)**, 629–649.

Safirova, E., Gillingham, K. and Houde, S. (2007), 'Measuring marginal congestion costs of urban transportation: Do networks matter?', *Transportation Research A* **41(8)**, 734–749.

Souza-Rodrigues, E. (2019), 'Deforestation in the amazon: A unified framework for estimation and policy analysis', *Review of Economic Studies* **forthcoming**.

Wenzel, T. and Fujita, K. S. (2018), 'Elasticity of vehicle miles of travel to changes in the price of gasoline and the cost of driving in texas', *LBNL Working Paper 2001138* .

Tables & Figures

Table 1: Summary Statistics of MetroMile Data

	(1) CA	(2) IL	(3) OR	(4) WA	(5) Total
Panel A: Number of Observations					
Number of Vehicles	3623	965	1208	1174	6970
Number of Tester Vehicles	3623	804	692	952	6071
Number of Customer Vehicles	257	293	710	416	1676
Vehicles that Switched from Tester to Customer	257	132	194	194	777
Panel B: Mean Number of Miles Driven					
Daily VMT for Tester Vehicles	27.9	26.1	17.6	19.4	25.2
Daily VMT for Customer Vehicles	15.5	16.3	13.7	13.2	14.3
Daily Trips for Tester Vehicles	2.7	2.8	2.4	2.4	2.6
Daily Trips for Customer Vehicles	1.9	2.0	2.1	1.9	2.0
Miles per Trip for Tester Vehicles	10.6	9.8	7.3	8.1	9.7
Miles per Trip for Customer Vehicles	8.9	9.0	7.0	7.4	7.8
Miles per Trip for Testers that Became Customers	8.9	9.1	6.6	7.2	8.0
Fraction of Free Miles for Tester Vehicles	0.04	0.05	0.03	0.01	0.04
Fraction of Free Miles for Customer Vehicles	0.04	0.06	0.04	0.01	0.04

Notes: Anonymized data from MetroMile as of 2015. Means in Panel B are taken over all observations, where an observation is a vehicle.

Table 2: Quantiles of Mean Daily VMT

	(1) Test Drivers	(2) Testers Who Became Customers
1%	0.7	0.2
5%	3.4	1.4
10%	5.9	2.6
25%	11.7	6.1
50%	20.2	11.6
75%	33.3	18.4
90%	49.3	28.5
95%	63.2	34.8
99%	96.4	62.8

Notes: These data are for the full sample, but similar differences exist in each state.

Table 3: Effect of Per-Mile Insurance on Driving by Testers Who Become Customers

	(1)	(2)	(3)	(4)	(5)	(6)
	log(VMT/Trip)		Trips/Day		log(VMT/Day)	
$1(\text{customer})_{it}$	-0.029	-0.030	0.175	0.175	0.073	0.071
	(0.009)	(0.009)	(0.038)	(0.038)	(0.027)	(0.027)
Vehicle fixed effects	Y	Y	Y		Y	Y
Day-of-week fixed effects	Y	Y	Y	Y	Y	Y
Month-of-year fixed effects	Y	Y	Y	Y	Y	Y
Balanced Panel		Y		Y		Y
R-squared	0.14	0.14	0.30	0.30	0.28	0.28
N	1,519,987	1,322,522	651,650	545,003	651,650	545,003

Notes: Dependent variable is listed above each of the columns, where VMT refers to vehicle-miles-traveled. An observation is a vehicle-day in columns (3)-(6). An observation is a trip in columns (1) and (2). To account for zeros, we are using the log of of the VMT variable plus one in columns (1)-(2) and (5)-(6). Standard errors clustered at the vehicle level in parentheses.

Table 4: Demand Estimation for Insurance

	(1) 5-digit zip code	(2) 4-digit zip code	(3) 3-digit zip code
$\delta_{Per-Mile}$	-11.59	-11.66	-11.86
	(0.109)	(0.088)	(0.080)
Daily VMT	-0.007	-0.008	-0.004
	(0.003)	(0.003)	(0.002)
State Farm premium	0.020	0.021	0.022
	(0.001)	(0.001)	(0.001)
MetroMile premium	-0.010	-0.007	-0.006
	(0.001)	(0.001)	(0.001)
R-squared	0.14	0.14	0.30
MetroMile Customers	4,316	6,191	7,308
N (Vehicles)	34,523,663	34,523,663	34,523,663

Notes: Logit estimations for the choice of being a MetroMile customer. An observation is a vehicle. Robust standard errors in parentheses.

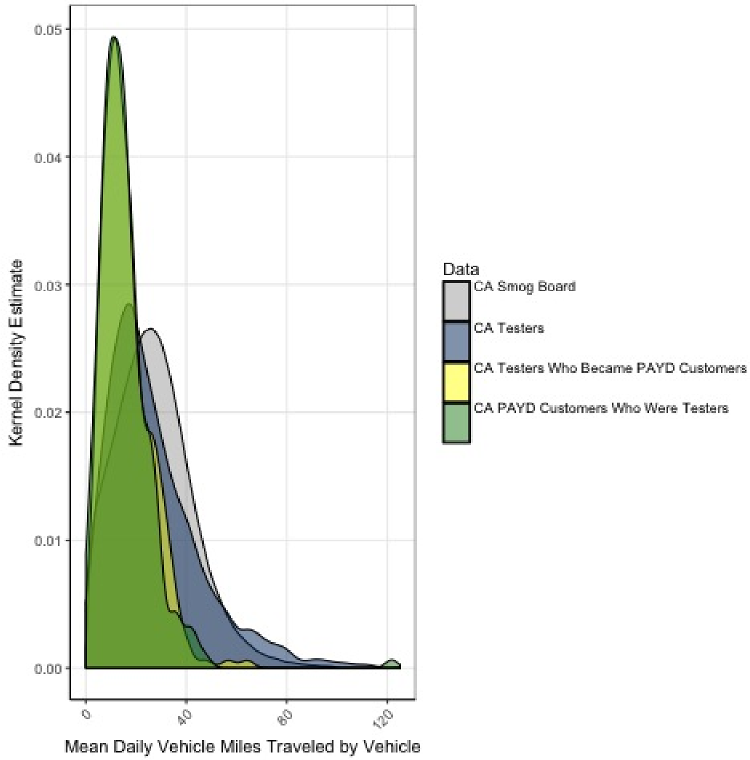


Figure 1: Comparison of California Air Resources Board and California MetroMile Driver Populations. Note that 19 Metromile testers have mean daily VMT greater than 125 with one driver having a mean daily VMT of 234.

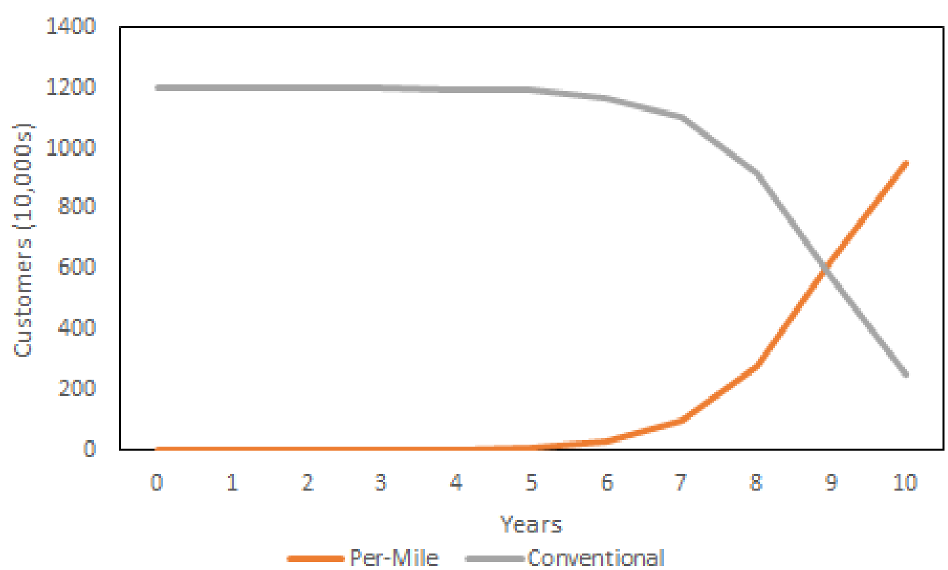


Figure 2: Unraveling in the California Insurance Market

ONLINE APPENDIX

A More Details on MetroMile

In 2011, the pay-per-mile auto insurance company Metromile was founded hoping to entice private vehicle owners away from the flat-rate monthly premiums traditionally offered by the auto insurance industry. Metromile provides customers with GPS-enabled devices to install in each vehicle covered by the insurance policy.[13] Every time a customer takes a trip in their vehicle, the GPS device, branded "Pulse", records the trip characteristics including trip mileage, duration, and timestamps for the trip's start and end. In addition to customers, Metromile also provided this device to over 6,000 testers who were not customers. Metromile continued to market its policies to these testers, and over 1,500 became customers.

The following two figures display the MetroMile Pulse unit and the introductory letter sent to testers:

[13]http://www.metromile.com/terms-conditions-pulse-device

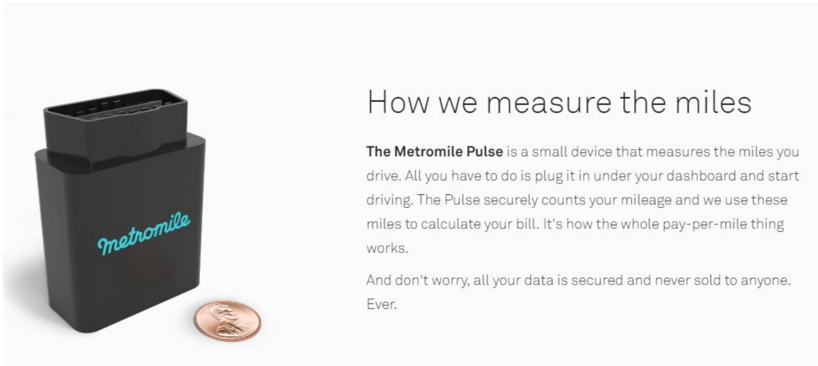
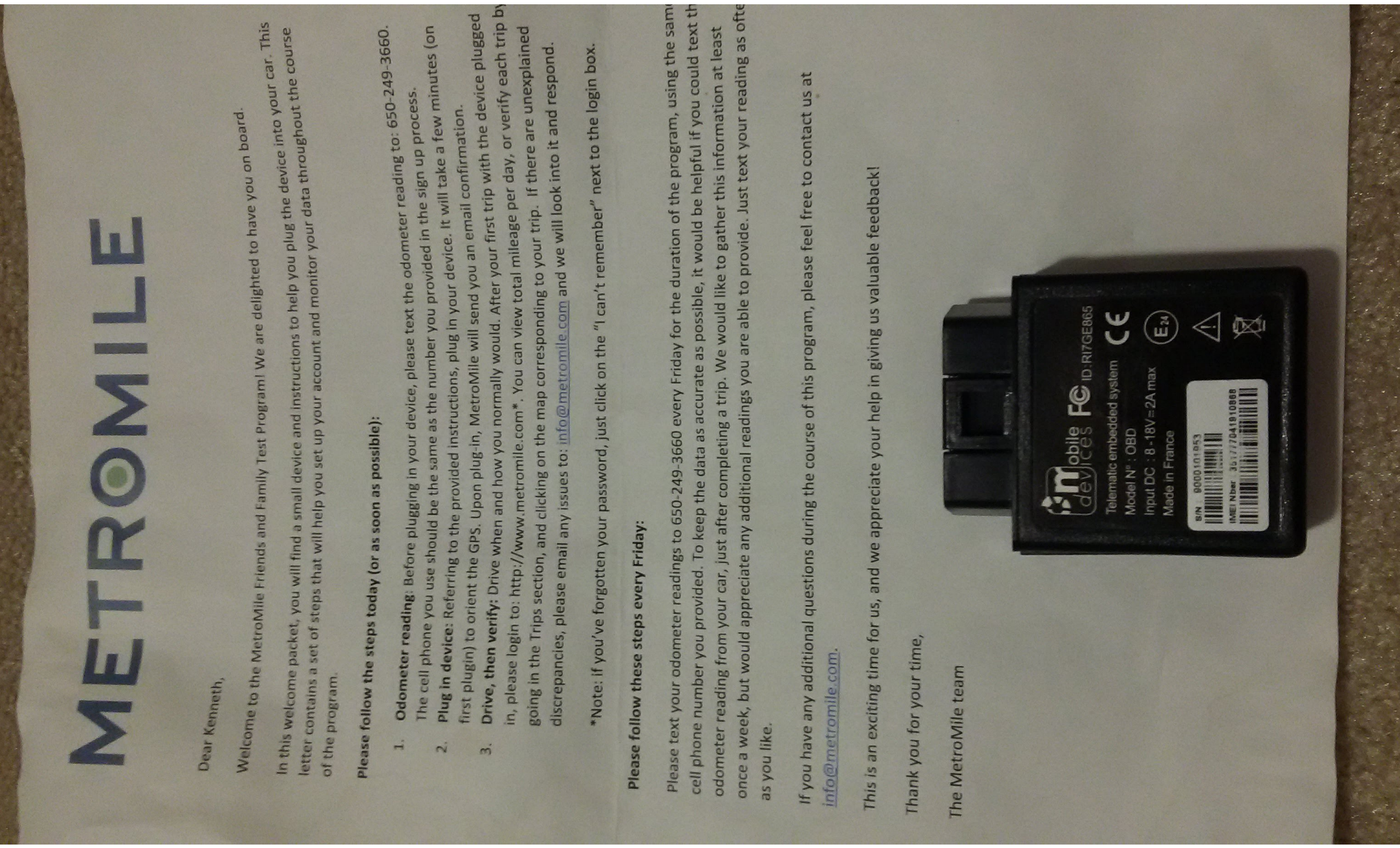


Figure A.1: MetroMile explanation of the Pulse unit.

Note that from the second figure it shows that there was a request to text odometer readings at the end of every week so that MetroMile could confirm the Pulse unit worked. This may have made the number of miles driven more salient for many of the testers and supports the behavioral explanations given in the text that there is a short-run effect that wears off. Note that after about a year, MetroMile stopped requesting weekly odometer readings.



METROMILE

Dear Kenneth,

Welcome to the MetroMile Friends and Family Test Program! We are delighted to have you on board.

In this welcome packet, you will find a small device and instructions to help you plug the device into your car. This letter contains a set of steps that will help you set up your account and monitor your data throughout the course of the program.

Please follow the steps today (or as soon as possible):

1. **Odometer reading:** Before plugging in your device, please text the odometer reading to: 650-249-3660. The cell phone you use should be the same as the number you provided in the sign up process.
2. **Plug in device:** Referring to the provided instructions, plug in your device. It will take a few minutes (on first plugin) to orient the GPS. Upon plug-in, MetroMile will send you an email confirmation.
3. **Drive, then verify:** Drive when and how you normally would. After your first trip with the device plugged in, please login to: http://www.metromile.com*. You can view total mileage per day, or verify each trip by going in the Trips section, and clicking on the map corresponding to your trip. If there are unexplained discrepancies, please email any issues to: info@metromile.com and we will look into it and respond.

*Note: If you've forgotten your password, just click on the "I can't remember" next to the login box.

Please follow these steps every Friday:

Please text your odometer readings to 650-249-3660 every Friday for the duration of the program, using the sam cell phone number you provided. To keep the data as accurate as possible, it would be helpful if you could text th odometer reading from your car, just after completing a trip. We would like to gather this information at least once a week, but would appreciate any additional readings you are able to provide. Just text your reading as ofte as you like.

If you have any additional questions during the course of this program, please feel free to contact us at info@metromile.com.

This is an exciting time for us, and we appreciate your help in giving us valuable feedback!

Thank you for your time,

The MetroMile team

Figure A.2: Introductory letter for the tester program that includes the MetroMile Pulse GPS unit.

B More Details on Data Cleaning

B.1 MetroMile Data

Under a non-disclosure agreement, Metromile provided our research team with trip-level data for over 7,000 tester and customer vehicles in California, Oregon, Illinois, and Washington between November 11, 2013, and April 7, 2015.[14] The data provided by Metromile includes three components: an accounts database, a vehicle database, and trips data based on each vehicle's individual trips. The accounts database includes information about tester and customer policies including policy start and end date, cancellations, and policy modifications, such as adding an additional vehicle. The accounts database also allows me to determine when an account switches from a tester to a customer status. The vehicles database includes vehicle characteristics including vehicle make, model, and vintage. Metromile provided a first round of these three datasets in June 2014 and a second round in April 2015. The two datasets were appended together.

As shown in Table A.1 Metromile had records for at least 117,399 accounts and 131,736 unique vehicles as of April 2015, of which 18,612 vehicles were associated with customer accounts. Metromile provided trips data for 7,208 of these vehicles, of which 2,073 were customers. In the process of merging and appending these various datasets, we identified some errors and inconsistencies, the most important of which we describe below. Fortunately, less than four percent of vehicles are removed from the sample as a result of these errors and inconsistencies. The final sample contains 6,970 vehicles, of which 6,071 were testers and 1,676 were customers. There were 5,034 tester vehicles that never became customers and 288 customer vehicles that were never testers. The records in the accounts database indicate that 1,648 of the vehicles were associated with accounts that switched from tester to customer. However, only 777 of the vehicles have trips observations as both a tester and as a customer.

In checking for errors and inconsistencies in the data, we first considered each of the

[14] Additional data going back to 2012 is available for a subset of the drivers and could be considered in future analyses.

			California	Illinois	Oregon	Washington	Other & Unknown	All States
[1]	**Raw Sample Description**	Number of Accounts in Accounts Database	70,548	16,143	13,919	16,824	41	117,399
[2]		Number of Vehicles in Vehicles Database	79,226	17,547	15,762	19,246	26	131,736
[3]		Number of Tester Vehicles in Vehicles Database	79,163	17,503	15,144	19,074	25	130,838
[4]		Number of Customer Vehicles in Vehicles Database	10,224	2,387	3,341	2,677	1	18,612
[5]		Number of Vehicles in Vehicles Database That Are Neither Customer or Tester	2	0	47	1	0	49
[6]		Number of Accounts With Trips Data	3,770	933	1,149	1,112	1	6,953
[7]		Number of Vehicles With Trips Data	3,770	978	1,270	1,198	3	7,208
[8]		Number of Tester Vehicles With Trips Data	3,770	972	1,002	1,154	0	6,887
[9]		Number of Customer Vehicles With Trips Data	346	374	855	504	1	2,073
[10]		Number of Vehicles With Trips Data That Are Neither Customer or Tester	0	0	19	0	1	20
[11]	**Cleaned Sample Description**	Number of Accounts	3,623	920	1,090	1,092		6,724
[12]		Number of Vehicles	3,623	965	1,208	1,174		6,970
[13]		Number of Tester Vehicles	3,623	804	692	952		6,071
[14]		Number of Customer Vehicles	257	293	710	416		1,676
[15]		Number of Tester Vehicles That Never Became Customers According to Accounts Database	3,316	607	415	696		5,034
[16]		Number of Customer Vehicles That Were Never Testers According to Accounts Database	0	5	239	44		288
[17]		Number of Vehicles that Switched From Tester to Customer According to Accounts Database	307	353	554	434		1,648
[18]		Number of Vehicles that Switched From Tester to Customer Observed in Cleaned Data	257	132	194	194		777

Table A.1: Comparison of Raw and Cleaned Sample accounts, vehicles, and trips datasets separately. In general, we resolved conflicts between the June 2014 and April 2015 data by choosing the more recent record. One of the most substantial assumptions was related to the accounts database. Every account, including tester-only accounts, has at least one binding date indicating when the policy took effect (or in the case of testers, when they entered the sample). Of the 15,398 customer accounts in the raw accounts database, 4,911 (over 32 percent) have at least one case where a tester record binds after a customer record. In these cases, we drop the customer records that occurred before a test drive and assume that the entire period prior to the last test drive was all testing. This assumption could be explored in future analysis.

Other errors and inconsistencies are more minor. We removed 42 accounts because of missing data in the field that distinguishes testers from customers and removed 19 accounts because of missing dates in the field that indicates a policy's binding date. We removed 36 accounts where the cancellation date occurred before any policy binding date. One vehicle number was associated with more than one account, so we removed that vehicle. We also removed one trip observation where the year is recorded as 2033.

Then we merged the trips data with the accounts and vehicles databases. The appended trips datasets (provided in June 2014 and April 2015) contain 1,604,917 unique trips. After merging these with the accounts and vehicles characteristics, we remove trips that occurred before the earliest binding date and after any cancellation date in the accounts database, regardless of whether the account was a customer or tester. After all of these corrections and restrictions on the data described above, only 44 vehicles and 28,303 trips were removed from the appended trips data, which represents less than two percent of all trips.

We then performed a final round of checks, adjustments, and restrictions. We removed 4,086 trips (0.003 percent) that covered less than 0.1 miles (rounded to zero in the data). For trips with a duration of zero seconds (rounded down from a value of less than 30 seconds), we assigned the duration to be 30 seconds. This only affected 529 trips. There were 532 trips across 426 vehicles where the implied average speed was greater than 125 miles-per-hour for the duration of the trip, so we removed all of these trips. Finally, we removed 189 vehicles with either missing or conflicting zip codes. The final dataset contains 1,519,987 individual vehicle trips.

C Regression Evidence on Selection by Testers

This section of the appendix provides regression evidence on which MetroMile testers select into being customers. The reason for performing this estimation is that a simple comparison of means of driving by testers and customers misses the fact that the percentage of testers versus customers in the sample changes over time. A regression approach allows us to control for changes over time using fixed effects. We employ a similar equation to 1 only we include variables for whether a driver of a vehicle switches from being a tester to a customer at some point in the sample ($1(\text{switcher})_i$) and a variable for the time period after that switch interacted with being a switcher ($1(\text{switcher})_i \times 1(\text{PostSwitch})_{it}$). This second variable is equivalent to the customer variable in 1. The specification is given as follows:

$$\log(VMT)_{it} = \beta_1 1(\text{switcher})_i + \beta_2 1(\text{switcher})_i \times 1(\text{PostSwitch})_{it} + \mu_i + \eta_t + \delta_t + \varepsilon_{it} \quad (4)$$

Note that in this specification, we include all of the fixed effects as in 1 except we cannot include vehicle fixed effects. Instead, we use state fixed effects (μ_i). Again, we present results using the full cleaned sample and the sample that drops testers or customers that only have one or less observations in either the customer or tester phase (the more balanced sample).

The results of the estimation are given below in Table **??**. We observe clear selection by testers, with a large and statistically significant coefficient on whether the tester is a switcher in all columns. For example, the coefficient for switchers in column (6) indicates that testers who switch have 58% less driving than other testers. After switching, they slightly increase their total driving per day, consistent with Table 3. This again is due to these switchers increasing the number of trips per day after becoming customers, as is shown in columns (3) and (4).

Table A.2: Evidence of Selection by Testers

	(1)	(2)	(3)	(4)	(5)	(6)
	log(VMT/Trip)		Trips/Day		log(VMT/Day)	
$1(\text{switcher})_i$	-0.170	-0.173	-0.679	-0.651	-0.586	-0.579
	(0.016)	(0.019)	(0.050)	(0.056)	(0.035)	(0.040)
$1(\text{switcher})_i \times 1(\text{PostSwitch})_{it})$	-0.050	-0.058	0.101	0.236	0.029	0.112
	(0.014)	(0.016)	(0.049)	(0.056)	(0.034)	(0.038)
Vehicle fixed effects	Y	Y	Y		Y	Y
Day-of-week fixed effects	Y	Y	Y	Y	Y	Y
Month-of-year fixed effects	Y	Y	Y	Y	Y	Y
Balanced Panel		Y		Y		Y
R-squared	0.02	0.02	0.05	0.05	0.05	0.05
N	1,448,173	1,322,522	611,308	545,003	611,308	545,003

Notes: Dependent variable is listed above each of the columns, where VMT refers to vehicle-miles-traveled. An observation is a vehicle-day in columns (3)-(6). An observation is a trip in columns (1) and (2). To account for zeros, we are using the log of of the VMT variable plus one in columns (1)-(2) and (5)-(6). Standard errors are clustered at the vehicle level.

D Further Details on Matching

Matching can help alleviate the concern that vehicles in the test group may have very different characteristics and follow different trends than vehicles that switch to Metromile insurance. Thus, we perform a series of different matches based on the zip code of each MetroMile customer and cars of the same make-model-model year in the same zip code. In many cases, we have a perfect match with a single vehicle in the inspection data that fits the criteria, but in many others there are multiple vehicles in the inspection data. When there are multiple vehicles, we take the average VMT from the odometer readings of the vehicles that fit the criteria. In some cases, we cannot find a vehicle that fits the criteria. In these cases, we drop the vehicle from the regression. This motivates us to perform the match at less disaggregated levels of geographic aggregation. For example, instead of using the full 5-digit zip code, we use the 4-digit zip code or 3-digit zip code. This brings in further vehicles into our sample. The trade-off is that we may be incorrectly matching some vehicles. Fortunately, our results appear robust to these restrictions.