Pay-As-You-Drive Auto Insurance in Massachusetts

A Risk Assessment and Report on Consumer, Industry and Environmental Benefits

November 2010

Authored by:

MIT Professor Joseph Ferreira, Jr.

&

Eric Minikel

Commissioned by:

Conservation Law Foundation

&

Environmental Insurance Agency

Executive Summary

Each year, Massachusetts drivers are driving more, and with each additional mile driven, levels of global warming pollution rise. The prospect of tying auto insurance rates to miles driven, called Pay-As-You-Drive auto insurance (PAYD), offers the opportunity to improve the accuracy of auto insurance rating while reducing vehicle miles traveled (VMT) and corresponding accident costs as well as reducing fuel consumption and greenhouse gas emissions.

Pay-As-You-Drive auto insurance is a win for consumers, insurers and the environment:

- Consumers can save money; they will only pay for the coverage needed based on how much they drive.
- Insurers can improve the accuracy of their rating plans while providing an incentive to reduce the number and cost of auto accident claims.
- The environment will benefit from the reduction in driving that PAYD incentivizes – less driving means reduced fuel usage and lower greenhouse gas emissions.

The Conservation Law Foundation (CLF) and the Environmental Insurance Agency commissioned a study to assess the risk-mileage relationship using actual insurance claims information in Massachusetts. This study ("Ferreira and Minikel 2010") offers the largest disaggregated analysis to date of the risk-mileage relationship and the actuarial basis for PAYD. The work analyzes data on $502 million worth of claims on almost 3 million cars driven an aggregate of 34 billion miles. The study confirms the statistical soundness of pay-as-you-drive auto insurance pricing and indicates that the PAYD approach would result in significant reductions in miles driven, green house gas emissions, and auto accident losses without adverse equity impacts to drivers.

PAYD Saves Money and is a More Accurate and Fairer Method to Price Auto Insurance

- By basing premiums at least partly on mileage, PAYD provides individual policyholders more control over their insurance costs and more accurate premiums for the type of driving they do.
- PAYD pricing reduces inequities by eliminating the subsidies low-mileage drivers currently pay for high-mileage drivers in the traditional pricing system.
- Even though suburban and rural car owners tend to drive more miles than urban car owners, their per mile charges would be lower. If they drive less than the average for their area, they would pay less for actuarially-priced PAYD insurance than they do today under the existing system.

PAYD Reduces Vehicle Mileage Traveled (VMT), Accidents and Fuel Consumption by 5-10%

- Switching all Massachusetts drivers to pure per mile auto insurance pricing would reduce mileage, accident costs and fuel consumption by about 9.5%. An alternative model with a flat yearly rate plus per mile pricing after the first 2,000 miles would reduce these measures by about 5%.
- These reductions could range between 3 and 14% depending on a number of variables like fuel prices. But even the study's lowest plausible VMT reduction (2.7%) would save more than a billion miles annually and millions of tones of GHG.
- Negative impacts of congestion will decrease under PAYD, particularly for urban driving.

Overview

Pay-As-You-Drive (PAYD) auto insurance converts the traditional lump-sum yearly insurance payment into a cents-per-mile rate, thus providing drivers with an opportunity to save money and an incentive to reduce mileage. For decades, researchers have touted PAYD's potential to reduce automobile accidents, congestion and greenhouse gas emissions while also improving equity over the current system. It appears that PAYD carries large potential benefits both for individual policyholders and for society as a whole, yet it has seen limited application to date, due in large part to economic and regulatory barriers. Congestion, pollution and some fraction of accident costs are all externalities, so any individual insurance company would see just a portion of the benefits of PAYD, even while incurring the full transaction and monitoring costs. Meanwhile, many state insurance regulations either prohibit or inhibit PAYD.

However, new technology has lowered transaction and monitoring costs and awareness of global warming has sparked a new state-level push for ways to reduce VMT without increasing consumer costs. From a policy standpoint, PAYD seems an increasingly appealing and feasible prospect, yet from an actuarial standpoint, it is still in need of further study. While it is clear that risk increases with mileage, the precise nature of the relationship at the individual level is not well understood. Most research on PAYD to date has examined mileage and risk data at a highly aggregated level, comparing, for instance, across U.S. states.

This report offers the largest disaggregated study to date of the risk-mileage relationship and the actuarial basis for PAYD. Linking recently released insurance and mileage data from the Commonwealth of Massachusetts for the 2006 policy year, we analyze the correlation between annual miles traveled and insurance risk for over three million individual vehicles insured on private passenger insurance policies and categorized by rating class and territory.

We begin by matching 2006 policy year earned exposure data to claims data for bodily injury and property damage liability, plus personal injury protection coverages— i.e., the compulsory, and therefore fairly uniform, types of insurance coverage. Next we create estimates of each vehicle's annual miles traveled based on odometer readings from mandatory safety checks. In addition, we obtain fuel economy estimates for each vehicle thanks to commercial Vehicle Identification Number (VIN) decoding services provided by VINquery.com. In all, we are able to analyze data on $502 million worth of claims on 2.87 million car years of exposure covering vehicles driven an aggregate of 34 billion miles.

We find a strong relationship between miles driven and auto accident claims frequency and loss costs. This relationship between risk and mileage is less than linear when all vehicles are considered together, but it becomes considerably more linear when class and territory are differentiated. Using pure premium as our measure of risk, we regress risk on mileage using a variety of models. We find that mileage is a highly significant predictor of risk but, used alone, provides less explanatory power than traditional class and territory factors, so a single, universal per mile insurance rate for all drivers would be inappropriate. However, a combined model using mileage along with class and territory groupings explains more risk variation than a similar model without mileage. In fact, mileage gains in its own explanatory power when used in conjunction with class and territory, probably because class and territory provide some control on where and how the miles are traveled. This suggests that telematic pricing that varies per mile rates based on when, where and how miles are traveled could improve actuarial accuracy even further and may even obviate the need for some traditional insurance rating factors.

Absent telematic pricing, PAYD insurance is most likely to be practical and effective with differential rates for customers in different classes and territories. Since low-mileage policyholders have higher per mile risk, and since there are fixed costs associated with writing an insurance policy, a pricing scheme where users purchase 2,000 miles for a flat yearly fee and then pay per mile thereafter may be more realistic and statistically sound than a strictly per mile pricing scheme.

Though PAYD is unlikely to eliminate existing class and territory distinctions, it appears to have positive equity implications. PAYD would improve fairness by shifting weight in insurance pricing towards an individually controllable factor, *mileage*, rather than involuntary groupings, and by reducing or eliminating the cross-subsidy from low to high mileage drivers. For low-income households, PAYD would create an opportunity to save money by choosing to reduce mileage, would make low-mileage car ownership more feasible, and would reduce the toll of auto-related externalities on the non-car owning poor.

Extrapolating from the per mile pure premiums we calculate for compulsory coverages, we estimate retail prices for full coverage for each class and territory. Under strictly per mile pricing, we estimate an average premium of 8.2¢ per mile statewide, ranging from 4.3¢ for the lowest-risk customers to 37¢ for the highest-risk customers. For statewide fuel economy we observe a 20 mile per gallon average, which at the current gasoline price of $2.70 translates into about 14¢ per mile. Assuming that drivers currently consider fuel to be the only marginal price of driving an additional mile, a switch to PAYD would represent more than a 50% increase in the perceived per mile price of driving for the average fully insured Massachusetts driver. Our literature review suggests a -0.15 elasticity of miles driven with respect to the marginal per mile price. Based on this, we estimate a 9.5% reduction in VMT if all drivers in Massachusetts switched to a strictly per mile PAYD insurance plan, and a 5.0% reduction if all drivers switched to a plan having 2,000 miles bundled into a yearly fee plus per mile pricing thereafter[1]. Fuel reductions are almost exactly proportional, as we find that average fuel economy exhibits almost no variation by class, territory or annual mileage. Depending on a number of variables, including the amount paid per mile, the types of coverage provided, and the availability of alternative modes of transportation to drivers, the VMT and fuel consumption reductions with PAYD could range between 3 and 14%.

Our 9.5% estimate is somewhat lower than what other researchers have estimated, probably because our differentiation of insurance rates by class and territory reveals that the highest-risk territories with the highest theoretical per mile rate already have the lowest annual mileage. The reduction in accidents resulting from the reduced mileage would be similar in percentage, but could be somewhat higher or lower depending upon the relative risk of the forgone miles and the additional benefit of reduced congestion and multi-car accident risk.

[1] The reduction from 9.5% to 5% for the 2K+per-mile scheme is due not only to the flat charge for the first 2000 miles but also to a lower, and more statistically justifiable, per mile price.

Overall, the risk analysis in this study confirms the statistical soundness of pay-as-you-drive auto insurance pricing and indicates that, if the per-mile charges are sufficiently timely and certain, then the approach would result in significant reductions in miles driven, green house gas emissions, and auto accident losses without adverse horizontal or vertical equity impacts.

Acknowledgements

This report was prepared by Joseph Ferreira, Jr. and Eric Minikel for the Conservation Law Foundation (CLF), with the support and participation of the CLF through the support of the Surdna Foundation, the Transportation Alliance and the Environmental Insurance Agency, and in collaboration with the Massachusetts Institute of Technology. The analysis uses a unique dataset of vehicle mileage and auto insurance exposure and claims for millions of private passenger Massachusetts vehicles. The dataset was made available to the public last March by the Massachusetts Executive Office of Energy and Environmental Affairs (EOEEA). MassGIS, the State GIS Agency, obtained the odometer readings from Registry of Motor Vehicle safety inspection data and the auto insurance claims and exposure data from the state's Commonwealth Automobile Reinsurer (CAR). MassGIS then integrated and anonymized these data to prepare a cross-referenced extract for public release[2]. The public dataset was further processed to prepare an analytic dataset for policy year 2006 by Prof. Joseph Ferreira, Jr. at the Massachusetts Institute of Technology (MIT) with partial support through University Transportation Center Region One research grant MITR22-5. This additional processing was required to convert the raw policy and claims transaction data into net paid losses and reserves (as of December 2008) and then match each claim (for bodily injury and property damage liability or personal injury projection) to the appropriate vehicle and earned exposure month. This processing of the raw CAR data was done for all policy and claim transactions involving vehicles insured during policy year 2006.

The analysis, opinions, and conclusions in this report are solely those of the authors and not necessarily those of the Conservation Law Foundation, the Transportation Alliance, MIT, EOEEA, or any other organization involved in the data preparation and analysis, or for whom the authors have worked.

The report and more information about PAYD may be found at CLF's website: http://www.clf.org/work/HCEJ/PAYD. A downloadable zip file with the Appendix 3 Analytic Dataset is available at MIT's website: http://mit.edu/jf/www/payd.

[2] Public notice of the dataset availability is posted on the EOEEA website at: http://www.mass.gov/?pageID=eoeeaterminal&L=3&L0=Home&L1=Grants+%26+Technical+Assistance&L2=Data+Resources&sid=Eoeea&b=terminalcontent&f=eea_data-resources_2005-08-auto-insur-data&csid=Eoeea

Biographic Sketch of Authors

Joseph Ferreira is Professor of Urban Planning and Operations Research in MIT's Urban Studies and Planning Department where heads the Urban Information Systems group and is also Associate Department Head. His undergrad and PhD degrees are also from MIT (in electrical engineering and operations research). Prof. Ferreira teaches analytical methods (including probability and statistics) and computer-based modeling at MIT. His research interests involve the use of geospatial services and interactive spatial analysis tools to model land use and transportation planning, build sustainable information infrastructures for supporting urban and regional planning, and develop decision support systems for assessing and managing risk. Prof. Ferreira has published widely on risk assessment as well as on urban planning uses of geographic information systems (GIS) and database management tools. He is a past-president of URISA (the Urban and Regional Information Systems Association).

Professor Joseph Ferreira has 40 years experience with risk assessment in auto insurance dating back to his work on the staff of the U.S. Dept. of Transportation Federal Auto Insurance and Compensation Study, 1968-1970. He served on the Mass State Rating Bureau in the Division of Insurance while on leave from MIT 1976-1978 and he has testified at numerous auto insurance rating and regulation hearings in Massachusetts and several other states. He has consulted for insurance companies as well as non-profits and Insurance and Consumer Affairs departments. Presently, he consults for the Plymouth Rock Assurance Corporation in Massachusetts on risk assessment and for the Transportation Alliance on this mileage-based auto insurance study.

Eric Minikel holds a Master in City Planning and M.S. in Transportation from the Massachusetts Institute of Technology and a B.S. in Chinese Language and Literature from the University of Wisconsin-Madison. While at M.I.T. he focused on transportation and land use economics and road safety. He received two awards for a paper on the land value of curb parking and wrote a thesis analyzing bicyclist collision rates on different street types. He now works as a Transportation Planner at IBI Group in Boston.

Contents

Introduction

In the years since economist William Vickrey first observed that lump-sum auto insurance pricing does not present users with a marginal insurance cost for each mile driven (Vickrey 1968), various proposals have been made for mileage-based pricing. In recent years, a model known as Pay-As-You-Drive (PAYD)[3], in which users pay a cents-per-mile rate based on actual mileage traveled, has gained particular currency in the literature. By confronting users with a marginal price for each mile driven in lieu of a yearly lump sum payment, PAYD offers the prospect of reducing vehicle miles traveled (VMT) and corresponding accident costs along with fuel consumption and greenhouse gas emissions.

Despite these benefits, PAYD has seen almost no real world implementation to date. One reason is doubtless the economic barriers elaborated by Edlin (2002). A large component of accident costs are external, as are congestion and greenhouse gas emissions, so the social benefits to reducing VMT are large. Yet insurance companies will benefit only from the reduction in their own customers' accident costs. This internal benefit is considerably smaller than the social benefit and might not outweigh transaction and monitoring costs.

Another barrier is state insurance regulations which in various ways hinder or prohibit PAYD policies. For example, Tennessee bans retrospective pricing schemes (Guensler et al 2002, 7), and California forbids the use of mileage *verification* as a rating factor (Bordoff and Noel 2008, 17). In addition to outright bans, insurance companies may be deterred simply by the specter of close regulatory scrutiny for any novel pricing plan such as PAYD (Ibid, 18). Finally, the insurance industry is heavily regulated and the resulting market entry barriers may keep new companies from entering the market to offer PAYD (Ibid, 18).

An additional reason why regulators might be slow to allow PAYD, and insurance companies slow to adopt it, is that the relationship between risk and mileage is still not well understood. While it is clear that higher mileage means higher risk, there have been few large-scale, highly disaggregated studies of the relationship between risk and mileage while controlling for other factors.

The Commonwealth of Massachusetts has recently released insurance policy and claims data together with odometer reading data at the individual vehicle level for the entire state[4]. The size and detail of

---

[3] PAYD, as we use the term here, indicates a *concurrent* pricing scheme in which users are directly charged a per mile fee for each mile driven. We do not consider so-called Pay-At-The-Pump (PATP) schemes in which insurance is added to the price of gasoline, nor *prospective* or "mileage rate factor" schemes in which past mileage is simply given a stronger weight in the setting of per car year insurance rates. We do discuss more elaborate GPS-based schemes which charge variable rates depending on when, where and how a car is driven, but we generally refer to these separately.

[4] In March, 2010, the Massachusetts Executive Office of Energy and Environmental Affairs (EOEEA) released a DVD containing 1.3 GB of compressed data from the state's Commonwealth Automobile Reinsurer (CAR) and the Registry of Motor Vehicles (RMV). The CAR data included several years of auto insurance policy and claims transaction records for all insured private passenger vehicles in the state. The RMV data included odometer readings from all state-mandated annual safety inspections of all private passenger vehicles. Public notice of the availability of the data is posted on the EOEEA website at: http://www.mass.gov/?pageID=eoeeaterminal&L=3&L0=Home&L1=Grants+%26+Technical+Assistance&L2=Data+Resources&sid=Eoeea&b=terminalcontent&f=eea_data-resources_2005-08-auto-insur-data&csid=Eoeea

this dataset[5], unprecedented in the PAYD literature, allow us to conduct a comprehensive analysis of the correlation between mileage and insurance risk on 2.87M car years of exposure. We use Poisson and linear regressions to characterize the correlation between mileage and risk and identify a role for mileage in insurance rating. We then consider possible pricing schemes and their implications for the adoption of PAYD. We analyze likely equity impacts and then model the expected reduction in VMT and fuel use if all drivers in Massachusetts switched to PAYD.

[5] For further details about the data sources and their usefulness in examining the statistical credibility of mileage-based rating, see the April 4, 2007, testimony of Joseph Ferreira, Jr. at a public hearing of the Joint Committee on Financial Services of the Massachusetts General Court (Ferreira, 2007).

Data

We base our analysis on the most extensive possible dataset linking insurance policies with mileage data for private passenger vehicles in the Commonwealth of Massachusetts. We link data for all insurance policies and associated claims in the 2006 policy year to mileage estimates that we create based on odometer readings at mandatory safety checks. In all, we are able to analyze data for 2.87M car years of exposure. This section describes in detail the datasets used in our study.

Insurance data

Commonwealth Automobile Reinsurers (CAR), an industry-operated entity created by Massachusetts state law, collects ratemaking data for all insurance policies issued in the state in accordance with the state-mandated plan for statistical reporting of auto insurance policies and loss claims. CAR has released extracts from their policy and claims records for 2004 through 2008 which detail the type of policy issued for each vehicle, along with claims tables where each row represents a transaction in which a claim was filed, paid or adjusted. These exposure and claims tables can be linked by Vehicle Identification Number (VIN), policy effective date and an anonymized policy ID. (These anonymized IDs were generated by the state agency, MassGIS, to protect the privacy and confidentiality of individual policyholders and companies).

For the purposes of this study, we examine only data from policy year 2006, which means policies issued with an effective date beginning during calendar year 2006. Each such Massachusetts auto insurance policy has a 12 month term. Our reasons for doing so, along with an explanation of some difficulties encountered along the way and how we handled them, are elaborated in Appendix 1. We also limit our study to compulsory bodily injury (BI) and property damage (PD) liability and personal injury protection (PIP) coverages, because these forms of coverage are fairly standard across different vehicles, while other types of insurance are optional.

In analyzing insurance claim data, we use two measures of risk: claim frequency and pure premium. Claim frequency, as we use the term here, refers to the number of *incidents*[6] per 100 car-years of exposure that resulted in one or more claims filed. We also compute pure premiums—actual losses paid to policyholders or incurred as allocated loss adjustment expenses, or still reserved for future payments as of the end of 2008. These 'pure premiums' have not been loaded with other costs that do not vary directly with losses such as insurance agent commissions, premium taxes, certain lawyer costs, and the like. Typically, pure premiums comprise 67% of the total premium charged to a customer. Since our figures are pure premiums for compulsory coverage only, we estimate that a multiplier of 5.5 might be appropriate to translate these premiums into a typical retail price for full coverage (including higher-than-required liability coverage limits plus collision and comprehensive coverages as well as the one-third of costs not included in 'pure premiums'.)

---

[6] In some cases, more than one (anonymized) claim number was assigned to various claims arising from a single 'accident.' We presume this occurred because some insurers assigned separate claim numbers for each vehicle or person filing a claim arising from a single incident. Further discussion of our assumptions and data processing is in Appendix I.

Insurance is rated on, among other factors, driver class categories and rating territories. Though the industry uses finer-grained distinctions, we limit our analysis to broad class and territory groupings. We choose five classes: experienced drivers (mostly "adults"), business use[7], two "youth" classes technically based on years of driving experience (0-3 and 3-6), and senior citizens. The classes are shown with their respective exposure levels, claim frequencies, and pure premiums[8] in Table 1. Adults comprise three-fourths of the market; all of the other classes are small by comparison. As the claim frequency and pure premium data in Table 1 show, the inexperienced driver and business classes are riskier than adults, while senior citizens are slightly less risky. Finer-grained class distinctions involve further dividing the already small "youth" classes by principal versus occasional operators and whether or not the policyholder has received driver training. Further subdividing the smaller classes would unduly complicate our analysis, which takes advantage of large sample sizes, without much likelihood of revealing new risk patterns.

Table 1. Five driver classes with their market share and claims experience

Class	Total Exposure (Car Years)	Percentage of all exposure	Claim frequency per 100 car years	Pure premium per car year (for basic BI/PDL/PIP)
Adult	2,141,668	75%	5.0	$ 160
Business	40,592	1%	6.0	$ 206
< 3yrs exp	114,929	4%	12.7	$ 421
3 -6 yrs exp	115,008	4%	9.6	$ 314
Senior citizen	459,695	16%	4.8	$ 144
ALL	2,871,892	100%	5.5	$ 175

We break the 351 Massachusetts towns and the 10 state-defined regions within the city of Boston into six territories[9] grouped according to their riskiness as measured by the Automobile Insurance Bureau's 2007 relativities estimated for each town and portion of Boston. Our Territory 1 is the least risky and Territory 6 the riskiest. This difference in risk is demonstrated by the claim frequency and pure premium figures in Table 2. Each of the six territories is roughly the same size in terms of car years of insurance "exposure". The numbers are not exactly the same because we assigned whole towns to a particular category and the 2007 exposures in the AIB study differ somewhat from the exposures in our study. The exposure levels are shown in Table 2.

[7] We are examining only vehicles with noncommercial license plates. Of these, only 1% are insured as being for business use, as seen in Table 1. If commercial vehicles were included, the percentage would be far larger.
[8] Note that the claim frequencies are for combined BI, PDL, and PIP coverages only, and the claims costs (paid and reserved) are capped at $25000. See Appendix I for further discussion of our assumptions and data processing.
[9] In 2006, the 351 Mass towns, including the 10 subdivisions of Boston, were ranked by estimated pure premium and grouped, based on estimated risk, into 33 rating territory mandated by the state. For simplicity, we further aggregated the towns and 10 subdivisions into the 6 low to high-rated territories summarized in Table 2.,.

Table 2. Six territories with their market share and claims experience

Territory	Total Exposure (Car Years)	Percentage of all exposure	Claim frequency per 100 car years	Pure premium per car year (for basic BI/PDL/PIP)
1 (low)	547,490	19%	3.9	$ 114
2	557,705	19%	4.5	$ 139
3	322,883	11%	4.8	$ 143
4	577,956	20%	5.4	$ 170
5	533,192	19%	6.6	$ 214
6 (high)	328,249	11%	9.2	$ 314
ALL	2,867,474	100%	5.5	$ 175

Mileage estimates

The Commonwealth's Registry of Motor Vehicles (RMV) has released odometer reading data from mandatory annual safety checks conducted over the last several years. These data were processed into annual mileage estimates by MassGIS and included as part of the publicly released dataset. However, the MassGIS mileage estimates proved unsuitable for our purposes since the between-inspection period used for their mileage estimation was far from the relevant 2006 policy year for many vehicles. Accordingly, we reselected pairings of the original odometer readings to recompute estimates for annual mileage traveled by each vehicle in between safety checks that overlapped the 2006 policy year. This process, and our reasons for creating our own estimates, are described in detail in Appendix 2. Both policy IDs in the CAR insurance datasets and the Massachusetts license plate numbers in the Registry data are anonymized. However, the Vehicle Identification Numbers (VINs) are real, and we take advantage of this fact to link insurance data to annual mileage estimates, and EPA fuel economy data, for each of the millions of vehicles in the dataset.

According to our mileage estimates, the average car in Massachusetts was driven 11,695 miles during its 2006 policy year. Figure 1 shows that mileage follows a roughly log-normal distribution that is truncated at zero and skewed to the right. The median car was driven about 10,500 miles. About 57% of cars were driven fewer miles than the average.

Annual mileage differs somewhat between class and territory groups, as shown in Table 3. Most notably, senior citizens drive their cars substantially fewer miles than other classes, and the highest-risk territories also have the lowest annual mileage per vehicle. High insurance risk territories are generally urban areas, where traffic density makes for more frequent claims but land use density makes for short car trips.

Figure 1. Histogram of annual mileage estimates for 3.25M policy-vehicle combinations in Massachusetts

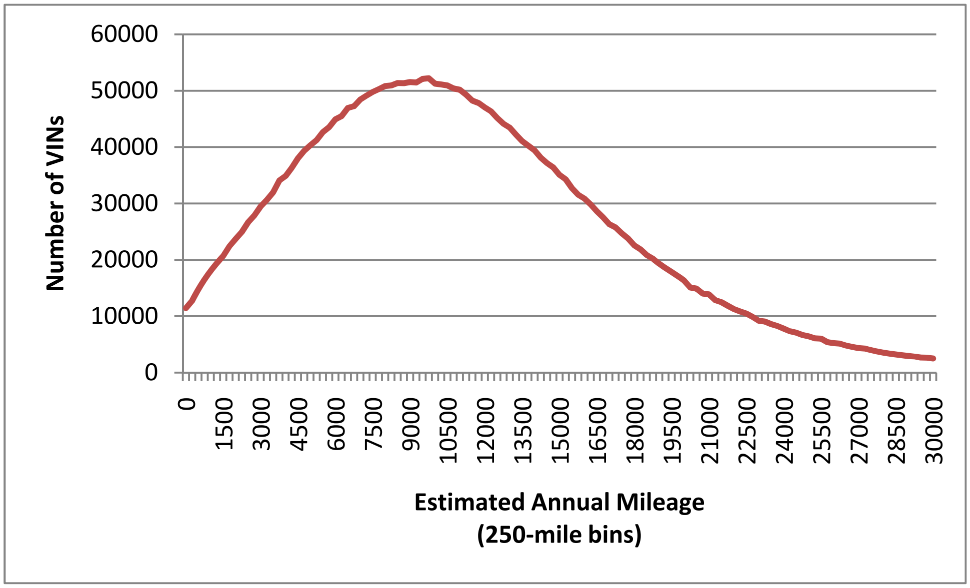


Table 3. Average annual mileage for five classes and six territories.

Class	Average annual mileage		Territory	Average annual mileage
Adult	12,398		1	12,456
Business	14,173		2	12,149
< 3yrs exp	12,911		3	12,262
3 -6 yrs exp	13,207		4	11,798
Senior citizen	7,519		5	10,702
			6	10,523
ALL	11,695		ALL	11,695

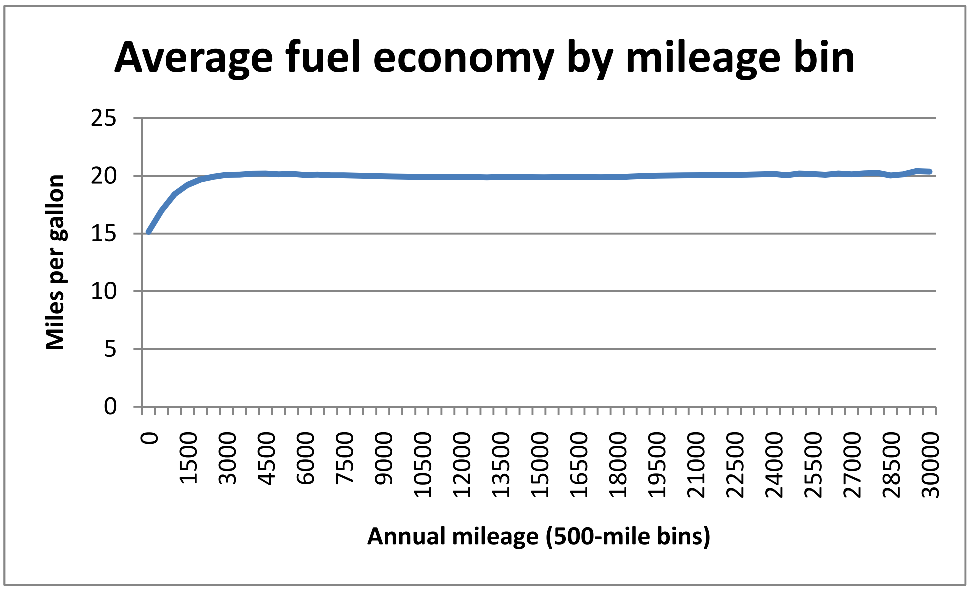
Fuel economy

Our mileage and insurance data are linked by the vehicle's 17-character Vehicle Identification Number (VIN), which contains a good deal of encoded information. We hired a commercial car data firm, VINquery.com, to parse make, model, year, and other data from the 3.05M unique VINs in our study and to provide corresponding EPA fuel economy estimates as well. 99.6% of our VINs proved valid and of these, VINquery.com was able to provide fuel economy data for 95.9%.

Fuel economy data consist of official EPA estimates for city driving and highway driving. While it is probable that cars in urban territories do more city driving and cars in suburban and rural territories do more highway driving, we avoid making an arbitrary assumption about proportions, instead taking a simple average of the two figures. We then multiplied by 88% to adjust for the fact that pre-2008 EPA estimates are inflated by at least 12%[10]. Our resultant figure is probably still higher than actual gas mileage but for the purpose of comparing across vehicles, this makes no difference.

Figure 2 shows that fuel economy exhibits almost no correlation with annual mileage[11], and Figure 3 shows that there is very little variation by territory. The high-risk territories have slightly higher average fuel economy, but only by a few percent (20.4 mpg in Territory 6 versus 19.7 in Territory 1). The only large variation by class is for business vehicles (which anyway account for just 1% of policies[12]), having substantially worse average fuel economy than household vehicles as shown in Figure 4. Between the age and driving experience groupings, average fuel economy varies only about 10% (19.8 mpg for adults versus 21.5 mpg for individuals with 3-6 years of driving experience).

Figure 2. Average fuel economy by annual mileage bin



[10] See, for instance, http://www.ucsusa.org/clean_vehicles/solutions/cleaner_cars_pickups_and_suvs/fixing-the-epas-fuel-economy.html

[11] The small dip at the left hand side of Figure 2 represents vehicles with an annual mileage estimate of fewer than 2,000 miles. These constitute only 4% of all car years of exposure in our study. The fact that these vehicles exhibit lower fuel economy than the others may indicate that this group consists largely of an assortment of "special cases"—campers kept in the driveway except for trips, summer cars in the garage over the winter, vehicles that are insured but being repaired, and so on.

[12] On vehicles with noncommercial license plates. Commercial vehicles are not included in our study but would raise this percentage significantly.

Figure 3. Average fuel economy by territory

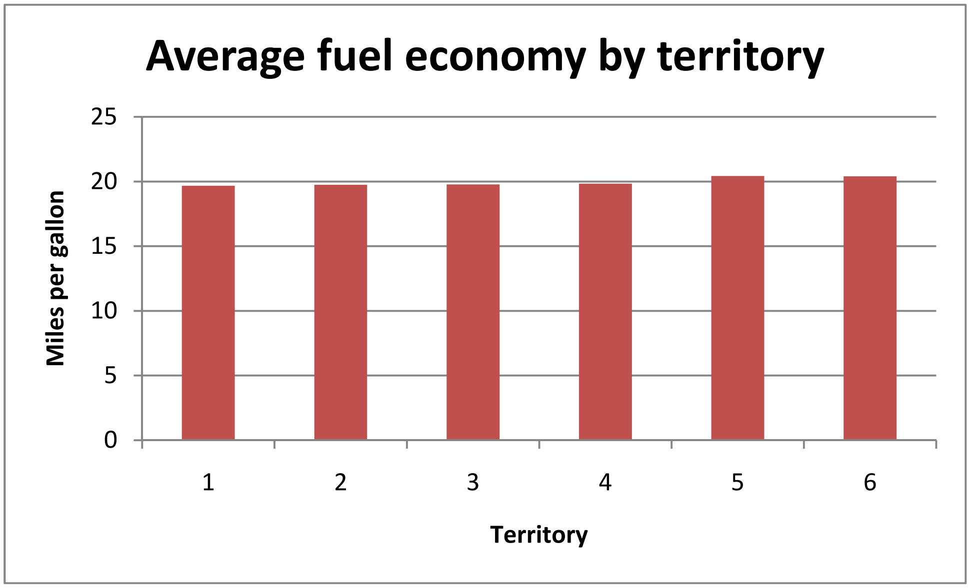
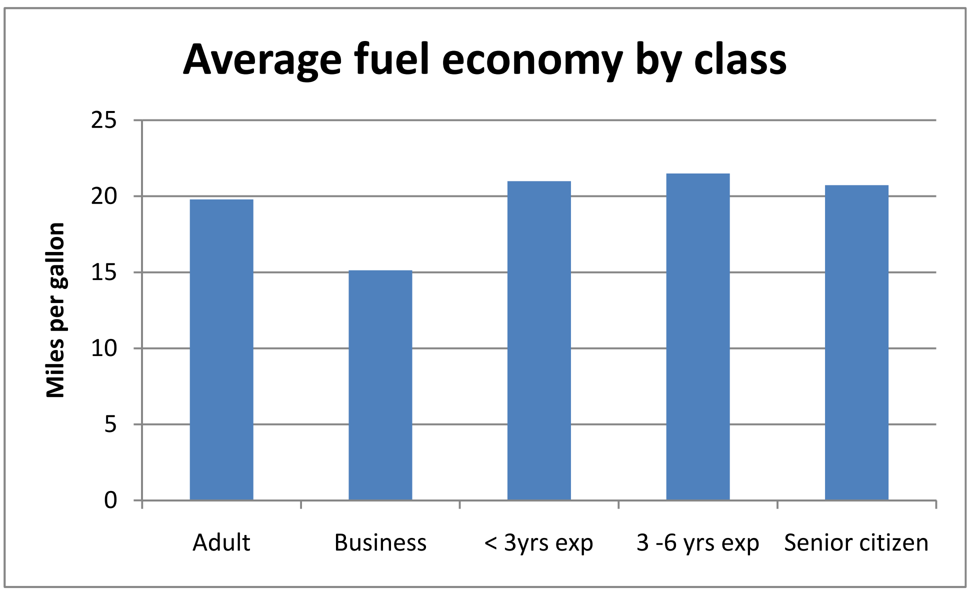


Figure 4. Average fuel economy by class



The lack of any strong correlation between fuel economy and class, territory or annual mileage indicates that PAYD will probably reduce fuel consumption by about the same proportion within all groups as it reduces VMT.

Sample size and bias

Our study encompasses 2.87M car years of exposure. There are 2.46M policies in our study, which insured a total of 2.79M vehicles as of each policy's effective date. Though only 2.79M vehicles were insured to begin with, 2.87M car years of exposure were earned because more vehicles were added to policies than dropped from policies over the course of the ensuing year. Due to vehicle turnover, 3.05M unique vehicles are included, and they resulted in a total of 3.25M unique policy-vehicle combinations. There are more policy-vehicle combinations than vehicles because some vehicles changed hands and so were insured under first one owner and then another. There are more vehicles than car years of exposure because some households may, for instance, consistently own two cars at a time but happen to replace one during the year, thus creating records for three distinct vehicles but only two car years of exposure.

The question of what sample size our study represents is nontrivial. Compared to CAR's entire dataset of private passenger vehicles in the 2006 policy year, we include 71% of exposures (2.87M out of 4.02M car years[13]), but only 62% of unique policy-vehicle combinations (3.25M out of 5.26M) and 65% of distinct vehicles (3.05M out of 4.68M). Of course, CAR's dataset only includes insured vehicles[14], and we have no data on what percentage of vehicles are uninsured.

The vast majority of the records that dropped out were eliminated for lack of a mileage estimate; only 0.1% were eliminated due to errors in the insurance data tables. Of the 4.68M vehicles represented by the original insurance policy data, we first limited our study to the 3.41M with MassGIS mileage estimates. Of these, we were then able to create our own mileage estimates for 3.05M[15]. In all, 1.63M (over a third) of the insured vehicles dropped out for lack of a mileage estimate. About 0.16M had no RMV inspections at all, and most of the remainder had only one inspection[16].

In many cases, it may simply be that the owner neglected to have an annual safety check as required by law. However, some vehicles were probably moved out of state, stolen or totaled, and so were simply not around to have a "second" inspection; others were purchased or moved into the state and so were not around to have a "first" inspection. In all of these scenarios, the policy on the vehicle is likely to last fewer than 12 months[17], and this explains why, by excluding them, we eliminated 35% of vehicles but only 29% of exposure.

While each of these scenarios could correlate with risk in some way, the most obvious source of bias is that cars that were totaled are more likely to have been eliminated from our study. Though we have not processed insurance policy and claims data for all 4.68M vehicles, we are able to get a ballpark idea of the bias by comparing the 3.25M policy-vehicle combinations in our study to the 0.36M policy-vehicle combinations for which there was a MassGIS mileage estimate but for which we could not create our own mileage estimate. When all policy-vehicle combinations (3.61M) are viewed together, claim

---

[13] 4.02M probably represents the approximate number of private passenger vehicles at any one time in Massachusetts. In the 2000 Census, households reported owning a total of 3.78M vehicles (Summary File 3, H46), so 4.02M would be a plausible figure for 2006. As explained above, the number of distinct vehicles insured over the course of a year is higher because some households replace a vehicle in the course of a year. Also, any households that owned a car for only part of the year or lived in Massachusetts for only part of the year would count fully towards the distinct vehicle count but only partly towards the aggregate car years of exposure.
[14] An exception is some vehicles which were insured but then had their policies canceled for nonpayment and so became uninsured.
[15] Applying our own mileage estimation method without limitation to only those vehicles for which MassGIS had an estimate would allow us to obtain estimates for 3.47M vehicles rather than 3.05M, so our sample could be expanded a bit in future research. However, some of the additional vehicles come simply from our using an extra six months of 2008 RMV inspection data which MassGIS did not have available at the time it created its estimates, so much of the new data would be of poor quality in that the mileage estimation period would have little or no overlap with the 2006 policy year.
[16] A smaller fraction had two or more inspections but either under different owners, with erroneous odometer readings, or spaced too closely or too far apart to create a reliable annual mileage estimate.
[17] The mentioned changes could also happen after the 12 month policy term ends but before a second safety inspection, so there is no guarantee that the policy term would last fewer than 12 months.

frequency averages 5.9 and pure premium averages $187 for the two types of compulsory coverage discussed above. However, the 90% of the combinations which we study exhibit a claim frequency of 5.6 and a pure premium of $175. The 10% for which we were unable to obtain an estimate have a much worse safety record, with a claim frequency of 9.0 and a pure premium of $323.

This bias is worth keeping in mind. While it demonstrates that our risk estimates are biased slightly downward, we have no reason to expect significant changes in our conclusions about the relationship between claims risk and miles driven. Nevertheless, the bias could strengthen or weaken the riskiness of miles across classes and territories. For instance, it is possible that in suburban and rural areas, where collisions are rarer but more severe when they do occur, a greater proportion of cars were totaled and thus dropped from the study than in urban areas. Identifying bias in the spatial pattern regarding which policies were eliminated from the study is a point for future research. Note that although the pure premiums we calculate are certainly biased downward, the multiplier of 5.5 which we apply to translate these pure premiums into retail insurance prices is inexact anyway, so it is difficult to say whether the per mile prices we use in modeling VMT impacts are ultimately affected by this bias.

As a final note, fuel economy estimates were available for just 96% of vehicles. This affects only the VMT reduction model—all other parts of our study use all 3.05M vehicles. When we extrapolate our VMT findings to reflect all private passenger vehicles in the state, we adjust accordingly.

The Correlation between Mileage and Risk

Figure 5 and Figure 6 provide a quick look at the aggregated data and indicate that mileage does correlate positively with risk. If mileage were irrelevant, one would expect flat horizontal lines in both graphs – that is, a constant claim frequency and pure premium risk regardless of how many miles each vehicle is driven during each policy year. Instead, there is a positive slope: claim frequency and pure premium both increase as annual mileage increases. Recall that these figures are for two types of compulsory coverage only— bodily injury and property damage liability, and personal injury protection.

Figure 5. Claim frequency per car year by annual mileage, all policies

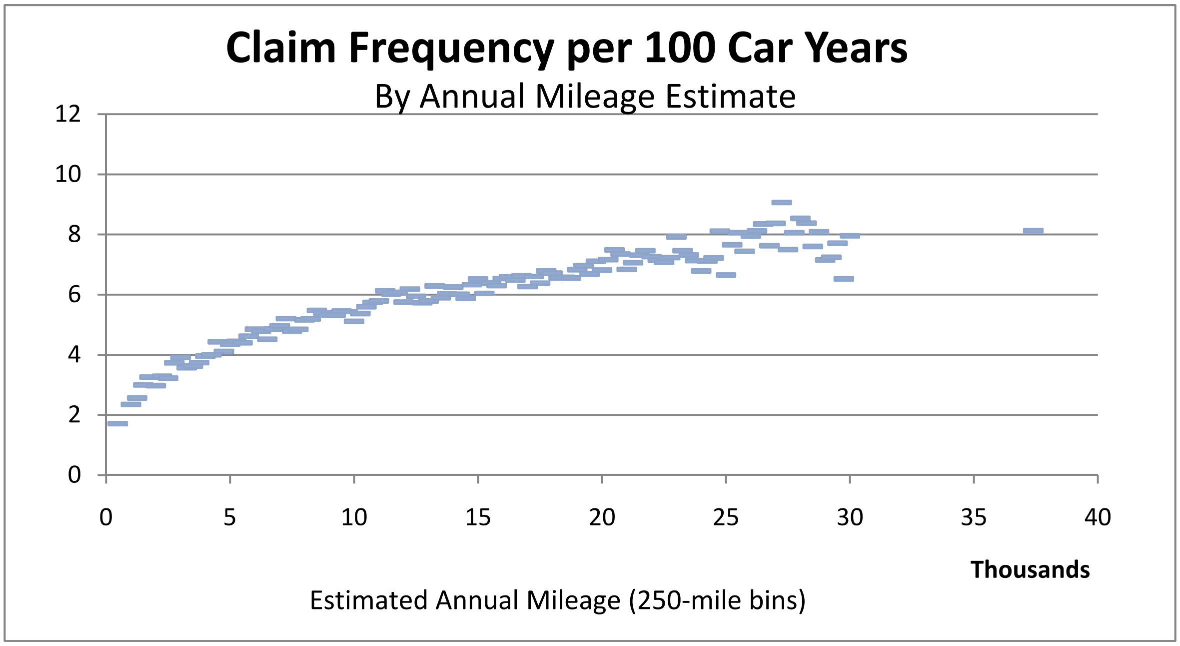
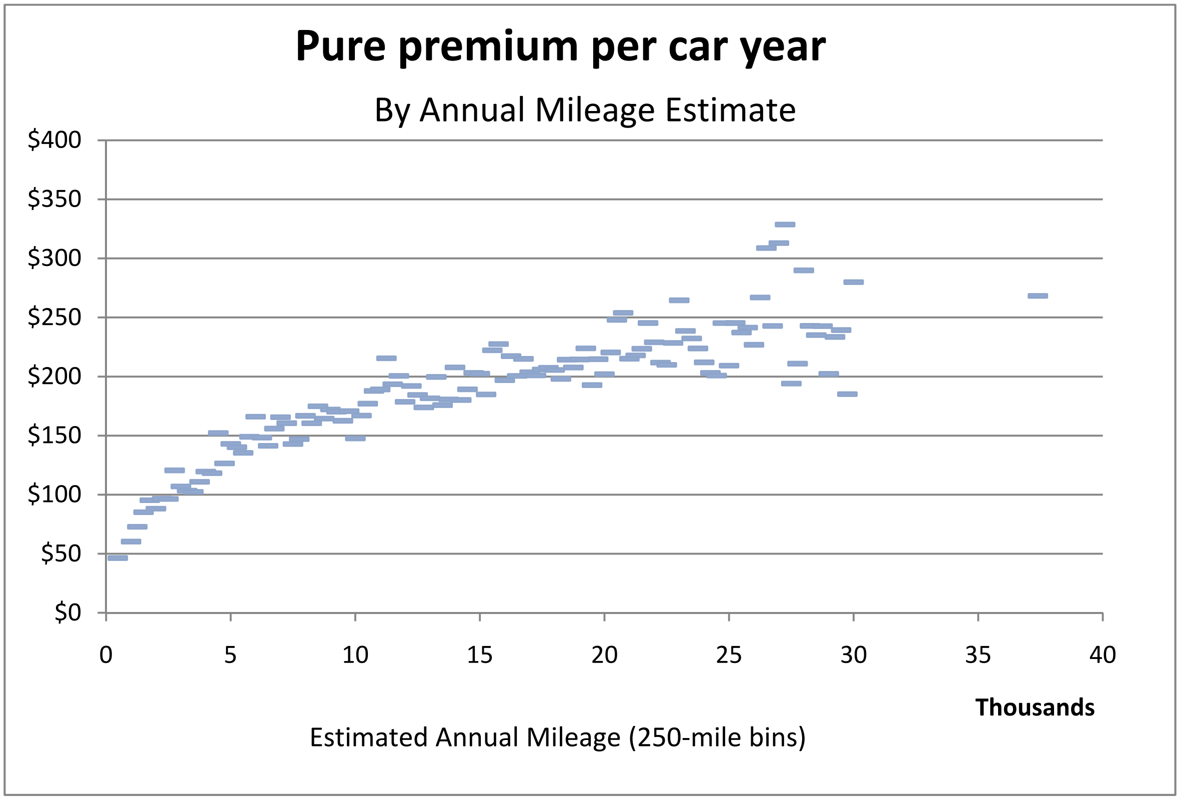
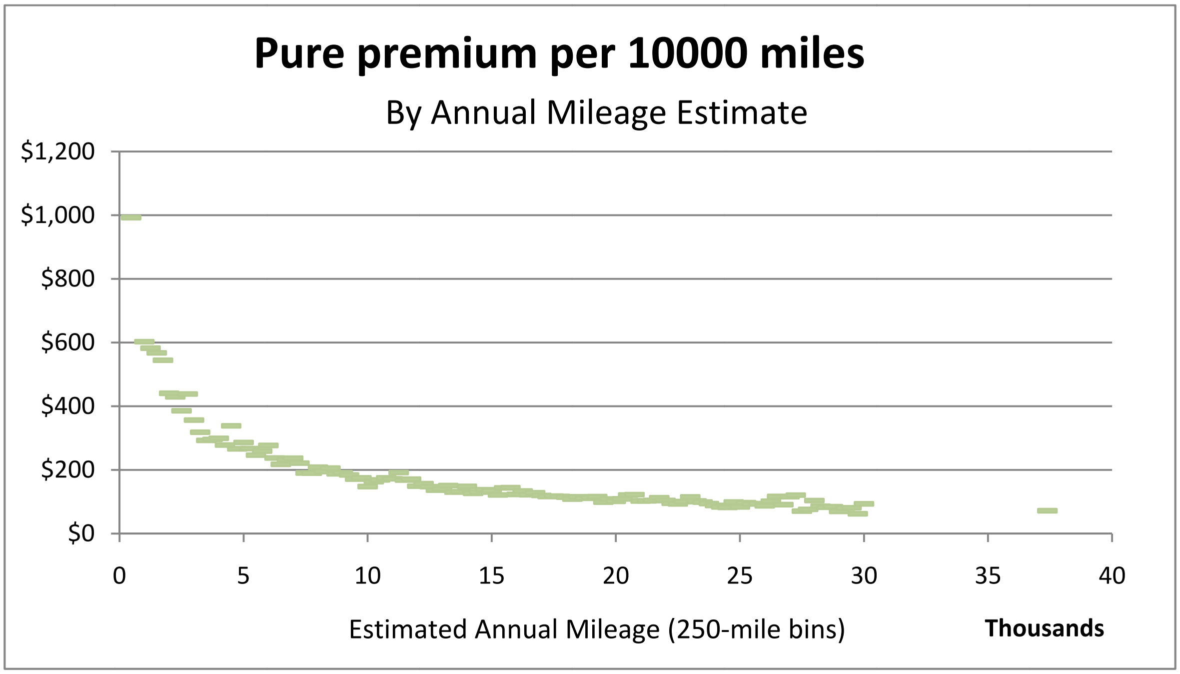


Figure 6. Pure premium per car year by annual mileage, all policies.



Before moving on, it is worth acknowledging the nature of the relationship between mileage and risk shown above. Although it is clear that mileage correlates with risk, two other phenomena are visible in these graphs. First, the relationship appears to be less-than-proportional: for instance, the threefold increase in mileage from 10,000 to 30,000 results in less than a doubling of claim frequency (from 5 to 8 claims per hundred car-years in Figure 5), and of pure premium (from $175 to $250 in Figure 6). Second, the relationship appears to be less-than-linear: the data form a curve with diminishing "returns." This fact is even clearer from Figure 7. If risk were proportional to mileage, one would expect the *per mile* pure premium to be constant—a flat horizontal line. Instead, it is a convex curve, indicating that a high annual mileage car costs somewhat less *per mile* to insure than a low mileage car.

Figure 7. Pure premium per 10,000 miles by annual mileage, all policies.



Other researchers have observed this before and have suggested a number of reasons why risk may not increase proportionally with mileage—for instance, high-mileage drivers may be more experienced, or they may do more of their driving on low-risk divided highways rather than high-risk city streets, and so on (Litman 2008a, 16; Bordoff and Noel 2008, 7). The graphs above suggest that a single per-mile fee for all drivers would not be very actuarially accurate. A central point of this paper will be to show that existing class and territory groupings capture enough of the variation in driving habits and skill such that the relationship between mileage and risk is substantially stronger within each grouping than it is over all drivers. In other words, mileage is better used to supplement, rather than substitute for, existing insurance rating factors.

While real differences in risk between different drivers account for much of the curvature in the above graphs, some of the diminishing returns effect is probably due to regression to the mean. Recall that we created mileage estimates by pairing two odometer readings; in most cases, the reading dates did not happen to fall on exactly the start and end date of the insurance policy. While most of the mileage estimation periods (92%) overlap at least partially with the period during which the vehicle was insured (and therefore for which claims data are available), they do not overlap the entire period—the average amount of overlap is about 8.5 months. Therefore, regression-to-the-mean effects are possible[18]. Vehicles estimated to have driven low or high annual mileage (e.g., 5,000 or 30,000 miles), based on

[18] When we paired odometer readings to create mileage estimates, we checked that the anonymized registration number was the same at both readings, indicating the same owner of the vehicle. So for the 92% of cases with some overlap between the mileage estimation period and the policy period, the mileage estimate is guaranteed to refer to the habits of the relevant insured drivers. Of the remaining 8%, there is probably some small subset where the vehicle was sold in the time between the insurance policy and the two odometer readings. Therefore the mileage estimate describes the behavior of a previous or later owner, whose driving habits are largely uncorrelated with those of the policyholder. While such cases are probably less than 1% of all data, they probably exhibit particularly strong regression to the mean.

their inspection period, are likely to have driven somewhat closer to the mean of 11,695 miles during the 12-month period during which we observed their insurance claims experience. Therefore, vehicles we estimate were driven 5,000 miles per year during the period between odometer readings probably have insurance claims commensurate with a higher mileage that they likely drove during the 2006 policy year included in our study, and we therefore overestimate their risk. Vehicles we estimate were driven 30,000 miles annually probably have insurance claims commensurate with a lower mileage during policy year 2006, and so we underestimate their risk.

Despite regression to the mean effects, our data show clearly that risk does correlate with mileage, and that the correlation is stronger once class and territory are considered. One way to show this is by conducting a Poisson regression over all 3.25M policy-vehicle combinations included in this study. We use the R statistical package's generalized linear model (GLM) function for this purpose. Poisson regression respects the underlying "rare event" nature of auto collisions and is therefore often used in studies of road safety. Another realistic feature of our Poisson regression is that it is constrained to predict zero risk when zero miles are traveled[19].

First, we regress pure premium or expected total losses per car year on mileage alone. The result is shown in Equation 1 and graphed in Figure 8.

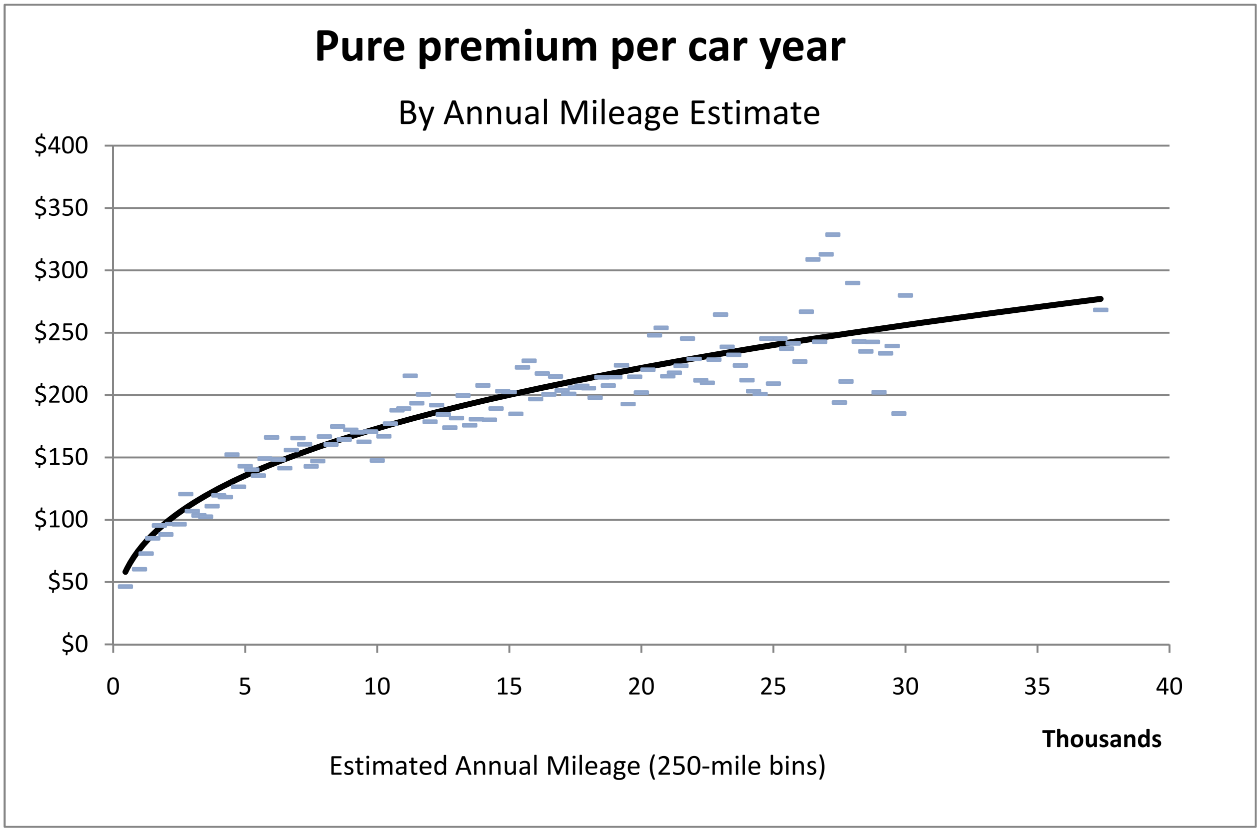
Equation 1. Poisson model for pure premium as a function of mileage alone.

$$Pure\ premium\ =\ 6.53\ *\ (ann\_miles^{0.36})$$

With a p-value of less than 2×10^{-16}, the exponent for mileage is extremely statistically significant, which is only to be expected with 3.25M data points. The relationship between mileage and risk is indeed found to be less-than-linear: risk is proportional to mileage raised to the 0.36 power.

[19] A vehicle with comprehensive coverage could, theoretically, stay parked all year and still file a claim for theft or damage. However, recall that we are addressing only property damage liability, bodily injury and personal injury protection.

Figure 8. Fit of Poisson regression on pure premium per car year by annual mileage estimate, all policies



Next, we regress pure premium on mileage, class, and territory grouping. The overall model is shown in Equation 2 with corresponding class and territory relativities[20] listed in Table 2.

Equation 2. Poisson model for pure premium as a function of mileage, class and territory.

$$Pure\ premium\ =\ 2.35 * (ann\_miles^{.40}) * (class\ relativity) * (terr\ relativity)$$

Table 4. Relativities for nine classes and six territories in Equation 2.

Class	Relativity		Territory	Relativity
Adult	1.00		1	1.00
Business	1.32		2	1.24
< 3yrs exp	2.65		3	1.28
3 -6 yrs exp	1.83		4	1.55
Senior citizen	1.17		5	2.04
			6	2.98

[20] We compute relativities with the lowest risk group, Territory 1 adults, as the reference. Industry practice is to use the overall average as the reference, such that low risk groups have relativities below one and high risk groups' relativities would be larger than one but not as large as those shown in Table 4.

Two things are interesting to note about Equation 2. First, the exponent on mileage has increased compared to Equation 1, from .36 to .40. This means that the relationship between risk and mileage is a bit closer to proportional once class and territory are considered. Second, note the range of relativities in Table 4: even when mileage is considered, risk varies as much as threefold between different class and territory groupings.

This model is imperfect, though, because the class and territory relativities only affect the magnitude of the curve, not its shape. The model is constrained to find a single exponent for mileage (in this case, 0.40) for all classes and territories. Suppose that the relationship between risk and mileage were linear within any one class/territory group, but that the slopes differed (to the extent that the per-mile risk differs across the class and territories). The model in Equation 2 would "compromise" between the different slopes by finding an exponent less than one, even though a regression on any one group would find an exponent closer to one, indicating a more proportional relationship.

For this reason we proceed to also fit various linear models to the data. Since a linear model would fit poorly on the original disaggregated 3.25M policy-vehicle combinations with their many zero-loss records, we aggregate the data into groups by class, territory, and 500-mile annual mileage bins. Of these we examine only the mileage bins between 2,000 and 30,000. We exclude very low and high annual mileage outliers for a few reasons. First, we expect regression to the mean effects to be strongest for the most extreme outliers and wish to limit the effect that this phenomenon will have on our results. Second, we will show in the next section that one possible pricing scheme would be to charge customers a flat yearly rate which includes coverage for the first 2,000 miles, and then to charge by-the-mile thereafter. A regression limited to vehicles which travel at least 2,000 miles per year will correspond more closely to such a pricing scheme than would a regression on all vehicles. This range also includes the vast majority of the data: 94.4% of all car-years of exposure in the analyzed data were earned on vehicles traveling between 2,000 and 30,000 miles per year.

Besides excluding outliers, we also weight the data points by number of car years of exposure, so that even though we have grouped by class, territory and mileage bin, each individual policyholder's outcomes affect the model equally. In other words, it's "one person, one vote" or, more accurately, "one car, one vote." A final feature of our regression is that we do not constrain the intercept to be zero. Admittedly, this allows some risk even at zero miles, which is in conflict with the reality that a policyholder has no chance of filing a property damage liability or injury claim if she literally leaves her car parked all year. However, in view of the scarcity of such policyholders in real life, along with evidence that the relationship between risk and mileage is less-than-proportional, we choose an unconstrained model. Our purpose in this section is simply to identify a role for mileage in insurance rating and pricing more generally; in the next section we compare the accuracy of different pricing schemes.

First, we model pure premium as a function of mileage bin alone with no class or territory effects. Since we do not constrain the intercept to be zero, the result shown in Equation 3 is a flat yearly fee plus a per-mile rate[21]. Since these figures are shown in dollars and cents, it merits reminding the reader that

[21] In this model, the intercept, $111.70, is the amount the user would pay simply to have a policy; it would not include any miles. Our purpose here is simply to compare the actuarial accuracy of various models in order to establish the proper role of mileage in insurance rating and pricing. However, as we discuss in the next section, a more realistic business model might be to increase the yearly fee so as to include the first, say, 2,000 miles.

these numbers represent only pure premiums for compulsory coverage, which constitute just a fraction of typical consumer prices for full coverage.

Equation 3. Linear model for pure premium as a function of mileage only.

$$Pure\ premium = \$111.70 + \ 0.55¢ \times ann\_miles$$

Once again, the coefficient for annual mileage is found to be highly statistically significant with a p-value less than 2×10^{-16}. One-way analysis of variance (ANOVA) reveals an F statistic of 171, where the 95th percentile of the F distribution is 3.8, and the p-value is, again, less than 2×10^{-16}, meaning that the mileage bin groupings provide more explanatory power than a random grouping would. To know *how much* explanatory power they provide, we turn to the adjusted R^2 statistic of 0.09, which indicates that just 9% of the variation in the observed 2006 pure premiums among class/territory/mileage bin groupings can be explained by mileage alone. Note that this number is somewhat arbitrary: since the regression is on mileage bins and not on the underlying individual-level data, and the amount of variation to be explained in the first place is an artifact of the number of mileage bins chosen. However, this 9% figure is meaningful in comparison with the amount of variation explained by other regression models on the same data.

As one such comparison, consider a model of pure premium as a function of class and territory alone with no mileage effects. This model corresponds roughly to actual present-day insurance pricing which takes little account of mileage. The basic formula is shown in Equation 4 with the adjustments (relative to the base case of adults in territory 1) elaborated in Table 5.

Equation 4. Linear model for pure premium as a function of class and territory only.

$$Pure\ premium = \$96.50 + class\ adjustment + territory\ adjustment$$

Table 5. Adjustments for five classes and six territories in Equation 4.

Class	Adjustment	Territory	Adjustment
Adult	$ 0.00	1	$ 0.00
Business	$ 52.84	2	$ 26.27
< 3yrs exp	$ 256.34	3	$ 31.40
3 -6 yrs exp	$ 143.16	4	$ 57.25
Senior citizen	- $ 3.87	5	$ 102.80
		6	$ 197.34

This model obtains an adjusted R^2 value of .57, meaning that class and territory alone can explain 57% of the variation in pure premium among the 1710 class/territory/mileage bin groups. This is a good deal more than the mere 9% explained by mileage alone. In this sense, class and territory are superior rating factors, though they are just proxies for driving patterns, whereas mileage is a direct individual measure of driving habits.

Even if mileage is a rather poor ***substitute*** for traditional insurance rating factors, though, a model using mileage, class and territory together shows that mileage is an excellent ***supplement***. In Equation 5, both the yearly fee and the per-mile rate are allowed to vary by class and territory. Table 6 is included for the sake of completeness in explaining the linear model shown in Equation 5. It should not be taken as a recommended pricing scheme[22].

Equation 5. Linear model for pure premium as a function of mileage, class and territory.

$$\begin{aligned} Pure\ premium = \$40.12 + class\ adjustment + territory\ adjustment \\ + (0.43¢ + class\ rate + territory\ rate) \times ann\_miles \end{aligned}$$

Table 6. Surcharges and additional per-mile rates for five classes and six territories in Equation 5.

Class	Yearly adjustment	Additional per-mile rate		Territory	Yearly adjustment	Additional per-mile rate
Adult	$ 0.00	0.00¢		1	$ 0.00	0.00¢
Business	$ 29.05	0.13¢		2	$ 22.81	0.04¢
< 3yrs exp	$ 246.40	0.06¢		3	$ 14.42	0.15¢
3 -6 yrs exp	$ 56.69	0.63¢		4	$ 48.59	0.10¢
Senior citizen	$ 25.90	-0.05¢		5	$ 59.80	0.48¢
				6	$ 123.40	0.81¢

This model achieves an adjusted R^2 statistic of 0.72, meaning that 72% of the variation among class/territory/mileage bin groups can be explained by all three variables together in this particular linear model. Note that ***the whole is better than the sum of the parts***: mileage explains 9% of the variation on its own, but 15% when used in conjunction with class and territory (for a total of 72% in Equation 5 versus 57% in Equation 4)[23]. This indicates that mileage is a more meaningful measure of risk when there is some control on what types of miles are being traveled and how they are being driven—most likely, territory provides a proxy for street setting and class provides a proxy for driver skill and behavior.

These results suggest that a more elaborate pricing scheme, based not only on how many miles are driven but when, where and how they are driven, might provide even more accuracy and perhaps replace some of the traditional rating factors altogether. Proposals for such telematic pricing usually involve an odometer monitor, a clock and Global Positioning System (GPS) to determine when and where the car is being driven, and perhaps even other sensors such as accelerometers to determine the

[22] The model that we present has a few peculiarities. For instance, the higher risk in Territory 4 compared to Territory 3 is entirely absorbed into a yearly adjustment; the per mile rate is actually lower in Territory 4 than Territory 3. The same is true of the drivers with less than three years of experience versus drivers with three to six years of experience. In practice it would probably be best to divide up the difference in risk more evenly between the yearly fee and the per mile rate, or, better yet, to fit a more complex model that does not assume independence between class and territory effects.

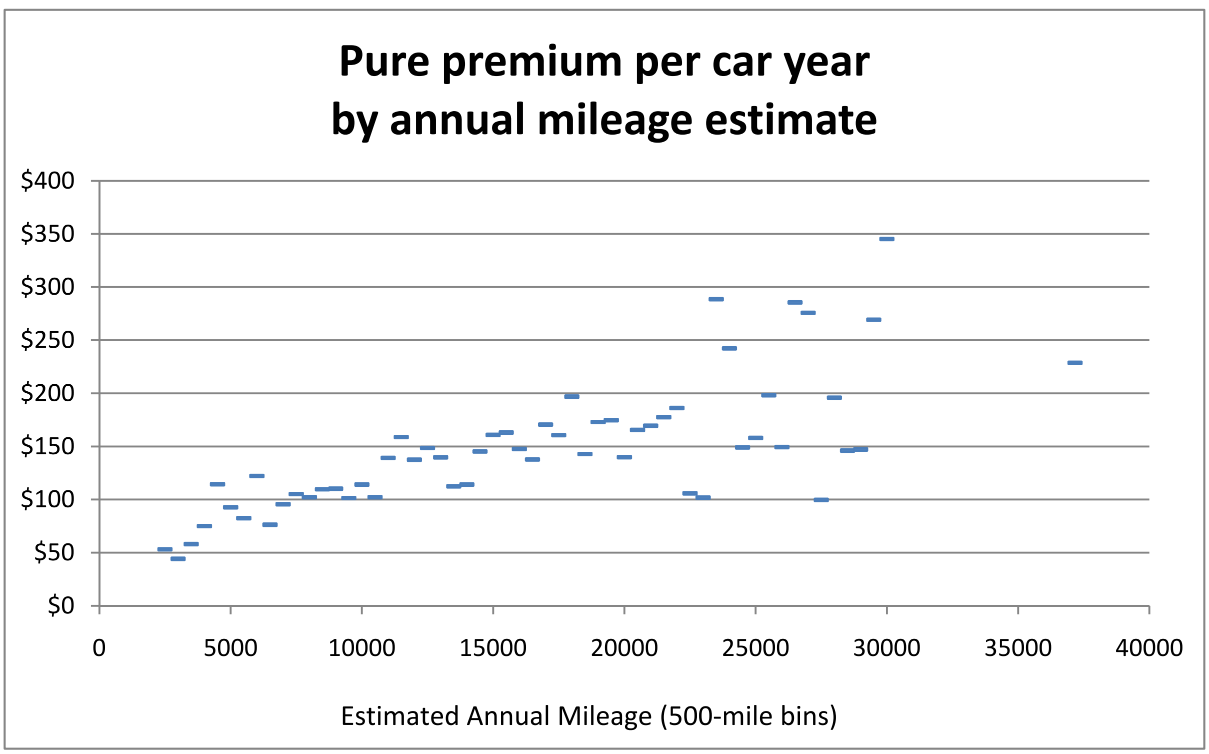
[23] Note that these R^2 statistics of 9%, 15%, 57% and 72% are only meaningful in relation to each other, and not in any absolute sense. If a model could achieve an R^2 of 100% then it would explain all of the variation *between the 1710 class/territory/mileage bin groups*, but still would not explain nearly all of the variation *between individual drivers*, most of which is due to random chance in any given year. For instance, we tested linear models similar to Equations 3, 4 and 5 but with nine class groupings instead of five, and found that mileage explained 7% of variation alone, class and territory explained 47% alone, and the two explained 59% together. So even with the same underlying data (the 3.25M policy-vehicle combinations), different models using different groupings of the data can find different R^2 statistics. However, the important conclusion from these results is the same regardless: the whole is better than the sum of the parts, and mileage provides more explanatory power when used with class and territory than when used alone.

driver's style of acceleration, braking and turning. This information could then be transmitted wirelessly to the insurance company and corresponding premiums would be billed to the user[24]. Needless to say, such tracking schemes raise a host of confidentiality, security, and locational privacy concerns, but their due concern is beyond the scope of this report.

The two Poisson models and three linear models shown here demonstrate the following. First, mileage is a significant predictor of insurance risk. Second, if used alone, mileage is inferior to traditional insurance rating factors. Third, if used in conjunction with traditional insurance rating factors, mileage can substantially improve actuarial accuracy.

It can be supposed that one reason why a linear model based on mileage alone (Equation 3) provides relatively little explanatory power is that, as seen in Figure 6, pure premium is not very linear with respect to mileage when *all* policies are considered. The substantially better linear fit obtained in Equation 5 can be explained in part by the fact that the risk-mileage relationship is more nearly linear within any one class and territory. Figure 9 emphasizes this point by showing the relationship between pure premium and mileage for one example rate group, Territory 3 adults[25]. Compared to Figure 6, it shows a much more linear relationship between risk and mileage. The relationship is also closer to proportional than that in Figure 6: pure premium doubles from about $100 at 10,000 miles to at least $200 at 30,000 miles.

Figure 9. Pure premium per car year by annual mileage, Territory 3 adults.



[24] For a more detailed discussion of telematic insurance pricing technologies, see Bordoff and Noel 2008.
[25] We choose experienced drivers ("adults") because that is the largest class and we choose Territory 3 because it is of intermediate risk.

As a further demonstration of this, we compare linear and Poisson regressions for Territory 3 adults. Equations 2 and 5 were both calibrated on the entire dataset rather than just Territory 3 adults, so neither accounts for class-territory interaction effects, and in the case of Equation 2, the exponent on mileage is reduced by the need to "compromise" between different rate groups. Therefore we conduct two new regressions on only Territory 3 adults, with results shown in Equation 6 and Equation 7.

Equation 6. Poisson model for pure premium as a function of mileage for Territory 3 adults only

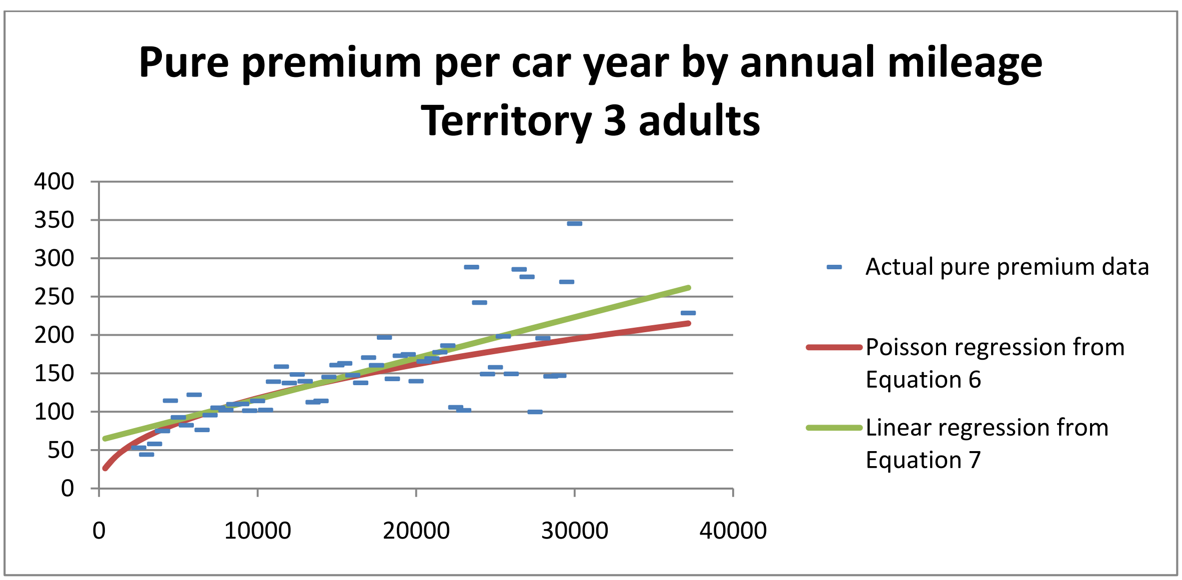
$$Pure\ premium = \$1.70 \times (ann\_miles^{0.46})$$

Equation 7. Linear model for pure premium as a function of mileage for Territory 3 adults only

$$Pure\ premium = \$62.81 + 0.53¢\ \times ann\_miles$$

Note that the exponent (0.46) on annual mileage in Equation 6 is even higher than that in Equation 2—again, the risk-mileage relationship is closer to linear within a class/territory grouping than across all groups. Figure 10 overlays these two new models on the same pure premium data for Territory 3 adults shown in Figure 9. Note that for this rate group, the linear regression and the Poisson regression both fit quite well. Across all data points, the Poisson regression has a lower exposure-weighted sum of squared errors (611 versus 659) because it better approximates the low and high outliers (which, after all, were excluded from the linear model). When only the mileage bins between 2,000 and 30,000 are considered, the linear model is actually a slightly tighter fit (597 versus 599).

Figure 10. Pure premium per car year by annual mileage, Territory 3 adults, with results from two regressions.



The observation that the risk-mileage relationship is more nearly linear *within* a class/territory group suggests that part of the reason for the curvature or diminishing "returns" exhibited in Figure 6 is that drivers in low-risk classes or territories tend to have higher annual mileage than drivers in high-risk classes or territories. The average annual mileage figures shown in Table 3 tend to reinforce this suspicion: in particular, the highest-risk territories (5 and 6) have lower annual mileage than the others. These are urban territories where high traffic density makes for high-risk driving[26] but destinations are relatively close to home[27].

[26] Here we refer to insurance risk. In terms of fatality and severe injury risk, researchers have found that cities are safer than suburban or rural locations. See, for instance, Marshall 2009. This is also discussed in Litman 2008a, 13.
[27] Of course, the average annual mileage in Territory 6 is not sufficiently lower than that in Territory 1 to cancel out the difference in per-mile risk, or else Territory 6 would not be considered riskier in the first place under traditional insurance rating.

Pricing Schemes Suggested by the Data

Under the current model of insurance pricing, users pay the same amount per car year (within their class and territory group) almost regardless of how many miles they drive[28]. This implies a cross-subsidy from low-mileage drivers to high-mileage drivers, and it also means that policyholders have no opportunity to save money by reducing mileage, nor any incentive to do so, despite the obvious fact that they impose additional costs both on their insurance company and on society with each marginal mile driven. This observation is what has motivated many researchers and advocates to propose Pay-As-You-Drive insurance.

The simplest pay-as-you-drive pricing model involves strictly per-mile pricing—if you don't drive, you don't pay and every mile driven incurs a constant marginal price —but other pricing models are imaginable as well and here we compare the implications of a few options. Our purpose in doing so is twofold. First, to use the data to illuminate some pitfalls and advantages of various schemes. Second, to settle upon a plausible pricing scheme to use in our predictions of how driver behavior would be affected by the introduction of Pay-As-You-Drive insurance.

For instance, the linear regressions shown above suggest that a flat yearly fee plus a per-mile fee is most actuarially accurate—if a zero intercept had been the best fit, the unconstrained regressions would have found a zero intercept. To be exact, Equation 3 and Equation 5 imply that the user pays a flat fee simply to have a policy at all, and then pays for every single mile thereafter. It is perhaps more realistic to imagine that insurance companies would bundle the first, say, 2,000 miles into a yearly rate and then charge per mile beginning with the 2001st mile.

In this section we will compare the status quo (strictly per car year pricing) with two alternate models: the "strictly per mile model" (no yearly fee) and the "2K+per-mile" model (a yearly fee that includes 2000 miles, with per-mile pricing thereafter). To compare these three models, we consider once again the example of Territory 3 adults.

To compute the pure premium for the status quo, we simply divide aggregate losses by aggregate car years of exposure. For Territory 3 adults the result is $130.64 per car year. Similarly, for the strictly per mile model we divide aggregate losses by aggregate miles traveled. For Territory 3 adults the answer is 1.0¢ per mile—again, this is a low per-mile rate because it only accounts for pure premiums for certain compulsory liability coverage.

For the "2K+per-mile" model we reason that the yearly fee must wholly cover losses for drivers who drive 2,000 miles in a year, so we set the yearly fee equal to the average loss for cars driven 1,750 to 2,250 miles per year[29]. Next we compute the aggregate losses for all drivers minus the total amount

[28] Discounts currently offered to low mileage customers are in the neighborhood of just 5 or 10% and are based on self-reported prospective mileage—i.e. how much the customer *expects* to drive, rather than how much the customer actually drives.

[29] Another method would be to set the price equal to the average loss for cars driven between 0 and 2,000 miles, but this would underestimate the average loss for cars driven exactly 2,000 miles, and so it would be rather risky for insurance companies to offer 2,000 miles packaged at such a price.

they would pay in yearly fees, and divide the difference by aggregate miles traveled in excess of 2,000 per year to obtain the per-mile rate. The result is $59.05 for the first 2,000 miles plus 0.68¢ per mile thereafter. Note that the implied rate for the first 2,000 miles is 2.95¢ per mile, more than four times the per mile rate thereafter. This contrast is in part reflective of the higher per-mile risk for low mileage drivers, and probably in part the result of regression to the mean.

Table 7 compares the three models. In the "estimated retail price" columns, pure premiums for compulsory coverage are multiplied by 5.5 to give an estimate of what a typical full coverage policy would cost the consumer. The average annual mileage for vehicles driven by adults in Territory 3 is 13,031, so by way of example, total premiums that would be paid for such a vehicle are shown, along with the proportion of that total premium that would consist of the flat (non-mileage-based) rate.

Table 7. Comparison of three pricing schemes for vehicles driven by adults in Territory 3.

	Pure premium for compulsory coverage		Estimated retail price for full coverage		Cost for average vehicle (adult in territory 3 driving 13,031 miles/year)	
	Yearly	Per mile	Yearly	Per mile	Total[30] premium	Proportion flat
Status quo	$ 130.64	0.00¢	$ 718.52	0.00¢	$ 718.54	100%
Strictly per mile model	$ 0.00	1.00¢	$ 0.00	5.51¢	$ 718.54	0%
2K+ model	$ 59.05	0.65¢	$ 324.74	3.56¢	$ 717.62	45%

As shown above, for an *average* vehicle, the total retail premium is exactly the same whether pure car year pricing or pure mileage pricing is used, and essentially the same in the 2K+ model as well[31]. In any event, someone in this rate group who drives less than 13,031 miles per year could expect to pay less under any of the mileage-based schemes than under the status quo, and a person with above-average mileage for their rate group could expect to pay more. None of this is terribly surprising.

Note, though, that the "proportion flat" column tells a bit about user incentives. Under the 2K+ model, an average driver in this group could reduce her mileage by half and still only save about 27% on insurance, because about half of her premium is non-mileage-based. Clearly, the "strictly per mile model" provides a stronger incentive to reduce mileage. From society's standpoint this is probably a good thing; from an insurance company's standpoint, it could be good or bad depending on whether the particular miles reduced (miles on particular trips forgone) are more risky or less risky than the average mile.

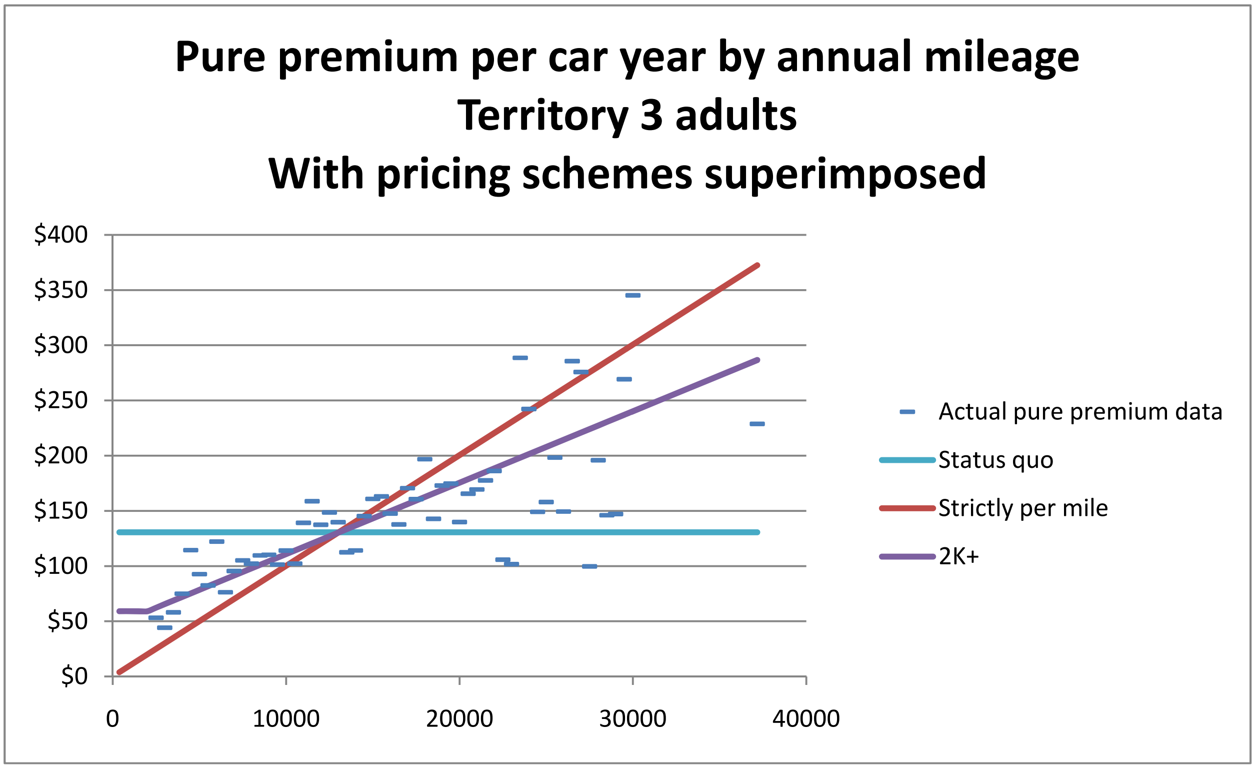
Figure 11 shows the three pricing schemes overlaid on the actual pure premium data first shown earlier in Figure 9 (for adults in territory 3). The graph illustrates a problem with strictly per-mile pricing: it appears to undercharge low-mileage drivers and overcharge high-mileage drivers. Though some of this effect may be due to regression to the mean, it is likely that much of it is genuinely due to driver

[30] All three schemes generate the same $718 total cost for 13,031 miles for an adult-driven vehicle in territory 3. Deviations are due to rounding errors in the calculations.

[31] A peculiarity of the particular 2K+ model that we use is that the average mileage user pays about 90¢ less than in the other models.

heterogeneity within Territory 3 adults. For instance, perhaps on average the low-mileage drivers tend to drive on relatively more urban roads, while the high-mileage drivers use limited-access highways. Though the degree of cross-subsidy is no worse than under the status quo (and probably a bit better), its reversal of direction presents a fresh problem to insurance companies: which customers would end up with strictly per-mile insurance? High-mileage drivers do not want it, but low-mileage drivers are not profitable. This is why the 2K+per-mile model, which fits the risk of the low-mileage drivers fairly well, may offer a more realistic entry point into the per-mile insurance market. For those users willing to use telematic devices to share information on when, where and how they drive with their insurance company and pay an accordingly variable rate, pure mileage-based pricing could probably be quite accurate and no yearly fee would be needed—though in that case, the equipment to monitor driving habits would still have some fixed installation and operating costs.

Figure 11. Illustration of three pricing schemes for Territory 3 adults.



Equity Impacts

Here we briefly discuss the implications of our findings for three types of equity. For our purposes, *fairness* refers to the ideal that a consumer's price for insurance should, to the degree possible, reflect individual and controllable traits of that consumer, rather than involuntary membership in a group. *Horizontal equity* refers to the ideal that similarly situated individuals should pay similar amounts. Finally, *vertical equity* refers to the ideal that low-income people should not shoulder undue financial burdens.

Fairness

In terms of *fairness*, there is much fault to be found in the status quo. Individuals are grouped into rate groups by, among other things, class and territory distinctions over which a driver has limited control[32]. Those rate groups each form separate risk pools and, if you find yourself in a pool with a lot of risky drivers, you will pay a lot for insurance even if your own driving behavior is low-risk—for instance, because you drive very little.

Pay-As-You-Drive insurance offers some improvement in terms of fairness, because at least you control how many miles you drive. However, our analysis shows that PAYD is most likely to be implemented in conjunction with existing class/territory distinctions so that the per-mile rate that you pay still varies depending on your address, age, gender, marital status and so on. Ultimately, though, territory is just a proxy for what kind of roads you drive on, and age (technically years of driving experience) is probably a proxy for how you drive. Telematic devices offer hope of charging individual customers varying per-mile rates depending on where, when and how they drive. Our analysis already shows that mileage carries more explanatory power when used with class and territory (15%) than when used alone (9%). This is suggestive that someday soon, drivers willing to share with their insurance company detailed data on their driving habits may be able to enjoy insurance pricing based more directly on controllable individual factors.

Horizontal equity

Our analysis suggests that, without detailed tracking of when, where, and how miles are driven, PAYD is most likely to be offered in conjunction with existing rating factors, so that class/territory groups are separate risk pools, just as today, and pay different per mile rates. As long as the risk pools remain separate, adding a PAYD component will not benefit one class/territory group at the expense of another, and so there will be no change in horizontal equity compared to the status quo. For instance, the often-expressed[33] concern that PAYD will benefit urban drivers but hurt rural drivers is moot. Rural

[32] The insurance industry uses even finer-grained distinctions than we do here but the additional factors, such as age and occupation, also tend to be broad categories bearing little relationship to an individual's driving behavior. So-called 'merit rating' plans do adjust future insurance premiums based on a driver's past record of at-fault accidents and traffic citations. While desirable and actuarially justified, accidents are (fortunately) too rare for such merit rating plans to account for the bulk of the statistical variability in risk across all drivers. See Ferreira (1972).

[33] See, for instance, Guensler et al 2002, 9; Bordoff and Noel 2008, 41; Litman 2008a, 53.

policyholders drive more miles but would pay a much lower per-mile rate than urban drivers, and the average driver in each group would be no better or worse off than today.

That said, the fact that under the status quo otherwise similar drivers in different territories pay such different insurance rates—differing as much as threefold—can be viewed as horizontally inequitable. It is generally urban territories that are more risky and therefore pricier, and much of this is simply because the higher traffic density on city streets makes for more insurance claims. However, another part of the reason could be that low-income urban neighborhoods have a higher incidence of insurance fraud and reckless driving than suburban or rural areas do. Even in those neighborhoods it is presumably quite a small percentage of individuals responsible for those problems, and while clearly *someone* has to pay the costs imposed by fraudulent or reckless drivers, it may be considered inequitable to collectively punish all the law-abiding drivers in the same neighborhood rather than spreading the costs out evenly over the whole state.

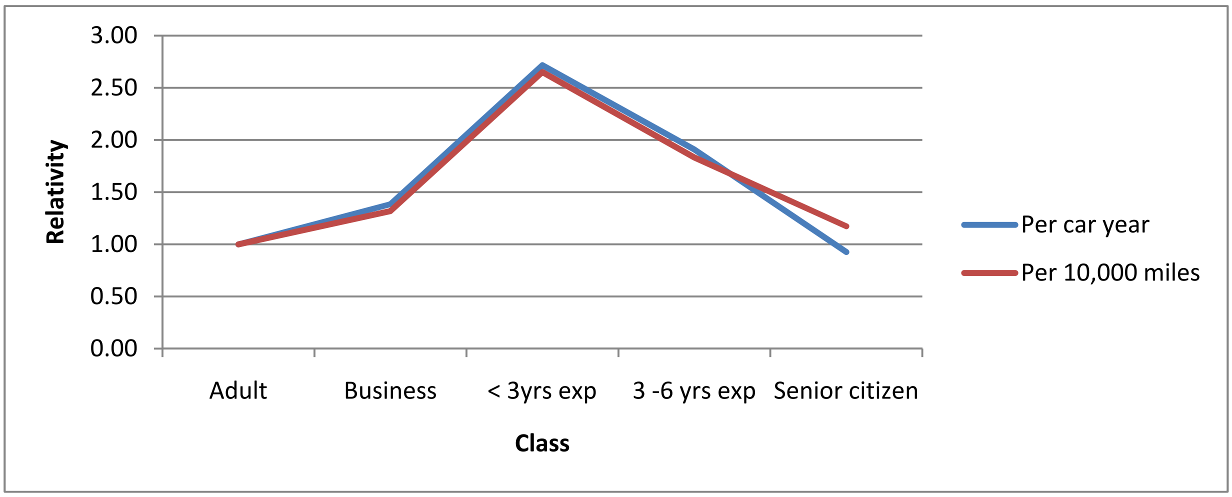
One hope for PAYD has been that it would eliminate or at least reduce the *relativity* or rate difference between different groups of drivers. Table 8 shows, however, that it does not. Based on property damage and personal injury pure premiums, we compute relativities[34] per car year (representing the status quo) and per 10,000 miles (representing strictly per mile pricing). Most of the relativities are quite similar either way, which is unsurprising since, as shown in Table 3, average annual mileage does not differ greatly across groups. Note that the numbers given in Table 8 are derived from a simple regression without interaction terms. So in this simple model, a less-than-3-years experience, Territory 6 driver has a per car year relativity of 2.72 * 2.77 = 7.53. In a more nuanced model, each class and territory combination will have an interaction term so that the risk for a given class/territory group may be a bit more or less than the product of its class relativity and its territory relativity.
The only class group with substantially different annual mileage than the others is senior citizens, as seen in Figure 12. Per car year, senior citizens are less risky than adults, but it turns out that this is because senior citizens drive so many fewer miles on average (7,500 miles per year versus 12,400 miles according to our estimates). Per mile, senior citizens are actually more risky, and so have a somewhat higher per mile relativity. Note that, as discussed above, this does not imply that senior citizens will be hurt by PAYD: the average senior citizen will pay the same amount under either pricing regime.

[34] Recall that our relativities use the low-rated adult in Territory 1 as the reference category, so our relativities for high-risk categories will appear to be larger than typical industry relativities that have been rebalanced to 1.00 for the statewide average.

Table 8. Relativities for nine classes and six territories per car year and per 10,000 miles

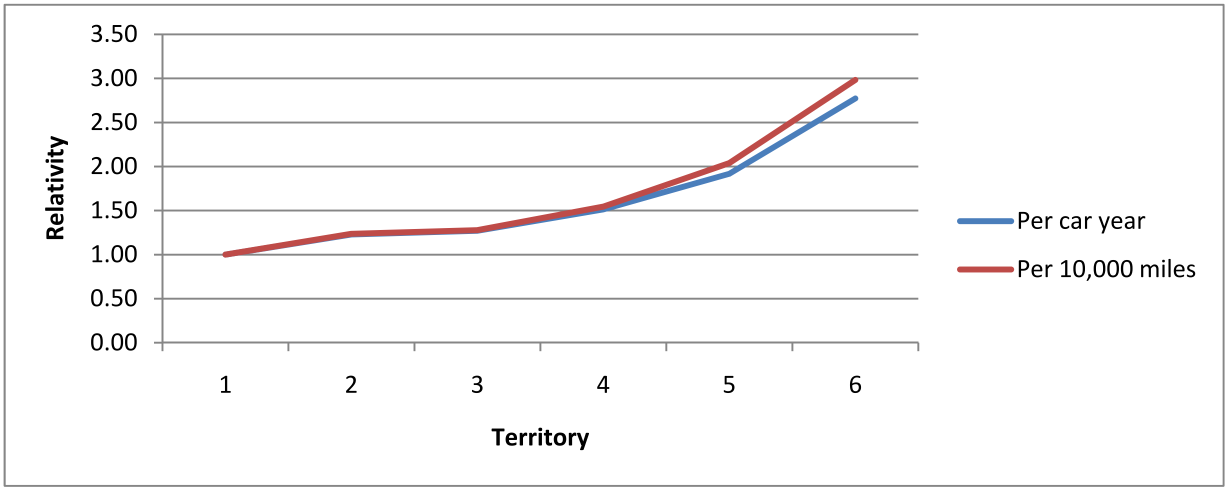
	Relativities			Relativities	
	Per car year	Per 10,000 miles		Per car year	Per 10,000 miles
CLASS			TERR		
Adult	1.00	1.00	1	1.00	1.00
Business	1.38	1.32	2	1.23	1.24
< 3yrs exp	2.72	2.65	3	1.27	1.28
3 -6 yrs exp	1.91	1.83	4	1.51	1.55
Senior citizen	0.93	1.17	5	1.92	2.04
			6	2.77	2.98

Figure 12. Class relativities per car year and per 10,000 miles



For territories, the relativities are actually heightened, though only slightly, as shown in Figure 13. This is because the riskier territories are more urban in character (for instance, most neighborhoods of Boston fall into Territory 6). Driving in cities is riskier per mile (in terms of insurance claims, not the risk of fatality) than in suburbs or in the country, but cars are also driven fewer miles in the city, which softens the relativity when insurance is priced per car year. According to our mileage estimates, the average car in Territory 6 is driven 10,500 miles per year while the average car in Territory 1 is driven 12,500 miles per year.

Figure 13. Territory relativities per car year and per 10,000 miles



Although territory relativities are very slightly higher under PAYD, per mile pricing is probably a plus for horizontal equity across class/territory groups. A law-abiding driver who pays a high per mile insurance rate due to being grouped together with a disproportionate number of fraudulent or reckless drivers at least has the option to save money by reducing mileage. Moreover, PAYD may reduce the number of uninsured drivers by making insurance more affordable to low-mileage drivers.

Horizontal equity effects *between* rate groups aside, PAYD will certainly improve horizontal equity ***within*** rate groups by reducing or eliminating the cross-subsidy from low-mileage drivers to high-mileage drivers. While more elaborate telematic pricing would probably be even more accurate, the simple version of PAYD that we present here improves actuarial accuracy, and would have each user pay a premium more tailored to their own expected losses given their driving behavior.

Vertical equity

To examine any effects PAYD might have on vertical equity, we link our data to 2000 U.S. Census data on household income. For obvious reasons, the Census does not release income data with any identifying information that would allow it to be linked to individual VINs in our study. Instead we take household income by census-designated place (CDP), and join it to insurance data by town listed for each insurance policy. Unfortunately, there is not a perfect match: of the 351 insurance towns plus the few within-town rating territories in Massachusetts, some are neighborhoods of Boston, and others are towns too small to have CDP status. After creating aliases to match different spellings or abbreviations, we find 163 matches between CDPs and insurance towns. With more effort a match could presumably be found for all 360 insurance towns; this could be a worthwhile point for further study[35].

One thing which is clear from our simple analysis is that the parts of Massachusetts that are highest-risk and carry the highest relativities—represented here by Territories 5 and 6—also have lower average

[35] MassGIS also has geocoded locations for individual VINs. This information is not part of the public dataset we use here, but if it were made available in the future, it would be possible to join our data to block group-level census data for a finer grained analysis of income and insurance or mileage data than we conduct here.

incomes than the lower-risk territories. Indeed, Table 9 shows that income decreases almost monotonically over territory as risk rises.

Table 9. Average household income for six territory groups.

Territory	Average household income[36]	Number of towns[37]
1	$ 60,859	43
2	$ 57,703	41
3	$ 51,505	15
4	$ 56,130	32
5	$ 45,025	22
6	$ 37,927	10

Above, we point out the horizontal equity issues that arise from grouping people by territory—a good driver in Territory 6 shoulders more of the cost of 'bad' drivers than a good driver in Territory 1. Table 9 shows that this is also a vertical equity issue: the Commonwealth's lowest income towns and neighborhoods are also the ones that pay the most for insurance under the status quo.

To reiterate, the fact that (as shown in Figure 13) Territories 5 and 6 have a higher mileage-based relativity than per car year relativity does *not* mean that PAYD would be regressive. The mileage-based relativities are higher than car year relativities precisely because cars in Territories 5 and 6 are driven fewer miles each year than cars in Territory 1, and the average driver in Territory 5 or 6 would be no better or worse off under PAYD than under the status quo—*if* she continued to drive the same number of miles. Therein lies one opportunity for PAYD to have a positive impact on vertical equity: low-income drivers are afforded a new opportunity to save money if they choose to drive less. Indeed, the urban drivers with the highest per mile rates are precisely the drivers with the most realistic alternatives to driving if they wish to reduce mileage.

Furthermore, when drivers reduce mileage in response to the incentives of per mile pricing, they will be reducing the toll of automobile externalities on non-car owners. Intuition suggests that non-car owners are disproportionately low-income, and the data in Figure 14 agree. While, again, Census data does not allow household-level comparisons, the regression across Massachusetts CDPs shown in Figure 14 strongly suggests that low income does indeed correlate with low car ownership[38]. A carless city dweller breathes in car exhaust fumes, sits on a bus stalled in the congestion caused by private automobiles, and

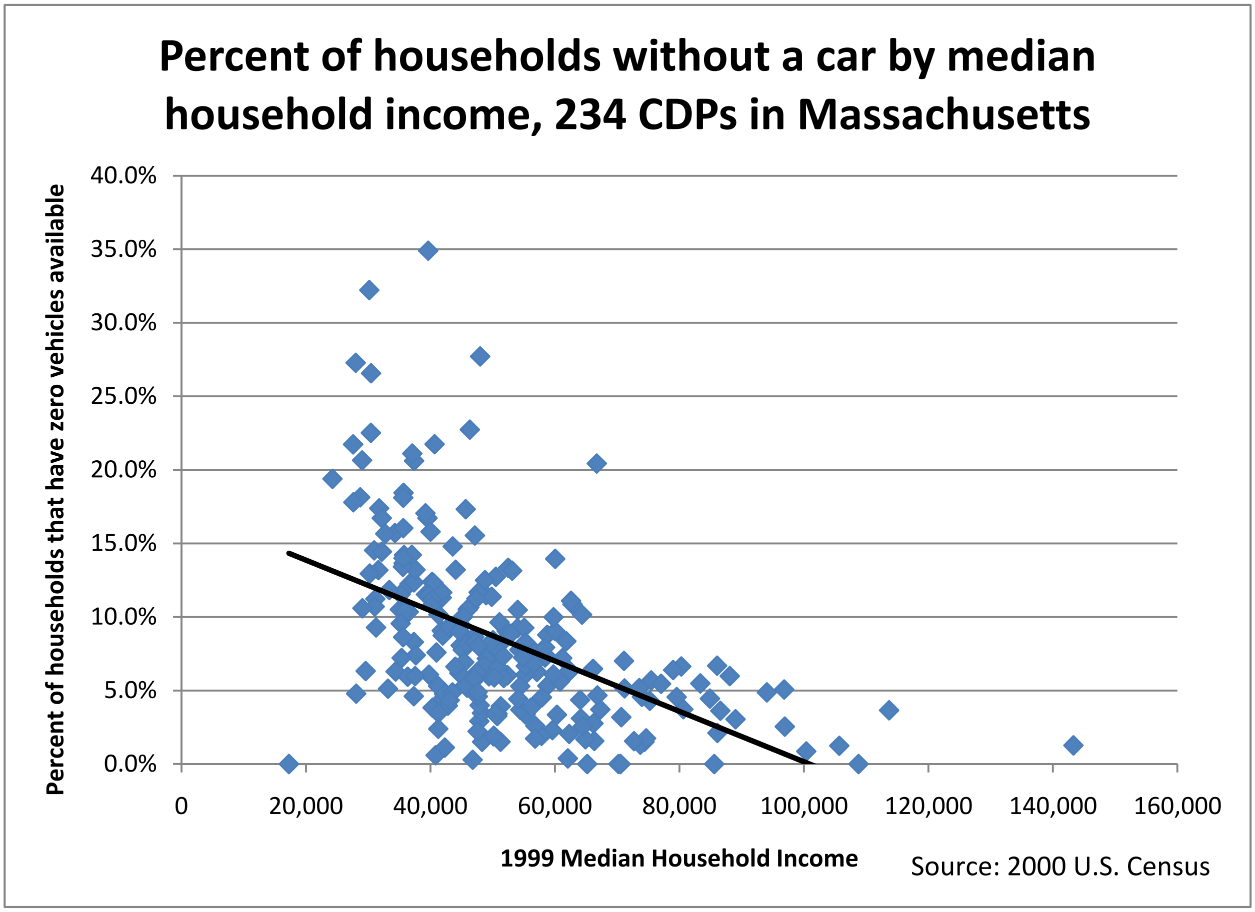
[36] This is actually the exposure-weighted average of each town's 1999 median household income according to the 2000 U.S. Census.

[37] This is the number of towns on which the average income is based. Recall from Table 2 that Territories 3 and 6 are smaller (in terms of exposure) than the others to begin with. Also, many of the Territory 6 insurance towns that failed to match a CDP are relatively low-income Boston neighborhoods such as Roxbury and Dorchester, so it should not be expected that a better match would raise the average income computed for Territory 6.

[38] Clearly, both variables are also correlated with urban form, as cities have more poverty and lower car ownership. But better econometrics would not change the observation that Massachusetts cities contain a large number of carless low-income households who would benefit from a reduction in other people's driving. In fact, they would particularly benefit because cities, where per mile premiums would be highest, could expect to see the largest reductions in driving.

risks being hit and killed by a motorist every time she crosses the street. To the degree that PAYD reduces driving—particularly urban driving, where per mile fees are highest and non-car owners most numerous—this individual will be better off.

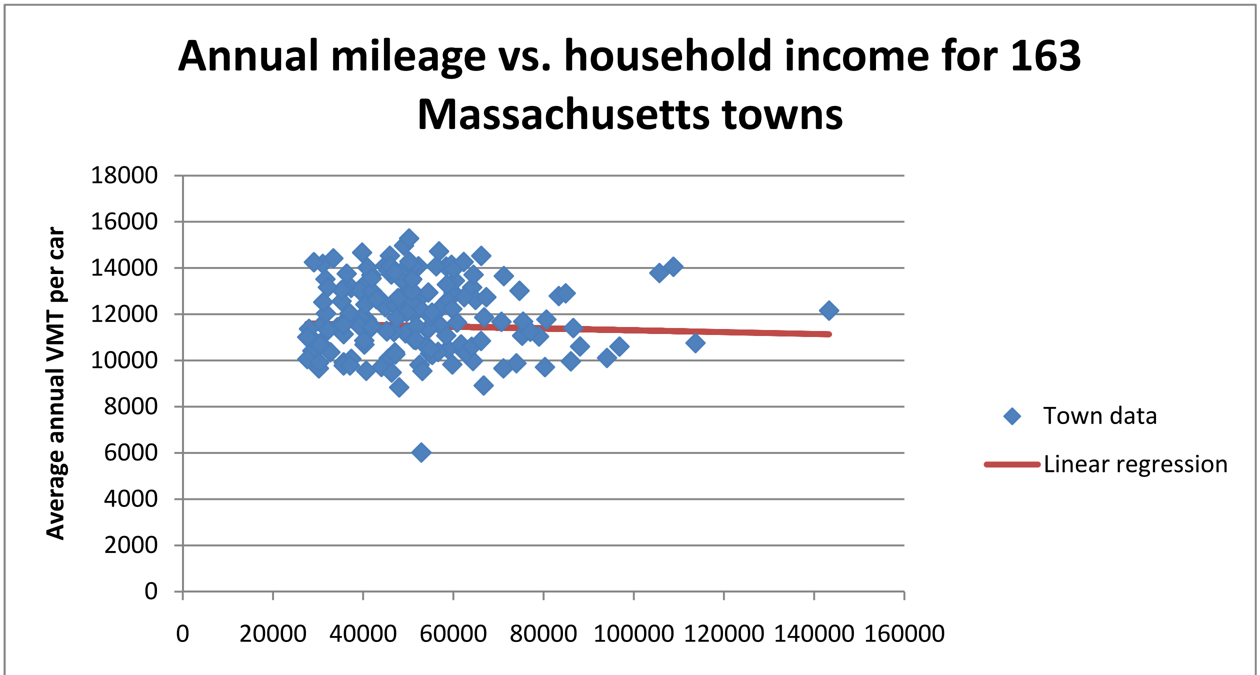
Figure 14. Linear regression shows that low income is correlated with non-car ownership.



To the extent that those individuals are unable to own a car precisely because of the high, fixed annual insurance cost, they may also benefit from the opportunity to own a very low mileage vehicle with PAYD insurance, thus opening up new employment opportunities.

Although distributional impacts across class/territory groups are not possible if PAYD is implemented in conjunction with those groupings, there is room for distributional impacts within a class/territory group. Available evidence suggests, though, that those impacts will be negligible. As discussed earlier, household-level data is not available, but Figure 15 below plots average annual mileage per vehicle (based on our estimates) against Census household income data for the 163 matching towns and shows no correlation between income and VMT/car. The data points appear as a cloud. The exposure-weighted regression actually has a negative slope, -.004, implying that for every $1000 of household income, a car is driven four miles fewer, but with a p value of .56, this is not at all statistically different from a slope of zero and the R2 of .002 shows that the regression explains none of the variation in annual mileage across towns.

Figure 15. Linear regression shows no correlation between income and annual VMT/car



Regression weighted by exposure.
Slope has p value .56; $R^2 = .002$

Of course, the regression in Figure 15 does not control for other variables, such as urban form, which probably account for much of the variation in annual mileage between towns. It is imaginable that income and VMT/car do exhibit some correlation either once controls are applied, or within towns rather than between towns[39]. However, it would be unusual if a strong correlation existed and yet were completely invisible at the town level without controls[40].

As a final note on equity issues, we expect that telematic pricing would be even more equitable, both horizontally and vertically, than simple PAYD. It seems unavoidable that urban territories will carry high per mile insurance rates because traffic density on urban streets makes for frequent insurance claims. However, while the per mile pricing of plain old PAYD discourages people who live in the city from driving anywhere, telematic pricing would discourage people who live anywhere from driving in the city. Whether or not residents in a high-risk urban residential neighborhood want to reduce their own mileage, they would probably agree that reducing other people's mileage in their neighborhood is a good thing.

[39] This relationship between urban form, income, and VMT has been examined by MassGIS and, in further detail, by studies at MIT in the first author's research group. These results have not yet appeared in the literature. They indicate that, after controlling for other factors, higher income households travel more miles per household - but the same or even fewer miles per car - than low income households.

[40] High income households travel more miles than low income households, but they also own more cars over which to spread that mileage. If high income households have slightly lower mileage per vehicle, this difference is probably already being priced by the current system through multi-car discounts offered by insurance companies. On the other hand, high income households may drive more miles per car than low income households, which is what Bordoff and Noel (2008, 10), examining national data, conclude.

Vehicle Miles Traveled and Environmental Impacts

We would like to know the impact that implementation of PAYD in Massachusetts might have on the state's greenhouse gas (GHG) emissions, specifically GHG emissions from private automobiles. To do this, we begin by modeling the effect of PAYD on VMT. Next we use fuel economy data to translate a VMT reduction into a fuel reduction, and finally we assume that GHG emissions are proportional to fuel use.

Central to any effort to model the effect of PAYD on VMT is an assumption about elasticity of VMT with respect to the marginal price per mile of private vehicle travel. Edlin (2002) and Bordoff and Noel (2008) have modeled the VMT impacts of PAYD, and both rely on elasticity estimates that other researchers have derived from studying the changes in VMT brought about by gasoline price changes. An extensive discussion of methods and empirical results for calculating transportation elasticities is available in Litman (2010a)[41].

The number that both the Edlin and the Bordoff & Noel studies settle on is -.15, but with different meanings. Bordoff and Noel appear to treat this as a vehicle's elasticity of ***miles*** with respect to gasoline price. Edlin takes -.15 as the short-run elasticity of ***aggregate gasoline demand*** with respect to gasoline price; he then translates this aggregate figure into a separate *miles* elasticity (a number he never states) based on the distribution of fuel economy over the vehicle fleet.

To select a value for use in this study, we review the sources cited for each approach. Bordoff and Noel borrow their figure from Parry (2005), who in turn reasoned as follows. He cites Parry and Small (2005), who find -.55 as the elasticity of fuel demand with respect to fuel prices, and find that 40% of this elasticity is due to reduced aggregate VMT (the other 60% being due to fuel economy improvements). He then cites Johansson and Schipper (1997) who find that of the elasticity of *aggregate* VMT, 33% is due to reduced vehicle ownership and 67% is due to reduced mileage per vehicle. Therefore the elasticity of individual vehicle mileage with respect to fuel price is $-.55 * .4 * .67 \approx -.15$.

[41] In modeling the VMT impacts of PAYD based on findings from gasoline price studies, we are assuming that PAYD consumers will perceive insurance as a direct marginal cost of driving an additional mile, much like they might view gasoline cost. This assumption should hold for *concurrent* pricing schemes where the user actually pays for each marginal mile driven. In this study we do not consider *prospective* schemes where one year's insurance rate is set based on the previous year's mileage, or "mileage rate factor" schemes where the user's self-reported or predicted mileage is used to set a per car year insurance rate. Under such 'prospective' or 'mileage factor' schemes, the marginal cost of each additional mile is less evident and considerably delayed, so the incentive effects are likely to be reduced compared with concurrent pricing schemes.

In choosing an aggregate price elasticity of demand for gasoline, Edlin reviews several studies, finds that the most inelastic estimate[42] is (negative) .2, and then chooses (negative) .15 as his own figure, apparently so as to be conservative, though he never states a reason. Though Edlin treats -.15 as a price elasticity of demand for gasoline, our review of the sources he cites indicate that -.15 could just as easily be taken as a miles elasticity and still be fairly conservative.

Bordoff and Noel (2008, Appendix A) obtain a mileage reduction by multiplying their elasticity of -.15 by the ratio of the per mile PAYD insurance rate (P_I) to the per mile price of fuel (P_F) as shown in Equation 8.

Equation 8. Model for per-vehicle mileage reduction implied by Bordoff and Noel (2008, Appendix A)

$$\Delta M = -.15\frac{P_I}{P_F}$$

Edlin, on the other hand, applies his miles[43] elasticity, e, to the per mile insurance rate (p_i) divided by the *total* per mile operating cost of the vehicle, which he takes to include 9.2 cents of maintenance and depreciation as shown in Equation 9.

Equation 9. Model for per-vehicle mileage reduction implied by Edlin (2002, 26)

$$\Delta M = e \times \frac{p_i}{4.2\ cents\ maintenance + 5\ cents\ depreciation + per\ mile\ fuel\ price}$$

We choose to use Bordoff and Noel's approach for two reasons. First, it is debatable whether drivers really consider the resale value of their car or the cost of an impending oil change when choosing whether to drive somewhere. Second and most importantly, all of the researchers that estimated the elasticities used by Bordoff and Noel or Edlin were estimating elasticity of miles or gasoline demand with respect to gasoline price only, not with respect to changes in total operating cost[44]. Since the -.15 figure was derived assuming gasoline price as the entire denominator, it should be applied to a model with gasoline price as the entire denominator. Note that in both models, the per mile insurance price does not appear in the denominator. This is because under the status quo, the user's marginal insurance price to drive an additional mile is zero.

Taking all of the above into consideration, we arrive at the model shown in Equation 10.

---

[42] In fact, the estimates he reviews are of different things. Some estimate the short run price elasticity of demand for gasoline: Dahl and Sterner (1991) give estimates between -.2 and -.3, Goodwin (1992) arrives at -.27 and Gallini (1983) comes to -.3 to -.4 and figures 84% of it is due to mileage reduction. In general, Edlin assumes that short run price elasticity of demand for gasoline is mostly due to mileage reduction, whereas long run price elasticity of demand for gasoline includes a shift to more fuel-efficient vehicles. Edlin also quotes a long run elasticity of miles per car with respect to gasoline price of -.2 from Johansson and Schipper (1997) and notes that Goldberg (1998) estimates almost zero miles elasticity but allows that for large price changes it might be -.2

[43] Edlin never explicitly states a value for his miles elasticity, but it must be close to his gasoline demand elasticity of -.15.

[44] Goldberg (1998) mentions maintenance and depreciation with regards to a household's decision to hold an old car or purchase a new one, but not with regard to per mile operating cost. Gallini (1983) considers car ownership costs as fixed costs not proportional to mileage. Parry (2005), Parry and Small (2005), Johansson and Schipper (1997) and Goodwin (1992) make no mention of maintenance or depreciation whatsoever with regards to their decision model.

Equation 10. Our model for each vehicle's mileage reduction in response to PAYD.

$$\Delta mileage_i = -.15 \frac{per\_mile\_insurance\_rate_i}{\frac{gasolineprice}{mpg_i}} mileage_i^0$$

Where:

$\Delta mileage_i$ = the change in vehicle i's annual mileage,
-.15 = the assumed elasticity of mileage with respect to per mile price,
$per\_mile\_insurance\_rate_i$ = the marginal rate (in cents per mile) a policyholder pays to drive vehicle i,
gasolineprice = the price of gasoline (in cents per gallon),
mpg_i = the fuel economy of vehicle I (in miles per gallon), and
$mileage_i^0$ = our estimated annual mileage for vehicle i

As a simplifying assumption, we apply a single elasticity value to all drivers. In reality, each individual has a different elasticity which may depend on, among other things, income (where lower-income people are more price sensitive and so have a higher elasticity) and the availability of alternatives to driving (where people who live in areas where transit, walking or biking are realistic options will be more able to reduce mileage and so will have a higher elasticity).

As of this writing, gasoline costs about $2.72 per gallon in Massachusetts[45], and the 12-month rolling average of past gasoline prices nationally is $2.71[46]. We therefore assume 270 cents per gallon as the gasoline price for our model.

Applying this model to the data requires a few additional assumptions. Though in the short term we expect that only low-mileage drivers would switch to PAYD, for the purposes of modeling, we wish to explore the maximum possible impact that PAYD could have, and so we assume that all policies are switched to PAYD.

We also need to assume a per mile marginal insurance price, p. We run the model twice, once with a strictly per mile pricing scheme and once with the "2K+per-mile" model described above, which has slightly lower per mile prices as well as a fixed cost for each vehicle's first two thousand (annual) miles. In each case, we apply a multiplier of 5.5 to translate our calculated pure premiums into typical retail insurance prices for full coverage.

For fuel economy we use, as described above, the average of the EPA city and highway gas mileage estimates, adjusted downward by 12% to account for some of the inflation in EPA estimates.

After computing a VMT reduction for each vehicle, we also use its fuel economy to compute an estimate of fuel consumption reduced and apply a simple formula provided by the EPA to compute greenhouse

[45] http://www.fuelgaugereport.com/MAavg.asp accessed July 7, 2010.

[46] $2.71 based on June 2009 – May 2010 figures given by the Energy Information Administration: http://tonto.eia.doe.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=mg_tt_us&f=M

gas reduction[47]. Finally we sum over all vehicles in the study to estimate the aggregate reduction in VMT and fuel use for Massachusetts.

Since our VMT reduction predictions are sensitive to our choice of elasticity (-.15) and our multiplier to convert pure premiums into full retail prices (5.5), we conduct a brief sensitivity analysis. We compare our results to those that would be obtained if the elasticity were -.22, a figure from Small and Van Dender (2006) which is among the largest mentioned in a literature review by Litman (2010a and 2010b)[48], and to results obtained if the multiplier were just 3.0, reflecting that many consumers may not choose full coverage[49].

VMT Reduction Results

For the strictly per mile pricing scheme, we obtain a VMT reduction of 9.5% and a corresponding fuel reduction of 9.3%[50]. As expected, fuel consumption is reduced almost exactly in proportion to VMT. These figures correspond to aggregate reductions of 4.1 billion VMT, 194 million gallons of fuel and 1.8 million metric tons of CO_2-equivalent in our model. However, those figures do not represent all private passenger vehicles in the state. In the first place, our study only includes insured vehicles (though this is probably realistic, as uninsured vehicles would see no per mile price unless their owners chose to begin buying insurance under PAYD), and then only the 71% of car years of exposure for which the vehicles had mileage estimates available. Our VMT reduction model is then applied only to the 96% of vehicles in the study for which fuel economy estimates are available. Including all insured drivers, the proportional reductions would not change but the aggregate reductions would probably be closer to 5.9 billion VMT, 286 million gallons of fuel and 2.6 million metric tons of CO_2-equivalent. Including uninsured drivers as well would lower the proportional reductions but leave the new aggregate figures intact.

For the "2K+per-mile" pricing scheme, we estimate a 5.0% reduction in VMT and 4.9% reduction in fuel use, corresponding to 2.1 billion VMT, 102 million gallons of fuel and 950,000 metric tons of CO_2-equivalent. Again, extrapolated to all insured drivers, these figures are probably closer to 3.1 billion VMT, 150 million gallons of fuel and 1.4 million metric tons of CO_2-equivalent. These figures are all lower than for the strictly per mile pricing scheme because, in the "2K+per-mile" scheme, the insurance cost of the first 2K miles is bundled into a flat yearly fee, and the average per mile price beyond the 2,000th mile is much lower than if all costs were charged per mile. This lower per mile price gives

[47] 8.8 kg of CO_2 per gallon of gasoline; CO_2 itself comprises 95% of the CO_2-equivalent emissions so we multiply by 100/95 to account for other GHGs emitted in exhaust. See http://www.epa.gov/oms/climate/420f05004.htm

[48] Litman reviews a range of fuel demand and VMT elasticities and argues that elasticities are probably higher today than during 1960-2000 when many studies of the subject were conducted. Small and Van Dender (2006) settle upon a long run VMT elasticity with respect to fuel price of -.22 but note that -.11 is the result obtained based on 1997-2001 data only. Since gasoline prices reached record lows during that time period, -.11 may be considered as a very low estimate. Indeed, we believe that our chosen value of -.15 is already fairly conservative.

[49] Pure premium is usually 2/3 of the retail premium for any given type of coverage, so a multiplier of 1.5 would indicate that *no* consumers choose full coverage. Allowing that many consumers will choose full coverage—and that our pure premiums are underestimated due to capping losses at $25,000—a multiplier of 3.0 appears to be on the low end of what might be imagined.

[50] Recall that these are percentage reductions of private passenger vehicle VMT and fuel consumption. Commercial vehicles are excluded from our study.

consumers a lesser incentive to reduce mileage, but does a better job of fitting the pure premium risk observed in the data (as shown above, for example, in Figure 11).

All of these predictions depend on an elasticity gleaned from literature on gasoline price changes. A key assumption of our analysis is that PAYD is implemented as a concurrent pricing scheme where the *mile* becomes the unit of exposure, instead of or in addition to car years, and users actually pay a price for each marginal mile driven. Pricing schemes where mileage is more indirectly related to the price consumers pay could not be expected to produce nearly as much of a reduction in VMT.

Our results might be taken as somewhat conservative for the following two reasons:

- Our assumed miles elasticity of -.15 is rather small compared to most estimates in the literature that we review.
- We apply this same elasticity to all drivers, even though the households with the highest insurance costs (both under the status quo and PAYD) tend to be lower income and to live in cities where alternatives to driving area available, meaning that their response to PAYD might be quite a bit more elastic.

On the other hand, there are also some factors that point in the other direction:

- The 12% adjustment we apply to reduce the EPA fuel economy estimates is probably not enough to bring pre-2008 EPA fuel economy estimates into line with reality. Therefore current fuel economy is likely to be a bit worse than we estimate, and per mile fuel costs higher than we suppose. Therefore the percent change in total per mile costs resulting from PAYD may be slightly smaller than we suppose.
- We do not model any increase in vehicle ownership resulting from PAYD[51]. The modeling would be rather complicated but could offset some of the mileage reduction from current drivers.
- Our model does not equilibrate all relevant factors. In reality, if VMT is reduced, insurance claims should be reduced more than proportionally, and as a result, per mile insurance rates will fall slightly.
- Our extrapolation factor of 5.5 translates the numbers in our study (that are based on data for compulsory coverages) into the cost for 'full insurance coverage." But some motorists will choose not to buy "full insurance coverage" thereby reducing some of the marginal pricing incentive, and some elements of 'full coverage' (such as theft insurance) may be priced differently if they do not exhibit the same per-mile risk.

On the whole it is difficult to say if our estimates are biased upward or downward. Our estimate of 9.5% is lower than that found by Bordoff and Noel (2008), who calculate an 8% reduction in VMT nationwide but a 10.7% reduction in Massachusetts due to its above-average insurance premiums. Our strictly per mile pricing model gives a mileage-weighted average premium of 8.2 cents (or an exposure-weighted average premium of 8.6 cents), not so different from the 8.7 cents per mile used by those researchers. However, Bordoff and Noel applied that per mile rate universally to all drivers in Massachusetts, whereas we divide by class and territory. In terms of territory, the riskiest (therefore highest per mile price) drivers are those who already drive fewer miles per year, so that probably played some role in

[51] Bordoff and Noel (2008) and Edlin (2003) also commit this omission.

dampening the VMT impact of PAYD in our analysis. However, allowing elasticity to vary by household income and availability of alternative transportation options might raise our estimate of VMT reduction.

The tables below briefly explore the sensitivity of our predictions to two input values: our elasticity of -.15 and our full premium multiplier of 5.5. We calculate VMT reduction impacts with Small and Van Dender's larger elasticity of -.22, in addition to our more conservative -.15, and with a multiplier of 3.0, which we believe is a lower bound, in addition to our estimate of 5.5.

Table 10. Percentage VMT reduction under strictly per-mile pricing scheme with various assumptions of elasticity and multiplier.

Elasticity	Multiplier	Predicted VMT reduction (with pure per mile pricing)
-.15	3.0	5.2%
-.15	**5.5**	**9.5%**
-.22	3.0	7.6%
-.22	5.5	14.0%

Table 11. Percentage VMT reduction under 2K+per-mile pricing scheme with various assumptions of elasticity and multiplier.

Elasticity	Multiplier	Predicted VMT reduction (with 2K+per-mile pricing)
-.15	3.0	2.7%
-.15	**5.5**	**5.0%**
-.22	3.0	4.0%
-.22	5.5	7.3%

As Table 10 and Table 11 show, VMT reductions could vary dramatically depending on a few key assumptions. This is one reason why it would be unwise to put too much faith in any particular estimate of driving reduction due to PAYD. Another reason is that consumer response to fuel prices (on which our elasticity estimates are based) may be a poor predictor of how consumers respond to PAYD. When the price at the pump changes, people are simply paying *more* for a good they are already used to paying for. PAYD would have people paying for something (a marginal mile of insurance) that they have never paid a cent for before, and their reaction could vary – at least in the short term. Nevertheless, even our lowest plausible VMT reduction (2.7%) would save more than a billion miles annually and millions of tones of GHG.

Any such reduction in VMT due to the incentive effect of per-mile pricing would yield a more or less proportional reduction in the number and cost of auto accidents. That is, if VMT drops 10%, then the reduction in accidents and accident costs will be about the same – 10%. The exact amount would depend on whether the forgone miles were more or less risky compared with the average mile[52].

[52] There could also be some additional reduction in accidents due to reduced congestion effects whereby a driver's mistake no longer results in a multi-car accident if no car is close by.

Conclusions

The unprecedented size and comprehensiveness of our dataset allows us to explore the implications of Pay-As-You-Drive insurance at a level of detail, and with attention to underlying accident risk, that has not been possible to date. Our analysis of insurance claims and annual mileage estimates indicate that:

- Mileage is positively correlated with risk, and the correlation is highly statistically significant.
- The relationship between risk and mileage is less-than-proportional and, when all vehicles are considered together without class or territory differentiation, less-than-linear.
- Mileage provides greater explanatory power when paired with some control on where and how the miles are driven. It is thus a powerful supplement to traditional insurance rating factors (such as driver experience and rating territory).
- More elaborate telematic pricing schemes (that more directly monitored when where and how a vehicle is driven) could probably supplant many traditional risk proxies.
- As an alternative to strictly per mile pricing, a flat rate for the first 2,000 miles and per mile pricing thereafter makes actuarial sense and may make PAYD appealing to low mileage customers while still covering expected losses for insurance companies.

Our analysis of equity issues suggests that:

- PAYD improves fairness by basing premiums at least partly on factors that are more individual and controllable.
- PAYD would not reduce relativities between classes and territories but would improve horizontal equity by eliminating a cross-subsidy from low-mileage drivers to high-mileage drivers.
- There appears to be little correlation between VMT per car and household income, so converting yearly premiums to per mile premiums would have little direct effect on vertical equity.
- PAYD would be a largely progressive measure in that it would afford low income households an opportunity to save money by choosing to reduce VMT, reduce the toll of automobile externalities on low income households that do not own a car, and bring low mileage auto ownership within financial reach of some low income households.

Our analysis of fuel economy data and our modeling of VMT reduction indicate that:

- Average fuel economy exhibits little variation by annual mileage, class or territory, so fuel reductions resulting from PAYD are likely to be proportional to VMT reductions.
- Switching all drivers to pure per mile pricing would reduce both VMT and fuel consumption by about 9.5%. A mixed model with a flat yearly rate plus per mile pricing after the first 2,000 miles would reduce both figures by about 5%.
- Depending upon various pricing, elasticity, and full-coverage assumptions, the VMT and fuel consumption reductions could range between 3 and 14%.
- Any such reduction in VMT due to the incentive effect of per-mile pricing would yield a more or less proportional reduction in the number and cost of auto accidents.

Bibliography

Bordoff, Jason E. and Pascal J. Noel. 2008. "Pay-as-you-drive auto insurance: a simple way to reduce driving-related harms and increase equity." The Brookings Institution. http://www.brookings.edu/~/media/Files/rc/papers/2008/07_payd_bordoffnoel/07_payd_bordoffnoel.ashx

Dahl, Carol and Thomas Sterner. 1991. "Analyzing gasoline demand elasticities: a survey." Energy Economics 13, no. 3: 203-210.

Edlin, Aaron S. 2002. "Per-mile premiums for auto insurance." University of California-Berkeley Department of Economics Working Paper No. E02-318. http://129.3.20.41/eps/le/papers/0303/0303001.pdf

Edlin, Aaron S. 2003. "Per-mile premiums for auto insurance." In *Economics for an imperfect world: essays in honor of Joseph E. Stiglitz*, edited by Richard Arnott, Bruce Greenwald, Ravi Kanbur and Barry Nalebuff. 53-82. Cambridge, MA: MIT Press.

Ferreira, Joseph Jr. 1972. "Driver accident models and their use in policy evaluation." In *Analysis of Public Systems*, edited by Alvin W. Drake. 287-316. Cambridge, MA: MIT Press.

Ferreira, Joseph Jr. 2007. Testimony concerning House #1079 before the Joint Committee on Financial Services of the Massachusetts General Court, April 4, 2007.

Gallini, Nancy T. 1983. "Demand for gasoline in Canada." Canadian Journal of Economics 16, no. 2: 299-324.

Goldberg, Pinelopi. 1998. "The effects of the Corporate Average Fuel Efficiency standards in the U.S." Journal of Industrial Economics 46, no. 1: 1-33.

Goodwin, P.B. 1992. "A review of new demand elasticities with special reference to short and long run effects of price changes." Journal of Transport Economics and Policy 26, no. 2: 155-169.

Greenberg, Allen. 2009. "Costs and benefits of varying per-mile insurance premiums based upon measured risks specific to each mile driven." Federal Highway Administration, http://www.vtpi.org/AG_PAYD.pdf

Guensler, Randall, Adjo Amekudzi, Jennifer Williams, Shannon Mergelsberg, and Jennifer Ogle. 2002. "Current State Regulatory Support for Pay-as-You-Drive Automobile Insurance Options." *Journal of Insurance Regulation* 21(3) (Spring).

Johansson, Olof and Lee Schipper. 1997. "Measuring the long-run fuel demand of cars: separate estimations of vehicle stock, mean fuel intensity and mean annual driving distance." Journal of Transport Economics and Policy 31, no. 3: 277-292.

Litman, Todd. 2008 (a). "Distance-based vehicle insurance feasibility, costs and benefits." Victoria Transport Policy Institute. http://www.vtpi.org/dbvi_com.pdf

Litman, Todd. 2008 (b). "Pay-As-You-Drive insurance: recommendations for implementation." Victoria Transport Policy Institute. http://www.vtpi.org/payd_rec.pdf

Litman, Todd. 2010(a). "Transportation Elasticities: How Prices and Other Factors Affect Travel Behavior," Victoria Transport Policy Institute. http://www.vtpi.org/elasticities.pdf

Litman, Todd. 2010(b). "Changing vehicle price sensitivities: the rebounding rebound effect." Submitted for the Transportation Research Board 2011 Annual Meeting.

Marshall, Wesley Earl. 2009. "Community Design, Road Safety & Transportation Sustainability." PhD Dissertation, Center for Transportation and Urban Planning, University of Connecticut.

Parry, Ian W.H. 2005. "Is Pay-As-You-Drive insurance a better way to reduce gasoline than gasoline taxes?" American Economic Review 95, no. 2: 288-293.

Parry, Ian W.H. and Kenneth A. Small. 2005. "Does Britain or the United States have the right gasoline tax?" American Economic Review 95, no. 4: 1276-1289.

Small and Van Dender. 2007. "Fuel efficiency and motor vehicle travel: the declining rebound effect." Energy Journal 28, no. 1: 25-51.

Vickrey, William. 1968. "Auto accidents, tort law, externalities and insurance: an economist's critique." Law and Contemporary Problems 33, no. 3: 464-487.

Appendix 1. Insurance Data Processing

The database released in March 2010 by the Massachusetts Executive Office of Energy and Environmental Affairs (EOEEA) includes auto insurance policy and claims data provided by the Massachusetts Commonwealth Automobile Reinsurers (CAR). The CAR data provides automobile insurance policy and claims data extracted from statistical plan reports for private passenger vehicles from 2004 through 2008[53]. The released dataset is comprehensive: an exposure table describes the policies written and the class, territory and other attributes of the insured, along with an anonymized policy ID and a real Vehicle Identification Number (VIN). A separate claims table records the filing, payment and adjustment of claims, along with dollar amounts of losses and reserves.

Although five years of data are included in the released dataset, we choose to examine only policy year 2006 due to time constraints and limitations in the corresponding mileage and claims data. As explained in Appendix 2, very little mileage data were included in the Registry of Motor Vehicle data that were released for 2004 and 2005. Meanwhile, policies beginning in 2006 might not end until as late as December 31, 2007 and so with claims data through 2008, we could be assured of at least a full year of development on all claims relevant to the 2006 policy year. If we had examined the 2007 policy year, some claims would not have had sufficient time to develop, and many claims for the 2008 policy year would not even be included in the claims data.

Both exposure and claims data are presented as transactions. For exposure, each cancellation or adjustment of a policy appears as a new record. To determine the total months of earned exposure for each policy and the start and end dates of each endorsement, it was necessary to aggregate these records into one record for each unique combination of VIN, policy ID and policy effective date and sum up the exposure earned under each phase of the policy as it was adjusted.

For claims, we include only those covered by compulsory forms of coverage— bodily injury (BI) and property damage liability (PDL), and personal injury protection (PIP)[54]. In aggregating over transactions, we consider each claim's total losses to be the sum of all its paid losses plus its last reported reserve amount (as of 2008 Q4) after capping paid and reserved losses at $25,000[55]. We also exclude claims with total net losses less than $50, considering these to have been settled without any net payment by the insurer. The capping of losses at $25,000 per claim ignores larger claims that can only arise if one or another of the involved vehicles have optional "excess limits" insurance coverages. By limiting our attention to that portion of each claim that falls within compulsory coverages plus the commonly held lower end of excess limit amounts, we avoid complications arising from having differences in coverage across the vehicles in our study[56]. We use a 5.5 multiplier to translate our figures into retail prices for full coverage (including expenses, commissions, and other loadings).

---

[53] The description and format of the Massachusetts Private Passenger Statistical Plan for various years is available from the CAR website at: http://www.commauto.com/manuals/ppstatplan/ppstatplan.aspx

[54] Specifically, we include only claims with subline code 1 or 5 and loss type codes corresponding to BI (01, 02), PDL (03), or PIP (11, 14, 23, 24, 34, 44, 45).

[55] This roughly corresponds to compulsory coverage limits. Compulsory coverage for BI is $20,000 per person and $40,000 per incident, for PDL is $5,000 and for PIP is $8,000.

[56] For the 2006 policy year, the compulsory coverages included $20,000 per person and $40,000 per accident for bodily injury liability, $5,000 for property damage liability, and $8,000 for personal injury protection.

We identified some data errors when we computed the months of earned exposure for each policy. The first transaction of every policy ought to contain a positive number of exposure months, since each policy starts out on the policy effective date as a 12-month policy. Subsequent transactions might have negative exposure in order to shorten the initial terms of the policy and then (in a new transaction) positive numbers extending the newly changed policy back out to the end of the year (with, for example, a replacement vehicle, a newly added driver, or a change of address). But in practice, about 0.1% of the 8.5 million exposure records for policy year 2006 either began with a negative number of exposure months (for a particular vehicle, policy, and policy effective date combination) or summed to a total number of months less than 0 or greater than 12. In some cases, a net of -1 or +13 could have resulted because all endorsement dates were rounded to the nearest month. But most such cases are likely to be coding errors, or corrections of previous coding errors. Overall, when processing the written premium transactions in order to reconstruct all the intervals of earned exposure for constant policy circumstances, around 0.5% of the exposure records had inconsistent records or duplicate vehicle/policy/effective-date entries concerning one or more mid-year endorsement circumstances. We removed from our analytic dataset all vehicle/policy/effective-date combinations that had one or more problematic exposure records.

Another issue is that the CAR database lists policy effective dates and transaction dates by month only—the actual date is not included. By default, we consider each date to be the first of the month, but when transaction dates and earned exposure months do not match, we add or subtract half months. For instance, if a transaction ends a policy in May but adding the earned exposure prior to that transaction to the policy effective date only adds up to April, then the transaction is considered to have taken place on April 15.

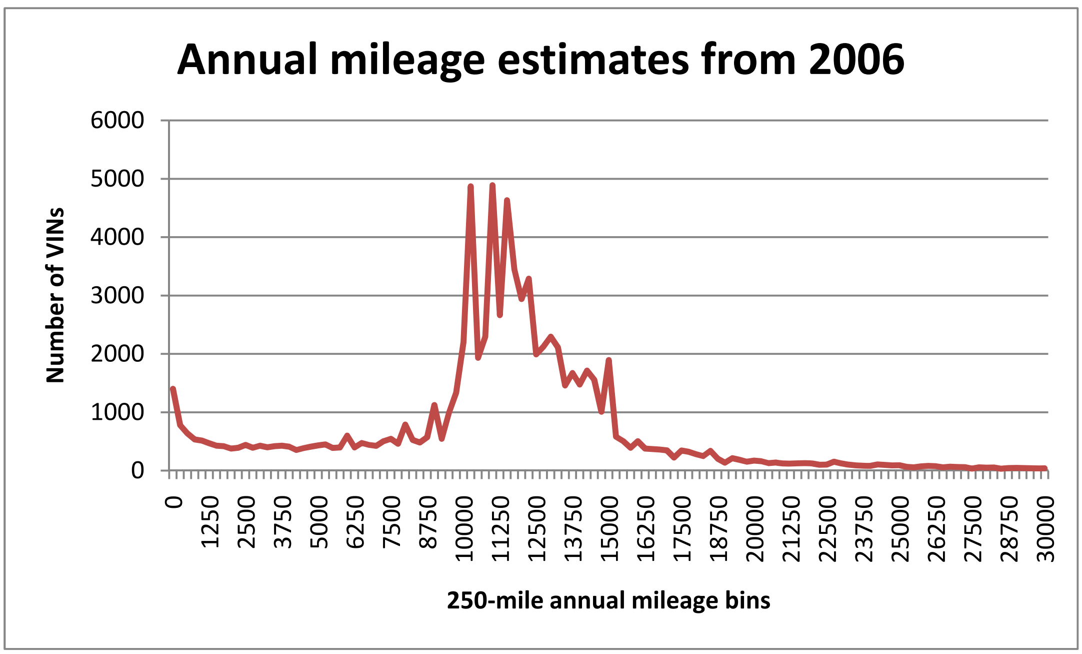
After eliminating the problematic exposure records, plus the 10% of insurance policies for which no mileage data are available, we are left with 3.25M policy-vehicle combinations (including 3.05M distinct VINs) which form the basis for this study. The difference in numbers does not indicate that some vehicles were double-insured. For vehicles that changed hands during the policy year, the first owner's insurance policy would have been cancelled and then the new owner would have insured the vehicle under a different policy number. For cases where two policies match one VIN, we consider both; as long as their durations are less than one year each and do not overlap. This does not constitute double-counting.

Appendix 2. Annual Mileage Estimation

We generated our own estimates of annual mileage for each VIN based on RMV odometer readings conducted at annual safety checks. MassGIS has also released its own mileage estimates for almost all VINs registered in the state, which we originally planned to use for our assessment of mileage-based pricing. The mileage estimates are generated by taking the difference between two odometer readings at RMV safety checks and extrapolating up or down to 365 days of driving.

Though the MassGIS estimates were helpful for most VINs, we created our own estimates in order to have estimates that are more specific to the 2006 policy year during which each VIN is insured. MassGIS's outline of how it processed the RMV data indicated that, due to errors in the original safety inspection data entry, some estimates came out negative or higher than 100,000 miles per year (a fairly implausible number of miles for one vehicle to be driven in a year). MassGIS threw out these estimates and replaced them with the average annual mileage for each town and registration type. Likewise, many VINs had only one odometer reading on record and so the miles driven between two inspections could not be computed. For these VINs as well, MassGIS filled in the average annual mileage obtained from other vehicles. As a result, in Boston there are 71,880 VINS listed as having annual mileage of exactly 10,374 miles, in Worcester there are 20,598 VINs listed with exactly 11,904 miles. There are smaller spikes for many other towns in Massachusetts, with the result that a histogram of annual mileage looks very abnormal. A histogram for just the 2006 estimates is shown in Figure 16.

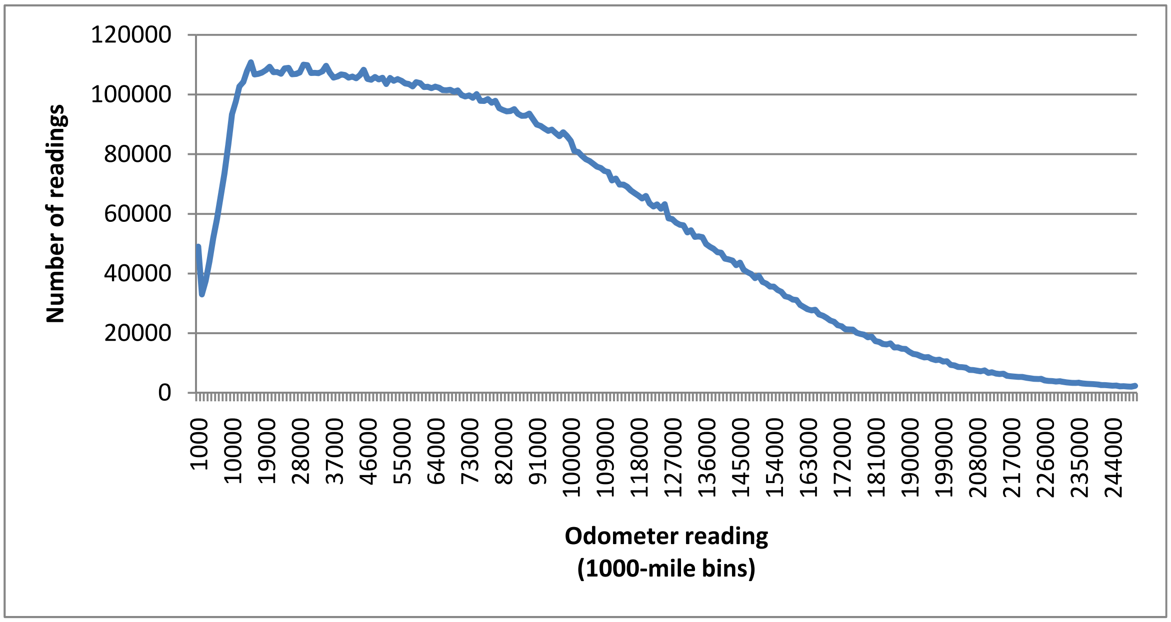
Figure 16. This histogram of annual mileage estimates shows spikes due to "filled in" values



Such "filling in" of data was crucial for MassGIS's original purpose of estimating aggregate VMT for different regions of the state under future growth scenarios, but proves detrimental to our purpose of matching claims data with mileage. Matching a VIN's actual claims amount with a "filled in" mileage estimate not specific to that VIN would dilute the significance of mileage in predicting risk and therefore distort our statistical model. For this reason we decided to compute our own mileage estimates based on the original odometer reading data.

To compute mileage estimates, we sorted through the original RMV odometer readings—about 15 million of them—and tossed out those which appeared to be in error, either because the odometer reading itself was implausible or because the odometer registered fewer miles than in a previous inspection[57]. On the whole, the RMV inspection data appear plausible and well-behaved. Vehicles are required to be inspected within seven days of changing ownership, including sales of new vehicles, so a huge number of readings showed fewer than 1,000 miles. Then there are relatively few readings up until the 10,000 mile mark (presumably since new vehicles, having been inspected immediately after purchase, do not need to be inspected for another year) after which point the number of inspections gradually tapers off, getting close to zero around the 250,000 mile mark. Figure 17 shows this trend, though the huge spike for readings between 0 and 1000 miles is not shown because it is about 10 times higher than the peak around 10,000 miles.

Figure 17. Histogram of odometer readings at different mileage levels.



Next we constructed every possible pairing of two odometer readings on the same VIN and registration number[58] which could be used to generate a mileage estimate, with the constraint that the time between readings be between six months and two years.

Finally we had to choose the *best* estimate of mileage for each vehicle (more precisely, for each unique identifier of VIN, policy ID and policy effective date in the insurance coverage data). Originally we hoped to develop two estimates for each vehicle. One would be "concurrent" or "retrospective," meaning that it would represent a best estimate of miles driven *during* the time the vehicle was insured under a given

---

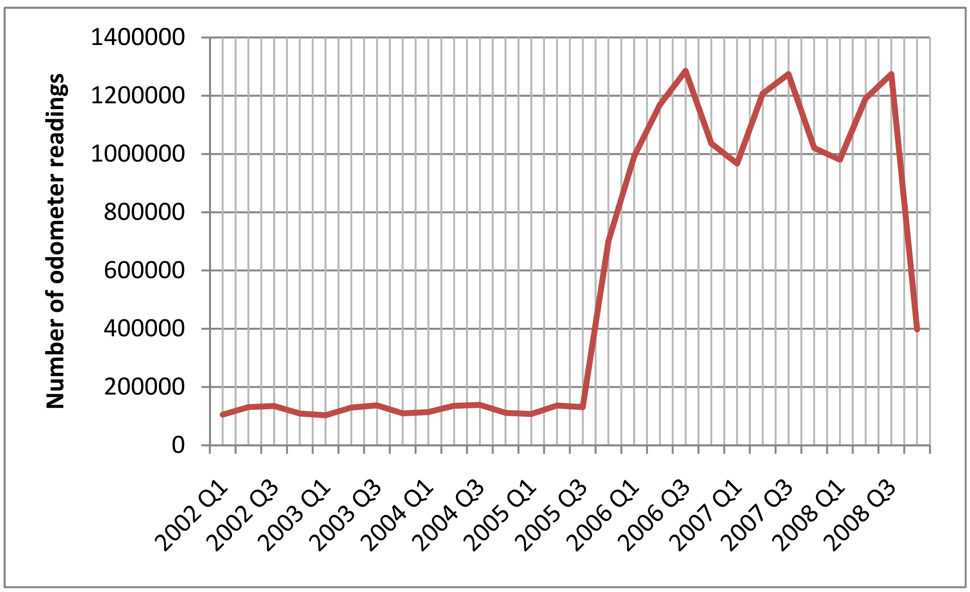
[57] In cases where there were more than two inspections for one VIN under the same registration number, we examined sets of three to determine which inspection(s) was (were) in error and, where possible, salvage an estimate out of the remaining two.

[58] Requiring that the registration number be the same means that the vehicle had the same owner during the two inspections, and therefore that the estimate we compute reflects that one driver's mileage. We later pair our mileage estimates with insurance policy data. When a mileage estimation period overlaps with the insurance policy effective period, this guarantees that we have estimated the mileage of the actual insured drivers. For the 8% of cases where there is no overlap, there is probably some small subset where the vehicle changed hands and our mileage estimate refers to the driving behavior of a different owner.

policy. This would represent the mileage that would be charged to the driver in a scenario where insurance were billed after the fact, on a cents-per-mile basis. Another estimate would be "prospective," meaning both inspections occurred before the policy was issued. This would represent the best estimate that the insurance company would have available *before* issuing the policy in the event that insurance were not billed on a per-mile basis but that monthly rates would be set in accordance with predicted mileage.

In practice, we were unable to develop "prospective" estimates. We are examining the 2006 policy year, meaning insurance policies that were issued any time in 2006 and lasted (for the most part) until one year later, sometime in 2007. A prospective estimate would require two odometer readings, at least six months apart, with the second one occurring in 2005 or early 2006, depending upon when the policy was issued. In fact, most RMV inspections for which we have data occurred in 2006 or later; only a handful of odometer readings are available before that time, as shown in Figure 18. As a test case, we tried creating prospective estimates for VINs in Cambridge, of which there are about 37,000 with odometer readings. Of these 37,000, only 187 had two valid odometer readings before the 2006 insurance policy was issued.

Figure 18. Few odometer readings from RMV inspections are available before 2005 Q4.

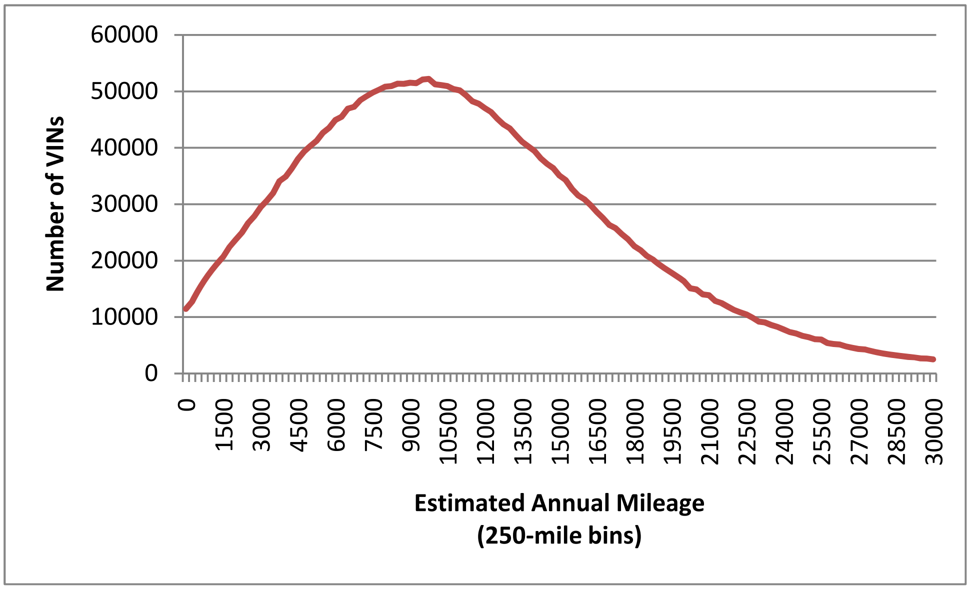


Instead, we decided to focus on best "concurrent" estimates—estimates of mileage driven during the policy period. We chose the "best" estimates based on two criteria. First, we chose the estimate which came from an inter-inspection period with the maximum number of days of overlap with the policy period (the time when the vehicle was insured). In the case of ties (usually when there was no overlap because the inspections both occurred after the policy ended, or when the inter-inspection period overlapped the entire policy period), we chose the inter-inspection period with a midpoint closest to the midpoint of the policy period.

Mileage estimates were available for 90% (3.25M out of 3.61M) of all VIN-policy id-policy effective date unique identifiers for which insurance data were deemed valid. Most of the best estimates ultimately chosen overlap the beginning (22%), end (57%) or entire (13%) policy period; less than 1% fall entirely within the policy period. Only 8% came from two inspections after the policy ended, and less than 1% each came from two inspections before the policy began.

The estimates we have computed form a (log) normal curve, as shown in Figure 19—the huge spikes visible in Figure 16 due to "filled in" values are gone.

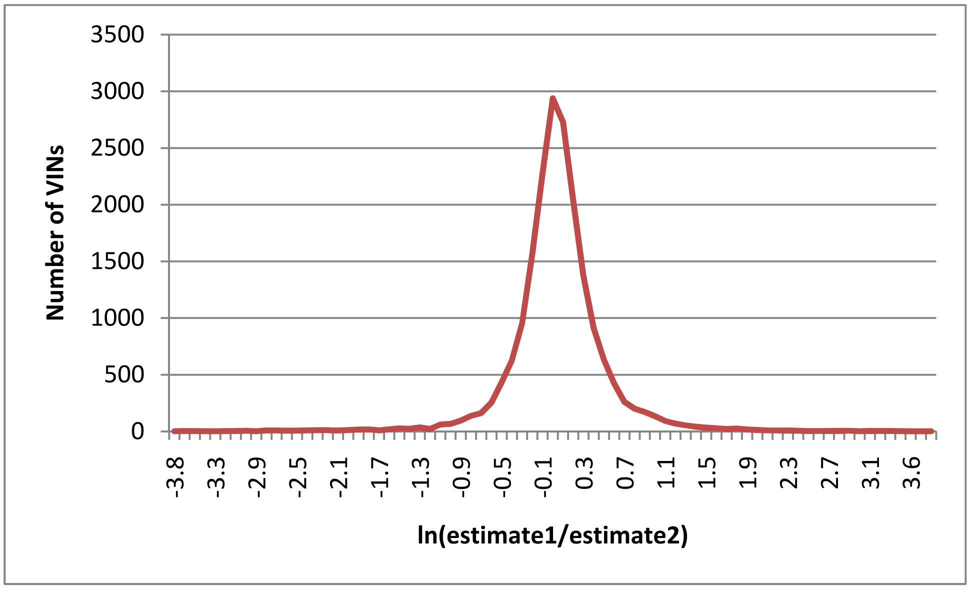
Figure 19. Histogram of annual mileage estimates we obtain for 3.25M policy-vehicle combinations in Massachusetts



The best estimates chosen are also reasonably consistent with the MassGIS estimates. Unfortunately, the publicly released EOEEA dataset does not include any indicator of which estimates were "filled in" and which were genuine, but MIT had access to MassGIS estimates of VMT for many VINs in this study where it was possible to determine which estimates are genuine and which are city averages. Of MassGIS' genuine estimates for the same VIN and with the same inspection period end date as used for one of our own estimates, 88% are within 2 miles of our estimates (many differ by one or two miles, presumably due to rounding and accounting for leap year). The remainder differ by reasonably small amounts (96% of the estimates are within 1000 miles of each other), so presumably the difference arises from a different decision of which start date to pair with a given end date, and so both estimates reflect reasonable estimates of the average annual mileage of the car in question.

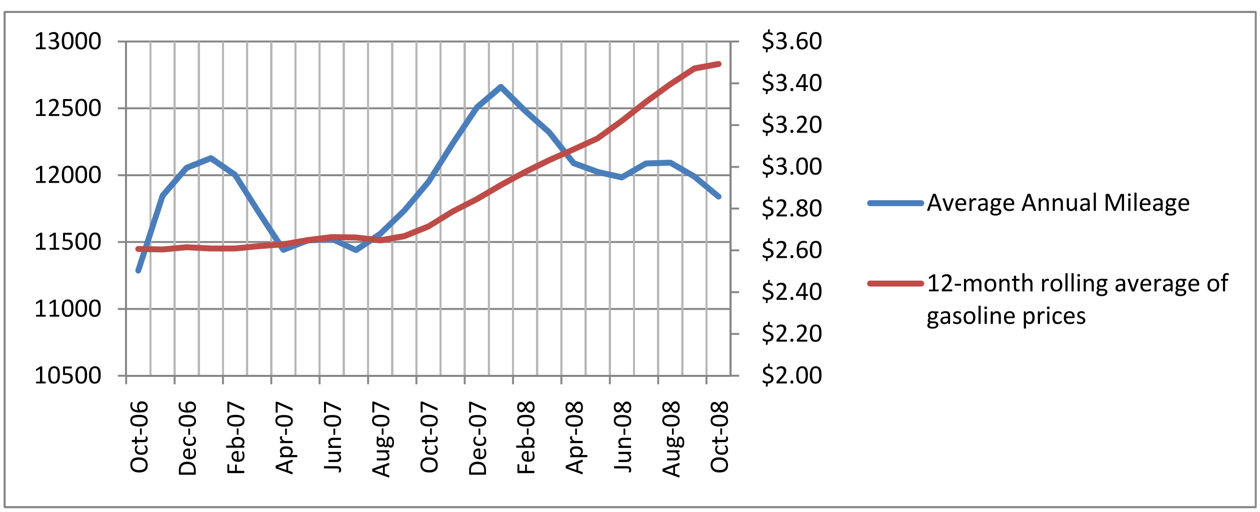
For vehicles with three inspections on different dates (call them a, b, and c), we also check how consistent is the annual mileage observed between dates a and b and between dates b and c. Wild variation of mileage estimates depending on which date pair is chosen would be worrisome, as it would imply that for vehicles where the inter-inspection period overlaps little with the policy period, our estimate would be a very poor predictor of mileage during that period. But in fact, the mileage estimates prove well-behaved. The histogram in Figure 20, computed only for VINs in Cambridge, shows that most vehicles get similar estimates no matter which pair of dates is used.

Figure 20. Histogram of log ratio of first estimate to second estimate for VINs in Cambridge with exactly three inspection dates



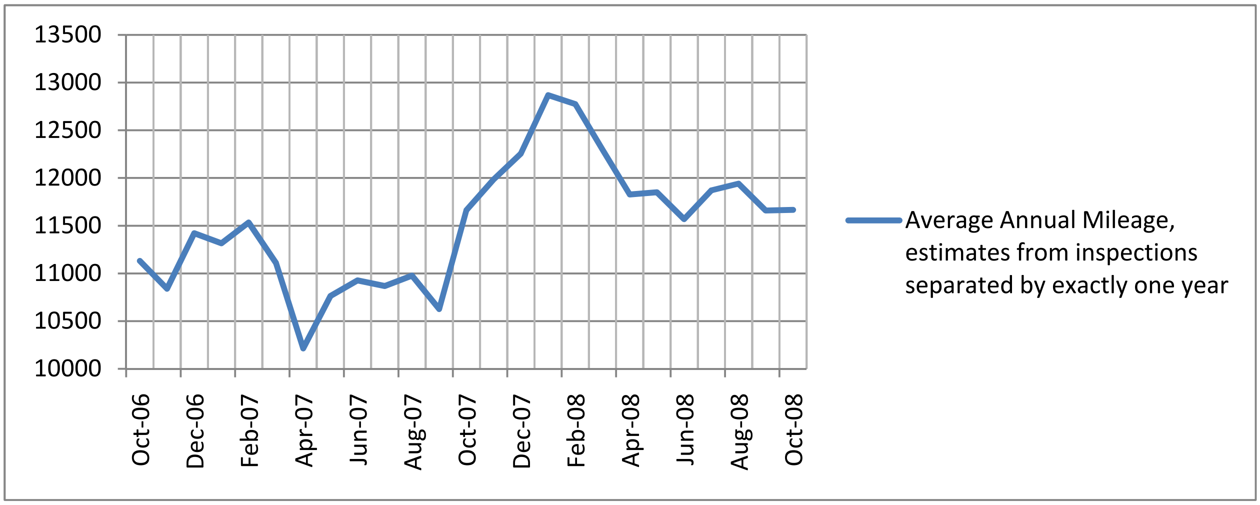
On average, the annual mileage implied by the first period is 1.05 times that of the second period. It is tempting to suppose that this is because more of the later inter-inspection periods covered mid-2008 when fuel prices rose to record levels, but further investigation reveals a more complicated relationship. Such an investigation requires first a comparable measure of gasoline prices, and in this case we choose a 12-month rolling average to match the roughly year-long inter-inspection periods—the mileage estimates that were considered to be most reliable come from periods lasting an average of 416 days. Figure 21 graphs estimated annual mileage by end date against this 12-month rolling average.

Figure 21. Best estimates of annual mileage per vehicle by end month versus 12-month rolling average of gasoline prices, October 2006 through October 2008. All vehicles in Massachusetts (3.05M) for which a mileage estimate is available.



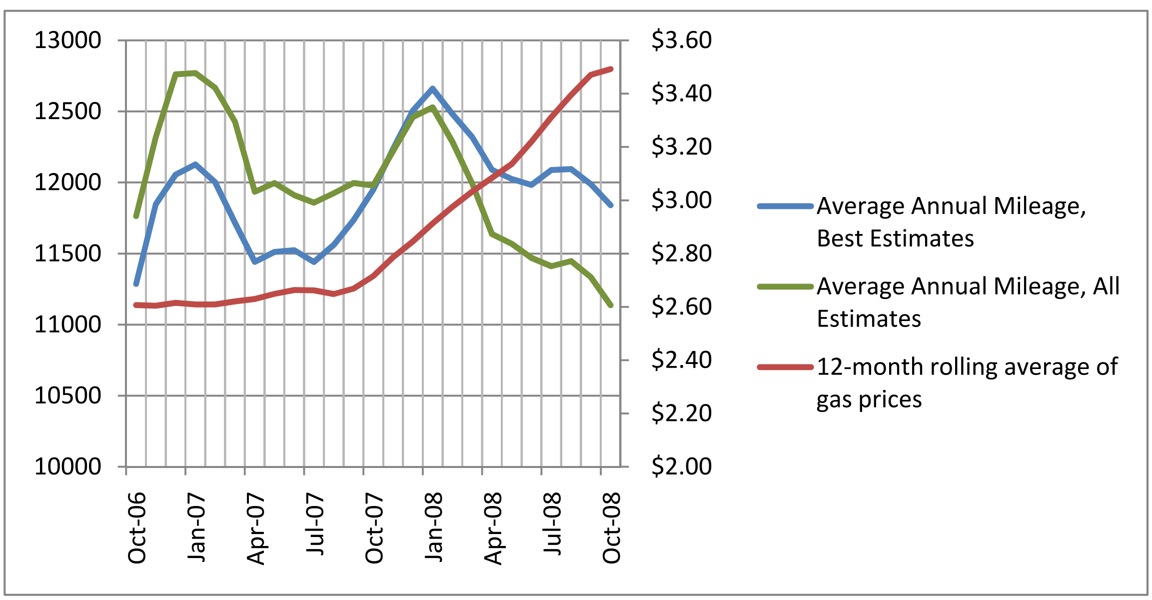
In fact, the average annual mileage seems to be obeying seasonal trends more than any cue from gas prices, and moreover, the average mileage is a bit higher in 2008 when gas prices are high. Yet it is unclear why there should be seasonality in an average annual mileage computed from inspections separated by about a year. Since the average inter-inspection period used in the 'best' estimates is actually 416 days, one explanation is that there is seasonality in the month and a half that are double-represented. However, choosing only those 'best' estimates (80,000 of them) that came from inspections separated by exactly 365 days yields a very similar pattern of seasonality, as shown in Figure 22.

Figure 22. Best estimates of annual mileage per vehicle by end month, for only the 80,000 estimates which came from inspections exactly 365 days apart. Seasonal patterns appear the same as for all best estimates.



Interestingly, if *all* available mileage estimates—not just the *best* estimate for each vehicle—are considered, then the seasonal bumps remain but the overall trend in mileage is downward rather than upward, as shown by the green line in Figure 23. The blue line representing best estimates from Figure 21 is left in for comparison.

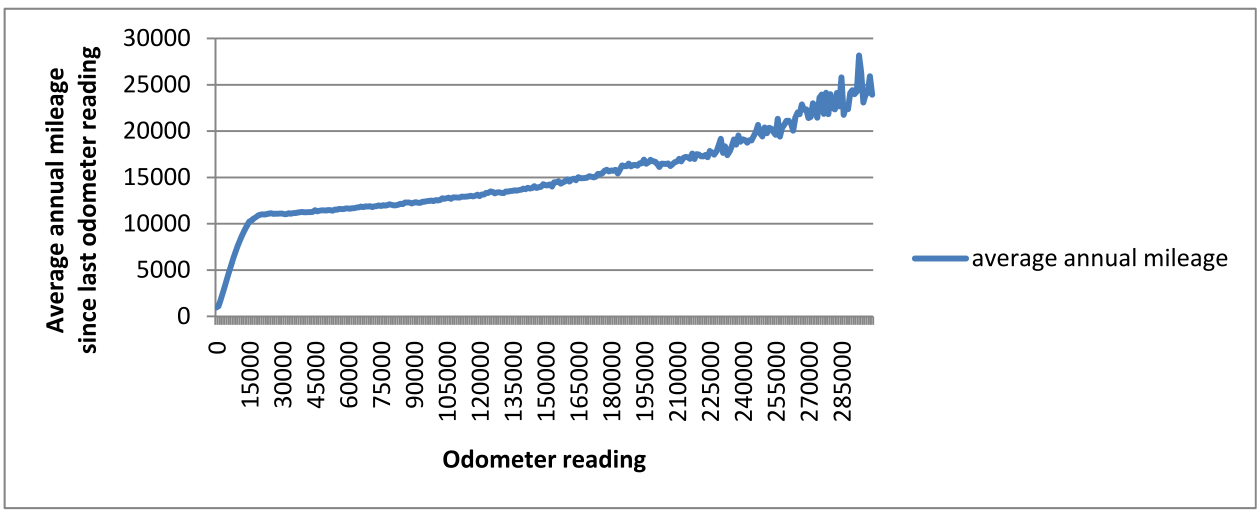
Figure 23. Average annual mileage from best estimates and from all estimates by end date, versus 12-month rolling average of gasoline prices.



Including *all* estimates changes the picture in two ways: first, it does not select for the estimate that overlaps best with a policy period falling sometime in 2006-2007, and second, it does not eliminate those vehicles which did not have an insurance policy during policy year 2006. This last point could be especially important. Households that bought an additional car in 2007 or 2008 despite rising gas prices would then spread their mileage out over a greater number of cars. This relatively low-mileage new (or used) car would be represented in the all estimates curve but not in the best estimates curve. On the other hand, households that sold or junked a car in that time period and did not replace it would spread their mileage out over fewer cars, so that even if household mileage decreased as gas prices rose, mileage per vehicle might rise.

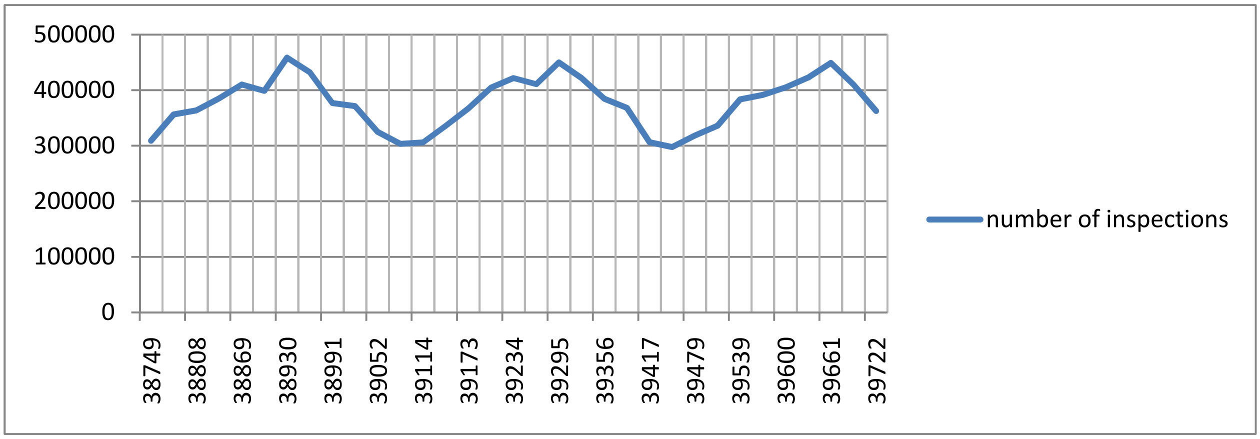
Odometer readings, it turns out, are also seasonal, and this is part of the answer to this puzzle. However, first it is important to establish that new cars are driven fewer miles annually than older cars with more miles on them. This may seem counterintuitive at first: people are excited about their new cars and drive them most everywhere they go, using their old car, if they keep it, as a backup. However, while an individual household may drive its new car for commuting and use its old car for low-mileage errands, that household is still spreading its total mileage over the two cars, whereas a similar household with just one car will load all its mileage onto that one. The RMV inspection data bear this out, as shown in Figure 24. Cars near the end of their life, with 200,000 miles on them, are driven about 17,000 miles per year on average, while young cars with, say, 30,000 miles, are driven just 11,000 miles per year. The sharp falloff at the far left of the curve is because only cars that are driven relatively few miles per year will actually get two inspections before, say, 12,000 miles, making a mileage estimate possible. The jagged behavior on the right is because only a few hundred vehicles are in each odometer bin, since few vehicles survive to their 300,000th mile.

Figure 24. Average annual mileage by odometer bin. Older cars are driven more miles each year than newer cars.



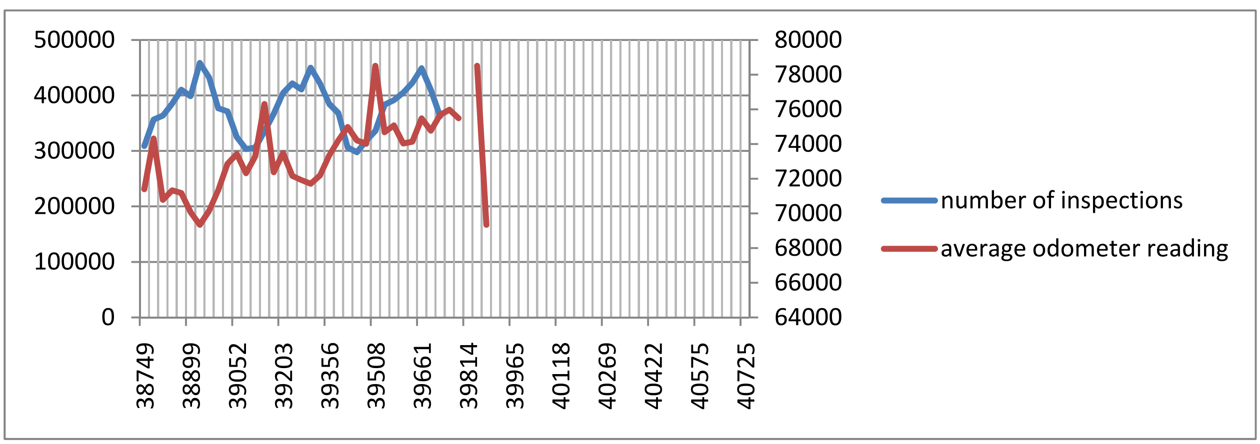
With the fact established that older cars are driven more miles each year than new cars, seasonality of odometer readings can provide insight into seasonality in annual mileage estimates. Some seasonality of odometer inspections is visible in Figure 18 above, which shows the number of readings by quarter, but when readings are broken down by month as in Figure 25, the seasonality becomes even starker: the peak in August is 50% higher than the trough in January. Far more safety inspections are conducted in the summer than in the winter.

Figure 25. Seasonality of RMV safety inspection odometer readings. About 450,000 checks are completed each August, compared to about 300,000 each January.



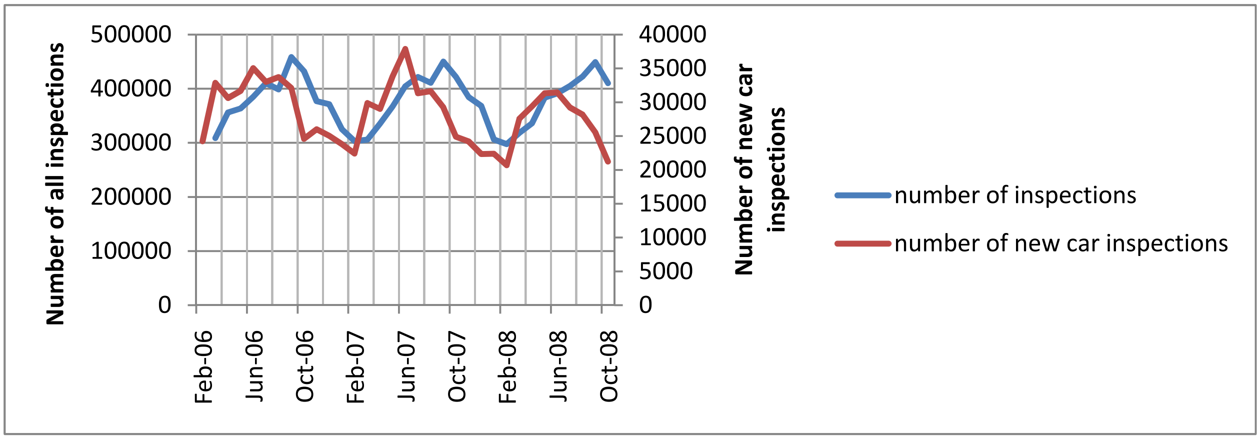
The juxtaposition of average odometer reading with the number of inspections as shown in Figure 26 suggests that a different profile of car is being inspected in the summer than in the winter. Though the difference is just 13% between the highest peak and lowest trough, it appears that on average, older or more heavily used cars are being inspected in winter and newer or more lightly used cars are being inspected in summer. Since brand new cars must be inspected within seven days of purchase, it could simply be that these new cars with odometer readings close to zero are being purchased predominantly in the summer.

Figure 26. Number of RMV safety inspections per month versus average odometer reading, February 2006 through October 2008.



To test this hypothesis, Figure 27 graphs the total number of inspections versus the number of those with fewer than 1000 miles on the odometer. The rough alignment of the peaks and valleys suggests that new car purchases may be part of the explanation. It cannot be the whole story, though, since the peaks don't line up exactly. However, a new car purchased in July 2006 will require inspection again after one year, and might actually get that second inspection late, in August 2007 rather than July 2007. Since households that buy new cars spread out their mileage over more cars, these new cars have lower annual mileage than the older cars belonging to households with fewer cars. So a new car purchase one summer could be responsible for repeated, low annual mileage inspections for several subsequent summers. A surge of new car purchases in the summer may ultimately be responsible for some of the seasonality observed both in the number of inspections and in the annual mileage computed from inspections.

Figure 27. Total number of inspections and inspections of new cars (with odometer reading below 1000 miles), February 2006 through October 2008.



The inspections of new cars as shown in Figure 27 appear to represent about 10% of all inspections, and the number of new car inspections nearly doubles in summer compared to winter. A new car, once purchased, will be due for inspections every summer for several years, only resetting its season of inspection when it is sold used to another owner who must then get it re-inspected within seven days. As a back-of-the-envelope calculation, if new car sales are double in summer compared to winter and a new car stays with its first owner and continues to be inspected in summer for half its life, after which time it switches to a random inspection date depending on the date of the used car sale, then there ought to be 50% more inspections in summer months than winter months. Further, if cars purchased new within the last five years account for 50% of cars inspected in a summer month and if these are driven 20% fewer miles per year than the average car, then the average annual mileage observed in summer months could be 10% less than in winter. So the seasonality of new car sales could actually explain much of the seasonality in computed annual mileage.

To the extent that this single phenomenon explains it, the seasonality in computed annual mileage need not be cause for concern about bias in the method of estimation we have employed. If an underlying seasonality in new car sales and inspection patterns is responsible, then annual mileage averaged over all vehicles can exhibit seasonality without suggesting that the mileage estimate for any one vehicle is biased. Therefore, modeling insurance claims for each vehicle as a function of mileage for each vehicle is not problematic.

Appendix III: Data Dictionary for PAYD Analytic Dataset

This appendix describes the PAYD analytic dataset used in this report. A downloadable zip file with the Appendix 3 Analytic Dataset is available at MIT's website: http://mit.edu/jf/www/payd.

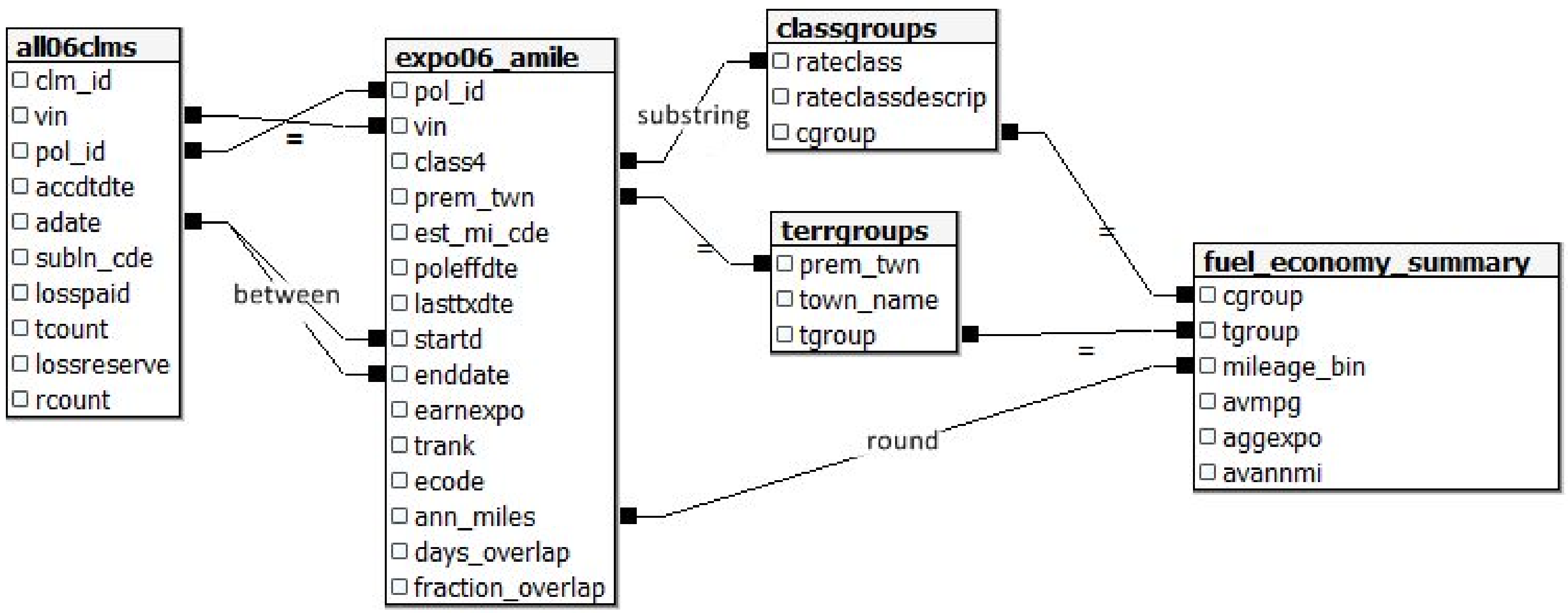
The dataset was developed from the DVD of raw data released in March, 2010, by the Massachusetts Executive Office of Energy and Environmental Affairs (EOEEA). That DVD contained 1.3 GB of compressed data from the state's Commonwealth Automobile Reinsurer (CAR) and the Registry of Motor Vehicles (RMV). The CAR data included several years of auto insurance policy and claims transaction records for all insured private passenger vehicles in the state. The RMV data included odometer readings from all state-mandated annual safety inspections of all private passenger vehicles. Public notice of the availability of these data is posted on the EOEEA website at: http://www.mass.gov/?pageID=eoeeaterminal&L=3&L0=Home&L1=Grants+%26+Technical+Assistance&L2=Data+Resources&sid=Eoeea&b=terminalcontent&f=eea_data-resources_2005-08-auto-insur-data&csid=Eoeea

Considerable processing was required to convert the raw policy transaction data from CAR into earned exposure records that matched each vehicle to an appropriate rating class, territory and mileage estimate during each insured month. Additional processing was also required to convert the raw claims transaction data into net paid losses and reserves (as of December 2008) and then match each claim (for bodily injury and property damage liability or personal injury projection) to the appropriate vehicle and earned exposure month. This processing of the raw CAR data was done for all policy and claim transactions involving vehicles insured during policy year 2006. The processing was done by Prof. Joseph Ferreira at the Massachusetts Institute of Technology (MIT) through the partial support of University Transportation Center Region One research grant MITR22-5.

The PAYD analytic dataset also includes annual mileage estimates for each policy year 2006 vehicle and fuel economy estimates for each class, territory, and mileage category. The DVD from EOEEA also contained annual mileage estimates developed by MassGIS from the safety inspection records. However, we redid the mileage estimates (using the same safety inspection records) in order to base each vehicle's estimate on the pair of inspections that had the largest overlap with that vehicle's 2006 policy year. The fuel economy estimates for each class, territory, and mileage category are averages of the adjusted EPA estimates (reported by VINquery.com) for each individual vehicle (out of the 3 million in our study) that fell into each class, territory, and mileage category. The mileage (re)estimation and fuel economy processing was done by the authors as part of this PAYD study.

The tables included in the PAYD analytic dataset are explained in the following entity-relationship diagram and table definitions:

Entity-Relationship Diagram for PAYD Analytic Dataset



all06clms

Summarizes claims data, including total paid losses and outstanding reserves, for all claims that matched an earned exposure period during the 2006 policy year.

681,423 records.

Primary key: vin||pol_id||clm_id||adate||subln_cde

clm_id	integer	Claim ID, anonymized by MassGIS
vin	varchar(17)	Vehicle Identification Number (VIN)
pol_id	varchar(14)	Insurance policy ID, anonymized by MassGIS
accdtdte	varchar(8)	Accident date in string format (YYYYMMDD)
adate	date	Accident date in date format
subln_cde	character(1)	Subline code: 1 = Liability (property damage and bodily injury) 5 = No fault (personal injury protection)
losspaid	bigint	Net total losses (and loss adjustment expenses) paid through 2008 Q4 for each unique claim ID and subline code. 2433 out of 625632 paid claim records are capped at $25000. 71,829 records in this table have net losspaid+lossreserve <= $50 (e.g., after subrogation) and so were excluded from frequency and pure premium calculations.
tcount	integer	Number of payment transactions related to this claim
lossreserve	bigint	Outstanding reserves as of 2008 Q4 for each unique claim ID and subline code. 2568 of 36635 records with outstanding loss reserves are capped at 25000. 71,829 records in this table have net losspaid+lossreserve <= $50 (e.g., after subrogation) and so were excluded from frequency and pure premium calculations.
rcount	integer	Number of transactions updating the reserve amount for this claim

expo06_amile

Earned exposure months and annual mileage estimates for each period of consistent policy endorsement conditions during policy year 2006 for 3.25M policy-vehicle combinations.

3,991,012 records.

Primary key: vin||pol_id||poleffdte||startd

pol_id	varchar(14)	Insurance policy ID, anonymized by MassGIS
vin	varchar(17)	Vehicle Identification Number (VIN)
class4	character(4)	Insurance rating class as defined in Code Descriptions.pdf
prem_twn	character(3)	3-digit code indicating the garage 'town' for insurance purposes. Includes 360 towns and neighborhoods in Massachusetts plus codes for out-of-state policies. Described in PremAccdtTownTables.pdf.
est_mi_cde	character(3)	Code for driver-reported annual mileage category described in Code Descriptions.pdf. This data comes from Commonwealth Automobile Reinsurers, not from our own mileage estimates.
poleffdte	varchar(8)	Year and month (YYYYMM) of policy effective date.
lasttxdte	varchar(8)	Year and month (YYYYMM) of last transaction modifying policy.
startd	date	Estimated start date for the period during which the policy endorsement conditions reported in this record are earned. (This is different from the policy effective date if an endorsement occurred).
enddate	date	Estimated end date of the period during which the policy endorsement conditions reported in this record are earned. (This is different from the policy end date if an endorsement occurred).
earnexpo	numeric(3,0)	The number of months of 'earned exposure' for the policy conditions reported in this data record.

trank	integer	Number of transactions modifying policy during the policy year.
ecode	integer	Indicates what type of adjustment to start and end dates was necessary in order to reconcile earned exposure months with transaction dates and policy effective date. 0: No endorsements, no adjustment needed 10: Endorsements made but no adjustment needed 22: End date moved earlier by 17 days 24: End date moved later by 14 days
ann_miles	numeric	Estimated annual miles traveled by the vehicle. New estimates developed from odometer readings in RMV safety inspection data from all_rmv_insp_06.txt, etc. 0 to 100,000 -1 = no mileage estimate available
days_overlap	integer	Number of days of overlap of the policy effective period and the two odometer inspections used to create the annual mileage estimate. A measure of the quality of the estimate. 0 to 366 -1: no mileage estimate available
fraction_overlap	numeric	The fraction of the policy effective period that is overlapped by the period between two odometer inspections used to create the annual mileage estimate. A measure of the quality of the estimate. 0.00 = No overlap ... 1.00 = Full overlap -1 = no mileage estimate available

terrgroups

Lookup table matching insurance 'towns' to six territory groupings.

360 records. Primary key: prem_twn

prem_twn	character(3)	3-digit code indicating the garage 'town' for insurance purposes. Includes 360 towns and neighborhoods in Massachusetts. Described in PremAccdtTownTables.pdf.
town_name	varchar(30)	Name of town or neighborhood.
Tgroup	character(1)	Six territory groupings: '1' = Least risky '2' '3' '4' '5' '6' = Most risky

classgroups

Lookup table matching the last character of the four-digit rating class code (class4) to the five class groupings.

9 records. Primary key: rateclass

Rateclass	character(1)	Indicates the rating class of the insurance policy. Corresponds to the fourth digit of class4 in expo06_amile. '1' through '9'. Detailed descriptions stored in rateclassdescrip.
rateclassdescrip	varchar(42)	Description of the rate class
Cgroup	character(1)	Five class groupings: 'A': Adults 'B': Business 'I': <3 yrs experience 'M': 3-6 yrs experience 'S': Senior citizens

fuel_economy_summary

Summarizes vehicle fuel economy data, aggregating vehicles into groups by class, territory and annual mileage bin.

5,170 records.

Primary key: cgroup||tgroup||mileage_bin

cgroup	character(1)	Five class groupings: 'A': Adults 'B': Business 'I': <3 yrs experience 'M': 3-6 yrs experience 'S': Senior citizens
tgroup	character(1)	Six territory groupings: '1' = Least risky '2' '3' '4' '5' '6' = Most risky
mileage_bin	integer	Vehicle's annual mileage by 500-mile bin
avmpg	numeric	Average fuel economy of vehicles in this group. Each vehicle's fuel economy is considered to be 88% of the simple average of its city and highway fuel economy. Average fuel economy for all vehicles in the group is then weighted by aggregate miles traveled (i.e. annual mileage times months of exposure)
aggexpo	numeric	Aggregate car years of exposure for vehicles in this group.
avannmi	numeric	Average annual miles traveled for vehicles in this group.