# THE USE OF ANNUAL MILEAGE AS A RATING VARIABLE

BY

JEAN LEMAIRE, SOJUNG CAROL PARK AND KILI C. WANG

## ABSTRACT

Auto insurance companies must adapt to ever-evolving regulations and technological progress. Several variables commonly used to predict accidents rates, such as gender and territory, are being questioned by regulators. Insurers are pressured to find new variables that predict accidents more accurately and are socially acceptable. Annual mileage seems an ideal candidate. The recent development in new technologies should induce insurance carriers to explore ways to introduce mileage-based insurance premiums. We use the unique database of a major insurer in Taiwan to investigate whether annual mileage should be introduced as a rating variable in auto third-party liability insurance. We find that annual mileage is an extremely powerful predictor of the number of claims at-fault. The inclusion of mileage as a new variable should, however, not take place at the expense of bonus-malus systems; rather, the information contained in the bonus-malus premium level complements the value of annual mileage. An accurate rating system should therefore include annual mileage and bonus-malus as the two main building blocks, possibly supplemented by the use of other variables like age, territory and engine cubic capacity. While Taiwan has specific characteristics (high traffic density, a mild bonus-malus system and limited compulsory auto coverage), our results are so strong that we can confidently conjecture that they extend to all developed nations.

## KEYWORDS

Auto liability insurance, Rating variables, Annual mileage

## 1. INTRODUCTION

Auto insurers, in order to remain competitive in risk selection and pricing, are constantly seeking better ways to measure risk. To this end, they adopt numerous rating variables — and, when unavailable, proxy variables — to better gauge how risky each particular customer is.

*Astin Bulletin* 46(1), 39-69. doi: 10.1017/asb.2015.25 

Auto insurers typically use a large number of variables in their ratings, including age, sex, marital status of principal driver, make, model, use of car, territory, moving violations, etc. Other factors that may improve risk classification are not used due to regulatory restrictions or practical reasons; a factor may be too costly to credibly observe or socially unacceptable. Consequently, in most developed countries, insurers have implemented bonus-malus systems (BMS), which modify the premium according to past claims history. One of the main goals of BMS is to reduce adverse selection by including indirectly information that could not be taken into account explicitly such as respect of the driving code, alcohol use, mileage driven, etc.

One of the potential classification variables that has not been widely used so far is annual mileage. It is intuitively clear that those who drive more will have more auto accidents, that each extra mile spent on the road creates a small additional chance of an accident. However, insurers have been reluctant to use annual mileage due to their inability to verify policyholders' statements and the relative easiness to tamper with odometers. This had led them to use proxy variables like the use of the car (e.g. personal, commuting or business) or the distance between home and work. Butler (2006) argues that no less than 12 widely used rating variables can be considered as proxies for odometer miles: gender, car age, previous accidents at-fault and not-at-fault, credit score, postal code, income, military rank, existence of a prior insurer, premium payment by installments, years with same employer, collision deductible and tort rights.

This reliance on proxy variables may change with the development of new technologies like telematics, on-board computers, sophisticated GPS transmitters, tampering-resistant odometers and their fast decrease in cost. Thanks to these advances, many auto insurers throughout the world have started to adopt annual mileage among their rating variables. As data recorded from GPS become available to actuarial researchers, opportunities to study previously unavailable variables will arise. Pioneering research using new variables include Ayuso *et al.* (2014) and Paefgen *et al.* (2013, 2014). Ayuso *et al.* (2014) analyze the driving patterns of 15,940 Spanish drivers under the age of 30 years; besides the daily distance travelled, they were able to record the percentage of total kilometers driven in urban areas, at night, or exceeding speed limits. They showed that the time until first crash is reduced by night driving, by speeding, and for inexperienced drivers, among other results. For 1,567 vehicles, Paefgen *et al.* (2013, 2014) studied the risk of an accident as a function of new variables like the time of day, the day of the week, and speed intervals, and discovered a non-linear relationship between annual mileage and claim frequencies.

While there is ample evidence that annual mileage positively correlates with claim rates (Ferreira and Minikel (2010), Jovanis and Chang (1986), Lemaire (1985), Litman (2011), Lourens *et al.* (1999), Progressive Insurance (2005), among others), there is a dearth of research in the actuarial literature that compares the accuracy of mileage as a rating variable with traditional pricing factors. A notable exception seems to be Ferreira and Minikel (2013), who study over three million individual car-years observed in 2006 in Massachusetts. Poisson and linear regression models are used to explain the pure premium as a function of annual mileage and two traditional rating factors: territory (six zones) and class (adults, senior citizens, business use, years of driving experience). The main conclusions are that, while mileage is a significant predictor of accident risk, it is inferior to the other rating factors if used alone; mileage can substantially improve rating accuracy if used in conjunction with other variables.

In this research, we investigate whether annual mileage is a potential rating variable using a unique database originating from Taiwan. We were able to merge the annual mileage recorded during routine maintenance and oil changes in a large network of specialized shops with auto insurance related data collected from the largest insurer operating in Taiwan. Our research extends the existing literature in several significant ways: (a) We use a large database, comprising over a quarter million policy-years; (b) We study claim severity in addition to claim frequency; (c) We include a large set of traditional classification variables as controls: gender, age, marital status of policyholder, vehicle age, type, use, engine cubic capacity, territory, urban/rural driving. We also use the BMS level of each policyholder, a variable that several studies (Lemaire (1985), among others) consider to be the best predictor of future accidents. One important rationale for the good accident predictability of BMS levels has been mileage: indeed BMS coefficients may partially reflect unobserved mileage driven. In addition, we use negative binomial regression models to evaluate the relationship between claim frequency and mileage, and linear regression to examine the relationship between claim severity and mileage.

By providing empirical evidence of a strong relationship between annual mileage and claim counts and a positive relationship between mileage and claim severity, this paper provides ample justification for the use of mileage as a rating variable. The remainder of the article proceeds as follows. Section 2 discusses criteria for auto insurance rating variables and evaluates annual mileage in light of these requirements. The data are presented in section 3. Section 4 presents the main results of regressions performed on the claim count and severity distributions. The robustness of results is discussed in section 5. Section 6 concludes.

## 2. Rating Criteria and Annual Mileage

### 2.1. Rating criteria for "fair" discrimination

Auto insurers openly practice discrimination in underwriting and pricing. Competition among insurers and adverse selection among policyholders trigger "fishing for good risks", the use of a large number of classification variables shown to affect claim frequency and severity. As long as regulators allow them, insurers are using variables like age, gender, marital status, territory (postal code), years licensed, credit score and occupation of the main driver; good-student discounts; driver training; participation in a traffic safety program; restricted usage; type, model, engine cubic capacity, horse power, age, use of the car; annual mileage; garage ownership; premium payment frequency; as well as past claims and moving violations. This process of segmentation, the subdivision of a portfolio of drivers into a large number of homogeneous rating cells, only ends when the cost of including more risk factors exceeds the profit that the additional classification would create, or when regulators rule out new variables.

Insurers have a preference for total freedom in selecting risk factors, so that they can charge appropriate premiums to all groups based on risk differentials. They claim that risk classification creates incentives for insureds to minimize risks. Accurate risk classification and incentives for risk reduction provide the main reasons why society lets insurers discriminate. Indeed, research consistently suggests that restrictions on risk classification result in cross-subsidizations: low-risk individuals choosing to reduce their coverage and more high-risk drivers on the road. As price subsidies weaken the link between risk and premiums, consumers' incentives for loss prevention are diminished. Insurance companies lose incentives to control costs and tend to send more applicants to assigned risk pools. As a result, whenever regulation prohibits or reduces the role of a rating variable, the resulting marginal premium decrease for high-risk drivers does not compensate the increase for low risks, and overall premiums tend to increase (Blackmon and Zeckhauser, 1991; Schwarze and Wein, 2005; Brown *et al.*, 2007; Regan *et al.*, 2008; Weiss *et al.*, 2010; Derrig and Tennyson, 2011; Sass and Siegfried, 2012).

Despite this evidence, society, represented by legislators and insurance regulators, has limited the types of discrimination insurers are allowed to practice. Indeed, in recent years, certain classifiers, including race, gender, age and territory, have been severely restricted or outright prohibited. For example, despite massive and undisputed proof that females cause fewer accidents, the Court of Justice of the European Union ruled that all insurance contracts entered on or after December 21, 2012, cannot price males and females differently. The use of gender is also prohibited in ten U.S. states, and limited in 22 others (Avraham *et al.*, 2013). Age is not used in six Canadian provinces and nine U.S. states, with strict restrictions in eleven other states. Two U.S. states ban the use of postal code in all property/casualty contracts. Other states limit the number of territorial rating cells that can be used, restrict premium ratios across two contiguous territories or between the highest-rated and lowest-rated districts, or force territory to be a secondary rating factor (Avraham *et al.* (2013), Brown *et al.* (2007), Derrig and Tennyson (2011), Harrington (1991), Jaffe and Russell (2001)).

"Discrimination" can be viewed in a positive or negative way. It may mean nothing more than recognizing a difference between groups, the cornerstone of insurance pricing, or it can construed as a prejudice, asserting that certain groups are morally inferior and undeserving of equal treatment. Everyone will agree that insurers should be permitted to deny coverage or charge a higher premium to drivers who have been convicted for drunk driving. Few would disagree that the use of race in rating should be disallowed, and even viewed as repugnant, despite significant differences in accidents costs, as race is a non-causal factor, not under the control of the insured, and historically linked to unfair treatment. What distinguishes "fair" discrimination from "unfair"? When should discrimination be deemed illegal? Which tests should be used to determine if a rating variable is socially acceptable?

The American Academy of Actuaries (1980), Avraham *et al*. (2013), Gaulding (1995), and mostly Kelly and Nielson (2006), have presented a variety of tests that ideal risk predictors should pass. Requirements can be subdivided into actuarial, operational, social and legal criteria.

2.1.1. *Actuarial criteria.* A classification variable is considered to be actuarially fair if it is accurate (the most important criterion, requiring a strong relationship between variable and claims), credible (sufficient data exist for all rating cells), reliable over time and shows homogeneity within cells.

The variables that have been questioned (race, age, gender, territory) easily pass the accuracy and reliability tests. There are some credibility issues for age and territory, as few data are available for very old drivers and some territories are sparsely populated. Age is subject to much criticism on the homogeneity issue: young and elderly drivers show greater heterogeneity of skills, driving abilities and accident rates.

2.1.2. *Operational criteria.* For each insured, the value of the variable must be objective (different underwriters will always classify in the same way), assessed at little cost, and not easy to manipulate. There must be an intuitive relationship with claim rates. Discontinuities between groups should be minimized.

Race, age and gender are objective, easily measured at no cost, and cannot be manipulated. The relationship with claim rates is not easily demonstrated: it is not evident that driving ability is a clear-cut function of age and gender. Age fails the continuity test, as there is often a big drop in premiums at the age of 25 years for females, and 30 years for males. Manipulation of territory constitutes one of the main causes of premium fraud, as it is not uncommon for car owners to register their car in a rural area when they actually live in a city; such deception is costly to detect, requiring insurers to patrol downtown areas during consecutive nights to identify out-of-town registrations.

As an example of a variable failing the cost criterion, an in-depth psychological profile could reliably predict accident risk, but the underwriting cost would be prohibitive.

2.1.3. *Social criteria.* Social acceptability is an important test to implement a rating variable, with the main requirements being privacy, controllability, affordability/availability and causality. Risk classification is easier to accept by the public if there is an intuitive and demonstrable cause-and-effect of the variable on claim rates, and if individuals are encouraged to take action to reduce their losses.

For age, gender and territory, privacy is generally not an issue, as individuals rarely mind revealing their age, sex or where they live. The other three requirements are probably at the origin of these variables' exclusion in many jurisdictions. Age and gender are obviously not under the control of policyholders. Contrary to variables like miles driven, model of car or traffic violations, drivers have no possible action to reduce their premium, thus no incentive for safer driving. Affordability is an issue, as the drivers getting penalized, the young and the elderly, are just those who can generally ill-afford to pay high premiums, and who, more often than others, have difficulties acquiring insurance. Causality is a major issue. The link between age and claims is indirect. Causality requires much more than correlation between the variable and claim rates. Younger drivers have high accidents rates; this is however not due to their age per se, but rather to risk-taking behavior, such as driving at night, under the influence of alcohol or drugs, or at excessive speeds, often without seatbelts buckled. Claim frequencies increase for the elderly, as some older drivers begin to lose their sensory skills (vision, hearing), their cognitive skills (memory, mental agility, processing of sensory information) and motor functions (muscle strength, flexibility, endurance); moreover, some medications impair driving ability. However, the cause-and-effect relationship is missing.

A variable commonly used in some European countries that clearly fails the causality test is garage ownership. While owners of a private garage are safer drivers, there is no clear explanation why this should be the case, except through correlation with a third variable, possibly income or a caring attitude towards the car. Similarly, the "good-student discount" used by many U.S. insurers is contentious due to the lack of causality.

Variables like Internet browsing, purchasing patterns, genetic information or sexual orientation would clearly violate the privacy requirement.

2.1.4. *Legal criteria.* Insurers should be prohibited from classifications that are socially suspect. According to the U.S. Supreme Court, suspect classifications have four factors in common: history of discrimination against the group; the characteristics that distinguish the group have no relationship to its ability to contribute to society; the characteristics are immutable; the subject class lacks political power. Any classification variable that perpetuates or reinforces social inequalities can be considered as suspect, as well as any characteristic associated with historical discrimination (Gaulding, 1995). The Supreme Court specifically characterized race, religion and national origin as definitely suspect factors, and gender and illegitimacy of birth as quasi-suspect (Avraham *et al.*, 2013).

While not going as far as prohibiting the use of age, gender or marital status, the Canadian Supreme Court has requested insurers to at least explore whether better, non-discriminating, variables exist (Kelly and Nielson, 2006).

Other variables currently used by insurers could be questioned given these legal criteria. Territory, credit scores and premium payment frequency, for instance, can be challenged as proxies for the more objectionable classifiers of income and race.

### 2.2. The evaluation of mileage as a rating variable

With the use of age, gender and territory prohibited or severely curtailed, and possibly other variables such as credit score and premium payment frequency next in line, insurers need to find new variables to maintain accuracy and possibly increase drivers' incentives to reduce risk. Annual mileage is an obvious candidate that has been suggested in numerous papers, dating as far back as Bailey and Simon's seminal paper (1960).

Annual mileage easily passes many of the criteria developed in section 2.1. It passes all actuarial tests. Many papers have established a strong relationship between annual mileage and claim frequencies, a relationship that has remained stable over time. Indeed, more time spent on the road translates into more traffic incidents and situations leading to claims. The relationship is, however, less than proportional: doubling annual mileage increases the claim frequency, but does not double it, possibly because high-mileage users are more experienced or drive more on low-risk highways rather than high-risk urban areas. Despite this, the variable is highly accurate as a predictor of claims, as it depends on individuals' own behavior and is directly based on exposure to risk, and not on the behavior of groups of people such as single males or inhabitants of a given township. In addition, within-cell heterogeneity is acceptable.

Mileage also passes several operational tests. It is a numerical, hence objective, variable. Rating discontinuities can be minimized as insurers are free to subdivide their portfolios into many mileage rating classes. Mostly, there is an obvious, intuitive relationship between mileage and claims, since each mile a car travels creates a small chance of an accident.

Mileage is a socially acceptable variable, mostly because of controllability: drivers have a strong incentive to affect their accident rate by reducing their driving. It improves fairness by shifting weight in pricing towards an individually controllable factor rather than based on involuntary membership in a group.

Causality is obvious: most policyholders should accept the idea that increased driving raises the chances of an accident. There should be few, if any, legal challenges to annual mileage, as this variable is not socially suspect in any way. High road users do not constitute a group that had to face historical discrimination.

Yet, in practice, mileage is a variable that is hardly used. Some insurers use one or two cut-off mileage points, with small surcharges and discounts. The main reason for its infrequent use is that the variable, while passing a majority of criteria, badly fails other requirements. Until the advent of GPS and on-board computers, significant moral hazard was present, as drivers had a strong and obvious incentive to under-report mileage. Incorrect mileage has been reported in numerous papers, especially when there is a financial incentive to under-report (Janke, 1991; Langford *et al.,* 2008; Staplin *et al*., 2008). Odometers were easily tampered with, and the cost to control this manipulation was prohibitive, requiring for instance inspector visits to policyholders' domicile, agreements with repair shops to report odometer readings and having policyholders forward a picture of their odometer — all leading to complaints about privacy issues. As a result, understatement of annual mileage is one of the major sources of auto insurance fraud.

This situation is rapidly changing, due to fast pace of introduction of telematics, on-board computers, and GPS transmitters, and the decreasing price of these new technologies. For instance, in May 2012, a large company introduced a voluntary program in Pennsylvania to monitor mileage using a telematics device. Using the catchy slogan "Just have your car send us your driving habits", the rating plan involves the use of a transmitter that comes factory-installed in all new vehicles sold by the largest U.S. car manufacturer, or can be professionally installed on existing cars at a cost of $ 100. A required subscription costing $ 200 per year provides automatic crash response, emergency services, roadside and stolen vehicle assistance and diagnostic and maintenance information. Odometer readings are recorded and e-mailed monthly to the subscriber and the insurer. Premium discounts are offered at each renewal, for instance 32% for 3,500 annual miles, 13% for 11,000 miles, 5% for 15,000 miles.

Other companies use telematics to monitor additional driving habits, such as the use of the car between midnight and 4 a.m., speeds over 80 miles per hour, acceleration and breaking behavior and the type of roads travelled (urban, country, motorway).

While customer tracking can be perceived as an invasion of privacy (policyholders may be leery of allowing their insurance company to track their location and driving hours), and affordability may remain an issue for some categories of drivers, the two main features of telematics — their inability to be manipulated by drivers and their generally low cost — lay to rest the main criticisms against using mileage in rating.

## 3. The Data

### 3.1. Background

Taiwan has a land area of 32,260 sq. km, the size of Belgium, and a population of 23,360,000 in July 2014 (CIA, 2015). Two thirds of the country consist mostly of rugged mountains, leading to a very high population concentration in the plains. Due to nightmarish driving conditions (high population density, 6.5 million motorcycles sharing the road with cars, unavailable parking in cities), only 4,826,000 non-commercial sedans were registered in 2012, a very low number for an affluent country, with a GPD per capita (corrected for purchasing power) of \$ 43,600, higher than France (Taiwan Insurance Institute, 2014). It is rare for young individuals to own a car. Very few couples own two cars.

Automobile insurance is organized in a somewhat different way than in most western countries. Compulsory liability only covers bodily injury losses up to a limit, currently NT\$ 2,200,000 per person (1 NT\$ = US\$ 31.27 as of April, 2015). The small increase of the limit during our observation period, from NT\$ 1,600,000 to NT\$ 1,700,000, is not expected to impact our study, as the vast majority of policyholders purchase coverage above the limit. Voluntary policies provide additional third-party bodily injury and property damage coverage. Our data pool all of these policies, which are subject to the same rating variables and BMS. First-party collision coverage is also available, but not considered in this study, as another BMS is used.

Only three a priori classification variables are used by insurers for rating purposes: use of car (personal/business), gender and driver age (< 20, 20–25, 25–30, 30–60, >60 years). As females receive a discount, a fact well known to Taiwanese households, it is a common practice for couples to register their car to the female driver. As a result, while the vast majority of drivers on the road are males, insurers report 70% of female drivers in their portfolios.

The BMS has no upper limit in the malus zone. However, no single driver in our sample pays more than a 60% surcharge. Therefore, we can model the Taiwanese BMS as a 10-class Markov Chain, with premiums levels 70, 74, 82, 100, 110, 120, 130, 140, 150 and 160. New drivers start in class 4, at level 100. Claim-free years are rewarded by a one-class discount. Each claim is penalized by three classes (Taiwan Insurance Institute, 2015).

### 3.2. Data

Our large database (over a quarter million policy-years) was produced by pooling claim and policy information from the largest auto insurer operating in Taiwan (market share: 20%) with maintenance records from a chain of repair shops operated by the largest car manufacturer (market share: 38%). Repair records resulting from an accident were excluded, to avoid introducing a bias in the database. Besides the number and severity of claims, insurance variables include gender, age and marital status of the main driver, territory, use of car, BMS class, engine cubic capacity and date of first registration of the car. As odometer readings are systematically collected by repair shops during each visit, interpolation or extrapolation of odometer values between visits allows us to estimate annual mileage. Data are available for seven policy years, 2001 to 2007. All policyholders purchased the compulsory policy; 88.82% bought additional voluntary insurance.

All claims, whether reported under the compulsory contract or one of the voluntary policies, are recorded. A claim may trigger a payment under a compulsory and/or a voluntary policy. To avoid double counting, claims reported on the same date under two or three policies are counted as a single claim.

TABLE 1

PERCENTAGE OF DRIVERS PURCHASING VOLUNTARY COVERAGE ACCORDING TO MILEAGE CLASS.

| Mileage Decile | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| % Voluntary | 87.86 | 87.94 | 88.16 | 89.08 | 89.38 | 88.93 | 89.02 | 89.14 | 89.62 | 89.12 |

Some claims may be missed, for instance, a property damage only claim, if the driver did not purchase the corresponding voluntary coverage — not a likely occurrence since nearly 89% of drivers in our sample bought it. This may raise a problem if high-mileage users are more prone to purchase additional insurance. If this is the case, more claims will be missed among the low-mileage drivers, and the impact of mileage on claim frequencies may be somewhat overstated. Such a behavior is well-known in collision coverage, but fortunately does not take place in our third-party sample, as shown in Table 1. (Policies are ranked by increasing mileage, and subdivided into ten equal-sized deciles.)

For all policyholders in our sample, the values of the following variables are recorded:

Gender is a classification variable used in rating. Only 29.49% of drivers are registered as males, a clear indication that policyholders take advantage of their knowledge of differential rates to get a premium discount. So it is all but certain that policies registered in the "female driver" category include a large number of cars owned by couples, often driven by males.

Age is also used in rating. While for rating purposes, the company uses five age categories (< 20, 20–25, 25–30, 30–60, >60 years), less than 1% of drivers are between ages 20 and 25 years, and only a handful are between 18 (the minimum driving age) and 20 years. Consequently, we combine the first three age groups and end up with three classes: under 30 (7.38% of drivers), 30–60 (88.76%), over 60 (3.86%) years.

Bonus-malus premium level, from 70 to 160, as described in section 3.1.

Vehicle type and use. Since 97.9% of cars are registered as non-commercial sedans, we discard the remaining categories (business use, trucks, passenger coaches, taxis).

Mileage is expressed in kilometers driven per day. Repair shop technicians know the date the car was put in service, which allows for a first estimate of mileage upon the first oil change. The date and odometer reading are recorded on each visit to the shop. Extrapolation or interpolation then yields an estimate of annual mileage. For instance, assume a driver has three visits to the repair shop. His odometer readings are 13,200 on October 1, 2001, 24,400 on April 1, 2002 (182 days later, 91 days into 2002, 274 days before January 1, 2003), and 37,400 on January

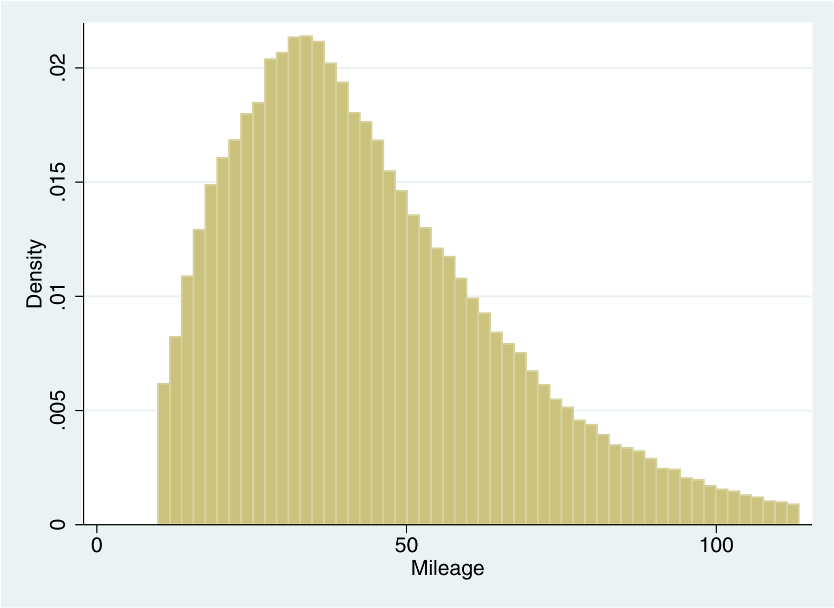



FIGURE 1: Distribution of daily kilometers driven. (Color online)

15, 2003 (289 days later, 15 days into 2003). The estimate of the number of kilometers driven in 2002 is

$$(24,400-13,200)\times(91/182)+(37,400-24,400)\times(274/289) = 17,925.$$

A visual inspection of the data shows numerous instances of obvious recording mistakes, with mileages like −44,581 km or +24,833 km. Truncating the upper and lower 1% of the data seems a conservative approach, eliminating all unrealistic figures. The truncation daily mileage varies across policy years, averaging 7.43 km and 133.37 km. After eliminating business users, trucks, and the unreasonable mileage figures, the total sample size is 259,029. Figure 1 shows the distribution of daily kilometers driven of our sample. The average annual number of kilometers driven per car is 16,167.

Several other variables are recorded for classification purposes.

Marital status. 92.03% of policy owners are married.
Car age. 26.45% of cars in the sample are under one year of age; 26.19% are between ages 1 and 2 years; 18.4% between ages 2 and 3 years; 12.38% between ages 3 and 4 years; 8.05% between ages 4 and 5 years; and 8.53% are older.
City. 49.99% of our sample drivers live in an urban area.

TABLE 2

DISTRIBUTION OF NUMBER OF CLAIMS FOR ENTIRE SAMPLE.

| Number of Claims | Number of Policies | Percentage |
|---|---|---|
| 0 | 247,955 | 95.72% |
| 1 | 8,222 | 3.17% |
| 2 | 2,689 | 1.04% |
| 3 | 136 | 0.05% |
| 4 | 25 | 0.01% |
| 5 | 2 | 0.00% |
| Total | 259,029 | 100% |

Territory. 47.45% of cars are registered in the north of Taiwan; 30.16% in the south; 17.31% in central Taiwan and 5.08% in the eastern part of the island.

Engine cubic capacity. The engine capacity is under 1,800 cc for 65.80% of cars; between 1,800 cc and 2,000 cc for 28.92% of cars and above 2,000 cc for the remaining 5.28%.

### 3.3. Summary statistics

Table 2 provides the distribution of the number of claims for the entire sample of 259,029 policies. As is common in observed claim count distributions in auto insurance, the sample variance (0.0768) is substantially larger than the sample mean (0.0545). This will require the use of negative binomial regression, rather than Poisson regression, in the statistical analysis. Figure 2 graphs the distribution of the natural logarithm of claim amounts.

The 259,029 policies are subdivided in ten deciles. Table 3 presents the mileage limits and mean number of kilometers driven in each mileage class (averaged across years), as well as the means and variances of the ten claim count distributions, and the means and variances of the logarithm of claim severities, expressed in U.S. dollars. Figure 3a plots the claim frequencies for the ten mileage classes and the 95% confidence intervals of the point estimates of the average claim frequency. Figure 3b graphs log (claim severity).

As expected, claim frequencies increase with mileage, but in a less-than-proportional way. Drivers in the top mileage decile have about three times as many accidents as those in the bottom decile. The variance of the claim number increases with mileage. There is no overlap between the confidence intervals for the upper and lower deciles, confirming a strong, significant, positive relationship between annual mileage and accidents. The fact that claim frequencies in deciles 1 and 2 are nearly identical provides support for the "low mileage bias" (Langford *et al.*, 2008; Staplin *et al.*, 2008), the observation that infrequent users of their cars, mostly elderly motorists, have a higher per-mile accident rate, as they mostly drive in congested urban areas.

TABLE 3

CLASS LIMITS AND MEAN DAILY MILEAGE FOR ALL TEN MILEAGE CLASSES, MEANS AND VARIANCES OF CLAIM COUNT DISTRIBUTIONS, MEAN CLAIM SEVERITY.

| Mileage Decile | Average Class Limit | Mean Daily km Driven | Claim Frequency | Variance of Claim Number | Log (Claim Severity) (USD) | Variance of Log Severity |
|---|---|---|---|---|---|---|
| 1 | 6.57–18.80 | 14.45 | 0.0349 | 0.0499 | 6.2139 | 1.3037 |
| 2 | 18.80–24.90 | 21.95 | 0.0354 | 0.0496 | 6.1718 | 1.3354 |
| 3 | 24.90–30.13 | 27.59 | 0.0416 | 0.0575 | 6.2300 | 1.3502 |
| 4 | 30.13–34.91 | 32.52 | 0.0491 | 0.0677 | 6.1985 | 1.2318 |
| 5 | 34.91–39.90 | 37.37 | 0.0501 | 0.0706 | 6.2443 | 1.3645 |
| 6 | 39.90–45.68 | 42.74 | 0.0562 | 0.0789 | 6.2975 | 1.3037 |
| 7 | 45.68–52.63 | 49.02 | 0.0578 | 0.0807 | 6.3057 | 1.4142 |
| 8 | 52.63–61.62 | 56.92 | 0.0631 | 0.0873 | 6.2637 | 1.3373 |
| 9 | 61.62–75.73 | 67.96 | 0.0726 | 0.1013 | 6.3577 | 1.3722 |
| 10 | 75.73+ | 92.41 | 0.0843 | 0.1219 | 6.3794 | 1.3076 |

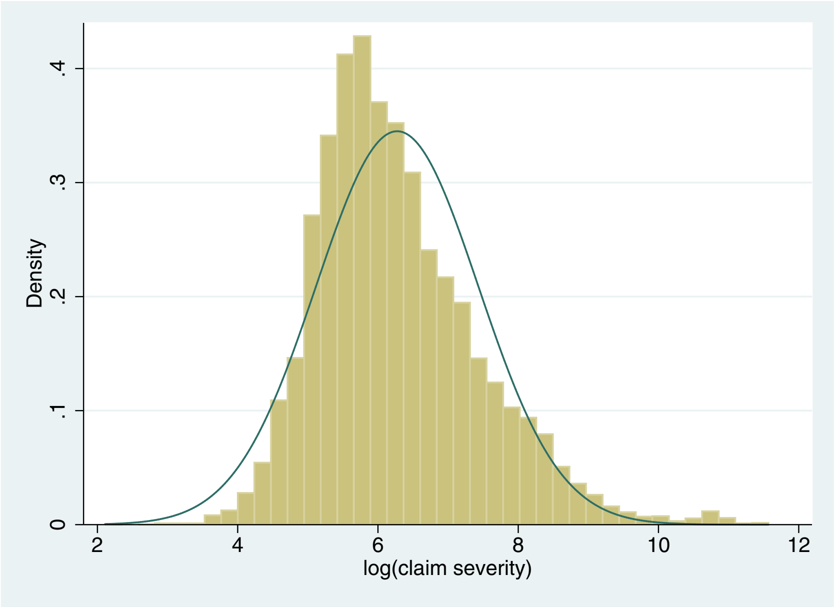
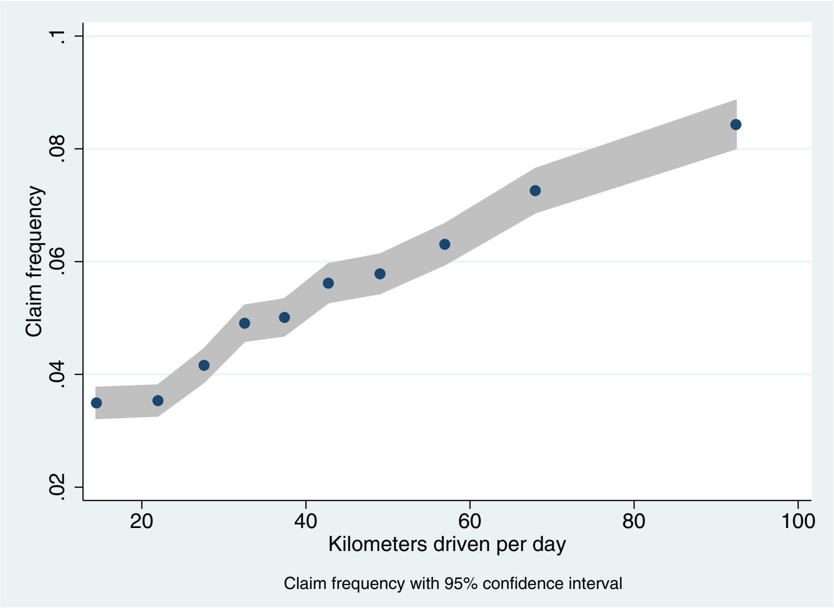
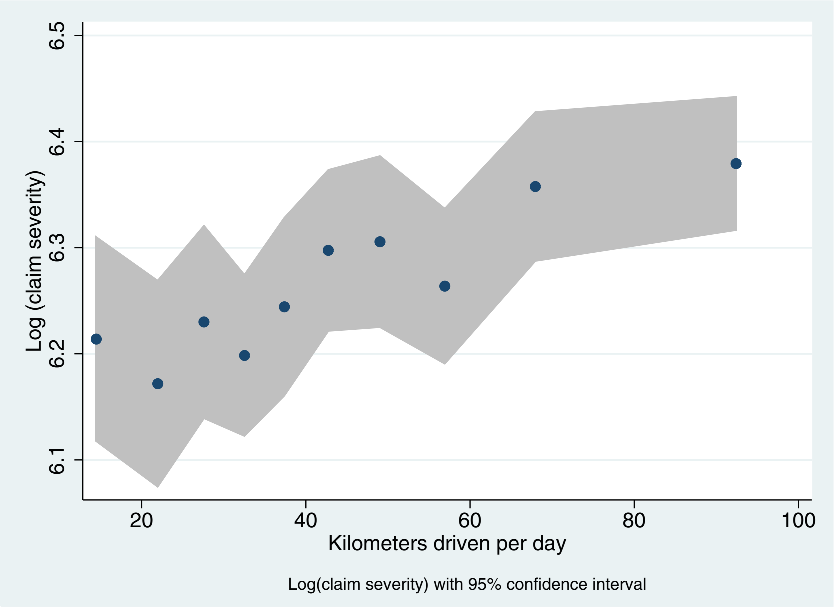



FIGURE 2: Distribution of claim severity (natural log of claim amount). (Color online)

For all available categorical variables, Table 4 provides the percentage of policyholders for each variable category, as well as claim frequencies and mean severities. Claim frequency differences across categories are smaller than differences found across mileage classes in Table 3. Means range from 0.0424 to 0.0742 in Table 4, whereas the mileage means range from 0.0349 to 0.0843, a larger variation. Females are at-fault in more accidents (female claim frequency:




(a)




(b)

FIGURE 3: Claim frequencies (3a) and severities (3b) as a function of daily kilometers, with 95% confidence intervals of point estimates of mileage decile average claim frequency and severity. (Color online)

TABLE 4

CLAIM FREQUENCIES AND SEVERITIES FOR RATING FACTORS.

| Category | Variable | Percentage (%) | Claim Frequency (%) | Claim Severity (USD) |
|---|---|---|---|---|
| Age | Age<30 | 7.38% | 6.57% | 1,644.49 |
| | Age30–60 | 88.79% | 5.37% | 1,364.26 |
| | Age60+ | 3.83% | 5.02% | 1,384.58 |
| Gender | Female | 70.51% | 5.67% | 1,331.71 |
| | Male | 29.49% | 4.93% | 1,547.11 |
| Married | Married | 92.05% | 5.41% | 1,388.32 |
| | Not Married | 7.95% | 5.91% | 1,402.5 |
| Car Age | Car age 0 | 26.45% | 7.42% | 1,569.13 |
| | Car age 1 | 26.19% | 5.24% | 1,331.2 |
| | Car age 2 | 18.40% | 4.58% | 1,310.77 |
| | Car age 3 | 12.38% | 4.60% | 1,242.28 |
| | Car age 4 | 8.05% | 4.24% | 1,179.25 |
| | Car age 5 | 8.53% | 4.24% | 1,283.57 |
| Capacity | Capacity 1 | 65.80% | 5.76% | 1,359.22 |
| | Capacity 2 | 28.91% | 4.93% | 1,461.41 |
| | Capacity 3 | 5.28% | 4.52% | 1,432.32 |
| Region | City | 5.08% | 6.20% | 1,837.63 |
| | North | 47.45% | 4.89% | 1,247.59 |
| | South | 30.19% | 5.83% | 1,449.32 |
| | Middle | 17.28% | 6.00% | 1,479.09 |

0.0567; male: 0.0493), but these accidents are on average less costly, which may justify the discount awarded to females.

Table 5 presents Pearson correlation coefficients for all continuous variables, using the logarithm of claim severity due to the high skewness of this variable. Spearman and Kendall correlations are very similar. Due to the large sample size, correlation coefficients between mileage and all other variables are statistically significant at the 1% level. Correlation coefficients show that young drivers tend to drive more, older drivers less, on average. As expected, urbanites drive less than rural policyholders. Mileage is positively related to the BMS coefficient, suggesting that, if mileage is not used as a rating variable, the information it contains is partially reflected through the BMS coefficient. The positive relationship between female and married further confirms our conjecture that married couples report the female as the main driver to get the insurance discount.

## 4. Regression Analysis of Claim Frequency and Severity

### 4.1. Claim frequency: Negative binomial regression results

Poisson or negative binomial regressions are typically used for count dependent variables. Negative binomial regression fits the data better when modeling

TABLE 5

PEARSON CORRELATIONS OF CONTINUOUS VARIABLES, WITH SIGNIFICANCE LEVELS.

| | Mileage | Claim Freq | Log (sev) | Driver_age | Car_age | Capacity | BMS |
|---|---|---|---|---|---|---|---|
| Mileage | 1 | 0.05512 | 0.04567 | −0.0735 | −0.05292 | 0.06728 | 0.05305 |
| | | (<0.0001) | (<0.0001) | (<0.0001) | (<0.0001) | (<0.0001) | (<0.0001) |
| Claim_freq | | 1 | 0.25112 | −0.00501 | −0.03445 | −0.01283 | 0.04160 |
| | | | (<0.0001) | (0.0108) | (<0.0001) | (<0.0001) | (<0.0001) |
| Log (sev) | | | 1 | 0.03529 | −0.02046 | 0.02058 | 0.02125 |
| | | | | (0.0002) | (0.0313) | (0.0303) | (0.0254) |
| Driver_age | | | | 1 | −0.0046 | 0.05913 | −0.11049 |
| | | | | | (0.0313) | (<0.0001) | (<0.0001) |
| Car_age | | | | | 1 | −0.03206 | −0.61857 |
| | | | | | | (<0.0001) | (<0.0001) |
| Capacity | | | | | | 1 | −0.01205 |
| | | | | | | | (<0.0001) |
| BMS | | | | | | | 1 |

Note: Numbers in parentheses show *p*-values.

over-dispersed count outcome variables, as is the case with our claim number distribution. As claim frequency shows marked over-dispersion, we use negative binomial regression with log link function. In order to allow for an individual specific dispersion parameter, we run a random effect negative binomial regression model, as suggested by Hausman *et al.* (1984) and Boucher and Guillen (2009). A more extensive discussion of model selection is found in section 5.

Table 6 reports the negative binomial regression results. Model (1) regresses only the mileage variable, which turns out to be highly significant. We add a mileage square term in Model (2) in order to test for a possible non-linear relationship between claim frequency and mileage, which is observed in Figure 3a. The significantly positive mileage term and the significantly negative square mileage term together indicate that claims increase with mileage less than proportionally; the curve plotting claim frequency as a function of mileage is increasing and concave. Model (3) regresses all current rating variables. While all variables are highly significant, the overall Chi-square is higher and the log likelihood is lower than in model (1), suggesting that mileage alone explains claim rates better than all current rating variables combined. Compared to drivers under the age of 30 years, discounts given to older drivers, particularly those between 30 and 60 years, are entirely justified.

In model (4), mileage is added to the current variables, and found, as expected, to have a hugely significant positive effect: it has the largest $z$-score (27.2) of all variables, followed by BMS (17.60). If only one variable is to be used, it should be mileage, consistent with the results in models (1)–(2) and the summary statistics section. The use of mileage in rating eliminates the need for discounts for older drivers; presumably, policyholders over the age of 60 years, many of

TABLE 6

CLAIM FREQUENCY PANEL NEGATIVE BINOMIAL REGRESSION RESULTS.

| Variables | (1) Mileage Only | (2) Mileage^2 | (3) Current Rating | (4) Current Rating with Mileage | (5) All Observable | (6) All Observable with Mileage | (7) Significant Variables Only |
|---|---|---|---|---|---|---|---|
| Mileage | 0.0107*** | 0.0137*** | | 0.0136*** | | 0.0140*** | 0.0139*** |
| | [0.0004] | [0.0005] | | [0.0005] | | [0.0006] | [0.0005] |
| Mileage^2 | | −0.0001*** | | −0.0001*** | | −0.0001*** | −0.0001*** |
| | | [0.0000] | | [0.0000] | | [0.0000] | [0.0000] |
| Age 30–60 | | | −0.1552*** | −0.1113*** | −0.1241*** | −0.0738** | −0.0676** |
| | | | [0.0336] | [0.0336] | [0.0351] | [0.0352] | [0.0299] |
| Age 60+ | | | −0.1797*** | −0.0626 | −0.1399** | −0.0141 | |
| | | | [0.0607] | [0.0608] | [0.0619] | [0.0621] | |
| Female | | | 0.1242*** | 0.1607*** | 0.1013*** | 0.1387*** | 0.1377*** |
| | | | [0.0216] | [0.0216] | [0.0218] | [0.0219] | [0.0218] |
| Bonus-Malus | | | 1.3768*** | 1.3030*** | 0.5529*** | 0.4939*** | 0.4897*** |
| | | | [0.0740] | [0.0740] | [0.1182] | [0.1174] | [0.1169] |
| Married | | | | | −0.0679* | −0.0719** | −0.0763** |
| | | | | | [0.0354] | [0.0354] | [0.0348] |
| Car Age 0–1 | | | | | 0.3527*** | 0.3500*** | 0.3280*** |
| | | | | | [0.0504] | [0.0503] | [0.0352] |
| Car Age 1–2 | | | | | 0.1523*** | 0.1366*** | 0.1162*** |
| | | | | | [0.0439] | [0.0439] | [0.0260] |
| Car Age 2–3 | | | | | 0.0298 | 0.0215 | |
| | | | | | [0.0453] | [0.0453] | |
| Car Age 3–4 | | | | | 0.0581 | 0.0483 | |
| | | | | | [0.0479] | [0.0480] | |
| Car Age 4+ | | | | | −0.0079 | −0.0108 | |

TABLE 6

*Contd.*

| Variables | (1) Mileage Only | (2) Mileage^2 | (3) Current Rating | (4) Current Rating with Mileage | (5) All Observable | (6) All Observable with Mileage | (7) Significant Variables Only |
|---|---|---|---|---|---|---|---|
| | | | | | [0.0533] | [0.0533] | |
| Eng. Capacity 2 | | | | | −0.1222*** | −0.1647*** | −0.1608*** |
| | | | | | [0.0222] | [0.0222] | [0.0221] |
| Eng. Capacity 3 | | | | | −0.1260*** | −0.1559*** | −0.1539*** |
| | | | | | [0.0458] | [0.0458] | [0.0457] |
| City | | | | | −0.0095 | 0.0536*** | |
| | | | | | [0.0200] | [0.0201] | |
| North | | | | | −0.1517*** | −0.1463*** | −0.1168*** |
| | | | | | [0.0436] | [0.0436] | [0.0241] |
| South | | | | | −0.0781* | −0.0794* | −0.0597** |
| | | | | | [0.0437] | [0.0437] | [0.0256] |
| Middle | | | | | −0.0444 | −0.0321 | |
| | | | | | [0.0468] | [0.0468] | |
| Wald Chi2 | 930.6 | 924.5 | 533.1 | 1,322 | 690.8 | 1,510 | 1,501 |
| Log Likelihood | −52,813 | −52,782 | −52,987 | −52,590 | −52,908 | −52,496 | −52,501 |
| Logarithmic Score 10-Fold Cross Validation | 0.2114 | 0.2113 | 0.2121 | 0.2107 | 0.2118 | 0.2105 | 0.2,105 |

Note: *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Numbers in parentheses provide standard errors. Unbalanced panel negative binomial regression is used. Chi2, log likelihood and logarithmic score 10-fold cross validation are presented as goodness of fit measures. Larger Chi2 and log likelihoods and smaller logarithmic scores indicate a better fit. Year dummy variables are included but the coefficients are not reported here.

them retired, spend less time on the road. So mileage reflects a large part of the age effect on claims, a useful outcome since age is one of the variables that regulators criticize. Although “Age 30–60 years” remains significant after the inclusion of mileage, its magnitude decreases when controlling for mileage.

Even after the inclusion of mileage, BMS remains a significant predictor. One of the main reasons for BMS is classification; BMS may pick up information not revealed to insurers or not used in rating. Therefore, there was a distinct possibility that the introduction of a powerful classification variable such as mileage would have lessened the need for the use of BMS in rating, maybe up to the point of making BMS insignificant. This did not occur: BMS remains a dominant variable, with its coefficient hardly decreased. Therefore, BMS contains important information not reflected by mileage, and should remain an important component in pricing.

BMS and mileage include important, but different, information about the risk a policyholder constitutes. Mileage carries present, or at least very recent, information: the latest knowledge about the amount of driving of the insured. BMS summarizes past information, as the current BMS level results from claim history since the inception of the policy. So BMS captures material from the past, including previous mileage, but also overall respect of the laws and the driving code, alcohol consumption, road rage behavior, ability to react to crisis situations, processing of dangerous circumstances, etc. Replacing current mileage by lagged mileage in our models did not affect results in any way: mileage and BMS remain highly significant. Therefore, the information contained in BMS reflects much more than past mileage.

Model (5) includes all available variables but mileage. Despite the large number of observations, few variables turn out to be highly significant. The claim frequency is a decreasing function of car age, with newer cars involved in more accidents. This could possibly be a spurious relationship, due to an omitted variable: our data do not include driving experience, a rating variable used in several countries. New drivers must start their driving career in the initial class of the BMS, at a premium level of 100. Given the lenient rules of the Taiwanese BMS, maluses never compensate bonuses, and the average driver has a premium level of 81. So it may be that that “new car” is somewhat synonymous with “recently licensed driver”, a conjecture supported by the large positive correlation between BMS and Car Age 0–1 years.

Compared to smaller cars, autos with a large engine have a reduced claim frequency. This point needs cautious interpretation because our sample consists of just one brand which does not build cars with super-sized engines. Among the geographical variables, only drivers from the northern part of the country could claim a discount, probably because of the better roads in this part of the state and the separate lanes for scooters. BMS remains significant at the 1% level, but its importance in model (5) is much decreased compared to model (3), as measured by the large reduction of its size of coefficient. The use of car age, engine cubic capacity and territory lessens the need for a sophisticated BMS. The Taiwanese BMS is fairly mild: penalties are not severe when compared to

TABLE 7

INCIDENCE RATE RATIOS (IRR).

| Variables | Coefficients | $z$-score | Standard Errors | IRR |
|---|---|---|---|---|
| Mileage | 0.0140*** | 25.40 | 0.0006 | 1.0141 |
| Mileage^2 | −0.0001*** | −8.27 | 0.0001 | 0.9999 |
| Age 30–60 | −0.0738** | −2.10 | 0.0352 | 0.9289 |
| Age 60+ | −0.0141 | −0.23 | 0.0621 | 0.9860 |
| Female | 0.1387*** | 6.34 | 0.0219 | 1.1488 |
| Bonus-Malus | 0.4939*** | 4.20 | 0.1174 | 1.6387 |
| Married | −0.0719** | −2.03 | 0.0354 | 0.9306 |
| Car Age 0–1 | 0.3500*** | 6.96 | 0.0503 | 1.4191 |
| Car Age 1–2 | 0.1366*** | 3.11 | 0.0439 | 1.1464 |
| Car Age 2–3 | 0.0215 | 0.47 | 0.0453 | 1.0217 |
| Car Age 3–4 | 0.0483 | 1.01 | 0.0480 | 1.0495 |
| Car Age 4+ | −0.0108 | −0.20 | 0.0533 | 0.9893 |
| Engine Capacity 2 | −0.1647*** | −7.41 | 0.0222 | 0.8481 |
| Engine Capacity 3 | −0.1559*** | −3.40 | 0.0458 | 0.8556 |
| City | 0.0536*** | 2.67 | 0.0201 | 1.0551 |
| North | −0.1463*** | −3.36 | 0.0436 | 0.8639 |
| South | −0.0794* | −1.82 | 0.0438 | 0.9237 |
| Middle | −0.0321 | −0.69 | 0.0468 | 0.9684 |
| Wald Chi2 | 1,510 | | | |
| Log Likelihood | −5,2496 | | | |

BMS in force in most other countries (Lemaire and Zi, 1994). Should Taiwanese companies decide to make transition rules and premium levels differentials more severe, the significance of BMS would certainly increase. Introducing new variables such as car age, territory and cubic capacity instead of a more severe BMS, while actuarially justified, would result in a complicated rating system with a large number of variables, which would be more difficult to understand by brokers and consumers. So BMS should remain an important component of auto insurance rating.

Adding mileage to all variables (Model 6), or regressing only significant variables (Model 7) hardly modifies the strong conclusions of this analysis.

Table 7 shows the Incidence rate ratios (IRR) and $z$-scores of Model (6), the full regression model from Table 6. The $z$-scores of mileage and squared mileage are the largest among all variables, indicating that, by far, mileage is the most accurate variable that insurers could introduce. The impact of mileage on claim frequencies surpasses the influence of all other variables, including BMS, by a wide margin. Mileage IRR show that driving an additional kilometer increases the chance of accident by 1.41%., everything else being equal. Note that in all models, we add a year fixed effect in order to control for year-specific events such as weather and road condition changes, which may affect everyone that year.

TABLE 8

CLAIM SEVERITY RANDOM EFFECT PANEL REGRESSION RESULTS.

| Variables | Coefficients | $z$-score | Standard Errors |
|---|---|---|---|
| Mileage | 0.0019*** | 4.18 | 0.0005 |
| Female | −0.0552** | −2.20 | 0.0251 |
| Engine Capacity 2 | 0.0505** | 1.98 | 0.0255 |
| City | −0.0795*** | −3.56 | 0.0223 |
| Wald Chi2 | 60.49 | | |
| R2 (overall) | 0.0056 | | |
| Rho | 0.2000 | | |
| Breusch and Pagan LM test Chi2 | 7.22 | | |
| RMSE 10-fold Cross Validation | 1.153 | | |

Note: *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Unbalanced panel random effect linear regression is used. Chi2, R2 and RMSE (Root mean square error) are presented as goodness of fit measures. Larger Chi2 and R2 and smaller RMSE indicate a better fit. Year dummy variables are included but the coefficients are not reported here.

### 4.2. Claim severity: Linear regression model results

Mileage clearly impacts claim frequency, but does it influence claim severity? Is the cost of an accident mostly random, or are high road users involved in more severe crashes, maybe because they drive more on freeways, and thus faster? We run the same set of random effect linear regressions as in Table 6, with log claim severity as a dependent variable. As most of the variables prove to be statistically insignificant (providing some support that the cost of an accident is for a large part random), only results with significant variables are presented in Table 8. The Breusch and Pagan (1979) LM test shows that the random effect model fits the data better than OLS.

Mileage turns out to be the most significant variable, in a set of only four. The claim severity of a driver in the top mileage decile driving 92 km a day is about 15% higher or about U.S. $ 200 more than a driver in the bottom mileage decile driving 14 km a day, everything else being equal. The effect of mileage on severity is positive, but much smaller than the effect on frequency. This further justifies the use of mileage as a rating variable. The squared mileage term is insignificant in the severity regression.

As intuitively expected, female drivers have more accidents on average but their severity is lower, by about 5%. In cities, where traffic density is higher, more accidents take place, but less severe accidents.

## 5. ROBUSTNESS CHECKS

### 5.1. Alternate model: Regressions with dummy mileage variables

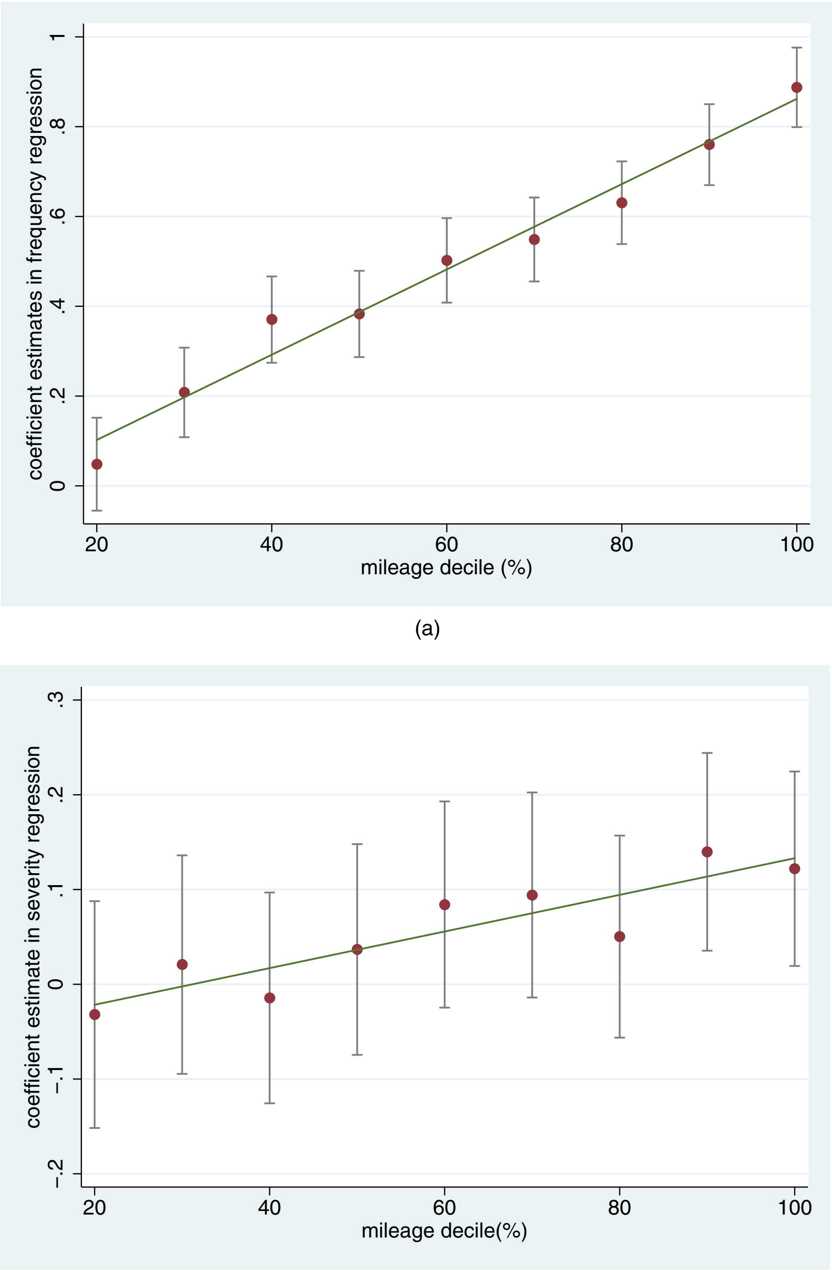
The use of mileage as a continuous variable implies a linear dependence. Non-linear or non-monotonic relationships are certainly possible, at least in certain mileage ranges. The positive association found in the previous section could be driven mostly by certain mileage levels. To rule out such a possibility, we run an alternative model using mileage decile dummies instead of a continuous mileage variable. Table 9 reports negative binomial regression results with all available variables, but with the continuous mileage variable replaced by nine dummy variables characterizing the ten mileage deciles. Regression coefficients for all other variables are barely affected. The Chi-square and log likelihood of the model with dummy variables are quite similar to the results in Model (6) of Table 6, indicating that the use of mileage deciles provides as much information as the continuous variable. In addition, categorical dummy variables reveal a slight non-linear relationship for low-mileage users, as shown in Figure 4. Controlling for all other factors, mileage exhibits a monotonically increasing relationship, both for frequency and severity. Therefore, every single mileage decile carries significant information, and the practice of some insurers to introduce in rating just one mileage cut-off point ("the low-mileage discount") is inefficient from an actuarial perspective, as valuable information is lost. Comparing IRR in Table 9, drivers in the top mileage decile have about 2.43 times more accidents per year than policyholders in the lowest mileage decile. None of the other categorical variables shows such a strong effect on claim frequency.

Our findings are comparable with the results in Paefgen *et al*. (2014), who analyzed a sample of 27,600 vehicle months over two years. Detailed In-Vehicle Recorders Data from a major European Pay-As-You-Drive insurance company enabled them to use variables such as time of day, day of week, velocity. Therefore, our study cannot be expected to provide the same results, due to omitted variables issues and major sample differences: we control for potential rating variables, Paefgen *et al*. (2014) control for the driving situation. However, the two studies overall provide very similar results, mostly a strong positive relationship between claim frequency and mileage. Paefgen *et al*. (2014) find a stronger non-linear relationship, with lower accident rates in the low-mileage area and a less-than-proportional increase for high mileage. Our results are similar for low mileage, but differ in the high-mileage zone. This seems to be largely due to sample differences. Our data is from Taiwan, a relatively small country with high traffic density. The average daily mileage of 92 km in the top decile corresponds to the eighth decile of Paefgen *et al*. (2014)'s sample. Truncating Paefgen *et al*. (2014)'s results at their eighth decile leads to very similar results.

### 5.2. Cross-validation

The purpose of our research is to evaluate mileage as a potential rating variable by comparing its predictive power to other classification variables. Cross-validation is an important component of predictive modeling, as for instance adding more variables reduces the training error but may result in sample overfitting, hence larger predicting errors (Geisser, 1993).

Among the various methods available, we run the widely-used 10-fold cross-validation as it is known to work well in model selection (Kohavi, 1995). We




(a)



(b)

FIGURE 4: Mileage dummy coefficients of frequency (4a) and severity (4b) regressions. (Color online)

TABLE 9

NEGATIVE BINOMIAL REGRESSION WITH DUMMY MILEAGE VARIABLES.

| | Claim Frequency | | | Claim Severity | |
|---|---|---|---|---|---|
| Variables | Coefficient | Standard Error | IRR | Coefficient | Standard Error |
| Mileage 1 | 0.0483 | [0.0528] | 1.0495 | −0.032 | [0.0611] |
| Mileage 2 | 0.2081*** | [0.0509] | 1.2313 | 0.0221 | [0.0588] |
| Mileage 3 | 0.3703*** | [0.0491] | 1.4482 | −0.0152 | [0.0567] |
| Mileage 4 | 0.3829*** | [0.0490] | 1.4665 | 0.0354 | [0.0567] |
| Mileage 5 | 0.5022*** | [0.0480] | 1.6524 | 0.0840 | [0.0555] |
| Mileage 6 | 0.5487*** | [0.0477] | 1.7311 | 0.0924* | [0.0552] |
| Mileage 7 | 0.6306*** | [0.0470] | 1.8788 | 0.0482 | [0.0544] |
| Mileage 8 | 0.7600*** | [0.0460] | 2.1382 | 0.1367** | [0.0532] |
| Mileage 9 | 0.8875*** | [0.0452] | 2.4292 | 0.1176** | [0.0522] |
| Age 30–60 | −0.0745** | [0.0352] | 0.9282 | −0.0333 | [0.0412] |
| Age 60+ | −0.0166 | [0.0621] | 0.9835 | 0.1024 | [0.0723] |
| Female | 0.1369*** | [0.0219] | 1.1467 | −0.0553** | [0.0256] |
| Married | −0.0709** | [0.0354] | 0.9316 | −0.0150 | [0.0415] |
| Bonus-Malus | 0.4943*** | [0.1175] | 1.6394 | −0.0569 | [0.1301] |
| Car age 0–1 | 0.3501*** | [0.0503] | 1.4193 | 0.0882 | [0.0575] |
| Car age 1–2 | 0.1370*** | [0.0439] | 1.1469 | 0.0180 | [0.0507] |
| Car age 2–3 | 0.0222 | [0.0453] | 1.0225 | 0.0264 | [0.0523] |
| Car age 3–4 | 0.0484 | [0.0480] | 1.0496 | −0.0245 | [0.0552] |
| Car age 4+ | −0.0107 | [0.0533] | 0.9894 | −0.0219 | [0.0612] |
| Engine capacity 2 | −0.1632*** | [0.0222] | 0.8494 | 0.0536** | [0.0258] |
| Engine capacity 3 | −0.1541*** | [0.0458] | 0.8572 | 0.0329 | [0.0537] |
| City | 0.0520*** | [0.0201] | 1.0534 | −0.0704*** | [0.0234] |
| North | −0.1470*** | [0.0436] | 0.8633 | −0.0673 | [0.0508] |
| South | −0.0801* | [0.0437] | 0.9230 | −0.0072 | [0.0512] |
| Middle | −0.0322 | [0.0468] | 0.9683 | 0.0157 | [0.0548] |
| Wald Chi2 | 1,475 | | | 93.83 | |
| Likelihood Ratio | −52,508 | | | | |
| R2 (overall) | | | | 0.0086 | |

Note: *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Numbers in parentheses provide standard errors. Unbalanced panel negative binomial regression is used for claim regression and random effect linear regression model is used for severity regression. Larger Chi2 and R2 indicate a better fit. Year dummy variables are included but the coefficients are not reported here.

randomly subdivide the data into a training sample and a testing (or hold-out) sample. 10% of the dataset becomes the hold-out sample. We fit the model using the training sample and evaluate accuracy using the estimated coefficients from that sample. We repeat this 10 times. To measure accuracy, we calculate the Root Mean Square Error (RMSE) in the severity regression and scoring rule in the negative binomial regression. We calculate logarithmic scores, as Bickel (2007) has shown that, overall, they outperform quadratic or spherical scoring. The score is defined as

$$\text{Score}(y,\ P) = -\log f(y),$$

where $f(y)$ is the predictive probability with mass function $\Pr(Y = y)$. We compute the average of this score over all ten testing samples. A forecast that is closer to the true probability receives a lower penalty. Therefore, the lower the score, the better the model. This metric is reported in Table 6. The logarithmic score is the lowest in model (6) where all variables are used as explanatory variables, and highest in model (3), implying that the model that includes the mileage variable is not over-fitted and improves predictability. Comparing models (2) and (3), mileage alone outperforms all current rating variables combined in terms of predictive power. The logarithmic score also indicates that using all other observable variables [model (5)] increases predictability but still underperforms mileage [models (1) and (2)].

### 5.3. Robustness of results and model selection

The Poisson model is widely used to model claim counts, but it fails to adjust for overdispersion. Overdispersion is taken into account through negative binomial regression. Both regression techniques do not factor in longitudinal data sets. Our sample has a somewhat limited number of years: the maximum duration of a policy is seven years, but the average is close to two. Therefore, the effectiveness of models utilizing the longitudinal feature of our data is somewhat doubtful (Gourieroux and Jasiak, 2004). Still, it is worthwhile to check the robustness of the model selection, as the relatively short observation period may possibly bias our results. For example, if high or low mileage drivers systemically move out, an attrition problem may result. To address this concern, we calculate the average mileage of the drivers who stay with the company, and of those who move out, and find that the difference is ignorable. Policyholders staying in the sample drive about 0.2 km more per day.

Boucher and Inoussa (2014) describe three types of models for longitudinal data. As Boucher *et al.* (2008) suggest that the conditional model performs poorly in fitting, we run two alternatives, the random effect model and the marginal model. First, we run a random effect Poisson regression where the unobserved heterogeneity among individuals is controlled. Second, we run a random effect negative binomial regression where the unobserved heterogeneity in dispersion is allowed. Third, we run a negative binomial regression controlling for the unobserved heterogeneity among individuals (Allison, 2005). Last, we run a population average model (marginal model), GEE (Generalized Estimating Equation) with negative binomial distribution and log link function, where error clusters within individuals are allowed. All regression results are provided in Table 10. Parameter estimates, especially concerning the mileage variable, are almost identical in all models. Test statistics show that the negative binomial model is superior to the Poisson model, and that the random effect Poisson model fits the data better than Poisson regression. However, all of these results

TABLE 10

ROBUSTNESS CHECKS: USING VARIOUS ESTIMATION MODELS.

| Variables | Poisson | Poisson Random Effect | Negative Binomial | NB Random Effect | GEE (Log Link, NB) | NB Random Effect NLMIXED |
|---|---|---|---|---|---|---|
| Mileage | 0.0140*** | 0.0143*** | 0.0141*** | 0.0140*** | 0.0140*** | 0.01437*** |
| | [0.0005] | [0.0006] | [0.0006] | [0.0006] | [0.0005] | [0.0006] |
| Mileage^2 | −0.0001*** | −0.0001*** | −0.0001*** | −0.0001*** | −0.0001*** | −0.0001*** |
| | [0.0001] | [0.0001] | [0.0001] | [0.0000] | [0.0001] | [0.0001] |
| Age 30–60 | −0.0713** | −0.0985** | −0.0678* | −0.0738** | −0.0754** | −0.1634*** |
| | [0.0312] | [0.0378] | [0.0395] | [0.0352] | [0.0326] | [0.0399] |
| Age 60+ | 0.0247 | −0.0042 | 0.0296 | −0.0141 | 0.0217 | −0.1930** |
| | [0.0543] | [0.0653] | [0.0668] | [0.0621] | [0.0565] | [0.0681] |
| Female | 0.1360*** | 0.1422*** | 0.1354*** | 0.1387*** | 0.1367*** | 0.04105* |
| | [0.0194] | [0.0243] | [0.0236] | [0.0219] | [0.0203] | [0.0237] |
| Bonus-Malus | 0.5339*** | −0.9246*** | 0.5095*** | 0.4939*** | 0.3179** | 0.07308* |
| | [0.1031] | [0.1036] | [0.1258] | [0.1174] | [0.1082] | [0.0402] |
| Married | −0.0595* | −0.0644* | −0.0591 | −0.0719** | −0.0597* | −1.1768*** |
| | [0.0315] | [0.0384] | [0.0392] | [0.0354] | [0.0328] | [0.1437] |
| Car Age 0–1 | 0.3820*** | 0.6806*** | 0.3869*** | 0.3500*** | 0.4302*** | 0.4544*** |
| | [0.0443] | [0.0484] | [0.0533] | [0.0503] | [0.0459] | [0.0548] |
| Car Age 1–2 | 0.1383*** | 0.2061*** | 0.1371*** | 0.1366*** | 0.1505*** | −0.1296*** |
| | [0.0389] | [0.0436] | [0.0463] | [0.0439] | [0.0401] | [0.04518] |
| Car Age 2–3 | 0.0432*** | 0.0540*** | 0.0421 | 0.0215 | 0.0459 | −0.2301*** |
| | [0.0401] | [0.0442] | [0.0475] | [0.0453] | [0.0412] | [0.0462] |
| Car Age 3–4 | 0.0617 | 0.0305 | 0.0644 | 0.0483 | 0.0584 | −0.2771*** |
| | [0.0425] | [0.0459] | [0.0503] | [0.0480] | [0.0435] | [0.0494] |
| Car Age 4+ | −0.0057*** | −0.0309 | −0.0063 | −0.0108 | −0.009 | −0.4187*** |
| | [0.0473] | [0.0497] | [0.0556] | [0.0533] | [0.0482] | [0.0556] |
| Eng. Capacity 2 | −0.1818*** | −0.1957*** | −0.1774*** | −0.1647*** | −0.1833*** | −0.1932*** |

TABLE 10

*Contd.*

| Variables | Poisson | Poisson Random Effect | Negative Binomial | NB Random Effect | GEE (Log Link, NB) | NB Random Effect NLMIXED |
|---|---|---|---|---|---|---|
| | [0.0197] | [0.0244] | [0.0240] | [0.0222] | [0.0206] | [0.0243] |
| Eng. Capacity 3 | −0.2025*** | −0.2306*** | −0.1968*** | −0.1559*** | −0.2060*** | −0.2406*** |
| | [0.0417] | [0.0517] | [0.0500] | [0.0458] | [0.0436] | [0.0502] |
| City | 0.0359** | 0.0295 | 0.0353 | 0.0536*** | 0.0353* | 0.04859** |
| | [0.0178] | [0.0221] | [0.0218] | [0.0201] | [0.0186] | [0.0221] |
| North | −0.1589*** | −0.1476*** | −0.1713*** | −0.1463*** | −0.1586*** | −0.4326*** |
| | [0.0386] | [0.0483] | [0.0481] | [0.0436] | [0.0404] | [0.0475] |
| South | −0.0547 | −0.0368 | −0.0609 | −0.0794* | −0.0534 | −0.3941*** |
| | [0.0385] | [0.0485] | [0.0483] | [0.0437] | [0.0404] | [0.0479] |
| Middle | −0.0123 | −0.0002 | −0.0298 | −0.0321 | −0.0117 | −0.3013*** |
| | [0.0413] | [0.0520] | [0.0519] | [0.0468] | [0.0433] | [0.0513] |
| Wald Chi2 | 2,002 | 1,482 | 1,320 | 1,510 | 1,841 | |
| Log Likelihood | −56,389 | −53,617 | −52,587 | −52,496 | | −52,776 |
| Alpha | | 4.44 | | | | |
| LR Test of Alpha, Chi2 | | 5,543.11 | | | | |
| Alpha | | | 8.49 | | | |
| LR Test of Alpha, Chi2 | | | 7,604.82 | | | |
| Likelihood-ratio Test vs. Pooled, Chi | | | | 0.00 | | |

Note: *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Numbers in parentheses provide standard errors. Unbalanced panel random effect negative binomial regression is used. Year dummy variables are included but the coefficients are not reported here.

show that our mileage result is robust and unlikely to be biased or exaggerated by the error structure.

## 6. Conclusions

In this research, we have used the unique database of a major insurance carrier in Taiwan to investigate whether annual mileage should be introduced as a rating variable in auto third-party liability insurance. Admittedly, several characteristics of Taiwan and its insurance market are quite different from other countries: the extreme traffic density, the low number of cars given the high average wealth level and compulsory insurance that only requires bodily injury coverage with fairly low policy limits. However, our results are so strong that we can confidently extend them to all developed countries. Annual mileage is an extremely powerful predictor of the number of claims at-fault. Its significance, as measured by $z$-score and its associated $p$-value, by far exceeds that of all other variables, including BMS. This conclusion applies independently of all other variables possibly included in rating. Cross-validation results show that a prediction model with the mileage variable alone performs better than models with all current rating variables and all other observable variables.

Insurance companies are facing difficult pricing decisions, as several variables commonly used are challenged by regulators. The EU now forbids the use of gender rating. Territory is being challenged in the U.S. as a substitute for race. Insurers are being pressured to find new variables that predict accidents more accurately and are socially acceptable. Annual mileage seems an ideal candidate, to be introduced whenever feasible. The recent development of telematics devices and their rapid decrease in price should induce carriers to explore ways to minimize the practical problems associated with mileage-based insurance premiums.

The inclusion of annual mileage as a new rating variable should, however, not take place at the expense of BMS. BMS are not a substitute for annual mileage; on the contrary, the information contained in the BMS premium level complements the value of annual mileage. An accurate rating system should therefore include annual mileage and BMS as the two main building blocks, possibly supplemented by the use of other variables like age and territory, where allowed.

## References

Allison, P.D. (2005) *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute.

American Academy of Actuaries. (1980) *Risk Classification: Statement of Principles*. Washington DC.

Avraham, R., Logue, K. and Schwarcz, D. (2013) *Understanding Insurance Anti-Discrimination Laws*. University of Michigan Law School Scholarship Repository, Ann Arbor, MI, Art 52.

AYUSO, M., GUILLEN, M. and PEREZ-MARIN, M. (2014) Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis and Prevention*, **73**, 125–131.

BAILEY, R. and SIMON, L. (1960) Two studies in automobile insurance ratemaking. *ASTIN Bulletin*, **1**(4), 192–217.

BICKEL, J.E. (2007) Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, **4**(2), 49–65.

BLACKMON, G. and ZECKHAUSER, R. (1991) Mispriced equity: Regulated rates for auto insurance in Massachusetts. *American Economic Review*, **81**(2), 65–69.

BOUCHER, J.P. and GUILLEN, M. (2009) A survey on models for panel count data with applications to insurance, RACSAM-Revista de la Real Academic de Ciencas Exactas, Fisicos y Naturales. Serie A. *Matematicas*, **103**(2), 277–294.

BOUCHER, J.P. and INOUSSA, R. (2014) A posteriori ratemaking with panel data. *ASTIN Bulletin*, **44**(3), 587–612.

BOUCHER, J.P., DENUIT, M. and GUILLEN, M. (2008) Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distribution. *Variance*, **2**(1), 135–162.

BREUSCH, T.S. and PAGAN, A.R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47**(5), 1287–1294.

BROWN, R., CHARTERS, D., GUNZ, S. and HADDOW, N. (2007) Colliding interest – age as an automobile insurance rating variable: Equitable rate-making or unfair discrimination? *Journal of Business Ethics*, **72**(2), 103–114.

BUTLER, P. (2006) Driver negligence vs. Odometer miles: Rival theories to explain 12 predictors of auto insurance claims, *American Risk & Insurance Association Annual Meeting*, Washington, D.C.

CIA. (2015) The World Factbook. www.cia.gov/library/publications/the-world-factbook/geos/tw.html.

DERRIG, R. and TENNYSON, S. (2011) The impact of rate regulation on claims: Evidence from Massachusetts automobile insurance. *Risk Management and Insurance Review*, **14**(2), 173–199.

FERREIRA, J. and MINIKEL, E. (2010) Pay-As-You-Drive auto insurance in Massachusetts. *Conservation Law Foundation and Environmental Insurance Agency*. http://www.clf.org/wp-content/uploads/2010/12/CLF-PAYD-Study_November-2010.pdf.

FERREIRA, J. and MINIKEL, E. (2013) Measuring per mile risk for Pay-As-You-Drive auto insurance. *Transportation Research Record: Journal of the Transportation Research Board*, **2297**(10), 97–103.

GAULDING, J. (1995) Race, sex, and genetic discrimination in insurance: What's Fair? *Cornell Law Review*, **80**, 1646–1694.

GEISSER, S. (1993) *Predictive Inference*. New York, NY: Chapman and Hall.

GOURIEROUX, C and JASIAK J. (2004) Heterogeneous INAR (1) model with application to car insurance. Insurance: Mathematics and Economics, **34**(2), 177–192.

GREEN W.H. (1994) Accounting for excess zeroes and sample selection in Poisson and negative binomial regression models. *Department of Economics, Stern school of Business, New York University*. https://archive.nyu.edu/bitstream/2451/26263/2/94-10.pdf.

HARRINGTON, S. (1991) Auto insurance in Michigan: Regulation, no-fault, and affordability. *Journal of Insurance Regulation*, **91**(10), 144–183.

HAUSMAN, J., HALL, B. and GRILICHES, Z. (1984) Econometric models for count data with application to the patents-R and D relationship. *Econometrica*, **52**(4), 909–938.

JAFFE, D. and RUSSELL, T. (2001) The regulation of automobile insurance in California. In *Deregulating Property-Liability Insurance: Restoring Competition and Increasing Market Efficiency* (ed. D. Cummins), Washington, DC: American Enterprise Institute – Brookings Institution Joint Center for Regulatory Studies, 195–236.

JANKE, M. (1991) Accidents, mileage, and the exaggeration of risk. *Accident Analysis and Prevention*, **23**(2-3), 183–188.

JOVANIS, P. and CHANG, H. (1986) Modeling the relationship of accidents to miles traveled. *Transportation Research Record*, **1068**, 42–51.

KELLY, M. and NIELSON, N. (2006) Age as a variable in insurance pricing and risk classification. *The Geneva Papers: Issues and Practice*, **31**(2), 212–232.

KOHAVI, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *The International Joint Conference on Artificial Intelligence*, **14**(2), 1137–1145.

LANGFORD, J., KOPPEL, S. MCCARTHY, D. and SRINIVASAN, S. (2008) In defence of the "low-mileage bias". *Accident Analysis and Prevention*, **40**(6), 1996–1999.

LEMAIRE, J. (1985) *Automobile Insurance: Actuarial Models*. Boston: Kluwer Nijhoff Publishing.

LEMAIRE, J. and ZI, H. (1994) A comparative analysis of 30 bonus-malus systems. *ASTIN Bulletin*, **24**(2), 287–309.

LITMAN, T. (2011) Distance-based vehicle insurance feasibility. *Costs and Benefits*. Victoria Transport Policy Institute. www.vtpi.org/dbvi_com.pdf.

LOURENS, P., VISSERS, J. and JESSERUN, M. (1999) Annual mileage, driving violations, and accident involvement in relation to drivers' sex, age, and level of education. *Accident Analysis and Prevention*, **31**(5), 593–597.

PAEFGEN, J., STAAKE, T. and THIESSE, F. (2013) Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach. *Decision Support Systems*, **56**, 192–201.

PAEFGEN, J., STAAKE, T. and THIESSE, F. (2014) Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data. *Transportation Research, Part A: Policy and Practice*, **61**, 27–40.

PROGRESSIVE INSURANCE. (2005) *Texas Mileage Study: Relationship between Annual Mileage and Insurance Losses*. http://www.nctcog.org/trans/air/programs/payd/PhaseI.pdf.

REGAN, L., WEISS, M. and TENNYSON, S. (2008) The relationship between auto insurance rate regulation and insured loss costs: An empirical analysis. *Journal of Insurance Regulation*, **27**(2), 23–46.

SASS, J. and SIEFRIED, F. (2012) Insurance markets and unisex tariffs: Is the European court of justice improving or destroying welfare? *Scandinavian Actuarial Journal*, online, http://dx.doi.org/10.1080/03461238.2012.683450.

SCHWARZE, R. and WEIN, T. (2005) Is the market classification of risk always efficient? Evidence from German third party motor insurance. *ESRC Centre for Analysis of Risk and Regulation, London School of Economics and Political Science*, Discussion Paper 32.

STAPLIN, L., GISH, K. and JOYCE, J. (2008) "Low mileage bias" and related policy implications – A cautionary note, *Accident Analysis and Prevention*, **40**(3), 1249–1252.

TAIWAN INSURANCE INSTITUTE. (2014) Insurance statistics. www.tii.org.tw/econtent/statistics/statistics01.asp.

TAIWAN INSURANCE INSTITUTE. (2015) Insurance laws and regulations database. http://law.tii.org.tw/Eng/new01.asp?LSID=FL006900&ldate=20131227.

WEISS, M., TENNYSON, S. and REGAN, L. (2010) The effects of automobile insurance rate regulation on loss costs and claim frequency: An empirical analysis. *Journal of Risk and Insurance*, **77**(3), 597–624.

JEAN LEMAIRE
*Wharton School, University of Pennsylvania,*
*459 JMHH, 3730 Walnut Street, Philadelphia,*
*PA 19104-6302, USA*
*E-Mail: lemaire@wharton.upenn.edu*
*Phone: +1-215-898-7765. Fax: +1-215-898-1280*

SOJUNG CAROL PARK (Corresponding author)
*College of Business Administration,*
*Seoul National University, Republic of Korea*
*E-Mail: park.sojung@gmail.com*

KILI C. WANG
*Research Fellow, Risk and Insurance ResearchCenter,*
*College of Commerce, National Chengchi University,*
*Tamkang University, Taiwan*
*E-Mail: kili@mail.tku.edu.tw*