Be Cautious in the Last Month: The Sunk Cost Fallacy Held by Car Insurance Policyholders*

Chao Ma[1]

Wang Yanan Institute for Studies in Economics & Department of Finance at School of Economics, Xiamen University, China.

Abstract

Investigating a unique large dataset, we find that automobile insurance policyholders are more likely to encounter accidents during the last month of the insurance policy term than during any other month. Our interpretation is that this effect is driven by the sunk cost fallacy held by policyholders, which exacerbates their moral hazard. The explanation is that in the last month, policyholders may become concerned that they may "waste" the premiums paid upfront if they have not encountered an accident before the policy expires; thus, they will reduce their accident-prevention efforts, although the premiums are sunk costs and cannot be reversed.

Keywords: sunk cost fallacy; moral hazard; mental accounting; salience; automobile insurance; car accidents; irrational behavior; behavioral bias

JEL codes: D9; L8; G4

Running Head: Be Cautious in the Last Month

---

* Manuscript received August 2019; revised June 2020.

[1] I thank the editor, an anonymous associate editor, two anonymous referees, Teck-Hua Ho, Kohei Kawaguchi, and Hang Wu for their helpful comments and suggestions. All errors are my own.

Please address correspondence to: Chao Ma, Wang Yanan Institute for Studies in Economics & Department of Finance at School of Economics, Xiamen University, Xiamen, China 361005. Email: ma.271@buckeyemail.osu.edu

1. Introduction

In traditional economic models, agents are rational and would not take into account sunk costs when making decisions because sunk costs have already been incurred and cannot be reversed. However, several empirical and experimental studies have noted that sunk costs do affect decisions by individuals and firms, which is referred to as the "sunk cost fallacy." These studies found that the sunk cost fallacy causes individuals or firms to make decisions that are not the first best for themselves, including Ho et al. (2018) on automobile purchases, Just and Wansink (2011) on restaurant consumption, Augenblick (2016) on auction bidding, Agarwal et al. (2015) on strategic defaults by mortgage borrowers, and Buchheit and Feltovich (2011) on price setting by duopolists.

In contrast, we examine the automobile insurance market and find that the sunk cost fallacy not only can cause individuals to make nonoptimal decisions for themselves but also can exacerbate the moral hazard problem. It can cause more property losses, bodily injuries, and even deaths, not only to the policyholders but also to other people, and it can incur more monetary losses to insurance companies. Using proprietary data from an automobile insurance company in China, we find that policyholders are more likely to encounter accidents during the last month before their one-year insurance contracts expire than at other times. This phenomenon is not due to the calendar-month effects because different policy contracts can start in different calendar months.

In China's automobile insurance market, the term of a policy contract is one year, which is a mandatory requirement by the regulator. Policyholders need to pay the premium at the beginning of the term; and then they need to pay another premium to renew the contract at the end of the current term if they would like to continue to drive the car. Regardless of whether the policyholder encounters an accident during the contract term, the premium paid at the beginning of the term cannot be refunded. Therefore, the premium paid at the beginning is a sunk cost that should not affect the policyholder's loss-prevention effort or moral hazard during the contract term if the policyholder is completely rational.

However, we find that policyholders for individual-owned vehicles are more likely to encounter accidents during the last month before the one-year insurance contracts expire than at other times. The explanation is that, due to the sunk cost fallacy, as the contract approaches the expiration date, policyholders may become concerned that they may "waste" the premium paid at the beginning of the policy term if they have not encountered an accident before the policy expires, and thus they will reduce their efforts to prevent accidents or mitigate accident hazards.

We also find that the last-month effect does not exist for vehicles owned by companies or governments. The reason is that the premiums for these vehicles are not paid by the drivers but by the companies or governments in which the drivers work. These drivers should not be susceptible to the sunk cost fallacy because, if encounter an accident, they actually "get the premium money's worth" for their employers rather than for themselves.

A rational individual policyholder without the sunk cost fallacy should not exert less loss-prevention or hazard-mitigation effort in the last month than in other months of a policy term. For policyholders, the benefits of reducing their loss-prevention or hazard-mitigation efforts include the time saved by driving at a higher speed, the gains from driving more miles or making more trips per day, the convenience of investing fewer efforts to avoid riskier routes or riskier activities (e.g., driving at night or chatting on cellphones or with other passengers), the novelty of driving on new (unfamiliar) routes, the relaxation of paying less attention while driving, and the enjoyment of "cool" but risky driving habits (e.g., rapid acceleration, hard braking, hard cornering, and rapid lane changes). The costs for policyholders to reduce their efforts are increases in the probability of an accident.[2] However, neither the benefits nor the costs of reducing the loss-prevention or hazard-mitigation efforts vary across different months within the policy term; thus, different months within the policy term should not differently affect policyholders' decisions to reduce the efforts if they are completely rational.

The phenomenon can also be explained within the framework of the mental accounting theory developed by Thaler (1985). In that theory, consumers take into account both acquisition utility and transaction utility when they make purchasing decisions: acquisition utility "depends on the value of the good," while transaction utility "depends solely on the perceived merits of the deal." In the policyholder's problem, because the total payment that the policyholder receives from the insurance company compared to the premium that she/he paid upfront may affect her/his perception of whether she/he received a good deal when purchasing the policy, of the fairness of the premium the company charged her/him, or of whether she/he "got the premium's worth," the policyholder may place the payment received from the insurance company into a different mental account from her/his out-of-pocket loss caused by accidents and use the sunk premium as a reference point.

We also demonstrate that the elevated accident risk in the last month is not fully driven by

---

[2] In the automobile insurance literature, Geyer et al. (2020) documented that the overall distance driven, the number of trips, and the average speed can increase the accident risk. Fan and Wang (2017) documented that rapid acceleration, hard braking, hard cornering, rapid lane changes, and when and where the vehicle is driven have significant influences on accident risk. Jin et al. (2018) found that drivers' route familiarity can reduce their accident risk.

the fraudulent-claims channel, selective reporting, selectivity caused by early termination, and measurement errors.

One may ask why policyholders would drive more dangerously to somehow recoup their sunk premiums. First, if policyholders only drive farther distances, make more trips, or drive more on new routes in the last month of the policy term, they may not think that they are driving more dangerously, but in fact, their daily accident probability is increased.

Second, if policyholders perform more frequent rapid acceleration or rapid lane changes in the last month, they may know that they are driving more dangerously, but the accident probability is still very small, and thus its increase is not salient to them (although the loss given an accident can be enormous). Irrational policyholders who are prone to taking chances may underestimate the increases in the expected losses. However, as the policies approach the expiration dates, the sunk premiums become salient to policyholders who have the sunk cost fallacy because they will obtain disutility from having "wasted" the premium if no accident occurs. Many experimental studies in economics and finance have suggested that people are relatively insensitive to a small change in the probability of an event when the probability is close to zero, whereas people are relatively sensitive to a small change in the probability of an event when the probability is close to one (e.g., the Allais Paradox in Allais, 1953).[3] In the last month, the probability of incurring loss by an accident to the mental account of acquisition utility is still close to zero, whereas the probability of "wasting the premium" in the mental account of transaction utility is sufficiently close to one if no accident has occurred.

Insurance coverage can cause rational drivers to reduce their efforts to prevent losses, which is a well-known textbook example of moral hazard. The mechanisms adopted by auto

[3] Also see Figure 4 in Kahneman and Tversky (1979).

insurance companies to mitigate moral hazard include deductibles and penalties that raise future premiums after accidents. However, if the sunk cost fallacy exists, the actual moral hazard will be higher than what is estimated based on rational agent models. As a result, the optimal amount of deductibles and penalties needed to mitigate moral hazard can also be underestimated. While in the traditional literature, moral hazard is generated by rational agents (Holmstrom, 1979; Shavell, 1979; Rubinstein and Yaari, 1983), our study shows that moral hazard can be exacerbated by the sunk cost fallacy held by irrational agents.

In the health insurance literature, Baicker et al. (2015) proposed a concept of behavioral hazard as compared to moral hazard. In their theoretical model, in patients' decision rules comparing costs and benefits, there is an error term ε representing patients' misperception of costs and benefits, which may lead to incorrect decisions. Starc and Town (2016) provided empirical evidence on patients' behavioral hazard and insurers' efforts in altering insurance designs to mitigate the behavioral hazard. In contrast, the behavioral bias in our study lies in drivers' objective function with a mental account that suffers from the sunk cost fallacy.

Based on our estimates, 2-3% of accidents of individual policyholders can be attributed to the sunk cost fallacy. The resulting losses can be tremendous because the automobile insurance industry is enormous. According to the China Banking and Insurance Regulatory Commission, during 2018, 448 million automobile insurance policies were sold with a total revenue of RMB 783.4 billion, and RMB 440.3 billion were paid by insurance companies as settlement proceeds. Moreover, the market size is rapidly growing because car ownership is rapidly increasing every year in China (by 8.83% in 2019).

The loss caused by the sunk cost fallacy is borne by three parties. First, insurance companies need to pay more money for settlements. Second, for policyholders, while reducing loss-prevention efforts cannot reverse a premium that has already been paid, it will increase the probability of accidents with property losses and bodily injuries that cannot be fully compensated by insurance proceeds. Third, these undue accidents can cause the other party in the collision property losses and bodily injuries that cannot be fully compensated by insurance proceeds.

The results provide two important implications to mitigate the losses caused by the sunk cost fallacy: one for policyholders and the other for insurance companies. First, policyholders should consciously make more loss-prevention or hazard-mitigation efforts in the last month of the policy term because the sunk cost fallacy can cause them to make fewer efforts in that month, which is not a rational decision. Education to improve awareness of the sunk cost fallacy could be helpful. Second, insurance companies could redesign policy contracts to mitigate the additional moral hazard in the last month due to policyholders' sunk cost fallacy, such as increasing the deductible or decreasing the per-accident coverage limit for accidents that occur in the last month of a policy term.

Our study mainly contributes to three strands of the literature. The first strand is the literature on the sunk cost fallacy. Previous research used industrial data, field experiments, or lab experiments to demonstrate the sunk cost fallacy in consumers' or firms' decision-making processes under different scenarios, including automobile purchases (Ho et al., 2018), restaurant consumption (Just and Wansink, 2011), price setting by firms (Buchheit and Feltovich, 2011; Al-Najjar et al., 2008; Shi et al., 2020), auction bidding (Augenblick, 2016), strategic defaults by mortgage borrowers (Agarwal et al., 2015), and computer game playing (Friedman et al., 2007). In those scenarios, the sunk cost fallacy causes consumers or firms to make nonoptimal decisions that are inconsistent with the predictions of rational agent models.[4] Different from those studies,

[4] On the other hand, Baliga and Ely (2011) built a model with limited memory to rationalize the sunk cost

our study provides evidence that, in addition to causing agents to make nonoptimal decisions for themselves, the sunk cost fallacy can exacerbate moral hazard and thereby cause direct losses to other parties.

The second strand of the literature to which we contribute is the literature on moral hazard. Multiple papers have developed different empirical strategies to identify moral hazard in the auto insurance markets (see Weisburd, 2015; Dionne et al., 2013; Abbring et al., 2003).[5] [6] Other papers analyzed the effects of state legislation or public policies on mitigating moral hazard in the auto insurance markets (see Dionne et al., 2011; Hoyt et al., 2006; Cohen and Dehejia, 2004). While in the context of those studies, moral hazard is generated by rational agents, our study shows that moral hazard can be exacerbated by the sunk cost fallacy held by irrational agents.

Third, we contribute to the literature on behavioral bias in insurance-related industries and automobile-related industries. Because the automobile manufacturing industry, the ride-providing industry, the auto insurance industry, and other insurance industries have enormous volumes each year, the behavioral bias in these industries has dramatic impacts.

Behavioral bias in the health insurance industry has been intensively studied. Baicker et al. (2015) built a theoretical model for patients' behavioral hazard as compared to moral hazard, and Starc and Town (2016) provided the empirical evidence. Abaluck et al. (2018) and Dalton et al. (2020) provided empirical evidence on patients' myopic spending behavior under Medicare Part

---

fallacy.

[5] Another group of papers studied moral hazard problems in other automobile-related industries, such as Schneider (2010) on the car-leasing market and Dunham (2003) on the market for used cars.

[6] Another line of the literature studied adverse selection in the auto insurance markets, including Chiappori and Salanie (2000), Puelz and Snow (1994), Cohen (2005), and Dionne et al. (2001).

D. Chang et al. (2018) studied the effect of daily air pollution levels on the demand for long-term health insurance. Sandroni and Squintani (2007) theoretically analyzed the overconfidence of policyholders about their risk levels and the corresponding behavioral rationale of compulsory insurance for a general insurance market.

Behavioral bias in other automobile-related industries has also been studied, but few studies have been conducted on behavioral bias in the automobile insurance industry. Gao et al. (2020) studied policyholders' reference-dependence in exaggerating the reported damage given that an accident occurs. Shum and Xin (2020) studied time-varying risk preferences among automobile drivers and found that drivers drive more conservatively following "near-miss" accidents (measured by hard brakes or hard turns). Busse et al. (2015) studied the psychological effect of weather on car purchases. Ho et al. (2017) studied the effect of the color yellow on taxi accident rates.

The remainder of this paper is organized as follows. In Section 2, we describe the data. In Section 3, we build a theoretical model. Sections 4.1 through 4.3 discuss the baseline empirical results. In Section 4.4, we provide evidence for the reducing-effort channel of the sunk cost fallacy effect, given the potential existence of the insurance-fraud channel. Section 4.5 discusses concerns about selective reporting. Section 4.6 discusses accident severity. In Section 4.7, we conduct a regression discontinuity (RD) design. Sections 5.1 through 5.3 address some concerns regarding the empirical results. In Section 5.4, we provide evidence that the elevated accident risk in the last month cannot be fully explained by rational behavior. In Section 5.5, we employ the salience theory to explain why the elevation in risk only appears in the last month of a policy term. In Section 6, we quantify the proportion of accidents that can be attributed to the sunk cost fallacy. Section 7 discusses the policy implications. Then, we conclude in Section 8.

2. Data

The data come from an auto insurance company in China. The data cover approximately 670,000 policies sold during 2010-2017 in nine major cities. The policy information in the data includes the starting and ending dates, the characteristics of the automobile (e.g., number of seats, manufacturer, model, and manufacture year), the characteristics of the policyholder (e.g., gender, age, and individual owner vs. company or government owner), and the characteristics of the policy contract (e.g., premium and coverage). The settlement information includes the accident date, the type of accident, the report date, the settlement date, and the settlement proceeds paid by the insurance company.

For each policy, the dataset provides the policyholder's previous accident history, which is classified into the following categories: no accident in the previous three or more years, no accident in the previous two years, no accident in the previous year, one accident in the previous year, two accidents in the previous year, three or more accidents in the previous year, and new driver.

After we delete policies with early termination and policies that miss essential information (such as whether the policyholder is an individual, company, or governmental organization), 630983 policies remain. Table 1 reports the descriptive statistics. The histogram of the number of policy cycles for each driver in the data is displayed in Figure C.1 in the online Supplemental Appendix.

[Insert Table 1 here]

Panel A of Figure 1 displays the accident counts in the data for each month within the one-year policy cycle. There is a spike in the last month of the cycle. It can also be seen that accident counts in the first two months are also abnormally high, which is mainly driven by new drivers (first-year drivers). New drivers have higher risk in the first two or three months and an additional month of driving experience is much helpful for them to reduce the risk. Panels B and C of Figure 1 display accident counts for new drivers and experienced drivers (with more than one year of driving experience), respectively. Figure 2 displays the accident counts for each month within the first two policy cycles for drivers that were served by the insurance company for at least two policy cycles. There is a spike in the last month of each policy cycle (the 12th and 24th months).

[Insert Figure 1 here]

[Insert Figure 2 here]

3. Theoretical model

In this section, we build a theoretical model in which a policy contract term has only two days. Although in reality, a policy contract term has 365 days, a simple model with a two-day contract is sufficient to show the following patterns: 1. a policyholder with the sunk cost fallacy exerts lower loss-prevention effort than a policyholder without the sunk cost fallacy; and 2. for a policyholder with the sunk cost fallacy, if no accident occurs on day 1, she/he will exert even lower effort on day 2 than on day 1.

Denote the premium of the policy as $prem$. On each day, a policyholder can either have no accident or encounter one accident. The probability of encountering an accident on a day is $p(e)$, where e is the effort or caution that the policyholder expends on the day to prevent an accident. Assume that $p'(e) < 0$ and $p''(e) > 0$, i.e., the effort will reduce the probability of an accident and the marginal effect of effort is diminishing. The effort level e on a day is determined by the policyholder at the beginning of the day; whether an accident occurs or not will be realized at the end of the day.

Conditional on an accident occurring, the severity level of the accident, s, is a random variable following the probability density function $g(s)$ defined on the support $(0, \bar{s}]$. The higher s is, the greater the damage or loss caused by the accident. Accordingly, the payment from the insurance company to the policyholder, $\phi(s)$, is a function of s; $\phi(s) > 0$ and $\phi'(s) \geq 0$ on the support $(0, \bar{s}]$.[7] Meanwhile, the out-of-pocket loss to the policyholder (after the payment from the insurance company), $\omega(s)$, is also a function of s; $\omega(s) > 0$ and $\omega'(s) \geq 0$ on the support $(0, \bar{s}]$. $\omega(s)$ may include not only the accident loss that is not covered by the insurance but also the present value of increases in future premiums caused by the current accident.

If a policyholder does not have the sunk cost fallacy, on day t $(t = 1, 2)$, she/he will choose the effort level e_t to maximize the expected utility on that day as follows:

$$(3.1) \qquad \max_{e_t} \ \hat{u}_t(e_t) = -e_t - p(e_t) \int_0^{\bar{s}} \omega(s) g(s) ds\,, \qquad t = 1, 2.$$

In equation (3.1), the first term on the right-hand side is the disutility from exerting effort, and the second term is the expected out-of-pocket loss. Denote the solution to the problem in (3.1) as $e_1 = e_2 = \hat{e}$.

---

7 We assume that $\phi(s) > 0$ on the support $(0, \bar{s}]$ for simplicity. In reality, if there is a deductible, for a small s, $\phi(s)$ could be zero. However, the results of the model are extendible to the situation in which there is a deductible.

If a policyholder has the sunk cost fallacy, on day 2, given that the payment she/he received from the insurance company on day 1 is equal to ϕ_1, she/he will solve the following utility-maximizing problem w.r.t. the effort level e_2:

$$\begin{aligned} v_2(\phi_1) = \max_{e_2} u_2(e_2|\phi_1) = -e_2 - p(e_2) \int_0^{\bar{s}} \omega(s) g(s) ds \\ - p(e_2)\lambda \int_0^{\bar{s}} \max\{prem - \phi_1 - \phi(s), 0\}\, g(s) ds \\ - [1 - p(e_2)]\lambda \max\{prem - \phi_1, 0\}. \end{aligned} \tag{3.2}$$

The third and fourth terms on the right-hand side in equation (3.2) indicate that the policyholder would obtain disutility if the total payment received from the insurance company during the entire policy term is less than the premium. If an accident occurs on day 2, the policyholder will obtain disutility $\lambda \max\{prem - \phi_1 - \phi(s), 0\}$, where $\phi(s)$ is the payment received from the insurance company on day 2 and s is a realization of the severity level of the accident on day 2; if no accident occurs on day 2, the policyholder will obtain disutility $\lambda \max\{prem - \phi_1, 0\}$. λ>0. Although the premium is a sunk cost, it enters the policyholder's utility function and works as a reference point. The policyholder not only cares about the loss-prevention effort and the out-of-pocket loss in an accident, but also cares about whether she/he can finally "get the premium money's worth."

The first two terms and last two terms on the right-hand side of equation (3.2) are similar to the concepts of acquisition utility and transaction utility, respectively, in the mental accounting theory developed by Thaler (1985). In that theory, acquisition utility "depends on the value of the good received," while transaction utility "depends solely on the perceived merits of the deal";

and consumers take into account both acquisition utility and transaction utility when they make purchasing decisions. In the driver's problem, the total payment that she/he receives from the insurance company during the entire policy term compared to the premium that she/he paid upfront can affect her/his perception of whether she/he received a good deal when purchasing the insurance, of the fairness of the premium the company charged her/him, or of whether she/he "got the premium's worth." λ in equation (3.2) measures the importance of one dollar obtained in a driver's transaction utility account relative to one dollar obtained in her/his acquisition utility account. λ can also be interpreted as the extent to which the driver is affected by the sunk cost fallacy.

Denote the optimal solution of the effort level to the problem in (3.2) given ϕ_1 as $e_2^*(\phi_1)$, which is a function of ϕ_1. $v_2(\phi_1)$ in (3.2) is the indirect utility function, which is also a function of ϕ_1. On day 1, the policyholder will choose the effort level on day 1 to maximize the total expected utility on day 1 and day 2:

$$\max_{e_1} V_1(e_1) = -e_1 - p(e_1)\int_0^{\bar{s}} \omega(s)g(s)ds + p(e_1)\int_0^{\bar{s}} v_2\big(\phi(s)\big)g(s)ds + [1-p(e_1)]v_2(0). \tag{3.3}$$

The first two terms and last two terms on the right-hand side of equation (3.3) are the expected utilities the policyholder can obtain on day 1 and day 2, respectively, given her/his effort on day 1. Denote the optimal solution to (3.3) as e_1^*.

We have the following two propositions.

Proposition 1: $e_1^* < \hat{e}$ and $e_2^*(\phi_1) \leq \hat{e}$, i.e., the loss-prevention effort on day 1 and day 2 when the sunk cost fallacy exists is lower than the effort on day 1 and day 2, respectively, when the sunk cost fallacy does not exist.

Proposition 2: $e_1^* > e_2^*(0)$, i.e., if no accident occurs on day 1, a policyholder with the sunk cost fallacy will reduce her/his loss-prevention effort on day 2 compared to her/his effort on day 1.

The intuition of Proposition 2 is that, on day 1, the policyholder has a two-day opportunity to "get the premium money's worth," whereas on day 2, if no accident occurred on day 1, the policyholder only has a one-day opportunity to "get the premium money's worth." Therefore, the driver does not need to reduce her/his loss-prevention effort on day 1 as much as on day 2.

We prove Propositions 1 and 2 in Appendix A.1.

4. Main empirical results

4.1. Baseline results

We estimate policyholder-day-level linear probability models as displayed in equation (4.1). The dependent variable $y_{i,t} = 1$ (rescaled to 10,000 basis points (bps)) if policyholder i had an accident on day t; $y_{i,t} = 0$ otherwise. $T_m(t)$ is a 0-1 dummy variable indicating whether day t is in the mth month of the policy term. $X_{i,t}$ is a rich set of control variables, including the vehicle-driver fixed effect, the calendar-year-month fixed effect, the vehicle owner type (individual owner vs. company or government owner), and the driver's accident history and months of driving experience.[8] $\varepsilon_{i,t}$ is the error term. We exclude the policies with early termination from the sample. We also exclude the policies that had not expired by the end of our sampling period.

$$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}. \tag{4.1}$$

When estimating α_m, $m = 1,2,\ldots,12$, we choose the last month of the contract term to be the omitted month, i.e., we set $\alpha_{12} = 0$; therefore, it is easy to see how different the last month is from any of the other months and whether the difference is statistically significant. The standard errors are clustered by vehicle-driver.

Column 1 of Table 2 displays the results of the pooled regression. After controlling for other factors, the accident probability on a day in the last month of the policy term is higher than that in any of the other months at a significance level of 0.1%. Overall, the daily accident probability for a policyholder in the last month is higher than those in other months by 2 to 4 bps, which is a considerable magnitude relative to the average daily accident probability over the

[8] We do not directly observe policyholders' months of driving experience; thus, we use the months since the start of their first contract with this insurance company as the proxy. Their true months of driving experience should equal the sum of months of driving before their first contract with this insurance company and months since the start of their first contract with this insurance company. The unobserved former part is controlled for by the vehicle-driver fixed effect. The proxy is equal to the true months of driving experience for the policyholders of whom the accident history for their first policies with this insurance company is "new driver." We also restrict the sample to these policyholders and obtain similar results. The results are available upon request.

entire policy term (7.51 bps). According to the data, for new drivers, 12 months of driving experience could reduce their daily accident probability by 3.31 bps (0.2755×12, see column 3 of Table C.1 in the online Supplemental Appendix);[9] the daily accident probability of drivers who had one accident in the last year is 4.43 bps higher than that of drivers with no accidents in the last three years or longer (11.5609-7.1352, see column 1 of Table 8 in Section 5.3).[10] Therefore, the magnitude of the sunk cost fallacy effect is also considerable relative to other effects on accident hazard.

[Insert Table 2 here]

The first diagram of Figure 3 plots the estimates of α_m, $m = 1,2,\dots,12$, in column 1 of Table 2. There is a sharp jump in the last month of the policy term. We also run a second regression,

---

[9] To control for the different nonlinear trends in the effects of additional months of driving experience on new drivers (first-year drivers) and on experienced drivers (with more than one year of driving experience) suggested by Figure 1 based on raw data, in Tables 2 and 3, we use fifth-degree polynomials for the two groups of drivers, respectively. In the online Appendix, Tables C.1 and C.2 are other versions of Tables 2 and 3 in which only linear terms of driving experience are controlled for. They clearly show that additional months of driving experience are helpful for new drivers in reducing the risk but are not helpful for experienced drivers. The coefficients for experienced drivers could even be positive at a 5% level of statistical significance, while very small in magnitude.

[10] In Table 8, we analyze how drivers' risks in the current policy cycle are correlated with their past accident histories without controlling for driver-vehicle fixed effects. If we add driver-vehicle fixed effects, then the variation in accident histories is within a driver, and there is no empirical pattern that, within a driver, a worse accident history will lead to higher accident probabilities in the current policy term.

(4.2)
$$y_{i,t} = \sum_{k=1}^{52} \theta_k T_k(t) + \beta X_{i,t} + \varepsilon_{i,t},$$

replacing the month effects α_m in equation (4.1) with the week effects θ_k, which are plotted in the second diagram of Figure 3. A third regression is

(4.3)
$$y_{i,t} = \sum_{d=1}^{365} \gamma_d T_d(t) + \beta X_{i,t} + \varepsilon_{i,t},$$

replacing the month effects α_m in equation (4.1) with the day effects γ_d, which are plotted in the third diagram of Figure 3. In equation (4.2), we choose the last week of a policy cycle as the omitted week, i.e., set $\theta_{52} = 0$; in equation (4.3), we choose the last day of a policy cycle as the omitted day, i.e., set $\gamma_{365} = 0$. The two diagrams show that, after controlling for other factors, the accident intensities start to increase slowly but abnormally around the tenth or the eleventh month of a policy term and start to increase rapidly at approximately the beginning of the last month of a policy term.

[Insert Figure 3 here]

4.2. Individual-owned vs. company-/government-owned vehicles

Column 2 of Table 2 displays the regression results for vehicles owned by companies or governments. In this subsample, the accident probabilities are determined by the drivers' loss-prevention efforts, but the premiums are paid by the companies or governments in which the drivers work rather than by the drivers themselves. These drivers should not be susceptible to the sunk cost fallacy because, if encounter an accident, they actually "get back" the sunk premium for their employers rather than for themselves. As shown in column 2, the accident probability on a day in the last month of the policy term is not significantly higher than that on a day in the previous months.

In contrast, column 3 of Table 2 displays the regression results for vehicles owned by individuals. In this subsample, the premiums are paid by the drivers themselves; therefore, they are susceptible to the sunk cost fallacy. As shown in column 3, the accident probability on a day in the last month of the policy term is significantly higher than that in other months.

Figure 4 displays the month effects in equation (4.1), the week effects in equation (4.2), and the day effects in equation (4.3) separately for government- or company-owned vehicles and individual-owned vehicles. The effect of the sunk cost fallacy exists for the latter group but not for the former group.

[Insert Figure 4 here]

The elevation of accident intensities in the last month of the policy cycle should not be driven by the calendar month effect because, although each policy has a one-year term, it can start in any calendar month. Figure C.2 in the online Appendix displays the distribution of policies with different starting calendar months. There are some variations across different calendar months, but they are within the normal range. In equation (4.1), since we already control for the vehicle-driver fixed effect, we do not need to control for the fixed effect for the calendar month in which the policy starts. We also divide the policies purchased by individual vehicle owners into 12 groups by the calendar month in which the policy starts and run the regression separately for each group. The results are robust and are reported in Table C.7 in the online Appendix.

4.3. Exogenous accidents

In column 4 of Table 2, we restrict the focus to exogenous accidents, including natural disasters, explosions, fires, and thefts. Drivers' accident-prevention efforts cannot influence the occurrence of these exogenous accidents, and thus the probability of such accidents should not be elevated in the last month by the sunk cost fallacy. The sample includes individual policyholders with coverage for these exogenous accidents. The dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an exogenous accident on day t; $y_{i,t} = 0$ otherwise. As shown in column 4 of Table 2, the daily probability of an exogenous accident in the last month of the policy term is not significantly higher than that in other months.

In column 5 of Table 2, the sample is the same as in column 4, but the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had a non-exogenous accident on day t and $y_{i,t} = 0$ otherwise. The results indicate that the daily probability of a non-exogenous accident is significantly higher in the last month than in other months, which is different from the pattern for exogenous accidents.

4.4. Less loss-prevention effort vs. more insurance fraud

The phenomenon of the elevated accident risk in the last month of a policy term can be explained by the sunk cost fallacy. As the contract approaches the expiration date, policyholders could start to be concerned that they may "waste" the premium paid at the beginning of the policy term if they have not encountered an accident before the policy expires, and thus they will reduce their efforts to prevent accidents. A rational policyholder without the sunk cost fallacy should not expend less effort in the last month than in other months of a policy term. The reason is that neither the benefits nor the costs of reducing loss-prevention or hazard-mitigation effort vary across different months within the policy contract term.

However, the elevated risk in the last month of a policy term can also be driven by more fraudulent claims. Policyholders might attempt to replace some auto parts damaged previously with new parts for free or at lower costs by fraudulently reporting an accident. In this subsection, we provide empirical evidence for the existence of the reducing-effort channel for the sunk cost fallacy effect, given the potential existence of the insurance-fraud channel.

We restrict the focus to accidents that involve collisions. Collisions have records in police offices, and proof from accident scenes and police reports are required to claim benefits from insurance companies. Therefore, it is difficult for policyholders to fraudulently report a collision to replace auto parts. Furthermore, policyholders would not make a real collision in order to replace auto parts because both the cost and the risk are too high. First, collisions may cause damages not only to the policyholders but also to the other party involved in the collisions. Second, policyholders may incur more damages than they wish to their vehicles because the level of damages is difficult to control in a collision; and some auto parts that can be damaged in a collision are, in fact, not replaceable. In column 1 of Table 3, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident that involved a collision on day t; $y_{i,t} = 0$ otherwise. The sample includes only policies for individual-owned vehicles. The daily probability of an accident involving collisions in the last month of the policy term remains significantly higher than that in other months.

[Insert Table 3 here]

We further restrict the focus to accidents with bodily injuries. In this case, in addition to proof from the accident scenes and police reports, hospital documentation is required to claim benefits from insurance companies. Therefore, it is even more difficult for policyholders to fraudulently report an accident with bodily injuries to earn insurance proceeds. Policyholders would not make a real accident that is sufficiently severe to cause bodily injuries in order to replace auto parts because both the cost and the risk are extremely high. In column 2 of Table 3, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident with bodily injuries on day t; $y_{i,t} = 0$ otherwise. The daily probability of an accident with bodily injuries in the last month of the policy term remains significantly higher than that in other months. The reason why the magnitude of the estimated monthly effects in this regression is much smaller than that in the regression for all the accidents is that the average daily probability of an accident with bodily injuries (0.2111 bps in column 2 of Table 3) is lower than the average daily accident probability overall. The ratio of the last month effect ($\alpha_{12} - \sum_{m=1}^{11} \alpha_m / 11$) to the average daily accident probability for accidents with bodily injuries is still approximately 23%.

One may argue that a policyholder and another party might collude to create a collision to replace auto parts or that a policyholder, another party in the collision, and a hospital might collude to file a fraudulent bodily injury claim to earn money. However, nowadays in China, monitoring cameras are widely and intensively installed along roads and streets; thus, most fraudulent collisions can be detected by investigations. According to the Insurance Law of the People's Republic of China, fraudulent claims of more than RMB 10,000 are criminal activities; if detected, the penalty includes not only fines but also imprisonment. Therefore, while some policyholders might file fraudulent claims for small monetary amounts to replace some auto parts, it is very uncommon for people to file fraudulent claims for large monetary amounts.

Accordingly, we restrict the focus to accidents with a settlement higher than RMB 10,000. In column 3 of Table 3, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t with a settlement amount greater than RMB 10,000; $y_{i,t} = 0$ otherwise. The sample includes only policies for individual-owned vehicles. The daily probability of an accident with a settlement above RMB 10,000 in the last month of the policy term remains significantly higher than that in other months.[11]

In column 4 of Table 3, we only include the policies of which the policyholders had no accidents in the previous three or more years. These policyholders are less likely to file fraudulent claims. The accident probability on a day in the last month of the policy term remains significantly higher than that in other months.

Even if policyholders conduct more fraudulent claims in the last month than in any other month of a policy term, this behavior can also be attributed to the sunk cost fallacy. For policyholders, the benefit of conducting fraud is the potential insurance proceeds; the cost is the possibility of being detected and the corresponding penalty ranging from a fine to imprisonment in addition to the time spent filing claims. However, neither the benefit nor the cost of conducting fraud varies across different months within the policy term; thus, different months within the policy term should not differently affect policyholders' decision to conduct fraud if they are completely rational. Correspondingly, insurance fraud should be distributed evenly across different months in a policy term if policyholders who conduct fraud are completely rational.

We obtain similar results if we use monthly-level regressions in which the dependent variable is the number of accidents that occurred to the policyholder during the month. Tables C.3 and C.4 in the online Appendix are the monthly versions of Tables 2 and 3.

4.5. Concerns about selective reporting

---

[11] We also use RMB 20,000 and RMB 30,000 as the threshold, respectively, and obtain similar results. The results are available upon request.

One possibility is that some drivers encounter accidents during earlier months within the policy term but do not file claims immediately because the damages are small; in the last month, they happen to encounter another accident and then claim all the damages together.

Another possibility is that some drivers do not report accidents in earlier months that resulted in car damages, for the thinking that, if another accident occurs at the end of the contract, they will report it and deal with all the damages from previous accidents. In this way, the car gets fixed but the driver has only one accident on her/his record.

If these two possibilities of selective reporting drive the last-month effect on accident intensity, we should also observe that the accidents that occur in the last month lead to higher payments from the insurance company than do accidents in other months. However, in column 1 of Table 4, using the accident-level data, we run a regression of the payment obtained from the insurance company and find that, conditional on having an accident, the payment obtained is not significantly higher for an accident that occurs in the last month.[12]

[Insert Table 4 here]

First, at the time of an accident in earlier months in a policy term, drivers do not know whether they will encounter another accident within the policy term. Given that accidents are rare events (in the data, only 4.25% of policies have 2 accidents, and only 1.59% of policies have more than 2 accidents), if drivers do not claim the damage immediately at an accident, the much greater likelihood is that they do not have another accident within the policy term and can never recover the damage.

Second, for many damages, technicians can detect when and in what situation the accident

[12] In fact, as shown in column 1 of Table 4, the payment obtained is even significantly lower for an accident that occurs in the last month than in other months (see Section 4.6 for why it is lower).

occurred. Claims for damages that were obviously inflicted several months earlier will not be accepted by insurance companies. Meanwhile, drivers cannot expect to repair past damage to a side door or the trunk through a current rear-end accident that damages the front part, for example. Moreover, drivers cannot predict which part of the car will be damaged by the next accident.

In column 2 of Table 4, we run an unconditional regression of the payment obtained using the driver-day-level data (payment equals zero for days without accidents). The results indicate that the last month has more payments. The reason is that, although conditional on having an accident, accidents in the last month incur smaller damages, insurance companies still lose more money in the last month because more accidents occur in the last month.

In addition, the daily intensity of accidents involving bodily injuries is also elevated in the last month of a policy term (see column 2 of Table 3), but people definitely would not wait until an accident in the last month to deal with all the bodily injuries from previous accidents.

4.6. Accident severity

One common belief is that policyholders would not risk their lives and the lives of others merely to “get back the premiums.” Therefore, the sunk cost fallacy should have a larger effect in making drivers pay less attention to prevent moderate accidents than in making drivers pay less attention to prevent severe accidents. Correspondingly, we should observe a higher proportion of accidents with small damages in the last month than in other months. As shown in columns 1 and 3 of Table 4, conditional on an accident occurring, it incurs lower payments from the insurance company on average and is less likely to cause bodily injuries if it occurs in the last month than if it occurs in other months.[13]

However, this result does not mean that the number of accidents with major damages or even bodily injuries is not increased at all in the last month. First, once an accident occurs, the severity is partially determined by exogenous factors, such as road conditions and other vehicles, which are not under the driver's control. Second, if some policyholders increase the overall distance driven and the number of trips in the last month, they will increase the probabilities of accidents at all severity levels by the same percentage.

The unconditional regression of the driver-day-level indicator on whether an accident with a payment above RMB 10,000 occurs (column 3 of Table 3) indicates that more accidents that incur a payment above RMB 10,000 occur in the last month than in any other month. The unconditional regression of the driver-day-level indicator on whether an accident with bodily injuries occurs (column 2 of Table 3) indicates that more accidents with bodily injuries occur in the last month than in any other month.

[13] One may ask why, in column 1 of Table 4, the coefficients of months 1 and 2 are much lower than those of months 3-11. The gap still exists even after controlling for the premium (which can affect the reported damages based on Gao et al., 2020), as can be seen in column 1 of Table C.6 in the online Appendix. This gap is actually driven by new drivers. We divide the sample into new drivers and experienced drivers and run the regressions separately. The coefficients of months 1 and 2 are higher than those of months 3-11 for new drivers (see column 2 of Table C.6) but not for experienced drivers (see column 3 of Table C.6). A possible explanation is that new drivers in the first two months may be reluctant to drive on highways or at high speeds and hence the damage given an accident tends to be small, although new drivers are more likely to encounter an accident in the first two months, as suggested by Figure 1.

4.7. Regression discontinuity design

If the higher daily accident intensity in the last month of a policy term is caused by the sunk cost fallacy rather than some other time trend, within two adjacent policy cycles of a driver, we should observe a downward discontinuous jump at the beginning of the succeeding policy cycle rather than finding that the daily accident intensity in the succeeding policy cycle starts at the same level as the end of the preceding policy cycle.

Therefore, we conduct an RD design. Let $\tau = 0$ denote for the last day of a policy term, $\tau = -1$ for the day before the last day of the policy term, $\tau = 1$ for the first day of the succeeding policy term for the same driver, $\tau = 2$ for the second day of the succeeding policy term, and so on. Using drivers with at least two policy cycles in the data, we run the following regression within an RD window $[-(\bar{\tau} - 1), \bar{\tau}]$ for $\bar{\tau}$ to be 30, 60, 90, 120, and 180, respectively.

$$y_{w,\tau} = \gamma_0 + \gamma_1 I(\tau \leq 0) + f(\tau) + \beta X_w + \varepsilon_{w,\tau}, \qquad \tau \in [-(\bar{\tau} - 1), \bar{\tau}]. \tag{4.4}$$

$y_{w,\tau}$ is the accident indicator (rescaled to 10,000 bps) on the τth day in vehicle-driver-window w. Each RD window consists of two adjacent policy cycles for a driver. $I(\tau \leq 0)$ is a 0-1 indicator of whether the day is in the preceding policy part of the RD window. $f(\tau)$ is a flexible polynomial of τ, allowing the coefficients of each term in the polynomial to be different between the domain where $\tau \leq 0$ and the domain where $\tau > 0$. In X_w, we control for vehicle-driver-window fixed effects rather than only vehicle-driver fixed effects because we want to ensure that the last 30 days of a driver's first policy cycle is compared with the first 30 days of her/his second policy cycle rather than with the first 30 days of her/his third policy cycle and so on.[14] We also control for calendar-month fixed effects in X_w.

The coefficient of main interest is γ_1. A significantly positive γ_1 indicates that there is a significant downward jump in the accident intensity at the beginning of the succeeding policy cycle. In panel A of Table 5, we set the degree of the polynomial $f(\tau)$ equal to 3 and alter the RD window bandwidth by changing $\bar{\tau}$ from 30 days to 180 days. In panel B of Table 5, we set $\bar{\tau} = 90$ and change the degree of the polynomial $f(\tau)$ from 1 to 5. We find a significantly positive γ_1 across all these specifications.

[Insert Table 5 here]

In Figure 5, after controlling for vehicle-driver-window fixed effects and calendar-month fixed effects, we plot the daily accident intensity (bps) per driver for the last 180 days of the preceding policy cycle and the first 180 days of the succeeding policy cycle. There is a sharp downward jump at the beginning of the succeeding policy cycle. Figure 5 also shows that, if there were no elevation in the last month of a policy cycle, the daily accident probability should have been slowly decreasing over time. The reason is that policyholders are accumulating more experience in driving over time.

[Insert Figure 5 here]

5. Other concerns

5.1. Selection into the last month

One concern is whether there is a selection effect into who remains for the entire contract term, i.e., drivers with lower risk may be more likely to obtain better offers from other insurance

[14] We also run the RD regression controlling for driver-vehicle fixed effects and obtain similar results. The results are available upon request.

companies and then terminate their current contracts earlier, and thus drivers who remain in the last month may have relatively high risk. Therefore, in the regressions of this study, we exclude the policies with early termination from the sample. We also exclude the policies that had not expired by the end of our sampling period.[15] [16]

---

[15] In the raw data, 1.21% of the policies early terminated their mandatory insurance, and 3.13% of the policies early terminated their optional insurance and still kept their mandatory insurance. These policies are all excluded from the regressions. First, mandatory insurance cannot be terminated early except for certain special situations, such as vehicle ownership changes or the vehicle being scrapped; and the prices of mandatory insurance are regulated such that, given the type of the car and the driver, different insurance companies charge the same price as determined by the regulator. Second, although optional insurance allows early termination, it has a high early termination fee (usually approximately 10-20% of the annual price), and policyholders need to undergo a time-consuming administrative process. On the other hand, the prices of optional insurance are regulated such that, given the type of the car and the driver, there exist a suggested price and a narrow floating range provided by the regulator, and each insurance company must set its prices within the floating range around the suggested price (usually -15% to +15%). Meanwhile, a low-risk driver may also be likely to obtain a better offer with the current company for the next term for being at low risk. Consequently, the main reasons for early termination of optional insurance are that the policyholders believe that they no longer need the additional coverage on top of the mandatory insurance (the minimum coverage required by the regulator), that the vehicle ownership has changed, or that the vehicle has been scrapped. Terminating the contract early and switching to another company for a better offer are less likely because the reduction in price by a better offer cannot cover the transaction cost, especially when it is close to the last month of the policy term. If policyholders want to switch providers, they will switch after the current contract expires.

[16] We also conduct regressions with those early terminated policies included in the sample. The results are similar and are available upon request.

5.2. Measurement errors

One concern is that policyholders might report an accident many days later after the accident date. However, the accident date and the report date are accurately recorded as different variables in the insurance company's database. When defining the monthly dummy $T_m(t)$, the weekly dummy $T_w(t)$, and the daily dummy $T_d(t)$ in the regression equations, we use the accident date rather than the report date.

Another concern is whether the recorded accident date could be later than the true accident date. In fact, the accident date is not only self-reported by drivers but also is verified by insurance companies. Drivers are required to call the police and the insurance company within 48 hours after an accident occurs. Thereafter, the insurance company will promptly send investigators to the site to check the involved vehicle or the traces, determine liability, and provide preliminary estimates for losses and repair recommendations. Usually, the insurance company will finish the investigation process within one day after the call. Figure 6 displays the histogram of the number of days from the accident date to the report date.

[Insert Figure 6 here]

If these concerns were true, we should also have observed a significantly positive last-month effect for the regression using the sample of company or government owners (column 2 in Table 2) and for the regression of exogenous accidents (column 4 in Table 2), but we do not observe such effects.

5.3. Do drivers know that an accident could increase their future premiums?

If policyholders know that their future premiums could be increased by an accident in the current policy term, the magnitude of the last-month effect caused by the sunk cost fallacy could be mitigated. In this subsection, we first use the premium data to examine how premiums change according to drivers' accident histories. Then, we provide an estimate for the premium increase in the next policy term caused by an accident in the current term given a driver's accident history. Finally, we analyze how the magnitude of the last-month effect differs across drivers with different accident histories.

To examine how premiums change according to drivers' accident histories, we run the following regression:

$$premium_i = \varphi \cdot accdnt\_hist_i + \beta X_i + \varepsilon_i. \tag{5.1}$$

Drivers' accident histories are categorized into the following levels: no accident for 3 or more years, no accident for 2 years, no accident for 1 year, new driver, 1 accident in the previous year, 2 accidents in the previous year, and 3 or more accidents in the previous year. $accdnt\_hist_i$ in equation (5.1) is a set of dummy variables for the seven categories. X_i is a rich set of control variables, including the driver's age, gender, driving experience, and location, as well as the car age and the fixed effect for the car model. The regression results displayed in Table 6 show that policyholders with a worse accident history need to pay higher premiums.

[Insert Table 6 here]

Based on the estimates in Table 6, for each category of accident history, we calculate the average premium difference in the next term between the case in which the driver encounters one accident in the current term and the case in which the driver does not encounter any accident in the current term, as displayed in the last column of Table 7. For example, if the accident history for the current policy term is "no accident for 2 years," the average next-term premium difference between one accident and no accident in the current term is RMB 706.19 (the premium for the history of "1 accident in the previous year" minus the premium for the history of "no accident for 3 or more years"). The better the accident history for the current term, the greater the increase in the next-term premium caused by an accident in the current term compared with no accident in the current term.

[Insert Table 7 here]

Next, we run a regression of drivers' daily accident indicators on their accident histories, controlling for other driver and vehicle characteristics. The results (reported in column 1 of Table 8) indicate that drivers with worse accident histories have higher probabilities of an accident.[17] In this regression, the coefficient of the last month indicator is significantly positive, which means that drivers have a higher accident probability in the last month of the contract term. In column 2 of Table 8, we add the interaction terms between the last month indicator and the accident history categories. There is a rough pattern that the better the accident history, the smaller the increase in the accident probability in the last month, probably because of a potentially larger future premium increase by a current accident. This pattern indicates that drivers might consider that an accident will increase the next-term premium, but this consideration is not sufficiently strong to fully stall reducing the loss-prevention effort in the last month due to the sunk cost fallacy.

[17] In this regression, we do not control for driver-vehicle fixed effects. If we add driver-vehicle fixed effects, then the variation in accident histories is within a driver, and there is no empirical pattern that, within a driver, a worse accident history will lead to higher accident probabilities in the current policy term.

[Insert Table 8 here]

There could be several reasons why the consideration of premium increases fails to fully stall the effect of the sunk cost fallacy. First, the increase in the next-term premium caused by one accident in the current term is usually small relative to the premium (approximately RMB 3,000 on average).

Second, policyholders can have some ambiguity regarding how much of an increase in the next-term premium can be caused by one accident in the current term. Although drivers should know the general principle that their premiums partially depend on their accident histories, they are unclear about the formula that the insurance company uses to determine their premiums and thus do not know the exact increase in the next-term premium caused by one accident in the current term. It is also difficult for drivers to infer the exact increase based on their own experience. They may experience a premium increase for many reasons, such as changes in driver ages, car ages, regulatory policies, insurance market competition, insurance companies' pricing strategies, and sales commission, as well as inflation.[18] It is difficult for drivers to attribute the correct proportion of the premium increase to the change in accident histories.

Third, the possible premium increase may not be sufficiently salient to drivers. Even though drivers reduce their efforts in the last month, the accident probability is still very low because

[18] In China, sales commission constitutes a large proportion of a premium. Even for the same insurance product provided by the same insurance company, the sales commission varies dramatically across different sales channels and agents. The sales channels include car dealerships (4S shops), insurance brokers, insurance companies' own branches, and the online and telephone channels. Sometimes sales agents yield part of their commissions to insurance purchasers as price discounts to promote insurance sales.

accidents are rare events. Many experimental studies in economics and finance have suggested that people are relatively insensitive to a small change in the probability of an event when the probability is close to zero and that people are relatively sensitive to a small change in the probability of an event when the probability is close to one (e.g., the Allais Paradox in Allais, 1953). In the last month, the probability of incurring an increase in the next-term premium by an accident to the mental account of acquisition utility is still close to zero, whereas the probability of "wasting the premium" in the mental account of transaction utility is sufficiently close to one if no accident has occurred.

5.4. Sunk cost fallacy vs. rational behavior

Most types of auto insurance in China only have per-accident coverage limits. For example, mandatory insurance only has coverage limits for each accident, including RMB 110,000 for death or disability, RMB 10,000 for medical payments, and RMB 2,000 for property damage if at fault, and RMB 11,000 for death or disability, RMB 1,000 for medical payments, and RMB 100 for property damage if not at fault. However, some types of auto insurance also have a cap on the total losses within a contract term in addition to per-accident caps: insurance compensating the policyholder for the days when the vehicle is being fixed has a cap on the total number of days compensated within a policy term; collision insurance has an annual coverage limit that is equal to the current value of the insured vehicle (new car price minus depreciation) determined at the beginning of the policy term; and scratch insurance also has an annual coverage limit.

If there is an annual coverage limit for the entire policy term, the elevated accident risk in the last month of the policy term can be generated by rational policyholders. In Appendix A.2, we build a rational agent model that can generate the elevated accident risk in the last month if there is an annual coverage limit. The intuition is that policyholders face uncertainty for a longer period during the early months of a policy term than they do later in the policy term. Given that a policyholder makes a low loss-prevention effort during the early months of a policy term and hence encounters an accident, if the policyholder encounters another accident later in the policy term, the accumulated loss may surpass the annual coverage limit, and the policyholder will have to bear some loss herself/himself. Therefore, a policyholder would make a higher effort during the early months of a policy term; if no accident is encountered, she/he will later reduce the effort level, which leads to a higher accident risk as the expiration date of the policy approaches.

However, if there are only per-accident coverage limits and no annual coverage limits, a fully rational policyholder would not reduce the effort as the policy contract approaches the expiration date.

We first restrict the sample to policyholders whose insurance only has per-accident coverage limits. The regression results of equation (4.1) for this subsample (291,592 policies) are reported in column 1 of Table 9. The accident probability on a day in the last month of the policy term remains significantly higher than that on a day in other months.

[Insert Table 9 here]

If the cumulative settlement proceeds within the policy term have surpassed the premium before the last month of the policy term, the policyholder should be less susceptible to the sunk cost fallacy that can elevate the last-month accident risk. The reason is that policyholders may think that they have already "gotten the premium money back" and do not need to reduce their loss-prevention efforts during the remaining contract term. Therefore, for each policyholder with only per-accident coverage limits, we further restrict the observations to the days in the policy term after the cumulative settlement proceeds within the policy term surpassed the premium; the results are reported in column 2 of Table 9. The accident probability on a day in the last month of the policy term is no longer significantly higher than that on a day in other months.

On the other hand, we restrict the sample to policyholders whose insurance has both per-accident coverage limits and annual coverage limits. As shown in column 3 of Table 9, for this subsample (276,180 policies), the accident probability on a day in the last month of the policy term is significantly higher than that on a day in other months, which is the same as column 1 of Table 9. However, for these policyholders, when we include only the observations for the days in the policy term after the cumulative settlement proceeds within the policy term surpassed the premium, as shown in column 4 of Table 9, the accident probability on a day in the last month of the policy term is significantly higher than that on a day in other months. This result pattern is different from that in column 2 of Table 9. The reason is that, for insurance with annual coverage caps, the elevated accident risk in the last month of a policy term is due to not only the sunk cost fallacy but also the rational behavior modeled in Appendix A.2.[19]

5.5. Why only the last month?

The theoretical model in Section 3 implies that the effort level should decrease gradually every day during the 365-day term and thus the accident probability should increase gradually. The intuition is that, at the beginning of a policy cycle, there are still many opportunities for drivers

[19] In Table 9, the magnitude of the negative coefficients of T^1 through T^{11} in column 4 is larger than that in column 3. The reason is that the regression in column 4 only includes the days after the cumulative settlement proceeds within the policy term surpassed the premium. Therefore, policyholders who have no accident in the policy term are excluded from the sample. If these policyholders were included, they would tend to reduce the magnitude of the negative coefficients of T^1 through T^{11}.

to "get the insurance money's worth" or "break even"; as time passes, the opportunities are diminishing.

However, the empirical results (the last diagram in Figure 4) show that the daily accident intensity increases gradually every day only in the last month of the term; the intensity holds almost constant during the first eleven months of the term (slightly increasing during the 10th and 11th months).

The salience theory may serve as a possible explanation. In the last month, the need to "get the insurance money's worth" or "break even" becomes fully salient to policyholders because they start to realize that there are only a few chances left if they have not yet encountered any accident or the insurance payment that they received has not surpassed the premium. Accordingly, the policyholders' loss-prevention effort levels follow the predictions of equations (3.3) and (3.2) and increase over time. At other times, the sunk cost is not salient to policyholders because there are still many chances to "get the premium back"; thus, they behave as if they were rational agents. Accordingly, their loss-prevention effort levels follow the prediction of equation (3.1) and do not change over time.

Dalton et al. (2020) studied the weekly medical spending of Medicare Part D enrollees and found a similar pattern. Medicare Part D has a nonlinear benefit structure: enrollees face modest out-of-pocket expenditures in the initial coverage region until their accrued total year-to-date drug spending reaches a threshold, after which they will pay the full prices of all drugs. Suppose that an enrollee is currently in the initial coverage region but forecasts that she/he will end the year above the threshold. If she/he is a rational dynamically optimizing enrollee, she/he should choose increasingly cheaper or fewer drugs as she/he approaches the threshold. However, Dalton et al. (2020) found that the weekly medical spending is flat in the initial coverage region and then starts to drop near the threshold, which can be explained by salience theory.

In the situation of our study, a rational policyholder should make a constant loss-prevention effort over the entire policy term; an irrational policyholder affected by the sunk cost fallacy but not the salience effect should gradually decrease her/his effort as she/he approaches the policy expiration date; and an irrational policyholder affected by both the sunk cost fallacy and the salience effect should only start to gradually decrease her/his effort at a time point close to the policy expiration date.

Salience effects were also empirically detected by several other studies. Busse et al. (2015) found that, when consumers purchase a car (a durable good), the choice to purchase a convertible or a four-wheel-drive is highly dependent on the weather at the time of purchase. Chang et al. (2018) found that the air pollution level in a day has a significant effect on the decision to purchase or cancel long-term health insurance. Chetty et al. (2009) found that increases in taxes included in posted prices reduce alcohol consumption more than increases in taxes applied at the register. Pan et al. (2019) studied land sales along two sides of the heating-service line in China. They found that, compared to transactions in the north where heating services are provided, land parcels in the south have a lower price only when the transactions occur in winter because land buyers to the south of the boundary of heating services factor in the disutility from cold winters only when they purchase land in winter. Abaluck et al. (2018) studied consumer salience in medical spending. Bordalo et al. (2012) developed a salience theory of choice under risk.

6. Quantifying the proportion of accidents caused by the sunk cost fallacy

In this section, we quantify the proportion of accidents that can be attributed to the sunk cost fallacy. Following the literature (e.g., Kleven and Waseem, 2013; Chen et al., 2019), we use a flexible polynomial to estimate the counterfactual daily accident intensity in the last three months of a policy term without the sunk cost fallacy for individual policyholders.

In the first step, we run a driver-day-level regression of the accident indicator on the fixed effects for each day of the policy term and control variables (as displayed in equation (4.3)). Then, we obtain the estimated daily fixed effects. Let c_t denote the estimated fixed effect for the tth day since the policy starts, $t = 1, 2, \dots, 365$.

In the second step, we use the estimated daily fixed effects excluding the days in the last three months (from the 271st day to the 365th day) to fit a flexible polynomial, by running the following regression:

(6.1) $$c_t = \sum_{j=0}^{p} \psi_j t^j + \sum_{\tau=271}^{365} \mu_\tau I(t = \tau) + \varepsilon_t,$$

where p is the degree of the flexible polynomial and μ_τ are daily fixed effects for the excluded range. We choose the last three months to be the excluded range because the daily accident intensity starts to increase slightly in the 10th month, as shown in the last diagram of Figure 4. The counterfactual daily fixed effects in the last three months without the sunk cost fallacy are obtained as the predicted values from (6.1) omitting the contribution of the dummies in the excluded range, i.e., $\hat{c}_t = \sum_{j=0}^{p} \hat{\psi}_j t^j$, $t = 271, 272, \dots, 365$. Then, the proportion of accidents that can be attributed to the sunk cost fallacy is estimated as

(6.2) $$\hat{A} = \frac{\sum_{t=271}^{365}(c_t - \hat{c}_t)}{365 \ \times \ mean \ \ of \ \ daily \ \ accident \ \ probability}.$$

Column 1 of Table 10 reports the quantified proportion of accidents that can be attributed to the sunk cost fallacy for policyholders with only per-accident coverage limits, with the degrees of the polynomial ranging from 0 to 2. The quantified proportion is 2-3%. Column 2 of Table 10 reports $\hat{A}$ for policyholders with annual coverage limits. For each degree of polynomial, $\hat{A}$ in column 2 is higher than $\hat{A}$ in column 1 because, for policyholders with annual coverage limits, the increase in the daily accident intensity in the last three months could be caused not only by the sunk cost fallacy but also by the rational behavior modeled in Appendix A.2.

[Insert Table 10 here]

To determine the degree of the flexible polynomial (p), we fit a flexible polynomial using the estimated daily fixed effects for the first 8 months and then check the out-sample fit of the polynomial for the estimated daily fixed effects in the 9^{th} month. We find that a zero-degree polynomial has a better out-sample fit. Therefore, in Table 10, the proportion that is quantified using a zero-degree polynomial is more reliable.

This 2-3% proportion of accidents could result in tremendous losses because the automobile insurance industry is enormous. According to the China Banking and Insurance Regulatory Commission, during 2018, 448 million automobile insurance policies were sold with total revenue of RMB 783.4 billion, and RMB 440.3 billion were paid by insurance companies as settlement proceeds. Moreover, the market size is rapidly growing because car ownership is rapidly increasing every year in China (by 8.83% in 2019).

7. Policy Implications

Because the sunk cost fallacy can cause drivers to irrationally make fewer loss-prevention or hazard-mitigation efforts in the last month of the policy term, which is nonoptimal for them, drivers should consciously make more efforts in the last month. Education to improve awareness of the sunk cost fallacy could be helpful.

In addition, we propose two types of methods for insurance companies to reduce losses caused by policyholders' sunk cost fallacies.

The first type of methods is simply sending a text message or an email to a policyholder in the last month of the policy term if the policyholder has not encountered any accident or the insurance payment has not surpassed the premium during the first eleven months, stating something like the following: "Congratulations! You have successfully maintained an excellent driving record in the previous eleven months. Please continue to maintain the record in the last month and then enjoy a premium discount of $x for your next policy term." The purpose of the message or email is to render the benefit of loss-prevention efforts relatively more salient and to make the sunk costs relatively less salient to policyholders in the last month.[20]

The second type of methods involves redesigning the policy contract, such as increasing the deductible or decreasing the per-accident coverage limit for accidents that occur in the last month of a policy term, to make policyholders expend more loss-prevention efforts when the sunk costs become salient. The mechanism of these methods is to increase policyholders' expected

[20] Field experiments have found that text message reminders are effective in multiple scenarios, including reducing drivers' traffic violations (Chen et al., 2017; Lu et al., 2016), improving borrowers' repayments for online P2P lending (Du et al., 2020; Huang and Bao, 2020), credit cards (Bursztyn et al., 2015), and microloans (Cadena and Antoinette, 2011), enhancing tax compliance (Hallsworth et al., 2014), increasing commitment attainment for banks' clients with commitment savings accounts (Karlan et al., 2016), and enforcing compliance with TV license fees (Fellner et al., 2013).

out-of-pocket loss from a last-month accident in the acquisition utility to catch up with the increasingly salient benefit of a last-month accident in the transaction utility in their mental accounts.

In future research, it would be contributive to conduct field experiments to investigate the effectiveness of the two types of methods proposed in this section to mitigate the effect of the sunk cost fallacy on the accident-risk elevation in the last month.

8. Conclusion

In traditional economic models, agents are rational and would not take into account sunk costs when making decisions because sunk costs have already been incurred and cannot be reversed. However, several empirical and experimental studies have noted that sunk costs do affect decisions made by individuals and firms, which is referred to as the "sunk cost fallacy." These studies found that the sunk cost fallacy causes individuals or firms to make decisions that are not the first best.

In this paper, we examine the auto insurance market and find that the sunk cost fallacy not only can cause agents to make nonoptimal decisions for themselves but also can exacerbate moral hazard and thereby cause direct losses to other parties. Using proprietary data from an auto insurance company, we find that policyholders are more likely to encounter accidents during the last month before the one-year insurance contracts expire than at other times. This phenomenon is not due to the calendar-month effects because different policy contracts can start in different calendar months. We also demonstrate that the elevated accident risk in the last month is not fully driven by rational behavior, the fraudulent-claims channel, selective reporting, selectivity caused by early termination, and measurement errors.

Due to the sunk cost fallacy, as the contract approaches the expiration date, policyholders could start to be concerned that they may “waste” the premium paid at the beginning of the policy term if they have not encountered an accident before the policy expires or the insurance payment received has not surpassed the premium, and thus they will reduce their efforts to prevent accidents or mitigate accident hazards. A rational policyholder without the sunk cost fallacy should not expend less effort in the last month than in other months of a policy term. The reason is that neither the benefits nor the costs of reducing effort vary across different months within the policy contract term. In contrast, an irrational policyholder with the sunk cost fallacy may have a mental account of transaction utility, in which the total settlement proceeds that the policyholder receives compared to the premium that she/he paid upfront may affect her/his perception of whether she/he received a good deal when purchasing the policy, of the fairness of the premium the company charged her/him, or of whether she/he “got the premium’s worth.”

Based on our estimates, 2-3% of accidents of individual policyholders are caused by the sunk cost fallacy. The resulting losses can be tremendous because the automobile insurance industry is enormous. According to the China Banking and Insurance Regulatory Commission, during 2018, 448 million automobile insurance policies were sold with total revenue of RMB 783.4 billion, and RMB 440.3 billion were paid by insurance companies as settlement proceeds.

The results provide two important implications: one for policyholders and the other for insurance companies. First, policyholders should consciously expend more loss-prevention or hazard-mitigation efforts when driving during the last month of the policy term because the sunk cost fallacy can cause them to expend fewer efforts in that month, which is not a rational decision. Education to improve awareness of the sunk cost fallacy could be helpful. Second, insurance companies could send a message to policyholders in the last month of the policy term to render the benefit of maintaining good records more salient to policyholders. Insurance companies could also redesign policy contracts to mitigate the additional moral hazard in the last month due to policyholders' sunk cost fallacy, such as increasing the deductible or decreasing the per-accident coverage limit for accidents that occur in the last month.

Appendix

A.1. Proofs of Propositions 1 and 2 stated in Section 3

Proposition 1: $e_1^*<\hat{e}$ and $e_2^*(\phi_1) \leq \hat{e}$, i.e., the loss-prevention effort on day 1 and day 2 when the sunk cost fallacy exists is lower than the effort on day 1 and day 2, respectively, when the sunk cost fallacy does not exist.

Proof:

To equation (3.1) where a driver does not have the sunk cost fallacy, the solution of effort $\hat{e}$ on day 1 and day 2, respectively, should satisfy the following first-order condition:

$$-p'(\hat{e})\int_0^{\bar{s}} \omega(s)g(s)ds = 1. \tag{A.1}$$

In the situation where a driver has the sunk cost fallacy, the solution to equation (3.2), $e_2^*(\phi_1)$ on day 2 given ϕ_1 received on day 1, should satisfy the following first-order condition:

$$\begin{aligned} &-p'\big(e_2^*(\phi_1)\big)\int_0^{\bar{s}} \omega(s)g(s)ds \\ &= 1 \\ &+ \lambda p'\big(e_2^*(\phi_1)\big)\left(\int_0^{\bar{s}} \max\{prem-\phi_1-\phi(s),0\}\, g(s)ds - \max\{prem-\phi_1,0\}\right). \end{aligned} \tag{A.2}$$

Because $\int_0^{\bar{s}} \max\{prem-\phi_1-\phi(s),0\}\, g(s)ds <$ $\max\{prem-\phi_1,0\}$ and $p'\big(e_2^*(\phi_1)\big) < 0$, we have

(A.3)
$$-p'\left(e_2^*(\phi_1)\right)\int_0^{\bar{s}}\omega(s)g(s)ds > 1.$$

Comparing (A.1) and (A.3), because $p''(.) > 0$, we have $e_2^*(\phi_1) < \hat{e}$.

The solution to equation (3.3), e_1^* on day 1, should satisfy the following first-order condition:

(A.4)
$$-p'(e_1^*)\int_0^{\bar{s}}\omega(s)g(s)ds = 1 + p'(e_1^*)\left(v_2(0) - \int_0^{\bar{s}}v_2(\phi(s))g(s)ds\right).$$

By the envelope theorem, from equation (3.2), we have

$$v_2{}'(\phi_1) = -\lambda p\left(e_2^*(\phi_1)\right)\frac{d\int_0^{\bar{s}}\max\{prem - \phi_1 - \phi(s), 0\}\, g(s)ds}{d\phi_1} - \lambda\left(1 - p\left(e_2^*(\phi_1)\right)\right)\frac{d\max\{prem - \phi_1, 0\}}{d\phi_1}.$$

Then, we have $v_2{}'(\phi_1) \geq 0$, and ">" holds for a certain domain of ϕ_1. Consequently, in (A.4), $v_2(0) - \int_0^{\bar{s}}v_2\left(\phi(s)\right)g(s)ds < 0$. Because $p'(.) < 0$,

(A.5)
$$-p'(e_1^*)\int_0^{\bar{s}}\omega(s)g(s)ds > 1.$$

Comparing (A.1) and (A.5), because $p''(.) > 0$, we have $e_1^* < \hat{e}$.

Proposition 2: e_1^*>$e_2^*(0)$, i.e., if no accident occurs on day 1, a policyholder with the sunk cost fallacy will reduce her/his loss-prevention effort on day 2 compared to her/his effort on day

1.

Proof:

Based on (A.2), if the payment from the insurance company on day 1 is zero (ϕ_1=0), then the effort on day 2, $e_2^*(0)$, should satisfy the following first-order condition:

$$-p'\big(e_2^*(0)\big)\int_0^{\bar{s}}\omega(s)g(s)ds = 1+\lambda p'\big(e_2^*(0)\big)\left(\int_0^{\bar{s}}\max\{prem-\phi(s),0\}\,g(s)ds-prem\right). \quad \text{(A.6)}$$

Comparing (A.4) and (A.6), because $p'(.)<0$ and $p''(.)>0$, to prove that $e_1^*>e_2^*(0)$, we only need to show:

$$v_2(0)-\int_0^{\bar{s}}v_2\big(\phi(s)\big)g(s)ds>\lambda\left(\int_0^{\bar{s}}\max\{prem-\phi(s),0\}\,g(s)ds-prem\right). \quad \text{(A.7)}$$

Based on equation (3.2), when ϕ_1=0,

$$v_2(0)=-e_2^*(0)-p\big(e_2^*(0)\big)\int_0^{\bar{s}}\omega(s)g(s)ds-p\big(e_2^*(0)\big)\lambda\int_0^{\bar{s}}\max\{prem-\phi(s),0\}\,g(s)ds$$
$$-\big[1-p\big(e_2^*(0)\big)\big]\lambda\cdot prem.$$

For $\forall s\in(0,\bar{s}]$, $v_2(0)=u_2(e_2^*(0)|0)>u_2\big(e_2^*\big(\phi(s)\big)|0\big)$. Therefore,

$$v_2(0)=\int_0^{\bar{s}}v_2(0)g(s)ds>\int_0^{\bar{s}}u_2\big(e_2^*\big(\phi(s)\big)|0\big)g(s)ds.$$

$$v_2(0) - \int_0^{\bar{s}} v_2(\phi(s))g(s)ds > \int_0^{\bar{s}} \Big(u_2\big(e_2^*(\phi(s_1))|0\big) - v_2\big(\phi(s_1)\big)\Big)\, g(s_1)ds_1$$

$$= \int_0^{\bar{s}} \Big(u_2\big(e_2^*(\phi(s_1))|0\big) - u_2\left(e_2^*(\phi(s_1))|\phi(s_1)\right)\Big)\, g(s_1)ds_1$$

$$= \lambda \int_0^{\bar{s}} \Bigg[-p\left(e_2^*(\phi(s_1))\right)\int_0^{\bar{s}} \max\{prem - \phi(s_2), 0\}\, g(s_2)ds_2$$

$$- \Big(1 - p\left(e_2^*(\phi(s_1))\right)\Big) prem$$

$$+ p\left(e_2^*(\phi(s_1))\right)\int_0^{\bar{s}} \max\{prem - \phi(s_1) - \phi(s_2), 0\}\, g(s_2)ds_2$$

$$+ \Big(1 - p\left(e_2^*(\phi(s_1))\right)\Big) \max\{prem - \phi(s_1), 0\}\Bigg] g(s_1)ds_1$$

$$= \lambda \int_0^{\bar{s}} \Bigg[-prem$$

$$+ \max\{prem - \phi(s_1), 0\} + p\left(e_2^*(\phi(s_1))\right)(prem - \max\{prem - \phi(s_1), 0\})$$

$$- p\left(e_2^*(\phi(s_1))\right)\int_0^{\bar{s}} \max\{prem - \phi(s_2), 0\}\, g(s_2)ds_2$$

$$+ p\left(e_2^*(\phi(s_1))\right)\int_0^{\bar{s}} \max\{prem - \phi(s_1) - \phi(s_2), 0\}\, g(s_2)ds_2\Bigg] g(s_1)ds_1.$$

Let

$$A(s_1) = \ p\left(e_2^*(\phi(s_1))\right)(prem - \max\{prem - \phi(s_1), 0\})$$

$$- p\left(e_2^*(\phi(s_1))\right)\int_0^{\bar{s}} \max\{prem - \phi(s_2), 0\}\, g(s_2)ds_2$$

$$+ p\left(e_2^*(\phi(s_1))\right)\int_0^{\bar{s}} \max\{prem - \phi(s_1) - \phi(s_2), 0\}\, g(s_2)ds_2.$$

Then,

$$v_2(0) - \int_0^{\bar{s}} v_2(\phi(s))g(s)ds$$

$$> \lambda\left(\int_0^{\bar{s}} \max\{prem - \phi(s), 0\}\, g(s)ds - prem\right) + \lambda \int_0^{\bar{s}} A(s_1)g(s_1)ds_1.$$

When $\phi(s_1) \geq prem$,

$$A(s_1) = p\left(e_2^*(\phi(s_1))\right)\left(prem - \int_0^{\bar{s}} \max\{prem - \phi(s_2), 0\}\, g(s_2)ds_2\right) > 0.$$

When $\phi(s_1) < prem$, given s_1, divide the domain of s_2 into three segments, $(0, c_1]$, $(c_1, c_2]$, and $(c_2, \bar{s}]$, such that $\phi(c_1) = prem - \phi(s_1)$ and $\phi(c_2) = prem$. Consequently, when $s_2 \in (0, c_1]$, $\phi(s_1) + \phi(s_2) - prem \leq 0$; when $s_2 \in (c_1, c_2]$, $\phi(s_1) + \phi(s_2) - prem > 0$ and $\phi(s_2) \leq prem$; when $s_2 \in (c_2, \bar{s}]$, $\phi(s_1) + \phi(s_2) - prem > 0$ and $\phi(s_2) > prem$. Then, we have

$$\begin{aligned}
A(s_1) = {} & p\left(e_2^*(\phi(s_1))\right)\Bigg[\phi(s_1) - \int_0^{c_1} (prem - \phi(s_2))g(s_2)ds_2 \\
& - \int_{c_1}^{c_2} (prem - \phi(s_2))g(s_2)ds_2 + \int_0^{c_1} (prem - \phi(s_1) - \phi(s_2))g(s_2)ds_2\Bigg] \\
= {} & p\left(e_2^*(\phi(s_1))\right)\Bigg[\phi(s_1) - \int_0^{c_1} \phi(s_1)g(s_2)ds_2 \\
& - \int_{c_1}^{c_2} (prem - \phi(s_2))g(s_2)ds_2\Bigg] \\
= {} & p\left(e_2^*(\phi(s_1))\right)\Bigg[\int_{c_1}^{c_2} (\phi(s_1) + \phi(s_2) - prem)g(s_2)ds_2 + \int_{c_2}^{\bar{s}} \phi(s_1)g(s_2)ds_2\Bigg] \\
> {} & 0.
\end{aligned}$$

Therefore, for $\forall s_1 \in (0, \bar{s}]$, $A(s_1)$>0. Then, we have

$$
\begin{aligned}
v_2(0) - \int_0^{\bar{s}} v_2\big(\phi(s)\big) g(s) ds \\
&> \lambda\left(\int_0^{\bar{s}} \max\{prem - \phi(s), 0\}\, g(s) ds - prem\right) + \lambda \int_0^{\bar{s}} A(s_1) g(s_1) ds_1 \\
&> \lambda\left(\int_0^{\bar{s}} \max\{prem - \phi(s), 0\}\, g(s) ds - prem\right).
\end{aligned}
$$

Thus, (A.7) is true, and $e_1^* > e_2^*(0)$.

A.2. Rational agent model with an annual coverage limit

As discussed in Section 5.4, if a policy only has a per-accident coverage limit, the elevated accident risk in the last month of the policy term can only be explained by the sunk cost fallacy. However, if a policy has an annual coverage limit, the elevated accident risk in the last month of the policy term can also be explained by rational behavior. The intuition is that policyholders face uncertainty for a longer period during early months of a policy term than they do later in the policy term. Given that a policyholder makes a low accident-prevention effort during early months of a policy term and hence encounters an accident, if the policyholder encounters another accident later, the accumulated loss may surpass the annual coverage limit, and the policyholder will have to bear some loss by herself/himself. Therefore, a policyholder would expend a higher effort at the beginning; if no accident is encountered, she/he will later reduce the effort level, which leads to a higher accident risk as the expiration date of the policy approaches.

In Appendix A.2, we build a rational agent model that can generate the elevated accident risk in the later part of a policy if there is a coverage limit for the cumulative loss in the entire policy term.

Without loss of generality, suppose that a policy contract term has only two days. On each day, the policyholder can either have no accident or encounter one accident. The probability of encountering an accident on a day is $p(e)$, where e is the effort or caution that the policyholder exerts on the day to prevent an accident. Assume that $p'(e) < 0$ and $p''(e) > 0$, i.e., the effort will reduce the probability of an accident and the marginal effect of effort is diminishing. Conditional on an accident occurring, a random loss L will be incurred, with the probability density function $f(L)$ defined on the support $(0, \bar{L}]$.

Assume that the coverage limit of the policy for the total loss in the entire contract term is

M. For simplicity, we assume that the policy has neither deductibles nor per-accident coverage limits. However, the conclusion is extendable to the case in which deductibles and per-accident coverage limits exist. The expected payoff for the policyholder on the first day is

$$u_1(e_1) = -e_1 - p(e_1)\int_M^{\bar{L}} (L_1 - M) f(L_1) dL_1.$$

Given the effort level on the second day, e_2, the expected payoff for the policyholder on the second day conditional on the loss on the first day L_1 is

$$u_2(e_2|L_1) = -e_2 - p(e_2)\int_{\max\{M-L_1,0\}}^{\bar{L}} [L_2 - \max\{M - L_1, 0\}] f(L_2) dL_2.$$

The policyholder's optimal effort levels in the two days will be solved recursively. On the second day, the optimization problem for the policyholder given L_1 is as follows:

$$v_2(L_1) = \max_{e_2|L_1} u_2(e_2|L_1).$$

The first-order condition w.r.t. e_2 is

$$-p'(e_2)\int_{\max\{M-L_1,0\}}^{\bar{L}} [L_2 - \max\{M - L_1, 0\}] f(L_2) dL_2 = 1. \tag{B.1}$$

$v_2(L_1)$ is the indirect payoff function given that the optimal effort level e_2^* is chosen.

On the first day, the policyholder's optimization problem is as follows:

$$v_1(L_1) = \max_{e_1} u_1(e_1) + E[v_2(L_1)|e_1],$$

where $E[v_2(L_1)|e_1] = \big(1 - p(e_1)\big)v_2(0) + p(e_1)\int_0^{\bar{L}} v_2(L_1) f(L_1) dL_1$. The first-order condition w.r.t. e_1 is

(B.2) $$-p'(e_1)\int_M^{\bar{L}}(L_1-M)f(L_1)dL_1 = 1+p'(e_1)\left[v_2(0)-\int_0^{\bar{L}}v_2(L_1)\,f(L_1)dL_1\right].$$

Proposition B.1: (i) the optimal effort on the second day e_2^* is increasing in the loss on the first day L_1, i.e., $\frac{de_2^*(L_1)}{dL_1} \geq 0$; (ii) at least, $e_1^* > e_2^*(0)$, i.e., if no accident occurs on the first day, the effort level on the second day will be lower than that on the first day.

Proof:

(i) Differentiate equation (B.1), the first-order condition of the optimization problem on the second day, w.r.t. L_1. After rearrangement, we have:

$$\frac{de_2^*(L_1)}{dL_1} = \begin{cases} \dfrac{p'(e_2^*)\int_{M-L_1}^{\bar{L}} f(L_2)\,dL_2}{-p''(e_2^*)\int_{M-L_1}^{\bar{L}}[L_2-(M-L_1)]\,f(L_2)dL_2} > 0, & M-L_1 \geq 0 \\ 0, & M-L_1 < 0 \end{cases}$$

(ii) When $L_1 = 0$, the first-order condition of the second day (equation (B.1)) degenerates to

(B.3) $$-p'\big(e_2^*(0)\big)\int_M^{\bar{L}}[L_2-\mathrm{M}]\,f(L_2)dL_2 = 1.$$

In the first-order condition of the first day (equation (B.2)),

$$v_2(0)-\int_0^{\bar{L}}v_2(L_1)\,f(L_1)dL_1 > 0,$$

because $v_2(0) > v_2(L_1)$ for $\forall\ L_1 > 0$. Therefore,

(B.4) $$-p'(e_1^*)\int_M^{\bar{L}}(L_1-M)f(L_1)dL_1 = 1+p'(e_1^*)\left[v_2(0)-\int_0^{\bar{L}}v_2(L_1)\,f(L_1)dL_1\right] < 1.$$

Compare equation (B.4) with equation (B.3); because $p''(.) > 0$, we have

$$e_1^* > e_2^*(0).$$

It is easy to show that if there are only per-accident coverage limits and no coverage limits for the entire policy term, a rational policyholder would not reduce loss-prevention effort as the policy contract approaches the expiration date. In this case, the rational policyholder faces a separate coverage limit on each of the two days; therefore, the decisions of effort levels on each of the two days are independent of each other.

References

ABALUCK, J., J. GRUBER, AND A. SWANSON, "Prescription Drug Use under Medicare Part D: A Linear Model of Nonlinear Budget Sets," *Journal of Public Economics* 164 (2018), 106-38.

ABBRING, J. H., P.-A. CHIAPPORI, AND J. PINQUET, "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association* 1 (2003), 767-820.

AGARWAL, S., R. K. GREEN, E. ROSENBLATT, AND V. YAO, "Collateral Pledge, Sunk-Cost Fallacy and Mortgage Default," *Journal of Financial Intermediation* 24 (2015), 636-52.

ALLAIS, M., "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine," *Econometrica* 21 (1953), 503-46.

AL-NAJJAR, N., S. BALIGA, AND D. BESANKO, "Market Forces Meet Behavioral Biases: Cost Misallocation and Irrational Pricing," *The RAND Journal of Economics* 39 (2008), 214-37.

AUGENBLICK, N., "The Sunk-Cost Fallacy in Penny Auctions," *The Review of Economic Studies* 83 (2016), 58-86.

BAICKER, K., S. MULLAINATHAN, AND J. SCHWARTZSTEIN, "Behavioral Hazard in Health Insurance," *The Quarterly Journal of Economics* 130 (2015), 1623-67.

BALIGA, S., AND J. C. ELY, "Mnemonomics: The Sunk Cost Fallacy as a Memory Kludge," *American Economic Journal: Microeconomics* 3 (2011), 35-67.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER, "Salience Theory of Choice under Risk," *The Quarterly Journal of Economics* 127 (2012), 1243-85.

BUCHHEIT, S., AND N. FELTOVICH, "Experimental Evidence of a Sunk-Cost Paradox: A Study of Pricing Behavior in Bertrand–Edgeworth Duopoly," *International Economic Review* 52 (2011) , 317-47.

BURSZTYN, L., S. FIORIN, D. GOTTLIEB, AND M. KANZ, "Moral Incentives: Experimental Evidence from Repayments of an Islamic Credit Card," Working Paper, The World Bank, 2015.

BUSSE, M. R., D. G. POPE, J. C. POPE, AND J. SILVA-RISSO, "The Psychological Effect of Weather on Car Purchases," *The Quarterly Journal of Economics* 130 (2015), 371-414.

CADENA, X., AND A. SCHOAR, "Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments," NBER Working Paper No. 17020, 2011.

CHANG, T. Y., W. HUANG, AND Y. WANG, "Something in the Air: Pollution and the Demand for Health Insurance," *The Review of Economic Studies* 85 (2018), 1609-34.

CHEN, Y., F. LU, AND J. ZHANG, "Social Comparisons, Status and Driving Behavior," *Journal of Public Economics* 155 (2017), 11-20.

CHEN, Z., Z. LIU, J. C. SUÁREZ SERRATO, AND D. Y. XU, "Notching R&D Investment with Corporate Income Tax Cuts in China," NBER Working Paper No. 24749, 2019.

CHETTY, R., A. LOONEY, AND K. KROFT, "Salience and Taxation: Theory and Evidence," *American Economic Review* 99 (2009), 1145-77.

CHIAPPORI, P.-A., AND B. SALANIE, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy* 108 (2000), 56-78.

COHEN, A., "Asymmetric Information and Learning: Evidence from the Automobile Insurance

Market," *The Review of Economics and Statistics* 87 (2005), 197-207.

COHEN, A., AND R. DEHEJIA, "The Effect of Automobile Insurance and Accident Liability Laws on Traffic Fatalities," *The Journal of Law and Economics* 47 (2004), 357-93.

DALTON, C. M., G. GOWRISANKARAN, AND R. TOWN, "Myopia and Complex Dynamic Incentives: Evidence from Medicare Part D," *Review of Economic Studies* 87 (2020), 822-69.

DIONNE, G., C. GOURIEROUX, AND C. VANASSE, "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment," *Journal of Political Economy* 109 (2001), 444-53.

DIONNE, G., P.-C. MICHAUD, AND M. DAHCHOUR, "Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France," *Journal of the European Economic Association* 11 (2013), 897-917.

DIONNE, G., J. PINQUET, M. MAURICE, AND C. VANASSE, "Incentive Mechanisms for Safe Driving: A Comparative Analysis with Dynamic Data," *The Review of Economics and Statistics* 93 (2011), 218-27.

DU, N., L. LI, T. LU, AND X. LU, "Prosocial Compliance in P2P Lending: A Natural Field Experiment," *Management Science* 66 (2019), 315-33.

DUNHAM, W. R., "Moral Hazard and the Market for Used Automobiles," *Review of Industrial Organization* 23 (2003), 65-83.

FAN, C. K., AND W.-Y. WANG, "A Comparison of Underwriting Decision Making between Telematics-Enabled UBI and Traditional Auto Insurance," *Advances in Management and Applied Economics* 7 (2017), 17-30.

FELLNER, G., R. SAUSGRUBER, AND C. TRAXLER, "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information," *Journal of the European Economic Association* 11 (2013), 634-60.

FRIEDMAN, D., K. POMMERENKE, R. LUKOSE, G. MILAM, AND B. A. HUBERMAN, "Searching for the Sunk Cost Fallacy," *Experimental Economics* 10 (2007), 79-104.

GAO, F., J. LIEN, AND J. ZHENG, "Exaggerating to Break-Even: Reference-Dependent Moral Hazard in Auto Insurance Claims," Woring Paper, 2020.

GEYER, A., D. KREMSLEHNER, AND A. MUERMANN, "Asymmetric Information in Automobile Insurance: Evidence from Driving Behavior," *Journal of Risk and Insurance* 87 (2020), 969–95.

HALLSWORTH, M., J. A. LIST, R. D. METCALFE, AND I. VLAEV, "The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance," *Journal of Public Economics* 148 (2017), 14-31.

HO, T.-H., J. K. CHONG, AND X. XIA, "Yellow Taxis Have Fewer Accidents Than Blue Taxis Because Yellow Is More Visible Than Blue," *Proceedings of the National Academy of Sciences*, 114 (2017), 3074-78.

HO, T.-H., I. P. L. PNG, AND S. REZA, "Sunk Cost Fallacy in Driving the World's Costliest Cars," *Management Science* 64 (2018), 1761-78.

HOLMSTROM, B., "Moral Hazard and Observability," *The Bell Journal of Economics* 10 (1979), 74-91.

HOYT, R. E., D. B. MUSTARD, AND L. S. POWELL, "The Effectiveness of State Legislation in

Mitigating Moral Hazard: Evidence from Automobile Insurance," *The Journal of Law and Economics* 49 (2006), 427-50.

HUANG, D., AND Z. BAO, "Gender Differences in Reaction to Enforcement Mechanisms: A Large-Scale Natural Field Experiment," Working Paper, 2020.

JIN, W., Y. DENG, H. JIANG, Q. XIE, W. SHEN, AND W. HAN, "Latent Class Analysis of Accident Risks in Usage-Based Insurance: Evidence from Beijing," *Accident Analysis & Prevention* 115 (2018), 79-88.

JUST, D. R., AND B. WANSINK, "The Flat-Rate Pricing Paradox: Conflicting Effects of "All-You-Can-Eat" Buffet Pricing," *The Review of Economics and Statistics* 93 (2011), 193-200.

KAHNEMAN, D., AND A. TVERSKY, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (1979), 263-91.

KARLAN, D., M. MCCONNELL, S. MULLAINATHAN, AND J. ZINMAN, "Getting to the Top of Mind: How Reminders Increase Saving," *Management Science* 62 (2016), 3393-411.

KLEVEN, H. J., AND M. WASEEM, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *The Quarterly Journal of Economics* 128 (2013), 669-723.

LU, F., J. ZHANG, AND J. M. PERLOFF, "General and Specific Information in Deterring Traffic Violations: Evidence from a Randomized Experiment," *Journal of Economic Behavior & Organization* 123 (2016), 97-107.

PAN, Y., Y. QIN, F. ZHANG, AND H. ZHU, "Salience and Learning in the Land Market: Evidence from Winter Heating," Woring Paper, 2019.

PUELZ, R., AND A. SNOW, "Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market," *Journal of Political Economy* 102 (1994), 236-57.

RUBINSTEIN, A., AND M. E. YAARI, "Repeated Insurance Contracts and Moral Hazard," *Journal of Economic Theory* 30 (1983), 74-97.

SANDRONI, A., AND F. SQUINTANI, "Overconfidence, Insurance, and Paternalism," *American Economic Review* 97 (2007), 1994-2004.

SCHNEIDER, H., "Moral Hazard in Leasing Contracts: Evidence from the New York City Taxi Industry," *The Journal of Law & Economics* 53 (2010), 783-805.

SHAVELL, S., "On Moral Hazard and Insurance," *The Quarterly Journal of Economics* 93 (1979), 541-62.

SHI, L., Q. LIU, AND M. KIM, "Fuel Hedging and Behavior Bias: An Empirical Look at the U.S. Airline Industry," Woring Paper, 2020.

SHUM, M., AND Y. XIN, "Time-Varying Risk Aversion? Evidence from Near-Miss Accidents," *Review of Economics and Statistics*, forthcoming.

STARC, A., AND R. J. TOWN, "Internalizing Behavioral Externalities: Benefit Integration in Health Insurance," NBER Working Paper No. 21783, 2016.

THALER, R., "Mental Accounting and Consumer Choice," *Marketing science* 4 (1985), 199-214.

WEISBURD, S., "Identifying Moral Hazard in Car Insurance Contracts," *The Review of Economics and Statistics* 97 (2015), 301-13.

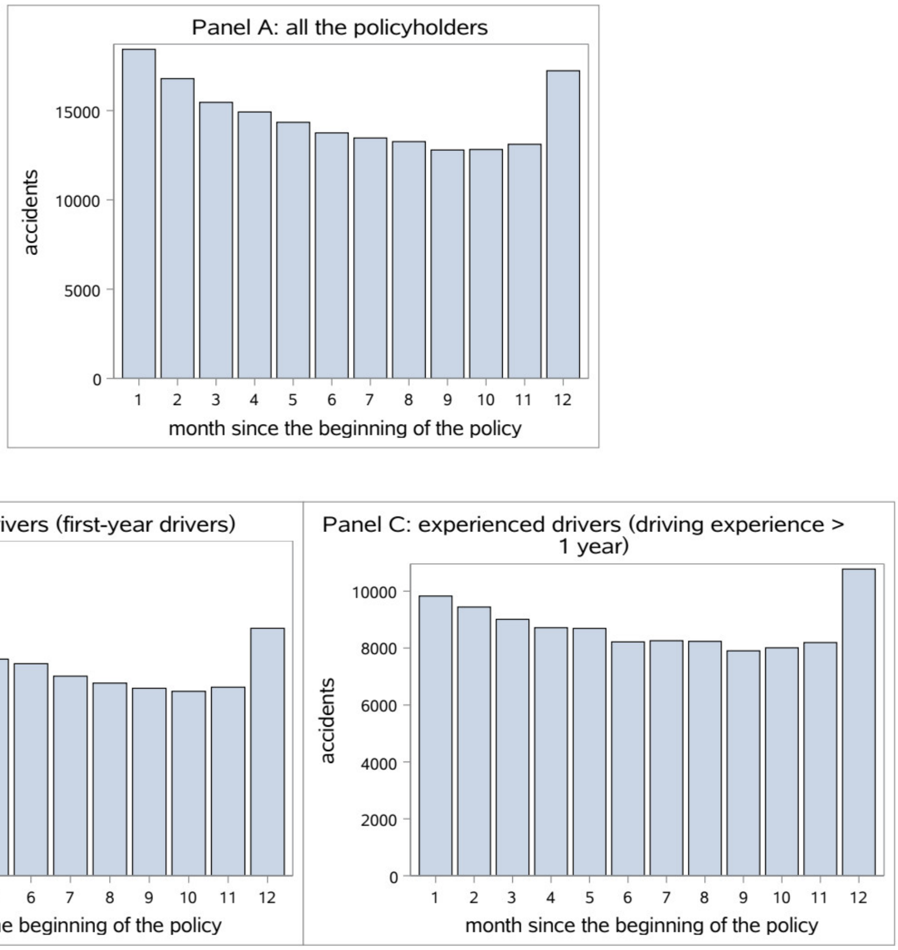

Figure 1. Accident counts. Panel A displays the accident counts in the data for each month within the one-year policy cycle. There is a spike in the last month of the cycle. The accident counts in the first two months are also abnormally high, mainly driven by new drivers (first-year drivers). New drivers have higher risk in the first two or three months and an additional month of driving experience is very helpful for them to reduce the risk. Panels B and C of Figure 1 report accident counts for new drivers and experienced drivers (with more than one year of driving experience), respectively. Each of the first eleven months of a policy term has 30 days; the remaining days of the policy term are assigned to the last month. Consequently, the last month can have more than 30 days (mostly 35 days). Correspondingly, to avoid exaggerating the number of accidents that occur in the last month, the accident counts of each policy in the last month are rescaled by dividing them by the number of days assigned to the last month and multiplying by 30.

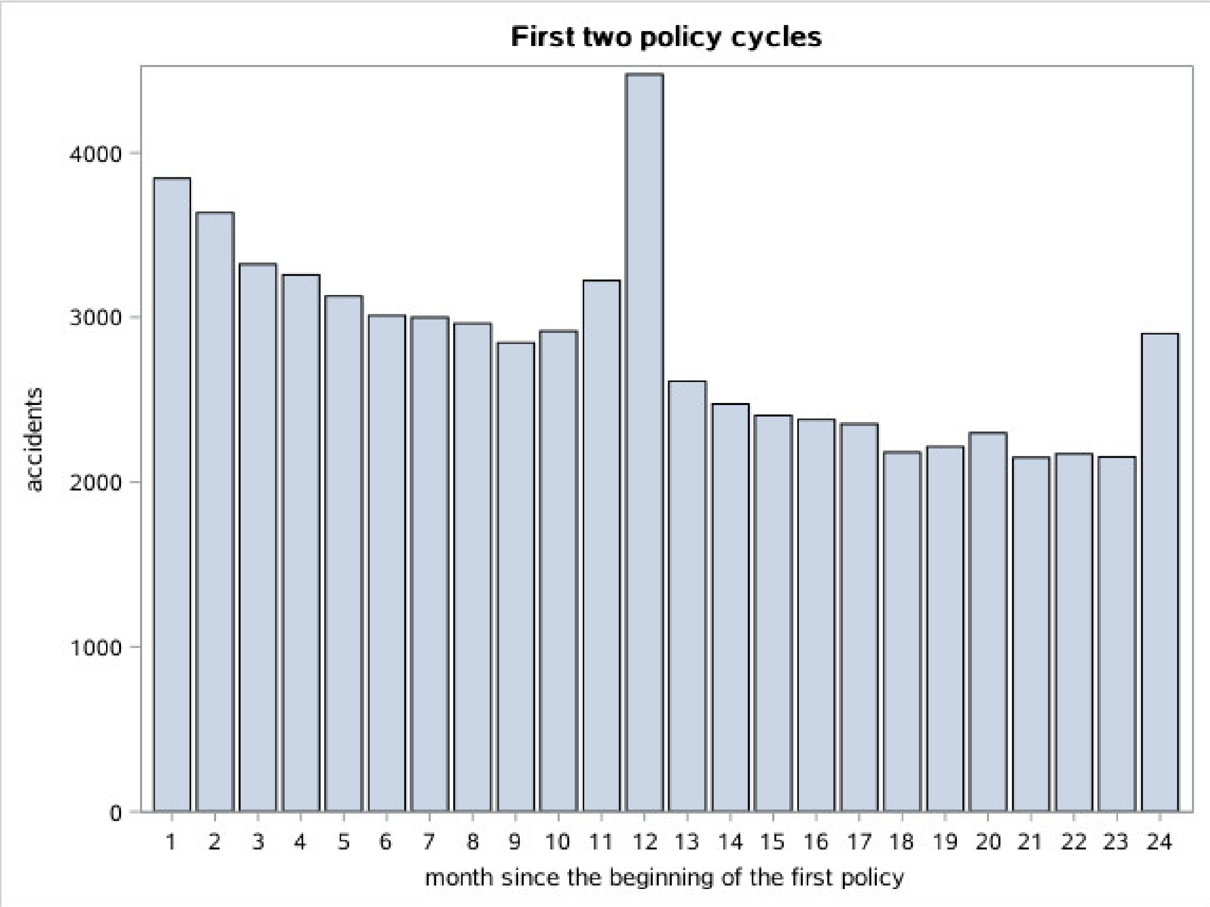

Figure 2. Accident counts in the first two policy cycles. The figure displays the accident counts for each month within the first two policy cycles for drivers that were served by the insurance company for at least two policy cycles. There is a spike in the last month of each policy cycle (the 12th and 24th months). Each of the first eleven months of a policy term has 30 days; the remaining days of the policy term are assigned to the last month. Consequently, the last month can have more than 30 days (mostly 35 days). Correspondingly, to avoid exaggerating the number of accidents that occur in the last month, the accident counts of each policy in the 12th and 24th months are rescaled by dividing them by the number of days assigned to the month and multiplying by 30.

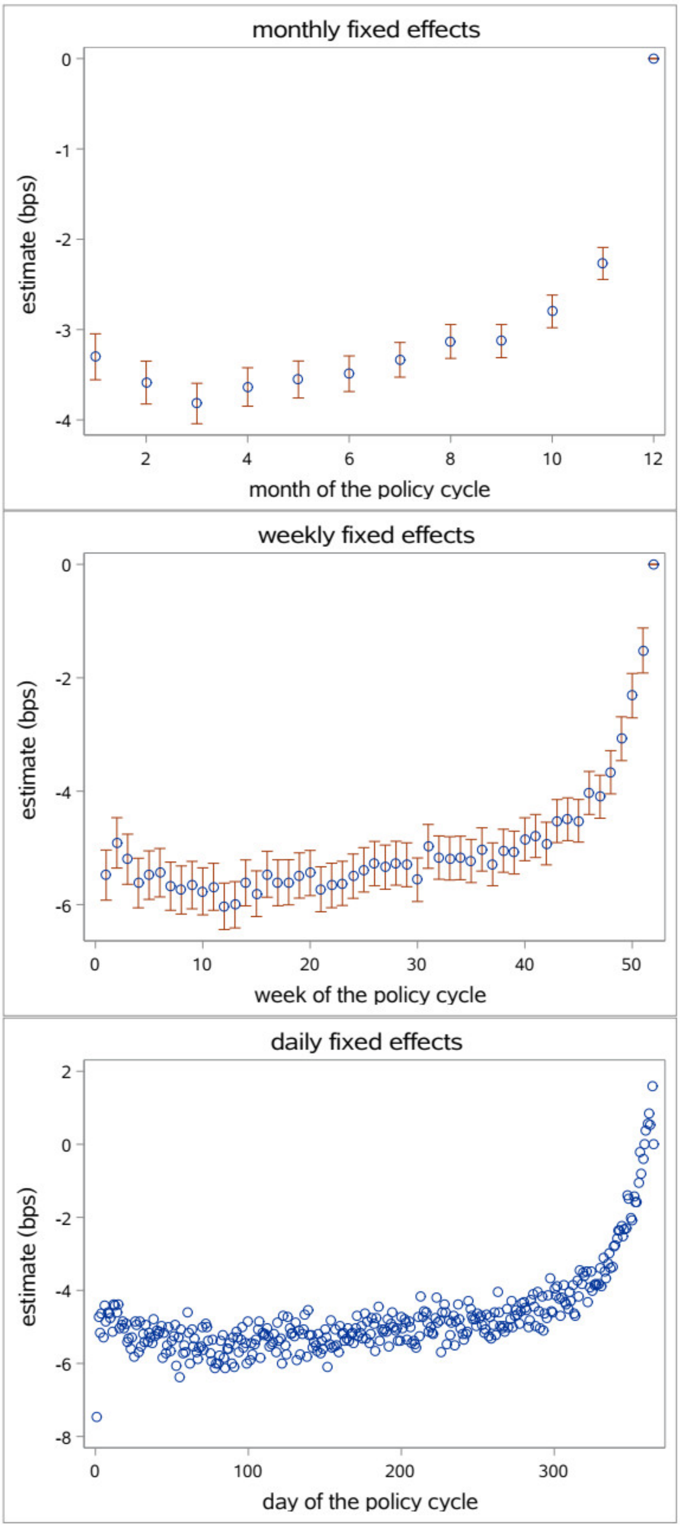

Figure 3. Display of estimates (all policyholders). Vertical bars in the first two diagrams represent the 95% confidence intervals. The last point in each diagram is not estimated; it is set

to be zero as the benchmark. All of the other points are estimated and are relative to the benchmark.

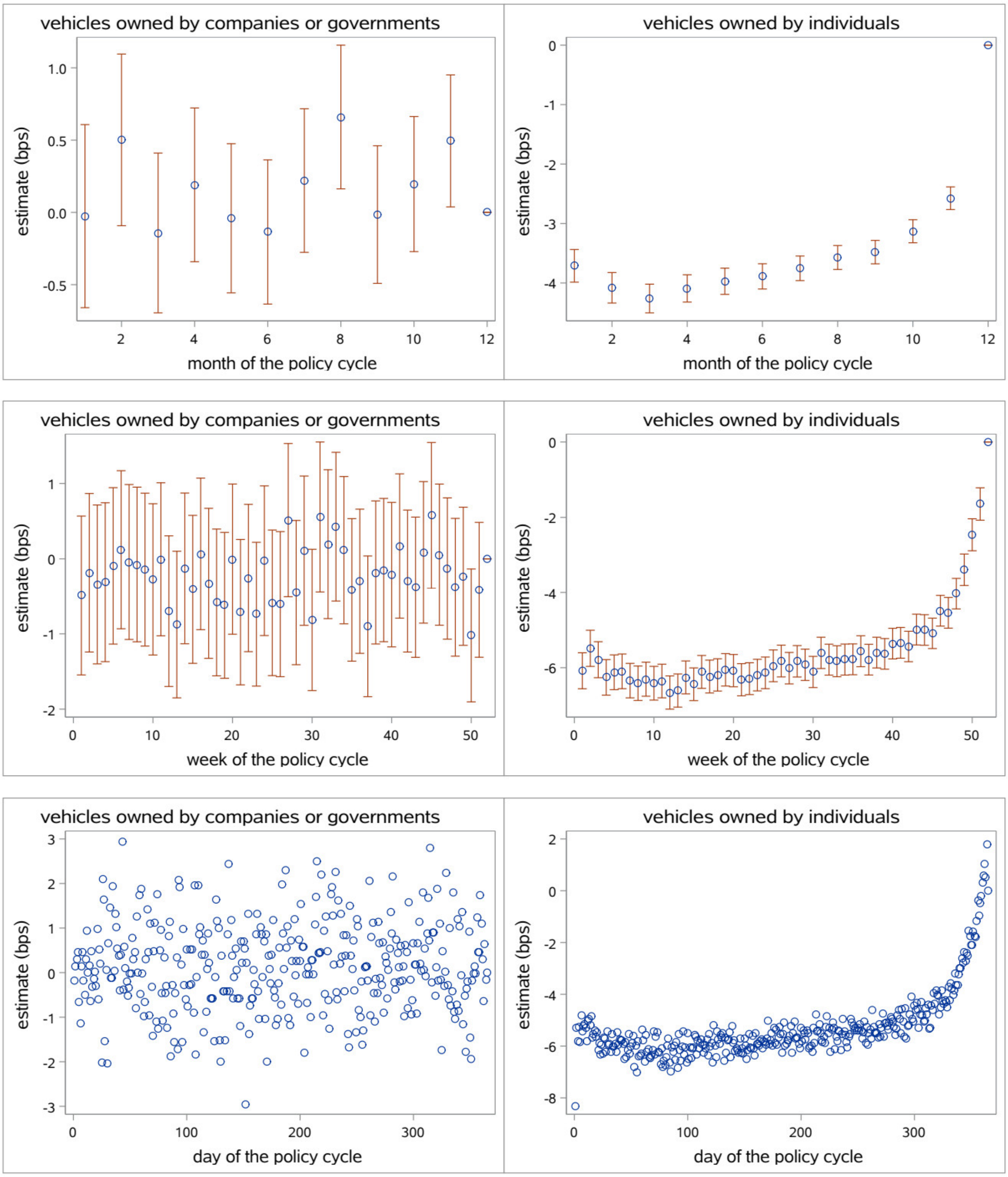

Figure 4. Display of estimates (individual-owned vehicles vs. company- or government-owned vehicles). Vertical bars in the diagrams in the first two rows represent the 95% confidence intervals. The last point in each diagram is not estimated; it is set to be zero as the benchmark. All of the other points are estimated and are relative to the benchmark.

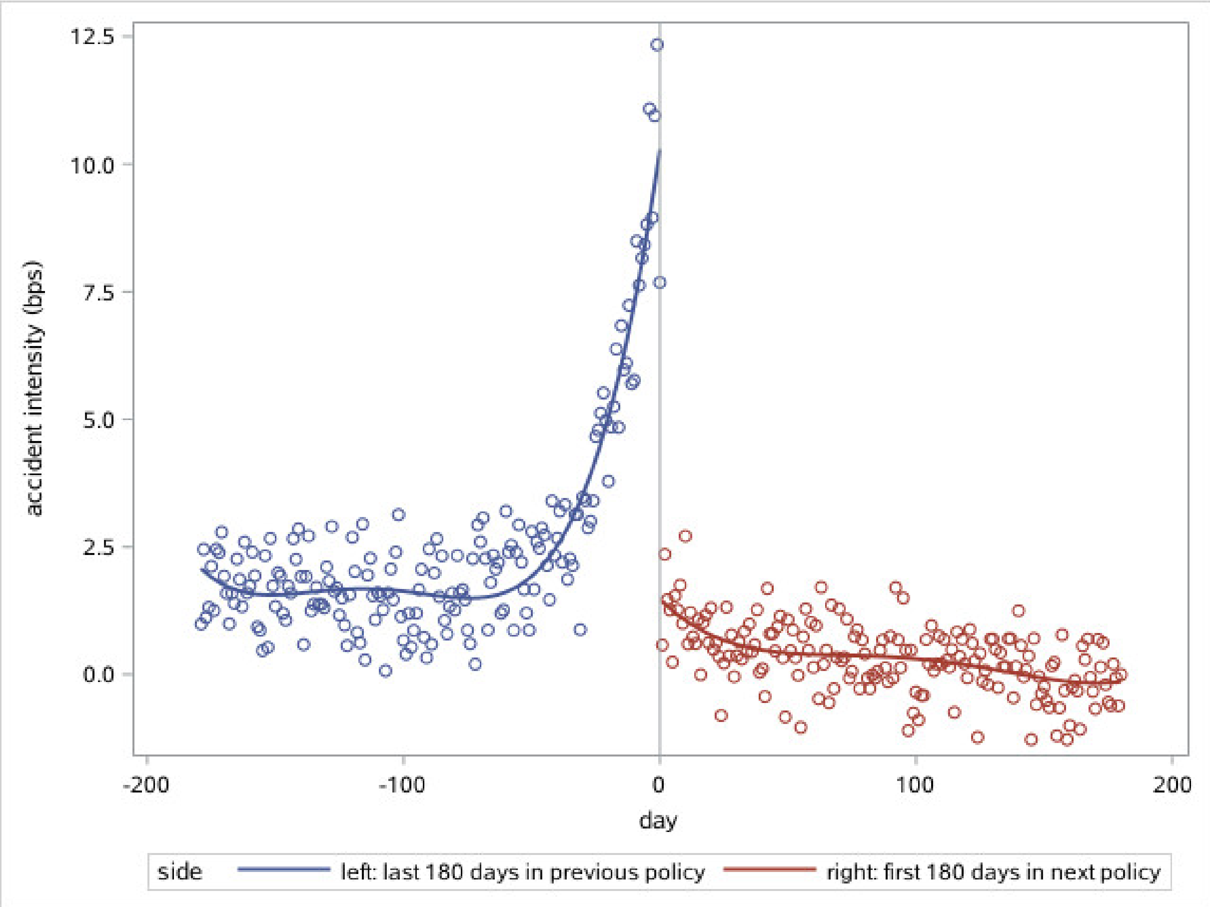

Figure 5. Regression of discontinuity. Equation (4.4) with $\gamma_0 + \gamma_1 I(\tau \leq 0) + f(\tau)$ replaced by the daily fixed effects is estimated. Each bubble represents the daily fixed effect for a day within the last 180 days of the preceding term and the first 180 days of the succeeding term. The last day of the preceding term is denoted as day 0, and the first day of the succeeding term is denoted as day 1.

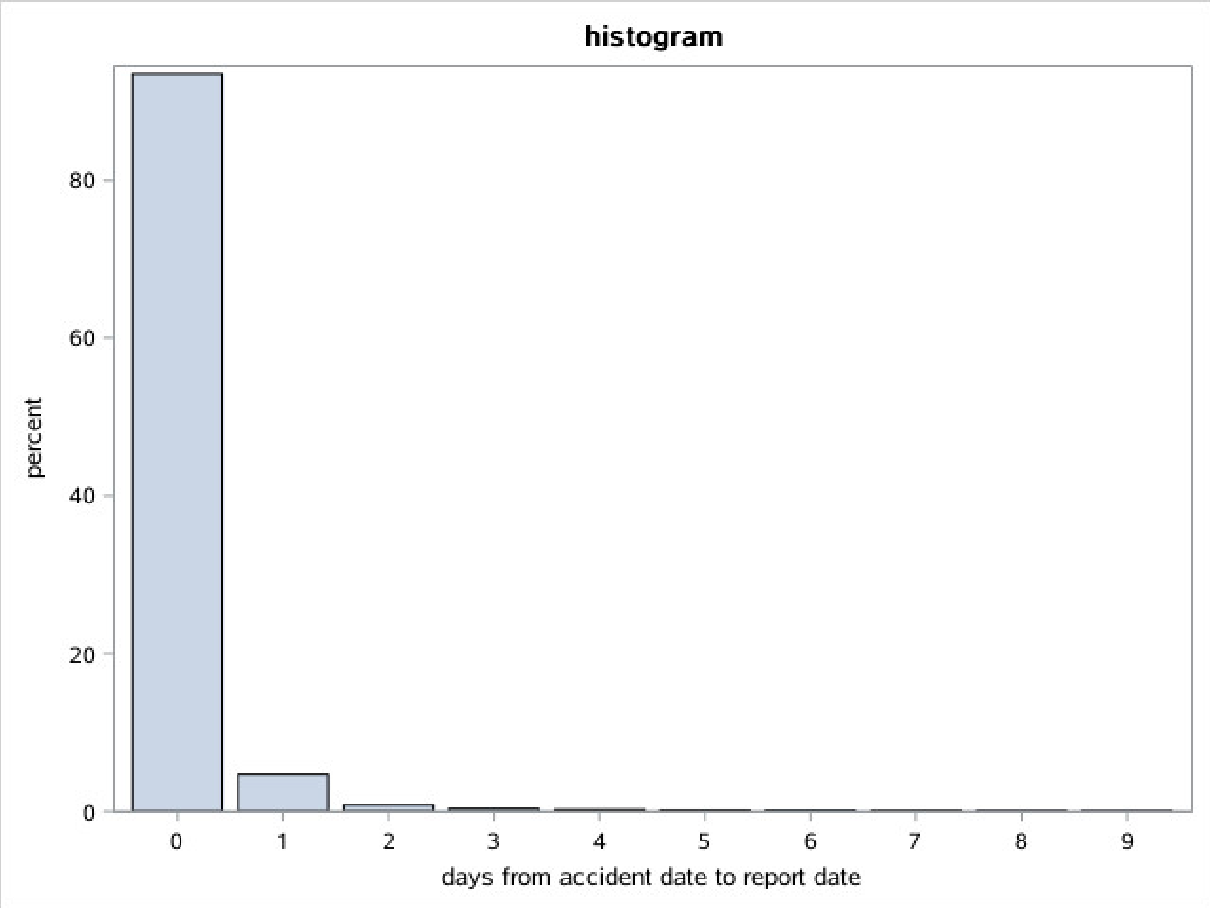

Figure 6. Distribution of the number of days from the accident date to the report date

Table 1. Descriptive statistics

Variable	N	Mean	Std Dev
Policyholder age (years)	630,983	39.88	9.52
Female	630,983	0.1974	0.3980
Seat number	630,982	4.74	1.71
Vehicle age (years)	596,608	4.21	3.44
Company- or government-owned	630,983	0.1000	0.3000
Premium (RMB)	630,983	2989.39	2814.84
Number of accidents per policy	630,983	0.2745	0.6519
Settlement payment per policy (RMB)	630,983	1778.96	15264.01
Settlement payment per accident (RMB)	173,202	6480.84	19285.58
Settlement payment per policy for policies with accidents (RMB)	121,954	9204.25	33721.52

Table 2. Baseline results

	Column 1: All car owners		Column 2: Company or government owners		Column 3: Individual owners		Column 4: Exogenous accidents		Column 5: Non-exogenous accidents	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-3.3012***	0.1297	-0.0258	0.3231	-3.7119***	0.1400	0.0049	0.0125	-7.1218***	0.3582
T_2	-3.5867***	0.1212	0.5016	0.3029	-4.0813***	0.1308	0.0114	0.0118	-7.9252***	0.3323
T_3	-3.8184***	0.1141	-0.1424	0.2820	-4.2619***	0.1232	0.0160	0.0128	-7.9710***	0.3197
T_4	-3.6354***	0.1085	0.1906	0.2714	-4.0917***	0.1170	0.0110	0.0111	-7.7675***	0.3076
T_5	-3.5528***	0.1045	-0.0408	0.2632	-3.9717***	0.1126	0.0107	0.0095	-7.7524***	0.3003
T_6	-3.4890***	0.1006	-0.1356	0.2545	-3.8894***	0.1083	0.0072	0.0089	-7.4814***	0.2928
T_7	-3.3350***	0.0981	0.2203	0.2534	-3.7535***	0.1054	0.0011	0.0084	-7.4170***	0.2849
T_8	-3.1319***	0.0957	0.6599**	0.2536	-3.5716***	0.1026	0.0073	0.0092	-7.3892***	0.2753
T_9	-3.1278***	0.0936	-0.0151	0.2429	-3.4812***	0.1006	0.0156	0.0106	-6.9927***	0.2700
T_{10}	-2.7988***	0.0921	0.1958	0.2384	-3.1302***	0.0989	-0.0031	0.0085	-6.6479***	0.2645
T_{11}	-2.2681***	0.0905	0.4943*	0.2329	-2.5749***	0.0972	0.0055	0.0082	-5.6508***	0.2595
Vehicle-driver fixed effects	✓		✓		✓		✓		✓	
Accident history	✓		✓		✓		✓		✓	
Driving experience	✓		✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓		✓	
Vehicle owner type	✓									
N	230,569,823		23,055,644		207,514,179		54,922,255		54,922,255	
Mean of dependent variable (bps)	7.5107		5.8029		7.7004		0.0259		13.9208	
$\frac{\alpha_{12} - \sum_{m=1}^{11} \alpha_m / 11}{Mean\ of\ dependent\ variable}$	43.63%				47.84%				52.32%	

$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$

$T_m(t)$ is a 0-1 dummy variable indicating whether day t is in the mth month of the policy term (T_{12} is omitted). In columns 1, 2, and 3, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t; $y_{i,t} = 0$ otherwise. In column 4, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an exogenous accident on day t; $y_{i,t} = 0$ otherwise. The exogenous accidents include natural disasters, explosions, fires, and thefts. These exogenous accidents are out of the drivers' control, and thus the probability of such accidents should not be elevated in the last month by the sunk cost fallacy. The sample for column 4 includes only individual policyholders with coverage for these exogenous accidents. The sample for column 5 is the same as that for column 4 but the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had a non-exogenous accident on day t; $y_{i,t} = 0$ otherwise. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 3. Evidence of reducing loss-prevention effort in the last month

	Column 1: Accidents with collisions		Column 2: Accidents with bodily injuries		Column 3: Accidents with settlements > RMB 10,000		Column 4: Policyholders with no accidents in previous 3 years	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-2.9664***	0.1357	-0.0653**	0.0208	-0.3120***	0.0529	-0.3057	0.2086
T_2	-3.3129***	0.1267	-0.0571**	0.0196	-0.3427***	0.0494	-0.6809***	0.2005
T_3	-3.4903***	0.1192	-0.0675***	0.0185	-0.3908***	0.0459	-0.7707***	0.1937
T_4	-3.3038***	0.1131	-0.0524**	0.0184	-0.3651***	0.0435	-1.0608***	0.1888
T_5	-3.1810***	0.1088	-0.0596***	0.0174	-0.3301***	0.0419	-1.0823***	0.1847
T_6	-3.1013***	0.1045	-0.0447**	0.0170	-0.2885***	0.0398	-1.3406***	0.1791
T_7	-2.9707***	0.1016	-0.0537***	0.0165	-0.3129***	0.0378	-1.2534***	0.1798
T_8	-2.7978***	0.0988	-0.0346*	0.0162	-0.2353***	0.0370	-1.3662***	0.1775
T_9	-2.7304***	0.0966	-0.0222	0.0160	-0.2789***	0.0355	-1.6026***	0.1735
T_{10}	-2.4046***	0.0950	-0.0302*	0.0153	-0.2147***	0.0348	-1.5681***	0.1718
T_{11}	-2.0001***	0.0930	-0.0371*	0.0146	-0.1305***	0.0341	-1.3200***	0.1685
Vehicle-driver fixed effects	✓		✓		✓		✓	
Accident history	✓		✓		✓		✓	
Driving experience	✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓	
N	207,514,179		201,264,158		207,514,179		46,589,363	
Mean of dependent variable (bps)	7.3725		0.2111		1.2324		4.9043	
$\frac{\alpha_{12}-\sum_{m=1}^{11}\alpha_m/11}{Mean\ of\ dependent\ variable}$	39.78%		22.58%		23.62%		22.90%	

$$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$$

$T_m(t)$ is a 0-1 dummy variable indicating whether day t is in the mth month of the policy term (T_{12} is omitted). In column 1, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t that involves collisions; $y_{i,t} = 0$ otherwise. In column 2, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident with bodily injuries on day t; $y_{i,t} = 0$ otherwise. In column 3, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t with a settlement amount higher than RMB 10,000; $y_{i,t} = 0$ otherwise. The sample for columns 1, 2, and 3 includes all the policies for individual-owned vehicles. These three columns only differ in the definition of dependent variables. The numbers of observations in these three columns are slightly different from each other because the three dependent variables have some missing values on different observations. In column 4, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t; $y_{i,t} = 0$ otherwise. The sample for column 4 includes the policies of which the policyholders had no accident in the previous three or more years. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 4. Payments and accident severity

	Column 1: Payments conditional on an accident		Column 2: Unconditional payments		Column 3: Bodily injury probability conditional on an accident	
	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	406.83*	173.32	-2.1331***	0.2367	0.0049**	0.0016
T_2	502.28**	174.83	-2.2567***	0.2218	0.0075***	0.0017
T_3	853.22***	200.53	-2.0669***	0.2257	0.0052**	0.0017
T_4	1052.78***	206.40	-1.7395***	0.2092	0.0075***	0.0018
T_5	831.50***	211.93	-1.8500***	0.2194	0.0060***	0.0018
T_6	941.74***	209.45	-1.6538***	0.2048	0.0072***	0.0019
T_7	771.39***	199.16	-1.6825***	0.1942	0.0063***	0.0019
T_8	912.42***	209.05	-1.5073***	0.1822	0.0092***	0.0019
T_9	997.95***	220.53	-1.3864***	0.1885	0.0110***	0.0020
T_{10}	799.35***	197.57	-1.2713***	0.1711	0.0076***	0.0019
T_{11}	790.56***	200.99	-0.8824***	0.1710	0.0038*	0.0019
N	176,985		207,514,179		170,528	
Mean of dependent variable	6477.00 (RMB)		5.0539 (RMB)		0.0292	

The regressions in columns 1 and 3 are based on the accident-level data. Each observation is an accident. T_m is a 0-1 dummy variable indicating whether the accident occurs in the mth month of the policy term (T_{12} is omitted). The dependent variable in column 1 is the payment by the insurance company for the accident; the dependent variable in column 3 is a 0-1 indicator of whether the accident causes bodily injuries. The regression in column 2 is based on the driver-day-level data. Each observation is a driver-day combination. T_m is a 0-1 dummy variable indicating whether the day of the observation is in the mth month of the policy term (T_{12} is omitted). The regression is the same as that in column 3 of Table 2 except that the dependent variable is the unconditional payment instead of the accident indicator. For a driver-day with an accident, the dependent variable equals the payment made by the insurance company for the accident; for a driver-day without an accident, the dependent variable equals 0. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 5. Regression discontinuity

Panel A: Degree of polynomial = 3					
RD window	[-29, 30]	[-59, 60]	[-89, 90]	[-119, 120]	[-179, 180]
γ_1	9.3769***	8.9925***	9.1058***	8.8131***	8.0823***
	(0.9146)	(0.6529)	(0.5336)	(0.4599)	(0.3702)
N	8,897,940	17,795,880	26,693,820	35,591,760	53,387,640

Panel B: RD window = [-89, 90]					
Degree of polynomial	1	2	3	4	5
γ_1	5.9974***	8.2513***	9.1058***	9.1038***	9.0119***
	(0.2563)	(0.3960)	(0.5336)	(0.6649)	(0.7915)
N	26,693,820	26,693,820	26,693,820	26,693,820	26,693,820

$$y_{w,\tau} = \gamma_0 + \gamma_1 I(\tau \le 0) + f(\tau) + \beta X_w + \varepsilon_{w,\tau}, \qquad \tau \in [-(\bar{\tau}-1), \bar{\tau}]$$

$y_{w,\tau}$ is the accident indicator (rescaled to 10,000 bps) on the τth day in vehicle-driver-window w. Each RD window consists of two adjacent policy cycles for a driver. $\tau = 0$ for the last day of the preceding policy term, $\tau = -1$ for the day before the last day of the policy term, $\tau = 1$ for the first day of the succeeding policy term, $\tau = 2$ for the second day of the succeeding policy term, and so on. $I(\tau \le 0)$ is a 0-1 indicator of whether the day is in the preceding policy part of the RD window. $f(\tau)$ is a flexible polynomial of τ, allowing the coefficients of each term in the polynomial to be different between the domain where $\tau \le 0$ and the domain where $\tau > 0$. In X_w, we control for vehicle-driver-window fixed effects and calendar-month fixed effects. In panel A, we set the degree of the polynomial $f(\tau)$ equal to 3 and alter the RD window bandwidth by changing $\bar{\tau}$ from 30 days to 180 days. In panel B, we set $\bar{\tau} = 90$ and change the degree of the polynomial $f(\tau)$ from 1 to 5. Standard errors in parentheses are clustered by vehicle-driver-window. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 6. The effect of accident history on premium

	Est.	S. E.
No accident for 3 or more years	-1549.4746***	77.8710
No accident for 2 years	-1529.6489***	77.9183
No accident for 1 year	-1406.6199***	77.8803
New driver	-914.6440***	77.9972
1 accident last term	-843.2802***	78.0493
2 accidents last term	-471.1739***	84.9891
Policyholder age	✓	
Seat number	✓	
Vehicle age	✓	
Driving experience	✓	
City fixed effects	✓	
Calendar year-month fixed effects	✓	
Gender	✓	
Coverage	✓	
Vehicle model fixed effects	✓	

Each observation is a policy. The sample includes all the individual policyholders with non-missing accident history categories. Policyholders in the category of “3 or more accidents last term” are omitted. The dependent variable is the premium of the policy. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 7. Estimated average next-term premium increase caused by an accident in the current term

Accident history for current term	Percentage of the sample	Accident history for next term if no accident in current term	Accident history for next term if 1 accident in current term	Premium increase next term relative to current term if 1 accident in current term	Premium increase next term caused by 1 accident in current term
No accident for 3 or more years	29.43%	No accident for 3 or more years	1 accident last year	RMB 706.19	RMB 706.19
No accident for 2 years	14.04%	No accident for 3 or more years	1 accident last year	RMB 686.37	RMB 706.19
No accident for 1 year	20.23%	No accident for 2 years	1 accident last year	RMB 563.34	RMB 686.37
New driver	27.04%	No accident for 1 year	1 accident last year	RMB 71.36	RMB 563.34
1 accident last year	8.71%	No accident for 1 year	1 accident last year	RMB 0	RMB 563.34
2 accidents last year	0.43%	No accident for 1 year	1 accident last year	RMB -372.11	RMB 563.34
3 or more accidents last year	0.12%	No accident for 1 year	1 accident last year	RMB -842.28	RMB 563.34

For each category of accident history for the current policy term, the last column reports the average premium difference in the next term between the case in which the driver encounters one accident in the current term and the case in which the driver does not encounter any accident in the current term; the penultimate column reports the average premium increase in the next term relative to the current term if one accident occurs in the current term.

Table 8. The sunk cost fallacy effects across policyholders with different accident histories

	Column 1:		Column 2:	
	Est.	S. E.	Est.	S. E.
No accident for 3 or more years	7.1352***	0.3819	7.0488***	0.3819
No accident for 2 years	7.3551***	0.3841	7.2233***	0.3844
No accident for 1 year	8.0767***	0.3832	7.8518***	0.3833
New driver	9.6021***	0.3825	9.4939***	0.3829
1 accident last term	11.5609***	0.3933	11.2942***	0.3941
2 accidents last term	16.6820***	0.4850	16.0811***	0.4892
3 or more accidents last term	20.5119***	1.2701	19.9292***	1.2771
Last-month indicator	3.0420***	0.0839	1.6422***	0.0487
Last-month indicator interacted with:				
No accident for 3 or more years			0.8670***	0.1347
No accident for 2 years			1.3208***	0.2107
No accident for 1 year			2.2708***	0.1947
New driver			1.0886***	0.1704
1 accident last term			2.7085***	0.3460
2 accidents last term			6.1488***	1.0103
3 or more accidents last term			5.8909	3.9419
Policyholder age	✓		✓	
Seat number	✓		✓	
Vehicle age	✓		✓	
Driving experience	✓		✓	
City fixed effects	✓		✓	
Calendar year-month fixed effects	✓		✓	
Gender	✓		✓	
Vehicle model fixed effects	✓		✓	
N	195,156,433		195,156,433	
Mean of dependent variable	8.1412		8.1412	

Each observation is a policyholder-day combination. The dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t; $y_{i,t} = 0$ otherwise. The policyholders with missing accident history categories are omitted. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 9. Sunk cost fallacy vs. rational behavior

	Column 1: Policies with per-accident coverage limits only		Column 2: In days after proceeds surpassed premium		Column 3: Polices with annual coverage limits		Column 4: In days after proceeds surpassed premium	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-0.2904*	0.1139	3.0276	6.4278	-6.5415***	0.2601	6.0707	5.2439
T_2	-0.3556***	0.1024	4.7594	4.8662	-7.0269***	0.2433	-4.5197	3.0403
T_3	-0.4549***	0.0960	-0.8577	3.9224	-7.2083***	0.2316	-11.1474***	2.3073
T_4	-0.3532***	0.0926	2.5874	3.5425	-6.9516***	0.2240	-9.2637***	1.9917
T_5	-0.2238*	0.0897	1.6388	3.0393	-6.8751***	0.2185	-11.5915***	1.7147
T_6	-0.3169***	0.0846	1.1345	2.6250	-6.7099***	0.2116	-9.7758***	1.5379
T_7	-0.2788***	0.0813	1.7169	2.2351	-6.5986***	0.2063	-10.5570***	1.3797
T_8	-0.3007***	0.0776	1.4973	1.8578	-6.3629***	0.1998	-8.8228***	1.2507
T_9	-0.3213***	0.0735	1.0779	1.5323	-6.3118***	0.1948	-11.5496***	1.1121
T_{10}	-0.1535*	0.0713	1.3763	1.2428	-5.9283***	0.1902	-10.4157***	0.9970
T_{11}	-0.1736**	0.0674	1.0881	0.9788	-4.9294***	0.1869	-7.7835***	0.8981
Vehicle-driver fixed effects	✓		✓		✓		✓	
Accident history	✓		✓		✓		✓	
Driving experience	✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓	
N	106,587,615		1,548,080		100,926,564		5,574,873	
Mean of dependent variable (bps)	2.2806		10.1933		13.4243		30.1280	

$$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$$

In column 1, the sample includes individual policyholders with per-accident coverage limits only. In column 2, the sample includes the same policyholders as column 1 but only includes the observations in the days after the cumulative settlement proceeds within the policy term surpassed the premium. In column 3, the sample includes policyholders with both per-accident coverage limits and annual coverage limits. In column 4, the sample includes the same policyholders as column 3 but only includes the observations in the days after the cumulative settlement proceeds within the policy term surpassed the premium. In these four columns, the dependent variable $y_{i,t} = 1$ (rescaled to 10,000 bps) if policyholder i had an accident on day t; $y_{i,t} = 0$ otherwise. $T_m(t)$ is a 0-1 dummy variable indicating whether day t is in the mth month of the policy term. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table 10. Percentage of accidents due to the sunk cost fallacy

Degree of polynomial (p)	With per-accident coverage limits only	With both per-accident and annual coverage limits
0	2.56%	6.53%
1	2.07%	5.70%
2	2.00%	3.91%

Online Appendix

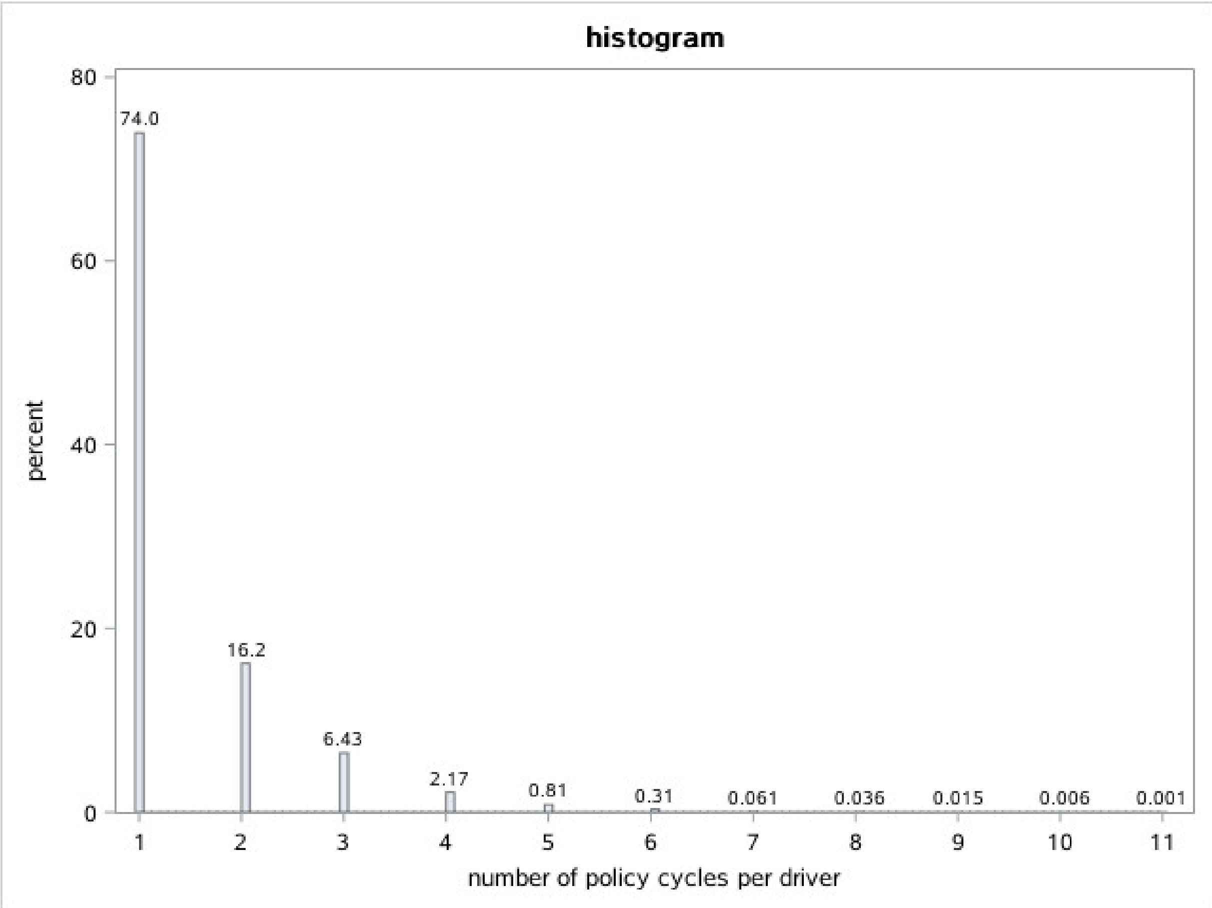

Figure C.1. Histogram of policy cycles per driver

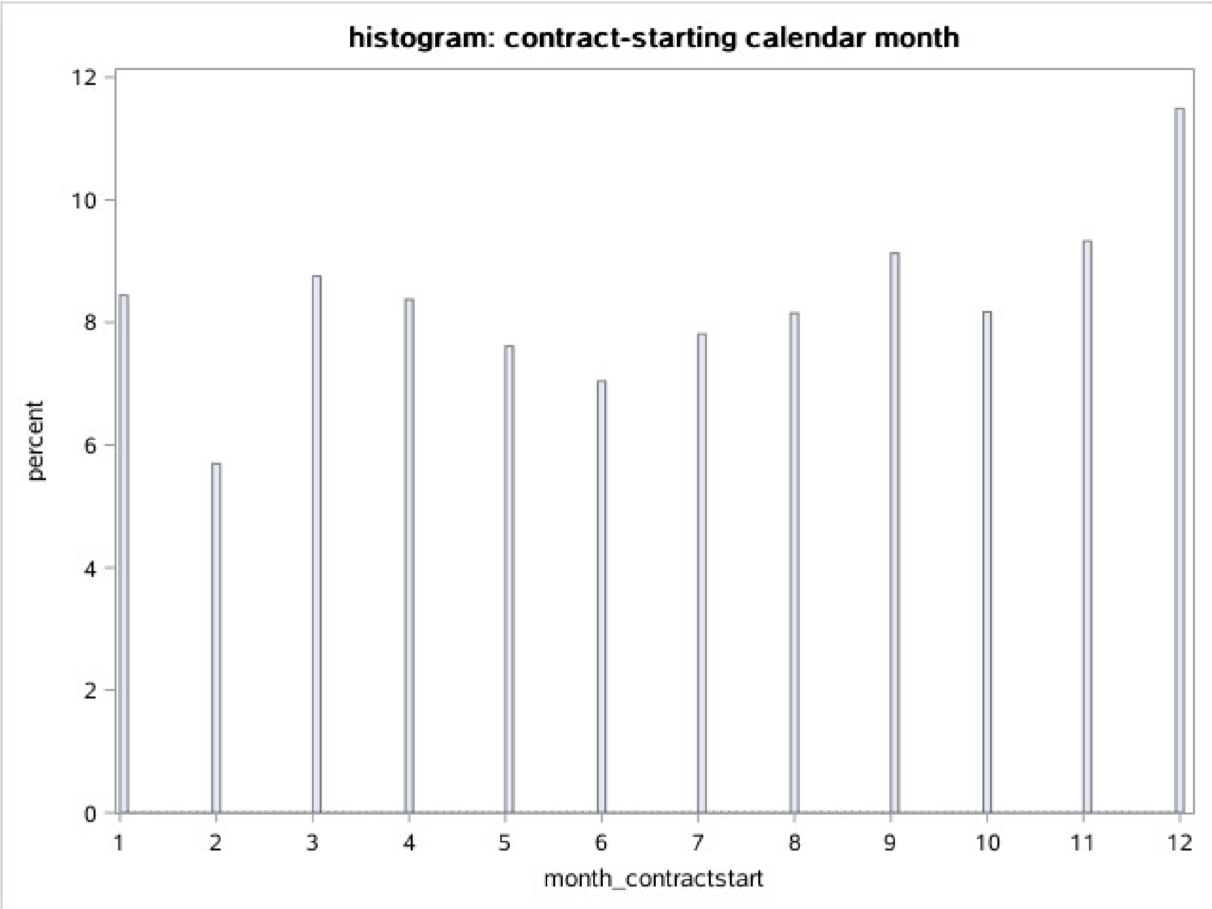

Figure C.2. Histogram of calendar months in which policies started

Table C.1. Baseline results: Linear terms for driving experience

	Column 1: All car owners		Column 2: Company or government owners		Column 3: Individual owners		Column 4: Exogenous accidents		Column 5: Non-exogenous accidents	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-2.5841***	0.1112	-0.1861	0.2730	-2.8536***	0.1200	0.0073	0.0104	-5.6808***	0.3208
T_2	-3.0947***	0.1077	0.3847	0.2710	-3.4795***	0.1160	0.0086	0.0103	-7.1564***	0.3092
T_3	-3.5079***	0.1047	-0.2241	0.2602	-3.8681***	0.1128	0.0097	0.0116	-7.7020***	0.3026
T_4	-3.4667***	0.1024	0.1368	0.2586	-3.8618***	0.1102	0.0026	0.0104	-7.8523***	0.2936
T_5	-3.4904***	0.1003	-0.0732	0.2535	-3.8663***	0.1079	0.0014	0.0093	-8.0685***	0.2865
T_6	-3.5015***	0.0979	-0.1521	0.2476	-3.8734***	0.1053	-0.0021	0.0089	-7.9271***	0.2805
T_7	-3.3951***	0.0963	0.2149	0.2488	-3.7962***	0.1034	-0.0076	0.0084	-7.9101***	0.2739
T_8	-3.2156***	0.0946	0.6615**	0.2504	-3.6460***	0.1014	-0.0002	0.0088	-7.8644***	0.2674
T_9	-3.2148***	0.0929	-0.0100	0.2404	-3.5645***	0.0998	0.0097	0.0104	-7.3994***	0.2650
T_{10}	-2.8720***	0.0918	0.2015	0.2372	-3.2028***	0.0985	-0.0072	0.0085	-6.9490***	0.2620
T_{11}	-2.3133***	0.0904	0.4983*	0.2328	-2.6206***	0.0971	0.0033	0.0084	-5.8204***	0.2589
Additional driving experience (months) × new-driver indicator	**-0.2566*****	**0.0145**	**-0.0371**	**0.0369**	**-0.2755*****	**0.0156**	**-0.0039**	**0.0024**	**-0.3328*****	**0.0709**
Additional driving experience (months) × experienced-driver indicator	**0.0235***	**0.0102**	**0.0009**	**0.0253**	**0.0275***	**0.0109**	**-0.0017**	**0.0022**	**0.0481**	**0.0671**
No accident for 3 or more years	8.1430***	1.7951	6.1452**	2.3557	8.5041***	2.1954	0.0106	0.0085	43.2918***	1.3796
No accident for 2 years	5.9866***	1.7916	4.9324*	2.3378	6.2379**	2.1920	0.0207	0.0134	37.2521***	1.3519
No accident for 1 year	3.9720*	1.7894	2.7729	2.3371	4.2194	2.1892	0.0006	0.0054	31.4800***	1.3390
New driver	7.7477***	1.7956	3.6717	2.3671	8.2431***	2.1947	0.0019	0.0170	32.1807***	1.3910
1 accident last term	-4.2633*	1.7967	-5.4658*	2.3716	-4.0127	2.1966	0.0085	0.0121	21.9184***	1.3543
2 accidents last term	-11.0734***	1.8616	-18.8146***	2.9274	-10.3181***	2.2536	0.0042	0.0072	11.3860***	1.4263
3 or more accidents last term	-3.0412	2.7247	-5.8066*	2.3726	-2.7720	3.0051	--	--	--	--
Vehicle-driver fixed effects	✓		✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓		✓	

Vehicle owner type	✓				
N	230,569,823	23,055,644	207,514,179	54,922,255	54,922,255

$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$

This table is another version of Table 2 in which we only control for linear trends in the effects of additional months of driving experience on new drivers (first-year drivers) and experienced drivers (with more than one year of driving experience), respectively, and do not control for the nonlinear trends, to better illustrate the effect of driving experience. The results indicate that additional months of driving experience are helpful for new drivers but are not for experienced drivers. The estimates of α_m are similar to those in Table 2. This table also reports the estimates for accident history categories. In contrast to Table 8 in which vehicle-driver fixed effects are not controlled for and the coefficients show that better accident histories lead to lower accident probabilities for the current policy, after controlling for vehicle-driver fixed effects, the variation in accident histories becomes within a vehicle-driver and there is no pattern that, within a vehicle-driver, a better accident history leads to lower accident probabilities for the current policy. There is even a rough pattern that, within a vehicle-driver, a worse accident history leads to lower accident probabilities for the current policy. One possible explanation is that it is a phenomenon of mean reversion. Another possible explanation is that, after accidents, drivers become more aware of the risk or more risk averse and then drive more cautiously. Shum and Xin (2020) found that drivers drive more conservatively following "near-miss" accidents (measured by hard brakes or hard turns). In columns 1, 2, and 3, policyholders with missing accident histories in the sample (approximately 13%) are omitted. In columns 4 and 5, policyholders with 3 or more accidents in the previous term are omitted because, in that sample, no policyholders' accident histories are missing. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table C.2. Evidence of reducing loss-prevention effort in the last month: Linear terms for driving experience

	Column 1: Accidents with collisions		Column 2: Accidents with bodily injuries		Column 3: Accidents with settlements > RMB 10,000		Column 4: Policyholders with no accidents in previous 3 years	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-2.0762***	0.1166	-0.0270	0.0183	-0.2446***	0.0450	-0.5160**	0.1921
T_2	-2.6750***	0.1124	-0.0267	0.0179	-0.2906***	0.0428	-0.8721***	0.1891
T_3	-3.0589***	0.1092	-0.0440**	0.0170	-0.3516***	0.0411	-0.9419***	0.1867
T_4	-3.0373***	0.1065	-0.0346*	0.0171	-0.3366***	0.0399	-1.2118***	0.1843
T_5	-3.0417***	0.1043	-0.0465**	0.0163	-0.3104***	0.0393	-1.2129***	0.1819
T_6	-3.0554***	0.1016	-0.0354*	0.0162	-0.2756***	0.0379	-1.4510***	0.1777
T_7	-2.9883***	0.0996	-0.0475**	0.0160	-0.3053***	0.0364	-1.3442***	0.1788
T_8	-2.8525***	0.0976	-0.0308*	0.0159	-0.2315***	0.0361	-1.4377***	0.1769
T_9	-2.7993***	0.0959	-0.0200	0.0157	-0.2777***	0.0349	-1.6557***	0.1732
T_{10}	-2.4680***	0.0947	-0.0292*	0.0152	-0.2149***	0.0345	-1.6036***	0.1717
T_{11}	-2.0411***	0.0929	-0.0368*	0.0146	-0.1310***	0.0340	-1.3390***	0.1685
Additional driving experience (months) × new-driver indicator	**-0.2951*** **	**0.0155**	**-0.0033**	**0.0028**	**-0.0427*** **	**0.0072**	**--**	**--**
Additional driving experience (months) × experienced-driver indicator	**0.0261***	**0.0109**	**0.0014**	**0.0022**	**0.0021**	**0.0052**	**0.0083**	**0.0155**
No accident for 3 or more years	8.5855***	2.1942	0.4756**	0.1444	1.6906**	0.5484	--	--
No accident for 2 years	6.2580**	2.1907	0.3499*	0.1418	1.3277*	0.5455	--	--
No accident for 1 year	4.1420	2.1880	0.2173	0.1409	1.0427	0.5437	--	--
New driver	8.3292***	2.1934	0.1682	0.1420	1.5821**	0.5484	--	--
1 accident last term	-4.2565	2.1954	-0.0625	0.1414	-0.3074	0.5493	--	--
2 accidents last term	-10.5268***	2.2526	-0.3267	0.1768	-3.2650***	0.6323	--	--
3 or more accidents last term	-2.6797	3.0108	0.0200	0.2533	-0.9728	1.4615	--	--
Vehicle-driver fixed effects	✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓	
N	207,514,179		201,264,158		207,514,179		46,589,363	

$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$

This table is another version of Table 3 in which we only control for linear trends in the effects of additional months of driving experience on new drivers (first-year drivers) and experienced drivers (with more than one year of driving experience), respectively, and do not control for the nonlinear trends, to better illustrate the effect of driving experience. The results indicate that additional months of driving experience are helpful for new drivers but are not for experienced drivers. The estimates of α_m are similar to those in Table 3. This table also reports the estimates for accident history categories. In columns 1, 2, and 3, policyholders with missing accident histories in the sample (approximately 13%) are omitted. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table C.3. Baseline results: Monthly-level regressions

	Column 1: All car owners		Column 2: Company or government owners		Column 3: Individual owners		Column 4: Exogenous accidents		Column 5: Non-exogenous accidents	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-99.4315***	3.8969	-0.0422	9.7414	-111.8556***	4.2053	0.1284	0.3810	-215.5323***	10.7622
T_2	-108.1219***	3.6392	15.3520	9.1105	-123.0253***	3.9258	0.2883	0.3541	-239.5314***	9.9708
T_3	-115.0794***	3.4258	-3.9309	8.4953	-128.4798***	3.6965	0.4401	0.3898	-241.2067***	9.5963
T_4	-109.9728***	3.2574	5.6599	8.1630	-123.7598***	3.5118	0.2921	0.3374	-236.0795***	9.2323
T_5	-107.2596***	3.1360	-0.6145	7.9161	-119.9726***	3.3784	0.2675	0.2914	-235.5178***	9.0140
T_6	-105.2416***	3.0205	-3.4523	7.6553	-117.3331***	3.2515	0.1706	0.2729	-226.9166***	8.7857
T_7	-100.6707***	2.9462	6.2813	7.6136	-113.2595***	3.1659	-0.0230	0.2582	-225.0619***	8.5501
T_8	-94.6344***	2.8742	19.7305*	7.6093	-107.8677***	3.0818	0.1759	0.2776	-224.1031***	8.2611
T_9	-94.2809***	2.8139	0.8248	7.3009	-104.9858***	3.0213	0.4256	0.3233	-211.8080***	8.1005
T_{10}	-84.1341***	2.7691	6.1173	7.1727	-94.1015***	2.9724	-0.1292	0.2595	-200.6317***	7.9372
T_{11}	-68.3609***	2.7286	15.2583*	7.0220	-77.6139***	2.9293	0.1377	0.2519	-170.5372***	7.8024
Vehicle-driver fixed effects	✓		✓		✓		✓		✓	
Accident history	✓		✓		✓		✓		✓	
Driving experience	✓		✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓		✓	
Vehicle owner type	✓									
N	7,570,296		757,032		6,813,264		1,804,080		1,804,080	
Mean of dependent variable (bps)	224.8190		174.3864		230.4226		0.7797		415.9839	
$\frac{\alpha_{12} - \sum_{m=1}^{11} \alpha_m / 11}{Mean\ of\ dependent\ variable}$	43.96%				48.22%				53.04%	

$$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$$

This table is the monthly version of Table 2. $T_m(t)$ is a 0-1 dummy variable indicating whether month t is in the mth month of the one-year policy cycle (T_{12} is omitted). In columns 1, 2, and 3, the dependent variable $y_{i,t}$ is the number of accidents that occurred within the tth month since the first contract of the policyholder started (rescaled to 10,000 bps for each accident). In column 4, the dependent variable $y_{i,t}$ is the number of exogenous accidents that occurred within month t (rescaled to 10,000 bps for each accident). The exogenous accidents include natural disasters, explosions, fires, and thefts. These exogenous accidents are out of the drivers' control, and thus the probability of such accidents should not be elevated in the last month by the sunk cost fallacy. The sample for column 4 includes only individual policyholders with coverage for these exogenous accidents. The sample for column 5 is the same as that for column 4 but the dependent variable $y_{i,t}$ is the number of non-exogenous accidents that occurred within month t (rescaled to 10,000 bps for each accident). In the monthly regressions, each of the first eleven months of a policy term has 30 days; the remaining days of the policy term are assigned to the last month. Consequently, the last month can have more than 30 days (mostly 35 days). Correspondingly, to avoid exaggerating the number of accidents that occur in the last month, we rescale the dependent variable for the last month of a policy term by dividing it by the number of days assigned to the last month and multiplying by 30. Because a contract can start in the middle of a calendar month, month t can cross two calendar months; correspondingly, the calendar year-month fixed effect represents the calendar year-month in which month t starts. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table C.4. Evidence of reducing loss-prevention effort in the last month: Monthly-level regressions

	Column 1: Accidents with collisions		Column 2: Accidents with bodily injuries		Column 3: Accidents with settlements > RMB 10,000		Column 4: Policyholders with no accidents in previous 3 years	
	Est.	S. E.	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-89.3537***	4.0778	-1.8984**	0.6264	-9.3945***	1.5916	-8.4954	6.2614
T_2	-99.7649***	3.8034	-1.6444**	0.5898	-10.2740***	1.4856	-20.2163***	6.0189
T_3	-105.0636***	3.5778	-1.9871***	0.5566	-11.7522***	1.3818	-22.8680***	5.8123
T_4	-99.8313***	3.3934	-1.5200**	0.5518	-10.9988***	1.3073	-31.8973***	5.6675
T_5	-95.9736***	3.2631	-1.7724***	0.5214	-9.9894***	1.2577	-32.3561***	5.5469
T_6	-93.3912***	3.1366	-1.3078**	0.5089	-8.6960***	1.1947	-39.7767***	5.3799
T_7	-89.5106***	3.0490	-1.5821***	0.4950	-9.4393***	1.1371	-37.6694***	5.3989
T_8	-84.3599***	2.9670	-1.0004*	0.4862	-7.0726***	1.1126	-41.0367***	5.3306
T_9	-82.2101***	2.9032	-0.6620	0.4776	-8.4000***	1.0668	-48.4584***	5.2123
T_{10}	-72.1430***	2.8556	-0.8658	0.4587	-6.3983***	1.0463	-46.9684***	5.1642
T_{11}	-60.1411***	2.8024	-1.0976*	0.4392	-3.8705***	1.0261	-40.1324***	5.0895
Vehicle-driver fixed effects	✓		✓		✓		✓	
Accident history	✓		✓		✓		✓	
Driving experience	✓		✓		✓		✓	
Calendar year-month fixed effects	✓		✓		✓		✓	
N	6,813,264		6,608,028		6,813,264		1,530,264	
Mean of dependent variable (bps)	220.8518		6.3382		32.1900		146.6668	
$\frac{\alpha_{12}-\sum_{m=1}^{11}\alpha_m/11}{Mean\ of\ dependent\ variable}$	39.99%		22.00%		27.19%		22.93%	

$$y_{i,t} = \sum_{m=1}^{12} \alpha_m T_m(t) + \beta X_{i,t} + \varepsilon_{i,t}$$

This table is the monthly version of Table 3. $T_m(t)$ is a 0-1 dummy variable indicating whether month t is in the mth month of the one-year policy cycle (T_{12} is omitted). In column 1, the dependent variable $y_{i,t}$ is the number of accidents that occurred within the tth month since the start of the policyholder's first contract and involved collisions (rescaled to 10,000 bps for each accident). In column 2, the dependent variable $y_{i,t}$ is the number of accidents with bodily injuries that occurred within month t (rescaled to 10,000 bps for each accident). In column 3, the dependent variable $y_{i,t}$ is the number of accidents that occurred within month t and incurred a settlement amount higher than RMB 10,000 (rescaled to 10,000 bps for each accident). The sample for columns 1, 2, and 3 includes all the policies for individual-owned vehicles. These three columns only differ in the definition of dependent variables. The numbers of observations in these three columns are slightly different from each other because the three dependent variables have some missing values on different observations. The sample for column 4 includes the policies of which the policyholders had no accident in the previous three or more years. The dependent variable $y_{i,t}$ is the number of accidents that occurred within month t (rescaled to 10,000 bps for each accident). Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table C.5. Alternative specification for the bodily-injury regression

	Est.	S. E.
T_1	-0.0812***	0.0202
T_2	-0.0722***	0.0190
T_3	-0.0811***	0.0180
T_4	-0.0652***	0.0178
T_5	-0.0708***	0.0168
T_6	-0.0547***	0.0164
T_7	-0.0617***	0.0160
T_8	-0.0415**	0.0157
T_9	-0.0276	0.0155
T_{10}	-0.0335*	0.0148
T_{11}	-0.0383**	0.0142
Vehicle-driver fixed effects	✓	
Accident history	✓	
Driving experience	✓	
Calendar year-month fixed effects	✓	
N	207,514,179	
Mean of dependent variable (bps)	0.2047	
$\frac{\alpha_{12} - \sum_{m=1}^{11} \alpha_m / 11}{Mean\ of\ dependent\ variable}$	27.88%	

In Table 3, the number of observations for column 2 (bodily injuries) is slightly smaller than that for column 1. The reason is that, for a small proportion of policyholders (approximately 3%), if an accident occurs, we do not know whether it caused bodily injuries or not. Therefore, we exclude these policyholders from the regression in column 2 of Table 3. The average daily accident probability for this excluded group is 8.7973 bps, which is not very different from that for all the individual vehicle owners (7.7004 bps in column 3 of Table 2). Alternatively, in this table, we include these policyholders in the regression and set the dependent variable for them always equal to zero (i.e., assume that all the accidents encountered by these policyholders did not cause bodily injuries). The results are robust. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table C.6. Payments conditional on an accident

	Column 1: All individual drivers		Column 2: New individual drivers		Column 3: Experienced individual drivers	
	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	919.83***	170.90	874.03**	301.07	1054.95***	201.52
T_2	933.84***	171.94	659.75*	291.34	1215.83***	213.99
T_3	1194.04***	198.01	1094.79**	341.05	1312.17***	240.98
T_4	1452.08***	203.33	1754.13***	383.21	1291.71***	218.60
T_5	1168.84***	208.70	1183.93**	379.87	1188.77***	241.41
T_6	1248.34***	206.47	1079.96**	349.34	1396.28***	255.81
T_7	1111.65***	195.13	1164.69***	339.11	1102.55***	235.86
T_8	1183.12***	205.24	1341.87***	369.62	1099.11***	241.86
T_9	1219.99***	217.27	1211.14***	354.67	1244.28***	275.73
T_{10}	992.24***	193.14	949.77**	347.73	1027.87***	227.69
T_{11}	976.44***	196.84	831.43*	345.13	1066.33***	237.53
Premium	1.0979***	0.0383	1.0559***	0.0623	1.1713***	0.0495
N	176,985		71,587		105,398	

This table reports alternative specifications for the regression in column 1 of Table 4. Each observation is an accident. The dependent variable is the payment by the insurance company for the accident. Compared to column 1 of Table 4, column 1 of this table adds the variable premium to the regression; column 2 restricts the regression sample to new drivers (first-year drivers); column 3 restricts the regression sample to experienced drivers (with more than one year of driving experience). The results indicate that the lower coefficients for months 1 and 2 compared to months 3-11 in column 1 of Table 4 are driven by new drivers. A possible explanation is that new drivers in the first two months may be reluctant to drive on highways or at high speeds and hence the damage given an accident tends to be small, although new drivers are more likely to encounter an accident in the first two months, as suggested by Figure 1. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.

Table C.7. Separate regressions by starting calendar months

Starting calendar month:	January		February		March	
	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-4.4479***	0.6435	-3.1332***	0.6600	-5.1028***	0.6245
T_2	-5.0761***	0.7161	-3.7688***	0.7181	-5.3723***	0.7046
T_3	-4.7525***	0.7714	-3.7305***	0.7521	-5.1821***	0.7273
T_4	-4.2132***	0.8037	-3.6227***	0.7800	-4.7035***	0.7473
T_5	-4.0517***	0.8305	-2.8022***	0.8047	-4.7881***	0.7622
T_6	-4.0848***	0.8288	-3.2971***	0.8034	-4.5137***	0.7775
T_7	-4.3080***	0.8130	-1.6695*	0.7992	-4.1332***	0.7808
T_8	-3.7814***	0.7714	-1.4521	0.7779	-3.0426***	0.7671
T_9	-2.7783***	0.6733	-1.1271	0.7267	-3.0980***	0.7442
T_{10}	-2.1371***	0.5526	-1.4760*	0.6077	-3.2933***	0.6662
T_{11}	-1.5361***	0.4291	-0.6169	0.4285	-1.3025*	0.5144
Starting calendar month:	April		May		June	
	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-3.6097***	0.6867	-4.5466***	0.6911	-3.1989***	0.7434
T_2	-3.4720***	0.7407	-4.0062***	0.7482	-3.1542***	0.8011
T_3	-3.4560***	0.7779	-3.9435***	0.7529	-3.2852***	0.7874
T_4	-3.2376***	0.7710	-4.0476***	0.7651	-2.9846***	0.7739
T_5	-2.5203***	0.7792	-3.8560***	0.7592	-2.8693***	0.7605
T_6	-2.3770**	0.7914	-3.8791***	0.7638	-3.0475***	0.7432
T_7	-2.6062***	0.8024	-3.5957***	0.7613	-2.5183***	0.7350
T_8	-2.0920**	0.7916	-2.5835***	0.7517	-2.3351***	0.7264
T_9	-1.8231*	0.7624	-2.5510***	0.7302	-2.1015**	0.7046
T_{10}	-1.1273	0.7193	-2.3520***	0.6817	-1.5466*	0.6576
T_{11}	-1.4000*	0.5859	-2.0360***	0.5745	-0.7699	0.5337
Starting calendar month:	July		August		September	
	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-4.3601***	0.7182	-3.1212***	0.7201	-4.5102***	0.7170
T_2	-3.8776***	0.7976	-1.5971*	0.7975	-4.4215***	0.8176
T_3	-3.1492***	0.8071	-2.1976**	0.8365	-4.7016***	0.8764
T_4	-2.6532***	0.7900	-2.0716*	0.8146	-5.0615***	0.8783
T_5	-2.9876***	0.7698	-1.9667*	0.7780	-4.7613***	0.8388
T_6	-2.4050**	0.7592	-1.6923*	0.7564	-4.5769***	0.8046
T_7	-2.5722***	0.7428	-1.6771*	0.7301	-4.2297***	0.7744
T_8	-2.0719**	0.7219	-1.4036*	0.6936	-3.6702***	0.7316
T_9	-2.2508***	0.6929	-1.4262*	0.6536	-3.4097***	0.6736
T_{10}	-1.4259*	0.6378	-1.1929*	0.6014	-2.4850***	0.6144
T_{11}	-1.6487***	0.5181	-0.8945	0.4995	-1.5678***	0.4846

Starting calendar month:	October		November		December	
	Est.	S. E.	Est.	S. E.	Est.	S. E.
T_1	-2.3347***	0.6880	-3.9394***	0.6769	-5.1959***	0.6613
T_2	-1.8552*	0.7656	-3.7947***	0.7512	-5.1156***	0.7390
T_3	-2.7048***	0.7998	-3.3937***	0.7868	-4.4843***	0.7866
T_4	-2.3530**	0.8185	-3.8009***	0.8048	-3.8469***	0.8118
T_5	-1.8336*	0.8131	-3.7557***	0.8342	-4.2744***	0.8220
T_6	-3.0010***	0.7669	-1.8980*	0.8161	-3.5616***	0.8330
T_7	-2.5981***	0.7202	-2.0158**	0.7594	-3.2072***	0.8073
T_8	-2.5029***	0.6809	-1.1893	0.6997	-2.8465***	0.7379
T_9	-2.0416***	0.6296	-0.9338	0.6529	-2.3057***	0.6569
T_{10}	-1.8238***	0.5647	-1.3885*	0.5712	-2.0480***	0.5832
T_{11}	-1.9209***	0.4719	-1.4048**	0.4371	-1.3181**	0.4417

For this table, the policies purchased by individual vehicle owners are divided into 12 groups by the calendar month in which the policy starts. The regression equation is the same as that for column 3 of Table 2 and is estimated separately for each group. Standard errors are clustered by vehicle-driver. * denotes significance at a 5% level. ** denotes significance at a 1% level. *** denotes significance at a 0.1% level.