Effects of Insurance Incentives on Road Safety:
Evidence from a Natural Experiment in China*

15 October 2019

Georges Dionne[a]
Ying Liu[b]

[a] Corresponding author. HEC Montréal, Montreal, Canada, H3T 2A7
Email: georges.dionne@hec.ca;

[b] Shandong University, Jinan, China, 250100
Email: liuying1030@sdu.edu.cn

Abstract

We contribute to the growing literature on moral hazard by offering empirical evidence of the effectiveness of insurance pricing incentives at improving road safety by comparing the claim frequency following a regulatory reform introduced in a pilot city in China with the experience of another city unaffected by the reform. By using the methodology of difference-in-differences, we find that improving insurance pricing on past claims and on traffic violations with full industry commitment reduces moral hazard and insured drivers' claim frequency by 12%. The treatment effects are, however, heterogeneous with respect to insured drivers' wealth and past claims history.

Keywords: Difference-in-differences; experience rating; traffic violation; past claim; moral hazard.

JEL Classification: C33; D81; D82; G22; R41.

I. Introduction

Traffic accidents cause serious injuries, disabilities and fatalities all over the world. It is therefore worthwhile to study what policy interventions can improve road safety, and how effective these policies are. Improvements in automobile safety equipment and highway design, strict enforcement of traffic laws and attempts to stimulate safe driving behavior via monetary or non-monetary incentives are three important channels to enhance road safety (Vukina and Nestić, 2015). Concerning private incentives, monetary mechanisms such as fines and non-monetary mechanisms such as point-record drivers' licenses have proven to be effective.[1] An experienced-rated premium based on past claims and on traffic violations in multi-period insurance contracts is another form of monetary incentive, which can be justified by the potential presence of asymmetric information between an insured driver and an insurer regarding individual risk (Dionne et al., 2013b). However, the causal effect of asymmetric information on automobile accidents is still not well documented because of the scarcity of appropriate data isolating the causality of incentive effects.

Asymmetric information goes in two directions in insurance contracting: adverse selection and moral hazard, both of which indicate a positive correlation between accident probability and insurance coverage chosen by the insured driver. Adverse selection means that high-risk insured drivers choose more coverage than do low-risk insured drivers, whereas moral hazard means that more coverage reduces the incentives for safe driving and therefore causes more accidents. The literature that tries to disentangle these two information problems dates back to the study by Arrow (1963). In the presence of moral hazard, past claims or traffic violations pricing may help prevent accidents (Abbring et al., 2003; Bourgeon and Picard, 2007). For adverse selection, risk

---

[1] On fines, see Bar-Ilan and Sacerdote (2004) and Makowsky and Stratmann (2011). On point-record drivers' licenses, see Abay (2018); Bourgeon and Picard (2007); Castillo-Manzano and Castro-Nuño (2012); De Paola et al. (2013) and Dionne et al. (2011).

classification seems more efficient (Crocker and Snow, 1985, 1986; Hoy, 1982). Empirically distinguishing moral hazard from adverse selection can therefore provide insight into the optimal policy scheme that can reduce inefficiencies associated with asymmetric information (Weisburd, 2015).

Chiappori and Salanié (2000; 2013) suggest that using cross-sectional data[2] cannot disentangle the two dimensions of asymmetric information, whereas employing dynamic panel data or a natural experiment[3] can. Although panel data were employed, some studies (Abbring et al., 2003; Dionne and Ghali, 2005; Rowell et al., 2017; Zavadil, 2015) did not find evidence of moral hazard while others did (Dionne et al., 2011; Dionne et al., 2013a; Israel, 2007; Wang et al., 2008; Weisburd, 2015). Dionne et al. (2013a) are the first to separate moral hazard from adverse selection and learning about risk within the context of a dynamic model that identifies different causality effects. Weisburd (2015), using an instrumental variable approach, estimates the effect of accident costs on driving behavior in a context where automobile insurance coverage is included as a fringe benefit offered by employers. Given that the selection mechanism for gaining insurance coverage is a function of occupation instead of individual preferences for insurance, this article estimates the effect of moral hazard from a plausibly exogenous variation in insurance coverage that is not related to adverse selection.

Compared with the prolific literature on panel data studies, systematic evidence of moral hazard based on natural experiments related to an exogenous regulatory change is lacking because appropriate data before and after the shock are scarce. Some exceptions are the studies by Dionne and Ghali (2005), Dionne et al. (2011), Lee (2013), and Li et al. (2007), which compare

---

[2] Using cross-sectional data, Chiappori and Salanié (2000) and Dionne et al. (2001) find no evidence of residual asymmetric information in insurance portfolios.

[3] Natural experiments where the population is randomly split into treated and control groups are valuable but scarce. If identical populations face different incentive schemes for exogenous reasons, this can be regarded as a quasi-natural experiment to test for moral hazard. In this paper, the experiment is mainly a quasi-natural experiment.

pre- and post-treatment claim frequency for insured drivers exposed to a treatment. In particular, Dionne et al. (2011) analyze agents' driving behavior as a function of past demerit points accumulated, and identify the presence of moral hazard. They further exploit the introduction of an experience-rating scheme in 1992 as a natural experiment and find that this reform reduces the frequency of traffic infractions by 15%. However, some temporal trends in accident frequency or other events than the treatment between the pre- and post-treatment periods may have biased their results because they had no access to a control group that was not affected by the reform. A satisfactory untreated control group is necessary to identify moral hazard from temporal variation in accident frequency that is not due to the experiment exposure. Based on this idea, the difference-in-differences (henceforth DID) approach is useful in the context of a natural experiment. Another underlying identifying assumption for DID is the parallel trend for the average accident frequency of the treated and control groups in the pre-treatment period. Without the verification of this assumption, the behavior of the control group cannot constitute a valid counterfactual for the treated group.

The paper most closely related to ours is that of Abay (2018), which examines the implementation of a demerit-point scheme in Denmark to investigate the differential behavioral responses of the drivers in the treated and control groups using the DID approach. Because of data limitations, the research design of this study separates drivers who face the common reform into treated and control groups based on their past driving behaviors, which is not the usual practice for conducting a DID study. The methodology requires the control group to be unaffected by the treatment, or the model specification would suffer from potential endogeneity.

Overcoming these limitations, our study provides estimates of a causal effect of insurance incentives on road safety. To this end, we exploit an exogenous new experience rating reform in

4

a pilot city in China in 2011 as a natural experiment. We obtained data from an insurer present in both the pilot city and another city that is unaffected by the reform. For robustness, we also use propensity score matching (henceforth PSM) to construct a matched sample, composed of comparable treated vehicles in the pilot city and untreated vehicles in the unaffected city. We then compare the effect of the reform on treated vehicles with the experience of untreated vehicles in the framework of DID.

Chiappori (2000) and Chiappori and Salanié (2013) suggest that when the population remains unchanged in the treated and control groups during the experiment period, the resulting change in driving behavior can be attributed to moral hazard. By concentrating on the vehicles (and on vehicle owners later) that remain in the sample for the whole period of analysis, we can assume that the populations of drivers in the two groups remain similar during the study period. Any improvement in road safety should then be more attributable to the variation in insurance incentives or moral hazard than to adverse selection.

By establishing the causal effect of insurance incentives on claim frequency and the presence of moral hazard in the studied market, we contribute to the insurance literature in three ways. First, we address (in some degree) the potential selection bias in the treatment relying on the DID matching method of Heckman et al. (1997; 1998). We can do so because we have panel data and both pre- and post-reform information on claim frequency. Our analysis employs a rich dataset on 20,484 vehicles across the treated and control city in China from 2009 to 2012. This dataset allows us to match vehicles based on all the available characteristics to obtain more comparable treated and control groups.

Second, we employ the DID to compare the difference in the behavioral responses before and after the reform for the treated group and the control group by exploiting the exogenous variation

in insurance incentives at the individual level generated by the reform. Blundell and Costa Dias (2000) and Smith and Todd (2005) show that combining DID and PSM can improve estimation quality significantly by matching the two groups in terms of observables or control variables. We demonstrate that the enforcement of the new experience-rating scheme based on past claims and on traffic violations in repeated insurance contracts has a significant impact on claim frequency. A DID study does not provide valid estimations when the time trend of claim frequency of the treated group differs significantly from that of the control group in the pre-reform period. We demonstrate that the assumption of a parallel trend holds for our data. We further limit our sample to vehicles that never change ownership during the study period, to better identify the moral hazard effect. The results remain consistent with this interpretation of our empirical results. We also show that the reform reduced road claims in the treated city for a larger sample of vehicles not restricted to be present during the four years of the study. Although this causal result cannot be interpreted as a moral hazard effect, it can be very useful to policy makers.

Third, we obtain important policy implications by conducting a detailed analysis of the treatment effects of insurance incentives on road safety. The unique setting of the two-stage reform in China allows us to compare the incentive effects of insurance pricing based on past claims and on traffic violations. We find that the effects of the programs are both economically and statistically significant. However, the treatment effects are heterogeneous based on vehicles' values (a proxy for wealth) and on insured drivers' past claims. Specifically, the reform is significantly effective for less wealthy insured drivers and insured drivers with low past risk experience, but ineffective for wealthy insured drivers and insured drivers with high past risk experience (reckless drivers). By quantifying the causal relationship between insurance incentives and road safety for different insured driver groups, our findings can help policy

6

makers design more effective experience rating schemes.

The paper proceeds as follows. Section 2 discusses the institutional background of the research setting and introduces the regulatory reform in vehicle insurance pricing. Section 3 presents the data and the DID methodology. Section 4 reviews the main estimation results. Section 5 presents the heterogeneity and robustness tests and Section 6 concludes the paper. An online appendix provides additional results.

II.   Institutional background and regulatory reform

Institutional background

All vehicle insurers in China offer almost identical contract options to the market under strictly regulated pricing rules. The vehicle insurance market consists of two parts: 1) compulsory third-party liability insurance and 2) commercial insurance. We investigate the commercial insurance market. The four main lines of commercial insurance are vehicle damage and loss insurance, theft insurance, third-party liability insurance, and driver and passenger liability insurance. As in many other countries, insurers in China use both a priori pricing and a posteriori pricing. A priori pricing is based on observable variables, whereas a posteriori pricing is based on a bonus-malus coefficient (BMC). Insurers that use a priori pricing compute the base premium at the start of the contract given observables such as the age and value of the vehicle. Drivers' characteristics are not widely considered in insurance pricing in this country. Before the reform, when contracts were renewed, insurers revised the base premium using the BMC adjusted on past claims, which was designed to work as an incentive scheme for safe driving.

Although the experience-rated premium has long existed in Chinese vehicle insurance, its efficiency was questioned. Given the fierce competition among vehicle insurers and the lack of an obligation to share claims information in a common platform, insurers were not committed to

7

enforcing the BMC because insured drivers could easily escape the cost of their bad record by switching to another insurer or even stay with the same insurer without any punishment. Dionne and Ghali (2005) analyzed the impact of introducing experience rating on road safety in Tunisia, and found that the effect is not statistically significant. One possible explanation for this result is that the new experience rating scheme has not been put into best practice because there is no sharing of information on past experience of insured drivers between insurers, and, more importantly, no commitment to use past driving experience for pricing.

Regulatory reform

To enhance incentives and fairness in vehicle insurance, a pricing reform was introduced in the vehicle insurance market of Shenzhen, a city in the province of Guangdong, China. This regulatory reform established a new vehicle insurance pricing mechanism based on insured drivers' past claims and on traffic violations. This reform applies to all individual vehicles in this city. The pricing mechanism in other cities in Guangdong remains unchanged. The reform city represents the treated city, and we consider the city of Foshan, also located in Guangdong, near the pilot city, as the control city. Table O1 in the online appendix shows that the two cities are comparable in terms of most of their pre-existing characteristics in the year prior to the reform.

The treated city introduced the first-stage reform on March 1, 2011 to improve the effectiveness of BMC pricing. In both cities, insured vehicles start at a BMC equal to one when they are insured for the first time. The two continuous lines in Figure 1 represent the BMC based on number of claims during the previous year, and the four discrete symbols at 0 on the left-hand side of the figure represent the BMC based on zero claims in the last two or three consecutive years in the two cities. Before the reform, the multiplicative coefficient to the base premium ranged from 0.7 (●) to 1.3 for both the treated and control cities. These numbers correspond to

those of the control city in the figure. Specifically, as Figure 1 indicates, the BMC based on number of claims in the last year ranged from 0.9 to 1.3. For no claims in the last two consecutive years the BMC was 0.8 (✳), and it was 0.7 (●) for no claims in the last three consecutive years.

[Figure 1 about here]

The first-stage reform changed the BMC in the treated city to a bigger range, from 0.5 (▲) to 2.0. The BMC based on number of claims in the last year ranges from 0.6 to 2.0. For no claims in the last two and three consecutive years the BMC decreases to 0.55 (■) and 0.50 (▲) respectively. The slope of the BMC line is much steeper and the malus (bonus) points are much higher (lower) for the treated city than for the control city, which suggests that the treated group faces bigger incentives for careful driving than the control group after the first-stage reform.

Another important characteristic of the new BMC in the treated city is the introduction of the commitment for all insurers to use past claims for insurance pricing. The regulatory commission requires all the insurers in the treated city to share all the claims records of their insured drivers on a common platform, and the BMC now follows the insured driver even if the insured driver switches to another insurer, as in France (Dionne et al., 2013a).

Table 1 presents the penalty multiplier (%) related to 7 levels of traffic violations for insurance pricing stipulated by the second-stage reform in the treated city. This second pricing system has no cumulative memory over time in the sense that only traffic violations committed during the previous year matter. The malus coefficient (MC) is multiplicative of the premium based on claims; it is equal to 1 plus the applicable cumulative penalty multiplier (%) during the previous year on the 7 levels of traffic violations with a maximum value. The cumulative penalty multiplier reaches its maximum at 50%, which corresponds to a maximum MC of 1.5. During the

9

first year of application, the individual's cumulative traffic violations were taken from the starting date of the second-stage reform to the starting date of the insurance contract. In the subsequent insurance periods, the traffic violations over the past 365 days are used for insurance pricing. The regulatory commission also requires all the vehicle insurers in the treated city to use publicly available information on past infractions to price insurance after the second-stage reform, as in the new BMC scheme of the first-stage reform. This represents the first time that the insurance premium in the Chinese vehicle insurance market is legally adjusted according to the record of insured drivers' traffic violations.

[Table 1 about here]

Claim frequency during the studied periods

We can summarize the two-stage reform as follows. The main administrative steps of the reform are illustrated in Figure 2, where the two vertical lines represent the start of each stage. During the pre-reform period the treated and control groups have the same BMC range of [0.7,1.3] and the insurers did not share the claim information on a common platform. After the first-stage reform on March 1, 2011, the BMC is steeper in the treated city ([0.5,2]) than in the control city ([0.7,1.3]) and no-claim records during the preceding two or three years are better rewarded. Moreover, all the insurers in the pilot city are now required by law to share insured drivers' claims records through a new Vehicle Insurance Information Exchange Platform, and to use this information for insurance pricing according to the new BMC formula. The second-stage reform started on Oct 15, 2011. Since that date, pricing depends not only on insured drivers' past claims but also on their past traffic violations. The additional multiplicative malus coefficient (MC) to the total premium based on past claims (base premium multiplied by BMC) varies from 1.0 to 1.5 depending on the seriousness of cumulative traffic violations during the previous year.

In Figure 2, we also label the dates for the public hearing, the announcement dates of the two-stage reform, and the introduction of the new subway lines in the treated city on the X-axis. The public hearing was a public consultation to collect ideas from the population during the period of Nov 4, 2010 to Nov 25, 2010. There was no policy announcement during this period. The two announcement dates were Feb 22, 2011 and Oct 11, 2011, only a few days before the starting dates of the two-stage reform. The Insurance Association announced the full content of the two-stage reform on the announcement date of the first-stage reform on Feb 22, 2011: the new BMC based on past claims started from the first-stage reform, and the introduction of the supplementary MC based on traffic violations started in the second stage, at a date not specified during the announcement of the first-stage reform.

[Figure 2 about here]

The proximity of the two announcement dates to their respective reform starting dates suggests that our data do not contain announcement effects, as in the study by Agarval and Qian (2014). Subway[4] line 2 in the treated city opened on Dec 28, 2010, and subway lines 3 and 5 opened simultaneously on Jun 28, 2011. Figure 2 also depicts the average claim frequency by group and by period based on the due date of each policy. Once is for at least one claim during an insurance period of one year. To facilitate visual comparison of claim distributions, each point plotted in Figure 2 represents a 30-day moving average of claim frequency surrounding a specific date. Over time, we observe a downward trend for claim frequency for both groups.

Comparisons of the raw data in Figure 2 indicate no significant difference for the time trend of

---

[4] The opening of a new subway would bias our estimations for the effects of insurance incentives on road safety upward if drivers choose to take the subway, which may reduce the claim frequency. We test the parallel trend between the two groups during the pre-reform period, and the results show that the two groups share a common trend. Subway line 2 was put into use in the treated city before the reform, so we conclude that the effects of the subway on claim frequency are negligible in our data. We used the interactions of the dummies for the subway opening dates and the treated city to run baseline regressions and found no significant effect.

claim frequency across the treated and control groups during the pre-reform period (we will revisit the assumption of the parallel trend in section IV by testing it empirically). After both the first-stage and second-stage reform, claim frequency in both groups declines continuously, while in the treated group it starts to deviate from the common trend with the control group by showing a greater drop. We obtained a similar result with a count variable measuring the total number of claims. This is consistent with our expectations that the new insurance incentives introduced in the treated city would reduce claim frequency accordingly.

It is important to know the weight of the BMC and MC values in the total insurance premium to get a sense of the cost differences for insured drivers before and after the reform in the two cities. We take the average of the base premium as an example. Table 2 reports the coefficient range and the actual premium range for the two groups during the pre-reform, when first-stage reform and second-stage reform periods are treated independently (there is no cumulative effect). It suggests that the relative differences for the premium ranges in the two groups may seem large enough to stimulate safe driving in the treated city after both the first-stage and second-stage reform.

[Table 2 about here]

III.   Data and methodology

Data

The data for the treated and control groups were obtained from a large vehicle insurer that operates in both cities, whose written premiums accounted for about 30% of their respective local markets in 2012. Some general statistics about the insurer and the local markets can be found in Table O2 of the online appendix. We observe that the insurer's loss ratios are not very different from those of the local markets.

We obtained the complete set of commercial vehicle policies (the whole sample) from the company's call center, which manages more than half of the company's total individual policies. The data span the years 2009 to 2012, and the insurer issues the annual policies on a daily basis from Jan 1, 2009 to Dec 31, 2012. In order to use the claims in the analyses, the real study period is from Dec 31, 2009 until Dec 30, 2013. We therefore perform our regressions and define the dummies for reform periods based on the end date of each insurance policy. During this four-year-period, insured drivers could join or leave the insurer freely. We keep only the vehicles that stay with this insurer for four consecutive years.[5] This balanced panel sample includes data on 20484 vehicles and 81936 insurance policies after addressing potential attrition issues [6] and excluding missing values, accounting for 9.3 percent of the whole sample. Descriptive statistics for the study sample and the whole sample are presented in Table O3 of the online appendix.

Each observation is a one-year vehicle insurance policy. Our sample contains detailed policy underwriting information and at-fault claims records. The underwriting data concern vehicles' characteristics such as cargo capacity (load), age, value, base premium, and type of vehicle. The claim frequency data record the claims for which insured drivers are fully or partially responsible. Therefore, the claims for which the third parties' insurers are fully responsible will not bias our estimations. For bonus-malus management, the insurer treats fully and partially responsible claims in the same manner.

DID matching estimators

One potential concern is that the control vehicles differ from the treated vehicles along most of the observable dimensions (see full sample in Panel A of Table O4 in the online appendix). For

---

[5] In section IV we further limit the population to vehicles that never change ownership before and after the reform. We also analyze the effect of the reform without these restrictions using the whole sample.

[6] We drop observations for two vehicles in our study sample that switched between the treated and control cities within the same insurer.

robustness, we use the PSM methodology to balance the two groups in terms of observables. We first estimate the propensity score or the likelihood of being treated by using a standard logit model in which the dependent variable equals 1 if a vehicle belongs to the treated group and 0 otherwise. We include all available covariates as matching variables (see Table A1) that may affect both the treatment and the claim frequency.

We impose the assumption of common support by removing all the observations outside the overlapping region of propensity scores for the two groups (bandwidth of 0.06). Panel A of Table O4 shows that after matching, 17931 vehicles are within the range of common support in the treated group and 2529 in the control group. We kept 99.9% of the vehicles in the full sample to obtain the matched sample. The balancing test in Panel A of Table O4 affirms that we achieve covariate balances between the matched treated and control groups for most of the dimensions we used.

Panel B of Table O4 reports DID matching estimators for Once (at least one claim) and Number (the total number of claims) by comparing differences in the pre- and post-reform claim frequency for the treated group with weighted averages of differences in the pre- and post-reform claim frequency for the control group based on matching obtained from characteristics in Panel A. Panel B indicates that Once and Number estimators drop by 0.035 and 0.067 respectively after the first-stage and second-stage reform in the treated group; they are both significant at the 1% level.

DID methodology

The DID approach allows us to test the dynamics of the treatment effect induced by each stage of the reform on road safety. We expect to observe a relative decline in claim frequency in the treated group compared with the control group after both stages of the reform. Equation (1)

presents our basic regression equation:

$$Claims_{it} = w + \sum_{s=1}^{2} \beta_s treat_i \times after_{st} + u_i + \eta_t + \varepsilon_{it} \qquad (1)$$

where $s = 1$ and $s = 2$ denote the first-stage and second-stage reform respectively. β_1 and β_2 are the main parameters of interest: they evaluate the differential effects of the two-stage reform at each stage across the treated and control groups. $Claims_{it}$ measures claim frequency for vehicle i in year t from Dec 31, 2009 until Dec 30, 2013. Given that the insurers can only observe claims instead of accidents, the variables Once (a dummy variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise) and Number (a count variable that represents the number of claims during the insurance period) serve as proxies for accident frequency. w is a constant term. $treat_i = 1$ when the policy is issued in the treated city, and 0 otherwise. $after_{1t} = 1$ and $after_{2t} = 1$ when the due date of the policy is after the first-stage and second-stage reform respectively, and 0 otherwise. u_i is the vehicle individual fixed effects, the proxy for driver fixed effects, which controls for vehicle/driver-level time-invariant heterogeneity. η_t is the year fixed effects, which accounts for the common aggregate shocks. During our study period, the government continually strengthened the road safety regulations nationwide,[7] which justifies controlling for the time fixed effects in the model.

In addition, to verify the parallel trend in claim frequency across these two groups before the reform, we estimate the time-varying effects of the reform by year using the following distributed lag model in Equation (2).

---

[7] For instance, an amendment for the provisions on the application and use of motor vehicle driving licenses came into force on April 1, 2010. One of the most important revisions was to increase the demerit points for serious traffic violations. On May 1, 2011, China began imposing criminal punishments on drivers found guilty of drunk driving.

$$Claims_{it} = w + \sum_{y=-1}^{2} \beta_y treat_i \times year_{yt} + u_i + \eta_t + \varepsilon_{it} \qquad (2)$$

where $year_{yt}$ represents a set of 4 yearly dummies from one year before the reform until two years after the first-stage reform year. β_{-1} measures the differential trend in claim frequency across the treated and control groups during the pre-reform period. β_0 captures the immediate responses of insured drivers in claim behaviors after the initiation of the reform. β_1 and β_2 give the additional marginal responses one year and two years after the reform, respectively.

We further estimate the dynamics of the responses in a slightly disaggregated version of Equation (2) to get the following specification.

$$Claims_{it} = w + \sum_{q=-2}^{11} \beta_q treat_i \times quarter_{qt} + u_i + \eta_t + \varepsilon_{it} \qquad (3)$$

where $quarter_{qt}$ represents a set of 14 quarter dummies for two quarters before the reform until eleven quarters after the first-stage reform quarter. For robustness, we also perform the parallel trend analysis with three quarters before the reform and semester periods.

IV.   Main results

Baseline estimates

We use both the full sample and the matched sample to estimate the separate effect of each stage of the reform on road safety. Table 3 confirms the effects of insurance incentives on road safety.[8] Columns (1) and (2) report the results for Once, and columns (3) and (4) do the same for Number. Columns (1) and (3) present the full sample results whereas columns (2) and (4) are related to the matched sample results. The conclusions are identical regardless of the sample utilized and the

---

[8] The results are robust to the whole sample analysis (883207 observations) as presented in Table O5 of the online appendix. These results cannot be interpreted as a pure measure of moral hazard because we do not control for the same vehicles before and after the two-stage reform.

definition of claim frequency. Both the first-stage and second-stage reforms are significant in reducing claim frequency. For the matched sample, Once in the treated group dropped by 0.018 and 0.017 (relative to the average of 0.463 during the pre-reform period) after the first-stage and second-stage reform respectively. For Number the drops are 0.049 and 0.041 (relative to the average of 0.750 during the pre-reform period) respectively.

These relative reductions of about 6.5% and 5.5% in the number of claims after the first-stage reform and second-stage reform respectively are both statistically and economically significant. They represent a total reduction of about 12% in the number of claims after the two-stage reform. This number is comparable with the 15% reductions in traffic violations obtained by Dionne et al. (2011) and with the 9-34 percent reduction obtained by Abay (2018) for speed related offences, according to the number of demerit points accumulated.

Underreporting

Because the insurer observes only the claims and possibly not all accidents, insured drivers may have chosen to underreport some past (minor) accidents in order to avoid an increase in their premium (Cohen 2005; Robinson and Zheng, 2010). In fact, insured drivers have a greater incentive to underreport past claims after the first-stage reform than before because of the new commitment rule and the steeper BMC. Given that the observed distribution of claims is a truncation of the true accident distribution, the observed effects of insurance pricing based on past claims may be biased (Chiappori, 2000; Dionne et al., 2013a). This potential bias should not be important in our data because we are limited to at-fault claims. In contrast, by definition, insurance pricing based on past traffic violations alone in the second stage cannot introduce a potential for insured drivers to underreport traffic violations. The F-test, shown in the last row of Table 3, reports whether the effects of the second-stage reform are significantly weaker than

those of the first-stage reform. The insignificant results indicate that the effects of road safety improvement induced by the first-stage reform are not statistically different from those of the second-stage reform, so we conclude that there should not be significant claim underreporting in the first-stage reform.

[Table 3 about here]

Validity of natural experiment

The validity of a natural experiment rests on the assumption of the parallel trend shared by the treated and control groups during the pre-reform period.[9] In this section, we explicitly test whether the claim frequency for the two groups behaved similarly in the absence of the treatment. We start with the analysis of the annual effects of the reform. Table 4 reports the annual time-varying effects of the reform, and specifically the test for differential trends between the two groups before the reform. The insignificant coefficients on the interactions of the treated group and one year before the reform in all model specifications consistently corroborate the key assumption of a common trend for the two groups in the year preceding the reform. The annual coefficients began to be significantly negative one year after the reform and remain negative and statistically significant two years after the reform. The excluded period corresponds to 2595 insurance policies. Because the excluded period in this annual analysis is rather short and contains few observations, we now consider semester and quarterly analyses.

[Table 4 about here]

Results with two pre-reform quarter coefficients in Figure 3 confirm the main results in Table 4. The excluded period has a total of 12600 insurance policies. Details of the regression analysis are presented in column (1) of Table O6 in the online appendix. Estimates of quarter coefficients

---

[9] See Giorcelli and Moser (2018) for a recent example of the parallel trend analysis.

indicate that the claim frequency before the first-stage reform is not statistically different between the two groups. The decline in claim frequency in the treated group starts three quarters after the first-stage reform quarter, $q = 0$, and the drop remains statistically significant after that in almost all quarters. The non-significance of some parameters after the reform can be explained by the high volatility of claims over short periods, as observed in Figure 2. As reported in column (1) of Table O6, the F-test indicates that we can rule out that post-reform parameters are jointly equal to 0 at the 1% level, in accordance with the results in Table 4. We also did the analysis with three excluded quarters and the results are robust, as shown in the online appendix (Table O6, column (2), and Figure O1). Finally, we performed an analysis with semesters and the results are the same regarding the validity of our natural experiment. Details are presented in Figure O2 of the online appendix.

[Figure 3 about here]

Further evidence of moral hazard

Up to now, we have retained the same vehicles before and after the reform in order to limit the adverse selection effect. However, owners with bad past claims histories may sell their vehicles to new owners, whom believe they are better drivers, to escape a large increase in their insurance premium, particularly in the treated city. The changes in ownership may bias the estimations for moral hazard in Table 3.

During our study period, 142 vehicles change ownership for the full sample. We delete them to run the same regressions as in Table 3. Results in Table 5 are identical to those in Table 3, so we conclude that our main policy result corresponds to a reduction in moral hazard in the treated city.

[Table 5 about here]

19

V. Heterogeneity analysis

We conduct a series of heterogeneity tests based on vehicles' characteristics (Table O7) and on insured drivers' past claims history (Table O8) to confirm the validity of our empirical findings. As we will see, the road safety reform effects are largely driven by less wealthy insured drivers and insured drivers with low risk. For rich people and reckless drivers, the reform has no effect.

Heterogeneous effects based on vehicles' characteristics

We now study the heterogeneous responses of drivers to the reform across different vehicles. The literature indicates that the capacity of monetary instruments to deter traffic violations is expected to vary depending on vehicle owners' income and wealth (Bar-Ilan and Sacerdote, 2004; Polinsky, 2006; Polinsky and Shavell, 1991). A natural question that follows is whether the impact of insurance incentives differs depending on the insured drivers' levels of income and wealth. Given that we do not have access to data on income and wealth, we consider the value of the vehicle (the purchase price of a new vehicle) as a proxy variable for the wealth of the insured driver. Using the median value of the vehicles in 2009 (100000 CNY), we split the matched sample ownership sample into two subsamples, namely low value and high value, to investigate the differential impact of the reform. Columns (1) and (2) of Table O7 in the online appendix report the results for the low-value (mean = 67382 CNY) and the high-value (mean = 167698 CNY) sample respectively. The results indicate that claim frequency measured by Once drops both after the first-stage and second-stage reform at the 10% significance level for the low-value sample. For the high-value sample the results are no longer statistically significant, which suggests that less wealthy people (low-value sample) respond more to the reform than wealthy people do (high-value sample). This result is partly consistent with the study by Bar-Ilan and Sacerdote (2004), which uses traffic data to test whether the deterrent effects of fines vary with

personal characteristics and finds that the elasticity of violations is smaller for wealthier drivers.

Similarly, we use the median age of vehicles in 2009 to split the matched same ownership sample into two subsamples of low age and high age, to test the potential heterogeneous treatment effects between them. Columns (3) and (4) of Table O7 report the results for the low-age and high-age sample respectively. Only the estimation for the low-age sample in Table O7 confirms our findings. The low-age sample responds more to the reform than the high-age sample. This may be due again to the difference in the average value of the vehicles (116869 CNY for low age and 121480 CNY for high age).

We then study the heterogeneous responses by owners with different types of vehicles. Columns (5) and (6) of Table O7 show the results for owners of regular automobiles and owners of trucks or minibuses respectively. The results for the regular automobiles confirm the main results in Table 3 again, whereas the effects of the reform disappear for trucks and minibuses. Note that the number of observations is much smaller for trucks and minibuses.

Based on the origin of the vehicles, we further split the matched same ownership sample into two subsamples of domestic and imported vehicles. The results are consistent for domestic vehicles, but the results are no longer statistically significant for imported ones. The mean values for domestic and imported vehicles are 112131 CNY and 346649 CNY respectively. If we regard the origin of the vehicle as the proxy for owners' wealth, the results in Columns (7) and (8) in Table O7 are highly comparable to the results we obtained based on vehicle value. There are only 610 imported vehicles, however.

Heterogeneous effects based on insured drivers' past claims history

In the previous regressions we grouped all the insured drivers together regardless of their heterogeneous claims history. Is it possible that the incentive effect of the reform varies across

insured drivers with different claims histories? This question has often been discussed in the literature. Insurance pricing based on past claims may not be a very effective tool for designing an optimal road safety policy because it may not create the appropriate incentives for reckless drivers (Sloan et al., 1995). Bourgeon and Picard (2007) show how point-record drivers' licenses provide incentives for road safety among normal drivers (those who respond to the usual incentive schemes) when the judicial system or insurance markets fail to provide optimal incentives. For example, point-record drivers' licenses may allow the regulator to incapacitate reckless or wealthy drivers when the BMC is bounded above. It is therefore important to verify how the new pricing scheme based on past claims and traffic violations creates adequate incentives for all risk categories.

We split the matched same ownership sample into the following three subsamples according to the number of claims in 2009 policies. 1) High-bonus sample: insured drivers who filed no claim or only one claim. Treated insured drivers face a lower bonus coefficient after the reform, as shown in Figure 1. 2) Equal BMC sample: insured drivers who filed two or three claims. Treated and control insured drivers face the same malus coefficient for two or three past claims before and after the reform, but the treated group faces more incentives both for more and less claims, as shown in Figure 1. 3) High-malus sample: insured drivers who filed more than three claims. Treated insured drivers face a higher malus than the control group for past claims after the reform.

We rerun the regressions using Equation (1) with the three subsamples separately, and the results are shown in Table O8. We find that the effects of the reform are significant only for the High bonus sample and Equal BMC sample, namely the less risky insured drivers. The two stages of the reform have similar effects to the results in the baseline regressions. The treated insured

drivers of the high bonus group want to keep their bonus. For the Equal BMC group, the treated insured drivers have more incentives than the control insured drivers do because they anticipate that having four claims will increase their malus more significantly. They may also want to take advantage of the bonus that the new reform offers. Reckless drivers do not change their behaviors after the reform, possibly because their marginal BMC or MC is not very sensitive to the number of claims.

VI.   Conclusion

This article provides evidence of the causal effect of moral hazard on claim frequency in China. A key contribution of our analysis is that we identify the effects of insurance incentives on road safety by controlling for potential endogeneity. To establish the desired causality, we exploit an exogenous insurance pricing reform introduced in a pilot city in China as a natural experiment, and then employ the DID approach to address some fundamental identification challenges found in the literature.

We obtained data from an insurer that operates in both the treated and control cities. Prior to the reform, the pricing mechanism was the same in both cities. We find strong behavioral effects arising from the reform. The results indicate that with industry commitment, the addition of an experience-rated premium based on past claims and on traffic violations reduces the number of claims by 12%. The F-test demonstrates that the effects of an experience-rated premium based on past claims and on traffic violations are not statistically different, which is indirect evidence of no severe underreporting of claims in the data. We performed a test for parallel trends over the pre-treatment period. The results suggest that our control group is a satisfactory reference group to estimate the treatment effects of the reform. We first restricted our study to the same vehicles over the period of analysis and to those that never change ownership during the same period to

isolate the moral hazard effects on road safety. We also estimated the effects of the reform on the whole sample of vehicles in the insurer's portfolio and obtained similar results on the reduction of claims. However, this last result cannot be interpreted as a pure reduction of moral hazard.

A set of robustness tests corroborates our identification strategies. Finally, we demonstrate that the treatment effects are heterogeneous across different vehicles and insured drivers with different past claims history. Less wealthy insured drivers respond to the insurance stimulus more than wealthy insured drivers do. Further, the reform is effective in reducing claim frequency for insured drivers with relatively good claims records, yet there is no effect for riskier insured drivers. Conducting a thorough and comprehensive economic analysis of insured drivers' behavioral responses would shed light on the feasibility of further targeted reform based on the characteristics of vehicles and insured drivers, in order to improve road safety more effectively.

References

Abay, K.A. (2018), How Effective Are Non-Monetary Instruments for Safe Driving? Panel Data Evidence on the Effect of the Demerit Point System in Denmark, *Scandinavian Journal of Economics 120*, 894-924.

Abbring, J., Chiappori, P.A., and Pinquet, J. (2003), Moral Hazard and Dynamic Insurance Data, *Journal of the European Economic Association 1*, 767–820.

Agarwal, S. and Qian, W. (2014), Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore, *American Economic Review 104*, 4205-4230.

Arrow, K. (1963), Uncertainty and the Welfare Economics of Medical Care, *American Economic Review 53*, 941-973.

Bar-Ilan, A., and Sacerdote, B. (2004), The Response of Criminals and Noncriminals to Fines, *Journal of Law and Economics 47*, 1-17.

Blundell, R., and Costa Dias, M. (2000), Evaluation Methods for Non-experimental Data, *Fiscal Studies 21*, 427-468.

Bourgeon, J.M., and Picard, P. (2007), Point-Record Driver's License and Road Safety: An Economic Approach, *Journal of Public Economics 91*, 235-258.

Castillo-Manzano, J. I., and Castro-Nuño, M. (2012), Driving Licenses based on Points Systems: Efficient Road Safety Strategy or Latest Fashion in Global Transport Policy? A Worldwide Meta-analysis, *Transport Policy 21*, 191-201.

Chiappori, P.A. (2000), Econometric Models of Insurance under Asymmetric Information. In: Dionne, G. (Ed.), Handbook of Insurance. Boston: Kluwer Academic Publishers, pp. 365-394.

Chiappori, P.A., and Salanié, B. (2000), Testing for Asymmetric Information in Insurance Markets, *Journal of Political Economy 108*, 56-78.

Chiappori, P.A., and Salanié, B. (2013), Asymmetric Information in Insurance Markets: Predictions and Tests. In: Dionne, G. (Ed.), Handbook of Insurance. New York: Springer, pp. 397-422.

Cohen, A. (2005), Asymmetric Information and Learning in the Automobile Insurance Market, *Review of Economics and Statistics 87*, 197-207.

Crocker, K., and Snow, A. (1985), The Efficiency of Competitive Equilibria in Insurance Markets with Asymmetric Information, *Journal of Public Economics 26*, 207-219.

Crocker, K., and Snow, A. (1986), The Efficiency Effect of Categorical Discrimination in the Insurance Industry, *Journal of Political Economy 94*, 321-344.

De Paola, M., Scoppa, V., and Falcone, M. (2013), The Deterrent Effects of the Penalty Points System for Driving Offences: A Regression Discontinuity Approach, *Empirical Economics 45*, 965-985.

Dionne, G., and Ghali, O. (2005), The (1992) Bonus-Malus System in Tunisia: An Empirical Evaluation, *Journal of Risk and Insurance 72*, 609-633.

Dionne, G., Gouriéroux, C., and Vanasse, C. (2001), Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment, *Journal of Political Economy 109*, 444-453.

Dionne, G., Michaud, P.C., and Dahchour, M. (2013a), Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France, *Journal of the European Economic Association 11*, 897-917.

Dionne, G., Michaud, P.C., and Pinquet, J. (2013b), A Review of Recent Theoretical and Empirical

Analyses of Asymmetric Information in Road Safety and Automobile Insurance, *Research in Transportation Economics 43*, 85-97.

Dionne, G., Pinquet, J., Maurice, M., and Vanasse, C. (2011), Incentive Mechanisms for Safe Driving: A Comparative Analysis with Dynamic Data, *Review of Economics and Statistics 93*, 218-227.

Giorcelli, M., and Moser, P. (2018), Copyrights and Creativity: Evidence from Italian Operas, Working paper, UCLA, 42 pages.

Heckman, J.J., Ichimura, H., and Todd, P.E. (1997), Matching as An Econometric Evaluation Estimator: Evidence from Evaluating A Job Training Programme, *The Review of Economic Studies 64*, 605-654.

Heckman, J.J., Ichimura, H., and Todd, P. (1998), Matching as An Econometric Evaluation Estimator, *The Review of Economic Studies 65*, 261-294.

Hoy, M. (1982), Categorizing Risk in the Insurance Industry, *Quarterly Journal of Economics 97*, 321-336.

Israel, M. (2007), Do We Drive More Safely when Accidents are More Expensive? Identifying Moral Hazard from Experience Rating Schemes, Working paper, Wharton School.

Lee, Y.W. (2013), Testing for the Presence of Moral Hazard Using the Regulatory Reform in the Car Insurance Market: Case of Korea, *Japanese Economic Review 64*, 414-429.

Li, C.S., Liu, C.C., and Yeh, J.H. (2007), The Incentive Effects of Increasing Per-claim Deductible Contracts in Automobile Insurance, *Journal of Risk and Insurance 74*, 441-459.

Makowsky, M.D., and Stratmann, T. (2011), More Tickets, Fewer Accidents: How Cash-strapped Towns Make for Safer Roads, *Journal of Law and Economics 54*, 863-888.

Polinsky, A.M. (2006), The Optimal Use of Fines and Imprisonment When Wealth is Unobservable, *Journal of Public Economics 90*, 823-835.

Polinsky, A.M., and Shavell, S. (1991), A Note on Optimal Fines When Wealth Varies among Individuals, *American Economic Review 81*, 618-621.

Robinson, C., and Zheng, B. (2010), Moral Hazard, Insurance Claims, and Repeated Insurance Contracts, *Canadian Journal of Economics 43*, 967-993.

Rowell, D., Nghiem, S., and Connelly, L.B. (2017), Two Tests for Ex Ante Moral Hazard in a Market for Automobile Insurance, *Journal of Risk and Insurance 84*, 1103-1126.

Sloan, F.A., Reilly, B.A., and Schenzler, C. (1995), Effects of Tort Liability and Insurance on Heavy Drinking and Driving, *Journal of Law and Economics 38*, 49-77.

Smith, J.A., and Todd, P.E. (2005), Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics 125*, 305-353.

Vukina, T., and Nestić, D. (2015), Do People Drive Safer When Accidents Are More Expensive: Testing for Moral Hazard in Experience Rating Schemes, *Transportation Research Part A: Policy and Practice 71*, 46-58.

Wang, J.L., Chung, C.F., and Tzeng, L.Y. (2008), An Empirical Analysis of the Effects of Increasing Deductibles on Moral Hazard, *Journal of Risk and Insurance 75*, 551-566.

Weisburd, S. (2015), Identifying Moral Hazard in Car Insurance Contracts, *Review of Economics and Statistics 97*, 301-313.

Zavadil, T. (2015), Do the Better Insured Cause More Damage? Testing for Asymmetric Information in Car Insurance, *Journal of Risk and Insurance 82*, 865-889.

Table 1 – Multiplier based on traffic violations in the treated city

	Type of traffic violation	Frequency	Multiplier (%)
1	Driving the wrong way on a one-way street or backwards	three or more times	10
2	Failure to observe a traffic light	three or more times	10
3	Exceeding the speed limit by more than 50%	three or more times	10
4	Driving without a license or with a revoked license	once	30
5	Hit-and-run accident	once	30
6	Driving after drinking	once	10
7	Drunk driving	once	30

Notes: This table presents the penalty multipliers based on traffic violations stipulated by the second-stage reform. The total variation in the insurance premium in year t uses the information on traffic violations in year $t – 1$ only, along with the information on past claims in previous years. The cumulative penalty multiplier for pricing insurance on past traffic violations has a maximum of 50%. The total malus coefficient (MC) is equal to the applicable multiplier accumulated during the previous year on the 7 levels of traffic violations plus 1. Since the cumulative multiplier reaches its maximum at 50%, the maximum MC is 1.5 and its minimum is 1.0. The criteria for driving after drinking and drunk driving are [20,80) mg/100ml and no lower than 80 mg/100ml respectively for drivers' blood alcohol levels.

Table 2 – Coefficient range and actual premium range by group and by period

Period	Treated group		Control group	
	Coefficient range	Premium range	Coefficient range	Premium range
Before reform	[0.7,1.3]	[3056,5676]	[0.7,1.3]	[3056,5676]
First reform (BMC)	[0.5,2.0]	[2183,8732]	[0.7,1.3]	[3056,5676]
Second reform (MC)	[1.0,1.5]	[4366,6549]	[1.0,1.0]	[4366,4366]

Notes: We calculate the actual premium range by applying the BMC and MC to the mean of the base premium (4366 CNY) by group, by period, and by reform independently. Specifically, we apply the MC to the mean of the base premium in the treated group for the second-stage reform, supposing the BMC at the first-stage reform equals 1. When we apply the BMC, we suppose MC equals 1. The premium range in the control group is similar to that of the treated group before the reform because the BMC is not affected by the first reform in this group. Moreover, the multiplicative coefficient range of the control group for the second reform remains at [1.0,1.0] because traffic violations are not taken into account for insurance pricing in this group.

Table 3 – Effects of the two-stage reform on claim frequency

	Once	Once	Number	Number
	Full	Matched	Full	Matched
	(1)	(2)	(3)	(4)
$treat_i \times after_{1t}$	-0.019**	-0.018**	-0.049***	-0.049***
	(0.009)	(0.009)	(0.018)	(0.018)
$treat_i \times after_{2t}$	-0.016**	-0.017**	-0.040***	-0.041***
	(0.008)	(0.008)	(0.014)	(0.014)
Driver FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.046	0.046	0.056	0.056
Observations	81936	81840	81936	81840
Vehicles	20484	20460	20484	20460
F-test	0.03	0.02	0.10	0.08
(p-value)	(0.871)	(0.896)	(0.746)	(0.776)

Notes: This table reports the results of the effects of the two-stage reform on claim frequency in the OLS regression $Claims_{it} = w + \sum_{s=1}^{2} \beta_s treat_i \times after_{st} + u_i + \eta_t + \varepsilon_{it}$. Full and Matched represent the full sample and the matched sample respectively. *Treat$_i$* is a dummy variable equal to 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. *After$_{st}$* is a dummy variable equal to 1 when the end of the insurance period of the vehicle is after the first-stage reform ($s = 1$) or after the second-stage reform ($s = 2$), and 0 otherwise. The dependent variable is measured by Once or Number. Once is an indicator variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. Number counts claims during the insurance period. Standard errors clustered at the individual level are reported in parentheses. For the F-test, the *p*-values are in parentheses. F-test verifies whether the effect of the first-stage reform is the same as that of the second-stage reform. *** indicates p<0.01 and ** p<0.05.

Table 4 – Annual estimates of the effects of insurance incentives on claim frequency

	Once Full	Once Matched	Number Full	Number Matched
	(1)	(2)	(3)	(4)
$treat_i \times year_{-1}$	-0.013	-0.013	0.000	-0.001
	(0.027)	(0.027)	(0.052)	(0.052)
$treat_i \times year_0$	-0.016	-0.016	-0.048	-0.048
	(0.027)	(0.027)	(0.051)	(0.051)
$treat_i \times year_1$	-0.049*	-0.050*	-0.109**	-0.109**
	(0.027)	(0.027)	(0.052)	(0.052)
$treat_i \times year_2$	-0.066**	-0.066**	-0.090*	-0.090*
	(0.028)	(0.028)	(0.054)	(0.054)
Driver FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.046	0.046	0.056	0.056
Observations	81936	81840	81936	81840
Vehicles	20484	20460	20484	20460

Notes: This table reports the results of time-varying effects of the insurance pricing reform on road safety in the OLS regression $Claims_{it} = w + \sum_{y=-1}^{2} \beta_y treat_i \times year_{yt} + u_i + \eta_t + \varepsilon_{it}$, where $year_{yt}$ represents a set of 4 yearly dummies from one year before until two years after the first-stage reform year, $y = 0$. *Treat$_i$* is a dummy variable equal to 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. The dependent variable is measured by Once or Number. Once is an indicator variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. Number counts claims during the insurance period. The period for policies terminated from Dec 31, 2009 until Feb 28, 2010 acts as the excluded period, which represents 2595 insurance policies. Standard errors clustered at the individual level are reported in parentheses. Full and Matched represent the full sample and the matched sample respectively. ** indicates p<0.05, and * p<0.10.

Table 5 – Further evidence of moral hazard: same ownership during all years

	Once	Once	Number	Number
	Full	Matched	Full	Matched
	(1)	(2)	(3)	(4)
$treat_i \times after_{1t}$	-0.020**	-0.020**	-0.051***	-0.050***
	(0.009)	(0.009)	(0.018)	(0.018)
$treat_i \times after_{2t}$	-0.016**	-0.016**	-0.041***	-0.042***
	(0.008)	(0.008)	(0.014)	(0.014)
Driver FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R^2	0.046	0.046	0.056	0.056
Observations	81368	81276	81368	81276
Vehicles	20342	20319	20342	20319

Notes: This table reports further evidence of moral hazard on claim frequency in the OLS regression $Claims_{it} = w + \sum_{s=1}^{2} \beta_s treat_i \times after_{st} + u_i + \eta_t + \varepsilon_{it}$. We limit the study sample to vehicles that never change ownership.

Treat$_i$ is a dummy variable equal to 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. *After$_{st}$* is a dummy variable equal to 1 when the insurance period of the vehicle ends after the first-stage reform ($s = 1$) or after the second-stage reform ($s = 2$), and 0 otherwise. The dependent variable is measured by Once or Number. Once is an indicator variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. Number counts claims during the insurance period. Standard errors clustered at the individual level are reported in parentheses. Full and Matched represent the full sample and the matched sample respectively. *** indicates p<0.01, and ** p<0.05.
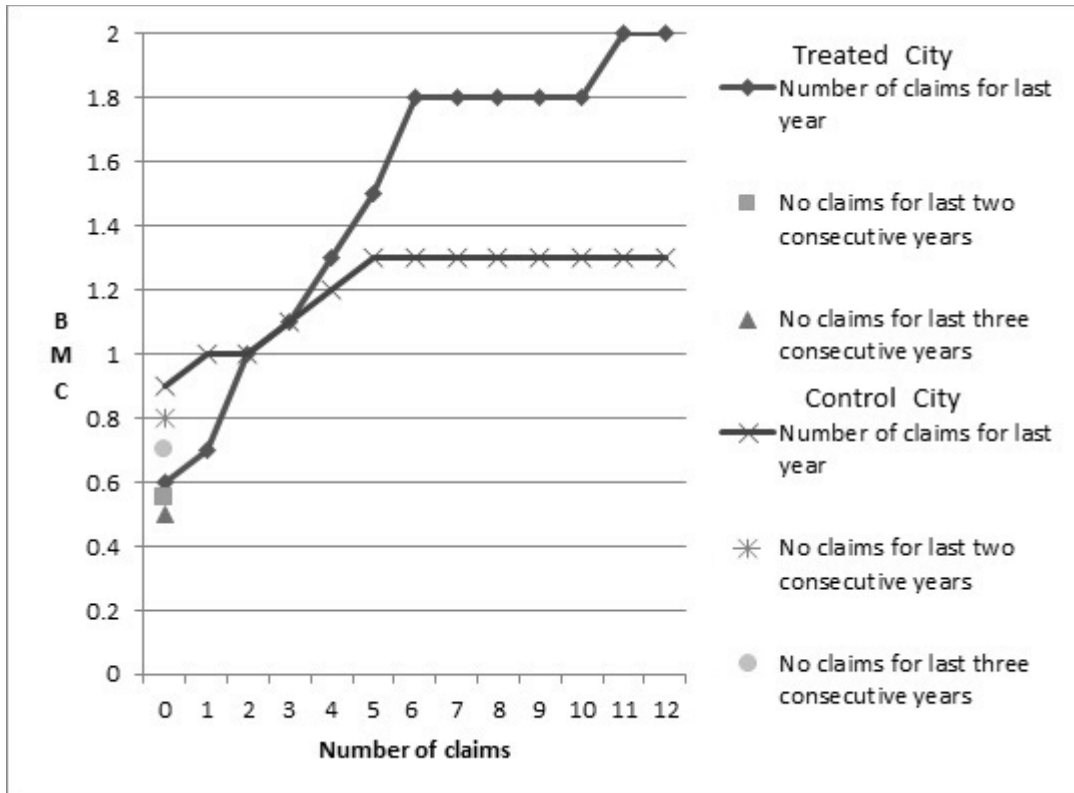
Figure 1 – BMC in the two cities before and after the first-stage reform

Notes: The two continuous lines represent the BMC based on number of claims during the previous year and the four discrete symbols on the left at zero represent the BMC based on zero claims in the last two or three consecutive years in the two cities. The BMC is equal to one for the first insurance contract of a vehicle in both cities. The BMC of the treated city was identical to that of the control city before the reform. The BMC of the control city is the same before and after the reform.

Figure 2 – Evolution of two-stage reform and claim frequency by group and period

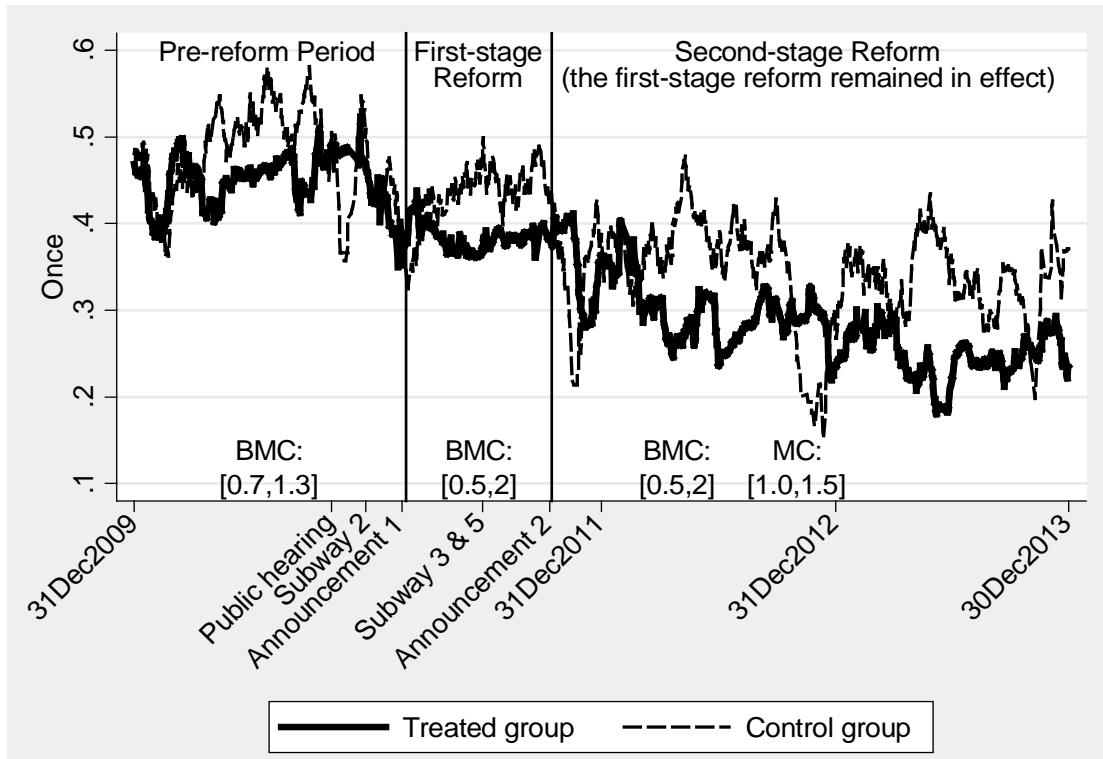*Notes:* The period of analysis is from December 31, 2009 to December 30, 2013. The two vertical lines represent the starting dates of the first-stage and second-stage reform, on March 1, 2011 and Oct 15, 2011, respectively. BMC is the bonus-malus based on past claims; MC is the total malus coefficient based on past traffic violations. Once is an indicator variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. We calculate the daily averages of Once based on the dates on the X-axis and then plot the 30-day moving averages to facilitate the visual comparison. The moving average of Once at date *d* is

$$Once(d) = 1/31 \sum_{i=-15}^{15} Once(d+i).$$ The Public hearing started on Nov 4, 2010 when the Insurance Association in the

treated city began to collect public opinions regarding the forthcoming pricing reform. Announcement 1 and 2 represent the announcement dates of the first-stage and second-stage reform on Feb 22, 2011 and on Oct 11, 2011 respectively, which were very close to their actual initiation dates. Subway 2 (Dec 28, 2010) and Subway 3&5 (June 28, 2011) are opening dates of new subway lines in the treated city.

33
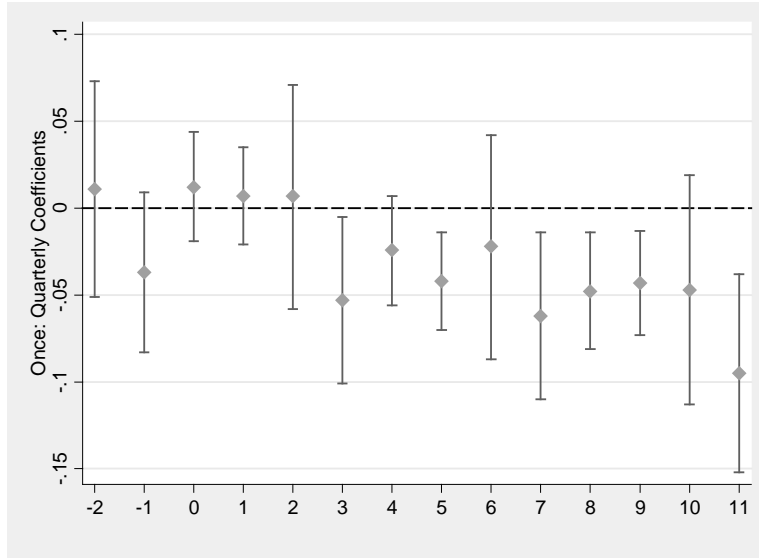
Figure 3 – Quarterly estimates of the effects of insurance incentives on road safety
with two pre-reform periods

Notes: This figure shows 95% confidence intervals for coefficients β_q using the matched sample in the OLS regression $Claim_{it} = w + \sum_{q=-2}^{11} \beta_q treat_i \times quarter_{qt} + u_i + \eta_t + \varepsilon_{it}$, where $quarter_{qt}$ represents a set of 14 quarterly dummies from two quarters before the first-stage reform quarter, $q = 0$, until 11 quarters after the reform. *Treat$_i$* is a dummy variable equal to 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. The dependent variable is measured by once. Once is an indicator variable equal to 1 when the insured has filed at least one claim during the insurance period, and 0 otherwise. The period for policies terminated from Dec 31, 2009 until Aug 31, 2010 acts as the excluded period. It contains 12600 observations. According to the F-test, we can rule out that post-reform parameters are jointly equal to zero at the 1% level.

Appendix

Table A1 – Definitions of variables

Variable	Definition
Outcome variables	
Once	A dummy variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise
Number	The number of claims during the insurance period
DID variables	
treat	A dummy variable equal to 1 when the insured driver vehicle is in the reform city, and 0 otherwise
$after_1$	A dummy variable equal to 1 when the end of the insurance period of the vehicle is after the first-stage reform, and 0 otherwise
$after_2$	A dummy variable equal to 1 when the end of the insurance period of the vehicle is after the second-stage reform, and 0 otherwise
Vehicle's characteristics	
Age	The age of the vehicle (in years)
Log(Value)	The log purchase price for a new vehicle (in thousands of CNY)
Log(Premium)	The log base premium during the insurance period (in thousands of CNY)
Load	The cargo capacity of the vehicle (in tons)
Import	A dummy variable equal to 1 when the vehicle is imported, and 0 otherwise
Nationwide driving	A dummy variable equal to 1 when the vehicle is driven within the whole country, and 0 otherwise (only driven within the province of the treated and control city)
Small truck	A dummy variable equal to 1 when the vehicle is a truck (2 tons or less), and 0 otherwise
Automobile	A dummy variable equal to 1 when the vehicle is a regular automobile (6 passengers or less), and 0 otherwise
Small minibus	A dummy variable equal to 1 when the vehicle is a minibus (7-10 passengers), and 0 otherwise
Big minibus	A dummy variable equal to 1 when the vehicle is a minibus (11-20 passengers), and 0 otherwise

Effects of Insurance Incentives on Road Safety:
Evidence from a Natural Experiment in China
Online Appendix

Georges Dionne[a]
Ying Liu[b]

[a] Corresponding author. HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine Montréal, Québec, Canada, H3T 2A7
Email: georges.dionne@hec.ca;
www.riskmanagementresearch.org

[b] School of Economics, Shandong University, No 27, Shanda South Road, Jinan, China, 250100
Email: liuying1030@sdu.edu.cn
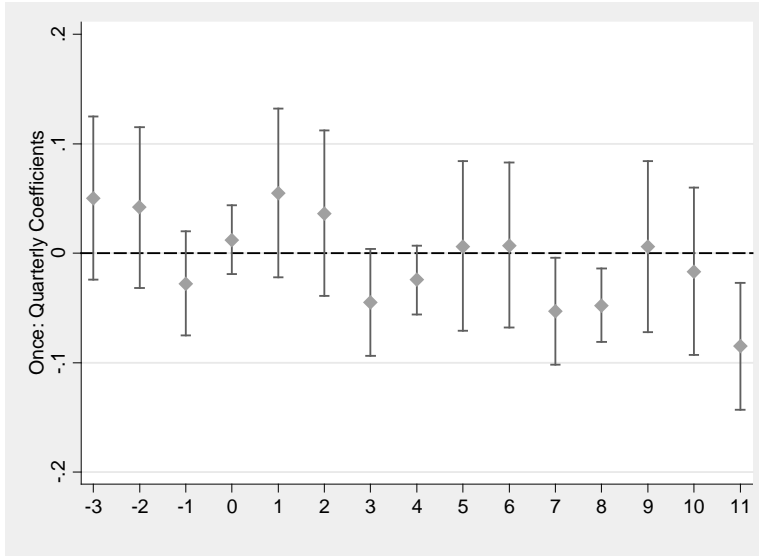
Figure O1 – Quarterly estimates of the effects of insurance incentives on road safety
with three pre-reform periods

Notes: This figure shows 95% confidence intervals for β_q coefficients using the matched sample in the OLS regression $Claim_{it} = w + \sum_{q=-3}^{11} \beta_q treat_i \times quarter_{qt} + u_i + \eta_t + \varepsilon_{it}$, where $quarter_{qt}$ represents a set of 15 quarterly dummies from three quarters before until eleven quarters after the first-stage reform quarter, $q = 0$. $Treat_i$ is a dummy variable that equals 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. $Claim_{it}$ is claim frequency. It is equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. The period for policies terminated from Dec 31, 2009 until May 31, 2010 acts as the excluded period. This period comprises 6195 insurance policies.
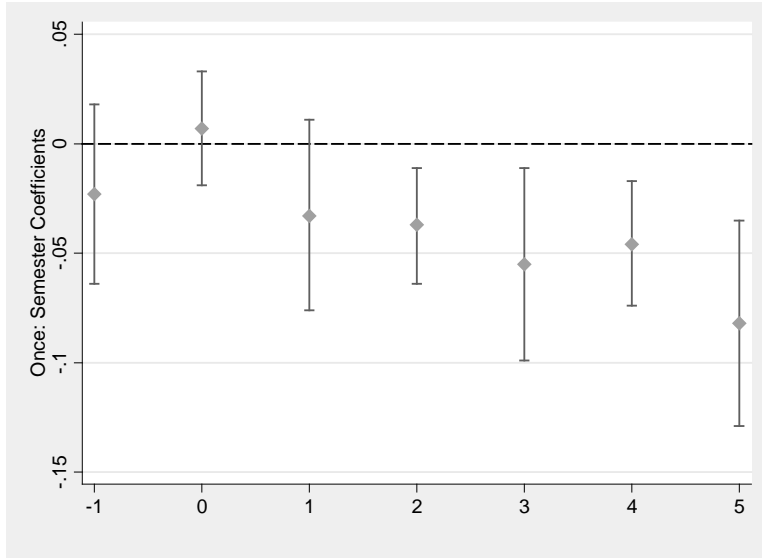
Figure O2 – Semester estimates of the effects of insurance incentives on road safety

Notes: This figure shows 95% confidence intervals for β_h coefficients using the matched sample in the OLS regression $Claim_{it} = w + \sum_{h=-1}^{5} \beta_h treat_i \times semester_{ht} + u_i + \eta_t + \varepsilon_{it}$, where $semester_{ht}$ represents a set of 7 semester dummies from half a year (one semester) before the first-stage reform semester, $h = 0$, until five semesters after the reform. *Treat_i* is a dummy variable that equals 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. *Claim_{it}* is claim frequency. It is equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. The period for policies terminated from Dec 31, 2009 until Aug 31, 2010 acts as the excluded period; the number of insurance policies during this period is 12600.

Table O1 – Comparison of treated and control city in the year prior to the reform (2010)

	Treated City	Control City
Population (in millions)	10.37	7.19
Male ratio	53.1%	50.0%
Years of education, workforce	10.93	10.0
GDP per capita (in ten thousand CNY)	10.69	9.40
Number of civil vehicles (in millions)	1.67	0.91

Notes: This table compares the two studied cities in terms of available characteristics in the year preceding the reform.

Table O2 – General statistics on the insurer studied and the local markets

	Premiums		Market share	No. of insurers	Firm's ranking	Claims		Loss Ratio	
	Firm	Market				Firm	Market	Firm	Market
Panel A: Treated city									
2009	1574.6	6791.77	23.2%	26	2	595.09	3523.56	37.8%	51.9%
2010	2387.23	8475.07	28.2%	27	1	942.85	3715.42	39.5%	43.8%
2011	2883.13	9244.27	31.2%	28	1	1251.17	4385.30	43.4%	47.4%
2012	3181.22	10333.89	30.8%	32	1	1674.37	5504.33	52.6%	53.3%
Panel B: Control city									
2009	715.11	2718.07	26.3%	19	2	263.55	1365.51	36.9%	50.2%
2010	1013.57	3467.24	29.2%	20	1	433.23	1524.58	42.7%	44.0%
2011	1240.86	4104.42	30.2%	22	1	543.97	1956.29	43.8%	47.7%
2012	1363.84	4686.54	29.1%	24	1	728.18	2348.41	53.4%	50.1%

Notes: This table compares data obtained from the insurer of the study to those of the local markets in the two cities. Premiums and claims are for vehicle insurance and are in ten thousand CNY. Market share is the firm's proportion of premiums in each market, and insurer's ranking is its ranking as a vehicle insurer in each market. Loss ratio is the ratio of total losses in claims divided by total premiums earned.

Table O3 – Summary statistics for the study sample and the whole sample

	Mean		Sd		Min		Median		Max		N	
	Study	Whole	Study	Whole	Study	Whole	Study	Whole	Study	Whole	Study	Whole
Outcome variables												
Once	0.358	0.370	0.479	0.483	0	0	0	0	1	1	81936	883207
Number	0.543	0.591	0.871	0.957	0	0	0	0	9	25	81936	883207
DID variables												
Treat	0.876	0.824	0.329	0.380	0	0	1	1	1	1	81936	883207
After_1	0.719	0.893	0.450	0.309	0	0	1	1	1	1	81936	883207
After_2	0.565	0.798	0.496	0.401	0	0	1	1	1	1	81936	883207
Vehicle characteristics												
Age	4.229	3.444	2.154	2.512	0.003	0	4.008	3	20.263	24.799	81936	883207
Log(Value)	11.516	11.486	0.574	0.637	9.770	9.002	11.473	11.435	14.457	15.701	81936	879424
Log(Premium)	8.284	8.268	0.465	0.497	5.558	4.384	8.361	8.332	10.677	11.259	81936	883207
Load	0.022	0.033	0.155	0.212	0	0	0	0	2.400	15	81936	883207
Import	0.031	0.043	0.173	0.204	0	0	0	0	1	1	81936	883207
Nationwide driving	0.991	0.989	0.092	0.103	0	0	1	1	1	1	81936	883207
Small truck	0.024	0.033	0.154	0.178	0	0	0	0	1	1	81936	883193
Automobile	0.867	0.830	0.339	0.375	0	0	1	1	1	1	81936	883193
Small minibus	0.106	0.134	0.308	0.340	0	0	0	0	1	1	81936	883193
Big minibus	0.002	0.003	0.047	0.055	0	0	0	0	1	1	81936	883193

Notes: Study and whole represent the study sample for the analysis of moral hazard and the whole sample for the analysis of the reform on all drivers available in the insurer's portfolio. All variables are defined in Table A1.

Table O4 – Balancing test and matching estimators based on DID-PSM

	Full sample			Matched sample		
	Treated	Control	Diff	Treated	Control	Diff
Panel A: Balancing test						
Age	2.778	2.381	0.397*** (0.000)	2.775	2.764	0.011 (0.651)
Log (Value)	11.571	11.353	0.217*** (0.000)	11.569	11.593	-0.024*** (0.001)
Log (Premium)	8.381	7.980	0.401*** (0.000)	8.379	8.216	0.163*** (0.000)
Load	0.013	0.091	-0 .078*** (0.000)	0.013	0.015	-0.003 (0.134)
Import	0.033	0.018	0.015*** (0.000)	0.032	0.031	0.001 (0.637)
Nationwide Driving	0.991	0.995	-0.004*** (0.047)	0.991	0.993	-0.002 (0.167)
Small truck	0.015	0.089	-0.074*** (0.000)	0.015	0.017	-0.002 (0.352)
Automobile	0.881	0.773	0.107*** (0.000)	0.881	0.882	-0.001 (0.839)
Small minibus	0.102	0.135	-0.033*** (0.000)	0.102	0.099	0.003 (0.469)
Big minibus	0.002	0.002	0.000 (0.799)	0.002	0.003	-0.001 (0.466)
N	17952	2532		17931	2529	
Panel B: DID matching estimators (standard errors in parentheses)						
Once	-0.035***					
	(0.009)					
R^2	0.06					
N	40920					
Number	-0.067***					
	(0.018)					
R^2	0.07					
N	40920					

Notes: This table presents the balancing test and DID matching estimators based on data in 2009 and in 2012. Panel A reports the balancing test for the full sample and the matched sample, and Panel B gives the DID matching estimators. We employ Gaussian kernel matching within a bandwidth of 0.06. Matching estimators are the differences in the pre- and post-reform claim frequency for the treated group to weighted averages of differences in the pre- and post-reform of the claim frequency for the control group based on matching obtained from characteristics in Panel A. Both estimators are significantly negative. Panel A reports *p*-values in parentheses and Panel B gives standard errors in parentheses. The variables are defined in Table A1.

Table O5 – Main results for the whole sample

	Once	Number
	(1)	(2)
$treat_i \times after_{1t}$	-0.010**	-0.054***
	(0.005)	(0.010)
$treat_i \times after_{2t}$	-0.031***	-0.081***
	(0.003)	(0.007)
Driver FE	Yes	Yes
Year FE	Yes	Yes
R^2	0.032	0.047
Observations	883207	883207
Vehicles	535576	535576

Notes: This table reports the results of the effects of the two-stage reform on claim frequency in the OLS regression $Claims_{it} = w + \sum_{s=1}^{2} \beta_s treat_i \times after_{st} + u_i + \eta_t + \varepsilon_{it}$ for the whole sample. *Treat$_i$* is a dummy variable equal to 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. *After$_{st}$* is a dummy variable equal to 1 when the end of the insurance period of the vehicle is after the first-stage reform ($s = 1$) or after the second-stage reform ($s = 2$), and 0 otherwise. The dependent variable is measured by Once or Number. Once is an indicator variable equal to 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. Number counts claims during the insurance period. Standard errors clustered at the individual level are reported in parentheses. *** indicates $p<0.01$ and ** $p<0.05$.

Table O6 – Quarterly analysis for parallel trend

	Once	
	(1)	(2)
$treat_i \times quarter_{-5}$	Excluded	Excluded
$treat_i \times quarter_{-4}$	Excluded	Excluded
$treat_i \times quarter_{-3}$	Excluded	0.05 (0.038)
$treat_i \times quarter_{-2}$	0.011 (0.032)	0.042 (0.038)
$treat_i \times quarter_{-1}$	-0.037 (0.024)	-0.028 (0.024)
$treat_i \times quarter_{0}$	0.012 (0.016)	0.012 (0.016)
$treat_i \times quarter_{1}$	0.007 (0.014)	0.055 (0.039)
$treat_i \times quarter_{2}$	0.007 (0.033)	0.036 (0.038)
$treat_i \times quarter_{3}$	-0.053** (0.024)	-0.045* (0.025)
$treat_i \times quarter_{4}$	-0.024 (0.016)	-0.024 (0.016)
$treat_i \times quarter_{5}$	-0.042*** (0.014)	0.006 (0.040)
$treat_i \times quarter_{6}$	-0.022 (0.033)	0.007 (0.039)
$treat_i \times quarter_{7}$	-0.062** (0.024)	-0.053** (0.025)
$treat_i \times quarter_{8}$	-0.048*** (0.017)	-0.048*** (0.017)
$treat_i \times quarter_{9}$	-0.043*** (0.015)	0.006 (0.040)
$treat_i \times quarter_{10}$	-0.047 (0.034)	-0.017 (0.039)
$treat_i \times quarter_{11}$	-0.095*** (0.029)	-0.085*** (0.030)
Driver FE	Yes	Yes
Year FE	Yes	Yes
R^2	0.047	0.047
Observations	81840	81840
Vehicles	20460	20460
Number of excluded	12600	6195

	policies		
F-test for all the pre-reform parameters (*p*-value)	1.59	1.62
	(0.205)	(0.181)
F-test for all the post-reform parameters (*p*-value)	3.06***	2.95***
	(0.0003)	(0.0004)

Notes: This table reports the results of time-varying effects of the insurance pricing reform on road safety in the OLS regression $Claim_{it} = w + \sum \beta_q treat_i \times quarter_{qt} + u_i + \eta_t + \varepsilon_{it}$, where $quarter_{qt}$ represents a set of quarterly dummies from two quarters before the reform in column (1) and three quarters before the reform in column (2) until eleven quarters after the first-stage reform quarter, $q = 0$. *Treat$_i$* is a dummy variable that equals 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. $Claim_{it}$ is claim frequency. It is equal to 1 when the insured has filed at least one claim during the insurance period, and 0 otherwise. The periods for policies terminated from Dec 31, 2009 until August 31, 2010 for results in column (1) and from Dec 31, 2009 until May 31, 2010 for results in column (2) act as the excluded periods respectively. Standard errors clustered at the individual level are reported in parentheses. *p*-values for the F-test are reported in parentheses. The matched sample is employed for analysis. *** indicates p<0.01, ** p<0.05 and * p<0.10. The F-test for pre-reform parameters indicates that we cannot reject that pre-reform coefficients are jointly equal to zero while that for post-reform coefficients indicates that we can reject that they are jointly equal to zero at the 1% level.

Table O7 – Heterogeneous effects based on vehicles' characteristics

	Low Value	High Value	Low Age	High Age	Regular Automobile	Trucks & Minibuses	Domestic Vehicles	Imported Vehicles
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$treat_i \times after_{1t}$	-0.025*	-0.012	-0.027**	-0.014	-0.018*	-0.025	-0.024***	0.129**
	(0.013)	(0.013)	(0.013)	(0.013)	(0.010)	(0.022)	(0.009)	(0.055)
$treat_i \times after_{2t}$	-0.020*	-0.011	-0.019*	-0.017	-0.016*	-0.016	-0.016**	-0.045
	(0.011)	(0.011)	(0.011)	(0.010)	(0.008)	(0.020)	(0.008)	(0.040)
Driver FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.043	0.049	0.057	0.036	0.047	0.041	0.047	0.032
Observations	39316	41960	40696	40580	70540	10736	78836	2440
Vehicles	9829	10490	10174	10145	17635	2684	19709	610

Notes: This table reports the results of the heterogeneous effects of the two-stage reform on claim frequency based on vehicles' characteristics in the OLS regression $Claims_{it} = w + \sum_{s=1}^{2} \beta_s treat_i \times after_{st} + u_i + \eta_t + \varepsilon_{it}$. $Treat_i$ is a dummy variable that equals 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. $After_{st}$ is a dummy variable that equals 1 when the end of the insurance period of the vehicle is after the first-stage reform ($s = 1$) or after the second-stage reform ($s = 2$), and 0 otherwise. $Claims_{it}$ is an indicator variable that equals 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. Standard errors clustered at the individual level are reported in parentheses. We split the matched same ownership sample to obtain the eight subsamples based on vehicles' characteristics. A low value car is below 100000 CNY in 2009. The mean value of this group is 67382 CNY, while the mean value of the high value cars is 167698 CNY. A low age car is below the age median in 2009. The mean value of low age cars is 2.92 years, and the mean of high age cars is 5.53 years. The vehicle types are described in Table A1. *** indicates p<0.01, ** p<0.05, and * p<0.10.

Table O8 – Heterogeneous effects based on insured drivers' pre-reform claims history

	High bonus	Equal BMC	High malus
	(1)	(2)	(3)
$treat_i \times after_{1t}$	-0.032***	-0.052***	-0.016
	(0.010)	(0.019)	(0.072)
$treat_i \times after_{2t}$	-0.014*	-0.045**	-0.086
	(0.008)	(0.018)	(0.065)
Driver FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
R^2	0.016	0.332	0.304
Observations	64136	15872	1268
Vehicles	16034	3968	317

Notes: This table reports the results of the heterogeneous effects of the two-stage reform on claim frequency based on insured drivers' pre-reform claims history in the OLS regression $Claims_{it} = w + \sum_{s=1}^{2} \beta_s treat_i \times after_{st} + u_i + \eta_t + \varepsilon_{it}$.

Treat$_i$ is a dummy variable that equals 1 when the insured driver's vehicle is in the reform city, and 0 otherwise. *After$_{st}$* is a dummy variable that equals 1 when the end of the insurance period of the vehicle is after the first-stage reform ($s = 1$) or after the second-stage reform ($s = 2$), and 0 otherwise. *Claims$_{it}$* is an indicator variable that equals 1 when the insured driver has filed at least one claim during the insurance period, and 0 otherwise. Standard errors clustered at the individual level are reported in parentheses. We split the matched same ownership sample to obtain the three subsamples of High bonus (less than two claims), Equal BMC (two or three claims), and High malus (more than three claims) in columns (1) to (3) based on the insured drivers' claim frequency in 2009. *** indicates p<0.01, ** p<0.05, and * p<0.10.