Technical Annex to Evaluation Paper 19/1: Econometric analysis

October 2019

Introduction

FCA Occasional Paper 12 (OP 12), published in December 2015, presented the results of a series of randomised controlled trials (RCTs) conducted on consumers renewing their annual home and motor insurance policies. These trials found that renewal notices that disclosed last year's premium next to the new premium led more consumers to switch provider or negotiate their policy at renewal in the home insurance market. The trials found no effect in the motor insurance market. Other changes to renewal notices, including simplifying renewal notices, sending information leaflets, and sending reminders had little or no impact on consumer behaviour.

Following these trials and following consultation, we made new rules and guidance in August 2016 that required firms to make changes to their general insurance renewal communications.[1]

Under the 'renewal requirements' firms proposing the renewal of a general insurance product to retail customers (consumers) must, in writing or a durable medium:

- show the premium to be paid on renewal and the previous year's premium[2] at each renewal (ie on the renewal notice)
- include text on the renewal notice encouraging consumers to check that the level of cover offered is appropriate against their needs and indicating that, if they wish, they can compare the prices and levels of cover from other providers
- include an extra prescribed message encouraging them to shop around where the proposal relates to a fourth or subsequent renewal

The rules came into effect on 1 April 2017. We also issued non-Handbook guidance to accompany the rules.

The aim of this intervention was to prompt consumers to pay greater attention at renewal, engage, shop around, and make better informed decisions. It was also expected to improve firms' treatment of existing customers by focusing on matters such as pricing strategies and improving renewal practices and, as a result, improve consumer outcomes. In doing so, the intervention was expected to promote effective competition in retail insurance markets by encouraging customers to shop around.

There are two characteristics of how firms could have implemented the rules that will feature in our approach for this ex post impact evaluation:

- firms had some flexibility in *how* to disclose and present the information
- firms decided *when* to implement the rules before they came into effect on 1 April 2017.

1 FCA, PS16/21: Increasing transparency and engagement at renewal in general insurance markets
2 This annualised figure should reflect any mid-term adjustments to the previous policy.

Our evaluation

This annex presents the econometric analysis we have undertaken as part of our evaluation of the April 2017 insurance renewal transparency intervention.[3]

The analysis attempts to estimate, in a way that controls for other factors outside of our intervention, the effect our intervention has had on switching and negotiating among consumers who received general insurance renewal notices following the intervention, as well as the potential effect on firms' pricing decisions.

We structure this annex as follows:

- Section 1 provides a summary of the expectations we are seeking to test and our hypotheses
- Section 2 describes our data collected from firms operating in the home, motor and pet insurance markets
- Section 3 details our methodology and the assumptions we have made
- Section 4 sets out our main empirical results
- Section 5 covers various robustness checks we have undertaken to test our results
- Section 6 discusses alternative econometric specifications we tested and their results

3 The evaluation was announced in our General Insurance Pricing Practices Terms of reference.

Section 1: Our expectations on the effect of our intervention and the role of econometric analysis

The aim of our econometric analysis in this evaluation was to identify if our intervention had a causal impact and, if so, of what magnitude. To do so, we had to compare actual observed outcomes with an estimate of what would have happened without the intervention (the 'counterfactual').

Our econometric analysis summarised in this Annex focused on only our expectations for which producing a counterfactual is viable. Table 1 outlines the questions that the analysis presented in this Annex seeks to answer[4], our expectations, and the broad approach that our econometric analysis takes to answering the question.

Table 1: Questions to answer, pre-intervention expectations to test and role of econometric analysis in this report

#	Question to answer	Our expectation	Role of econometric analysis
1	Has the number of consumers switching or negotiating at policy renewal increased?	Based on the results of OP 12, we expected that our intervention would prompt consumers to pay greater attention at renewal. This would cause a greater level of engagement, shopping around and negotiation by consumers	Our econometric analysis estimates if and by how much switching and negotiating rates for retail insurance products in the home, motor and pet insurance markets have changed because of our intervention.
2	Are consumers paying less for similar levels of coverage at renewal?	We expected that our intervention would improve firms' treatment of existing customers by focusing on matters such as pricing strategies	Our econometric analysis estimates if and by how much insurance retail prices have changed because of our intervention

Source: FCA

4 A full list of questions can be found in the Evaluation Paper main report.

Section 2: Data

To support our econometric analysis, we needed to collect data from the insurance firms and brands that implemented our rules. Since our rules apply to all general insurance contracts within the defined scope, they apply to multiple product lines, and tens of millions of insurance contracts. To be proportionate, we chose to analyse a sample of markets and firms.

Market selection

We selected markets based on market size, continuity with analysis undertaken before our intervention, the scope of our rules, and representativeness. Based on the criteria above we selected 3 markets[5]:

1. Home insurance
2. Motor insurance
3. Pet insurance

The field trial in OP 12 found evidence of a statistically significant effect for the renewal letter disclosure in the home insurance market but not the motor insurance market.

Firm data request

We shortlisted insurance firms within each market before using a random sampling approach to select firms to submit data.[6] Some firms and brands were then removed firms from this shortlist if they dealt exclusively with brokered business (since renewal notices and data collection systems were not consistent with the rest of the sample), or if they conducted a very small number of transactions. Our sample does not capture business conducted through price comparison websites, which account for approximately 10% of the market.[7]

Our final sample consisted of 17 general insurance firms in the home, motor and pet insurance markets. Some firms were selected to provide data for more than one market. Since firms can have multiple brands, in some cases we asked firms to provide data on multiple brands separately (eg where their renewal documentation or pricing systems differed). Taking these cases as separate observations, we analysed 29 brands across the home, motor and pet insurance markets. Table 2 sets out the number of

5 Markets were defined based on the underlying asset being insured. Our Implementation Annex contains more detailed information on the motivations for the choice of the 3 markets.

6 See Annex 1 for our implementation assessment for further detail.

7 We considered that collecting reliable data on brokered and intermediated consumers that would allow us to estimate the effect of our intervention would be problematic because of the way customer information is recorded. Our expectations would be for brokered consumers to be similar in characteristics to the consumers in our sample, but customers using price comparison results to have different characteristics. Therefore, our results are likely to be representative for a large proportion of but not the entire home, motor and insurance markets.

firms and brands by market, as well as an estimate for the proportion of gross written premium (GWP) they represent.[8]

Table 2: Summary of data request firms and brands

Market	Number of firms	Number of principal brands	Estimated proportion of the market
Home	7	10	49%
Motor	10	14	56%
Pet	4	7	N/A[9]
Total	17	29	-

Source: FCA

The same brands referred to in Table 2 were included in our implementation assessment (Annex 1).

Transaction data request

We asked the 17 selected firms to provide consumer transaction-level data on all policies with renewal notices sent between 1 February and 30 June 2017, and 1 February and 30 June 2018. This allowed us to observe a reasonable period just before and just after our rules came into effect, as well as the same periods a year later. We collected this data between February and April of 2019.[10] The transaction-level data request was designed to help us estimate a counterfactual outcome and conduct econometric analysis of the intervention's effects on key outcomes.

For each consumer-level transaction, we requested data on several dimensions including consumer details, policy characteristics, policy and renewal notice dates, policy prices (including details of any negotiation), and expected claims cost to the firm. These are set out in detail in Table 3:[11]

Table 3: Data variables requested

Category	Variable	Description in data request template
Details on the consumer	Consumer ID	Unique consumer identification number
	Consumer postcode	Consumer's full postcode
	Consumer date of birth	Consumer's date of birth
	Existing consumer	Indication of whether the consumer is an existing or new consumer. Consumers who move to a new product within the same firm are treated as existing consumers

8 The 'market' excludes price comparison websites but includes brokered business.

9 GWP pet insurance market data were not available, but the firms selected made up 42% of all complaints made to pet insurance firms.

10 Separately, we asked firms in our sample to provide initial transaction-level data in September 2018 to support our implementation analysis work. Our analysis only concerns the data collected in 2019.

11 Note that for the last four categories, we also asked respondents for previous period variables. This has not been included in the table for simplicity.

	Number of previous policy renewals	Number of times that the consumer has renewed their policy with the firm, excluding the current renewal
	Number of other policies held by consumer with your firm	Number of other products that the consumer has with the firm that are not an add-on to this base policy
	Auto-renewing	Indication of whether the consumer is signed up to auto-renew their coverage at the end of their coverage period
Details on the underwriters, brand and policy type	Underwriter name	Name of the firm which underwrote the policy
	Brand name	Name of brand or distributor through which the policy was sold
	Insurance policy type	The policy product type: i) home insurance, ii) motor insurance, iii) pet insurance
	Policy coverage type	Type of policy coverage
Timing of policy and renewal notices	Policy start date	The date on which the new policy or renewal comes into effect
	Renewal notice sent date	The date the renewal notice was sent to existing consumers
	Policy paid date	Date on which the new or renewed insurance policy was paid
Policy prices	Premium offered[12]	The premium offered to existing consumers in their renewal invitation document
	Premium negotiated[13]	Indication of whether the consumer made contact with the firm to negotiate the quoted premium or coverage amount for the policy
	Negotiated price offered	The new premium offered to an existing consumer if they negotiated the price
	Premium paid[14]	The total premium paid by the consumer for the policy
	Fees paid	Fees paid by the consumer for the policy (for example, administration and credit card fees)
	Interest paid[15]	Where applicable, total interest to be paid by the consumer across the length of the policy
	Policy coverage amount[16]	The coverage amount included in this period's policy
	Policy excess[17]	The excess amount included in this period's policy
Policy characteristics	Negotiated policy coverage changes	An indication of whether, and how, a consumer made changes to the policy's details between the renewal notice and renewing
	Policy renewed	An indication of whether an existing consumer renewed their policy

12 This includes tax but does not include fees.

13 If this was not recorded, we asked the firm to provide whether the consumer made contact in the period between the renewal documentation being sent and the date the renewed policy took effect.

14 This includes tax but does not include fees and interest. For monthly policies, this would be the sum of all monthly premiums.

15 For example, a policy with a premium of £100, that is paid through 11 monthly instalments of £10, would have £10 of interest.

16 If there are various coverage types, this would be provided as an aggregate figure.

17 If there are different excess levels for different aspects of the cover, we asked firms to provide the highest.

	Payment timing	An indication of whether this period's premium was paid in full or monthly
	Payment method	An indication of how this period's premium was paid
	Policy length[18]	The length of time, in months, for which a policy is taken out
	Number of people covered by policy	The number of people covered by this period's policy
	Number of add-ons to basic policy	The number of add-ons to the policy's basic coverage
Policy cost	Expected claims cost	The expected claims cost of providing the insurance policy[19]

Source: FCA

We also requested monthly aggregate data on average premiums and policy non-renewals over from May 2014 to April 2018. This provided context on market dynamics leading up to and following on from our intervention.

Data management and data quality

We carried out the following tasks upon receipt of respondents' data:

- An initial data review and quality assessment of each individual response and importing data into statistical software.
- An in-depth objective and subjective review and cleaning of all responses. This included, but was not limited to:
  - Establishing common variable formats (for instance for postcodes)
  - Distinguishing between zeroes and missing values
  - Ensuring logical consistency between prices
  - Ensuring logical sequences of dates (for instance renewal notice date to precede policy start date)
  - Examining the distribution of variables to examine possible errors
  - Removing duplicated observations[20]
- If issues were discovered in either of these first two steps, the respondent was re-contacted and asked to clarify the submission or to provide an updated version of the data.
- We merged all data into a single dataset. Due to the differences in the nature of home, motor and pet insurance, we analysed these markets separately.

There remain some data quality issues to note, primarily due to differences in data recording systems. In particular, we are not able to use most variables related to policy

18 For example, a 3-year policy would be 36 months

19 We did not gather information on the claims history of a policy, but would expect this variable to reflect all risk factors.

20 We produced a unique identifier for every transaction in our dataset and removed duplicate entries of these unique identifiers. We did not remove duplicates of consumer ID by year where it was possible that consumers could hold multiple policies (for example more than one car). We also found cases where consumers details were matched but IDs were different, suggesting that churning (cancelling and taking out a new policy with the same firm) may be present; these observations were not removed.

coverage changes and policy excess since firms did not record this information in a consistent manner. One firm only recorded partial information with regards to consumer date of birth and postcode, which may mean they are under-represented in our matched sample. A few firms were unable to populate the previous policy's expected claims costs, meaning we could not use change in expected claims cost as a control in our regression on premium differences.

We integrated the results of our renewal notice implementation assessment (see accompanying Annex) into the transaction data. While our implementation assessment scored renewal letters at different points in time, we were not able to observe in our transaction data which version of the renewal letter the consumer received.[21] Instead, we matched observations according to the policy's renewal notice sent date, and the dates firms told us they implemented the renewal notice modifications. Where firms introduced the modifications gradually, we coded the compliance score as missing (as it is not certain which version of the renewal notice the consumer received). We excluded from our analysis policies where the renewal notice was sent within four days of the firm's stated modification date since. This is because some firms indicated there was a gap of several days between a firm implementing a change to its renewal notices template and those notices being sent to consumers.

To inform our econometric analysis of whether consumers who switch or negotiate at renewal pay less for similar levels of coverage, we identified consumers that rejected a renewal notice with one firm and took out another policy with another firm within our sample. To find such matches, we matched policies by postcode and date of birth within the same market and within the renewal period of the old policy. For firms that could only provide age and not date of birth, we used a fuzzy match by postcode and potential dates of birth. In both cases, we used only unique matches for our analysis.[22]

To find cases where consumers 'churned' (ie took out a new policy with the same insurer), we used the above methodology but looked for matches within firms.

We have not rebased prices to account for the effects of inflation. The main reason is that our rules required firms to present a comparison of a consumer's previous insurance premium in nominal terms. If we were to adjust last year's premium for inflation, then our measure of premium difference between the current and previous year would be distorted relative to that which the consumer saw on their renewal letter. Not adjusting for inflation should have no bearing on our results that do not include premium difference, since our main empirical method, difference-in-differences, compares relative and not absolute differences in outcomes between groups.

Switching and negotiating

The first question our analysis seeks to answer in this Evaluation Paper concerns consumers that switch or negotiate. In our data request, we defined switching as when a consumer did not renew their insurance policy. The consumer research undertaken

21 We also do not have information on the content of renewal notices before our intervention.

22 This method will only capture individuals who switched to another firm or brand in our sample. Since our sample of firms was randomly selected, we would not expect the resulting matched sample to be systematically biased. The matching method could also result in some false matches (ie different customers with the same date of birth and postcode); because of the narrow time period in which we seek a match, this probability is small and we do not consider that false matches could non-negligibly affect our estimates.

for OP 12 found that all home and motor consumers who did not renew their insurance policy switched to an alternative provider. We therefore take non-renewing and switching to be equal. We note, however, that some consumers who do not renew their insurance will not switch to another provider, for instance because they no longer require insurance coverage. OP 12 did not consider the pet insurance market, where for several reasons it is possible that a higher proportion of non-renewal represents cancellation. Our econometric approach ensures we only estimate *changes* in the rate of non-renewal or switching caused by our intervention (see Section 3).

For negotiation, we use responses to the question in our data request as to whether the consumer made contact to negotiate the policy. This question will include cases of attempted as well as successful negotiation. While most respondents were able to provide this data (or said that their firm did not permit negotiation), several firms provided proxies such as to whether a consumer contacted the firm in the renewal period or whether the premium or coverage details were amended during the same period. A small number of firms were not able to supply any information on whether the consumer negotiated. Because of these data quality issues, we have also tested our analysis using an alternative definition of negotiation (see robustness checks in Section 5).

In our regression analysis, we combine responses for switching and negotiating into a single dummy variable. There is some overlap between switching and negotiating, for instance because some consumers attempted to negotiate but then rejected their renewal offer. However, we also estimate the causal impact of our intervention on negotiating and switching separately.

Summary statistics

After cleaning, our dataset contains data on 21.6 million insurance policies. This comprises 5.6 million for the home insurance market, 13.9 million for motor and 2.0 million for pet. The number of observations for brands in our sample ranged from 40,000 to 2.7 million. Table 4 provides a summary of some of the key aspects of the data.

The results show differences in the structure of our home, motor and pet insurance samples. Motor insurance has the highest mean and median annual premium (all policies in our data are 12 months long), and also has the highest switching and negotiating rates. As a result, the average consumer in the motor insurance market has renewed with their current firm 2.4 times, compared with 3.1 times in the pet insurance market and 4.2 times in home. Pet insurance has the largest median premium increase in renewal offers compared to last year's premium.[23] This paper does not examine profit margins in detail, but the estimated median margin appears to be higher for the home insurance market in our sample.

We also note that there is substantial diversity among the firms in the same market. Reported switching and negotiating rates by firm vary substantially.

23 Pet insurance may have the largest median premium increases because of different market characteristics, eg the increasing risk of ill health increasing with a pet's age.

Table 4: Summary statistics of selected and derived variables by market (all data combined)

Statistic	Home	Motor	Pet
Number of observations	5.6 million	13.9 million	2.0 million
Mean annual premium	£271	£447	£375
Median annual premium	£210	£360	£311
Mean number of previous renewals (among existing consumers only)	4.2	2.4	3.1
Median premium increase at renewal	£17.20	£22.70	£51.70
Median premium increase at renewal (as % of the median premium)	6.3%	5.1%	16.6%
Switching rate (all periods combined)	16.5%	26.2%	13.8%
Negotiation rate (all periods combined)	12.7%	24.7%	3.5%
Median margin (premium minus expected claims cost as % of premium)	46%	32%	35%

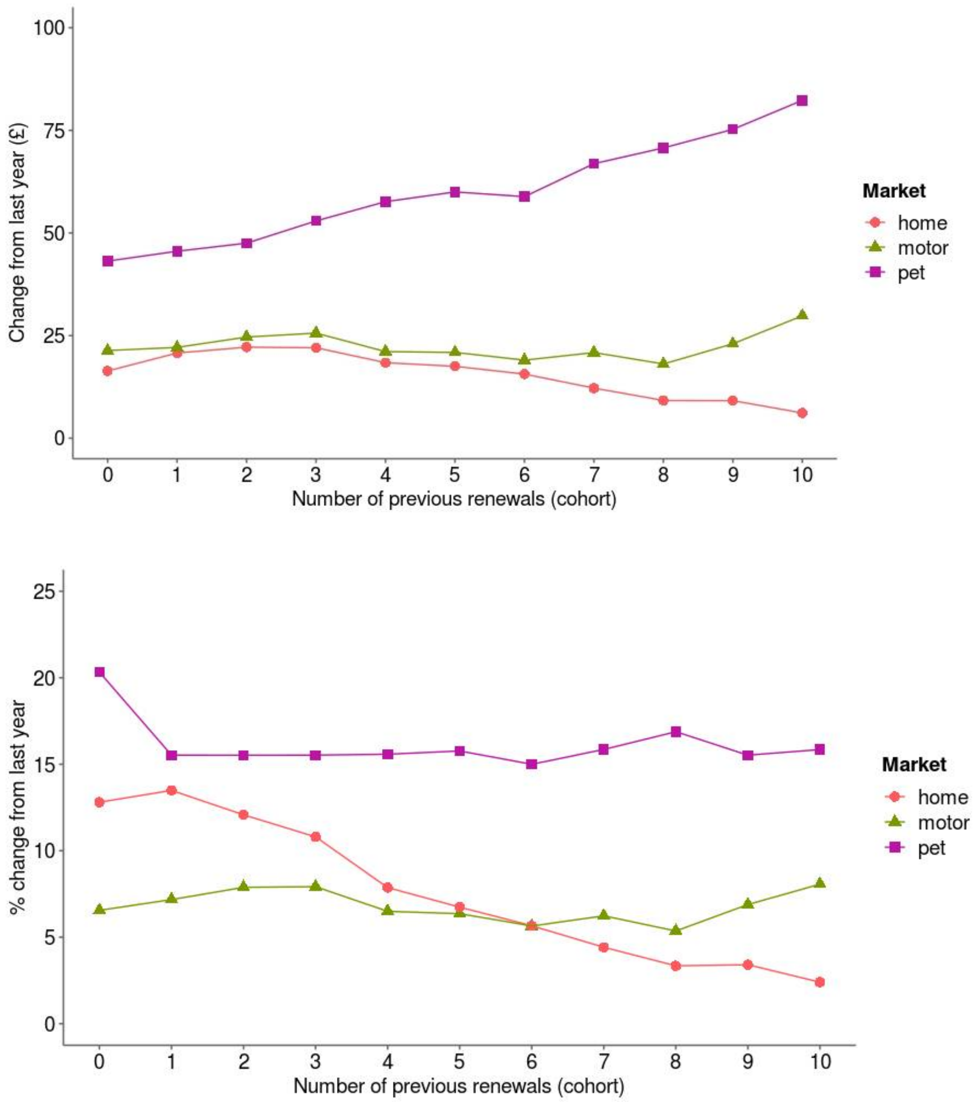
Source: FCA analysis of transaction data

To illustrate insurance premium changes, Figure 1 shows the median change in insurance premium among customers with between 0 and 10 consecutive renewals. It is important to note that this is based on cross-section of customers, not longitudinal data, and therefore the chart does not allow inference for changes over the lifetime of a policy.[24] One reason for the pattern is that nearly a third of motor insurance premiums in our sample fall from one year to the next, potentially driven by risk changes such as no claims bonuses. From our data, premium falls are much rarer in the pet insurance market, which is likely again related to characteristics and risk profile of the insurance product.

In absolute and relative terms, the median change in premium is highest in the pet insurance market for a given number of previous renewals. The home insurance market exhibits a potential pattern of premium increases reducing over time (noting the data is not longitudinal), whereas pet and motor exhibit a flatter pattern.

24 There are compositional differences between the customers at each point on the curve. For instance, some firms do not have customers with more than a certain number of renewals, and the nature of the policies and insured risks will also vary. Therefore, the shape of the curve should be taken as illustrative of the lifetime of a given policy.

Figure 1: Median change in premium offered compared to last year's premium, absolute and relative, by number of previous renewals and market



Note: Median premium offered minus last year's premium in absolute terms (top) and as a percentage of last year's premium (bottom). All data in our sample combined.
Source: FCA analysis of transaction data.

Section 3: Econometric approach

In this section, we outline our econometric approach including the methodology and core assumptions.

Methodology

Our main methodology to estimate the causal effect of our intervention on consumer and firm behaviour is a difference-in-differences (DiD) model.[25] This method is often used when assessing the impacts of policy interventions and regulation, especially under the conditions of a natural or quasi-experiment.[26]

The concept of DiD is to compare the outcomes over time of two or more similar groups, some of which are exposed to a treatment and some of which are not. The simplest case is where there are two groups and two time periods. As a result of an exogenous event[27] (eg a policy change), one group (the treatment group) receives the treatment in the second period, but not in the first. The second group (the control group) does not receive the treatment in either period. The DiD estimator is equal to the difference in the outcome between the first and second periods of the treatment group minus the difference experienced by the control group (hence the term difference-in-differences).

The counterfactual outcome for the treatment group - what would have happened without the exogenous event – is defined by the path of the control group's outcome. The DiD method does not measure the change experienced by the treatment group per se, but the change relative to the control group. Assuming the two groups follow the same path (see assumptions below), any statistically significant deviation can be attributed to a causal effect of the exogenous event.

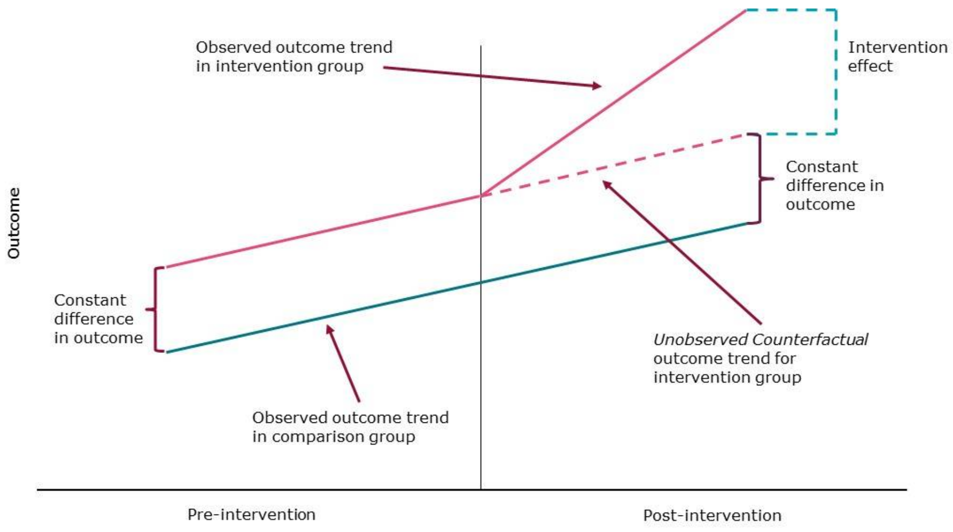
Figure 2 represents the theory of the DiD approach in graphical form. The DiD estimator can be derived from an econometric analysis of data on the control and treatment groups.

25 For further information on the general DiD methodology see: Angrist & Pischke, 2009, 'Mostly Harmless Econometrics: An Empiricist's Companion', p.227-243; and Wooldridge; 2009; 'Introductory Econometrics: A Modern Approach. Fourth Edition', p.450-455

26 For selected examples of DiD approaches used in the academic literature in a range of different scenarios see: Ashenfelter & Card, 1985, 'Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs'; Meyer, Viscusi & Durbin, 1990, 'Workers' Compensation and Injury Duration: Evidence from a Natural Experiment'; and Card, 1994, 'Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania'.

27 An exogenous event is one that comes from outside of a system and is not driven by the system itself.

Figure 2: Representation of differences-in-differences approach



Source: Adapted from the Columbia University Mailman School of Public Health

Implementing the difference-in-differences model

We apply the difference-in-differences model by using the variation of when firms in our sample complied with our intervention. The firms in our sample first complied with our intervention between December 2016 and the end of March 2017 (see Annex 1). This variation provides an opportunity to compare how the outcomes of the firm that has just complied with the intervention (our treatment group) compares with firms that have yet to comply (our control group). The trend in outcomes among firms that have yet to comply provide the counterfactual.

Our methodology means that we focus primarily on the short-term impact of the intervention when it was first implemented in early 2017. Given the pattern of firms' compliance, we consider it more difficult to estimate the impact over a longer time period.[28]

Our analysis departs from the simple two-group, two-period DiD set up, as we have multiple groups (firms) that complied with our intervention at different times. We therefore implement a variation of the DiD model that allows for multiple groups and multiple time periods. We then estimate a regression model known as a 'two-way fixed effects model' that contains dummy variables for each group and each time period. The two-way fixed effects model allows us to estimate a weighted average of the two-group, two-period (2x2) DiD comparisons. These weights are provided by a combination of sample size, here the number of policies, and variance.[29]

Importantly, with multiple groups and time periods, the control groups include not only firms that have yet to comply with the intervention but also firms that complied

28 We have attempted to estimate the causal impact of our intervention over a longer time period in Section 5, though there is greater doubt over the assumptions in this analysis.

29 For further information see Goodman-Bacon, 2018, Difference-in-differences with variation in treatment timing.

previously. Firms that complied with the intervention in an earlier time period act as a control group in later time periods because, once complied, they have no further variation in their compliance status. This characteristic creates an additional assumption for our model (see below).

Defining time periods

We manually constructed time periods for the two-way fixed effects model according to firms' compliance dates. One condition we imposed is that each firms' compliance level does not change within a period, and any variation only occurs between periods.[30] This ensures that in the binary case each firm has either complied or not complied, and in the continuous case the compliance level is static in each period.

We use a minimum of three time periods. This exploits the fact that firms complied at different times and provides the required variation between periods.

Another condition we imposed was to avoid periods consisting of 1 or 2 days. From preparatory work, we noted that brand-level outcomes on a single-day basis are subject to noise, but are smoother over longer periods. To reduce the risk of this noise affecting our results, we limited our choice of time periods to those with a minimum of 3 days. The average length of time period in our selection is 17 days.

There are multiple ways to construct time periods according to these conditions. Because of the overlapping pattern of firm compliance dates, not every firm can be included in each model, and the treatment group might comprise several firms. We selected a small number of time period models that provided variation in the firms included, and the number and lengths of the periods used. As an illustration, model 2 uses data from 3 time periods (1 Feb – 10 March, 17-20 March and 5-30 April inclusive). The time periods we selected are summarised in Table 5.

Table 5: Time periods used for our 2-way fixed effects regression models

Model	Market	Number of days in each period					Number of firms included (out of total)
		Period 1	Period 2	Period 3	Period 4	Period 5	
1	Home	38	4	3	26	-	7/10
2	Home	38	4	26	-	-	9/10
3	Home	19	5	26	-	-	7/10
4	Motor	39	3	5	25	-	9/14
5	Motor	39	3	5	26	-	8/14
6	Motor	21	5	25	-	-	10/14
7	Pet	25	10	9	6	22	7/7

Source: FCA analysis of transaction data

30 We have also avoided drawing time periods that include firms during the assumed 4-day gap between their implementation and renewal letters being sent. Similarly, we have avoided periods or excluded firms in cases where the firm implemented the change gradually over a longer time period (because it is ambiguous which renewal letter a consumer would have received).

If our intervention had a similar causal effect on all firms in a market, we would expect the different models in Table 5 to provide similar results within each market, since they differ only in time period and firms studied. If the models provided different results in the same market it could imply variation among treatment effects among the firms in our sample. Differences could also be caused by other factors such as statistical noise that is captured by one model's time periods but not another's.

Specification

Our econometric specification is set out below. We estimate an ordinary least squares regression with group and time fixed effects.[31] The impact of our intervention, controlling for other factors, is estimated by the coefficient (labelled as 'DiD' in our regression results in Section 4) on a variable representing the presence or intensity of treatment (D_{it}). A statistically significant difference-in-differences variable (D_{it}) would indicate that our intervention has had a causal effect on our outcomes.

$$\boldsymbol{y_{it} = \alpha_i + \gamma_t + \beta^{DiD} \cdot D_{it} + \theta \cdot X_{it} + \varepsilon_{it}}$$

Where:

- i is the individual consumer observation
- t is the time period
- y_{it} is the outcome or dependent variable for individual *i* at time *t*. For switching and negotiating, this is 1 if an individual switched or negotiated, and 0 otherwise. For premium difference, this is the difference in renewal offer in pounds for individual i.
- α_i are insurance brand (ie group) fixed effects
- γ_t are time fixed effects representing the time periods in our two-way fixed effects models
- D_{it} is the 'treatment' variable. Depending on specification, this is either: a treatment dummy variable equal to 1 if the consumer was exposed to the post-intervention renewal letter, and 0 otherwise; or a continuous variable equal to the brand's compliance score. Where we use compliance score we also add the square of the compliance score in an additional specification.
- X_{it} represents the set of control variables
- ε_{it} is the error term

Brand fixed effects control for time-invariant characteristics of insurance brands and their consumers.[32] Section 2 demonstrated variation in the characteristics of firms in our sample. Time fixed effects control for factors that are common to all brands at a point in time or a shared time trend. We include a set of control variables, X_{it}, to control for observable systematic factors that may differ between brands in a way that also

31 Controlling for fixed effects is equivalent to adding a series of dummy variables for each group or time period.

32 We have controlled for brands rather than groups of brands that comply in the same time period. Since compliance scores vary at the brand level, this ensures a consistent approach to group fixed effects between the 'binary' and 'compliance score' regressions. We included the alternative definition of groups as a robustness check.

varies over the time period of our analysis.[33] We include only control variables that are reported by all the firms in our sample.

As set out in Annex 1, we attributed each firm a 'compliance score' between 0 and 100 indicating the degree to which a sample of renewal notices matched the requirements of our intervention. This leads to two main options for specifying the treatment variable in our regression model, which we discuss below[34]:

- A binary variable where firms are split into two groups of either being compliant or non-compliant. The interpretation of the DiD coefficient in this case is the estimated effect of the intervention as it was implemented in 2017. One downside is that all implementation of our rules is grouped under the same term, but we know from our compliance analysis that renewal letters varied to a large degree.
- Use each brand's compliance score (scaled to between 0 and 1) to indicate the intensity of how the intervention was applied. The interpretation of the DiD coefficient in this case is the estimated effect of the intervention as if it had scored 100 points on our scale, but the coefficient can also be used to estimate the average treatment effect for firms at different levels of compliance.[35] One downside of this approach, however, is that it imposes a linear impact of compliance score, ie an increase from 0 to 10 points has the same impact as from 90 to 100 points.[36]

To answer the evaluation questions in Section 1, there are two groups of dependent variables or outcomes of interest for our analysis:

1. The rate of switching or negotiating (either combined or separately)
2. The difference between the premium offered and previous premium paid, or the difference between the previous and current premium paid.

Assumptions

The use of a DiD methodology relies on a number of identifying assumptions, which we discuss in turn below:

- There are parallel trends between the control and treatment groups in the absence of treatment
- The average treatment effect is assumed constant
- Strict exogeneity
- The stable unit treatment value assumption (SUTVA)
- No serial correlation

33 Including these control variables also reduces the size of the error variance and therefore the standard error of the variable of interest, which is the measure of the impact of our intervention.

34 Although both of these dependent variables are scaled between 0 and 1, we use ordinary least squares regression rather than a limited-dependent variable model. This is common in the DiD literature, for example in studies in labour economics examining employment.

35 For further information on 'fuzzy' DiD designs with varying treatment intensity, see: De Chaisemartin and D'Haultfoeuille, (2017), "Fuzzy Differences-in-Differences", Review of Economic Studies.

36 In preliminary work, we also included a variable representing compliance score squared, that would allow us to estimate the marginal effect of our intervention for firms at different levels of compliance. However, the results are not easy to compare, and the policy conclusions are uncertain as there are multiple potential explanations behind the findings. In addition, extrapolating results to fit a quadratic curve is only meaningful over the range of compliance scores we observed.

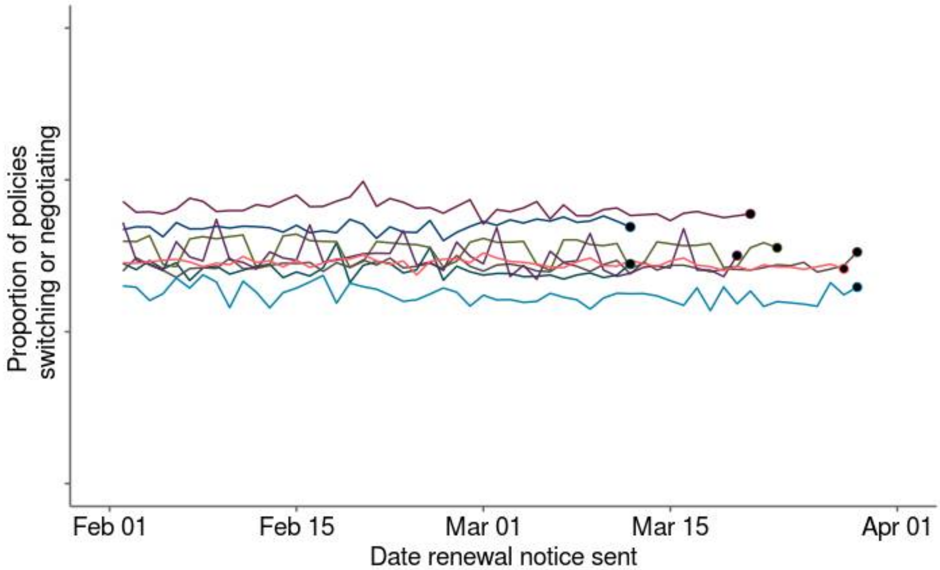
Parallel trends

The 'parallel trends' assumption is the main identifying assumption for a DiD model. It requires that in the absence of the intervention (treatment), the difference between the treatment group and the control group is constant over time. Violation of the conditional parallel trend assumption leads to a biased estimation of the causal effect. Our 'placebo tests' provide an additional check on the parallel trends assumption (see Section 5).

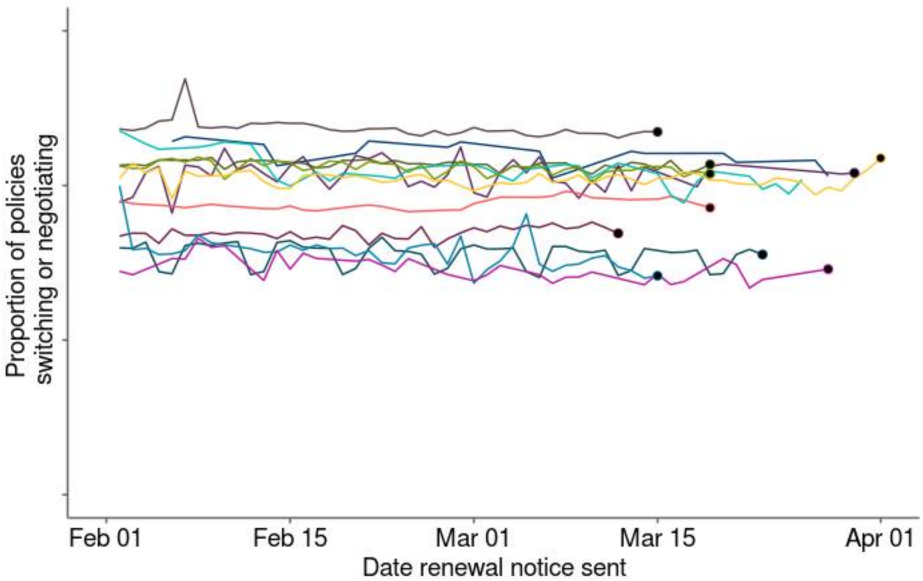
Figure 3 plots the pre-intervention switching and negotiating rate by brand as a visual check of pre-intervention parallel trends. These plots are for unconditional parallel trends, ie without controlling for other factors, whereas our regression results present specifications with and without control variables.

Figure 3: Pre-intervention trends in switching and negotiating rate (anonymised), by brand and market, 2017

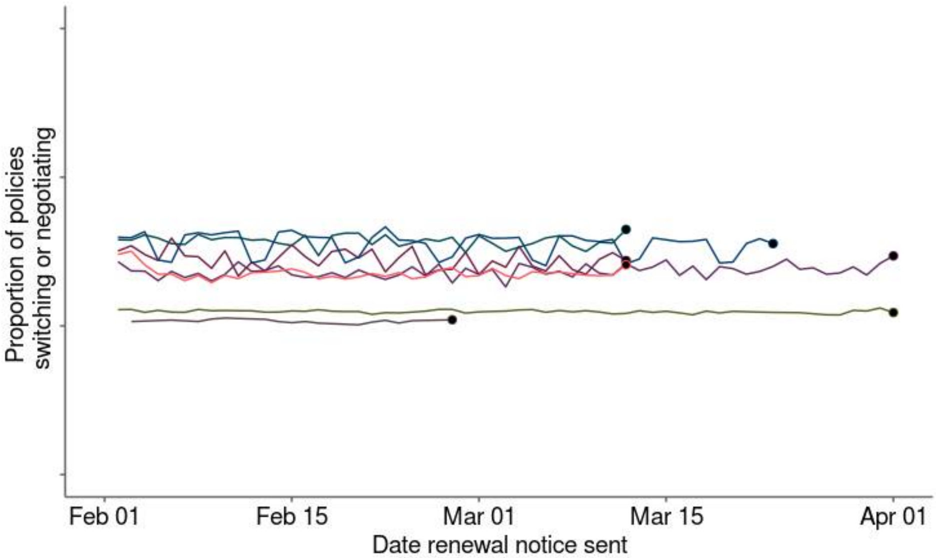
Home



Motor



Pet



Notes: Lines represent brands in our analysis. The black dot represents the implementation date of the brand. To maintain anonymity, the scales and limits of the vertical axes are not labelled – the scales do not start at zero. Dates with fewer than 20 observations per brand are excluded. Source: FCA analysis of transaction data.

We can see that the pre-intervention trends in switching and negotiating rates are broadly parallel. This is especially case in our pet insurance sample. While broadly parallel, there is some considerable volatility within brands upon examining the data on a daily basis. There are short periods in which outcomes for brands appear to converge and diverge, most obvious in our motor insurance sample. We can conclude there do not appear to be large-scale infractions of the parallel trends assumption but that the noise in the data warrants caution in analysing our main results. In our robustness checks, we run tests on pairs or groups of firms that have the closest unconditional pre-intervention trends to test if these suggest a different estimated treatment effect.

Treatment effects

An additional assumption required by our use of a two-way fixed effects DiD model with multiple time periods is that the average treatment effect is constant over time. Because, in our methodology, already treated groups act as a control group in later periods, if the intervention has a lagged effect then it can bias the size and direction of post-intervention 2x2 DiD comparisons. Therefore, we must assume a constant treatment effect, with no lagged effect. This assumption seems intuitive when applied to consumer decisions. Each time period in our analysis consists of a new cohort of consumers who have no opportunity to learn over the period of time of our analysis. The treatment intensity also appears constant - some firms made post-compliance refinements to their renewal notices during our 2017 period of analysis, but their compliance scores were within 2 points of the original score.[37]

Strict exogeneity and stable units

The strict exogeneity assumption requires that the timing of firms' implementation of our intervention is independent from the outcomes, conditional on group and time fixed effects.[38] Firms in our sample did implement our intervention in a varied manner, and this variation could have been driven by their expected effect on our outcomes of interest. But this self-selection would only present a problem if treatment intensity depended on factors that vary over time. Our inclusion of specifications that account for the intensity of treatment attempts to account for this scenario.

Another way strict exogeneity could be violated would be if the treatment could be anticipated in advance. We can be reasonably confident that consumers have little opportunity to anticipate and act upon the content of their upcoming renewal notice before it arrives, especially since a firm's implementation date is unknown in advance. On the supply-side, however, firms may have been able to anticipate the introduction of the renewal notice disclosure and its effect on their consumers. Our final rules were announced in August 2016, ahead of an implementation date of 1 April 2017. Firms could theoretically have optimised their renewal pricing to maximise profits based on an anticipation of the intervention's effects (for example, as described in OP 12) and to avert the risk of an adverse effect on their renewal rate.

Any such anticipation would only be captured by our DiD if it happened at the same time as the firm complied with the intervention. We do not have any evidence to draw on as to the likelihood of anticipation, but consider it unlikely that firms would change their pricing ahead of the intervention coming into effect. We also note that if intervention timing is correlated with the anticipated impact on renewals, our DiD estimates could potentially suffer from downwards bias if certain firms only act as a control group in our specifications.

Stable units

The stable units assumption is that the treatment status of one unit (here either firms or consumers) must not affect the outcomes of other units, ie that there are no spillovers. Our use of transactional data means we should meet this condition. It is

37 Our specification that uses compliance score as the dependent variable will already account for these minor changes.

38 This assumption is not concerned with self-selection into treatment groups per se, but rather with selection that violates the common trends assumption, ie the difference in error terms must not be correlated with treatment, but their level can be.

possible that supply-side pricing reactions among firms could alter competitive dynamics and thus the reactions of other firms, but we do not consider this a major risk over our short time periods.

Related to stable units, we also verified that there are no large changes in the composition of the treatment and control groups over time, for instance due to sample erosion. In our DiD design we ensured the periods are sufficiently wide, so that there is no large reduction in sample size among certain groups (as might happen, for instance, near the end of the period over which we requested data).

Serial correlation

Since our analysis uses more than two time periods, we have considered whether the outcomes we use could be serially correlated. Where one period's outcome is correlated with the preceding period's, for instance because of a lasting shock, the parallel trends assumption may not hold. In DiD specifications, serial correlation is more likely with a longer time series, the nature of the dependent variable being analysed, and where the treatment variable changes little within a group over time.[39]

We attempt to correct for serial correlation by reporting standard errors clustered by brand.[40] Although our analysis uses customer-level data and a relatively short time series of between 3 and 5 periods, it is possible that a brand's mean insurance premium or the mean premium difference could nevertheless be serially correlated. Our decision to cluster was also informed by initial results from our placebo tests (see Section 6).[41]

Correcting for potential serial correlation by clustering standard errors by brand makes it more difficult to detect real effects of our intervention, especially with a small number of groups (we use a maximum of 10).[42] Therefore the downside of our approach is that our analysis may suffer from low power, and even a real effect of our intervention may not be found to be statistically significant. We therefore interpret our point estimates with care and do exclude any results based on statistical significance.

Alternative econometric approaches

We have also drawn on alternative methodological approaches to answer the questions in our evaluation. Undertaking this analysis provides us with results to compare to the main analysis and so enables us to assess the validity of our difference-in-differences model results.

The first approach we looked at was a sharp regression discontinuity design (RDD). Our RDD analysis looked at individual firms' outcomes within the tight time range around each firms' renewal notice change. By comparing consumers who received the

39 Serial correlation can lead to biased standard errors, affecting significance testing though not the magnitude of estimated coefficients. See Bertrand, Duflo & Mullainathan (2004), "How much should we trust differences-in-differences estimates?", Quarterly Journal of Economics.

40 This helps address serial correlation since the standard errors are calculated according to a variance-covariance matrix that is consistent in the presence of any correlation pattern within brands over time.

41 Another reason to cluster standard errors by firm is that our models group together treated firms, ie our unit of observation (a brand) is more detailed than the unit of variation (a group of brands).

42 We used small-sample adjusted clustered standard errors, but even with this method our standard errors could suffer from some bias. There are other methods available to attempt to correct for serial correlation in small groups. However, we were not able to implement these because of the large size of our dataset.

old renewal notice just before the firm implemented a change, with those receiving the new-style renewal notices just after the firm implemented the change, it is possible to estimate the causal effect of our intervention.

The RDD approach assumes that the individuals in each group are similar (ie there are no unobserved differences in the makeup of consumers renewing just before and just after the implementation date, after controlling for other variables). Since our RDD estimation contains different periods, it is possible that this assumption could be violated. However, focusing on narrow time range minimises the possibility of other time-varying factors driving observed differences in outcomes between the two sets of consumers. The downside of a narrow time range is that the approach could be more vulnerable to noise within our data. Section 6 contains our RDD results.

Section 4: Empirical results

Overview

This section presents the results of our analysis and discusses the interpretation of the findings.

We present separately our analysis of consumer switching and negotiating, and premium differences upon renewal. We present our two versions of the DiD estimator:

- a ‘binary’ measure of whether or not the firm complied with our intervention
- A ‘treatment intensity’ measure equal to the compliance score of the firm

We present the results of our two-way fixed effects difference-in-differences model here, with separate results for the home, motor and pet insurance markets. We have run a variety of different specifications to test the stability of our estimates and how they change in the presence of different control variables. We present those results in Section 5. Our main report uses only our binary regression estimates.

Our baseline model includes only brand and time fixed effects. A second version adds our full list of controls (see Table 6).[43]

Table 6: Variables included in the regression specification

Variable	Description	Baseline	Baseline with all controls
DiD estimator	One of our three treatment variables	✓	✓
Brand[44] fixed effects	Dummy variable for each brand in the sample	✓	✓
Time period fixed effects	Dummy variable for each of the time periods in the two-way fixed effects	✓	✓
Expected claims cost (*premium difference only*)[45]	Expected cost of the insurance policy to the firm		✓

43 This list does not include variables that we would have wanted to include as control variables, but data quality issues would have meant excluding firms from our sample. We have not included the coefficients from all the control variables in our empirical results but summarise which controls were used.

44 We use brand as our grouping variable in our specification, rather than a dummy variable that groups together firms that comply during the same period. The main reason is to maintain consistency with the treatment intensity regressions where each brand has a distinct compliance score and therefore cannot be grouped with other brands. For the binary regressions, we tested the alternative group dummy and found it had a negligible impact on our findings.

45 Our preferred control for consumer risk in our premium difference specification would be the difference between current and previous expected claims cost as it captures the full effect of the difference over time. We use the current expected claims cost only, as many firms have been unable to provide an accurate version of the previous expected claims cost. In many cases, firms have been unable to apply mid-term adjustments to the previous expected claims cost so it is inconsistent with the previous premium paid which is adjusted in this way.

Number of previous renewals	How many times the consumer has renewed before		✓
Auto-renewing	Whether the contract automatically renews at termination		✓
Consumer age	Age of consumer		✓
Policy coverage type	Dummy variable for each policy coverage type outlined in our data request		✓
Weekday	The weekday the renewal notice was sent		✓

Source: FCA

Regression results on switching and negotiating

Binary results

Table 7 outlines the results of our analysis on the effect of our intervention on switching and negotiating by consumers. The interpretation of the coefficient (labelled as "DiD (binary)" in the tables) is the estimated average effect of our intervention as it was implemented on the combined switching and negotiating rate (ie 0.01 would represent a 1 percentage point increase).

The results suggest that, on average, our intervention is associated with a positive effect on the switching and negotiating rate in the motor insurance market (of 1.3 to 2.3 percentage points), a smaller positive effect in the pet insurance market (1.1 to 1.2 percentage points), and a negative effect in the home insurance market (-0.7 to -3.1 percentage points). Only the results for the pet insurance market are statistically significant, however, so there is greater uncertainty as to the effect in the home and motor insurance markets. Introducing control variables such as consumer age changes the size but not the direction of the estimated DiD coefficients.

The results of the different models show consistent results for motor, but show some variation for the home insurance market. This could suggest, for example, that the intervention affected firms in different ways or that the results are not robust to the change in firm population within each specification/period.

Table 7: Estimated effect on combined switching/negotiating rate - binary regression results[46]

Home

Binary	Model 1		Model 2		Model 3	
Model	**Baseline**	**Baseline with all controls**	**Baseline**	**Baseline with all controls**	**Baseline**	**Baseline with all controls**
DiD (binary)	-0.019	-0.018	-0.031	-0.03	-0.01	-0.007
	(0.0138)	(0.013)	(0.0198)	(0.0193)	(0.0164)	(0.0137)
Number of Observations	648,629	644,976	775,893	772,345	316,601	315,119
With other controls	No	Yes	No	Yes	No	Yes
Other controls	No	Yes	No	Yes	No	Yes

Motor

Binary	Model 4		Model 5		Model 6	
Model	**Baseline**	**Baseline with all controls**	**Baseline**	**Baseline with all controls**	**Baseline**	**Baseline with all controls**
DiD (binary)	0.02	0.015	0.022	0.017	0.023	0.013
	(0.0222)	(0.0175)	(0.0209)	(0.0157)	(0.0246)	(0.0155)
Number of observations	1,754,792	1,753,990	1,371,883	1,371,074	1,280,132	1,279,584
Other controls	No	Yes	No	Yes	No	Yes

Pet

Binary	Model 7	
Model	**Baseline**	**Baseline with all controls**
DiD (binary)	0.011***	0.012***
	(0.0035)	(0.0034)
Number of observations	371,118	208,809
Other controls	No	Yes

Note: * significant at 1%, ** significant at 5%, * significant at 10%. Brand-clustered standard errors in parentheses.**
Source: FCA analysis of transaction data

Binary results for switching and negotiation separately

We have also estimated the impact of our intervention on switching rates and negotiation rates separately. Table 8 sets out the DiD coefficients for each model above. We did not analyse negotiation in the pet insurance market since only two firms in our sample permitted or recorded negotiation.

The results suggest that our intervention may have reduced both switching and negotiating in the home insurance market, but for the motor insurance market mainly affected switching rather than negotiation. Our intervention is associated with a 0.4 to 1.2 percentage point fall in switching in the home insurance market, and a 0.9 to 2.5 percentage point fall in negotiation. By contrast in the motor insurance market our intervention is associated with a 1.1 to 1.3 percentage point increase in switching, and a negligible change in negotiation. We only find a statistically significant effect for

46 Model numbers correspond to those set out in Table 5 – because of our method of drawing time periods, we run 3 models for the home and motor insurance markets and 1 for the pet insurance market.

switching in the pet insurance market and one of the models for the home insurance market.

Table 8: Estimated impact on switching rate and negotiating rate separately – binary coefficient estimates

Market	Model	Switching rate	Negotiation rate
Home	**Model 1**	-0.006	-0.017
	Model 2	-0.012**	-0.025*
	Model 3	-0.004	-0.009
Motor	**Model 4**	0.011	0
	Model 5	0.013	-0.001
	Model 6	0.011	-0.001
Pet	**Model 7**	0.012***	-

Note: * significant at 1%, ** significant at 5%, * significant at 10%.**
Source: FCA analysis of transaction data

Regressions using treatment intensity

This section presents our results on switching and negotiating when the dependent variable is each brand's compliance score. Where we include only the compliance score, the interpretation of the regression coefficient is the estimated effect of our intervention if it was implemented in a way that scored 100 points on our compliance scale, assuming a linear relationship between compliance and switching/negotiating. To provide more policy-relevant figures, at the bottom of each table we have evaluated this coefficient at two compliance scores: 50 points and at the weighted average compliance score in that market.[47]

Table 9 outlines the results of our analysis. As expected, the results are generally similar in direction to those where compliance is defined as a binary variable only. Because the coefficient represents the effect of a hypothetical 100-point renewal letter, the results tend to magnify those in the binary regressions. A greater proportion of the results, though not all, are statistically significant at the 5% or 1% levels.

The final two rows in Table 9 represent the estimated average effect of our intervention as if implemented at 50 compliance points, or at the weighted average level we observe in each market. At the average compliance level our intervention was associated with a 2.1 to 3.0 percentage point reduction in the switching and negotiating rate in home, a 1.9 to 2.5 percentage point increase in motor and a 0.2 percentage point increase in the pet insurance market.

47 The weighted averages are 60 points in home, 34 points in motor and 27 points in pet.

Table 9: Difference in switching/negotiating rate – compliance score regression results

Home

Binary	Model 1	Model 2	Model 3
DiD (compliance score)	-0.049**	-0.035	-0.041***
	(0.0214)	(0.0246)	(0.0103)
Number of observations	648,629	775,893	316,601
With other controls	Yes	Yes	Yes
Effect at 50 points	-0.025	-0.018	-0.021
Effect at weighted average score	-0.030	-0.021	-0.025

Motor

Binary	Model 4	Model 5	Model 6
DiD (compliance score)	0.063***	0.055*	0.073*
	(0.0238)	(0.0292)	(0.0385)
Number of observations	1,753,990	1,371,074	1,279,584
With other controls	Yes	Yes	Yes
Effect at 50 points	0.032	0.028	0.037
Effect at weighted average score	0.021	0.019	0.025

Pet

Binary	Model 7
DiD (compliance score)	0.009*
	(0.005)
Number of observations	208,809
With other controls	Yes
Effect at 50 points	0.005
Effect at weighted average score	0.002

Note: * significant at 1%, ** significant at 5%, * significant at 10%. Brand-clustered standard errors in parentheses. Note we do not estimate significance for the interpretations of our estimated effect at 50 points or the average compliance score.**
Source: FCA analysis of transaction data

Regression results on premium difference

Binary results

Table 10 outlines the results of the analysis of our intervention on the difference in premium quoted at renewal compared to the premium paid in the previous year. The coefficient here can be interpreted as the estimated monetary effect of our intervention, as implemented, on the premium offered to consumers when their policy comes up for renewal. For example, a coefficient of 2.5 would mean a £2.50 increase in premium difference associated with our intervention.

These results indicate that, on average, firms have responded to our intervention by reducing renewal premium increases in the home (-£1.60 to -£8.60) and motor insurance markets (-£3.70 to -£6.50) and increasing them in the pet insurance market (by £4.70 to £5.30). In the home and pet insurance markets, most specifications are significant at 5% or 10%. None of the specifications in motor reach this level of significance, so the impact is more uncertain. Our control variables have a small impact on the size of the DiD coefficients, but do not affect the direction of the estimated coefficient. The results of the different specifications (periods) show variation across both the home and motor insurance markets which suggests that the intervention affected firms in different ways.

Table 10: Difference in premium quoted at renewal - binary regression results

Home

Binary	Model 1		Model 2		Model 3	
Model	Baseline	Baseline with all controls	Baseline	Baseline with all controls	Baseline	Baseline with all controls
DiD (binary)	-6.13**	-6.121*	-8.581**	-9.124**	-1.692	-2.378
	(3.127)	(3.325)	(4.018)	(4.41)	(2.731)	(2.931)
Number of observations	648,556	635,091	751,129	738,147	316,412	308,276
With other controls	No	Yes	No	Yes	No	Yes

Motor

Binary	Model 4		Model 5		Model 6	
Model	Baseline	Baseline with all controls	Baseline	Baseline with all controls	Baseline	Baseline with all controls
DiD (binary)	-6.512	-4.577	-5.738	-3.665	-6.401	-0.122
	(8.981)	(4.944)	(8.965)	(4.898)	(9.123)	(4.571)
Number of observations	1,160,394	1,131,643	1,108,975	1,080,485	943,347	923,366
With other controls	No	Yes	No	Yes	No	Yes

Pet

Binary	Model 7	
Model	**Baseline**	**Baseline with all controls**
DiD (binary)	4.757	5.33**
	(3.165)	(2.502)
Number of observations	360,535	202,195
With other controls	No	Yes

Note: * significant at 1%, ** significant at 5%, * significant at 10%. Brand-clustered standard errors in parentheses.**
Source: FCA analysis of transaction data

Compliance score results

Table 11 outlines the results of the analysis of our intervention on the difference in premium quoted at renewal compared to the premium paid in the previous year. The coefficient can be interpreted as the estimated effect of our intervention if it was implemented in a way that scored 100 points on our compliance scale, with the coefficient representing a change in pounds.

The estimated coefficients for the home and pet insurance markets are similar to those where compliance is defined as a binary variable. As expected, the effect of the renewal notice which scored 100 points is larger than the estimated effect of the intervention as implemented in these markets. However, the results for motor are much smaller in magnitude.[48] The results evaluated at a compliance score of 50 points or at the weighted average level of compliance are broadly of a similar magnitude to our binary results. These results should be treated with caution as almost all specifications in all markets do not reach the 10% significance threshold.

48 We would normally expect the coefficient on compliance score to be in the same direction and the have a larger absolute value compared to the equivalent binary coefficient. This is not the case in our estimates for the motor insurance market. We anticipate that this is being driven by weighting of the treatment effects; with variation in treatment timing and treatment heterogeneity, weights can be negative as well as positive. See De Chaisemartin and D'Haultfoeuille (2017) for a discussion of weighting issues when using continuous treatment variables.

Table 11: Difference in premium quoted at renewal – compliance score results

Home

Binary	Model 1		Model 2		Model 3	
Model	Baseline	Baseline with all controls	Baseline	Baseline with all controls	Baseline	Baseline with all controls
DiD (compliance score)	-6.436	-8.602	-7.183	-18.821	0.537	3.676*
	(6.1)	(9.88)	(5.045)	(17.03)	(5.507)	(1.958)
Number of observations	648,556	635,091	751,129	738,147	316,412	308,276
With other controls	No	Yes	No	Yes	No	Yes
Effect at 50 points	-3.22	-4.30	-3.59	-9.41	0.27	1.84
Effect at weighted average score	-3.89	-5.20	-4.35	-11.39	0.32	2.22

Motor

Binary	Model 4		Model 5		Model 6	
Model	Baseline	Baseline with all controls	Baseline	Baseline with all controls	Baseline	Baseline with all controls
DiD (compliance score)	-9.134	-4.598	-0.976	-0.959	2.53	0.824
	(17.192)	(7.681)	(21.029)	(7.157)	(22.142)	(7.524)
Number of observations	1,160,394	1,131,643	1,108,975	1,080,485	943,347	923,366
With other controls	No	Yes	No	Yes	No	Yes
Effect at 50 points	-1.32	0.25	-1.50	0.36	-0.36	2.42
Effect at weighted average score	-0.89	0.17	-1.01	0.24	-0.24	1.63

Pet

Binary	Model 7	
Model	Baseline	Baseline with all controls
DiD (compliance score)	5.855	10.569
	(12.841)	(6.619)
Number of observations	360,535	202,195
With other controls	No	Yes
Effect at 50 points	2.93	5.28
Effect at weighted average score	1.60	2.90

Note: * significant at 1%, ** significant at 5%, * significant at 10%. Brand-clustered standard errors in parentheses. Note we do not estimate significance for the interpretations of our estimated effect at 50 points or the average compliance score.**
Source: FCA analysis of transaction data

Interpretation of our findings

We have found some evidence consistent with our intervention, on average, being associated with firms adjusting the scale of the premium renewal they offer to consumers. We have also found some evidence consistent with our intervention causing the proportion of consumers switching or negotiating their insurance policy to increase or decrease.

Our findings are our best estimates of the impact of our intervention. While using clustered standard errors, some of our results are not statistically significant. This should not be taken as 'no effect', but means the point estimates are subject to uncertainty and we have presented ranges in the main report. There is some variation when we examine different time periods and firm populations, which could suggest the intervention had an uneven effect on firms.

One interpretation of our findings could be that firms first adjusted their renewal premium offers in response to our intervention, and then consumers reacted in new ways to a combination of the effects of the renewal premium offer and our intervention on the renewal letter. This could explain, in particular, why we observe both a decrease in renewal premium changes and switching and negotiating rates in the home insurance market – consumers saw lower premiums and were less inclined to switch or negotiate in response to these lower premiums.

However, to the extent that our results do represent causal effects on the supply and demand-side of the market, the sequencing of impacts cannot be observed. Because our results are formed of multiple time periods, it is not necessarily the case that the estimated premium change and a consumer reaction occurred at the same time. And it is not clear that the firms that drive the average results on switching are the same firms that drive the average results for switching.

It is important to note that our analysis focuses on the short-term effects of our intervention, and the reaction of firms and consumers just after implementation. It is possible that the longer-term effects could differ, though our robustness checks in this area were inconclusive (see following section).

Further robustness checks of our analysis are contained in Section 5, with results from our alternative econometric analysis in Section 6. Having applied these robustness tests, we consider that our analysis gives a good estimation of the likely impact, in terms of direction and scale, of our intervention.

Section 5: Robustness checks

Overview

The empirical results presented above outline the main findings from our analysis. To ensure these results are robust, we have undertaken several additional checks including:

- Placebo tests
- Variation in the definition of 'negotiation'
- Other robustness checks

Overall, we conclude that our robustness checks support our main findings. The results of the placebo tests for our pet sample suggest the parallel trends assumption may not hold as well as it does for home and motor, so invites some caution in the interpretation of the pet findings.

Placebo tests

Our DiD estimator measures the differences in mean outcomes between firms that had or had not yet implemented our intervention. We assume that the firm that implemented the intervention would otherwise have had outcomes that followed the same trend as the untreated firm. However, it might be that a third factor, which changes over time but is unrelated to our intervention, is the underlying cause of any differences in treated and untreated outcomes. We have therefore undertaken 'placebo tests' of our intervention.

Our placebo tests repeat our analysis on a subset of a dataset in which no intervention took place. They attempt to test whether any treatment effect can be estimated where one would not be expected to exist. If a placebo test estimates a statistically significant treatment effect, it can suggest the presence of time-varying factors or underlying differences in trends between firms that may be affecting our results. Placebo tests are therefore a form of diagnostic tool for the parallel trends assumption that underpins DiD models.

To undertake placebo tests for our analysis, we analyse time periods where no firms implemented any changes to their renewal notices. We have performed our placebo tests over two periods:

1. February 2017, a period before most firms implemented the intervention.
2. April 2017, a period just after the rules came into effect, where no firms made any modifications to their renewal notices.

To mimic our two-way fixed effects model, we split each period into 3 randomly assigned sub-periods of between 3 and 8 days. We then randomly allocate firms an artificial compliance date at the start of one of the sub-periods. We repeat this random allocation process 100 times, and in each case we run our binary models without covariates. The construct allows us to test whether our randomly allocated artificial treatment is associated with changes in our dependent variables.

Our placebo tests are not directly comparable to our main result, but provide a good approximation. Firstly, there could have been greater volatility in outcomes in the placebo test periods than in the time period used for our main analysis. Secondly, in our two-way fixed effects models the number of 3-day sub-periods, which we hypothesise to be the most susceptible to statistical noise, is higher in our placebo tests than in our main analysis.[49] Overall, however, these differences are relatively minor and a priori we would expect the results of our placebo tests to be able to indicate whether our control groups are appropriate.

Table 12 presents the coefficients on the DiD estimator from the 100 iterations of our placebo tests. For switching/negotiating as the dependent variable, we find a statistically significant coefficient in 5%, 6% and 9% of cases for the home, motor and pet insurance markets respectively before firms implemented the intervention. In the period after firms implemented the intervention, we find a significant coefficient in 6%, 3% and 10% of tests. In both cases, the average coefficient is close to zero, as would be consistent with results driven by statistical noise.

For premium difference, the estimated DiD coefficient is statistically significant at the 5% level in 3%, 2% and 10% of cases for the home, motor and pet insurance markets respectively before firms implemented the intervention. In the period after firms implemented the intervention, we find a significant coefficient in 6%, 2% and 7% of tests. As with switching and negotiating, the average estimated effects are small.

Table 12: Coefficients on placebo test results for switching and negotiating rate and 'price difference paid'

Switching and negotiating

Market	February 2017 pre-intervention		April 2017 post-intervention	
	Average coefficient	% of stat. sig. results	Average coefficient	% of stat. sig. results
Home	-0.0002	5%	0.0012	6%
Motor	0.0006	6%	0.0006	3%
Pet	0.0010	9%	-0.0010	10%

49 In addition, because of the relatively short period in which we run the placebo tests, we also use a total of 3 time periods in the placebo test regressions, whereas our main results use between 3 and 5.

Premium difference

Market	February 2017 pre-intervention		April 2017 post-intervention	
	Average coefficient	% of stat. sig. results	Average coefficient	% of stat. sig. results
Home	-0.2372	3%	0.3092	6%
Motor	0.1447	2%	-0.1356	2%
Pet	-0.0591	10%	-0.0079	7%

Note: Statistical significance is defined at the 5% level. The coefficient for switching/negotiating represents a change to a rate (multiply by 100 for percentage point change), and for premium difference represents a £ change.

Source: FCA analysis of transaction data

The results of our placebo tests are largely in line with those that would be expected if the parallel trends assumption holds. We would expect roughly 5% of DiD coefficients to be statistically significant at the 5% level as the result of chance. This broadly matches our results.[50] Notably, the percentage of significant results in the pet insurance market is higher than 5% across all of our tests, suggesting that the parallel trends assumption may not hold as well in our pet insurance sample. We therefore should interpret our results for the pet insurance market with some caution. Overall, the fact that average coefficients are an order of magnitude less than are main results provides some further reassurance of our main results.[51]

Variation in the definition of 'negotiation'

As another test of the robustness of our main results, we used an alternative definition of our negotiation variable. The definition of 'negotiation' in our data request was whether the consumer made contact with the firm to negotiate the quoted premium or coverage amount for the policy. Some firms did not record this information and provided a proxy such as whether the consumer made contact with the firm for any reason between the renewal notice being sent and the policy renewal date. A few firms were not able to supply any information about negotiation.

Given these data quality issues, we checked whether an alternative definition of consumer negotiation would affect our findings. Our alternative measure is to record any difference of more than £2 between the renewal notice offer and the eventual premium paid as a negotiation. The £2 difference should allow for any recording inaccuracy in our transaction data.

50 We have not attempted to quantify the probability of achieving our main results by chance. Since we have run multiple two-way fixed effects models to take advantage of variation in our sample, the probability of achieving those multiple results would be different to the probability of achieving a single result. In addition, the necessary differences in the construction of our main results and our placebo tests would affect the probability in unknown ways.

51 In preliminary work we ran the same placebo tests without clustering standard errors by brand. We found a much higher proportion of statistically significant results – between 7% and 50% in the regressions on premium difference, and between 15% and 35% for switching and negotiation. However, the average coefficient for the statistically significant placebo tests was low compared to our main results. The difference between the un-clustered results and clustered results could be consistent with serial correlation in our outcome variables, high volatility in outcomes, or a deviation from the parallel trends assumption. We therefore clustered standard errors in our main analysis, even though this may reduce our chance of finding a statistically significant 'true' effect of our intervention.

Our alternative measure captures all non-negligible price changes between offer and renewal, regardless of the reason. The definition includes premium negotiations by consumers but also policy coverage changes that change the price. Since our intervention included text to encourage consumers to check the appropriateness of their insurance cover, one outcome could be consumers adjusting their coverage details (such as providing updated information on their situation, or choosing options that match their preferences). Disadvantages of the alternative measure are that it is unable to detect negotiation where price stays the same, or unsuccessful negotiation attempts. It is also vulnerable to inaccuracies in responses to our data request.

Table 13 provides the results of our switching and negotiating model using the alternative definition of negotiation, comparable with Table 7. The results are of the same direction and of very similar magnitude as in our main analysis so do not cause us to amend our conclusions. As with our main results, only the results for the pet insurance market are statistically significant after clustering standard errors.

Table 13: Estimated effect on combined switching/negotiating rate - binary regression results using alternative negotiation measure

Home

Binary	Model 1		Model 2		Model 3	
Model	Baseline	Baseline with all controls	Baseline	Baseline with all controls	Baseline	Baseline with all controls
DiD (binary)	-0.018	-0.017	-0.031	-0.03	-0.011	-0.009
	(0.0141)	(0.0135)	(0.0203)	(0.0198)	(0.0166)	(0.014)
Number of observations	648,629	644,976	775,893	772,345	316,601	315,119
Other controls	No	Yes	No	Yes	No	Yes

Motor

Binary	Model 14		Model 5		Model 6	
Model	Baseline	Baseline with all controls	Baseline	Baseline with all controls	Baseline	Baseline with all controls
DiD (binary)	0.019	0.015	0.021	0.018	0.023	0.016
	(0.0225)	(0.0186)	(0.0213)	(0.0171)	(0.0253)	(0.0177)
Number of observations	1,754,792	1,753,990	1,371,883	1,371,074	1,280,132	1,279,584
Other controls	No	Yes	No	Yes	No	Yes

Pet

Binary	Model 7	
Model	Baseline	Baseline with all controls
DiD (binary)	0.009***	0.015**
	(0.0033)	(0.0069)
Number of observations	371,118	208,809
Other controls	No	Yes

Note: * significant at 1%, ** significant at 5%, * significant at 10%. Brand-clustered standard errors in parentheses.**
Source: FCA analysis of transaction data

Other robustness checks

Reduced sample

An additional robustness check, we tested was to remove individual firms from our DiD regression specifications which we considered to be at risk of violating the parallel trends condition. As discussed in Section 3, the parallel trends condition is the main assumption for a DiD model and its violation leads to a biased estimation of the causal effect. We removed both firms which displayed differing pre-intervention trends and those which showed a high degree of post-intervention volatility. In cases where the removal of firms resulted in a lack of variation, we adjusted the models' time periods accordingly.

The results of these regressions were largely consistent with our main findings so do not cause us to amend our conclusions. We removed two, three and three firms from the home, motor and pet insurance markets respectively. In all markets, the results were of the same direction although with a slightly higher magnitude for our premium difference binary results in motor and pet.

Group-specific linear trends

We tested whether introducing a group-specific linear time trend would affect our results. This is a common robustness check in difference-in-differences analysis that helps test whether treatment and control groups were already on differential paths before the intervention. If they were, the DiD results could represent a change that would have happened anyway. Adding group-specific linear time trends to the regression specification can control for this possibility.

The addition of brand-specific linear time trend does not affect our conclusions. We tested adding brand-specific time trends by splitting the first pre-intervention period in our two-way fixed effects models into two separate periods. This allows two pre-intervention observations on which to base the time trend. The direction of every coefficient was the same as in our main results and the magnitude was always similar (within 0.5 percentage points for the switching/negotiating binary results). Standard errors often changed by a larger amount, which affects the statistical significance and confidence intervals for some of our results.

Longer time period

Our DiD approach exploits the different dates that firms implemented our renewals rules. Because of the implementation pattern, our analysis is focused on estimating the short-term effects of our intervention in the period just following implementation. But it is possible the longer-term effects varied. We were not able to replicate our approach over a longer time frame but as a robustness check we changed the final time period in our multi-period DiD designs from April 2017 to April 2018 (the same period one year later). The impact of including a later time period on estimated switching/negotiation and premium offered was not conclusive. Some findings suggested the estimated impact when including a 2018 period was more modest than our main approach, but the difference depended on the model and market studied, with no consistent pattern.

Section 6: Alternative econometric analysis

Overview

We used an alternative econometric approach to complement our main econometric analysis presented above. The reasons for this were to:

- validate the results of our main analysis (ie whether we estimate similar results by using a different approach)
- understand the impact of our remedies at a firm level

This alternative econometric approach uses the regression discontinuity design (RDD) method.

Overall, our analysis using the RDD method:

- found estimates of the treatment effect which suggested that our intervention's impact varied by firm. The results were mostly within the range of expectations (ie based on those identified in the main analysis). Direct comparisons, however, should be made noting that our main econometric analysis produces a weighted average of impacts across data from several firms whereas the RDD analysis produces results on an individual firm basis.
- produced a number of non-statistically significant results. The likelihood of a non-statistically significant result increased as we reduced the time window over which we conducted our analysis and added more control variables. Although it is difficult to say, we believe the prevalence of non-significant results were partially driven by noise in the data and the tight time window over which the data for the analysis was restricted, which was required to ensure other time varying effects did not impact on the estimate of the treatment effect.
- leads us to consider that the findings from the alternative econometric analysis help validate our choice of approach in the main analysis and emphasise the value of using a method where a control and treatment group can be compared over time.

Methodology

Similar to DiD, the RDD method[52] is commonly used by practitioners to assess the impacts of policy interventions in non-experimental settings.[53]

52 The regression discontinuity design was first introduced in: Thistlethwaite & Campbell, 1960, 'Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment'. For further information on the general RDD methodology see: Angrist & Pischke, 2009, 'Mostly Harmless Econometrics, An Empiricist's Companion', p.187-202; and Wooldridge, 2009, 'Introductory Econometrics: A Modern Approach. Fourth Edition', p.954-959. For guidance on the implementation of RDD and its use in economics more broadly see: Imbens & Lemieux, 2008, 'Regression discontinuity designs: A guide to practice'; and Lee & Lemieux, 2010, 'Regression Discontinuity Designs in Economics. Discussion of regression discontinuity in time' can be found in: Hausman & Rapson, 2017, 'Regression discontinuity in time: considerations for empirical applications'.

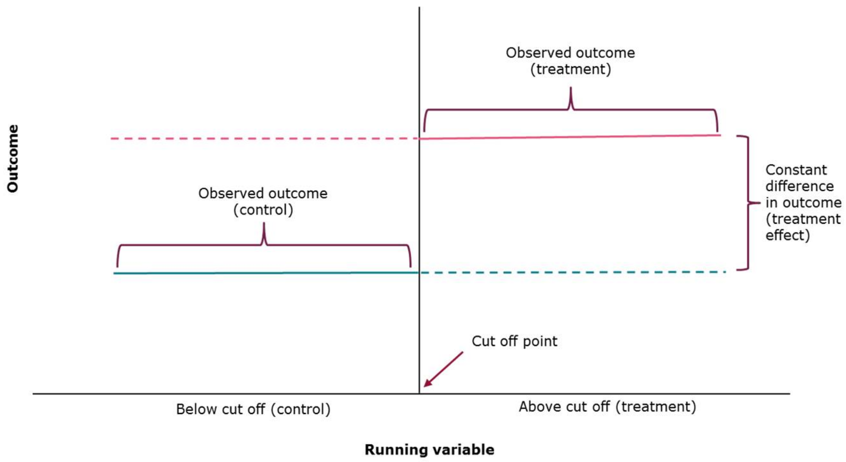
53 For selected examples of the RDD method used in the academic literature see: Lee, 2008. Randomized experiments from non-random selection in U.S. house elections; and Angrist & Lavy, 1999,

Our analysis uses the 'sharp' RDD method.[54] This method relies on the treatment being decided by some observed running variable (also known as the forcing variable). In the case of the sharp RDD, this variable is deterministic (ie an individual is either treated or not treated).

Individuals are classified as treated when the running variable exceeds some known level (also known as the cut-off). At the point of the cut-off, we would expect a discontinuity in the outcome of interest for individuals either side of the threshold. However, we would expect such individuals to be otherwise similar (ie have the same or very similar characteristics). If this is the case, then the treatment assignment can be viewed as being as good as random. To estimate the treatment effect, we would compare outcomes for individuals just before the cut-off (ie the untreated) to those just after the cut-off (ie the treated).

Figure 4 illustrates the theory described above. The difference at the cut-off between the observed outcome before the cut-off and the observed outcome after the cut-off can be interpreted as the treatment effect. Practically, this treatment effect can be estimated using econometric analysis, either parametric or non-parametric.[55]

Figure 4: Representation of sharp RDD method



Source: Adapted from Lee & Lemieux (2008)

Implementing the method

We apply the RDD approach by creating a running variable based on when firms in our sample updated their renewal notices to implement our intervention. The date this

Using Maimonides' rule to estimate the effect of class size on scholastic achievement. A list of papers using a regression discontinuity in time approach can be found in Table 1 of Hausman & Rapson (2008).

54 There are 2 types of RDD method, the 'sharp' RDD and the 'fuzzy' RDD. In both cases, the method relies on the treatment being decided by some observed running variable (also known as the forcing variable). However, in the case of the fuzzy RDD method, the running variable may not be deterministic (ie an individual is more or less likely to be treated but it cannot be determined if they have been treated for sure). We have not used the fuzzy RDD method as we have been given dates by firms when they updated their renewal notices for all customers and, subsequently, filtered out those where a sharp implementation was not applied.

55 Parametric analysis makes an assumption about the distribution of the data from which the sample analysed is gathered (eg that the data are normally distributed). Non-parametric analysis makes no assumption about the distribution of the data from which the sample analysed is gathered.

came into effect was provided to us by firms. For each firm, the running variable was calculated as the number of days away from the date a consumer was sent a renewal notice. Those consumers that were sent the updated renewal notice would have a positive running variable equal to the number of days between the firm's implementation and when their renewal notice was sent.[56] Correspondingly those receiving the original notice would have a negative variable. By design, the value of the running variable for the implementation date itself was set to 0.[57]

Understanding how far away an individual is from the cut-off is an important factor in our analysis. When conducting analysis using a sharp RDD method, we should consider the trade-off between comparing individuals who are alike and the number of individuals who we are able to compare. As the running variable moves further away from the cut-off, individuals are less likely to be similar. There is, therefore, a higher chance of there being confounding factors biasing our estimate of the treatment effect. However, being closer to the cut-off is likely to limit, or reduce, the number of individuals that we can include in our analysis. This means that there is a greater chance of not finding a treatment effect when one exists.

In our analysis, this trade-off is particularly relevant as our running variable is based around time (ie the number of days before or since a firm updated their renewal notice to implement our rules). The further away individuals are from the cut-off, the more likely it is that time-varying effects may lead to bias in the estimate of our treatment effect (eg through serial correlation).[58]

As a result, to balance the need for more data with limiting any potential bias in our results due to time-varying effects, we used graphical and regression analysis, using data restricted to 2 different time periods:

- 4 weeks either side of the cut-off
- 1 week either side of the cut-off

Before carrying out any analysis, we filtered brands that we considered could be suitable for our RDD analysis. We did this based on how these brands implemented the intervention and the data provided to us. Starting with all brands in our sample, we removed brands where:

- there was a phased implementation of the updated renewal notice (ie no sharp change in treatment)
- there was no change in implementation of the intervention over the time period for which we had requested data (ie implementation didn't take place, or took place outside of the period for which we requested data)
- there were significant data issues in the analysis time window which resulted in inconclusive graphical or regression analysis

56 For some firms, there is some uncertainty over whether consumers sent their notices around or just after the date renewal notices were updated received the updated or original renewal notice. In our DiD analysis we have accounted for this by undertaking additional regressions where this potential uncertainty is accounted for. For the purposes of the RDD method we have sought to remove firms where this fuzziness is present based on the implementation dates that were provided to us and chosen to assume that such fuzziness is not significant for other firms.

57 This is in line with the suggested adjustment for the running variable to rebase the cut-off as equal to 0 outlined in Imbens & Lemieux (2008).

58 For further discussion on this issue in relation to regression discontinuity in time see Hausman & Rapson (2017).

Following this process, there were 16 brands remaining which we considered could be viable candidates for us to conduct the RDD analysis.

Specification

Our econometric specification for the sharp RDD method is set out below.

We estimate the treatment effect, parametrically, using a pooled ordinary least squares regression.[59] The impact of our intervention, controlling for other factors, is estimated by the coefficient (labelled as 'RDD' in our regression results presented below) on a binary variable representing whether an individual received a renewal letter before or after the cut-off (D_{it}).

We also include a polynomial transformation of the running variable to account for the general trend in outcomes but-for the treatment effect and other control variables to account for individuals further from the cut-off.

A statistically significant RDD variable (D_{it}) would indicate that our intervention has had a causal impact on the outcomes of interest.

$$\boldsymbol{y_{it} = \alpha + \beta^{RDD} \cdot D_{it} + f(\delta_{it}) + \theta \cdot X_{it} + \varepsilon_{it}}$$

Where:

- i is the individual consumer observation
- t is the time period
- α is a constant
- y_{it} is the outcome or dependent variable for individual *i* at time *t*. For switching and negotiating, this is 1 if an individual switched or negotiated, and 0 otherwise. For premium difference, this is the difference in renewal offer in pounds for individual i.
- $f(\delta_{it})$ is a polynomial function of the running variable. The running variable is the number of days before or after the date that the firm changed their renewal letter to implement our intervention.
- D_{it} is the 'treatment' variable. This is a treatment dummy variable equal to 1 if the consumer was exposed to the post-intervention renewal letter, and 0 otherwise.
- X_{it} represents the set of control variables
- ε_{it} is the error term

As the estimate of the treatment effect is sensitive to the polynomial transformation applied to the running variable, we ran regressions with polynomial transformations ranging from linear (ie order 1 polynomial) to quartic (ie order 4 polynomial). We include a set of control variables, X_{it}, to control for observable systematic factors that

59 Another parametric approach frequently used is to estimate separate regressions either side of the cut-off without including an intercept. The treatment effect is then estimated as the residuals of the regression above the cut-off minus the residuals of the regression conducted below the cut-off. Estimating the treatment effect using an RDD method is also regularly estimated non-parametrically through local linear regression, localised around the cut-off. We chose our selected approach to allow for an easier estimation of standard errors and for clearer interpretability of the coefficient of interest. For further details on types of estimation, see Imbens & Lemieux (2008) and Lee & Lemieux (2010).

may differ before and after the cut-off for the time period of our analysis. We include only control variables that are reported by all the firms in our sample. This approach is in line with the approach taken in the main econometric analysis.

For comparability with the main econometric analysis, we use the same two groups of dependent variables (ie outcomes of interest) for our analysis:

1. The rate of switching or negotiating (either combined or separately)
2. The difference between the premium offered and previous premium paid, or the difference between the previous and current premium paid.

Assumptions

In comparison to the DiD and other non-experimental methods, the sharp RDD method relies on relatively limited assumptions. When estimating causal treatment effects, there are two standard assumptions which are generally applied:[60,61]

- the unconfoundedness assumption (also known as the ignorability or exogeneity assumption) assumes that if we are comparing outcomes between two groups, we can ignore how individuals ended up in each group (ie by controlling for everything else that may be different between the two groups, any remaining difference is a treatment effect)
- the overlap assumption requires that we observe individuals who are both treated and untreated for a given value of any covariate (ie we see individuals who received the updated renewal notice and the original renewal notice for any day in our dataset)

Although the sharp RDD method meets the unconfoundedness treatment (by design as individuals just either side of the cut-off are assumed to be almost the same but for the different type of renewal notice received), it violates the overlap assumption. By definition of the running variable, which is binary based on whether an individual received the original or updated notice, there are no individuals who receive both the original renewal notice and the updated renewal notice. As a result, to estimate a treatment effect, we need to compare those who received the original notice with those who received the updated notice and assume they are otherwise the same.

The main assumption needed to allow for this comparison is the continuity assumption. All other factors are assumed to be continuous with respect to the running variable (ie other consumer characteristics and demographics have the same impact either side of the cut-off). As a result, although we do not observe individuals who received the original and updated notice for a given day in our data set, we can, under the continuity assumption, compare outcomes for individuals who received the original notice just before it was changed and for individuals who received the updated notice just after it was changed. We then estimate the average treatment effect as the average difference between these 2 groups.

60 These assumptions follow from work undertaken by: Rubin, 1974, 'Estimating causal effects of treatments in randomized and non-randomized studies'.

61 For further information on the definitions of these assumptions and how they apply in an RDD setting, see Imbens & Lemieux (2008) and Lee & Lemieux (2010).

Another important assumption for our RDD analysis is no manipulation. This requires that individuals cannot select themselves into a control or treatment group (ie a consumer doesn't request a renewal notice early or switch firms ahead of renewal to avoid receiving the updated form of the renewal notice). As explained previously in this annex, we are reasonably confident that consumers have little opportunity to anticipate and act upon the content of their upcoming renewal notice before it arrives. This is especially because a firm's implementation date was unknown in advance, as firms decided when they were implementing the change to their renewal letters.[62]

Finally, it could be that other coincidental, but otherwise unrelated, discontinuities took place spuriously around the time of the cut-off. For example, firms may have made a change to their strategy, or an external event may have driven a change in outcomes. This may confound our treatment effects. To test for this, we have conducted analysis looking at explanatory variables, such as the consumer's age, to separate any discontinuity around the time of the cut-off with any discontinuities found for the outcome variables.

Serial correlation & time varying treatment effects

As our running variable is based on a variation in time, as opposed to variation on a cross-sectional variable, there is the potential for serial correlation and time-varying treatment effects.[63]

To account for this, we have assumed a generally smooth and constant treatment effect. We have, therefore, run our analysis within a relatively tight window of 1 week and 4 weeks either side of the cut-off. This limits the potential for observing time-varying effects in our estimation of the treatment effect.

Graphical analysis results

Ahead of running the RDD regressions, it is common and helpful to, first, plot charts of the outcome variables against the running variables (what we subsequently refer to as an RDD plot). This is typically done by dividing the running variable into fixed 'bin sizes' (ie intervals) and calculating averages of the outcome variable over each of these bins. We present this as a scatter chart, where the outcome is presented on the vertical axis and the running variable is presented on the horizontal axis. A polynomial trend line, estimated separately for observations below and above the cut-off, is shown to outline the general functional form of the data.

62 Although consumers were unlikely to be able to manipulate their behaviour to avoid or delay the intervention firms may have had an incentive to implement the intervention in such a way as to impact the outcomes. This issue is explored previously in the annex in the 'strict exogeneity and stable units' sub section.

63 Hausman & Rapson (2017) outline these two factors as issues which are specific to the regression discontinuity in time model and which need to be considered.

The purpose of undertaking this analysis is to observe any discontinuity generated around the cut-off[64] and to ascertain if the functional form is the same both above and below the cut-off.[65]

Before running any regressions, we produced RDD plots for each of the 16 brands analysed across each of the 4 outcome variables of interest. These charts were put together in different forms, including:

- varying the polynomial transformation on the running variable between linear, quadratic, cubic and quartic
- estimating a linear regression of the outcome of interest including all controls, saving the residuals and then reproducing the charts using the residuals (ie showing the outcome after controlling for other factors)

We produced the same graphs by replacing the outcome variables with the various explanatory variables that we used as controls in our regression. The aim of this was to see if there were any instances of spuriousness of any discontinuity seen in the outcome variables. We also did this to see if there was any sorting behaviour by consumers, which may bias our treatment effect estimates.

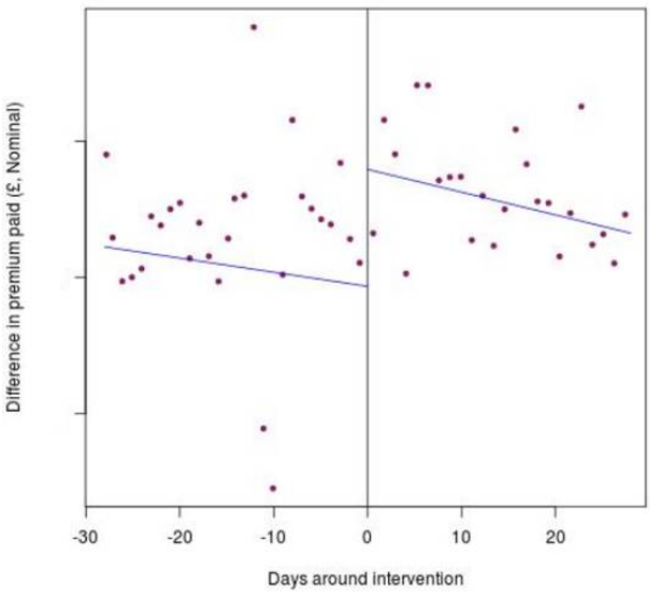
Figure 5 gives an example of the RDD plots that we produced for the difference between the premium offered this year and premium paid last year (ie the premium difference outcome variable) with a linear trend line. Figure 6 shows the same, but with a quartic trend line.[66] Although the scatter plots are the same, the polynomial transformation applied to the trends visually indicate different conclusions about the discontinuity and general pattern of the data.

64 A formal test of the presence of a discontinuity can be conducted using the Mcrary (2008) test. This test helps to check for sorting behaviour. However, this approach is not valid in a world where the running variable is based around time rather than a cross section. As such it is not presented here. Instead as suggested by Hausman & Rapson (2017) we have checked discontinuities for other explanatory variables to check for sorting behaviour.

65 The RDD method, parametrically estimated using OLS is particularly sensitive to the choice of functional form. Misspecification of the functional form can result in an incorrect estimate of the treatment effect.

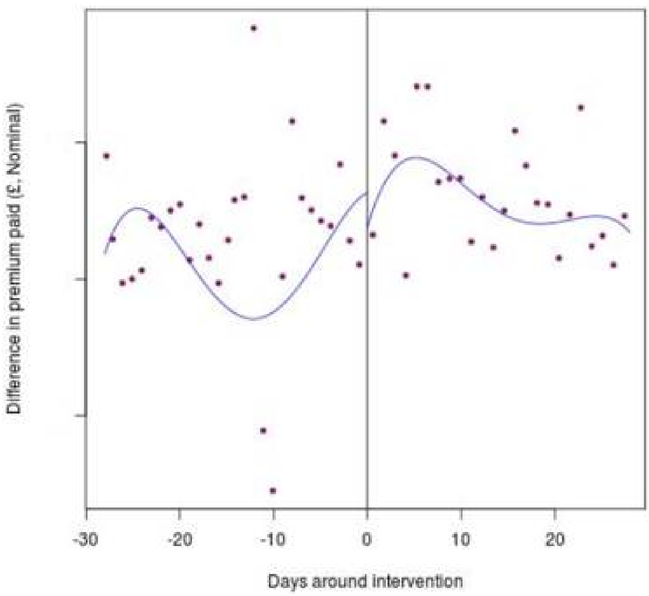
66 In total we produced over 500 charts studying outcome variables and over 250 charts studying explanatory variables. We have not presented all the charts for ease of reading.

Figure 5: Example of RDD plot for difference in premium paid including a linear polynomial trend line



Notes: To maintain anonymity, the scales and limits of the vertical axes are not labelled.
Source: FCA analysis of transaction data.

Figure 6: Example of RDD plot for difference in premium paid including a quartic polynomial trend line



Notes: To maintain anonymity, the scales and limits of the vertical axes are not labelled.
Source: FCA analysis of transaction data.

We reviewed the charts by firm and outcome to identify, visually, any discontinuity in the outcome around the cut-off. We also did this for the explanatory variable charts. Generally, the results of the graphical analysis were that:

- there was considerable noise around the cut-off, even with a small time window, and many firms showed mostly different trends in outcomes either side of the cut-off

- based on visual inspection, the choice of polynomial had a significant impact on the presence and size of a discontinuity around the cut-off
- across all outcomes and most firms, it was generally difficult to see a clear discontinuity in outcomes, suggesting such discontinuities may not be present
- in a small number of cases, the presence of a discontinuity was clearer, suggesting that these might be more suitable candidates for the RDD method

Regression analysis results

This section presents a summary of the results of our RDD regression analysis and sets out possible interpretations of the findings.

We present results for 2 models. Our baseline model includes only the RDD estimator and polynomial transformations of the running variable. A second version adds our full list of controls (see Table 14; these are similar to those used in our main analysis).[67]

Given the output of the graphical analysis, we ran regressions for all brands and outcome variables with a range of polynomial transformations. We present a summary of regression results below for:

- switching and negotiating
- difference between premium offered this year and premium paid last year (premium difference)

These outcomes were chosen to align with the main analysis.[68]

The summary results tables presented below, by market, summarise:

- the number of regressions estimated
- the number of statistically significant findings for the RDD variable from these regressions
- the minimum and maximum value of the coefficients for these regressions which were statistically significant

We present results for a linear polynomial transformation and a quartic polynomial transformation of the running variable.[69] We used data from 4 weeks either side of the cut-off.[70]

67 This list does not include variables that we would have wanted to include as control variables, but data quality issues would have meant excluding firms from our sample. We have not included the coefficients from all the control variables in our empirical results but summarise which controls were used.

68 Results for switching and premium paid were also produced but have not been included for ease of reading this annex.

69 Results were also produced using quadratic and cubic polynomial transformations. Graphical analysis indicated that linear and quartic transformations fit the data better than quadratic, whilst quartic transformations showed similar findings to cubic transformations. The results of these regression did not fundamentally differ from the findings presented here and these results have not been included for ease of reading this annex.

70 We conducted analysis for shorter time periods including 1 week either side of the cut-off and 3 days either side. In both cases, the results were less likely to be statistically significant than those presented here and have not been included for ease of reading.

Table 14: Variables included in the regression specification

Variable	Description	Baseline	Baseline with all controls
RDD estimator	Dummy variable identifying those above and below cut-off	✓	✓
Transformed running variable	Polynomial transformation of the running variable	✓	✓
Expected claims cost (*premium difference only*)[71]	Expected cost of the insurance policy to the firm		✓
Number of previous renewals	How many times the consumer has renewed before		✓
Auto-renewing	Whether the contract automatically renews at termination		✓
Consumer age	Age of consumer		✓
Policy coverage type	Dummy variable for each policy coverage type outlined in our data request		✓
Weekday	The weekday the renewal notice was sent		✓

Source: FCA

Switching and negotiating results

Table 15 shows the results of the RDD analysis for the outcome of switching and negotiating, where the running variable is transformed by a polynomial of order 1. Table 16 contains the same as Table 15, but with the running variable transformed by a polynomial of order 4.

There were:

- 3 statistically significant findings in the home insurance market, using baseline regression model, and 2 statistically significant findings using all controls in the regression model.
- 2 statistically significant findings in the motor insurance market.
- no statistically significant findings in the pet insurance market.

The range of coefficient magnitudes for those regressions where the RDD variable was statistically significant was largely within the range of expectations (ie from our main analysis for the DiD variable). Increasing the order of the polynomial transformation

71 Our preferred control for consumer risk in our premium difference specification would be the difference between current and previous expected claims cost as it captures the full effect of the difference over time. We use the current expected claims cost only, as many firms have been unable to provide an accurate version of the previous expected claims cost. In many cases, firms have been unable to apply mid-term adjustments to the previous expected claims cost so it is inconsistent with the previous premium paid which is adjusted in this way.

generally decreased the number of statistically significant findings in the home and motor insurance markets. However, the number of statistically significant findings in the pet insurance market baseline increased to 2, before falling back to 0 when we added more controls. Magnitudes for these results were still largely in keeping with the main analysis.

Table 15: Regression results for switching and negotiating with running variable polynomial transformation of order 1

Market	Total number of brands analysed	Total number of statistically significant RDD variable	RDD coefficient value for statistically significant regressions	
			Minimum	**Maximum**
Home (baseline)	6	3	0.01	0.05
Home (all controls)	6	2	-0.02	0.03
Motor (baseline)	7	2	-0.02	0.02
Motor (all controls)	7	2	-0.02	0.02
Pet (baseline)	3	0	0.00	0.00
Pet (all controls)	3	0	0.00	0.00

Source: FCA analysis of transaction data.

Table 16: Regression results for switching and negotiating with running variable polynomial transformation of order 4

Market	Total number of brands analysed	Total number of statistically significant RDD variable	RDD coefficient value for statistically significant regressions	
			Minimum	**Maximum**
Home (baseline)	6	1	0.03	0.03
Home (all controls)	6	1	0.03	0.03
Motor (baseline)	7	1	0.03	0.03
Motor (all controls)	7	2	-0.01	0.04
Pet (baseline)	3	2	0.02	0.03
Pet (all controls)	3	0	0.00	0.00

Source: FCA analysis of transaction data.

Premium difference results

Table 17 shows the results of the RDD analysis for the outcome of premium difference, where the running variable is transformed by a polynomial of order 1. Table 18 contains the same as Table 17, but with the running variable transformed by a polynomial of order 4.

There were:

- 2 statistically significant findings in the home insurance market.
- 1 statistically significant finding in the motor insurance market
- 2 statistically significant findings in the pet insurance market.

These findings used the baseline model. In all cases, the range of coefficient magnitudes for those regressions where the RDD variable was statistically significant was largely within the range of expectations (ie from our main analysis for the DiD variable). Introducing controls and increasing the order of polynomial transformation did affect the statistical significance of the findings and led to a change in the magnitude of the coefficients.

Table 17: Regression results for premium difference with running variable polynomial transformation of order 1

Market	Total number of brands analysed	Total number of statistically significant RDD variable	RDD coefficient value for statistically significant regressions	
			Minimum	Maximum
Home (baseline)	6	2	3.92	6.32
Home (all controls)	6	1	4.84	4.84
Motor (baseline)	7	1	-2.84	-2.84
Motor (all controls)	7	0	0.00	0.00
Pet (baseline)	3	2	6.61	11.69
Pet (all controls)	3	2	7.70	11.40

Source: FCA analysis of transaction data.

Table 18: Regression results for premium difference with running variable polynomial transformation of order 4

Market	Total number of brands analysed	Total number of statistically significant RDD variable	RDD coefficient value for statistically significant regressions	
			Minimum	Maximum
Home (baseline)	6	1	3.40	3.40
Home (all controls)	6	0	0.00	0.00
Motor (baseline)	7	1	-4.49	-4.49
Motor (all controls)	7	1	-6.66	-6.66
Pet (baseline)	3	2	7.92	8.24
Pet (all controls)	3	2	9.51	9.54

Source: FCA analysis of transaction data.

12 Endeavour Square, London E20 1JN
Telephone: +44 (0)20 7066 1000 Website: www.fca.org.uk