OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305–9049
doi: 10.1111/obes.12204

What Drives Fraud in a Credence Goods Market? – Evidence from a Field Study*

Alexander Rasch†,‡ and Christian Waibel§

†*Duesseldorf Institute for Competition Economics (DICE), University of Duesseldorf, Universitaetsstrasse 1, 40225 Duesseldorf, Germany (e-mail: rasch@dice.hhu.de)*
‡*Centre for European Economic Research, L 7, 1, 68161 Mannheim, Germany*
§*ETH Zurich, Zuerichbergstrasse 18, 8092 Zurich, Switzerland (e-mail: cwaibel@ethz.ch)*

Abstract

This paper investigates the impact of competition on an expert firm's incentive to defraud its customers in a credence goods market. Controlling for the competence of car repair shops, their financial situation, and reputational concerns, we use and complement the data set from a nationwide field study conducted by the German Automobile Association that regularly checks the reliability of garages in Germany. We find that more intense competition lowers a firm's incentive to defraud its customers.

I. Introduction

Making use of a field study in the German market for car repairs, we analyse the impact of competition on an expert firm's incentive to defraud its customers in a credence goods market.[1] In credence goods markets (Darby and Karni, 1973), fraud may arise due to asymmetric information between the expert and the customer: The expert knows the quality of the good the customer needs and, in most cases, gives a treatment recommendation based on a diagnosis and provides a treatment. The customer, however, does not know the quality required and therefore must rely on the expert's advice.

Our empirical analysis of corporate car repair firms shows that a higher degree of competition lowers a firm's incentive to defraud its customers. A larger number of competitors in the market reduces customers' search costs for obtaining a second opinion and hence lowers experts' incentives to defraud their customers. In our analysis, we control for the

JEL Classification numbers: D82, L15.

*We thank Hans-Jürgen Andreß, Florian Gössl, David Jaeger, Rudolf Kerschbamer, Wanda Mimra, Michael Pfaffermayr, Lamar Pierce, Bettina Rockenbach, Henry S. Schneider, Matthias Sutter, Achim Wambach, Roberto Weber, Achim Zeileis, and seminar audiences in Cologne, Duesseldorf (DICE), Munich (LMU), and Vallendar (WHU) for their helpful comments and discussions. We also thank two anonymous referees for their helpful suggestions and Lisa Boxberg, Sarah Dahmen, Nicolas Fugger, Björn Haist, Marlene Scholz, and Dennis Uiberall who provided excellent research assistance. Part of this research was done while Christian Waibel was visiting the University of Innsbruck. He would like to express his sincere thanks to the University of Innsbruck for its hospitality throughout his stay.

[1] See Dulleck and Kerschbamer (2006) for an overview of these markets.

financial situation of garages, their competence, and reputational concerns. The results suggest that a critical financial situation and low reputational concerns increase expert fraud, whereas a high level of competence reduces such fraud.

Fraudulent behaviour and faulty repairs are major issues in the car repair market (Wolinsky, 1993, 1995; Titus, Heinzelmann and Boyle, 1995; Consumer Federation of America *et al.*, 2011). The market for auto repairs and the scope of fraud therein are important for two reasons: Firstly, the market itself is an important economic sector in industrialized countries, and secondly, the insights from the functioning of this particular credence goods market may help us to better understand the occurrence of fraud in other expert markets. Many important markets exhibit credence goods properties, including the markets for healthcare, taxi services, legal advice, and other so-called 'professional services'. To analyse the fraudulent behaviour of experts, we make use of the results of a field study in the German car repair market that is carried out on an annual basis by the German Automobile Association (*Allgemeiner Deutscher Automobil-Club e. V.*, ADAC), Europe's largest automobile club.

ADAC's database contains information on fraud and the competence of firms. Specifically, the automobile club has recorded whether firms illegally charged customers for the provision of services, including charging for more repairs than were provided and for providing more services than were necessary. We extend this database by collecting the number of car repair shops in a 10-km distance from each firm's location in order to quantify the intensity of competition, as the level of competition among car repair shops is often regarded as an important issue in the competition policy debate.

The seminal theoretical contribution on fraud in the car repair market is Taylor (1995), who studies an expert's incentive to overcharge his or her customers. The author shows that under short-term contracts, experts will charge all customers for a treatment independent of whether a car is faulty or not. Consequently, all customers whose cars are not faulty are defrauded.

There have only been a few field studies focusing on the determinants of dishonest behaviour in markets for credence goods.[2] Balafoutas *et al.* (2013) conducted a field experiment on credence goods by investigating taxi services in Athens, Greece. Their analysis reveals that when passengers had poor information about optimal routes, they were taken on longer detours. The authors also point out that a higher (perceived) customer income increased the level of fraud. In a follow-up study, Balafoutas, Kerschbamer and Sutter (2017) show that when passengers explicitly stated that they would be reimbursed by their employer, they were almost twice as likely to be charged excessively compared to situations in which drivers did not have such information.

Similar to our paper, Schneider (2012) analyses data from a field experiment in which he visited garages undercover in order to investigate whether expert reputation can alleviate the efficiency problems arising from asymmetric information. He finds both pervasive overtreatment and undertreatment but no evidence that reputation helps to reduce these

[2] In addition to the literature on credence goods, our analysis is related to the literature on supplier-induced demand (see, e.g., Evans, 2000 for a seminal contribution). Empirical studies investigating the direct impact of informational asymmetries on physicians' behaviour include Domenighetti *et al.* (1993), Currie, Lin and Zhang (2011), and Schmid (2015).

problems.[3] Our study differs from the contribution of Schneider (2012) in that we study the impact of competition on the level of fraud. Dulleck, Kerschbamer and Sutter (2011) were the first to experimentally study the aspect of competition in a credence goods market, showing that neither competition nor reputation decreases an expert's incentive to overcharge in a market with liability. Liability itself, however, turns out to be a crucial instrument for the reduction expert fraud.[4]

The remainder of the paper is organized as follows. In the next section, we derive our hypothesis from the theoretical literature on credence goods. We describe the data set in section III. In section IV, we present our results and compare them to the theoretical predictions. We check the robustness of our results in section V. The final section concludes and discusses implications for other credence goods markets.

II. Hypothesis

We make use of the model developed by Wolinsky (1993) and slightly modify it following Dulleck and Kerschbamer (2006) in order to derive our hypotheses. Below, we present the basic underlying incentives that help to explain firms' incentives to defraud their customers.[5] In the market under consideration, there is a mass one of homogeneous customers (car owners) who may have a major problem that occurs with a certain probability. With the complementary probability, they experience a minor problem. The minor problem can be fixed through a major or a (cheaper) minor treatment, whereas the major problem can only be fixed through a major treatment.[6] Customers do not know which type of treatment they require. Customers have access to liable expert firms that are able to recommend the treatment needed. Liability implies that experts cannot provide a minor treatment to customers with a major problem, i.e., experts cannot undertreat their customers. Moreover, customers have the possibility to perform a costly search for a second opinion after receiving the first treatment recommendation. Depending on the verifiability of the actual treatment performed and the price markups for the two treatments, our framework allows for two types of fraud – overcharging (no verifiability) and overtreatment (verifiability and a higher markup for the major treatment).

In this market, two incentive-compatibility constraints play important roles. Firstly, in equilibrium, an expert consulted by a customer with a minor problem must be indifferent between recommending the minor and the major treatment. This trade-off is due to the fact that a truthful minor-treatment recommendation entails a lower markup with certainty,

[3] He also shows that there is a positive relationship between the level of capacity available at a garage at the time of the visit and the probability of a repair recommendation. Moreover, there is a repeat-business effect for the diagnosis fee.

[4] Other experimental studies include Kerschbamer, Neururer and Sutter (2016) and Mimra, Rasch and Waibel (2016a, b) (see Kerschbamer and Sutter, 2017 for an overview).

[5] An extensive review of the theoretical literature and a unifying model are given in Dulleck and Kerschbamer (2006).

[6] We apply the notion of minor and major treatment used frequently in the credence goods literature. In the real-life market that we analyse, the minor treatment corresponds to opting not perform a treatment, whereas the major treatment corresponds to performing a treatment.

whereas a major-treatment recommendation gives a higher markup[7] but may not always be accepted, as customers may reject the recommendation and search for a second opinion.

Secondly, in equilibrium, a customer faced with a major-treatment recommendation must be indifferent between rejecting and accepting the diagnosis. Hence, the customer compares the additional costs of searching for a second opinion with the expected savings from visiting a second expert firm.

We can now use these basic trade-offs to analyse the relationship between competition and fraud. Consider a situation in which a customer's search costs depend on the number of firms located in the customer's neighbourhood: The more garages in the neighbourhood, the lower the search costs. This may be due to the fact that customers can spend less time and effort searching for suitable experts. Then, ceteris paribus, customers can seek out for a second opinion at a lower cost as the number of firms increases. As a consequence, customers are more likely to reject a major-treatment recommendation. This in turn decreases the firms' incentive to defraud their customers. We can therefore state the following hypothesis:

Hypothesis. As the degree of competition among expert firms intensifies, firms tend to defraud their customers less.

III. Data

Sample

We make use of pooled cross-section data from ADAC's tests of car repair shops from the years 2006 and 2008–10; in 2007, there was no test.[8] The automobile club's data set provides information on 303 garages. We disregard 25 garages that belong to the same corporate entity because these observations are not independent with respect to their financial situation. We further restrict the sample to 134 corporate enterprises because of data availability and firm characteristics. Crucially, only corporate enterprises are required to publish data on their financial situation. As we shall see below, a garage's financial situation is an important predictor of its incentive to defraud. Thus, not considering the financial situation would lead to an omitted variable bias in the estimates. Moreover, we derive our theoretical predictions based on a model that assumes that firms operate under limited liability. This is the case for almost all corporate garages, but not for non-corporate garages. Hence, restricting the data set to corporate enterprises appears reasonable.

ADAC's data collection for the tests proceeds as follows.

1. Club members from all over Germany are asked whether they would like to participate in the garage test.
2. The automobile club determines whether customers' cars fit the test criteria. Specifically, the cars must be similar with respect to maintenance-related characteristics that could affect the effort and time required on the part of garages: All cars must

[7] Note that the markup is either higher per se (as is the case under verifiability and overtreatment) or because the expert performs the cheaper minor treatment but charges for the major treatment (as is true for non-verifiability and overcharging).

[8] See http://www.adac.de/infotestrat/tests/autohaus-werkstatt (accessed on 11 July 2017) for details.

have been registered for the first time during the same time period, run on a gasoline engine (of the most popular performance type), and be due for their main inspection; in addition, the owners must present a detailed record of previous inspections.

3. Motor vehicle experts outfit the test cars with the same five defects: The license plate light does not work, the air pressure in the spare tire is too low, the exhaust pipe is loose, the coolant level is too low, and the front-right headlight is displaced to the very bottom. If for any reason one of these defects cannot be introduced, the screen wiper blade on the passenger side is cut down to 2 cm. These potential faults are all listed in the inspection guidelines for all car makers, meaning that they should be easily detected.
4. The automobile club sends these cars to garages located in the vicinity of the car owner's residence. There is a maximum of one vehicle test per garage.
5. Upon completion of the inspections, the automobile club assesses each garage's performance according to a detailed evaluation scheme that also includes issues related to service, etc. The results are published in the club's monthly magazine (*ADAC Motorwelt*) and can be readily accessed online. The automobile club reports for each garage how many of the five introduced defects were fixed and whether the garage charged for any services illegally.[9]

Our binary dependent variable, fraud, indicates whether a garage illegally charged for the provision of services. Illegal charging includes charging for more repairs than were provided and for providing more services than were necessary. Note that the data do not allow us to distinguish different types of illegal charging. Furthermore, the data only cover part of the garages' potential defrauding behaviour, as we cannot account for cases in which customers were charged for more expensive repairs than those that were actually made. Thus, our data provide a lower boundary for garages' defrauding behaviour. For our controls, we consider the number of faults detected by a garage in the automobile club's data set as an indicator of the garage's competence: The fewer faults a garage finds, the less competent it is.[10] Although one might expect competence to be highly correlated with the other explanatory variables, the maximum absolute correlation we observe is between competence and a critical financial situation (-0.2011).

This very basic data set does not allow us to study the impact of competition on the garages' incentive to defraud, nor can we control for a firm's financial situation or its reputational concerns. We therefore complement the automobile club's dataset with our main variable of interest and the two controls: (i) We introduce a measure of the competitiveness of the environment in which each of the garages does business, (ii) we observe an indicator

[9] The original criteria as stated by the automobile club is as follows. 'Punktabzug gab es für unzulässige Berechnung von 'Systemreinigern' und anderen, nicht näher spezifizierten Mitteln, für die seitens der Hersteller keine Vorschriften bestehen, keine zugehörige Diagnose gestellt sowie auf der Rechnung vermerkt wurde und für die vom Kunden kein Auftrag vorlag'. ('Points were deducted for illegal charges for 'system cleaners' and other not precisely specified goods for which there are no manufacturer guidelines, for which no diagnosis was performed, and that appeared on the bill without the customer's approval'.) Note that in previous versions of the paper, we considered such illegal charging as 'overcharging'. Now that we have access to the exact definition of illegal charging used by ADAC, we refer more generally to fraud, as we cannot exclude the case of overtreatment.

[10] Note that the data do not allow us to distinguish between garages that did not find a fault and those that found a fault but forgot to charge for it. However, we are confident that the incentives to charge for all repairs performed are sufficiently high that forgetting to charge for an item will scarcely ever occur.

TABLE 1

Overview of variables

Variable	*Proxy*	*Source*
Fraud	Services charged but not performed or unnecessary services performed	ADAC experiment, 2006 & 2008–10
Competition intensity	# of competitors within 10 km is above median	*Gelbe Seiten* from 2011
Financial situation	Negative equity	*Elektronischer Bundesanzeiger*, 2006 & 2008–10
Competence	# of faults found out of 5	ADAC experiment, 2006 & 2008–10
Low reputation	Distance to next interstate less than 1,500 m	Google Maps Distance Calculator, 2010

of each garage's financial status, and (iii) we suggest a proxy for reputational concerns. Table 1 provides an overview of the variables, the proxies, and the respective data sources.

Ad (i): In order to evaluate the strength of the competition that a garage faces, we analyse the number of competitors in a garage's neighbourhood. We opt for the number of competitors as an indicator of competition over other measures such as the Hirschman–Herfindahl Index (Hirschman, 1964) and the price-cost margin (Boone, 2008) because of data availability. Note that the number of competitors has been used as a proxy for competition by previous studies in credence goods markets (see, e.g., Pike, 2010).

We determine the number of garages located within a distance of 10 km from the garage in question. We consider 10 km to be the average distance a potential customer is willing to travel to a competitor.[11] We obtain the data on the number of competitors of each garage from the publicly available directory of businesses sorted by sector, the German version of Yellow Pages (*Gelbe Seiten*). *Gelbe Seiten* provides one of the largest lists of phone numbers and addresses for companies in Germany.[12] The great advantage of this database compared to, e.g., *Google Places*, is that the editing process ensures that the businesses listed actually exist and that they can be properly categorized as car repair shops. We perform a search for *Autowerkstätten* ('car repair shops') within a radius of 10 km from each garage's address and count the number of results. Finally, we divide the group of garages into those facing a number of competitors above the median and those below. By dichotomizing competition intensity, we account for the fact that garages' defrauding behaviour most likely depends upon whether there are few or many competitors but not on whether there are one or two additional competitors within close proximity. It is worth emphasizing that our results do not rely on the dichotomization of the variable, as shown by the robustness checks described below.

Ad (ii): We extend the automobile club's data set by adding information on the garages' financial situations at the beginning of the test year. These financial data are publicly available through the Electronic Federal Gazette for corporate enterprises in Germany (*Elektronischer Bundesanzeiger*).[13] According to German corporate law, enterprises are required to publish basic financial information for the benefit of potential shareholders.

[11] Our results do not change if we use 5 or 20 km as the maximum distance a customer is willing to travel (see section V for robustness checks).

[12] See http://www.gelbeseiten.de (accessed on 11 July 2017) for details.

[13] See http://www.bundesanzeiger.de (accessed on 11 July 2017) for details.

We divide the garages into those with positive equity and those with negative equity.[14] A firm has negative equity if its debts exceed its assets. Such firms are in a critical financial situation because banks are no longer willing to lend them additional money, although they are not yet bankrupt (bankruptcy only applies if one of the debts is due and cannot be repaid to the lender). Because the amount of a firm's equity is correlated with firm size, we dichotomize the equity variable. Hence, we only capture the firm's financial status without confounding the status with firm size. We opt to use equity as a proxy for a firm's financial situation over other indicators (such as profit) because it is not subject to annual upturns and/or downturns.

Ad (iii): We extend the database by adding the garages' distance to the nearest interstate. We consider this distance to be a good proxy for a garage's reputational concerns. Cars that break down on the interstate are usually towed to the nearest garage,[15] and thus garages that are located close to an interstate experience more one-time interactions, implying a lower chance of repeat business. As a consequence, these firms are less concerned about building up a reputation in comparison to garages that are located further away from an interstate. We consider garages that are located less than 1,500 m from an interstate to be 'close' to an interstate and all others to be 'not close'.[16] We dichotomize the distance to the nearest interstate because cars are hardly ever towed to a garage that is far away from the interstate. This holds irrespective of whether the garage is 10 or 30 km away from the nearest interstate. We complement the data set by introducing each garage's exact distance to the nearest interstate, which we calculate using *Google Maps Distance Calculator*. Using *Google*'s geographic database via APIs, the *Google Maps Distance Calculator* enables the user to select two arbitrary points on the map in order to calculate the air-line distance.[17] We take the garage's address as the reference point and the closest point on the nearest interstate as the second point.[18]

Identification

Given the variables and their measurements above described, the main identification challenge is reverse causality. Below, we briefly comment on measurement errors and possible omitted variables. The relationship between reputational concerns and fraud as well as that between the level of competition and fraud might be reverse causal, as the choice of a garage's location and thus the distance to the nearest interstate and the level of competition might not be exogenous in explaining fraud. However, there are three reasons that a garage's location may indeed be exogenous. Firstly, the average age of the garages that defrauded

[14] We considered garages to have negative equity if their equity was negative at the beginning of the test year or if their equity was positive at the beginning but negative at the end of the test year.

[15] The vast majority of towings in Germany are conducted by ADAC. ADAC always tows cars to the nearest garage as a free service for members. Towing to any other garage is subject to a service fee (see http://www.adac.de/mitgliedschaft/leistungen/default.aspx, accessed on 11 July 2017).

[16] Our results are robust if we consider garages less than 1,000 m or less than 2,000 m from the nearest interstate to be 'close' to the interstate (see section V).

[17] Note that our results are robust to using different distance measures, such as the distance of the actual route from the nearest interstate exit to a garage (see Table A4).

[18] See http://www.daftlogic.com/projects-google-maps-distance-calculator.htm (accessed on 11 July 2017) for details.

in the test was 20 years (the minimum age was ten years). If endogeneity concerns were to hold, a garage's defrauding behaviour today would have to be correlated with the choice of location 20 years ago. Hence, reverse causality does not seem very plausible. Secondly, one cannot simply open a garage anywhere; rather, it must operate within an area specifically zoned for that purpose. Thus, garages are restricted in their choice of location. Thirdly, business insiders confirm that maximizing customer visits is the main criterion in selecting the location for a new garage.[19] These three reasons support our contention that garage locations are not chosen based on the type of interaction (i.e., repeated or one-time) or the number of competitors.

Reverse causality between the incentive to defraud and a garage's financial situation might also exist. Since defrauding influences a firm's financial situation, we might encounter endogeneity with respect to a firm's equity at the end of the year. Note, however, that defrauding increases equity compared to an honest repair. Consequently, if there were reverse causality between fraud and a firm's equity, we would underestimate the effect of the financial situation on the probability of defrauding. Thus, reverse causality with respect to the financial situation would weaken our results.[20]

Minor identification challenges might arise due to potential measurement errors and omitted variables. One possible concern regarding the measurement of fraud is that garages might defraud by mistake and not intentionally. Because we cannot distinguish between intended and unintended fraud, we must assume that garages are fully aware of the services for which they bill. As for omitted variables, we perform extensive robustness checks with respect to the properties of garages and their annual variations. Due to the limited number of observations, we are unable to account for regional differences that might arise from divergent customer populations. It would be interesting to determine, for example, whether garages are more likely to operate in areas in which customers' knowledge of car repair is limited.

Descriptives

After restricting the data set, it contains 134 corporate garages, of which 128 did not defraud; that is, we find that six (4.5%) of the garages defrauded their customers (see Table 2). This number is in accordance with Schneider (2012), who observes that defrauding occurred in three out of 51 visits (or 6%).[21] Although 4.5% cases of defrauding might not seem to be a lot, the issue of fraud is an important problem, as discussed in the Introduction. The yearly turnover in the market for car repairs is about 30 billion € in Germany alone (Zentralverband Deutsches Kraftfahrzeuggewerbe (Ed.), 2012). Extrapolations from our data indicate that the amount of fraud in the car repair market is far from negligible.

[19] See, e.g., Johnson, D.L.: '6 tips to start your auto repair shop business today' (see http://ezinearticles.com/?6-Tips-To-Start-Your-Auto-Repair-Shop- Business-Today&id=1176780) (accessed on 11 July 2017) or eHow: 'How to open an auto repair shop' (see http://http://www.ehow.com/how_2387498_open-auto-repair-shop.html, accessed on 11 July 2017).

[20] One might argue that garages that frequently defraud may face decreased equity in the long run. Remember, however, that customers do not observe fraud. Hence, it is difficult for them to punish garages that defraud, even in the long run.

[21] The average amount overcharged was US$32 per incident in the study conducted by Schneider (2012). The sum of fraud incidents across all visits accounted for 2% of total charges.

TABLE 2

Descriptives

Variable	*Mean*	*SD*	*Min*	*Max*	*Observations*
Fraud (=1 if true)	0.045	0.208	0	1	134
Intense competition (=1 if # of competitors is above median)	0.500	0.502	0	1	134
Critical financial situation (=1 if equity is negative)	0.134	0.342	0	1	134
Competence (# of faults found out of 5)	4.239	1.125	0	5	134
Low reputation (=1 if distance < 1,500 m)	0.284	0.452	0	1	134

TABLE 3

Mean comparisons between garages that did and did not defraud their customers

	*Intense competition**	*Critical financial situation****	*Competence****	*Low reputation*
Fraud = 1	0.167	0.500	3.000	0.500
Fraud = 0	0.516	0.117	4.297	0.273

Notes: χ^2 and Mann–Whitney U Test, respectively (the latter test for the comparison of means of competence), two-tailed: $^*P < 0.1$, $^{**}P < 0.05$, $^{***}P < 0.01$.

Table 2 also provides the descriptives for the four explanatory variables. By the construction of the variable, about half of the garages face intense competition; 13.4% are in a critical financial situation. The high average level of competence (4.24 faults found out of five) is due to the fact that the faults are all listed on the mechanics' checklists for inspections issued by all car makers. Overall, 27.3% of the garages are located close to the interstate and therefore have low reputational concerns.

Table 3 illustrates that the two groups – garages that did and did not defraud – differ considerably in their characteristics: Garages that defrauded face intense competition less often than garages that did not defraud. This difference in competition intensity is weakly significant (χ^2 Test, two-tailed: $P = 0.095$). For the controls, we find that 50% of the garages that defrauded are in a critical financial situation, whereas significantly fewer of garages that did not defraud are in financial difficulty (11.7%, χ^2 Test, two-tailed: $P = 0.007$). The average competence level of garages that defrauded is significantly lower than the average level of garages that did not defraud (Mann–Whitney U Test, two-tailed: $P = 0.003$). Table 3 suggests that garages that defrauded have low reputational concerns more often than garages that did not defraud; however, this difference is not statistically significant (χ^2 Test, two-tailed: $P = 0.229$).

IV. Result

The small sample of our empirical analysis, the skewed distribution of our dependent variable, and the quasi-separation of the data with respect to competition intensity represent challenges for the derivation of meaningful conclusions. To address these issues, we make use of a well-established method – namely, the Firth logit regression (Firth, 1993) – that is typically used in other research fields in which small samples, skewed distributions, and quasi-separations are frequently observed phenomena. Most importantly, note that our results do *not* depend on the choice of the regression model used, as we will show in the robustness checks (see section V).

First, we offer a brief comment on the advantages of the Firth regression. The standard maximum likelihood estimation used in binary regression models assumes the sample to be large. As the sample size converges to infinity, the parameter estimates converge to the true parameter values; hence, estimates may be biased in smaller samples. The Firth regression uses a penalized likelihood estimation that eliminates the first-order bias that occurs due to small samples (Heinze, 2006). The Firth approach also regularizes the data and thereby circumvents the separation problem (Zorn, 2005). Consequently, the Firth regression always results in finite parameter estimates, which is not the case for regressions based on the standard maximum likelihood estimation. This approach is frequently used in medical research[22] and has been proven to outperform alternative small-sample models such as the exact logistic regression (Heinze, 2006). Heinze (2006) emphasizes that for small samples, 'penalized likelihood confidence intervals for parameters show excellent behaviour in terms of coverage probability and provide higher power than exact confidence intervals'.

Needless to say, the fact that only six out of our sample of 134 garages defrauded their customers makes the identification of effects more difficult than if the dependent variable had exhibited a higher variance. Note, however, that at the given level of the type-I error, the probability that we falsely reject the null hypothesis of 'no effect' amounts to 5%. Thus, if we can identify effects of the explanatory variables on fraud, the differences between garages that did and did not defraud must be considerably large. In that case, we can expect a systematic difference between the two groups of garages and not merely a difference that arises by chance.

Given the four explanatory variables – competition intensity, financial situation, competence, and reputation – our Firth logit model is specified as follows:

$$\begin{aligned} firth\_logit(fraud) = {} & \beta_0 + \beta_1 intense\_competition + \beta_2 critical\_financial\_situation \\ & + \beta_3 competence + \beta_4 low\_reputation + \epsilon \end{aligned} \tag{1}$$

We report the results of the Firth regression in Table 4. We also present the results of the linear probability model in order to ease interpretation. To evaluate the model fit, we calculate McFadden's R^2 for the binary response models and the ordinary R^2 for the linear model. We opt to use McFadden's R^2 as a measure for the binary model fit because it can also be applied to the Firth logit regression. McFadden's R^2 is defined as $1 - L1/L0$, where $L1$ is the log-likelihood of the fully specified model and $L0$ is the log-likelihood of the null model. Interpreting $L0$ as the total sum of squares in the linear regression analysis and $L1$ as the residual sum of squares, McFadden's R^2 provides a measurement for the model fit similar to that for the ordinary R^2 (Wooldridge, 2009). McFadden (1979) suggests that models with an R^2 between 0.2 and 0.4 indicate an excellent fit. The McFadden R^2 of our Firth regression is 0.412 and is therefore close to an excellent fit.

We now turn to our main result on the impact of competition on expert fraud.

Result. Garages facing intense competition defrauded less often than those in a weakly competitive environment.

[22] As an example, George *et al.* (2010) apply the Firth logit regression to the question of how a certain medication (phenylephrine) impacts spinal anaesthesia-induced hypertension. Their work is based on a sample size of 45 participants. Only nine participants did not show a positive reaction to the medication.

TABLE 4

What drives fraud?

Fraud	*Firth logit*	*OLS*
Intense competition (= 1 if # of competitors > median)	−2.049**	−0.078**
	(1.040)	(0.035)
Critical financial situation (= 1 if true)	1.757**	0.114**
	(0.891)	(0.051)
Competence (# of faults found out of 5)	−0.765**	−0.041***
	(0.315)	(0.015)
Low reputational concerns (= 1 if distance < 1,500 m)	2.078**	0.077**
	(0.999)	(0.039)
Constant	−0.510	0.220***
	(1.125)	(0.071)
McFadden R^2	0.412	–
R^2	–	0.142
Observations	134	134

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$. P-values are based on two-tailed tests.

In line with theory, we find that a high level of competition decreases the level of fraud. According to the OLS estimates, a (highly) competitive environment decreases the probability of being defrauded by an expert by 7.8 percentage points. In fact, five out of the six garages that defrauded face a competition level that is lower than the median; only every second garage that did not defraud faces a similar competition level (see Table 3).[23] Our results thus add to the experimental findings of Dulleck *et al.* (2011); whereas Dulleck *et al.* (2011) do not find a significant impact of competition on fraud under liability, our results indicate that more competition leads to less fraud. These divergent results may be explained by the different market environments. In Dulleck *et al.* (2011), customers cannot search for a second opinion. Competition thus leads to lower prices but not to a change in experts' defrauding behaviour. In contrast, in our study, a higher level of competition implies lower search costs in the market for car repairs and thus reduces garages' incentives to defraud their customers.

Population

Note that in our estimation, we have not included information on population, i.e., we have not normalized our measure of competition by the population in the respective area. Clearly, one could argue that the number of garages is correlated with the population and thus car density. Recall that our theoretical argument for why competition should reduce expert fraud is based on the idea that an increase in competition (i.e., a greater number of garages

[23] Note that the effect of competition crucially depends on whether experts' and customers' interests with respect to fraudulent behaviour are aligned or not. In their empirical study, Bennett *et al.* (2013) find that competition among experts for vehicle emissions tests increases fraud. This is due to the fact that in their case, car owners whose cars are passed at higher rates due to fiercer competition may benefit from fraud, as they save money on costly repairs. This gives experts a greater incentive to generate a competitive advantage through illicit actions, raising the question of whether competition is necessarily the ideal market structure in such an environment.

in a customer's vicinity) reduces search costs. When second opinions become cheaper, this disciplines experts. As a consequence, our measure of competition should not depend on population density but only on the absolute number of garages in the area. Of course, this argument only holds as long as population density is sufficiently low, such that garages do not hit their capacity constraints. If that were the case, incentives not to defraud customers would not be affected by a change in the number of garages, since searching for a second opinion would be a non-credible threat.

Upon examining the degree of capacity utilization in the German car repair market, we conclude that capacity constraints are not very likely to be of relevance. In a press statement on business in the year 2012, the *German Federation for Motor Trades and Repairs* [*Zentralverband Deutsches Kraftfahrzeuggewerbe* (accessed on 11 July 2017), ZDK], which represents the professional interests of 38,500 accredited automotive companies on the national level, reported that the average degree of capacity utilization of 83% in the service sector 'fell only short of the degree of the strong year 2011'.[24] This implies that in the years before 2011, firms' capacities were utilized to a lesser extent. Other sources have confirmed this observation.[25] We take this as evidence that garages in Germany are not even close to working at full capacity. As a consequence, accounting for population density is not warranted in our case.[26]

Controls

With regard to the controls, we observe that a critical financial situation increases a garage's incentive to defraud: The OLS model estimates that a critical financial situation increases a customer's probability of being defrauded by 11.4 percentage points. Moreover, highly competent mechanics have a lower incentive to defraud their customers: The OLS regression results indicate that the probability of being defrauded decreases by 4.1 percentage points for each additional fault a garage detects.[27] The regression results also show that low reputational concerns increase a garage's incentive to defraud. The intuition is as follows: Garages that have a low reputational concern experience many one-time interactions. Hence, they can defraud their customers without risking any loss of future earnings. The OLS results suggest that the probability of a garage defrauding its customers increases by 7.7 percentage points if the garage has low reputational concerns. Note that this result is only partially in line with previous field evidence: Schneider (2012) shows that low reputational concerns lead to higher diagnosis fees, but he does not find a significant difference in the level of overtreatment. A possible explanation for this divergence in findings is that

[24] https://www.kfzgewerbe.de/presse/pressemeldungen/archiv/kfz-gewerbe-durchwachsene-bilanz-verhaltene-aussichten.html (accessed on 11 July 2017).

[25] See, e.g., http://www.stuttgarter-nachrichten.de/inhalt.vertragshaendler-mit-auto handel-laesst-sich-kaum-snoch-geld-verdienen.f17922de-15c9-4b08-99bf-402f5953d1a9.html or http://www.stuttgarter-nachrichten.de/inhalt.konkurrenzkampf-am-automobilmarkt-endspiel-um-das-deutsche-autohaus. 75f4b3b7-ec76-493a-8aff-dc6aefbf112c.html (accessed on 11 July 2017).

[26] Under a competition measure normalized by the population in each garage's postcode, the competition effect diminishes, whereas the controls' effect is still present.

[27] Because a low level of competence may lead to unintentional fraud, including competence as an explanatory variable may confound the results. Our robustness check shows that our findings are robust also to a regression specification that excludes competence as an explanatory variable (see Table A6).

TABLE 5

Robustness against different models

Fraud	*OLS*	*Logit*	*Probit*	*Scobit*
Intense competition (= 1 if # of competitors > median)	−0.078**	−2.593**	−1.253**	−2.539**
	(0.035)	(1.262)	(0.605)	(1.162)
Critical financial situation (= 1 if true)	0.114**	1.966*	0.884*	2.014**
	(0.051)	(1.010)	(0.535)	(0.870)
Competence (# of faults found out of 5)	−0.041***	−0.887**	−0.454**	−0.835***
	(0.015)	(0.367)	(0.191)	(0.316)
Low reputational concerns (= 1 if distance < 1,500 m)	0.077**	2.423**	1.190**	2.264**
	(0.039)	(1.157)	(0.559)	(1.047)
Constant	0.220***	−0.540	−0.282	−15.006
	(0.071)	(1.263)	(0.717)	(1,878.318)
McFadden R^2	–	0.352	0.345	0.365
R^2	0.142	–	–	–
Observations	134	134	134	134

Notes: Standard errors in parentheses, $*P < 0.1$, $**P < 0.05$, $***P < 0.01$. P-values are based on two-tailed tests.

Schneider considered unnecessary repairs to be overtreatment only when they cost more than 50US$. Our data set exclusively provides information on whether there was evidence of overcharging and/or overtreatment.

V. Robustness checks

Our results turn out to be robust against alternative models such as the logit model with a regular maximum likelihood estimator, the probit model, and the scobit regression (see Table 5).[28] This last model accounts for the skewed distribution of the fraud variable but is not significantly different from the logit regression. The significance levels of our explanatory variables remain practically unchanged when we apply these alternative models. The only decrease in a significance level (from 5% to 10%) occurs for the critical financial situation variable in the logit and probit model.

Because our results are based on six garages that defrauded and 128 garages that did not defraud, a possible concern may be that individual garages are driving the results. Consequently, we perform six Firth logit regressions, each excluding one of the six defrauding garages from the data set. Table A1 in Appendix A shows that the overall results do not change, but the significance levels decrease somewhat (as expected).

The results are also robust against choosing different parameters as cut-off points. In the above analysis, we measured the number of competitors within 10 km and then divided the garages into two categories: those facing fewer or more competitors than the median level. As Table A2 in Appendix A indicates, measuring the number of competitors within 5 or 20 km instead of 10 km does not change our results. Our results are also robust against including competition intensity as a continuous variable instead of using the dichotomized variable (see also Table A2 in Appendix A). For the variable of low reputational concerns, Table A2 in Appendix A shows that when we consider garages within 1,000 or 2,000 m of

[28] In order to improve the readability of this section, some robustness checks have been relegated to Appendix A.

the nearest interstate to be 'close' (instead of 1,500 m), we do not obtain results that differ from the above analysis.

Table A3 in Appendix A presents the results of our robustness checks with respect to alternative specifications. We control for annual effects in order to ensure that the financial crisis has not affected garages' behaviour. The results remain unchanged. Furthermore, we show that whether a garage is an authorized or an independent garage does not change any of our results.

In addition, we check the robustness of our results against inclusion of the 25 garages belonging to the same chain in our analysis. Note that the financial situation of the garages in this chain does not vary. Thus, we must exclude the garages' financial situation as a predictor for fraud. This might lead to biased results, as we have seen that the financial situation plays a crucial role in explaining the experts' defrauding behaviour. The other characteristics of the chain's garages are on average similar to the 134 garages in our main sample. One out of the 25 garages defrauded, which reflects almost exactly the mean fraud level of the 134 garages that do not belong to the chain. Upon analyzing the extended data set, our results again turn out to be fairly robust (see Table A5 in Appendix A). The intensity of competition and the level of competence continue to be significant predictors of fraud. The coefficient of low reputational concerns is still positive, as expected, but is no longer significantly different than zero anymore.

VI. Conclusion

Making use of a field study, we analyse the impact of competition on the incentive to defraud of car repair shops. We control for reputational concerns and the financial situation of the garages under investigation. In accordance with theory, we find that a high level of competition reduces experts' incentive to defraud. However, when we control for the number of inhabitants or the number of registered cars, the competition effect diminishes. With respect to our controls, we observe that firms that care little about their reputation and those struggling with a critical financial situation have a greater incentive to defraud their customers, whereas firms with a high level of competence are less likely to defraud. The limitations of our study are mainly due to the low number of defrauding incidents that are used to identify the impact of the explanatory factors. We account for this problem by using small-data methods. Nevertheless, more research in this area is needed.

Appendix A: Tables for Robustness Checks

Tables A1–A6.

Final Manuscript Received: May 2017

TABLE A1

Robustness with respect to individual garages

Fraud	*Firth logit w/o garage 1*	*Firth logit w/o garage 2*	*Firth logit w/o garage 3*	*Firth logit w/o garage 4*	*Firth logit w/o garage 5*	*Firth logit w/o garage 6*
Intense Competition (= 1 if # of competitors > median)	−2.099*	−1.853*	−1.796*	−1.824*	−1.776*	−3.250**
	(1.137)	(1.038)	(1.059)	(1.027)	(1.075)	(1.631)
Critical financial situation (=1 if true)	2.214**	1.561*	1.429	1.449	1.972**	2.254**
	(1.014)	(0.929)	(0.975)	(0.943)	(0.926)	(0.992)
Competence	−0.972**	−0.690**	−0.755**	−0.724**	−0.737**	−0.743**
	(0.411)	(0.331)	(0.311)	(0.327)	(0.321)	(0.343)
Low Reputational Concerns (=1 if distance < 1,500 m)	2.975**	2.073**	1.774*	2.194**	1.673	1.754
	(1.371)	(0.993)	(1.047)	(1.033)	(1.061)	(1.079)
Constant	−0.715	−0.791	−0.416	−0.749	−0.729	−0.617
	(1.206)	(1.217)	(1.128)	(1.174)	(1.174)	(1.225)
Observations	133	133	133	133	133	133

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$. P-values are based on two-tailed tests.

TABLE A2

Robustness against different cut-off points

Fraud	*Firth logit competition 5k*	*Firth logit competition 20k*	*Firth logit competition continuous*	*Firth logit reputation 1,000 m*	*Firth logit reputation 2,000 m*
Intense competition (= 1 if # of competitors within 5k > median)	−1.933*				
	(1.035)				
Intense competition (= 1 if # of competitors within 10k > median)				−1.759*	−2.327**
				(1.006)	(1.075)
Intense competition (= 1 if # of competitors within 20k > median)		−1.844*			
		(1.019)			
Intense competition (continuous)			−0.014*		
			(0.008)		
Critical financial situation (= 1 if true)	1.546*	1.580*	1.876**	1.811**	1.864**
	(0.861)	(0.862)	(0.901)	(0.907)	(0.887)
Competence (# of faults found out of 5)	−0.800**	−0.667**	−0.707**	−0.782**	−0.754**
	(0.318)	(0.301)	(0.301)	(0.317)	(0.312)
Low reputational concerns (= 1 if distance < 1,000 m)				2.278**	
				(1.031)	
Low reputational concerns (= 1 if distance < 1,500 m)	1.985**	2.126**	2.274**		
	(0.991)	(0.981)	(1.026)		
Low reputational concerns (= 1 if distance < 2,000 m)					1.885*
					(1.019)
Constant	−0.339	−0.891	−0.365	−0.563	−0.529
	(1.136)	(1.086)	(1.163)	(1.121)	(1.120)
McFadden R^2	0.400	0.389	0.620	0.426	0.392
Observations	134	134	134	134	134

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$. P-values are based on two-tailed tests.

TABLE A3

Robustness against different specifications.

Fraud	*Firth logit controlling for authorized garages*	*Firth logit controlling for years*
Intense competition (= 1 if # of competitors > median)	−2.043**	−1.956*
	(1.036)	(1.160)
Critical financial situation (= 1 if true)	1.720*	1.596*
	(0.887)	(0.933)
Competence (# of faults found out of 5)	−0.747**	−0.713**
	(0.312)	(0.317)
Low reputational concerns (= 1 if distance < 1,500 m)	2.017**	2.286**
	(0.984)	(1.056)
Authorized garage	1.037	
	(1.728)	
Year 2006		−0.260
		(1.555)
Year 2008		0.179
		(1.295)
Year 2009		−1.190
		(1.397)
Constant	−0.507	−0.257
	(1.119)	(1.226)
McFadden R^2	0.375	0.426
Observations	134	134

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$. P-values are based on two-tailed tests.

TABLE A4

Robustness against a different distance measure

Fraud	*Firth logit*
Intense competition (= 1 if # of competitors > median)	−2.113**
	(1.063)
Critical financial situation (=1 if true)	2.558**
	(1.088)
Competence (# of faults found out of 5)	−0.676**
	(0.297)
Low reputational concerns (= 1 if driving distance to next interstate exit < 1,500 m)	3.457***
	(1.243)
Constant	−0.971
	1.250
Observations	134

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$. P-values are based on two-tailed tests.

TABLE A5

Robustness against including the chain into the analysis

Fraud	*Firth logit*
Intense competition (= 1 if # of competitors > median)	−1.760*
	(0.949)
Competence (# of faults found out of 5	−0.672**
	(0.273)
Low reputational concerns (= 1 if distance < 1,500 m)	1.153
	(0.848)
Chain (=1 if true)	−0.678
	(1.093)
Constant	−0.012
	(1.014)
McFadden R^2	0.251
Observations	159

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$. P-values are based on two-tailed tests.

TABLE A6

Robustness against excluding competence from the analysis

Fraud	*Firth logit*	*OLS*
Intense competition (= 1 if # of competitors > median)	−1.748*	−0.074**
	(1.007)	(0.035)
Low Reputational Concerns (=1 if distance < 1,500 m)	1.442*	0.067*
	(0.865)	(0.039)
Critical financial situation (=1 if true)	2.021**	0.141***
	(0.837)	(0.051)
Constant	−3.280***	0.044*
	(0.671)	(0.026)
Observations	134	134

Notes: Standard errors in parentheses, *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

References

Balafoutas, L., Beck, A., Kerschbamer, R. and Sutter, M. (2013). 'What drives taxi drivers? A field experiment on fraud in a market for credence goods', *Review of Economic Studies,* Vol. 80, pp. 876–891.

Balafoutas, L., Kerschbamer, R. and Sutter, M. (2017). 'Second-degree moral hazard in a real-world credence goods market', *Economic Journal*, Vol. 127, pp. 1–18.

Bennett, V., Pierce, L., Snyder, J. and Toffel, M. (2013). 'Customer-driven misconduct: how competition corrupts business practices', *Management Science*, Vol. 59, pp. 1725–1742.

Boone, J. (2008). 'A new way to measure competition', *The Economic Journal*, Vol. 118, pp. 1245–1261.

Consumer Federation of America, National Association of Consumer Agency Administrators, North American Consumer Protection Investigators. (2011). 2010 Consumer Complaint Survey Report.

Currie, J., Lin, W. and Zhang, W. (2011). 'Patient knowledge and antibiotic abuse: evidence from an audit study in China', *Journal of Health Economics*, Vol. 30, pp. 933–949.

Darby, M. R. and Karni, E. (1973). 'Free competition and the optimal amount of fraud', *Journal of Law and Economics*, Vol. 16, pp. 67–88.

Domenighetti, G., Casabianca, A., Gutzwiller, F. and Martinoli, S. (1993). 'Revisiting the most informed consumer of surgical services', *International Journal of Technology Assessment in Health Care*, Vol. 9, pp. 505–513.

Dulleck, U. and Kerschbamer, R. (2006). 'On doctor, mechanics, and computer specialists: the economics of credence goods', *Journal of Economic Literature*, Vol. 44, pp. 5–42.

Dulleck, U., Kerschbamer, R. and Sutter, M. (2011). 'The economics of credence goods: an experiment on the role of liability, verifiability, reputation, and competition', *American Economic Review*, Vol. 101, pp. 526–555.

Evans, R. G. (2000). 'Supplier-induced demand: some empirical evidence and implications', in Perlman, M. (ed.), *The Economics of Health and Medical Care*, London: Macmillan, pp. 162–173.

Firth, D. (1993). 'Bias reduction of maximum likelihood estimates', *Biometrika*, Vol. 80, pp. 27–38.

George, R. B., McKeen, D., Columb, M. O. and Habib, A. S. (2010). 'Up-down determination of the 90% effective dose of phenylephrine for the treatment of spinal anesthesia-induced hypotension in parturients undergoing cesarean delivery', *Anesthesia & Analgesia*, Vol. 110, pp. 154–158.

Heinze, G. (2006). 'A comparative investigation of methods for logistic regression with separated or nearly separated data', *Statistics in Medicine*, Vol. 25, pp. 4216–4226.

Hirschman, A. O. (1964). 'The paternity of an index', *American Economic Review*, Vol. 54, pp. 761–762.

Kerschbamer, R. and Sutter, M. (2017). 'The economics of credence goods – a survey of recent lab and field experiments'. *CESifo Economic Studies*, Vol. 63, pp. 1–23.

Kerschbamer, R., Neururer, D. and Sutter, M. (2016). 'Insurance coverage of customers induces dishonesty of sellers in markets for credence goods', *Proceedings of the National Academy of Science*, Vol. 113, pp. 7454–7458.

McFadden, D. (1979). 'Quantitative methods for analysing travel behaviour of individuals: some recent developments', in Hensher, D. A. and Stopher, P. R. (eds), *Behavioural Travel Modeling*, London: Croom Helm, pp. 279–318.

Mimra, W., Rasch, A. and Waibel, C. (2016a). 'Second opinions in markets for expert services: experimental evidence', *Journal of Economic Behavior & Organization*, Vol. 131, Part B, pp. 106–125.

Mimra, W., Rasch, A. and Waibel, C. (2016b). 'Price competition and reputation in expert markets: experimental evidence', *Games and Economic Behavior*, Vol. 100, pp. 337–352.

Pike, C. (2010). *An Empirical Analysis of the Effects of GP Competition*, Cooperation and Competition Panel Working Paper Series. Vol. 1, No. 2.

Schmid, C. (2015). 'Consumer health information and the demand for physician visits', *Health Economics*, Vol. 24, pp. 1619–1631.

Schneider, H. S. (2012). 'Agency problems and reputation in expert services: evidence from auto repair', *Journal of Industrial Economics*, Vol. 60, pp. 406–433.

Taylor, C. R. (1995). 'The economics of breakdowns, checkups, and cures', *Journal of Political Economy*, Vol. 103, pp. 53–74.

Titus, R. M., Heinzelmann, F. and Boyle, J. M. (1995). 'Victimization of persons by fraud', *Crime & Delinquency*, Vol. 41, pp. 54–72.
Wolinsky, A. (1993). 'Competition in a market for informed experts' services', *RAND Journal of Economics*, Vol. 24, pp. 380–398.
Wolinsky, A. (1995). 'Competition in markets for credence goods', *Journal of Institutional and Theoretical Economics*, Vol. 151, pp. 117–131.
Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach*, Cengage Learning Services, Mason.
Zentralverband Deutsches Kraftfahrzeuggewerbe (Ed) (2012). *Zahlen & Fakten 2011: Ausgabe 2012*, Self-published, Bonn.
Zorn, C. (2005). 'A solution to separation in binary response models', *Political Analysis*, Vol. 13, pp. 157–170.