

What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods

LOUKAS BALAFOUTAS

University of Innsbruck

ADRIAN BECK

University of Innsbruck

RUDOLF KERSCHBAMER

University of Innsbruck

and

MATTHIAS SUTTER

University of Innsbruck, University of Gothenburg, IZA Bonn, CESifo Munich

First version received September 2011; final version accepted November 2012 (Eds.)

Credence goods are characterized by informational asymmetries between sellers and consumers that invite fraudulent behaviour by sellers. This article presents a natural field experiment on taxi rides in Athens, Greece, set up to measure different types of fraud and to examine the influence of passengers' presumed information and income on the extent of fraud. We find that passengers with inferior information about optimal routes are taken on significantly longer detours, while lack of information on the local tariff system increases the likelihood of manipulated bills by about fifteen percentage points. Passengers' perceived income seems to have no effect on fraud.

Key words: Credence goods, Expert services, Natural field experiment, Taxi rides, Fraud, Asymmetric information, Distributional preferences

JEL Codes: C93, D82

1. INTRODUCTION

Many goods and services, such as car or computer repairs, medical treatments or taxi rides in an unknown city, share the characteristic of an informational asymmetry between the seller and the consumer. Typically, the seller knows more than the consumer about the quality that yields the highest surplus from trade, and moreover the consumer is frequently not even *ex post* able to observe the quality of the good or service he received. These features have led Darby and Karni (1973) to create the term “credence good”, because consumers are unable to identify the appropriate quality on their own and have to rely on the judgement of an expert who is also the seller of the good.

The informational asymmetries present on credence goods markets open the door to different types of fraudulent behaviour on the sellers' side, including *overtreatment* (providing a higher quality than the surplus maximizing one), *undertreatment* (choosing a quality that is insufficient to satisfy the consumer's needs), and *overcharging* (charging for a higher quality than has been provided). Fraud in markets for credence goods has potentially large efficiency costs for an economy. For instance, the U.S. Department of Transportation has estimated that more than half of the total expenses for car repairs are for unnecessary repairs, which is a rough estimate of the efficiency costs of overtreatment (Wolinsky, 1993). Referring to the health care sector, a Swiss study has found that the average person's probability of receiving a particular surgical intervention is one-third above that of a physician or a member of a physician's family, indicating that a consumer's (presumed) information affects the quality of the received treatment and hence the likelihood of under- and overtreatment (Domenighetti *et al.*, 1993).¹

Yet, the complexity of services such as medical treatments or car repairs remains an obstacle to unambiguously measuring the extent of fraud. In fact, "mistreatment" in such markets could be driven by causes unrelated to fraud. For instance, an overly cautious car mechanic (physician) might be prone to conducting repairs (performing treatments) prematurely in the best intention to prolong the lifetime of the car (patient). Moreover, it is difficult to detect and quantify the amount of overcharging in such markets either because the customer is typically not present during the service (as in most cases of car repairs), or because he lacks the expertise to judge which "treatment" has been performed (in the health care market this seems to be rather the rule than the exception).

In this article, we present the results of a natural field experiment designed to identify the extent *and* type of fraud in the market for taxi rides. In this market, confounding factors present in other credence goods markets—such as issues of misdiagnosis or incompetence of the expert seller, overly cautious behaviour, or the unavailability of services that solve a consumer's problem—are practically non-existent. In addition, undertreatment by failing to reach the requested destination is hardly an issue, and overtreatment is easy to identify for an informed external observer since unnecessary detours are unlikely to add any value for the customer. Given those features of the market for taxi rides, the present article contributes to the literature on fraudulent behaviour on markets for credence goods in two substantial ways.

First, running a field experiment in a market for taxi rides allows us to identify the extent of fraud *and* to disentangle it into the two dimensions *overtreatment* and *overcharging*. We do so by letting undercover experimenters take taxi rides in the capital city of Greece, Athens, a city with approximately 4 million inhabitants and 14,000 taxis. For each single ride, a portable GPS satellite logger enables us to precisely record the chosen route and the taxi's exact position and speed at each point in time. With these data, we can quantify overtreatment in the form of detours. The GPS data also allow calculating the correct fare for a given distance. The difference between the total fare charged by the taxi driver and the correct fare then measures the amount of overcharging.

Second, and more important, we study how the extent and the type of fraud in the market for taxi rides depend on characteristics of a passenger that are important for a driver's provision and charging behaviour in theory. In our experimental treatments, we manipulate the taxi driver's perception about the passenger's (i) information about the city; (ii) information about the tariff system; and (iii) income.

1. In health economics overtreatment is referred to as "physician-induced demand" and its detection is a key topic in the field—see McGuire (2000) for a review. On a more general level, the problem of uncovering evidence of hidden behaviour emanating from informational asymmetries is part of what is emerging as "forensic economics"—see Zitzewitz (2012) for an excellent review. Our study contributes to this new field by using GPS-technology to identify and measure the extent of two dimensions of fraud in a credence goods market.

The purpose of the first treatment variation—the manipulation of the *driver's perception of the passenger's familiarity with the city*—is to test the theoretical prediction that there is not much room for overtreatment of consumers who know their needs (Dulleck and Kerschbamer, 2006). It is implemented by letting some passengers only state the destination and others state the destination and ask the driver whether he knows the destination, adding as an explanation for asking that they are not familiar with the city. We hypothesize that the latter passengers are more likely to be taken on detours than the former.

The second treatment variation—the manipulation of a *driver's perception of the passenger's familiarity with the details of the local taxi tariff system*—aims at testing the theoretical prediction that there is not much room for overcharging of consumers who can verify whether the correct tariff has been applied (Dulleck and Kerschbamer, 2006). It is implemented by varying the passenger's spoken language (Greek versus English). We hypothesize that taxi drivers try to exploit their informational advantage in the charging dimension more extensively with English-speaking than with Greek-speaking passengers, since an English-speaking passenger is arguably less likely to be perceived as familiar with the details of the (nationwide regulated) Greek taxi tariff system.²

Finally, we manipulate a *driver's perception of the passenger's income* by varying the passenger's clothes and the requested destination. There is impressive evidence from laboratory experiments that (i) distributional preferences are behaviourally relevant; and (ii) the overwhelming majority of non-selfish decision makers has convex distributional preferences (see, *e.g.*, Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000, Charness and Rabin, 2002). Given that theory predicts that decision makers with convex distributional preferences tend to treat low-income agents better than high-income ones, we hypothesize that passengers in the role of low-income consumers are less prone to fraud.

Our field experiment with taxi drivers provides a number of advantages compared to other approaches to examine expert sellers' behaviour on credence goods markets.³ Compared to *non-experimental field studies* like those discussed in the beginning, it allows for a systematic variation of treatment variables, while still investigating the behavior of real expert sellers in their natural working environment. Another advantage of our study is that the experts in our sample are always residual claimants, meaning that they reap the potential profits from fraud themselves. This is typically not the case for car mechanics or doctors, for instance, who are frequently employed for a fixed wage, which weakens their financial incentives for fraud, thus making it very difficult to observe and measure it unambiguously.

Compared to recent *laboratory experiments* on the impact of institutional and market conditions on the extent of fraud in credence and experience goods markets (Huck *et al.*, 2007, 2012, Dulleck *et al.*, 2011) our study has the advantage that external validity is of lesser concern, because observations are collected in a real credence goods market. For instance, one of the findings in Dulleck *et al.* (2011) is that sellers' distributional preferences have a large impact on their provision and charging decisions, but it is not clear how this result from the laboratory translates to real world credence goods markets. By manipulating taxi drivers' perception of a passenger's income in our natural field experiment we can address this question. It must still be acknowledged, however, that our findings may to some extent be affected by factors idiosyncratic

2. It is a long established pattern in linguistics that people tend to infer (correctly or incorrectly) private information from speech. In a classic field experiment, Kingsbury (1968) asked pedestrians on a Boston street for directions to a department store, either in the local Boston dialect, or employing a dialect spoken in rural Missouri. In the latter case the directions given were significantly longer and more detailed than in the former case. Apparently, from the dialect alone people inferred a lower level of local expertise.

3. See List (2006) and List and Reiley (2008) for surveys on the general advantages of field experiments.

to the particular market and environment under consideration (such as selection of participants or cultural factors). This implies that the results might vary in settings that are different to ours.

Our experiment has also advantages over the only other *field experiment* aimed at measuring fraud in a credence goods market—a paper by Schneider (2012) on the impact of reputation on fraudulent behaviour in the car repair sector. While Schneider (2012) reports that overtreatment and undertreatment are pervasive, he acknowledges that one cannot rule out incompetence as a factor contributing to under- and overtreatment. Moreover, unnecessary repairs might have been caused by overly cautious mechanics who replaced parts to avoid possible malfunctioning in the near future. As mentioned earlier, such potential confounds are hardly present in the market for taxi rides.

We now turn to a short preview of our main findings. In total, the experimenters took 348 rides, with a total driving distance of more than 4,400 km and an overall duration of 128 h of taxi driving. The average length of a ride was 12.7 km, of which on average 1.3 km (10%) was an unnecessary detour. Those passengers conveying the impression of being unfamiliar with the city were, on average, taken on detours of more than double length compared to passengers who did not. In 11% of rides passengers were overcharged through the application of incorrect tariffs. However, those passengers presumably perceived as unfamiliar with the tariff system (*i.e.*, those speaking in English) were overcharged in 22% of cases, while this happened in only 6% of cases to passengers in the role of Greek citizens. This shows that two different types of informational asymmetries between taxi drivers and customers give rise to two different types of fraud. Conveying the impression of having high income had no significant effect on either type of fraud.

We proceed as follows. Section 2 provides some background information on the taxi market in Athens. Section 3 introduces the experimental design and derives our hypotheses. Section 4 presents the results. Section 5 discusses our findings and Section 6 concludes.

2. THE MARKET FOR TAXI RIDES IN ATHENS

The market for taxi rides is regulated nationwide in Greece. The regulation concerns both market entry and the tariff system.⁴ Market entry is regulated through taxi licenses that are issued by a governmental authority as a perpetuity for their holders. At the time of running the field experiment (first wave in July 2010; second wave in March 2012), approximately 14,000 taxi licenses were valid in Athens. This means that there were roughly 350 taxis per 100,000 inhabitants. To compare with, the respective number in London is 280 taxis, in Berlin 210 taxis, in New York City 160 taxis (yellow cabs only), and across the U.S. around 110 taxis per 100,000 inhabitants.⁵ This comparison shows that the supply of taxi rides is relatively large in Athens.

The tariff system is regulated such that there is a fixed fee for entering a taxi plus a variable fee, dependent on the time of the day, the distance travelled and the duration of a ride. This corresponds to a time-varying two-dimensional two-part tariff, a tariff type in place also in many other major cities (*e.g.*, New York City). All over Greece, the tariff looks as follows (effective both in July 2010 and March 2012): The fixed fee is €1.16. The distance-dependent tariff yields €0.66 per kilometer during daytime (*i.e.*, from 5 a.m. until midnight) and €1.16 per kilometer

4. This is common in many major cities, *e.g.*, in New York City, Paris, Brussels, Helsinki, or Quebec, to name just a few (Bekken, 2003; The New York City Taxicab Fact Book, 2006; OECD, 2007, http://www.iedm.org/uploaded/pdf/aout2010_en.pdf).

5. *Source*: Own calculations, based on numbers from The New York City Taxicab Fact Book (2006), “Transportation for London” (<http://www.tfl.gov.uk/businessandpartners/taxisandprivatehire/1380.aspx>), “Taxi Innung Berlin” (<http://www.taxiinnung.org/Taxi-Bestellen.24.0.html>), and Schaller (2005).

during nighttime. In contrast, the duration-dependent tariff is invariable and yields €0.1775 per minute. The algorithm for charging is standardized nationwide in all taximeters and switches automatically to the counting method (distance dependent versus duration dependent) that is more profitable for the driver. That is, if the vehicle's speed during daytime is above 16 km/h, only distance is charged, whereas below 16 km/h only duration is charged. Under reasonable assumptions, it is then straightforward to show that *overtreatment* by taking a detour is more profitable for a taxi driver than choosing a shorter route with a traffic jam.⁶ Also, given the supply and demand conditions in Athens, drivers typically have to queue for passengers—even for long periods of time. This implies that it is generally by far more profitable for a taxi driver to take a passenger on a detour than to choose the shortest and quickest route in the hope of accumulating many fixed fees. As far as *overcharging* is concerned, the incentives are straightforward, since overcharging increases the driver's revenue without affecting the cost of service. It is important to stress that the marginal incentives of taxi drivers to engage in fraud of any type are practically identical both for owners of the taxi and those drivers who lease the vehicle, because leasing a taxi comes at a fixed cost (of roughly €35 per shift; see www.satataxi.gr) and no part of the revenue collected during the shift is shared with the owner.⁷ Therefore, taxi drivers are always residual claimants, meaning that they reap the possible profits from fraud themselves.

3. THE FIELD EXPERIMENT

3.1. Method and procedure

The experiment involved five experimenters switching between three different information and two different income roles. In order to minimize the potential for confounding effects of a passenger's age or gender on driver behaviour, all experimenters were male and in their late twenties. For each route, we had always three (out of the five) experimenters taking a ride from the same starting point to the same destination. They approached the taxi stand one by one in intervals of about two min, such that taxi drivers could never see them together.⁸ The short intervals were chosen in order to control for a host of unforeseeable factors, such as variations in traffic, road works, or accidents, that may influence the optimal route. In the following, we denote the three rides that were taken at practically the same time as a *triple*.

Each experimenter always took a seat in the back of the taxi in order to avoid conversation. For this purpose, we also chose destinations that are easy to find, so that taxi drivers did not have to ask back. In fact, all taxi drivers knew the requested destinations. In case the driver asked which route to take, the choice was explicitly left to the driver.

All experimenters were equipped with a GPS satellite logger. This small device (see Picture A.1 in the Supplementary Appendix) is easy to hide and allowed the experimenters to record the exact route driven, the total distance travelled, the exact duration of the ride, and the location and speed of the taxi at each point in time. In addition to the GPS data, the experimenters collected data on the total fare, as well as the sex and approximate age of the driver.

6. Assuming an average speed of 40 km/h, a fuel consumption of 8 l per 100 km while driving and 1 l per hour while waiting, and a gas price of €1.3 per liter (the price in July 2010; in 2012 it was €1.7), a minute spent on a detour during daytime yields €0.37 [= (0.67 km/min × 0.66 €/km) − (0.67 km/min × 0.08 l/km × 1.3 €/l)], which is considerably more than the estimated earnings of €0.16 [= (0.1775 €/min − 1/60 l/min × 1.3 €/l)] for a minute spent in a traffic jam.

7. Vehicle owners may have a minimally lower marginal net benefit from overtreatment due to depreciation and increased maintenance costs. These costs are very small, however, given the average length of 12.7 km per ride. Moreover, randomization ensures that the minimal differences in marginal incentives between taxi owners and those who lease them do not affect any of our findings.

8. Each experimenter always entered the taxi that was waiting at the front of a queue, as is the rule in Athens.

TABLE 1
Treatments and locations in the experiment

[A] Treatments and number of observations			
Passenger's information role	Passenger's income role		Total
	Low income	High income	
Local	58	58	116
Non-local native	58	58	116
Foreigner	58	58	116
Total	174	174	348

[B] Description of origins and destinations	
Name	Description
Airport	E. Venizelos International Airport
Glyfada	High-income suburb, southern Athens
Karaiskaki square	Run-down neighbourhood (central)
Kifissia	High-income residential suburb, northern Athens
Port (Piraeus)	Main commercial and tourist port
Syntagma	Central square, foreigner area
Train station	Main train station, all intercity trains
Evangelismos	Central Athens
Abelokipi	Middle-income neighbourhood, close to city centre
Bus station	Main bus station, services mainly to southern and central Greece
Pagrati	Central residential area, starting point only

Our treatment variations were implemented as follows. To manipulate a taxi driver's perception about the passenger's information about the city and the tariff system, each passenger had one of three different "information roles". We refer to them in the following as *local*, *non-local native*, and *foreigner*, respectively. In all three roles, an experimenter instructed the driver upon entering the taxi to take him to a particular destination. Passengers in the roles of locals and non-local natives did this in Greek, whereas passengers in the role of foreigners spoke in English. Passengers in the role of non-local natives and of foreigners then asked the driver whether he knew the destination, adding as an explanation for asking that they were not familiar with the city. The question whether the driver knew the destination (plus the added explanation) is the only difference between locals and non-local natives, since both types of passengers spoke in Greek. The language is the only difference between non-local natives and foreigners, both of whom had the same text when entering the taxi.

In addition to an information role, each passenger also had an "income role". Passengers intended to be perceived as having *high income* were dressed in a suit and carried a briefcase, whereas *low-income* passengers were dressed casually and carried a backpack. For routes with a hotel as destination, a high-income passenger would drive to a top-end hotel, whereas a low-income passenger would have a low-end hostel as his destination.⁹ Panel [A] of Table 1 summarizes our treatments and the number of observations per treatment.

We collected observations during 2 weeks in July 2010 and 1 week in March 2012, covering every day of the week and every time of day between 8 a.m. and midnight. The observations were not collected on a single route, but on 16 different ones, covering large parts of Athens and including rich and poor neighbourhoods, as well as typical tourist spots, the international airport, the port, and the main train station. Panel [B] of Table 1 gives a short description of the points of

9. Both addresses were very close to each other in the same street, meaning that the route was practically identical.

origin and destination.¹⁰ To ensure that the main treatment variable (the passenger's information role) is orthogonal to the point in time (*i.e.*, day of the week and time of the day) and in space (*i.e.*, the route taken) where the data were collected, the three experimenters in a triple were always in three different information roles (implying that we have exactly the same number of observations for each information role in any given point in time \times space). On average, the three experimenters in a triple entered the taxis in random order in intervals of 117 s one after the other.

3.2. Hypotheses

Our first hypothesis concerns the influence of the informational asymmetry regarding the optimal route to the destination on the extent of overtreatment by taking detours. Given that non-local native and foreign passengers revealed that they were unfamiliar with the city, whereas local passengers did not, and given that theory predicts that there is not much room for overtreatment of passengers who know the shortest route to their destination, we expected the former two groups to be more prone to be taken on detours than the latter group. We did not expect to see differences between non-local natives and foreigners in this regard, since both were arguably perceived as equally poorly informed about the optimal route to the destination.

H1 (Information on the City): *Non-local native passengers and foreign passengers are more prone to overtreatment than local passengers. The extent of overtreatment does not differ between non-local native passengers and foreign passengers.*

Our second hypothesis refers to the influence of the informational asymmetry regarding the details of the local taxi tariff system on the likelihood and the amount of overcharging. Since taxi tariffs are subject to the same regulation all over Greece, speaking in Greek was meant to convey to the driver that the passenger is likely to know the general rules for charging. Since an English-speaking passenger is arguably less likely familiar with the details of the Greek taxi tariff system, and since theory predicts that there is not much room for overcharging of consumers who can verify whether the correct tariff has been applied, we expected that taxi drivers try to exploit their informational advantage in the charging dimension more extensively with passengers in the role of foreigners than with passengers in the role of locals or non-local natives. We did not expect to see differences between locals and non-local natives in this regard, since both were arguably perceived as equally well informed about the details of the Greek taxi tariff system.

H2 (Information on Tariffs): *Foreign passengers are more prone to overcharging than local passengers and non-local native passengers. The extent of overcharging does not differ between local passengers and non-local native passengers.*

Our third hypothesis is motivated by the large evidence from laboratory experiments that distributional preferences are behaviourally relevant in many important market and non-market transactions (see Cooper and Kagel, 2012, for a survey). Convex distributional preferences imply that a decision maker's benevolence towards another individual increases (or that malevolence decreases) as the income of the other individual decreases along an indifference curve (Cox *et al.*, 2008). Since the overwhelming majority of decision makers who are not exclusively interested in the maximization of their own material income has convex distributional preferences, we expected taxi drivers to overtreat or overcharge low-income passengers less than high-income ones.

H3 (Income): *High-income passengers are more prone to overtreatment and overcharging than low-income passengers.*

10. Table A.1 in the Supplementary Appendix lists all routes. Picture A.2 illustrates them in a map.

TABLE 2
Overtreatment Index (normalizing each route by the shortest one in the triple)

Passenger's role	Low income	High income	Total
Local	1.021	1.037	1.029
Non-local native	1.066	1.087	1.077
Foreigner	1.079	1.096	1.087
Total	1.055	1.073	1.064

4. RESULTS

In total, the five experimenters took 348 taxi rides—174 in the first wave in July 2010 and 174 in the second wave in March 2012—adding up to 4,417 km of travelling through Athens and 128 h of driving.¹¹ The total cost for all rides was €4,347. On average, a ride was 12.7 km long, lasted for 22 min, and cost €12.49. All except six taxi drivers in our sample were male (98%). Each single ride ended at the requested destination, meaning that undertreatment did not occur.

4.1. Overtreatment

We calculate an *Overtreatment Index* by taking, for each triple of rides, the shortest trip and normalizing the other two trips by the shortest one. Table 2 presents the results, showing an index of 1.03 for locals, 1.08 for non-local natives, and 1.09 for foreigners. The difference between passengers in the role of locals and each of the other two passenger roles is statistically significant ($p < 0.01$; two-sided Wilcoxon signed ranks tests), but there is no significant difference between non-local natives and foreigners ($p > 0.3$).¹² These results support hypothesis H1, because they show that passengers conveying unfamiliarity with the city are taken on significantly longer detours while conveying unfamiliarity with the tariff system does not lead to more overtreatment.¹³

Figure 1 plots the cumulative distribution function of the Overtreatment Index for each of the three information roles. It confirms that local passengers experience significantly less overtreatment than both non-local natives and foreign passengers ($p < 0.01$ in each case, two-sided Kolmogorov–Smirnov tests), but there is no significant difference between non-local natives and foreigners ($p > 0.2$).

Of course, a longer and more costly route could, in principle, be driven by a desire of the driver to save on the passenger's time by taking a quicker, albeit longer and thus more expensive, route. This is not what we find in the data, however. In fact, the average duration of a ride was shortest for locals (21 min and 36 s), intermediate for non-local natives (22:06 min), and longest for foreigners (22:20 min). Hence, to say the least, the differences in the length of detours shown

11. In the following analysis, we pool the data of the two waves because the results regarding overtreatment and overcharging—conditional on an experimenter's role—do not differ significantly from each other: both, non-parametric tests and regressions with a dummy for the second wave, show p -values > 0.3 .

12. For these tests, we match the observations of two passengers with different information roles in a triple of rides and apply the Wilcoxon signed ranks test to the resulting pairs of matched observations in all triples.

13. In the Supplementary Appendix (Table A.2 and Figure A.1) we present—as a robustness check—tests using an alternative overtreatment index, one that normalizes each route by the shortest possible route; they yield exactly the same insights.

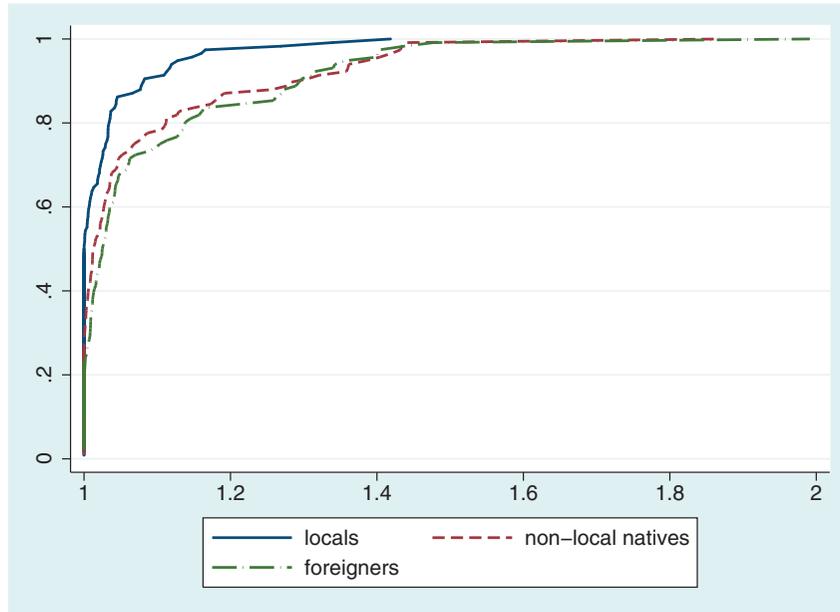


FIGURE 1

Cumulative distribution function of Overtreatment Index

in Figure 1 are certainly not compensated by shorter travel times of those passengers that are being taken on longer detours.¹⁴

Turning to the influence of the income role, Table 2 shows that high-income passengers have an average Overtreatment Index of 1.073, and low-income passengers an index of 1.055. The difference is not significant ($p > 0.2$; two-sided Wilcoxon signed ranks test), meaning that we fail to find support for hypothesis H3 in the overtreatment dimension.¹⁵

4.2. Overcharging

Overcharging occurs when a passenger pays more than he should for a given distance. Three different manifestations of overcharging were identified: (i) the driver switched to the more profitable night tariff even though all rides were during daytime; (ii) the driver did not switch on the taximeter at all, but at the end of the ride demanded a higher price than justified by the distance travelled and the duration of the ride; and (iii) the driver demanded an amount higher than the one shown on the taximeter, with the justification of bogus surcharges.

In total, we observed overcharging in 39 out of 348 rides (11.2%). In four cases the taximeter was not switched on, ten cases were due to the unjustified usage of the night tariff, and bogus surcharges accounted for 25 cases. Panel [A] of Table 3 shows that overcharging occurred in only 3.4% (7.8%) of rides with local (non-local native) passengers, but happened in 22.4% of

14. Indeed, the opposite seems to be true. A Jonckheere-Terpstra-Test shows that rides are quickest for local passengers, intermediate for non-local natives, and slowest for foreign passengers ($p = 0.08$). See also Figure A.2 in the Supplementary Appendix.

15. For the Wilcoxon signed ranks tests along the income dimension, we compare the mean value of the index for the two high-income passengers in each triple (or the two low-income passengers, depending on the distribution of income roles within the triple) with the value of the index for the third passenger.

TABLE 3
Overcharging

Passenger's role	Low income	High income	Total
[A] Relative frequency of overcharging			
Local	0.034	0.034	0.034
Non-local native	0.034	0.121	0.078
Foreigner	0.259	0.190	0.224
Total	0.109	0.115	0.112
[B] Amount of overcharging in €, conditional on overcharging			
Local	2.97	2.43	2.70
Non-local native	1.38	2.28	2.08
Foreigner	6.57	5.21	5.99

TABLE 4
Fare Index

Passenger's role	Low income	High income	Total
Local	1.020	1.055	1.037
Non-local native	1.113	1.110	1.112
Foreigner	1.281	1.200	1.241
Total	1.138	1.122	1.130

cases to foreigners, providing support for hypothesis H2 ($p < 0.01$ for locals versus foreigners, and for non-local natives versus foreigners; $p > 0.1$ for local versus non-local natives; χ^2 -tests). Overcharging is slightly more frequent for high-income passengers (11.5%) than for low-income passengers (10.9%), but the difference is not significant ($p > 0.8$), meaning that we find no support for hypothesis H3 in the overcharging dimension either.

Conditional on overcharging having taken place, the average amount of overcharging was €4.75, which corresponds to 38% of the average fare. Panel [B] of Table 3 shows that the amount of overcharging is highest with foreign passengers ($p < 0.05$ when testing against the pooled data of Greek-speaking passengers; two-sided Mann–Whitney U-test; $N = 39$), while there is no significant difference between local and non-local native passengers ($p > 0.3$), lending further support to hypothesis H2.

4.3. Total fare

Table 4 presents a *Fare Index* as an indicator of the overall amount of fraud. It is calculated as the ratio of a passenger's fare over the minimum fare in a triple. As implied jointly by hypotheses H1 and H2, we see that non-local natives paid higher fares than locals (Fare Index of 1.11 versus 1.04; $p < 0.01$; two-sided Wilcoxon signed ranks test), and that foreigners paid higher fares than non-local natives (Fare Index of 1.24 versus 1.11; $p < 0.01$). The former result is largely driven by differences in overtreatment and the latter by differences in overcharging. The perceived income does not play a role, since the Fare Index is 1.12 for high-income and 1.14 for low-income passengers ($p > 0.9$).

TABLE 5
OLS Regressions

	(1) Overtreatment Index	(2) Amount of overcharging (in €)	(3) Fare Index
Non-resident [#]	0.049*** (0.017)	0.123 (0.093)	0.084*** (0.024)
Foreign [#]	0.002 (0.028)	1.499** (0.660)	0.130** (0.055)
High income [#]	0.017 (0.011)	-0.209 (0.189)	-0.024 (0.020)
Driver female	0.024 (0.026)	-0.395 (0.390)	-0.035 (0.068)
Driver age	-0.000 (0.001)	0.007 (0.008)	0.000 (0.001)
Rush hour [#]	0.024 (0.017)	-0.140 (0.359)	0.026 (0.038)
R^2	0.188	0.186	0.171
Prob > F	0.000	0.000	0.000

$N = 348$. **, *** denotes significance at the 5%, 1% level, respectively.

OLS regressions with route and experimenter fixed effects. Standard errors in brackets, clustered by route.

[#]*Non-resident* is a dummy for passengers revealing that they were not familiar with the city (i.e., for *non-local native* passengers and *foreign* passengers); *foreign* is a dummy for passengers speaking in English; *high income* is a dummy for passengers in the role of high-income passengers; *rush hour* is a dummy with the value 1 on the following times of day: every weekday 8–10 a.m. and 2–4 p.m., as well as Tuesday, Thursday, and Friday 6–8 p.m. (these are the commercial shops' opening hours that were retrieved from the Athens Traders Association).

4.4. *Econometric analysis*

Table 5 presents three different OLS regressions.¹⁶ All regressions include route and experimenter fixed effects, to account for unobserved characteristics of certain routes, as well as for potential confounds due to the appearance of our undercover travellers unrelated to our experimental manipulation. We also cluster standard errors by routes. In the top row of Table 5 we indicate the dependent variables.¹⁷ As independent variables we use, first, a dummy *non-resident* for having expressed unfamiliarity with the city. This dummy captures passengers in the role of non-local natives and foreigners, and thus their presumed informational disadvantage concerning the optimal route to a destination. Second, we include a dummy *foreign* for passengers speaking in English. This dummy is intended to reflect the effects of being perceived as less likely familiar with the taxi tariff system. Third, we add the dummy *high income* for passengers in the high-income role. In addition to these variables we insert further controls, such as the driver's gender and approximate age and a dummy variable for rides taken during rush hours (based on commercial shops' opening hours that were retrieved from the Athens Traders Association).

Specification (1) presents a regression on the Overtreatment Index. Local passengers constitute the benchmark group. We see that *non-resident* is significant and adds an estimated 4.9% of detours. Thus, conveying the impression of unfamiliarity with the city increases overtreatment considerably. Being perceived as a foreign passenger does not further add to overtreatment (in comparison to a non-local native), which is in line with our hypothesis H1. High-income passengers are taken on slightly longer detours than low-income passengers. Yet, the effect is not significant, leading us to refute hypothesis H3.

16. Using Tobit regressions yields practically the same results.

17. We have estimated further specifications as robustness checks, replacing the Overtreatment (Fare) Index by the log of the distance (fare), without any qualitative changes in the results.

Specification (2) looks at overcharging (in €). While conveying the impression of unfamiliarity with the city does not have a significant impact in itself (see *non-resident*), being in the role of a foreigner increases the amount of overcharging significantly by an estimated €1.50 (see *foreign*). This confirms hypothesis H2. Perceived income does not play a role for overcharging (see *high income*), again refuting hypothesis H3.

Concerning the Fare Index, we find in specification (3) that both, *non-resident* and *foreign*, are significantly positive. This shows that both dimensions of informational asymmetry (on the optimal route and on the tariff system) contribute to a higher fare, while again the dummy for high income is insignificant.

5. DISCUSSION

5.1. *Explaining treatment differences*

5.1.1. Differences in detection probabilities. In deriving our hypotheses H1 and H2 we have implicitly assumed that taxi drivers suffer a cost if they are detected as cheaters. Given such a cost, differences in passengers' perceived information and the associated differences in detection probabilities translate into differences in the incentives for fraud.

It seems plausible that a cost of detection exists in material and in non-material terms. Detection may have a material cost through penalties for fraudulent behaviour, either as monetary fines or as the loss of a driver's taxi license. It is important to note that there is an asymmetry between overtreatment and overcharging with respect to potential material costs. A driver accused of overtreatment can argue that he took a detour in the best interest of the passenger to evade a traffic jam. There is no such excuse for overcharging, making overcharging less attractive if material costs play a role. In fact, 46% (30%) of passengers were taken on detours that accounted for at least 5% (10%) of the shortest route, but only 11% of passengers were overcharged. Beyond the potential material costs, there exist most likely some non-material utility losses from being detected as a cheater. These may arise from unpleasant discussions with passengers about the route taken or the tariff applied. Moreover, a driver may feel ashamed or guilty (à la Charness and Dufwenberg, 2006) if a passenger finds out that he was overtreated or overcharged.

5.1.2. Differences in reporting probabilities. There is an alternative explanation for parts of our data patterns that relies on differences in reporting rather than differences in detection probabilities. Assume that the costs of reporting fraud to the authorities are smallest for local passengers, intermediate for non-local natives and highest for foreigners. Such an ordering arises, for instance, when locals know where to report, whereas non-local natives have to find out, and foreigners additionally face a language barrier in communicating with the Greek authorities. Under these assumptions one would expect the lowest level of fraud for local passengers, an intermediate level for non-local natives, and the highest level for foreigners, which matches our observations concerning the Fare Index in Table 4. However, this alternative explanation is hard to reconcile with the disentangled data on overtreatment and overcharging. Specifically, there is a treatment difference in *overtreatment* between locals, on the one hand, and non-local natives and foreigners on the other hand, with no treatment difference between the latter two groups. At the same time, there is a treatment difference in *overcharging* between the set of locals and non-local natives and the set of foreigners, yet no treatment difference in overcharging between locals and non-local natives. This pattern is hardly consistent with an explanation relying on differences in reporting probabilities, because differences in those probabilities are unlikely to depend on the fraud dimension. Note that our line of reasoning in this paragraph—and the following one—depends on our ability to disentangle overtreatment from overcharging through the GPS data.

5.1.3. Reputation concerns. An alternative explanation for the effects of a passenger's information role on the extent of fraud is related to reputation concerns of taxi drivers. The probability of repeated business might be perceived as highest for passengers in the role of locals, intermediate for non-local native passengers, and lowest for foreign passengers. If this were the case, then a driver's incentive to give up short-term gains from fraud in expectation of larger future benefits from repeated business would vary with a passenger's information role, implying the pattern of the Fare Index reported in Table 4. Yet, this explanation does not fit the data on overtreatment and overcharging, the reason being that differences in the perceived probabilities of repeated business across passengers should not result in different behavioural implications depending on the type of fraud. It is also worth mentioning that not a single driver ever offered a business card, mobile phone number, or the like in an attempt to pave the way for another ride with this passenger in the future, indicating that reputational concerns are hardly an important factor in this market. At any rate, existing empirical evidence does not point towards an important effect of reputation in credence goods markets: Dulleck *et al.* (2011) find that allowing sellers to build up reputation has little influence on market efficiency in their lab experiments. This is consistent with the field results reported by Schneider (2012), who finds no evidence of an impact of reputation on under- or overtreatment in the car repair sector.

5.1.4. Refund for expenses. One might argue that passengers in the role of non-local natives or foreigners were defrauded more often because taxi drivers expected them to be able to collect a refund for the travel expenses from a company. This could make them more willing to accept fraud than local passengers, because the latter would have to pay the consequences of the fraud out of their own pockets.¹⁸ We do not think that this hypothesis explains our data well. Specifically, the finding that local passengers are taken on significantly shorter detours than non-local natives, whereas the frequency of overcharging does not differ between the two groups, combined with the finding that non-local natives are less often overcharged (but not less overtreated) than foreigners, is difficult to bring in line with a plausible refund story.

5.2. *Evidence from a survey among taxi drivers in Athens*

After the second wave of our experiment (in March 2012) we interviewed 124 taxi drivers in Athens. Each interview took about 5 minutes, and we paid each driver €4 for participation. The purpose of the interviews was 2-fold:

First, we intended to find out whether parts of our results are driven by a correlation between one of our treatment variations and a driver's perception that the customer is likely to get his expenses reimbursed (as discussed in the previous subsection). Our questionnaire (available in the Supplementary Appendix) therefore included a question (Q2) asking, for each destination in our field experiment, how likely the interviewee would consider a ride to that destination as being business related. We elicited answers with a discrete grading system, ranging from 1 (for "very unlikely") to 5 (for "very likely"). The results (in the Supplementary Appendix) show that three destinations were judged as quite likely to be business related (average score of answers above 4 for each of them) and three destinations as quite unlikely to be business related (average score below 3 for each). Comparing key variables (Overtreatment Index, Overcharging Frequency, Fare Index) between these two subsamples of destinations reveals that there are no significant

18. This hypothesis would be consistent with experimental evidence by Gneezy (2005) who shows that subjects tend to cheat more on others when the costs imposed on others through cheating are smaller.

differences between routes to destinations considered as likely and those considered as unlikely to be business related.

A second purpose of the interviews was to collect data on taxi drivers' perceptions of how other drivers behave. In a generally framed question (Q3) on the perceived level of service provision, each interviewee was asked to assess how many of his colleagues take the shortest possible route to the requested destination (mean answer 3.67 on a scale ranging from 1 for "no one or almost no one" to 5 for "everyone", with 3 for "some", and 4 for "most of them") and how many charge the appropriate amount (mean answer 3.51; same scaling).

In a first of two more specific questions (Q4) we elicited perceptions about likely *determinants of overtreatment* as follows: "There are rumors that some taxi drivers do not always take the shortest route to the destination. Indicate for each of the following explanations how likely you consider it." The answers to this question clearly indicate that—on average—interviewees have correct expectations regarding the behaviour of their colleagues: 55% of the interviewees think that those drivers who take detours do so predominantly when they think the passenger is unfamiliar with the city, while only 31% of the drivers think that income is a likely determinant of overtreatment. The only misperception is the assessment of 64% of interviewees that saving the passenger's time is a likely determinant of overtreatment. Contrary to this assessment, we have found empirically that taking detours leads to longer, not shorter, rides.

A second specific question (Q5) asked for likely *determinants of overcharging* as follows: "There are rumors that some taxi drivers do not always charge the appropriate amount for a ride. [...] Indicate for each of the following explanations how likely you consider it." The answers to this question clearly indicate that the interviewees correctly expect that unfamiliarity with the tariff system is a likely determinant for overcharging, while (high) income is not. Indeed, unfamiliarity with the tariff system as an explanation for overcharging receives the highest approval rate of all proposed explanations (see Supplementary Appendix).

6. CONCLUSION

This article has presented the results of a controlled field experiment on the extent and the determinants of fraudulent behaviour in the provision of a frequently consumed credence good: taxi rides. The first contribution of this article to the literature on credence goods markets has been the exact *and* separate measurement of two common and serious problems in the provision of credence goods, namely overtreatment and overcharging. Using portable GPS loggers we have been able to keep track of the taxi drivers' routes in the city of Athens, Greece, meter by meter, allowing to measure the extent of costly overtreatment by taking unnecessary detours. By letting a triple of passengers—each in a different role—ask for the same service at practically the same time we were able to control for a variety of unforeseeable factors, such as traffic jams. Given the data on the exact length of a route and the information on the local taxi tariff system, we have then been able to identify the amount of overcharging by charging more than justified by the chosen route.

Overall, we have found that 46% of passengers were taken on detours that accounted for at least 5% of the shortest possible route. The overall average detour was 10%, or roughly 1.3 km of the average total length of 12.7 km. Overcharging through manipulating fares was observed in only 11% of possible cases. Recall that overcharging—once detected—is typically much easier to verify than overtreatment, because there are always possible excuses for taking a detour. Thus, the expected material costs of being detected (*e.g.*, the risk of being fined or losing one's license) are probably higher for a taxi driver in the overcharging than in the overtreatment dimension, making the latter more attractive.

The second main contribution of this article has been a controlled manipulation of the driver's perception of a passenger's (i) familiarity with the city, and thus with the optimal route; (ii) information about the local taxi tariff system; and (iii) income. This allowed studying how these factors affect the extent of fraud in the two dimensions overtreatment and overcharging. Consistent with hypothesis H1, we found that taxi drivers exploit their informational advantage over passengers perceived as having less information on the optimal route (*i.e.*, non-local natives and foreigners) by taking them on longer detours. In line with hypothesis H2, we discovered that taxi drivers exploit their informational advantage over passengers perceived as less informed about the tariff system (*i.e.*, foreigners) by overcharging them more frequently and by a higher amount. The manipulation of a driver's perception of the income of the passenger did not yield any significant results, suggesting that distributional preferences play, at best, a minor role.

Our findings have several practical implications. From an individual's point of view, conveying to an expert seller the impression of possessing relevant information (be it true or not), or at least refraining from revealing one's lack of information, can alleviate the problems associated with the provision of credence goods. With car repairs, memorizing some technical terms might help, and in the case of medical treatment the existence of a (fictional) doctor in one's family can be the key to an appropriate treatment. In the case of taxi rides, instructing the driver which route to take might be helpful to demonstrate an ability to verify the optimal route. Some passengers may also be able to calculate approximate fares for a given route, thanks to a number of new apps for smart phones—an example of markets being able to solve severe shortcomings stemming from informational asymmetries without any government interference. Concerning possible policy interventions, it seems noteworthy that some cities have imposed fixed fares for routes disproportionately often requested by less informed consumers, such as the one from the airport to downtown. Another potentially useful intervention could be an obligation for drivers to display in the taxis the results of a car navigation system for the shortest route from the starting point to the requested destination to reduce the informational advantage of expert sellers over buyers of the credence good studied in this article, *i.e.*, taxi rides.

Acknowledgments. We thank the editor, Imran Rasul, four anonymous referees, and Tim Barmby, Colin Camerer, Gary Charness, Nick Feltovich, Uri Gneezy, Martin Kocher, John List, Charles Noussair, and Jan Potters, as well as audiences at APESA 2011 in Kuala Lumpur, IMEBE 2011 in Barcelona, RES 2011 in London and the universities of Aberdeen, Arizona, Athens, Chicago, Edinburgh, Innsbruck, Lyon, Pittsburgh, Tilburg, and York for very helpful comments. Special thanks are due to Konstantinos Konstantakis, Vasileios Selamis, and Dimitrios Mavridis for assisting in the experiment. Financial support from the Austrian Science Fund (FWF grant P20796) is gratefully acknowledged.

SUPPLEMENTARY DATA

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- BEKKEN, J.-T. (2003), *Taxi Regulation in Europe – Executive Summary* (Norway: Institute of Transport Economics).
- BOLTON, G. E. and OCKENFELS, A. (2000), “ERC – A Theory of Equity, Reciprocity and Competition”, *American Economic Review*, **90**, 166–193.
- CHARNESS, G. and DUFWENBERG, M. (2006), “Promises and Partnerships”, *Econometrica*, **74**, 1579–1602.
- CHARNESS, G. and RABIN, M. (2002), “Understanding Social Preferences with Simple Tests”, *Quarterly Journal of Economics*, **117**, 817–869.
- COOPER, D. J. and KAGEL, J. H. (2012), “Other Regarding Preferences: A Selective Survey of Experimental Results”, In Kagel, J. H. and Roth, A. (eds) *Handbook of Experimental Economics*, Vol. 2 (Princeton: Princeton University Press), forthcoming.
- COX, J. C., FRIEDMAN, D. and SADIRAJ, V. (2008), “Revealed Altruism”, *Econometrica*, **76**, 31–69.
- DARBY, M. R. and KARNI, E. (1973), “Free Competition and the Optimal Amount of Fraud”, *Journal of Law and Economics*, **16**, 67–88.

- DOMENIGHETTI, G., CASABIANCA, A., GUTZWILLER, F. and MARTINOLI, S. (1993), "Revisiting the Most Informed Consumers of Surgical Services: The Physician-Patient", *International Journal of Technology Assessment in Health Care*, **9**, 505–513.
- DULLECK, U. and KERSCHBAMER, R. (2006), "On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods", *Journal of Economic Literature*, **44**, 5–42.
- DULLECK, U., KERSCHBAMER, R. and SUTTER, M. (2011), "The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition", *American Economic Review*, **101**, 526–555.
- FEHR, E. and SCHMIDT, K. (1999), "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics*, **114**, 817–868.
- GNEEZY, U. (2005), "Deception: The Role of Consequences", *American Economic Review*, **95**, 384–394.
- HUCK, S., LÜNSER, G. and TYRAN, J.-R. (2007), "Pricing and Trust" (UCL Working Paper).
- HUCK, S., LÜNSER, G. and TYRAN, J.-R. (2012), "Competition Fosters Trust", *Games and Economic Behavior*, **76**, 195–209.
- KINGSBURY, D. (1968), *Manipulating the Amount of Information Obtained From a Person Giving Directions* (Unpublished Honors Thesis, Department of Social Relations, Harvard University).
- LIST, J. A. (2006), "Field Experiments. A Bridge Between Lab and Naturally Occurring Data", *Advances in Economic Analysis and Policy*, **6**, Article 8.
- LIST, J. A. and REILEY, D. R. (2008), "Field Experiments", In: Durlauf, S. N. and Blume, L. E. (eds) *The New Palgrave Dictionary of Economics*, 2nd edn. (Palgrave Macmillan).
- MCGUIRE, T. (2000), "Physician Agency", In Culyer, A. J. and Newhouse, J. P. (eds) *Handbook of Health Economics*, Vol. 1 (Elsevier) 461–536.
- OECD. (2007), "Taxi Services: Competition and Regulation 2007" (Policy Roundtables, Department of Competition Law and Policy, OECD).
- SCHALLER, B. (2005), "A Regression Model of the Number of Taxicabs in U.S. Cities", *Journal of Public Transportation*, **8**, 63–78.
- SCHNEIDER, H. (2012), "Agency Problems and Reputation in Expert Services: Evidence from Auto Repair", *Journal of Industrial Economics*, **60**, 406–433.
- THE NEW YORK CITY TAXICAB FACT BOOK. (2006), "Schaller Consulting [online]" <http://www.schallerconsult.com/taxi/taxifb.pdf> (Accessed February 12, 2011).
- WOLINSKY, A. (1993), "Competition in a Market for Informed Experts' Services", *RAND Journal of Economics*, **24**, 380–398.
- ZITZEWITZ, E. (2012), "Forensic Economics", *Journal of Economic Literature*, **50**, 731–769.