Price Discrimination Factors for Competitive Non-Regulated Taxi Markets

By

Rolando José Arellano Bahamonde

A Dissertation Presented in Partial Fulfillment of the Requirements for the Degree of

Doctor in Business Administration

CENTRUM CATÓLICA GRADUATE BUSINESS SCHOOL

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

MAASTRICHT SCHOOL OF MANAGEMENT

Santiago de Surco, September, 2018

Abstract

The lack of information on price discrimination regarding which characteristics of the client are used and how they influence the definition of the initial price offered in a competitive non-regulated taxi market is the main problem that encouraged this investigation. The study differs from other studies in its use of an experimental research method which allowed analysis of the problem as close as possible to the natural context of the phenomenon. Interviews with 10 taxi drivers produced six variables affecting the process of price definition. A group of 16 people matching those variables collected rates offered by a random sample of taxi drivers. Due to the lack of normality in the distribution of the prices collected, an ordered regression model was implemented. The findings are that price discrimination exists in a non-regulated market such as that of taxis in Lima and that phenotype and the accent of the client are individual characteristics that have a significant influence on the initial price offer. The results confirm that price discrimination is applied in a context like the one of the study, but the question remains as to why it is naturally present and what conditions make it work.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Price determination is one of the basic aspects of the marketing mix, perhaps the one considered as the most important feature for business success (Kotler & Armstrong, 2008). In this context of pricing, price discrimination is an important issue to analyze in markets characterized by a non-regulated large supply with a heterogeneous demand where this pricing strategy can represent a vital factor for the existence of some services like the taxi service, a service that can be defined as inelastic in a short-run analysis (Anderson, McLellan, Overton, & Wolfram, 1997). In these types of markets, price discrimination is commonly present; a clear example is the non-regulated taxi market (no taximeter) in some developing economies. For example, in the city of Lima in Peru, the same service in the same point of sale (a street corner) is offered for a higher price to one person and a lower price to another. This context is very different from the more stable, more regulated, and seldom non-monopolistic markets usually studied by economists that analyze price discrimination (Armstrong, 2005).

Furthermore, most of the marketing studies on price discrimination focus on the characteristics and reactions of the buyers to the different prices and not on the characteristics of the clients that the sellers take into account to discriminate prices. The aim of this research was to try to identify the characteristics used by sellers to discriminate between customers when defining the initial price offered. With this purpose, this study aimed to confirm that price discrimination exists in this market, identify the main criteria for price discrimination (characteristics of the supplier and the demander), and then validate the relationship between all the variables identified.

Background of the Problem

When working with price, marketers have been concerned primarily with production costs even before thinking about sales, customers, or competition (Varble, 1980). Historically, transaction prices have been defined through bargaining processes, but nowadays firms are concerned primarily about management convenience prioritizing fixed price policies. This management convenience of fixed price policies is also supported by a supposed more efficient transaction cost and a perceived fairness. People perceive greater fairness compared to the result of a bargaining process that usually implies some sort of price discrimination. In the United States of America, 31 states have some form of prohibition of price discrimination, either through general laws (such as the Unfair Practices Acts) or through special anti-price discrimination statutes (Grether, 1941; Turow, Feldman, & Meltzer, 2005). However, negotiated prices are used in agriculture as a possible means of gaining higher prices and, hopefully, higher incomes (Sullivan, 1969), and they are still used extensively in developing countries.

"There is growing evidence that the assumption of pure self-interest in a bargaining situation is an inadequate explanation of behavior in many contexts. Often, people behave as if they care not only about their own well-being but also about the well-being of others" (Zwick & Chen, 1999). In a negotiating context, negotiators' offers are often higher than the amount truly necessary to provoke the other party to accept (Corfman and Lehmann, 1993). The failure to predict subjects' behavior in a bargaining process can reasonably be explained if the unobserved and uncontrolled elements of the bargainers' utilities are connected with subjects' perceptions of "fairness," which involve comparing their share of the obtainable wealth to that of the other bargainer (Ochs and Roth, 1989). If the pricing rule (e.g., price discrimination strategy) is believed as fair, a price is judged fair (Dickson & Kalapurakal, 1994). Negotiating seems to have a socially positive impact. In agriculture, for example, the economic effect seems to be superior for sellers, something that could be a consequence of the unobservable "fairness" that seems to be part of the bargaining process (Sullivan, 1969). A similar conclusion should apply to street taxi drivers.

The subject of this study is price discrimination, a subject closely related to the bargaining process. Bargaining is a process that has been previously investigated in the literature. Most of the studies have been realized in the United States and Western Europe where bargaining is not an extensive commercial phenomenon but is used only in the acquisition of products like cars, houses and secondhand goods. In these countries, a considerable part of the population probably never experiences price bargaining, and those who do may have only a few lifetime occasions to bargain (Abdul-Muhmin, 2001). Furthermore, researchers have sought to adopt laboratory experimental approaches to the study of bargaining behavior in these situations, with less emphasis on external validity and greater on internal validity (Cook & Campbell, 1979). Thus, the inquiry remains as to whether the results achieved in these studies have external validity in the sense of supporting across all bargaining situations. Specially, it is still unclear whether these results will be reproduced in a flexible-price market context where bargaining is a common commercial phenomenon. In flexible-price market environments as the described, consumers tend to have vast experience with bargaining and are prone to have developed rules-of-thumb to guide their behavior and expectations in future bargaining circumstances (Abdul-Muhmin, 2001). Most customers notice the concept of charging different amounts to different customers as unfair and often consider it to be illegal, particularly in online settings (Turow et al., 2005). Regardless of these consumer perceptions, price discrimination is legal in most circumstances, as long as the implementation is not centered on a "suspect category" such as race. Therefore, many firms use price discrimination strategies although the risk of adverse customer reactions (Haws & Bearden, 2006; Ramasastry, 2005). Third degree price discrimination is implemented in revenue management policies, by which firms selectively specify higher or lower prices to different segments of consumers (Ferguson, 2014; Wirtz & Kimes, 2007).

Statement of the Problem

On two occasions, the municipality of the city of Lima has unsuccessfully tried to implement taximeters. The measure aimed to reduce the important problem of traffic congestion by standardizing rates with the purpose of eliminating the need of price negotiation before getting a taxi service (Aguirre, 2008). On both occasions, negotiations with price discrimination prevailed; a pricing strategy seems essential for the taxi services market to operate in a context like the city of Lima. It is therefore necessary to confirm that price discrimination exists in this commercial situation and to know which variables taxi drivers use and how they are used to differentiate their rates. These criteria could be very beneficial for the commercialization of other products and services in markets with high levels of competition and high heterogeneity of purchasing power among consumers.

Purpose of the Study

The purpose of this experimental study was to test the theory of price discrimination by determining customer characteristics taken into consideration by independent street taxi drivers (sellers), without price list or taximeter, when defining fares in the city of Lima, Peru. The main objective was to understand the price discrimination policy that remains in most markets in developing countries. An important aspect of the process is that sellers are usually meeting with the customer for the first time and have only a few seconds to gather visual information on which to offer an initial discriminated price.

Significance of the Problem

The aim of this research was to identify the customer- and seller-related factors for initial price discrimination that influence the bargaining processes in markets found in developing countries, which are characterized by an important economic heterogeneity. Countries like Mexico, where higher price elasticities were found among households living in rural areas (for soft drinks), in more marginalized areas and with lower income (Colchero, Salgado, Unar-Munguia, Hernandez-Avila, & Rivera-Dommarco, 2015). In this sense, the aim of the research was to make a contribution to the academic field of marketing by studying a traditional system of price definition not yet fully understood. Another aim was to make a contribution to society, especially to developing countries, by finding factors that could be used as proxy of "economic power". The researcher also wanted to make a contribution to the retail profession by highlighting some fundamental aspects of an existing price discrimination policy present in highly competitive markets. This might allow the development of a mechanism of price discrimination that could be applied to modern retail channels.

Nature of the Study

The focus of the research was the process of price discrimination occurring in transportation services, specifically, empirical taxi fare definition, taxis without taximeters, mobile applications or predefined tariff where excluded. The research consisted of two parts. The first stage was exploratory, in order to identify the main criteria for customer characteristics used by taxi drivers for initial price discrimination. To explore which criteria are taken into consideration to define initial prices, the investigation started with interviews with 10 taxi drivers who, it was assumed, practiced price discrimination to define the initial price offered to each potential client. Applying methodologies derived from the field of psychology, conscious and unconscious parameters used to discriminate prices were identified. It is important to remember that the seller is usually meeting the customer for the first time and for a limited amount of time and has only visual information in order to offer an initial discriminated price.

Next, an experimental quantitative research was developed to validate and rank all the variables identified in the exploratory stage. With this purpose, if the initial prices, the dependent variable, were normally distributed, the intention was to implement a multiple regression with categorical variables to determine the statistically significant variables used by sellers and the degree to which each variable affected the initial price offered to a customer. The alternative (that was adopted) was to conduct an analysis with an ordered regression model such as ordinal logit that respects the dependent variable as an ordinal multimodal outcome (Long, 1997).

Price discrimination is present at the start of the bargaining process when an initial offer is made by the seller (taxi driver), and the bargaining process continues until the seller and the customer agree upon a price; this is how most taxi fares are defined in Lima. To study this process, an experimental research was designed with the purpose of preserving the natural context in which this process occurs. In a natural environment (a street corner), a group of interviewers stopped taxis and asked them the fare for a common journey from the starting point, and the initial price offer was noted. This process was repeated several times in order to complete a sample for each customer prototype, matching each customer prototype with a set of initial prices. Each customer prototype was characterized by an interviewer according to the experimental design based on the variables identified in the previous stage.

Research Questions

The research questions that guided the research were as follows:

1. Does discrimination in the initial price in a non-regulated taxi market exist?
2. What are the characteristics of customers that sellers consider when defining the initial price offer?
3. Is there a significant relationship between external characteristics of customers and the price initially offered to them by sellers?

These questions were used to help understand what characteristics sellers take into consideration when they discriminate initial prices across their customers, customers they meet for the first time and for only a few seconds before offering them the initial price of the service. This first evaluation done by the driver seemed to be very efficient, and most of the time, any error in judgment should be corrected through the bargaining process. The research also helped understand what characteristics of the seller, if any, had an effect on initial prices. As important as it was to identify the variables used to discriminate prices, it was essential to determine the relative importance of these variables in the process. To know which variables were more important than others could help build a general model to discriminate prices based on the most relevant characteristics. In the future, this could also help derive the general model into specific instruments taking into consideration the precision and simplicity needed to be applied. A more precise instrument should include all the variables, but it becomes more complex to apply. Less precision is achieved with fewer variables (variables with less power of discrimination are not considered).

Hypotheses

The following hypotheses were tested:

$H_0 1$: Discrimination in the initial price does not exist in a non-regulated taxi market.

$H_a 1$: Discrimination in the initial price exists in a non-regulated taxi market.

$H_0 2$: There is no difference in the level of discrimination generated between the identified characteristics of customers.

$H_a 2$: There is a difference in the level of discrimination generated between the identified characteristics of customers.

$H_0 3$: There is no relationship between the external characteristics of customers and the initial price offered by sellers.

$H_a 3$: There is a relationship between the external characteristics of customers and the initial price offered by sellers.

Theoretical Framework

The general topic of this research is price discrimination in the empirical practice of price fixing. Pricing strategies for services or products take into consideration three main ways to increase profits; the company owner can cut costs, sell more, or look for more profit with an improved pricing strategy. When sales are hard to increase and costs are already at their lowest, implementing a superior pricing strategy is a key decision to stay economically viable (Tellis, 1986).

Raising prices is not always the way to go, especially in a poor economy. Too many businesses have failed because they priced themselves out of the marketplace. On the other hand, too many businesses leave “money on the table.” One pricing strategy does not fit all, so adopting the right one is a learning curve as we understand the needs and behaviors of customers and clients (Gregson, 2008). In this scenario, price discrimination plays an important role. As Samuelson and Marks (2008) described, first degree price discrimination implicates monopolistic pricing to sell at each customer’s maximum price. Second degree price discrimination is related to quantity discounts. Third degree price discrimination (the focus of this study) occurs when a business charges different prices to different customer groups, also known in the literature as variable consumer pricing (Heyman & Mellers, 2008). Finally, in fourth degree price discrimination prices are the same for different customers nevertheless costs to the firm may vary, also referred as reverse price discrimination as the effects are visible on the producer.

The term price discrimination as used in this research refers to third degree price discrimination, closely related to dynamic pricing or flexible pricing mechanisms made possible by advances in technology and employed mainly online (Clay, Krishnan, Wolff, & Fernandes, 2002). Instead of cost structure or transactional characteristics that must be defensible to all customers relative to other prices and seller costs (e.g., Bolton & Alba, 2006; Bolton, Warlop, & Alba 2003), these firms often price discriminate based upon factors like location of service, temperature, time of purchase, or randomized price components between different customer segments with some paying more and some less for the same service or product (Heyman & Mellers, 2008; Kimes & Wirtz, 2002).

The discriminated initial prices fixed manually by the seller as a response to the characteristics of the client are related with the general topic. A cross-sectional quantitative research that includes experimental survey techniques to collect data was attempted. The dependent variable for the study was the initial price fixed by the seller; the independent variables were the characteristics of the need (the product or service required), and the moderating variables were the set of characteristics of the client, vehicle (the seller), and of the data collection (day and shift) (see Figure 1). As the independent variable, the characteristics of the need (the service required), remained fixed for the entire analysis, all moderating variables studied were treated as independent (see Figure 2). Those two models represent the hypothesis that the set of characteristics of the client influence the initial price.

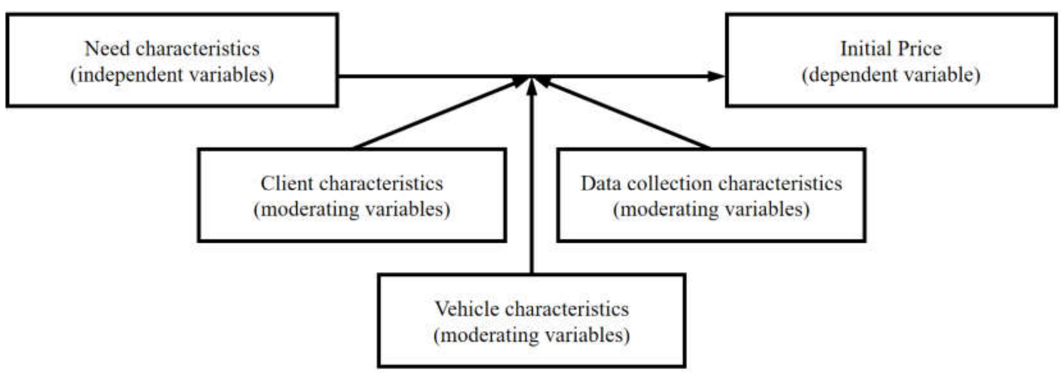


Figure 1. Conceptual framework of the relationship between initial price and need characteristics, moderated by client, vehicle (seller), and data collection characteristics.

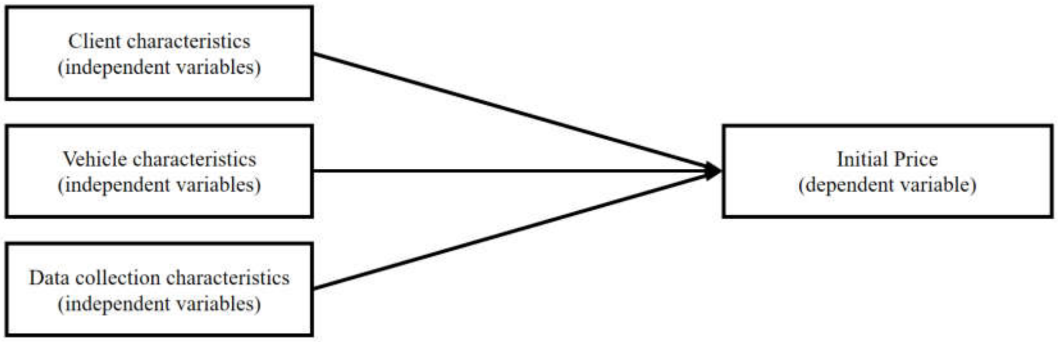


Figure 2. Conceptual framework of the relationship, where all moderating variables are treated as independent while the characteristics of the need remained fixed for the entire analysis.

This kind of pricing strategy used to be present in flexible-price markets, which are defined as markets where most prices are negotiable. In developing countries, such contexts are commonly found, where prices for anything from shirts to refrigerators to cars and houses are negotiable (Kassaye, 1990). One of those countries is Peru, where the study was conducted. One of the first commercial rules a visitor to Peru learns from experiences is always to ask for a rebate or to make a counteroffer when shopping in traditional markets or taking a taxi.

As mentioned by Abdul-Muhmin (2001), in such markets, a classic bargaining situation between a buyer and a seller occurs as follows: the buyer asks the seller how much is the product going for; the seller quotes a price which for this study is known as the initial price or initial offer; then follows a process of offers and counteroffers between the seller and the buyer until an agreement is made. In this study, the focus is on the first step of the bargaining process: the fixation of the initial price offer, a differentiated initial price according to the characteristics of the buyer.

Definition of Terms

Third degree price discrimination refers to a price strategy in which prices vary by individual customers' identity; the attribute in question is used as a proxy for ability and/or willingness to pay. For this kind of price discrimination, the supplier is capable of differentiating between consumer classes (Frank, 2010).

The *initial price or initial offer* corresponds to the seller's quoted price in answer to the customer's question how much the product is going for. After this stage, the bargaining process continues (Abdul-Muhmin, 2001).

Client characteristics refer to every characteristic of the customer noticeable in a few seconds and declared as relevant for fixing the initial price by the seller. Demographic characteristics such as sex and age, ethno-racial markers such as phenotype (physical complexion) and accent, and external appearance such as tidiness and attire were considered relevant for this research (Quijano, 2007).

Customer prototype alludes to a customer with a specific set of the variables relevant for price discrimination. For this study, it was presumed there would be a vast variety of prototypes made up of different combinations of these variables.

The *vehicle characteristics* indicate a set of external features of the seller that could have an influence in the initial price. Variables such as the color, brand, and year of the vehicle were considered relevant for the experiment.

The *data collection characteristics* refer to the day (Saturday or Sunday) and the shift of data collection (morning, afternoon, and late afternoon and evening).

The *need characteristics* refer to the set off variables defining the request of the client. For this study, the researcher fixed those characteristics.

Assumptions

The first assumption was that taxi drivers interviewed would respond honestly to the interview process. Also, it was assumed that differences in operating cost between taxi drivers are non-significant due to the short distance of the ride. Another assumption was that interviewers representing different prototypes of taxi customers would be credible as customers to the taxi drivers.

Scope and Limitations

1. This study was limited to the competitive non-regulated taxi markets.
2. This study was limited to subjects who agreed to participate voluntarily.
3. This study was limited to the number of subjects surveyed and the amount of time available to conduct the study.
4. The validity of this study is limited to the reliability of the instrument used.

Delimitations

This study was confined to a survey of street taxi drivers selected randomly by intercepting them at a street corner. The quantitative study focused on the key variables identified in the previous qualitative stage. Only independent street taxi drivers (those who do not belong to a taxi service company and without a price list, taximeter or mobile applications) were included in the study. The experiment took place on weekends because of the limited availability on weekdays of interviewers representing the different prototypes. The experiment extended over four weekends, instead of shorter period, to avoid producing an unnatural scenario of 16 interviewers in the same place simultaneously asking for prices to the same destination. The experiment was scheduled to be applied in a context where the supply of taxi service was higher than the demand, an extremely competing scenario where initial price definition was particularly important. Taking into account the conditions of the study, it is assumed that there is no bias in the selection of taxis and that all have the same operating costs.

Summary

Price definition is one of the basic aspects of the marketing mix, perhaps the one considered as the most important feature for business success. This aspect is in clear contradiction of the fact that it is still an issue scarcely studied in the area of marketing.

In this context, this research sought to understand, from a marketing point of view, which are the criteria used for price discrimination in markets characterized by a wide supply and a very heterogeneous demand. In these markets, price discrimination is commonly present; a clear example of this kind of market is the non-regulated taxi service market (no taximeter or mobile applications) where the same service in the same point of sale (a street corner) is offered for a higher price to one person and a lower price to another. This price discrimination seems to be an accepted and important mechanism that allows affordability in markets with economically heterogeneous customers.

This research was developed in two key sections. The aim of the first section of the study was to identify the main criteria for initial price discrimination, and the purpose of the second section was to validate and rank all the variables identified in the previous stage.

The first stage was exploratory, consisting of interviews with 10 taxi drivers who offer different initial prices to their customers, in order to explore what criteria they take into consideration to define initial prices. Methodologies derived from the field of psychology were applied to try to identify conscious and unconscious parameters used to discriminate prices. It is important to remember that the seller usually meets the customer for the first time for a limited amount of time and only gets visual information in order to decide the price.

In the second stage, an experimental quantitative research was developed to validate and rank all the variables identified in the exploratory stage. To preserve the natural context,

in a natural environment (a street corner), a group of interviewers stopped taxis and asked them the fare for a common journey from the starting point. This process was repeated several times with the purpose of completing a sample for each customer prototype. Each customer prototype was characterized by an interviewer according to the experimental design based on the variables identified in the previous stage. The strategy of price discrimination studied in this research had the initial price as a dependent variable fixed according to the characteristics of the need, the independent variable, and varying according to each individual on the basis of a set of attributes used as a proxy for affordability or willingness to pay, a set of attributes of the vehicle, and a set of characteristics of the data collection, the moderating variables.

Chapter 2: Literature Review

In order to research the factors used for price discrimination in non-regulated and highly competitive markets, a literature review centered on the following three subjects is considered: (a) dependent variable and price fixing strategies, (b) market segmentation, and (c) moderating variable and third generation price discrimination.

Dependent Variable and Price Fixing Strategies

Alt (1949) presented a short review about how price definition policies have evolved. It begins with the "total cost" policy, used to give sellers a very clear way to fix prices and to give customers a way to understand the price fixing. Simultaneously, the "average cost" policy was used in order to establish an adequate profit level. Later, before the Second World War, the development of monopolistic competition theories permitted the identification of several price policies for enterprises. At this time, economists tried to understand how the enterprise's internal factors such as (a) organizational structure, (b) size, (c) type of property, and also its external factors such as (a) type of products, (b) industry costs, (c) industry maturity, (d) technology, (e) entry – out barriers, and (f) type of distribution channels may affect price fixation. These investigations permitted the development of price fixation policies such as the basing-point system, price leadership, zone pricing, base rating, and price stabilization, centered on the price definition policies from the producer's point of view, lacking the retailer's perspective. Most researchers have studied these policies in a monopolistic situation, and just a few of them have considered a context of competitive markets. The lack of studies of this issue makes it appropriate to recall the comment that Phillips (1946) made referring to the Second World War period:

> Pricing has long been considered as a—perhaps—central marketing and business problem. A large amount of material has been accumulated on the history of price movements. Economics texts are full of discussions of price setting under various conditions—competition, monopoly, and, in recent years, monopolistic competition. Government records, particularly those of the Federal Trade Commission and the Department of Justice, containing testimony on pricing practices are also voluminous.
> In spite of much discussion and research in this field, it is a fair statement that we still lack the basic, detailed case studies of price making which are necessary to a thorough understanding of the problems of pricing and, in turn, of the problems which grow out of various pricing methods. Researchers have not yet sat in with management groups as pricing problems are being settled, and recorded what actually takes place.
> Such case studies would be particularly valuable at the present time when both business and government are making important decisions based on assumptions as to how prices are made and as to which pricing methods are "good" and which are "bad." (p. 21)

Walker (1950) mentioned, "The most successful practitioners of this art [pricing] usually find it difficult to formalize their thinking on price making because intuitive judgments bulk large in pricing decisions." A crucial contribution of consumer research to the matter of pricing is the founding that price perceptions are as much an issue of psychology as of economics (Bolton, Keh, & Alba, 2010). Pricing fixing and negotiation (bargaining) since the beginning of commerce have been closely related. Recently, even if prices used to be fixed, negotiation and bargaining have once again come into vogue. The economic recession in North America and Europe has made people and enterprises more focused on cost reduction. Cost reduction is obtained through more efficient processes and the decrease of supply costs. As Cressman (2006) noted, "when price is the primary focus, customers exhibit price aggression—demanding price concessions from their suppliers—and prices get lower," a negotiation procedure with a negative effect on profit. Such influences could be provoked partially by the existence of a non-discriminated initial price to start the bargaining process. On the other hand, in a commodity market, pricing higher than the competition, in the short term, is a good tactic, but the seller will be confined to a niche market, those willing to pay more, reducing the volume of work available, and minimizing the ability to use volume to reduce pricing (Kehoe, 2004).

In the 21st century, this situation has changed because of the more important role assigned to clients and also because of the development and decrease in price of new data treatment technologies. These changes have permitted the creation of price models that allow the adaptation of prices to each specific client, mostly based on historic buying behavior. This kind of pricing policy looks to maximize the profit obtained with each customer and, consequently, with the whole market. The main goal in this way is then to arrive to the maximum segmentation level: marketing adapted to each specific consumer.

Market Segmentation

Market segmentation can be defined as dividing a market into distinct groups of customers with different characteristics, needs, or behavior, who might need separate products or who may respond differently to various mixes of marketing (Kotler & Armstrong, 2008). Some features of segmentation that may be used include geographic, demographic, psychographic, and behavioral. Effective segmentation typically needs that each segment is assessed on certain criteria such as size, growth potential, stability, accessibility, and responsiveness and whether the customers in that segment and the marketing efforts directed towards them are consistent with company goals and resources.

Because a company has limited resources and must focus on how best to identify and serve its customers, segmentation is crucial. Each segment is characterized by a certain degree of within-group homogeneity that helps ensure that the members of it will respond in similar ways to marketing efforts. This allows firms to apply marketing resources to each segment more efficiently. Of course, companies are interested to undertake segmentation strategies only if these efforts provide a positive return on investment.

Although the benefits of segmentation are now extensively known, this must be weighed against the resource needed to put it in practice (Weinstein, 2004). The obstacles to implementation which practitioners are exposed to are diverse, ranging from lack of data and inappropriate personnel to operational problems and resistance to change. Even overcoming all of these problems, managers are under great pressure to demonstrate the impact and effectiveness of their segmentation plan (Dibb & Simkin, 2009).

The goal of any segmentation is to make a better adaptation of the offer to the demand by identifying groups of consumers (segments) that are more prone to accept one's products and services. Customer relationship management (CRM) looks to understand individual-level behavior, allowing firms to customize marketing campaigns to gradually smaller segments or even to individual customers (Peppers & Rogers, 1993). The development of customer databases and communication technologies (Xie & Shugan, 2001) has allowed firms to begin implementing tailored marketing strategies. This agrees with a growing body of empirical studies focused on the development of individual-level marketing policies (Lewis, 2005; Rust & Verhoef, 2005; Zhang & Krishnamurthi, 2004).

Moderating Variable and Third Generation Price Discrimination

Setting prices to consumers is one of the most critical decisions for a retailer as it is a primary driver of his profitability. To increase profitability, retailers usually employ some sort of price discrimination. For successful price discrimination, economists used to say that the following conditions are necessary: (a) the company must be capable to discriminate between different market segments, such as industrial users and domestic users; (b) each segment must have a different price elasticity; (c) markets must be kept separate, either by physical distance, time, or nature of use; (d) there must be no seepage between two markets, which means that a consumer cannot purchase at the low price in the elastic sub-market and then re-sell to other consumers in the inelastic sub-market, at a higher price; and (e) the firm must have some degree of monopoly power. The three later conditions are strictly necessary in an analysis from the marketing point of view. For instance, in very competitive contexts, retailers often vary prices across stores to exploit demand differences between trading areas, consistent with a strategy of third degree price discrimination. For example, higher prices are often found in stores that are situated in areas with a smaller number of shopping alternatives (Goodman, 2003). Researchers have study how setting optimal retail prices based on competitive factors and observed demographic that can be related to demand characteristics (Chintagunta, Dubé, & Singh 2003; Montgomery, 1997). Cowan (2016) found that, when demand functions in different markets are derived from distributions of reservation prices that differ only in their means, conditions exist such that third-degree price discrimination leads to greater total output and greater total welfare compared to uniform price. For example, Graddy (1995) noticed that sellers of the New Fulton Fish Market were quoting lower prices to Asian customers for the same box of fish compared to white customers. Lii (1995) noted that Asian customers seemed to be price oriented because they resell the product they buy in Chinatown, and they need to maintain the reputation of the cheapest place to buy seafood in New York City. Store owners claimed they must keep their price low due to fierce competition and the fact that “most of their customers’ blue-collar workers simply cannot pay more.” A change in customers’ needs or in the competitive environment can justify the need for segmentation and even price discrimination. Chen, Hu, Szulga, & Zhou (2018) noted that in the Chinese automobile market gender has a large and statistically significant conditional effect on car price and local consumers pay significantly less for vehicles than non-local consumers. Fabra (2018) found that in market characterized with lower search costs, price discrimination benefits small and large buyers compare to medium ones, and smaller ones are more benefited than larger ones. Namata, Ostaszewski, and Sahoo (1990) indicated that despite the importance of pricing strategy for retailer profitability, there is a limited understanding of the relative effectiveness of the price discrimination mechanisms available to the retailer.

A common and implicit notion in the literature on imperfect price competition is that buyers who compare prices across different products are able to remember perfectly all the prices they come across and use them in their decision making. However, there is a considerable body of psychological research that examines the consequence of memory limitations on consumer choice among existing alternatives. Limitations on short-term memory involve that consumers would not be capable to recall exactly relevant price information, and consumers are more likely to face greater limitations of short-term memory in environments with higher levels of information. Imperfect short-term memory of prices is well-documented for consumers buying products purchased routinely or products with low involvement. (Dickson & Sawyer, 1990; Monroe & Lee, 1999). Faced with memory constraints, consumers make choices using heuristics that help them shape suitable price impressions. A relevant heuristic to deal with the large quantity of information is the grouping of objects, events, or numbers into categories based on their perceived similarities (Rosch & Mervis, 1975).

The analysis found quite a few effects of limited consumer recall that are remarkably uniform across the different categorization processes and market environments. When consumers compare either a label to an observed price (asymmetric categorization) or category labels (symmetric categorization), the ideal strategy for the consumers calls for accurate categorization toward the bottom of the equilibrium price distribution. This implies that in equilibrium, consumers should allocate greater memory resources to encoding lower prices to encourage firms to put greater importance on charging more favorable prices.

The literature review shows that economists have studied the price discrimination issue in depth, but that is not the same for marketers. As several authors have noted, there are not many studies on this subject from the marketing point of view. In fact, although some researchers have analyzed some general items that may be used as price determination criteria, no one has focused on identifying scientifically what these criteria are, and no one has analyzed this issue in a competitive market situation (such as the non-regulated taxi market in some developing countries). That is why an aim of this research was to seek to determine how the characteristics of the client, the seller, and the data collection conditions play a moderating role in the definition of the initial price fixed for each client. The dependent variable was the initial price fixed for each client. The independent variable of the study was a set of variables describing the characteristics of the need. All the characteristics of the need were controlled and remained the same for the whole study. Only the characteristics of the client, the seller, and the data collection considerations could vary. The client characteristics refer to every characteristic of the customer noticeable in a few seconds and declared as relevant for fixing the initial price by the seller. Demographic characteristics such as sex and age, ethno-racial markers such as phenotype (physical complexion) and accent, and external appearance such as tidiness and attire were considered relevant for this research. The vehicle characteristics indicated a set of external features of the seller that could have an influence in the initial price. Variables such as the color, brand, and year of the vehicle were considered relevant to be gathered for the experiment. The data collection characteristics refer to the day (Saturday or Sunday) and the shift of data collection (morning, afternoon, and late afternoon and evening). The strategy of price discrimination studied in this research was based on how the relationship between the independent variables and the dependent variable was moderated by a set of attributes of the client, used as a proxy for affordability or willingness to pay, the seller, and of the data collection.

Summary

In order to research the factors used for price discrimination in non-regulated and highly competitive markets, the literature review focused on the following three subjects: (a) dependent variable and price fixing strategies, (b) market segmentation, and (c) moderating variable and third generation price discrimination. A price fixing strategy in the 21st century is more oriented to the role assigned to the clients and is able to process much more data due to the development and decrease in price of new data treatment technologies. These changes have permitted the creation of price models that allow the adaptation of prices to each specific client, mostly based on historic buying behavior. This kind of pricing policy looks to maximize the profit obtained with each customer and, consequently, with the whole market. The main goal is to arrive at the maximum segmentation level: marketing adapted to each specific consumer.

Market segmentation can be defined as dividing a market into different groups of customers with distinct characteristics, needs, or behavior, who might require separate products or who may react differently to various marketing mix efforts (Kotler & Armstrong, 2008). Effective segmentation typically needs that each segment is assessed on certain criteria such as size, growth potential, stability, accessibility, and responsiveness and whether the customers in that segment and the marketing efforts directed towards them are consistent with company goals and resources. The goal of segmentation is for a company to make a better adaptation of the offer to the demand by identifying groups of consumers (segments) that are more prone to accept its products and services. The development of customer databases and communication technologies (Xie & Shugan, 2001) has allowed firms to begin implementing tailored marketing strategies. This agrees with a growing body of empirical studies focused on the development of individual-level marketing policies (Lewis, 2005; Rust & Verhoef, 2005; Zhang & Krishnamurthi, 2004).

Third degree price discrimination is a usual way used by retailers to increase profitability. In very competitive contexts, retailers often vary prices across stores to exploit demand differences between trading areas, consistent with a strategy of third degree price discrimination. Namata et al. (1990) indicated that despite the importance of pricing strategy for retailer profitability, there is a limited understanding of the relative effectiveness of the price discrimination mechanisms available to the retailer. The literature review shows that economists have studied the price discrimination issue, but there is not much research on price discrimination from the marketing point of view. Studies that exist tend to analyze some general criteria to determine prices, but the literature did not reveal anyone focused on identifying scientifically what these criteria are, and no one has analyzed this issue in a competitive market situation (such as the non-regulated taxi market in developing countries).

Chapter 3: Methodology

The purpose of this experimental study was to test the theory of price discrimination by determining customer characteristics taken into consideration by independent street taxi drivers (sellers), without price list or taximeter, when defining fares (initial prices) in the city of Lima, Peru. The main objective was to understand the price discrimination policy that exists in most markets in developing countries. In this process, it is important to remember that sellers are usually meeting customers for the first time and have only visual information on which to offer an initial discriminated price. The research consisted of an exploratory qualitative first stage to identify the main customer characteristics taken into account by the sellers, followed by an experimental quantitative research using statistics to validate and rank the factors found for initial price discrimination in a competitive non-regulated market. The findings may be useful for better understanding of price discrimination, which is a common practice in competitive markets in countries with important economic heterogeneity, particularly in developing countries.

Research Design

To answer the question which customer characteristics independent street taxi drivers (sellers) take into consideration when defining initial prices in the city of Lima, an experimental research was done to preserve the natural context in which this process occurs. In a natural environment (a street corner), a group of interviewers stopped taxis and asked them the fare for a common journey from the starting point. This process was repeated several times in order to gain a complete and representative sample for each customer prototype. Each customer prototype was characterized by an interviewer according to the experimental design based on the variables identified in the first stage. In this experiment, the dependent variable was the initial price fixed by the seller, the independent variables were the characteristics of the need (the service or product required), and the moderating variables were the characteristics of the client. The moderating variables could be nominal or continuous, for example, the type of clothing, skin color, sex, or age, while the dependent variable could be continuous in nature or categorical if not normally distributed. Variables such as the brand and year of production of the taxi were also gathered to analyze whether they had a moderating effect on the relationship. In order to identify qualitatively the main features taken into account in determining a price, 10 taxi drivers were interviewed. The 10 interviews were conducted with drivers recruited at the same corner where the quantitative research was performed. By analyzing these interviews, the variables and levels that explain the variability in the initial price offered were identified. With this information, interviewers were recruited who represented the largest number of combinations of these variables. Each combination of these variables corresponded to what was called a customer prototype.

Even after careful selection of the factors and levels for a study, the overall number of potential prototypes is frequently too large and unmanageable in an experiment like this one. To solve this problem, a suitable fraction of all possible combinations of the factor level was used by implementing a full-profile approach using what is termed a fractional factorial design. The resulting set, named an orthogonal array, was designed to capture the main effects for each factor level. Interactions between levels of one factor with levels of another factor were assumed to be negligible. After identifying the set of prototypes to evaluate, one of the most delicate steps of the experiment ensued, the recruitment of interviewers. Each interviewer recruited was chosen to match with each prototype resulting from fractional factorial designs. Before proceeding with the experiment, a quality control stage of the prototypes took place. For this purpose, a sample of drivers was recruited and asked to evaluate whether each prototype represented the required characteristics. Each potential interviewer recruited, corresponding to one of the prototypes, was maintained only if a significant number of drivers answered according to expectations, so that one could conclude that there was consistency between the prototype and observation. All prototypes that received a significantly different evaluation from what was desired were optimized or replaced and then evaluated again to achieve consistency. After validating the set of prototypes, the experiment was performed.

The experiment took place on weekends, due to the availability of all interviewers, and in three shifts from 08:00 to 12:00, 12:00 to 16:00, and 16:00 to 20:00 in a general context where the supply of taxi service is higher than demand, an extremely competing scenario where initial price definition is extremely important. In a natural environment, a specific street corner, each prototype stopped taxis and, following a strict questionnaire (discourse), they asked them the fare for the same common journey from that starting point. The initial price offered by the taxi driver was recorded, and this process was repeated several times in order to complete a sample of initial prices for each customer prototype. Cases where a taxi driver refused to give an initial price or before giving an initial price asked how much the customer was willing to pay were discarded from the experiment. As a result of this last stage of gathering information, data were obtained with the following characteristics listed in columns: a column represented the dependent variable, the price indicated by the driver, and each of the following columns corresponded to the feature set of the prototype that had requested the price. With these data, a multilevel ordinal logistic regression analysis was conducted. The regression weights provided a quantitative measure of the effect of each on the price variation. To perform a multilevel ordinal logistic regression analysis requires coding responses recorded as text. To code the responses, first they were grouped based on similarity with the aim of detecting responses that differed only due to typographical errors or extra spaces. Because the dependent variable was not normally distributed, an analysis with an ordered regression model such as ordinal logit that respects the dependent variable as an ordinal multimodal outcome (Long, 1997) was conducted.

To answer the question which characteristics of their customers are taken into consideration by independent street taxi drivers (sellers), without price list or taximeter, when defining fares (initial prices) in the city of Lima, Peru, first, a preliminarily identification of the characteristics used by taxi drivers to discriminate prices between one client and another was made, and then the relevance of each of them in this process was quantitatively validated and measured. For the quantitative part of the study, a field experimental research was executed to preserve the natural context in which this process occurs. An experimental research was necessary because the aim was to identify the effect existing between the customer's characteristics (moderating variables) and the initial price proposed to a customer (dependent variable). The brand and year of production of the taxi car were also gathered to analyze whether they were significant in the relationship as moderating variables.

Research Questions

The following research questions guided the research:

1. Does discrimination in the initial price in a non-regulated taxi market exist?
2. What are the characteristics of customers that sellers consider when defining the initial price offer?
3. Is there a significant relationship between external characteristics of customers and the price initially offered to them by sellers?

The purpose of these questions was bring about understanding of what characteristics are taken into consideration by sellers when they discriminate initial prices across their customers, remembering that they are meeting customers for the first time and for only a few seconds before offering them the initial price for the service. This evaluation seems to be very efficient and most of the time, the deviation can be corrected through the bargaining process. The research was also designed to understand what characteristics of the seller and of the data collection, if any, influenced the relationship.

As important as it was to identify the variables used to discriminate prices, it was very important to determine the relative importance of these variables in the process. To know which variables are more important than others could help one to build a general model to discriminate prices based on the relevant characteristics. In the future, this could also help derive the general model into specific instruments taking into consideration the precision and simplicity needed to be applied. A more precise instrument should include all the variables, but if it becomes more complex to apply, less precision is achieved with fewer variables (variables with less power of discrimination are taken off).

The following hypotheses were tested:

$H_0 1$: Discrimination in the initial price does not exist in a non-regulated taxi market.

$H_a 1$: Discrimination in the initial price exists in a non-regulated market.

$H_0 2$: There is no difference in the level of discrimination generated between the identified characteristics of customers.

$H_a 2$: There is a difference in the level of discrimination generated between the identified characteristics of customers.

$H_0 3$: There is no relationship between the external characteristics of customers and the initial price offered by sellers.

$H_a 3$: There is a relationship between the external characteristics of customers and the initial price offered by sellers.

Population

The study universe corresponded to the population of taxi drivers in Lima, Peru, who do not have a meter or tariff to price for their services. According to the Federación Nacional de Taxis y Colectivos, the estimated population of taxis is 240,000 units (Federación de Taxis del Perú, 2011). The participants in the research were required to be active taxi drivers and to participate voluntarily.

Informed Consent

Because the collection of information was the result of field work, it was impossible to require informed consent. If consent had been requested before starting the experiment, it would have skewed the information. Neither could it be applied immediately after obtaining the information, given that taxi drivers were doing their job (driving) and would immediately be looking for a new customer.

Sampling Frame

To carry out the interviews, convenience sampling was used to choose a group of drivers who agreed to participate. For the experiment, taxis who were interviewed were selected according to a systematic sampling technique; the interviewers asked the rates for a specific journey from every third taxi passing through the point of data collection. This process was repeated several times to complete a sample over 100 cases for each customer prototype. For generalization of the results, the ratio of observations to independent variables should never fall below 5:1, implying that five observations are made for each independent variable in the variate. Even though the minimum ratio is 5:1, the objective in this research was to have between 15 to 20 observations for each independent variable (Hair, Black, Babin, Anderson, & Tatham, 2010).

Confidentiality

The data were kept confidential respecting the anonymity of all participants. All information obtained from the interviews was kept in a file without any information allowing the identification of the informant. In the field experiment, it was almost impossible to gather information that allowed identification of individual taxi drivers, so their anonymity and confidentiality were assured.

Geographic Location

Given the complexity of the experiment and to have the maximum control in it, the data collection took place on one corner that is characterized by having a high number of taxis and a large variety of clients. The experiment was conducted at a busy street corner located in the Miraflores district in the city of Lima.

Instrumentation

For the purpose of this research, an interview guide (see Appendix A) was employed to identify the parameters used by taxi drivers to discriminate prices between one client and another, and during the experiment, a questionnaire, including a strict discourse, was used to record the characteristics of the prototype and the initial price offered for the journey. Before proceeding with the experiment, there was a quality control stage of the prototypes. For this purpose, a sample of drivers was recruited to whom each of the prototypes was shown in order to evaluate whether each prototype represented the required characteristics. Each potential interviewer recruited, corresponding to one of the prototypes, was maintained only if a significant number of drivers answered according to expectations, so that one could conclude that there was consistency between the prototype and observation. All prototypes that received a significantly different evaluation from what was desired were optimized or replaced and then evaluated again to achieve consistency before the experiment was performed.

Data Collection

In order to identify qualitatively the main features taken into account in determining an initial price, interviews were conducted with 10 taxi drivers. These interviews were with drivers recruited on the same corner where the quantitative research was performed. The interviews were recorded using a digital recorder. Then, these recordings were transferred to a computer to be transcribed manually into a grid for analysis. Finally, all the features used by taxi drivers to discriminate the initial price were coded in order to identify the main characteristics used for this purpose. In order to validate these variables and quantify their importance in the process of price discrimination, a field experiment was conducted in a natural environment, a specific street corner, where a group of prototypes (people with a set of features, observing a fractional factorial design) stopped taxis and asked them the fare for a specific journey starting from the experiment location. All asked for the same destination and followed the same discourse (according to a strict questionnaire). This process was repeated several times throughout the schedule to complete a sample for each customer prototype. All initial prices received were recorded on a paper questionnaire. These questionnaires were typed in order to build a database to perform appropriate statistical analysis.

Data Analysis

Multiple regression is the suitable method of analysis when the research problem alludes to a single metric dependent variable (price) supposed to be related to two or more metric independent variables, or non-metric if coded. As the independent variable, the characteristics of the need (the service required), remained fixed for the entire analysis, all moderating variables were treated as independent. The objective of the multiple regression analysis was to forecast the changes in the dependent variable (price) as a reaction to changes in the independent variables (customer characteristics).

Because categorical predictor variables were identified, and they cannot be entered directly into a regression model and be meaningfully interpreted, they were coded. To code them, first, the responses were grouped based on similarity, aiming to detect responses that differed only due to typographical errors or extra spaces. After identifying the main variables used by taxi drivers to discriminate prices among potential customers, the general linear model program in Stata was used to execute a multiple regression.

Prior to multiple regression analysis, it is necessary to identify the existence of possible outliers in the data. To do this, the univariate, bivariate, and multivariate diagnostic methods were applied, as described by Hair et al., 2010:

> Univariate methods examine all metric variables to identify unique or extreme observations. For small samples (80 or fewer observations), outliers typically are defined as cases with standard scores of 2.5 or greater. For larger sample sizes, increase the threshold value of standard scores up to 4. Bivariate methods focus their use on specific variable relationships, such as the independent versus dependent variables. We will use a scatterplot with confidence intervals at a specified alpha of 5%. Multivariate methods, best suited for examining a complete variate, such as the independent variables in regression. Threshold levels for the tf/df measure should be conservative (.005 or .001), resulting in values of 2.5 (small samples) versus 3 or 4 in larger samples. (p. 66)

To maximize the prediction from the selected independent variables, one should look for independent variables that have low multicollinearity with the other independent variables but also have high correlations with the dependent variable. It is necessary to analyze the linearity of the relationship between dependent and independent variables, representing the degree to which the change in the dependent variable is associated with the independent variable:

> The regression coefficient is constant across the range of values for the independent variable. The concept of correlation is based on a linear relationship, thus making it a critical issue in regression analysis. Linearity of any bivariate relationship is easily examined through residual plots. (Hair et al., 2010, p. 180)

The next step in the analysis was look for a constant variance of the error term with residual plots, plotting the residuals (studentized) against the predicted dependent values and comparing them to the null plot and Levene test. The next test was to look for the independence of the error terms that each predicted value is independent, not related to any other prediction; that is, they are not sequenced by any variable (Hair et al., 2010). To accomplish this, the residuals were plotted against any possible sequencing variable. If the residuals are independent, the pattern should appear random and similar to the null plot of residuals. The final test before proceeding to execute the multiple regression was to test normality of the error term distribution and normality of the independent or dependent variables or both. For this purpose, the normal probability plots were used to compare the standardized residuals with the normal distribution (Hair et al., 2010).

To select the independent variables, the stepwise estimation method was used because it enables one to examine the contribution of each independent variable to the regression model, adding first the independent variable with the greatest contribution followed by variables with decreasing contribution to the equation (Hair et al., 2010). The multiple regression model obtained was tested to examine its statistical significance, testing the coefficient of determination. Then the significance tests of regression coefficients were carried out.

The following possibilities were considered. If the dependent variable (price) showed responses were highly concentrated and overlapping normal distribution in the histogram showed a huge deviation from a normal distribution, the dependent variable would be considered as not normally distributed. If transforming of the dependent variable (logarithms) did not solve the problem, then transforming the variable into a dichotomy using the mode as a threshold value would be considered. It would lose variability, but it would allow estimating a less restrictive model without imposing normality assumptions. In the case of a lack of normality in the dependent variable, some of the responses would be categorized in the dependent variable and transformed into a categorical variable. One characteristic of the new categorical variable would be that it would retain its cardinal properties; that is, categories can be ranked from low to high except that one cannot note the distances between adjacent categories. This situation would justify conducting an analysis with an ordered regression model such as ordinal logit that respects the dependent variable as an ordinal multimodal outcome (Long, 1997).

In a typical ordinary least squares (OLS) model, the assumption is that responses collected are not biased because they are all within one level: every price collected gave the characteristics of the client, and the assumption was that such characteristics were independent of each other. If they were not, one possible solution was to ignore the existence of such a variation. If the prices were, in effect, correlated with clients, it would always be possible that coefficients associated with higher level-unit of analysis characteristics might be unbiased and the standard errors underestimated, leading to spurious estimates. Another alternative solution to explore was to collapse the Level 1 data to Level 2, ignoring the within-surveyors' variation and then run a classical OLS model. The consequence would be to lose variation and more likely to inflate the coefficients of the relationship and be trapped into an ecological fallacy, leading to the conclusion that the relationships observed between higher-level units could be extended to lower-level units. To take advantage of the implicit research design of this research, a multilevel strategy was contemplated with a given set of random prices (Level 1 unit) to be collected by the same client, a surveyor in the field (Level 2 unit). The nesting (clustering) of prices within the same client made it more likely to expect that the dependent variable would lack independence, that is, that the prices would tend to be similar rather than to exhibit more variation. Multilevel modeling allowed the appropriate addressing of one of the key questions of the research: whether the prices were correlated with the traits of the clients. It is possible to decompose what part of the total variation observed in prices could be attributed to traits that correspond to the taxi drivers or the characteristics of the unit and what part could be attributed to the traits of the surveyor. If the latter share of variation was statistically significant different from zero, the initial hypothesis on the role of client characteristics to set the price could be tested. In this context, fixed effects were the proper model because the focus of the study was on the effects of the Level 2 units in the sample. If the Level 2 units were a sample of a larger universe, then the random coefficients would make sense, and the results could be generalized to such a population (Snijders & Bosker, 2012). The multilevel ordered regression analysis was implemented in Stata 14 using the command mixed effects ordinal logistic regression (meologit).

Validity and Reliability

The internal validity of the research was grounded, taking into consideration the eight factors of Campbell and Stanley (1963): (a) during the experiment, care was taken that no special event influenced the results, (b) during the experiment, no psychological changes occurred within the subjects, (c) all testing of instruments was performed to different subjects having no direct influence on those members of the main sample, (d) all instruments were tested several times to ensure they would work well during the experiment, (e) internal validity related to statistical regression was ensured by identifying and excluding all abnormal cases, outliers, and influential observations, (f) the experiment was performed on 32 samples, the minimum number of samples possible taking into consideration the number of prototypes, and with a 12-hour period of data gathering divided in three shifts, to limit the existence of a differential selection, (g) experimental mortality could be represented by the fact that a taxi driver might not agree to give a price for the journey, something difficult to control but something very unlikely to happen, and finally (h) the experiment was performed as naturally as possible in such a way that one subject would not influence others.

Summary

The purpose of this experimental study was to test the theory of price discrimination by determining customer characteristics taken into consideration by independent street taxi drivers (sellers), without price list or taximeter, when defining fares (initial prices) in the city of Lima, Peru. The main objective of the study was to understand the price discrimination policy that exists in most markets in developing countries. An important aspect of the process is that sellers are usually meeting the customer for the first time and have only visual information on which to offer an initial discriminated price. The research consisted of an exploratory qualitative first stage to identify the main customer characteristics taken into account by the sellers. Interviews were conducted with 10 drivers recruited on the same corner where the quantitative research was performed. This was followed by an experimental quantitative research to statistically validate and rank the factors found for initial price discrimination in competitive non-regulated markets. In a natural environment (the street corner), a group of interviewers stopped taxis and asked them the fare for a common journey from the starting point; the initial price offered by the taxi driver was recorded. This process was repeated several times to complete a sample for each customer prototype. Each customer prototype was characterized by an interviewer according to the experimental design based on the variables identified in the previous stage. In this experiment, the independent variables were the characteristics of the client, which might be nominal or continuous, for example, the type of clothing, skin color, sex, or age, while the dependent variable which corresponds to the initial price offered by the driver for the journey is continuous in nature. The number of the prototypes and their characteristics were defined according to a fractional factorial design, which presents a suitable fraction of all possible combinations of the factor levels. Each prototype (interviewer) was recruited and tested to match correctly with each prototype resulting from fractional factorial designs. After validation of the set of prototypes, the

experiment was performed. This last stage of gathering provided the following data listed in columns: a column represented the dependent variable, the price indicated by the driver, and each of the following columns corresponded to the feature set of the prototype that requested the price. Using these data, a multilevel ordinal logistic regression analysis was conducted. The regression weights provided a quantitative measure of the influence of each on the price variation. To perform a multilevel ordinal logistic regression analysis requires coding responses recorded as text. To code the responses, they were grouped first based on similarity to detect responses that differed only due to typographical errors or extra spaces. The research questions that guided the research were as follows:

1. Does discrimination in the initial price in a non-regulated taxi market exist?
2. What are the characteristics of customers that sellers consider when defining the initial price offer?
3. Is there a significant relationship between external characteristics of customers and the price initially offered to them by sellers?

The study universe corresponded to the population of taxi drivers in Lima, Peru, who do not have a meter or tariff to price their services. The estimated population of taxis is 240,000 units (Federación de Taxis del Perú, 2011). Because the collection of information was the result of field work, it was impossible to obtain informed consent; if requested before starting the experiment, it would have skewed the information collected, and it could not be requested immediately after obtaining the information because taxi drivers were doing their job (driving or looking for a new customer). Convenience sampling was used to choose a group of drivers for the 10 initial interviews. Taxis to be interviewed were selected according to a systematic sampling technique. This process was repeated several times to complete a sample over 100 cases for each customer prototype. Although the minimum ratio for generalization of the results is 5:1, between 15 to 20 observations were obtained for each independent variable (Hair et al., 2010). Given the complexity of the experiment and the need for greater control, it took place at one street corner with a large number of taxis and a high variety of clients. The experiment was performed on weekends, due to the availability of all interviewers, and in three shifts from 08:00 to 12:00, 12:00 to 16:00, and 16:00 to 20:00 in a general context where the supply of taxi services was higher than demand and extremely competitive where initial price definition is imperative.

In order to identify qualitatively the main features taken into account in determining an initial price, 10 taxi drivers were interviewed employing an interview guide. In order to validate these variables and quantify their importance in the process of price discrimination, a field experiment was conducted using a questionnaire to record the characteristics of the prototype and the initial price offered. Because the independent variable, the characteristics of the need (the service required), remained fixed for the entire analysis, all moderating variables studied were treated as independent. For this purpose, a multiple regression was the appropriate method of analysis because the research problem involved a single metric dependent variable (price) presumed to be related to two or more metric independent variables or non-metric if dummy coding was performed. This objective was achieved through the statistical rule of least squares. Prior to multiple regression analysis, it is necessary to identify the existence of possible outliers in the data. To do this, the univariate, bivariate, and multivariate diagnostic methods were applied. To maximize the prediction from the selected independent variables, independent variables were sought that had low multicollinearity with the other independent variables but also had high correlations with the dependent variable. The linearity of the relationship between dependent and independent variables was analyzed, representing the degree to which the change in the dependent variable is associated with the independent variable. The next step was to look for a constant variance of the error term with residual plots, plotting the residuals (studentized) against the predicted dependent values and comparing them to the null plot and Levene test. Next, the independence of the error terms was tested to find out whether each predicted value was independent, not related to any other prediction; that is, they were not sequenced by any variable (Hair et al., 2010). The final test before proceeding to performing the multiple regression was to test normality of the error term distribution, normality of the independent or dependent variables, or both (Hair et al., 2010). The stepwise estimation method was used to select the independent variables (Hair et al., 2010). The multiple regression model obtained was tested to examine its statistical significance, testing the coefficient of determination. Then, the significance tests of regression coefficients were carried out. If the dependent variable (price) were considered as not normally distributed and transforming of dependent variable (logarithms) did not solve the problem, then transforming the variable into a dichotomy using the mode as a threshold value would be considered, looking for a dependent variable transformed into a categorical variable. This situation would justify conducting an analysis with an ordered regression model such as ordinal logit that respects the dependent variable as an ordinal multimodal outcome (Long, 1997). To take advantage of the implicit research design, a multilevel strategy, a given set of random prices (Level 1 unit) were collected by the same client, a surveyor in the field, (Level 2 unit). The nesting (clustering) of prices within the same client made it more likely that the dependent variable would lack of independence, that is, that the prices would tend to be similar rather than to exhibit more variation. Multilevel modeling allowed appropriate addressing of one of the key questions of the research: whether the prices were correlated with the traits of the clients. It was possible to decompose what part of total variation observed in prices could be attributed to traits that corresponded to the taxi drivers or the characteristics of the unit and what part could be attributed to the traits of the surveyor. If the latter share of variation was statistically significant different from zero, the initial hypothesis on the role of client characteristics to set the price could be tested. In this context, fixed effects were the proper model because the focus of the research was on the effects of the Level 2 units in the sample. If the Level 2 units were a sample of a larger universe, then the random coefficients would make sense, and the results could be generalized to such a population (Snijders & Bosker, 2012). The multilevel ordered regression analysis was implemented in Stata 14 using the command mixed effects ordinal logistic regression (meologit).

The internal validity of the research was grounded taking into consideration the eight factors of Campbell and Stanley (1963): (a) during the experiment, care was taken that no special event influenced the results, (b) during the experiment, no psychological changes occurred within the subjects, (c) all testing of instruments was performed to different subjects having no direct influence on those members of the main sample, (d) all instruments were tested several times to ensure they would work well during the experiment, (e) internal validity related to statistical regression was ensured by identifying and excluding all abnormal cases, outliers, and influential observations, (f) the experiment was performed on 32 samples, the minimum number of samples possible taking into consideration the number of prototypes, and a 12-hour period of data gathering divided in three shifts limited the existence of a differential selection, (g) experimental mortality could be represented by the fact that a taxi driver would not agree to give a price for the journey (something difficult to control but unlikely to happen), and finally (h) the experiment was performed as naturally as possible in such a way that subjects would not influence each other. Regarding the reliability of the study, a split-half was performed, testing the significance of the Spearman-Brown coefficient.

Chapter 4: Presentation and Analysis of Data

This chapter contains the presentation of the results of the analysis of the data gathered for the experimental study whose purpose was to test the theory of price discrimination by determining the customer characteristics taken into consideration by independent street taxi drivers (sellers), without price list or taximeter, when defining fares (initial prices) in the city of Lima, Peru. The main objective was understanding the price discrimination policy that exists in most markets in developing countries. This chapter is structured in the following order: (a) the data collection procedures, (b) the development of the experiment, (c) the pilot procedures, (d) the gathering of the data, (e) the setup of an analytical dataset, (f) the data diagnostics, (g) the modeling of price offer, and (h) conclusions.

Data Collection Procedures

Convenience sampling was used to choose a group of 10 taxi drivers who were interviewed with an interview guide. All the answers were recorded and then transcribed into a grid for the analysis. The main features identified that drivers took into account in determining an initial price were (a) demographic: sex and age, (b) ethno-racial markers: phenotype (physical complexion) and accent, and (c) external appearance: tidiness and attire. The levels for each factor found were (a) sex: female and male, (b) complexion: white and mestizo, (c) accent: Peruvian and foreign, (d) tidiness: neat and tacky, and (e) attire: formal and casual. The main features identified as vehicle characteristics were (a) color of the vehicle, (b) brand of the vehicle, and (c) year of the vehicle. With the customer variables identified, a fractional factorial design was run in Stata to find the fraction of all possible combinations of the factor levels that had to be represented by each of the prototypes of clients (interviewers).

Table 1

Characteristics of the 16 Prototypes of Client for the Experiment

Prototype	Sex	Age	Complexion	Accent	Tidiness	Attire
1	Female	Elder (65+)	Mestizo	Peruvian	Neat	Formal
2	Male	Elder (65+)	White	Peruvian	Tacky	Casual
3	Male	Young (-18)	Mestizo	Foreign	Tacky	Formal
4	Male	Young (-18)	Mestizo	Peruvian	Neat	Casual
5	Male	Elder (65+)	White	Foreign	Neat	Formal
6	Female	Elder (65+)	Mestizo	Foreign	Tacky	Casual
7	Male	Young (-18)	White	Foreign	Tacky	Formal
8	Female	Adult (19-64)	White	Peruvian	Neat	Formal
9	Male	Adult (19-64)	Mestizo	Foreign	Neat	Formal
10	Male	Young (-18)	White	Peruvian	Neat	Casual
11	Female	Adult (19-64)	White	Foreign	Tacky	Casual
12	Female	Young (-18)	White	Peruvian	Tacky	Formal
13	Female	Young (-18)	White	Foreign	Neat	Casual
14	Male	Adult (19-64)	Mestizo	Peruvian	Tacky	Casual
15	Female	Young (-18)	Mestizo	Foreign	Neat	Casual
16	Female	Young (-18)	Mestizo	Peruvian	Tacky	Formal

Development of the Experiment

After identifying the set of prototypes to evaluate, interviewers were recruited and asked to respect the level of tidiness and attire of the prototype they had to represent. Next, a team of three supervisors was recruited, who were responsible for ensuring that the interviewers on their schedule respected the procedure and for assisting them if anything unexpected happened. The 16 interviewers were assigned shifts, as tabulated in Table 2, according to their availability and to avoid to repeating a schedule. Each prototype in both shifts was able to gather between 120 and 316 initial prices. Thus, they gathered a total of 3538 observations in all (see Table 3).

Table 2

Schedule of Data Gathering

	Shift 1		Shift 2	
Prototype	Day	Schedule	Day	Schedule
1	Sunday 12	08:00 to 12:00	Saturday 25	12:00 to 16:00
2	Saturday 16	12:00 to 16:00	Sunday 17	16:00 to 20:00
3	Saturday 6	16:00 to 20:00	Sunday 7	08:00 to 12:00
4	Sunday 12	08:00 to 12:00	Saturday 25	12:00 to 16:00
5	Saturday 6	12:00 to 16:00	Sunday 7	08:00 to 12:00
6	Saturday 6	16:00 to 20:00	Sunday 7	08:00 to 12:00
7	Saturday 11	08:00 to 12:00	Sunday 26	12:00 to 16:00
8	Saturday 16	12:00 to 16:00	Sunday 17	16:00 to 20:00
9	Sunday 12	12:00 to 16:00	Saturday 25	16:00 to 20:00
10	Saturday 6	12:00 to 16:00	Sunday 7	16:00 to 20:00
11	Saturday 11	08:00 to 12:00	Sunday 26	12:00 to 16:00
12	Saturday 16	12:00 to 16:00	Sunday 17	08:00 to 12:00
13	Saturday 16	16:00 to 20:00	Sunday 17	08:00 to 12:00
14	Sunday 12	16:00 to 20:00	Saturday 25	08:00 to 12:00
15	Saturday 11	16:00 to 20:00	Sunday 26	12:00 to 16:00
16	Saturday 11	08:00 to 12:00	Sunday 26	16:00 to 20:00

Pilot Procedures

For the quality control stage, the interviewers were asked to submit a photograph that included their entire body. The background of the picture was removed and replaced with plain white. Then, a sample of drivers evaluated the prototype in each picture according to each of the six characteristics to determine how they perceived the prototype. The taxi drivers evaluated all the prototypes as expected, so there was consistency between the prototype and observation.

Data Gathering

In order to qualitatively identify the main features taken into account in determining an initial price, 10 interviews were conducted with taxi drivers. The interviews were recorded using a digital recorder, and these recordings were transferred to a computer to be transcribed manually into an Excel grid for the analysis.

In order to validate these variables and quantify their importance in the process of price discrimination, a field experiment was conducted in a natural environment, at the corner of Schell and Porta Streets in the district of Miraflores, Lima, Peru. In that location, a group of prototypes (people with a set of features, observing a fractional factorial design) stopped taxis and asked them the fare for a specific journey from the corner of Schell and Porta Streets to Larcomar, an important shopping mall a distance of 1.6 kilometers away. All asked for the same route and followed the following discourse in Spanish: *Buenos días/tardes, ¿me podría decir cuánto me cobra hasta Larcomar?* (Good morning/evening, could you tell me how much you charge to Larcomar?) This process was repeated several times throughout the schedule to gather the maximum sample for each customer prototype (see Table 3). All initial prices received and the vehicle characteristics (color, brand, and the number plate of the vehicle) were recorded on a paper questionnaire. In some cases, the number plate of the vehicle was missing due to an unreadable plate or because the taxi went off too fast; in both cases *VACIO* was registered in the corresponding field of the questionnaire. These questionnaires were typed in order to build an Excel database to perform appropriate statistical analyzes.

Table 3

Number of Initial Prices (Observation) Gathered by Prototype

Prototype	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Observation	289	264	152	226	128	179	177	120	212	246	203	316	172	256	300	298	3538

Analytical Dataset Setup

The data gathered in questionnaires were transcribed into an Excel format, and the Excel file (“consolidado taxis (3).xlsx”) was imported into Stata version 14. A total of 3538 observations were imported and read into a new raw dataset. Then, all string/character variables were coded into a format suitable for analysis. Numeric variables were preserved. Most responses were recorded as text. To code them, first, the responses were grouped based similarity aiming to detect responses that differed only due to typographical errors or extra spaces. The variable *shift*, which reflects the schedule in which the collection was done, initially was coded into 10 categories to match the shift when data were collected. Based on the goals of the analysis, they were grouped further into 3 categories, morning (08:00 to 12:00), midday-early afternoon (12:00 to 16:00) and evening (16:00 to 20:00). The variable *year* comes from the number plate registered in the questionnaire, with a search being made for each number plate in the public vehicle register (https://www.sat.gob.pe/Websitev9). Sometimes, when the car was too old, was newly bought, or came from outside Lima, it did not appear in the public vehicle register; in those cases, *No se encuentra placa* was registered in the data. When the year of the vehicle was blank or text-coded as *No se encuentra placa* ("a") and *VACIO* ("b"), the data were treated as missing values. A sizeable part of the sample (47%) lacked a valid response in this item. For the variables *color* and *brand*, some categories were explicitly coded as *VACIO* with 135 and 134 cases. These cases were also coded with the same text in the variable *year*. The variable *price*, the key dependent variable showed a concentration in two categories, 5 and 6 Soles (~89% of the sample). One case (under 0.1% of the sample) showed a value of 35 Soles and was considered an outlier, more than two times higher than the standard deviation (*SD*). Mean value of price with the outlier case was 5.51 Soles (*SD*=0.75, median=5.0). Mean value after removing the outlier was 5.50 (*SD*=0.75, median=5.0). After removing outlier cases, the analytical database had 3537 valid cases.

Data Diagnostics

After exploring the distribution of dependent and independent variables, they were tested for normality and for significance of differences between groups. The dependent variable (price offer) had the following characteristics. Responses of the dependent variable (price) were highly concentrated in just two categories, 5 and 6 Soles, and represented 88.83% of the cases (see Table 4). Overlapping normal distribution in the histogram shows a huge deviation from a normal distribution, and quantile plots show divergence (see Figure 3). The conclusion was that the dependent variable was not normally distributed. A logarithmic transformation of the dependent variable does not solve the problem (see Figure 3). Despite the dependent variable being measured in an interval scale, it does not appear show two other properties: a true continuous and unbounded variable. Transforming the variable into a dichotomy using the mode as a threshold value was considered. It would lose variability, but it would allow estimating a less restrictive model without imposing normality assumptions. Descriptive statistics and distribution of predictors are presented in detailed results in Appendix B and C.

Table 4

Frequency Distribution of Dependent Variable, PRECIO (Initial Offer)

Valid	Freq.	Percent	Valid	Cum.
3	3	0.08	0.08	0.08
4	87	2.46	2.46	2.54
5	1955	55.27	55.27	57.82
6	1187	33.56	33.56	91.38
7	247	6.98	6.98	98.36
8	54	1.53	1.53	99.89
10	4	0.11	0.11	100.00
Total	3537	100.00	100.00	

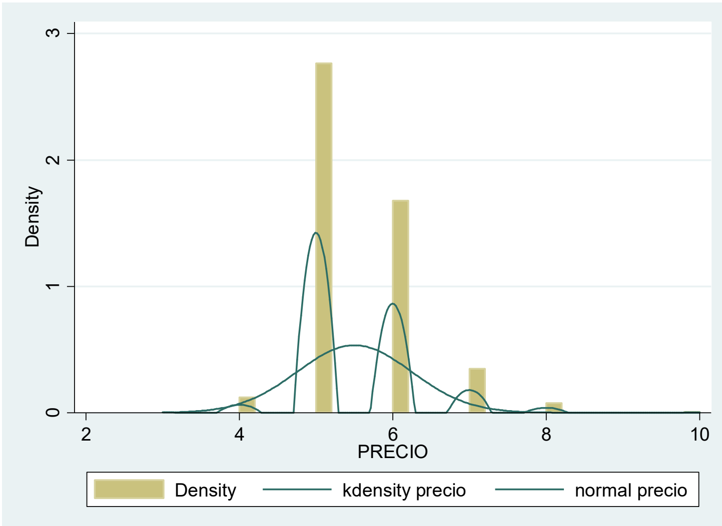


Figure 3. Histogram, normal and empirical (kdensity) distribution Shapiro-Wilks test.

The value of *W* is high but the *p*-value indicates rejection of the null hypothesis (H_0) that values are normally distributed. There is evidence of non-normality in the data. A similar pattern is observed in the results of the Shapiro-Francia test (see Table 5).

Table 5

Results of Normality Tests

Normality tests					
Shapiro-Wilk *W* test for normal data					
Variable	Obs	*W*	*V*	*z*	Prob>*z*
PRECIO	3537	0.97572	48.261	10.059	0.00000
Shapiro-Francia W' test for normal data					
Variable	Obs	*W'*	*V'*	*z*	Prob>*z*
PRECIO	3537	0.97543	52.087	9.803	0.00001
Skewness/Kurtosis tests for normality					
Variable	Obs	*Pr*(Skewness)	*Pr*(Kurtosis)	adj chi2(2)	Prob>chi2
PRECIO	3537	0.0000	0.0000	.	0.0000

Modeling of the Initial Price

The initial model used is the ordinary least squares (OLS), due to the propriety that the dependent variable is not required to be normally distributed (Williams, Grajales, &

Kurkiewicz, 2013). For documentation purposes and to explore the relationship between predictors and independent variables, a regression was run on the price of the ride controlling by characteristics of the surveyor, time of the data collection, and characteristics of the vehicle. These results should be taken with caution and only as a referral due to the problems of non-normality detected. Then, an OLS regression was used with a common specification and 3 models: (a) classical OLS, (b) robust OLS to control for possible deviation of the data, and (c) clustered OLS to adjust for potential clustering effects due to the design of the project, that is, responses within the same surveyor tend to be correlated. Results show magnitude and direction of the relationship and, as expected, estimated coefficients did not change, but standard errors did because they are sensitive to the model specified (see Appendix E). Notice that the most restrictive model (OLS with clustered data) turned all the predictors into statistically not significant. Another important finding was that residuals analysis after controlling for different variables still exhibited a non-normal distribution, making the case for a different approach to model the relationship between the price and the predictors.

Due to the lack of normality in the dependent variable reported in previous sections, a categorization of the responses in the dependent variable was necessary to transform it into a categorical variable (see Table 6 and Figure 4). One characteristic of the new categorical variable is that it retains its cardinal properties, that is, categories can be ranked from low to high except that one cannot determine the distances between adjacent categories. This situation justified conducting an analysis with an ordered regression model such as ordinal logit that respects the dependent variable as an ordinal multimodal outcome (McCullagh, 1980).

Table 6

Distribution of Original and Recode Price Offer

Panel A: Original price offer precio (PRECIO)		Freq.	Percent	Valid	Cum.
Valid	3	3	0.08	0.08	0.08
	4	87	2.46	2.46	2.54
	5	1955	55.27	55.27	57.82
	6	1187	33.56	33.56	91.38
	7	247	6.98	6.98	98.36
	8	54	1.53	1.53	99.89
	10	4	0.11	0.11	100.00
	Total	3537	100.00	100.00	
Panel B: Recoded price offer precio_cat -- RECODE of precio (PRECIO)					
Valid	1 <5	90	2.54	2.54	2.54
	2 5	1955	55.27	55.27	57.82
	3 6	1187	33.56	33.56	91.38
	4 7+	305	8.62	8.62	100.00
	Total	3537	100.00	100.00	

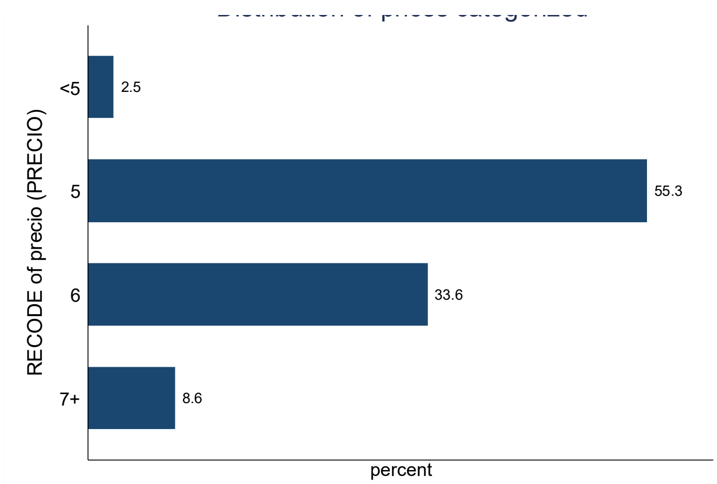


Figure 4. Distribution of price offers after recategorization.

In a typical OLS model, it is assumed that responses collected are not biased because they are all within one level: for every price collected, the characteristics of the client are obtained, and the assumption is that such characteristics are independent of each other, but such a situation was not strictly true in this research. One possible solution is to ignore that such variation exists, but if the prices are in effect correlated with clients, it is always possible that coefficients associated with higher level unit of analysis characteristics might be unbiased and the standard errors underestimated, leading to spurious estimates. Another alternative solution is to collapse the Level 1 data to Level 2, ignoring the within-surveyors' variation and then run a classical OLS model. The consequence is to lose variation and more likely to inflate the coefficients of the relationship and be trapped into an ecological fallacy, leading to the assumption that the relationships observed between higher level units are to be extended to lower level units. In this situation and to take advantage of the implicit research design of this project, a multilevel strategy, a given set of random prices (Level 1 unit) were collected by the same client, a surveyor in the field (Level 2 unit). The nesting (clustering) of prices within the same client makes it more likely that the dependent variable would lack independence, that is, that the prices tend to be similar rather than to exhibit more variation. According to the multilevel modeling framework proposed by Long (1997), this model allows one appropriately to address one of the key questions of the research: whether the prices are correlated with the traits of the clients. It is possible to decompose what part of total variation observed in prices can be attributed to traits that correspond to the taxi drivers or the characteristics of the unit and what part can be attributed to the traits of the surveyor. If the latter share of variation is statistically significant different from zero, the initial hypothesis on the role of client characteristics in setting the price can be tested. Fixed effects are the proper model because the research was only interested in the effects of the Level 2 units in the sample. If the Level 2 units were a sample of a larger universe, then the random coefficients would make sense, and the results could be generalized to such population (Snijders & Bosker, 2012). The multilevel ordered regression analysis was implemented in Stata 14 using the command mixed effects ordinal logistic regression (meologit). Specification for the models is as follows:

- Dependent variable
    - Categorized price offers
- Independent variables: 3 blocks of variables distributed at 2 levels
    - Level 1: characteristics associated with the price offer
        - Vehicle characteristics
            - Color of the vehicle
            - Brand of the vehicle
            - Year of the vehicle
        - Data collection characteristics
            - Day of data collection
            - Shift of data collection
    - Level 2: characteristics of the client
        - Demographic

    - Sex
    - Age
- Ethno-racial markers
    - Phenotype (physical complexion)
    - Accent
- External appearance
    - Tidiness
    - Attire

First, we start with a simple (no levels) model for ordinal dependent variables with a single independent variable (Winship & Mare, 1984):

$$y_i{}^* = \alpha + \beta x_i + \varepsilon_i$$

where $y_i{}^*$ is a latent variable ranging from -∞ to ∞, i is an observation, and ε_i is a random error (with a standard logistic distribution, hence the ordered logit model). The measurement model is expanded to divide $y_i{}^*$ into J ordinal categories:

$$y_i = m \text{ if } \tau_{m-1} \leq y_i{}^* < \tau_m \text{ for } m = 1 \; to \; J$$

where τ_1 and τ_{J-1} are cut-points or thresholds and need to be estimated. It is assumed that $\tau_1 = -\infty$ and $\tau_J = \infty$. Then, the notation is extended to accommodate a multilevel representation of a simple model, with the following equations for difference specifications (Guo & Zhao, 2000; Raudenbush & Bryk, 2002):

Model 0 is a null model used to estimate the intra-class correlation and comes from the combined model:

$$y_{ij}^* = \beta_0 + u_j + e_{ij}$$

where y_{ij}^* is the outcome variable for the ith unit at Level 1 (vehicle and data collection) and jth unit at Level 2 (client); β_0 is the grand intercept; u_j is a random effect accounting for the random variable at Level 2 (note the term s β_0 and u_j can be reexpressed as β_{0j}); and e_{ij} is the Level 1 random effect. The within-cluster or intra-class correlation is obtained from:

$$\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$$

for ordinal logistic models, the σ_e^2 is assumed to be equal to $\pi^2/3$; hence the estimated variance at Level 2 is 0.08321 that indicates that there is a chance of 8% of finding 2 similar prices within each nested level unit (the higher the variance at this level, the more correlated are the responses in the nested level).

Model 1 shows bivariate regression with each separate client (Level 2) characteristic without controls:

$$y_{ij}^* = \beta_0 + \beta_1 c_j + \ u_j + e_{ij}$$

where c_j is an individual explanatory variable that indicates a client characteristic.

Model 2 shows multivariate regression with each separate client (Level 2) characteristic controlled by Level 1 characteristics:

$$y_{ij}^* = \beta_0 + \beta_1 c_j + \beta_2 V_{ij} + \beta_3 D_{ij} + \ u_j + e_{ij}$$

where $\beta_2 V_{ij}$ is a vector of characteristics of the vehicle (Level 1) and $\beta_3 D_{ij}$ is a vector of characteristics of the data collection process (Level 1).

Model 3, multivariate regression, has all client (Level 2) and vehicle and data collection (Level 1) characteristics in a saturated model:

$$y_{ij}^* = \beta_0 + \beta_1 C_j + \beta_2 V_{ij} + \beta_3 D_{ij} + \ u_j + e_{ij}$$

where C_j is a vector of explanatory variables related to the client characteristics (Level 2).

Findings

Regarding the results of the multilevel ordinal logistic regression, Table 7 shows the standardized effects of the client's characteristics on the independent variable (price offer categorized). Due to the dichotomous nature of the predictor variables, standardized coefficients were used to describe the findings, hence preserving the original unit of measurement of the variables. An overall finding is that among the set of personal characteristics, those that might be proxies of ethnic and/or racial categories might have an influence on the price offer. Thus, only the phenotype (physical complexion) has a statistically significant effect on the price. A white client received a price offer between 0.36 and 0.38 standard deviations higher than a mestizo client. This finding is consistent across the different models although, in Model 2, the level of significance is slightly under the conventional 95%. Another ethnic marker such as client's accent has an influence in terms of increasing the price offer. Though the results are marginally significant, the effects of having a foreign accent ranges from .239 to .266 standard deviations, being significant only in Model 3. Demographic traits such as sex and age exhibit a negative effect on the price offer. Thus, a female is more likely to receive a lower fare than a male (.22 to .23 *SD*, not significant). An increase in the age of the client also has a negative impact on price: in particular, for those in the 19-64 age group (.288 *SD*, only significant in Model 3). The remaining characteristics are related to personal image traits such as tidiness and attire. The effect of both characteristics is in opposite directions; a client with a less groomed style (tacky) is more likely to receive a higher price: the effects range from .13 to .16 standard deviations but are not statistically significant for any model. Finally, a client wearing more casual attire is more likely to receive a price offer that is lower than a client with more formal attire (work-related attire), but in all cases, the statistical significance of the coefficients is under the conventional thresholds (under 90%).

Table 7

Standardized Effects of the Client's Characteristics on the Price Offer

	Model 1		Model 2		Model 3	
	Direct effect (no controls)		Effect controlling by Level 1 characteristics		Effect controlling by Level 1 and Level 2 characteristics	
	(b)	(t)	(b)	(t)	(b)	(t)
Panel A: Sex						
Male (reference category)						
Female	-0.225	(-1.13)	-0.236	(-1.16)	-0.235	(-1.60)
Panel B: Age						
Under 19 (reference category)						
19-64	-0.279	(-1.40)	-0.283	(-1.38)	**-0.288+**	**(-1.90)**
65+	-0.162	(-0.79)	-0.159	(-0.75)	-0.164	(-1.06)
Panel C: Complexion						
Mestizo (reference category)						
White	**0.380***	**(2.08)**	**0.369+**	**(1.93)**	**0.362***	**(2.46)**
Panel D: Accent						
Peruvian (reference category)						
Foreigner	0.239	(1.22)	0.262	(1.31)	**0.266+**	**(1.82)**
Panel E: Tidiness						
Neat (reference category)						
Tacky	0.132	(0.64)	0.161	(0.77)	0.160	(1.08)
Panel F: Attire						
Formal (reference category)						
Casual	-0.120	(-0.59)	-0.096	(-0.46)	-0.094	(-0.64)
Controls						
Level 1 (characteristics of the vehicle and data collection)	No		Yes		Yes	
Level 2 (client characteristics)	No		No		Yes	
Observations	3537		3537		3537	

Note. Dependent variable is the ordinal categorization of price of the ride. Results are estimated using a mixed effects ordinal logistic regression. Effects are expressed as standardized coefficients. *T* statistics reported in parentheses.
Significance levels: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

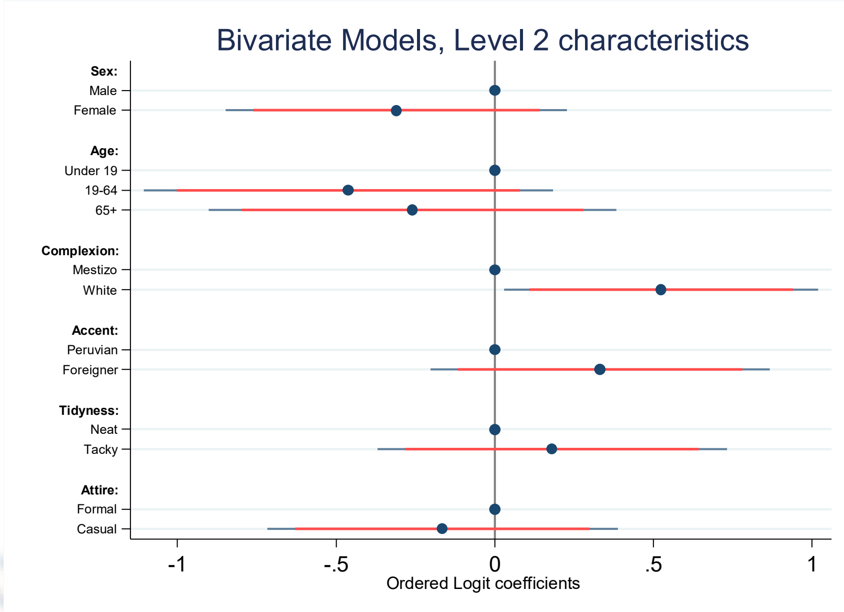


Figure 5. Effects of Level 2 characteristics on price offer based on bivariate regression models.

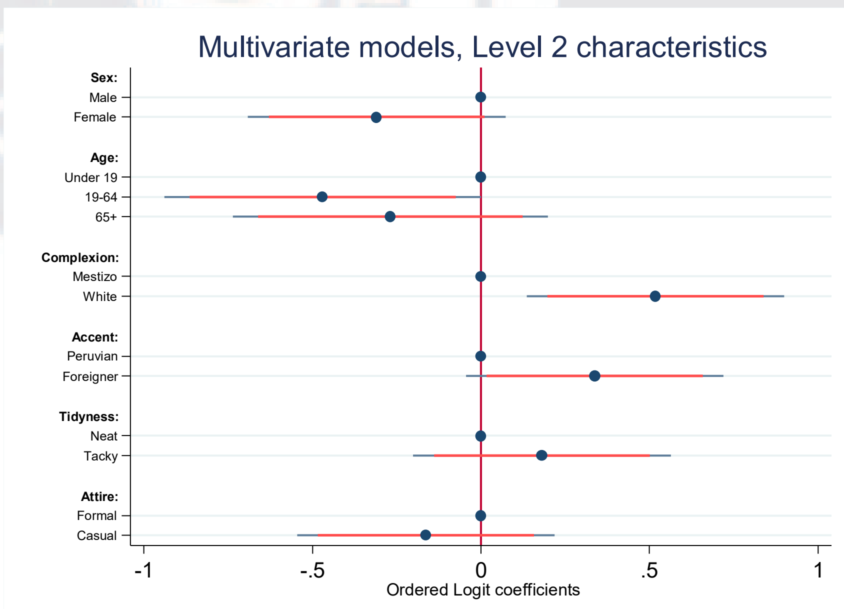


Figure 6. Effects of Level 2 characteristics on price offer based on multivariate regression without control.

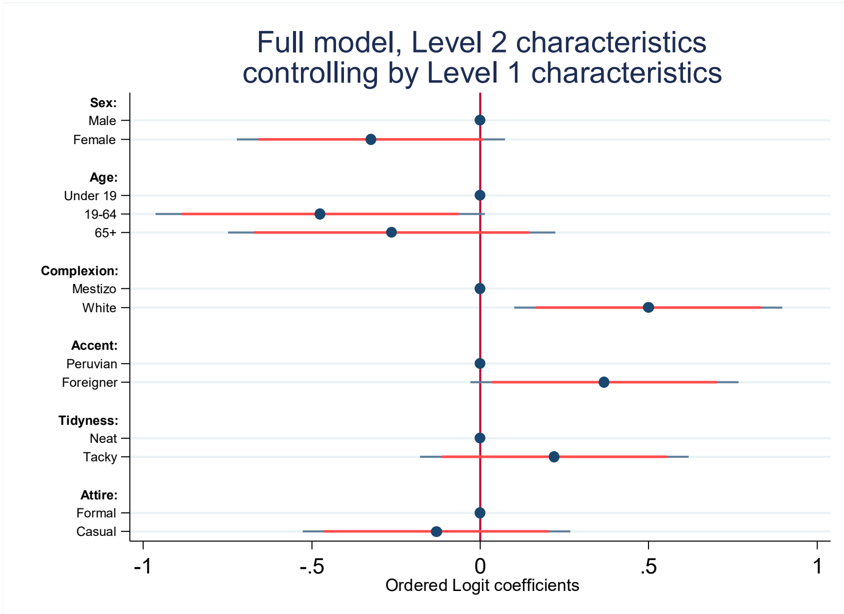


Figure 7. Effects of Level 2 characteristics on price offer based on multivariate regression models controlling by Level 1 characteristics.

Regarding the predicted effects of characteristics on price, we started by analyzing the role of the client's phenotype (physical complexion). To help interpret the effect of the significant variable on the price offer, we estimate the change in the predicted probabilities due to a discrete change in the client's physical traits:

$$\Pr(y = m|x) = F(\tau_m - x\beta) - F(\tau_{m-1} - x\beta)$$

where $F(.)$ indicates the cumulative probability of cut-point τ at outcomes 1 to *J*. Our predictor of interest is a dummy variable so the discrete change in the probability of observing a given price offer for a change of phenotype can be expressed as the change from the mestizo to the white category. The following equation expresses the discrete change while holding all other variables constant at their mean values:

$$\frac{\Delta \Pr(y = m|\bar{x})}{\Delta x_k} = \Pr(y = m|\bar{x}, x_k = x_E) - \Pr(y = m|\bar{x}, x_k = x_s)$$

where m indicates any outcome possible for the dependent variable y; x_E and x_S indicate the start and end value of the predictor variable, in this case 1 and 0 respectively; x_k indicates the changes from x_S to x_E; and $\bar{x}$ indicates the mean values of the other variables.

The overall results are reported in Table 8 and showed in Figure 8: (a) a client with a mestizo complexion has a higher probability of receiving a lower price than a client with a white complexion; (b) a client with a mestizo complexion has a 50% higher probability of getting a price offer of less than 5 Soles in contrast with his white counterpart (3% vs. 2%); (c) at a higher price (5 Soles), such gap decreases, but it still favors mestizo clients: 60% of mestizo clients are likely to receive a price offer of 5 Soles, while among white clients, only 50% get such price; (d) clients who receive a price offer of 6 Soles are more likely to be white than mestizo (probability of 38% vs. 30%); (e) and similarly, if a client gets a price offer of 7 Soles or more, it is more likely that he has a white complexion (10% vs. 7%).

Table 8

Predicted Probability of Price Offer Adjusted by Change in the Client's Phenotype

	(1) <S/.<5	(2) S/.5	(3) S/.6	(4) S/.7+
Mestizo	0.0304***	0.600***	0.301***	0.0688***
	(5.83)	(21.44)	(13.02)	(7.04)
White	0.0188***	0.497***	0.377***	0.107***
	(5.62)	(15.71)	(17.17)	(7.48)
N	3537	3537	3537	3537

Note. *t* statistics in parentheses.

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

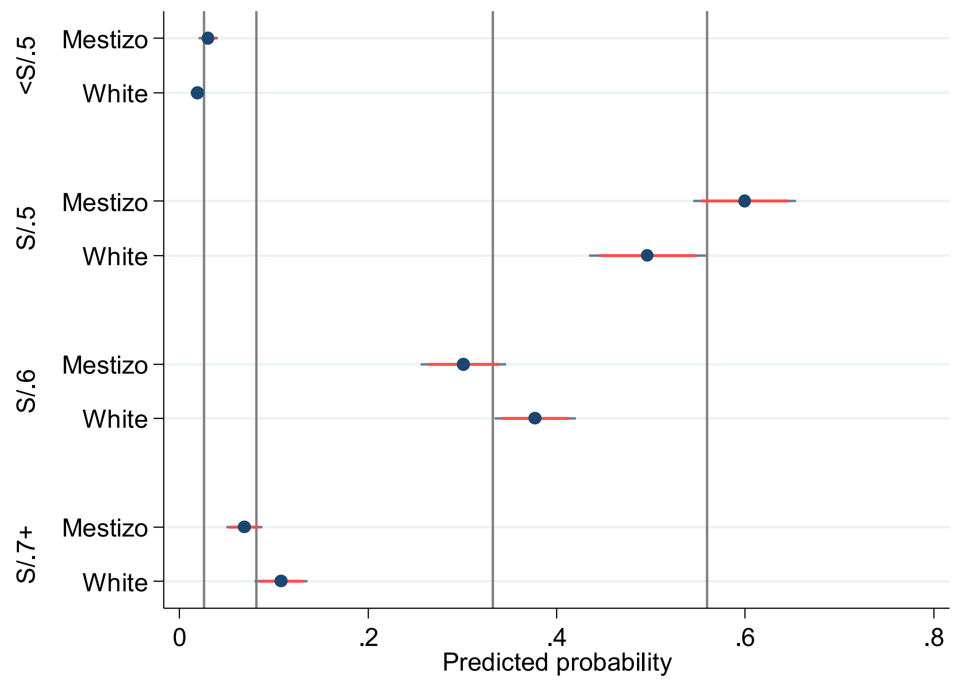


Figure 8. Predicted probability of price offer adjusted by completion (phenotype).

Note. Vertical gray lines indicate the observed probability for each price offer (<5=.025; 5=.553; 6=.336; and 7+= .086).

Conclusions

The conclusions regarding the research questions are as follows:

1. Does discrimination in the initial price in a non-regulated taxi market exist? Discrimination exits, but it is limited.
2. What are the characteristics of customers that sellers consider when defining the initial price offer? Phenotype (proxy for race/ethnicity) and accent (foreigner) have a significant effect on the initial price offer.
3. Is there a significant relationship between external characteristics of customers and the price initially offered to them by sellers? A negative relationship was found for the phenotype (statistically significant) and a positive relationship for a foreign accent (but marginally significant).

Summary

The purpose of this chapter was to present the results of the analysis of the data gathered in the experimental study to test the theory of price discrimination by determining the customer characteristics taken into consideration by independent street taxi drivers (sellers), without price list, taximeter or mobile application, when defining fares (initial prices) in the city of Lima, Peru. The main objective was to understand the price discrimination policy that exists in most markets in developing countries.

In the first stage of the research, 10 interviews were conducted to identify the main features taken into account in determining an initial price: (a) demographic: sex and age, (b) ethno-racial markers: phenotype (physical complexion) and accent, and (c) external appearance: tidiness and attire. The levels for each factor found were: (a) sex: female and male, (b) complexion: white and mestizo, (c) accent: Peruvian and foreign, (d) tidiness: neat and tacky, and (e) attire: formal and casual. The main features gathered for the vehicle characteristics were (a) color of the vehicle, (b) brand of the vehicle, and (c) year of the vehicle. With the customer variables identified, fractional factorial design in Stata was run which yielded 16 possible combinations of the factor levels that had to be represented by each of the prototypes of clients (interviewers).

After identifying the set of prototypes to evaluate, interviewers were recruited and asked to respect the level of tidiness and attire of the prototype they had to represent. All the interviewers recruited were evaluated as the prototypes they had to represent so that there was consistency between the prototype and observation. During the experiment, each of the prototypes gathered between 120 and 316 initial prices, gathering a total of 3538 observations. These observations were obtained under particular circumstances: a controlled location, on weekends.

The data gathered in questionnaires were transcribed into an Excel format, then imported into Stata version 14. A total of 3538 observations were imported and read into a new raw dataset. Then all string/character variables were coded into a format suitable for analysis. Numeric variables were preserved. After removing outlier cases, the analytical database had 3537 valid cases.

The distribution of dependent and independent variables was explored and the variables tested for normality and for significance of differences between groups. The dependent variable (price offer) was highly concentrated in just two categories (prices S/. 5 and 6/. ; ~89% of the cases). This led to the conclusion that the dependent variable was not normally distributed. Due to the lack of normality, some of the responses were categorized in the dependent variable and transformed it into a categorical variable (<5, 5, 6, and >7), retaining its cardinal properties.

This situation justified conducting an analysis with an ordered regression model such as ordinal logit that respects the dependent variable as an ordinal multimodal outcome. In a typical OLS model, the assumption was that responses collected are not biased because they are all within one level: for every price collected, the characteristics of the client were obtained and assumed to be independent of each other. Taking advantage of the implicit research design of this project, a multilevel strategy, a given set of random prices (Level 1 unit) were collected by the same client, a surveyor in the field (Level 2 unit). The nesting (clustering) of prices within the same client made it more likely that the dependent variable would lack independence, that is, that the prices would tend to be similar rather than to exhibit more variation. This multilevel modeling allowed appropriately addressing one of the key questions of the research: whether the prices were correlated with the traits of the clients. It was possible to decompose what part of total variation observed in prices could be attributed to traits that corresponded to the taxi drivers or the characteristics of the unit and what part could be attributed to the traits of the surveyor. With the latter share of variation statistically significant different from zero, the initial hypothesis on the role of client characteristics to set the price could be tested. The multilevel ordered regression analysis was implemented in Stata 14 using the command mixed effects ordinal logistic regression (meologit).

The test started with a simple (no levels) model for ordinal dependent variables with a single independent variable. A multilevel ordered regression analysis followed using three models: (a) Model 1, bivariate regression, with each separate client (Level 2) characteristic without controls; (b) Model 2, multivariate regression, with each separate client (Level 2) characteristic controlled by Level 1 characteristics; and (c) Model 3, multivariate regression, with all client (Level 2) and vehicle and data collection (Level 1) characteristics in a saturated model.

Due to the dichotomous nature of the predictor variables, standardized coefficients were used to describe the findings, hence preserving the original unit of measurement of the variables. An overall finding is that among the set of personal characteristics, those that might be proxies of ethnic and/or racial categories might have an influence on the price offer. Thus, only the phenotype (physical complexion) had a statistically significant effect on the price. A white client received a price offer between 0.36 and 0.38 standard deviations higher than a mestizo client. Another ethnic marker, the client’s accent, had an influence in terms of increasing the price offer. A foreign accent ranges from .239 to .266 standard deviations, being significant only in Model 3. Demographic traits such as sex and age exhibited a negative effect on the price offer; a female was more likely to receive a lower fare than a male (.22 to .23 *SD*, not significant). An increase in the age of the client also had a negative effect on price: in particular, for those in the 19-64 age group (.288 *SD*, only significant in Model 3). The remaining characteristics are related to personal image traits such as tidiness and attire. The influence of both characteristics was in opposite directions. A client with a less groomed style (tacky) was more likely to receive a higher price: the effects ranged from .13 to .16 standard deviations, but they were not statistically significant for any model. Finally, a client wearing more casual attire was more likely to receive a price offer that was lower than a client with more formal attire (work-related attire), but in all cases, the statistical significance of the coefficients was under the conventional thresholds (under 90%).

The analysis shows that a client with a mestizo complexion had a higher probability of receiving a lower price than a client with a white complexion: (a) a mestizo had a 50% higher probability of getting a price offer of less than 5 Soles in contrast with his white counterpart (3% vs. 2%), (b) 60% of mestizo clients were likely to receive a price offer of 5 Soles, while among white clients only 50% were offered such price, (c) clients who received a price offer of 6 Soles were more likely to be white than mestizo (probability of 38% vs .30%), and (d) if a client received a price offer of 7 Soles or more, it was more likely that he had a white complexion (10% vs. 7%).

The findings regarding the research questions were that discrimination in the initial price in a non-regulated taxi market does exist, but it is limited, that phenotype (proxy for race/ethnicity) and accent (foreigner) have a significant effect on the initial offer price, and that a negative relationship exists for the phenotype (statistically significant) and a positive relationship exists for a foreigner accent (but marginally significant).

Chapter 5: Summary and Recommendations

The problem that gave rise to this study was the lack of information regarding which variables are used and how they influence the definition of the initial price offered to a customer for products or services that do not have a defined tariff. Therefore, the purpose of this research was to test the theory of price discrimination and to show that price discrimination exists by identifying the variables of discrimination taken into consideration to define an initial price and how each variable affects the initial price given. Unlike other studies, in this case, the intention was to carry out an experimental study as close as possible to the natural context of the phenomenon. The experiment considered a group of 16 people, each of whom represented a set of defined characteristics: (a) sex: female and male, (b) complexion: white and mestizo, (c) accent: Peruvian and foreign, (d) tidiness: neat and tacky, and (e) attire: formal and casual. Each of these people proceeded to ask a sample of taxi drivers to offer a rate for the same destination from the same departure point. The data, composed of the set of characteristics and the prices of the initial offers, were analyzed to answer three research questions:

1. Does discrimination in the initial price in a non-regulated taxi market exist?
2. What are the characteristics of customers that sellers consider when defining the initial price offer?
3. Is there a significant relationship between external characteristics of customers and the price initially offered to them by sellers?

The main limitations of the research were that (a) it could only be done on weekends, (b) it was only feasible to analyze the taxi market, and (c) only the variation of prices for a single journey was analyzed. This chapter contains a detailed review of (a) the conclusions, (b) the implications, and (c) the recommendations resulting from the investigation.

Conclusions

The experimental method used for this research, based on the manipulation of the characteristics of the potential customer of the taxi driver, proved a feasible and relatively simple mechanism to implement. This is of great importance because it has left evidence that it is possible to continue investigating a subject using techniques closer to the reality in which the phenomenon naturally occurs. In this case, it has been possible to measure the phenomenon in its natural environment: the street with individuals used to ask for taxi fares and expect deviations and independent taxi drivers that work without taximeters or established fares. This is important considering that, as Abdul-Muhmin (2001) indicated, much of the research in the field of price discrimination has been carried out under circumstances where price discrimination is not frequently present on a day-to-day basis. On the contrary, the great majority of studies have greater levels of internal validity than external. In addition, the productivity of the method was much higher than originally expected, obtaining a total of 3,538 observations (initial prices), more than double that expected.

As for the quality of the data collected, the high concentration of the prices collected (89% were 5 or 6 Soles) limited the analysis to be made. This result implied that the data did not meet the normality assumptions, and in order to continue with the research objective, it was necessary to transform the dependent variable from continuous to categorical, with four levels: (a) less than 5 Soles, (b) 5 Soles, (c) 6 Soles, and (d) 7 or more Soles. This result led to the question what could be happening that led to this situation and, on the other hand, what could be done to avoid it. From the point of view of design, a path of greater distance between origin and destination could have been defined in such a way as to increase the possibility of a greater dispersion in prices. Although this was contemplated when choosing the route, a route was defined as long enough for dispersion but also short enough to control external factors to the study. In particular, it was important to avoid factors such as traffic, greater number of possible routes, and the experience of the taxi driver, among others, influencing the initial price. Additionally, in countries where it is customary to use cash and credit cards are unusual, the phenomenon of the round currency exists. For example, in Peru, people usually deal with whole units of Soles and on very rare occasions use cents. Therefore, although it was possible that the initial prices could be in fractions of Soles, this did not happen; no taxi driver offered a price of 4 Soles and 30 cents or 5 Soles and 50 cents, for example. Finally, it is also possible that price discrimination is present within low dispersion margins to avoid being seen as abusive and also to avoid the fare becoming impractical, consequences that could lead to the loss of the marginal contribution that could be generated.

Regarding the results of the analysis, it was found that discrimination in the initial price in a non-regulated taxi market exists and the set of individual characteristics that could be a proxy of ethnic or racial categories seemed to show a greater influence on the initial price offer. Thus, the phenotype (physical complexion) had a statistically significant effect on price. A white customer received a price offer between 0.36 and 0.38 standard deviations higher than a mestizo customer. Similarly, the accent of the client had an influence in terms of increasing the price offer. Although the results are marginally significant, the effects of having a foreign accent ranged from 0.239 to 0.266 standard deviations. These two characteristics are the only ones with a significant effect on the initial prices.

On the other hand, other characteristics of the clients showed an effect in the expected direction but not reaching a level of significance high enough to assert that they should be considered in a price discrimination model for contexts such as the one studied. For example, demographic traits, such as sex, had a negative effect on price. A woman was more likely to receive a lower rate than a man (0.22 to 0.23 *SD*, not significant). The remaining features are related to personal image traits such as cleanliness and attire. The effect of both features was in opposite directions. A customer with a less formal (tacky) style was more likely to receive a higher price: effects range from 0.13 to 0.16 standard deviations, but are not statistically significant for any model. Finally, it was more likely that a customer wearing more informal clothes would receive a lower price offer than a client with more formal attire (clothes related to work), but in any case, the statistical significance of the coefficients was found to be under the conventional threshold (less than 90%).

Results indicate that all the characteristics of the clients evaluated are intended to be a proxy of the economic power of the person, some of them evidencing a significant relationship with the initial price offered by the seller. For some characteristics that could imply greater economic power, the direction of the relationship is in favor of higher prices, and for characteristics that tend to imply lower economic power, the direction is in favor of lower prices.

Implications

From a cultural perspective, the price discrimination studied is evidenced as a natural phenomenon in economic transactions between people. This process of discrimination has existed for a long time (Pigou, 1920, Serrano, 1947) and is part of the empirical learning passed from one generation to another, which suggests the existence of favorable factors in its existence. This is of great importance and reinforces that price discrimination in the field of commercial transactions should be studied further, before it disappears, given the current trend where discrimination is seen as an abusive method, existing with the main purpose that bidders can take advantage of those demanding their products or services. This situation is magnified in a context where, with the purpose of providing facilities for the control and management of business processes, the use of fixed or tariff prices is increased as an alternative to pricing through empirical discrimination policies where prices are perceived as subjective and less auditable.

From a social perspective, price discrimination plays a very important economic role. Discrimination of the initial price seems to comply with a principle of social justice, suggesting that those who are able to pay more for a good or service do so in favor of those who cannot pay as much, achieving a balance between both parties and making possible the continuation of the system. As a result of the investigation, factors related to the purchasing power of people emerged as the ones with the greatest influence on the initial price that the drivers suggested. Foreigners, people with a European phenotype, or those with a foreign accent, all variables that in a context such as Peru are directly related to greater purchasing power (Quijano, 2007), usually received initial offers of a higher price. In this context, the idea that bidders take advantage of the price discrimination mechanism to abuse their position in order to obtain higher income is not entirely correct, given that the use of their position exists but, at the same time, allows balancing an economic system in which the respect of the demand is just as important as the subsistence of the offer. That is why there is natural discrimination with the purpose that the surplus achieved with a public can cover the deficit maintained with another public, achieving convenient results so that the offer can continue to exist, therefore fulfilling its role in the satisfaction of needs present in the demand.

From an academic perspective, price discrimination is a poorly studied scheme compared to what corresponds to the management of fixed prices, which contradicts the rationale that the offer must adapt to the customer and not vice versa. The fixed price or tariff schemes are based largely on the logic of the company towards the customer. In many cases, prices are fixed based on variables linked to production costs and, in the best case, framed within economic accessibility ranges. On the other hand, in discriminated price schemes, prices are based on the characteristics of the client while the costs and the contribution to fixed costs only serve as an indicator of the lower limit to be respected for the viability of the transaction.

Additionally, from the academic point of view related to the research carried out within the framework of price discrimination, many writers focus on understanding the illegitimacy of the model, delving into the analysis of the injustice perceived by the consumer. Few have investigated the procedure and policies applied by the merchants to implement this pricing scheme that, as mentioned above, can represent one of the best alternatives for the survival of the offer, through the implementation of a cross subsidy scheme managed by the bidder: price discrimination. This leads to a proposal that, within the framework of price theory linked to discrimination, there should be a subdivision as regards the existing third degree discrimination types: on one hand, there should be a positive, fair, or pro-market discrimination, corresponding to this positive mechanism based on the cross-subsidy implemented through the definition of prices that may vary according to the economic possibilities of the customer. On the other hand, there should be a negative, abusive, or anti-market discrimination, corresponding to mechanisms of price variation based on negative, hurtful factors, such as the establishment of prices based on gender, sexual orientation, or race, which have no purpose aligned with the improvement of commercial relationships leading to an improvement in the general quality of life.

Finally, from a business management perspective, it is essential to evaluate the design and implementation of price discrimination mechanisms for business models as an opportunity to achieve part of the growth sought, identifying in this mechanism the tool that would allow them to enter into new markets. These could be markets with an insufficient number of potential clients capable of affording the actual standard price. In situations such as cities with important differences in the purchasing power of their population, a scheme based on price discrimination would allow having an offer that suits or approaches the ability to pay of a larger number of people. This opportunity is only viable if the implementation of price discrimination mechanisms is accompanied by education and awareness campaigns that price discrimination could be favorable, protecting the seller from any reputational crisis due to the actual perception that price discrimination is an abusive mechanism. Discrimination schemes must support both supply and demand so that the offer can reach a greater number of people requesting the product or service necessary to cover their needs and therefore increase their quality of life.

This investigation has shown that, in the non-regulated taxi market, price discrimination exists and occurs naturally in commercial relationships that are not framed by an offer with fixed prices or subject to tariffs and that this discrimination is modeled according to characteristics of those making the demand that seem to be linked to their economic capacity. All these elements are of great value in the framework of continuing to enrich the knowledge of the community of scientists and management professionals. This research opens the way to continue understanding the mechanism of price discrimination in other services or products and to continue to study them in other locations with wide differences in the purchasing power of their population. Researchers could seek a greater understanding of which physical characteristics of those who demand a service influence the price and of the magnitude of their influence, as well as whether it changes according to the service, product, or location. Such investigations would certainly reinforce the idea that price discrimination is favorable for the economic development of companies and for the contribution of a greater number of products and services to currently inaccessible populations. Finally, in order to improve the image of price discrimination in the framework of better business management, it would be relevant to coin a specific concept for price discrimination in favor of a market based on fair relationships.

Recommendations

The research carried out for this study represents an additional step in a subject of study that could be useful in the near future. Although price discrimination has been limited or prohibited with the aim of improving commercial management, price discrimination can be very useful if it is well understood and carefully applied using the correct variables and affecting the price to the correct extent.

Furthermore, this research shows that price discrimination seems to be unconsciously accepted as fair, or at least its marginal impact on prices does not seem to merit an effort by the population to fight against it; rather it could be seen as an egalitarian mechanism of opportunity. It is for this reason that what has been identified in the market of taxi drivers in Lima should be studied for more categories and in a greater number of locations with the purpose of constructing more robust, practical, effective, and accepted models of price discrimination. Further studies, such as: a) to study the characteristics of the offer (product or service) needed for a feasible and beneficial price discrimination. b) To study the characteristics of the demand (target market) needed for a feasible and beneficial price discrimination. c) To study the characteristics of the distribution channels to identify in which of them price discrimination is feasible and beneficial. d) To study which rules (code of ethics) should be respected in a price discrimination mechanism so that it is perceived as positive and has consequences in favor of market development. e) To study what should be the characteristics that must be met in advertisement, so that the promotion of products or services sold with discriminated prices is within the framework of the law, is economically feasible for companies and does not generate a rejection from the target market. Finally, f) to study what should be the internal characteristics of a company so that it is feasible to implement a mechanism of beneficial price discrimination. Future steps in this field of research could be very helpful to an increasing number of companies that are constantly accumulating more information about their customers with the clear purpose of using it to achieve higher sales by adapting their offerings to them. For example, it could especially be helpful for data scientists in the construction of models for loyalty programs with discounts or models for e-commerce with discriminated price offers.

In particular, this research confirms that price discrimination is real and leaves for further research the understanding of why it exists. A plausible explanation is that it is a necessary mechanism of cross-subsidy between customers, managed by the supplier, providing marginally greater access to the demand, and allowing greater volumes of sales for the supplier. With regard to this last point, it would be useful to understand further the direct and indirect, conscious and unconscious benefits of this differentiated pricing mechanism in order to understand the factors that need to be taken into consideration for the acceptance of a price discrimination model based on the demand and the supply.

References

Abdul-Muhmin, A. G. (2001). The effect of perceived seller reservation prices on buyers' bargaining behavior in a flexible-price market. *Journal of International Consumer Marketing*, *13*(3), 29-45. doi:10.1300/J046v13n03_03

Aguirre, K. (2008, November 15). Informe.21: Del regateo al taxímetro [Report.21: From haggling to the taximeter]. Retrieved from http://peru21.pe/noticia/221324/informe21-regateo-al-taximetro

Alt, R. M. (1949). The internal organization of the firm and price formation: An illustrative case. *Quarterly Journal of Economics*, *63*(1), 92-110. doi:10.2307/1882735

Anderson, P. L., McLellan, R. D., Overton, J. P., & Wolfram, G. L. (1997). Price elasticity of demand. *McKinac Center for Public Policy. Accessed October*, *13*, 2010.

Armstrong, M. (2005). *Recent developments in the economics of price discrimination*. Retrieved from https://www.microfinancegateway.org/organization/economics-working-paper-archive-econwpa

Bolton, L. E., & Alba, J. W. (2006). Price fairness: Good and service differences and the role of vendor costs. *Journal of Consumer Research, 33*(2), 258-265. doi:10.1086/506306

Bolton, L. E., Keh, H., & Alba, J. W. (2010). How do price fairness perceptions differ across culture? *Journal of Marketing Research*, *47*(3), 564-576. 10). doi:10.1509/jmkr.47.3.564

Bolton, L. E., Warlop, L., & Alba, J. W. (2003). Explorations in price (un)fairness. *Journal of Consumer Research, 21*, 393-407.

Campbell, D. T., & Stanley, J. C. (1963). Experimental designs for research on teaching. *Handbook of research on teaching*, 171-246.

Chen, L. Z., Hu, W. M., Szulga, R., & Zhou, X. (2018). Demographics, gender and local knowledge—Price discrimination in China's car market. *Economics Letters*, *163*, 172-174.

Chintagunta, Pradeep K. (2002), Investigating Category Pricing Behavior at a Retail Chain, *Journal of Marketing Research*, 39 (May), 141–54.

Clay, K., Krishnan, R., Wolff, E., & Fernandes, D. (2002). Retail strategies on the web: Price and non-price competition in the online book industry. *The Journal of Industrial Economics, 50*(3), 351-367. doi:10.1111/1467-6451.00181

Colchero, M. A., Salgado, J. C., Unar-Munguia, M., Hernandez-Avila, M., & Rivera-Dommarco, J. A. (2015). Price elasticity of the demand for sugar sweetened beverages and soft drinks in Mexico. *Economics & Human Biology*, *19*, 129-137.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis for field setting*. Boston, MA: Houghton Mifflin.

Corfman, K. P., & Lehmann, D. R. (1993). The importance of others' welfare in evaluating bargaining outcomes. *Journal of Consumer Research*, *20*(1), 124-137. doi:10.1086/209338

Cowan, S. (2016). Welfare-increasing third-degree price discrimination. *The RAND Journal of Economics*, *47*(2), 326-340.

Cressman, G. E., Jr. (2006). Fixing prices. *Marketing Management, 15*(5), 33-37.

Dibb, S., & Simkin, L. (2009). Implementation rules to bridge the theory/practice divide in market segmentation. *Journal of Marketing Management, 25*(3/4), 375-396. doi:10.1362/026725709X429809

Dickson, P. R., & Kalapurakal, R. (1994). The use and perceived fairness of price-setting rules in the bulk electricity market. *Journal of Economic Psychology, 15*(3), 427-448. doi:10.1016/0167-4870(94)90023-X

Dickson, P. R., & Sawyer, A. G. (1990). The price knowledge and search of supermarket shoppers. *The Journal of Marketing, 54*(3), 42-53. doi:10.2307/1251815

Fabra, N. (2018). *A Model of Search with Price Discrimination* (No. 12823). CEPR Discussion Papers.

Federación de Taxis del Perú. (2011, September 20). Mañana 21 - 09 - 22 se Inicia la mesa de trabajo para formalizar el servicio de taxi [Tomorrow 21 - 09 - 22 begins the work-table to formalize the taxi service]. Retrieved from http://fentacperu.blogspot.com/search?q=taxis

Ferguson, J. L. (2014). Implementing price increases in turbulent economies: Pricing approaches for reducing perceptions of price unfairness. *Journal of Business Research, 67*(1), 2732-2737. doi:10.1016/j.jbusres.2013.03.023

Frank, R. H. (2010). *Microeconomics and behavior* (8th ed.). New York, NY: McGraw-Hill Irwin.

Goodman, L. (2003). A rotten deal. *Self*, 34-7.

Graddy, K. (1995). Testing for imperfect competition at the Fulton Fish Market. *RAND Journal of Economics, 26*, 75-92. doi:10.2307/2556036

Gregson, A. (2008). *Pricing strategies for small business*. Vancouver, Canada: International Self-Counsel Press.

Grether, E. T. (1941). Current trends affecting pricing policies. *Journal of Marketing, 5*(3), 222-223. doi:10.2307/1246662

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology, 26*(1), 441-62. doi:10.1146/annurev.soc.26.1.441

Hair, J. F., Black, W. C., & Babin, B. J. (2010). RE Anderson Multivariate data analysis: a global perspective. *New Jersey. Pearson. Ed*, *7*, 816.

Haws, K. L., & Bearden, W. O. (2006). Dynamic pricing and consumer fairness perceptions. *Journal of Consumer Research, 33*(3), 304-311. doi:10.1086/508435

Heyman, J. E., & Mellers, B. A. (2008). Perceptions of fair pricing. In C. P. Haugtvedt, P. M. Herr, & R. K. Frank (Eds.), *Handbook of consumer psychology* (pp. 683-697). New York, NY: Lawrence Erlbaum.

Kassaye, W. W. (1990). The role of haggling in marketing: An examination of buyer behavior. *Journal of Consumer Marketing, 7*(4), 53-62. doi:10.1108/EUM0000000002589

Kehoe, K. (2004). Make your price fit. *Landscape Management, 43*(10), 46.

Kimes, S. E., & Wirtz, J. (2002). Perceived fairness of demand-based pricing for restaurants. *Cornell Hospitality Quarterly, 43*(1), 31. doi:10.1016/S0010-8804(02)80006-4

Kotler, P., & Armstrong, G. (2008). *Princípios de marketing* [Principles of marketing]. Madrid, Spain: Pearson Educación.

Lewis, M. (2005). Research note: A dynamic programming approach to customer relationship pricing. *Management Science*, *51*(6), 986-994. doi:10.1287/mnsc.1050.0373

Lii, J. (1995, November 29). Chinatown, where are cheaper. *New York Times*, Sec. C, p. 1.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological), 42*(2), 109-142.

Monroe, K. B., & Lee, A. Y. (1999). Remembering versus knowing: Issues in buyers' processing of price information. *Journal of the Academy of Marketing Science*, *27*(2), 207-225. doi:10.1177/0092070399272006

Montgomery, Alan L. (1997), Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data, *Marketing Science*, 16 (4), 315–37.

Namata, B., Ostaszewski, K., & Sahoo, P. K. (1990). Direction of price changes in third-degree price discrimination. *The American Economic Review, 80*(5), 516-524.

Ochs, J., & Roth, A. E. (1989). An experimental study of sequential bargaining. *The American Economic Review, 79*(3), 355-384.

Peppers, D. (1993). *The one to one future: Building relationship one customer at a time*. New York, NY: Doubleday.

Phillips, C. F. (1946). Major areas for marketing research. *Journal of Marketing*, *11*(1), 21-26. doi:10.2307/1246802

Pigou, A. C. (1920). *The economics of welfare*. London, UK: Macmillan.

Quijano, A. (2007). Colonialidad del poder y clasificación social. In S. Castro-Gómez & R. Grosfoguel (Eds.), *El giro decolonial: Reflexiones para una diversidad epistémica más allá del capitalismo global* (pp. 93-126). Retrieved from http://www.unsa.edu.ar/histocat/hamoderna/grosfoguelcastrogomez.pdf

Ramasastry, A. (2005, June 24). Web sites change prices based on customers' habits. Retrieved from http://www.cnn.com/2005/LAW/06/24/ramasastry.website.prices

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605. doi:10.1016/0010-0285(75)90024-9

Rust, R. T., & Verhoef, P. C. (2005). Optimizing the marketing interventions mix in intermediate-term CRM. *Marketing Science*, *24*(3), 477-489. doi:10.1287/mksc.1040.0107

Samuelson, W. F., & Marks, S. G. (2008). *Managerial economics*. New York, NY: John Wiley & Sons.

Serrano, J. L. (1947). La discriminacion de precios [Price discrimination]. *Boletín de Estudios Económicos, 2*(11), 115-132.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles, CA: Sage.

Sullivan, G. D. (1969). Some thoughts on bargaining. *American Journal of Agricultural Economics, 51*(4), 960-961. doi:10.2307/1237802

Tellis, G. J. (1986). Beyond the many faces of price: An integration of pricing strategies. *The Journal of Marketing, 50*(4), 146-160. doi:10.2307/1251292

Turow, J., Feldman, L., & Meltzer, K. (2005). Open to exploitation: America's shoppers online and offline. A report from the Annenberg Public Policy Center of the University of Pennsylvania. Retrieved from http://repository.upenn.edu/asc_papers/35

Varble, D. (1980). Flexible price agreements: Purchasing's view. *California Management Review*, *23*(2), 44-51. doi:10.2307/41164916

Walker, Q. (1950). Some principles of department store pricing. *Journal of Marketing, 14*(4), 529-537. doi:10.2307/1247582

Weinstein, A. (2004). *Handbook of market segmentation: Strategic targeting for business and technology firms*. New York, NY: The Haworth Press.

Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research and Evaluation*, *18*(11), 1-14.

Winship, C., & Mare, R. (1984). Regression models with ordinal variables. *American Sociological Review, 49*(4), 515-525. doi:10.2307/2095465

Wirtz, J., & Kimes, S. E. (2007). The moderating role of familiarity in fairness perceptions of revenue management pricing. *Journal of Service Research, 9*(3), 229-240. doi:10.1177/1094670506295848

Xie, J., & Shugan, S. M. (2001). Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Science*, *20*(3), 219-243. doi:10.1287/mksc.20.3.219.9765

Zhang, J., & Krishnamurthi, L. (2004). Customizing promotions in online stores. *Marketing Science*, *23*(4), 561-578. doi:10.1287/mksc.1040.0055

Zwick, R., & Chen, X. P. (1999). What price fairness? A bargaining study. *Management Science*, *45*(6), 804-823.

Appendix A: Interview Guide

Guía de Entrevistas

Taxistas

I. Introducción

Buenos días, mi nombre es ____________ y estoy participando del recojo de información para un trabajo de investigación doctoral. Esta investigación es sobre los hábitos y rutas de taxis, por lo que le pediría que pudiera brindarme 15 minutos de su tiempo para realizar una pequeña entrevista. Toda la información que nos dé va a ser estrictamente confidencial. Recuerde que no hay respuestas correctas ni incorrectas.

II. Hábitos y rutinas

1. **Presentación:** ¿Cómo es su día a día, su rutina diaria? ¿En qué horarios suele trabajar? ¿Difiere en los días normales y los fines de semana?
2. **Rutas usuales:** Ahora me gustaría saber ¿suele tener rutas usuales en su trabajo? *(E: indagar si va a todos los destinos o solo circula por ciertos distritos)*.

III. Variaciones en el precio

3. **Horarios:** Y cuénteme, ¿usted suele variar sus tarifas en base a algún indicador? Por ejemplo, respecto a los horarios. ¿En qué horarios considera que las tarifas deben incrementarse? ¿A qué se debe esto? Y, por el contrario, ¿en qué horarios considera que las tarifas deben ser menores? ¿A qué se debe esto?
4. **Punto de partida y destino:** Y así como hay horarios en los que las tarifas se incrementan, ¿pasa lo mismo con las rutas? ¿cómo así? *(E: indagar si el punto de origen o de destino impactan en la fijación de precios, p.e. taxistas que incrementan su tarifa porque no desean ir a ciertos lugares)*.
5. **Características del pasajero:** ¿Usted considera que hay taxistas que fijan sus precios en base a las características de los pasajeros? ¿Por qué cree que se da esto? Y, ¿cuáles son las características que hacen que le cobren más a unos que a otros? *(E: enumerar las características una a una y profundizar en cuáles incrementan el precio y cuáles mantienen/bajan el precio, p.e. Sexo, Edad, Vestimenta, Tono de piel)*.
6. **Otras categorías:** Ahora que ya hemos determinado que en esta categoría "taxis" no hay precios fijos, sino que se van definiendo en base a varios factores, ¿considera que existen otras categorías que tampoco tienen precios fijos? ¿Cuáles? ¿Cómo así?

Muchas gracias

Appendix B: Descriptive statistics and distribution of dependent and predictor variables

Descriptive Statistics: Price Offer (dependent variable)

```
.  fre precio

precio -- PRECIO
-----------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+--------------------------------------------
Valid   3     |          3       0.08       0.08       0.08
        4     |         87       2.46       2.46       2.54
        5     |       1955      55.27      55.27      57.82
        6     |       1187      33.56      33.56      91.38
        7     |        247       6.98       6.98      98.36
        8     |         54       1.53       1.53      99.89
        10    |          4       0.11       0.11     100.00
        Total |       3537     100.00     100.00
-----------------------------------------------------------

. fsum precio, stat(mean sd min max p50)

 Variable |        N     Mean       SD   Median      Min       Max
----------+-------------------------------------------------------
   precio |     3537     5.50     0.75     5.00     3.00     10.00
```

Descriptive Statistics by predictors: client characteristics

```
.  foreach var of varlist c_fsexo c_fedad c_ftez c_facento c_fimagen c_fvestimenta {
.         table `var' , c(mean precio sd precio n precio ) format(%7.2f)
. }

----------------------------------------------------
       Sex | mean(precio)    sd(precio)     N(precio)
----------+-----------------------------------------
      Male |         5.54          0.72         1,660
    Female |         5.47          0.77         1,877
----------------------------------------------------

----------------------------------------------------
       Age | mean(precio)    sd(precio)     N(precio)
----------+-----------------------------------------
  Under 19 |         5.57          0.80         1,886
     19-64 |         5.41          0.72           791
       65+ |         5.42          0.63           860
----------------------------------------------------

----------------------------------------------------
Phenotype | mean(precio)    sd(precio)     N(precio)
----------+-----------------------------------------
   Mestizo |         5.41          0.67         1,911
     White |         5.61          0.82         1,626
----------------------------------------------------

----------------------------------------------------
    Accent | mean(precio)    sd(precio)     N(precio)
----------+-----------------------------------------
  Peruvian |         5.46          0.74         2,015
 Foreigner |         5.56          0.75         1,522
----------------------------------------------------

----------------------------------------------------
Appearanc |
e         | mean(precio)    sd(precio)     N(precio)
----------+-----------------------------------------
      Neat |         5.45          0.66         1,693
     Tacky |         5.55          0.82         1,844
----------------------------------------------------

----------------------------------------------------
```

```
   Attire | mean(precio)     sd(precio)      N(precio)
----------+----------------------------------------
   Formal |         5.52           0.80          1,691
   Casual |         5.48           0.70          1,846
-----------------------------------------------------
```

Descriptive Statistics by predictors: vehicle characteristics

```
.  fre taxi_color taxi_marca taxi_anho

taxi_color -- Vehicle color
---------------------------------------------------------------------
                        |      Freq.    Percent      Valid       Cum.
------------------------+--------------------------------------------
Valid   1  Yellow       |        311       8.79       8.79       8.79
        3  Blue         |        165       4.66       4.66      13.46
        4  Beige        |         74       2.09       2.09      15.55
        5  White        |        710      20.07      20.07      35.62
        8  Gray         |        395      11.17      11.17      46.79
        12 Black        |        644      18.21      18.21      65.00
        13 Silver       |        545      15.41      15.41      80.41
        15 Red          |        268       7.58       7.58      87.98
        17 Green        |         79       2.23       2.23      90.22
        88 Other color  |         81       2.29       2.29      92.51
        90 No data      |        265       7.49       7.49     100.00
        Total           |       3537     100.00     100.00
---------------------------------------------------------------------

taxi_marca -- Vehicle brand
---------------------------------------------------------------------
                        |      Freq.    Percent      Valid       Cum.
------------------------+--------------------------------------------
Valid   3  BYD          |         43       1.22       1.22       1.22
        4  CHEVROLET    |        252       7.12       7.12       8.34
        5  DAEWOO       |         56       1.58       1.58       9.92
        11 HONDA        |         29       0.82       0.82      10.74
        12 HYUNDAI      |        327       9.25       9.25      19.99
        15 KIA          |        333       9.41       9.41      29.40
        17 MAZDA        |         42       1.19       1.19      30.59
        19 MITSUBISHI   |         49       1.39       1.39      31.98
        20 NISSAN       |        914      25.84      25.84      57.82
        21 RENAULT      |         22       0.62       0.62      58.44
        23 SUZUKI       |         37       1.05       1.05      59.49
        24 TOYOTA       |        853      24.12      24.12      83.60
        25 VOLKSWAGEN   |         97       2.74       2.74      86.34
        88 Other brand  |        224       6.33       6.33      92.68
        90 No data      |        259       7.32       7.32     100.00
        Total           |       3537     100.00     100.00
---------------------------------------------------------------------

taxi_anho -- Vehicle year
-------------------------------------------------------------------
                      |      Freq.    Percent      Valid       Cum.
----------------------+--------------------------------------------
Valid   0    No data  |       1694      47.89      47.89      47.89
        2010 2010     |        146       4.13       4.13      52.02
        2011 2011     |        309       8.74       8.74      60.76
        2012 2012     |        272       7.69       7.69      68.45
        2013 2013     |        278       7.86       7.86      76.31
        2014 2014     |        246       6.96       6.96      83.26
        2015 2015     |        307       8.68       8.68      91.94
        2016 2016     |        285       8.06       8.06     100.00
        Total         |       3537     100.00     100.00
-------------------------------------------------------------------
```

Descriptive Statistics by predictors: data collection characteristics

```
.  fre c_dÍadeobservac c_horario c_horariogr
```

c_dÍadeobservac -- Day of data collection

		Freq.	Percent	Valid	Cum.
Valid	1 Saturday	2045	57.82	57.82	57.82
	2 Sunday	1492	42.18	42.18	100.00
	Total	3537	100.00	100.00	

c_horario

		Freq.	Percent	Valid	Cum.
Valid	101 08:00 - 12:00	734	20.75	20.75	20.75
	102 08:15 - 12:00	192	5.43	5.43	26.18
	103 08:15 - 12:15	184	5.20	5.20	31.38
	201 12:00 - 16:00	911	25.76	25.76	57.14
	202 12:30 - 16:30	120	3.39	3.39	60.53
	203 14:00 - 16:00	120	3.39	3.39	63.92
	301 16:00 - 20:00	1105	31.24	31.24	95.17
	302 16:30 - 20:00	120	3.39	3.39	98.56
	303 16:50 - 20:00	3	0.08	0.08	98.64
	304 15:00 - 20:00	48	1.36	1.36	100.00
	Total	3537	100.00	100.00	

c_horariogr -- Shift of data collection

		Freq.	Percent	Valid	Cum.
Valid	100 Morning (08-12)	1110	31.38	31.38	31.38
	200 Mid-afternoon (12-16)	1151	32.54	32.54	63.92
	300 Afternoon (16-20)	1276	36.08	36.08	100.00
	Total	3537	100.00	100.00	

Appendix C: Descriptive statistics of prices against independent (predictors) variables

Characteristics of the client

Sex

Sex	mean(precio)	sd(precio)	N(precio)
Male	5.54	0.72	1,660
Female	5.47	0.77	1,877

Test of differences: Male vs Female

	diff.	
precio	0.0696**	(2.77)
N	3537	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Age	mean(precio)	sd(precio)	N(precio)
Under 19	5.57	0.80	1,886
19-64	5.41	0.72	791
65+	5.42	0.63	860

Age

Test of differences: Under 19 vs 19-64

	diff.	
precio	0.158***	(4.80)
N	2677	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Test of differences: 19-64 vs 65+

	diff.	
precio	-0.00627	(-0.19)
N	1651	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Test of differences: Under 19 vs 65+

	diff.	

precio	0.152***	(4.91)
N	2746	

t statistics in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Phenotype

Phenotype	mean(precio)	sd(precio)	N(precio)
Mestizo	5.41	0.67	1,911
White	5.61	0.82	1,626

Test of differences: Mestizo vs White

	diff.	
precio	-0.198***	(-7.94)
N	3537	

t statistics in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Accent

Accent	mean(precio)	sd(precio)	N(precio)
Peruvian	5.46	0.74	2,015
Foreigner	5.56	0.75	1,522

Test of differences: Peruvian vs Foreigner

	diff.	
precio	-0.103***	(-4.07)
N	3537	

t statistics in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Appareance

Appearanc e	mean(precio)	sd(precio)	N(precio)
Neat	5.45	0.66	1,693
Tacky	5.55	0.82	1,844

Test of differences: Neat vs Tacky

	diff.	
precio	-0.0932***	(-3.71)
N	3537	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Attire

Attire	mean(precio)	sd(precio)	N(precio)
Formal	5.52	0.80	1,691
Casual	5.48	0.70	1,846

Test of differences: Formal vs Casual

	diff.	
precio	0.0462	(1.84)
N	3537	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Vehicle characteristics

Vehicle color

Vehicle color	mean(precio)	sd(precio)	N(precio)
Yellow	5.52	0.78	311
Blue	5.47	0.70	165
Beige	5.49	0.62	74
White	5.46	0.70	710
Gray	5.49	0.72	395
Black	5.50	0.77	644
Silver	5.54	0.77	545
Red	5.52	0.70	268
Green	5.49	0.83	79
Other color	5.53	0.71	81
No data	5.52	0.84	265

Vehicle brand

Vehicle brand	mean(precio)	sd(precio)	N(precio)
BYD	5.37	0.66	43
CHEVROLET	5.49	0.75	252
DAEWOO	5.43	0.68	56
HONDA	5.55	0.78	29
HYUNDAI	5.46	0.77	327
KIA	5.57	0.81	333
MAZDA	5.45	0.67	42
MITSUBISHI	5.47	0.62	49
NISSAN	5.48	0.71	914
RENAULT	5.50	0.96	22
SUZUKI	5.49	0.77	37
TOYOTA	5.53	0.76	853
VOLKSWAGEN	5.54	0.68	97
Other brand	5.47	0.68	224
No data	5.52	0.85	259

Vehicle year

Vehicle year	mean(precio)	sd(precio)	N(precio)
No data	5.48	0.75	1,694
2010	5.49	0.68	146
2011	5.48	0.69	309
2012	5.51	0.77	272
2013	5.48	0.73	278
2014	5.55	0.73	246
2015	5.50	0.75	307
2016	5.59	0.84	285

Data collection characteristics

Day of data collection

Day of data collectio n	mean(precio)	sd(precio)	N(precio)
Saturday	5.54	0.79	2,045
Sunday	5.45	0.67	1,492

Day of data collectio n	mean(precio)	sd(precio)	N(precio)
Saturday	5.54	0.79	2,045
Sunday	5.45	0.67	1,492

Test of differences: Saturday vs Sunday

	diff.	
precio	0.0865***	(3.41)
N	3537	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Shift of data collection

c_horario	mean(precio)	sd(precio)	N(precio)
08:00 - 12:00	5.52	0.79	734
08:15 - 12:00	5.65	0.87	192
08:15 - 12:15	5.22	0.52	184
12:00 - 16:00	5.53	0.81	911
12:30 - 16:30	5.47	0.62	120
14:00 - 16:00	5.57	0.78	120
16:00 - 20:00	5.54	0.70	1,105
16:30 - 20:00	5.10	0.30	120
16:50 - 20:00	5.00	0.00	3
15:00 - 20:00	5.21	0.46	48

Shift of data collection	mean(precio)	sd(precio)	N(precio)
Morning (08-12)	5.49	0.78	1,110
Mid-afternoon (12-16)	5.53	0.79	1,151
Afternoon (16-20)	5.49	0.68	1,276

Shift of data collection	mean(precio)	sd(precio)	N(precio)
Morning (08-12)	5.49	0.78	1,110
Mid-afternoon (12-16)	5.53	0.79	1,151
Afternoon (16-20)	5.49	0.68	1,276

Test of differences: Morning (08-12) vs Mid-afternoon (12-16)

	diff.	
precio	-0.0364	(-1.11)
N	2261	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Test of differences: Mid-afternoon (12-16) vs Afternoon (16-20)

	diff.	
precio	0.0382	(1.28)
N	2427	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

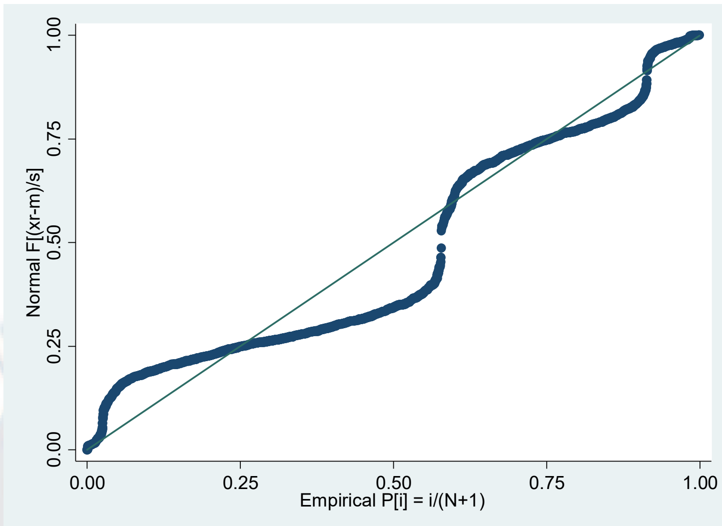
Test of differences: Morning (08-12) vs Afternoon (16-20)

	diff.	
precio	0.00173	(0.06)
N	2386	

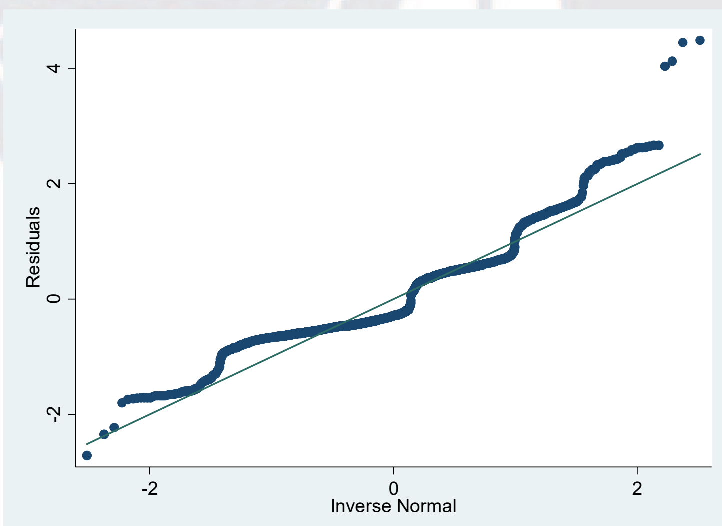
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Appendix D: Normality of residuals after applying a OLS regression model with clustered standard errors

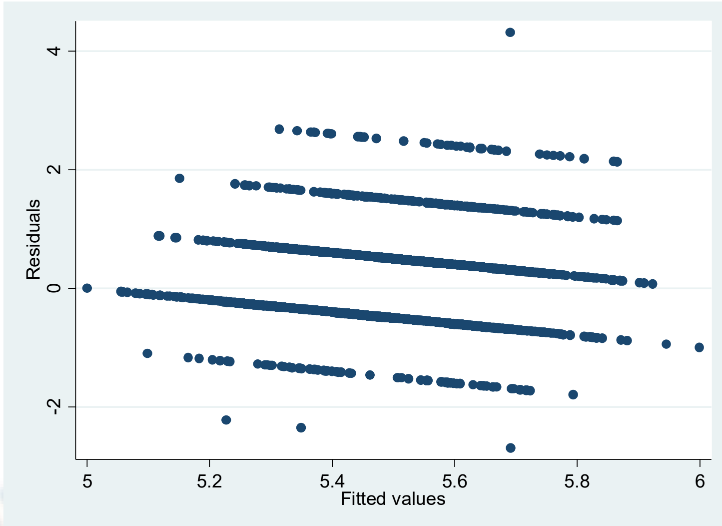
Standardized normal probability plot of residuals



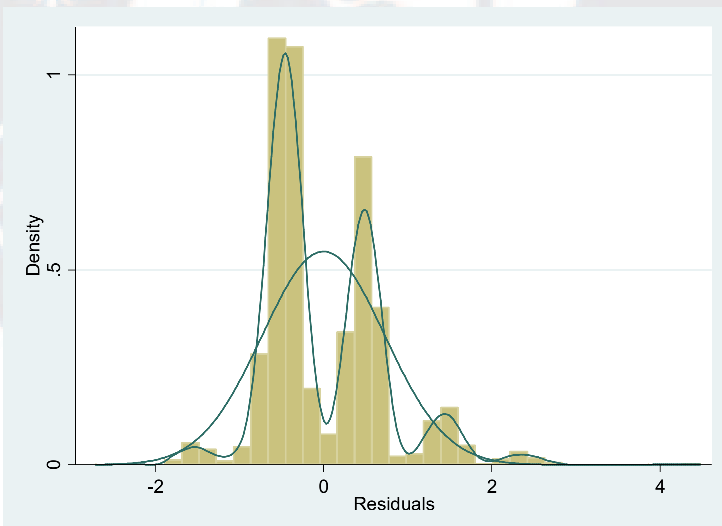
Quantiles of residuals against quantiles of normal distribution



Residuals against fitted values



Histogram, density distribution and normal distribution of residuals



Appendix E: Exploratory linear regression analysis

Regression of price of taxi ride controlling by surveyor characteristics, time of data collection, and characteristics of the vehicle

	Null Model (1)	OLS (2)	OLS Robust (3)	OLS Clustered (4)
Saturday		0	0	0
		(.)	(.)	(.)
Sunday		-0.0634*	-0.0634*	-0.0634
		(-2.32)	(-2.39)	(-0.96)
Morning (08-12)		0	0	0
		(.)	(.)	(.)
Mid-afternoon (1~16)		0.0277	0.0277	0.0277
		(0.82)	(0.79)	(0.37)
Afternoon (16-20)		0.00284	0.00284	0.00284
		(0.09)	(0.09)	(0.03)
Male		0	0	0
		(.)	(.)	(.)
Female		-0.106***	-0.106***	-0.106
		(-4.05)	(-4.01)	(-1.51)
Under 19		0	0	0
		(.)	(.)	(.)
19-64		-0.170***	-0.170***	-0.170
		(-5.23)	(-5.37)	(-1.84)
65+		-0.134***	-0.134***	-0.134
		(-4.33)	(-4.77)	(-2.06)
Mestizo		0	0	0
		(.)	(.)	(.)
White		0.172***	0.172***	0.172*
		(6.60)	(6.72)	(2.16)
Peruvian		0	0	0
		(.)	(.)	(.)
Foreigner		0.131***	0.131***	0.131
		(5.03)	(5.15)	(2.10)
Neat		0	0	0
		(.)	(.)	(.)
Tacky		0.0883***	0.0883***	0.0883
		(3.37)	(3.54)	(1.29)
Formal		0	0	0
		(.)	(.)	(.)
Casual		-0.0513*	-0.0513*	-0.0513
		(-1.98)	(-2.09)	(-0.75)
Constant	5.500***	5.408***	5.408***	5.408***
	(437.79)	(41.60)	(44.96)	(67.74)
N	3537	3537	3537	3537
r2_a	0	0.0371	0.0371	0.0371

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Appendix F: Unstandardized effects of the client's characteristics on the price offered

	Model 1		Model 2		Model 3	
	Direct effect (no controls)		Effect controlling by level 1 characteristics		Effect controlling by level 1 and level 2 characteristics	
	(b)	(t)	(b)	(t)	(b)	(t)
Panel A: Sex						
Male (reference category)						
Female	-0.310	(-1.13)	-0.326	(-1.16)	-0.324	(-1.60)
Panel B: Age						
Under 19 (reference category)						
19-64	-0.461	(-1.40)	-0.467	(-1.38)	**-0.475**	**(-1.90)**
65+	-0.259	(-0.79)	-0.254	(-0.75)	-0.263	(-1.06)
Panel C: Complexion						
Mestizo (reference category)						
White	**0.524***	**(2.08)**	**0.509+**	**(1.93)**	**0.499***	**(2.46)**
Panel D: Accent						
Peruvian (reference category)						
Foreigner	0.332	(1.22)	0.365	(1.31)	**0.369+**	**(1.82)**
Panel E: Tidyness						
Neat (reference category)						
Tacky	0.181	(0.64)	0.221	(0.77)	0.220	(1.08)
Panel F: Attire						
Formal (reference category)						
Casual	-0.165	(-0.59)	-0.132	(-0.46)	-0.130	(-0.64)
Controls						
Level 1 (characteristics of the vehicle and data collection)	No		Yes		Yes	
Level 2 (client characteristics)	No		No		Yes	
Observations	3537		3537		3537	

Notes: Dependent variable is the ordinal categorization of price of the ride. Coefficients estimated using a mixed effect ordinal logistic regression. Effects expressed in logit units. T statistics reported in parentheses

Significance levels: + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Appendix G: Results of multilevel ordinal logistic regressions using variables only at level 1

Estimation using level 1 predictors

Characteristics of the taxi/vehicle

*color of the vehicle: taxi_color

*brand of the vehicle: taxi_marca

*year of the vehicle //notice ~45% missing year: taxi_anho

Characteristics of date and time of observation

*day of observation (2 days): c_dĺadeobservac

*time (recoded into 3 shifts) of observation: c_horariogr

Level 1 predictors, bivariate models

Level 1 predictor variable: taxi_color Vehicle color

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3530.7636
Iteration 2:   log likelihood = -3530.7632

Refining starting values:

Grid node 0:   log likelihood =  -3432.368

Fitting full model:

Iteration 0:   log likelihood =  -3432.368  (not concave)
Iteration 1:   log likelihood = -3430.0002  (not concave)
Iteration 2:   log likelihood =  -3428.328
Iteration 3:   log likelihood = -3425.3472
Iteration 4:   log likelihood = -3425.3369
Iteration 5:   log likelihood = -3425.3369

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(10)     =       4.58
Log likelihood = -3425.3369                     Prob > chi2       =     0.9176
------------------------------------------------------------------------------
```

```
   precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
              |
   taxi_color |
         Blue |  -.0844015   .1915236    -0.44   0.659    -.4597809    .2909779
        Beige |  -.0155895   .2545144    -0.06   0.951    -.5144285    .4832494
        White |  -.1325203   .1362858    -0.97   0.331    -.3996356    .1345951
         Gray |  -.0708287   .1513507    -0.47   0.640    -.3674706    .2258132
        Black |  -.0116134   .1383612    -0.08   0.933    -.2827965    .2595696
       Silver |   .0634761   .1414741     0.45   0.654    -.2138082    .3407603
          Red |  -.0065184   .1654141    -0.04   0.969    -.3307241    .3176874
        Green |  -.0826679   .2592004    -0.32   0.750    -.5906913    .4253555
  Other color |   .1752106   .2461273     0.71   0.477    -.3071901    .6576113
      No data |  -.0675577   .1682917    -0.40   0.688    -.3974034    .2622881
--------------+----------------------------------------------------------------
        /cut1 |  -3.863992   .2073207   -18.64   0.000    -4.270333   -3.457651
        /cut2 |   .2656596    .179821     1.48   0.140    -.0867831    .6181022
        /cut3 |   2.409058   .1866151    12.91   0.000     2.043299    2.774817
--------------+----------------------------------------------------------------
c_observador  |
    var(_cons)|   .3061185   .1156317                      .1460026    .6418276
-------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 210.85        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 4.58 | degrees of freedom: 10 | p-value: 0.92

```
Likelihood-ratio test                                 LR chi2(10)  =      4.57
(Assumption: m_null nested in m_taxi_color)           Prob > chi2 =    0.9178
```

Level 1 predictor variable: taxi_marca Vehicle brand

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3528.4589
Iteration 2:   log likelihood = -3528.4571
Iteration 3:   log likelihood = -3528.4571

Refining starting values:

Grid node 0:   log likelihood = -3430.9225

Fitting full model:

Iteration 0:   log likelihood = -3430.9225  (not concave)
Iteration 1:   log likelihood = -3428.5547  (not concave)
Iteration 2:   log likelihood = -3426.8535
Iteration 3:   log likelihood = -3423.7693
Iteration 4:   log likelihood =  -3423.759
Iteration 5:   log likelihood =  -3423.759

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7
```

```
                                                Wald chi2(14)     =       7.70
Log likelihood =  -3423.759                     Prob > chi2       =     0.9045
------------------------------------------------------------------------------
   precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
              |
   taxi_marca |
   CHEVROLET  |   .2692156   .3353769     0.80   0.422    -.3881111    .9265424
      DAEWOO  |   .0267051   .4136877     0.06   0.949    -.7841078     .837518
       HONDA  |   .1837676   .4792312     0.38   0.701    -.7555084    1.123044
     HYUNDAI  |   .1043929   .3310018     0.32   0.752    -.5443586    .7531445
         KIA  |   .4189453   .3295734     1.27   0.204    -.2270068    1.064897
       MAZDA  |   .1424656   .4387114     0.32   0.745    -.7173931    1.002324
  MITSUBISHI  |   .3346914   .4164853     0.80   0.422    -.4816048    1.150988
      NISSAN  |   .2417399   .3180463     0.76   0.447    -.3816194    .8650992
     RENAULT  |    .239794   .5453868     0.44   0.660    -.8291444    1.308732
      SUZUKI  |   .1765907   .4580458     0.39   0.700    -.7211626    1.074344
      TOYOTA  |   .3415528   .3187547     1.07   0.284    -.2831949    .9663005
  VOLKSWAGEN  |   .3661461   .3676105     1.00   0.319    -.3543572    1.086649
 Other brand  |   .2360135   .3382326     0.70   0.485    -.4269102    .8989372
     No data  |   .2261746   .3358536     0.67   0.501    -.4320863    .8844354
--------------+---------------------------------------------------------------
        /cut1 |   -3.56463   .3551186   -10.04   0.000     -4.26065    -2.86861
        /cut2 |   .5661334   .3408576     1.66   0.097    -.1019352    1.234202
        /cut3 |   2.711329   .3449377     7.86   0.000     2.035263    3.387394
--------------+---------------------------------------------------------------
c_observador  |
    var(_cons)|   .3051933   .1153245                      .1455219    .6400612
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 209.40        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 7.70 | degrees of freedom: 14 | p-value: 0.90

```
Likelihood-ratio test                                 LR chi2(14) =      7.73
(Assumption: m_null nested in m_taxi_marca)           Prob > chi2 =    0.9029
```

Level 1 predictor variable: taxi_anho Vehicle year

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3530.4082
Iteration 2:   log likelihood = -3530.4076
Iteration 3:   log likelihood = -3530.4076

Refining starting values:

Grid node 0:   log likelihood =  -3431.308

Fitting full model:

Iteration 0:   log likelihood =  -3431.308  (not concave)
Iteration 1:   log likelihood = -3428.9428  (not concave)
Iteration 2:   log likelihood = -3427.3137
Iteration 3:   log likelihood = -3424.5244
Iteration 4:   log likelihood = -3424.5147
Iteration 5:   log likelihood = -3424.5147

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316
```

```
Integration method: mvaghermite                Integration pts.  =          7

                                               Wald chi2(7)      =       6.25
Log likelihood = -3424.5147                    Prob > chi2       =     0.5104
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
   taxi_anho |
       2010  |   .0679638    .169671     0.40   0.689    -.2645852    .4005128
       2011  |   .0295876   .1228305     0.24   0.810    -.2111558    .2703311
       2012  |    .068612   .1307071     0.52   0.600    -.1875691    .3247932
       2013  |  -.0189941   .1303824    -0.15   0.884    -.2745389    .2365507
       2014  |    .186974   .1331269     1.40   0.160      -.07395    .4478979
       2015  |   .0760422   .1234543     0.62   0.538    -.1659237    .3180082
       2016  |   .2703596   .1269213     2.13   0.033     .0215984    .5191208
-------------+----------------------------------------------------------------
       /cut1 |  -3.779315   .1795158   -21.05   0.000    -4.131159    -3.42747
       /cut2 |   .3504765   .1475354     2.38   0.018     .0613124    .6396406
       /cut3 |   2.495185   .1559582    16.00   0.000     2.189512    2.800857
-------------+----------------------------------------------------------------
c_observador |
   var(_cons)|    .307335   .1160668                      .1466057    .6442777
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 211.79        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 6.25 | degrees of freedom: 7 | p-value: 0.51

```
Likelihood-ratio test                                 LR chi2(7)  =      6.22
(Assumption: m_null nested in m_taxi_anho)            Prob > chi2 =    0.5145
```

Level 1 predictor variable: c_dÍadeobservac Day of data collection

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3529.7036
Iteration 2:   log likelihood = -3529.7031
Iteration 3:   log likelihood = -3529.7031

Refining starting values:

Grid node 0:   log likelihood = -3432.4683

Fitting full model:

Iteration 0:   log likelihood = -3432.4683  (not concave)
Iteration 1:   log likelihood = -3430.0968  (not concave)
Iteration 2:   log likelihood = -3428.3761
Iteration 3:   log likelihood = -3425.6735
Iteration 4:   log likelihood =  -3425.664
Iteration 5:   log likelihood =  -3425.664

Mixed-effects ologit regression                Number of obs     =      3,537
Group variable:    c_observador                Number of groups  =         16

                                               Obs per group:
                                                             min =        120
                                                             avg =      221.1
                                                             max =        316

Integration method: mvaghermite                Integration pts.  =          7

                                               Wald chi2(1)      =       3.91
```

```
Log likelihood =  -3425.664                      Prob > chi2       =     0.0479
--------------------------------------------------------------------------------
     precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
                |
c_dÍadeobservac |
        Sunday  |  -.1413981    .071469    -1.98   0.048    -.2814747   -.0013214
----------------+---------------------------------------------------------------
          /cut1 |  -3.890482   .1784428   -21.80   0.000    -4.240223    -3.54074
          /cut2 |   .2378317   .1454495     1.64   0.102     -.047244    .5229075
          /cut3 |   2.381543    .153546    15.51   0.000     2.080598    2.682488
----------------+---------------------------------------------------------------
c_observador    |
      var(_cons)|   .3020786   .1141903                      .1439968    .6337049
--------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 208.08        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 3.91 | degrees of freedom: 1 | p-value: 0.05

```
Likelihood-ratio test                                 LR chi2(1)  =      3.92
(Assumption: m_null nested in m_c_dÍadeobs~c)         Prob > chi2 =    0.0477
```

Level 1 predictor variable: c_horariogr Shift of data collection

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3531.4873
Iteration 2:   log likelihood = -3531.4871

Refining starting values:

Grid node 0:   log likelihood = -3431.0097

Fitting full model:

Iteration 0:   log likelihood = -3431.0097  (not concave)
Iteration 1:   log likelihood = -3428.6493  (not concave)
Iteration 2:   log likelihood = -3427.0433
Iteration 3:   log likelihood = -3422.6699
Iteration 4:   log likelihood = -3422.6546
Iteration 5:   log likelihood = -3422.6546

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(2)      =       9.90
Log likelihood = -3422.6546                     Prob > chi2       =     0.0071
------------------------------------------------------------------------------------------
              precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+------------------------------------------------------------------
                       |
           c_horariogr |
Mid-afternoon (12-16)  |   .2965707   .0945828     3.14   0.002     .1111918    .4819495
    Afternoon (16-20)  |   .1885246   .1023127     1.84   0.065    -.0120047    .3890539
-----------------------+------------------------------------------------------------------
                 /cut1 |  -3.674113     .18701   -19.65   0.000    -4.040646   -3.307581
                 /cut2 |   .4609196   .1575844     2.92   0.003     .1520599    .7697792
                 /cut3 |   2.606477   .1658924    15.71   0.000     2.281334     2.93162
```

```
-----------------------+----------------------------------------------------------------
c_observador           |
             var(_cons)|   .3215344   .1214448                      .1533647    .6741084
----------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 217.67        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 9.90 | degrees of freedom: 2 | p-value: 0.01

```
Likelihood-ratio test                                 LR chi2(2)  =      9.94
(Assumption: m_null nested in m_c_horariogr)          Prob > chi2 =    0.0069
```

__82502300

__ i.taxi_color i.taxi_marca i.taxi_anho i.c_dÍadeobservac i.c_horariogr

Level 1 predictors (full model)

Level 1 all predictor variables: taxi_color taxi_marca taxi_anho c_dÍadeobservac c_hor ariogr .

meologit precio_cat $L1_list || c_observador:,

```
Fitting fixed-effects model:

Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3518.8046
Iteration 2:   log likelihood = -3518.7932
Iteration 3:   log likelihood = -3518.7932

Refining starting values:

Grid node 0:   log likelihood = -3421.2006

Fitting full model:

Iteration 0:   log likelihood = -3421.2006  (not concave)
Iteration 1:   log likelihood =  -3418.844  (not concave)
Iteration 2:   log likelihood = -3417.1054
Iteration 3:   log likelihood = -3411.6739
Iteration 4:   log likelihood = -3411.6578
Iteration 5:   log likelihood = -3411.6578

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(34)     =      31.74
Log likelihood = -3411.6578                     Prob > chi2       =     0.5791
----------------------------------------------------------------------------------------
            precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+----------------------------------------------------------------
                       |
            taxi_color |
                 Blue  |  -.1088401   .1969873    -0.55   0.581    -.4949281     .277248
                Beige  |  -.0595859   .2588706    -0.23   0.818     -.566963    .4477912
```

```
              White  |  -.1583387    .139661    -1.13   0.257    -.4320692    .1153919
               Gray  |  -.1422276   .1576222    -0.90   0.367    -.4511614    .1667063
              Black  |  -.0722203   .1464911    -0.49   0.622    -.3593376     .214897
             Silver  |   .0354097   .1461653     0.24   0.809    -.2510689    .3218884
                Red  |  -.0258221   .1695313    -0.15   0.879    -.3580974    .3064531
              Green  |   -.102001   .2637486    -0.39   0.699    -.6189387    .4149368
        Other color  |   .1705732   .2494718     0.68   0.494    -.3183826     .659529
            No data  |   .2232924   .7586232     0.29   0.768    -1.263582    1.710167
                     |
          taxi_marca |
          CHEVROLET  |   .2910729    .338908     0.86   0.390    -.3731746    .9553205
             DAEWOO  |   .0417839   .4248939     0.10   0.922    -.7909928    .8745606
              HONDA  |   .1207322   .4866667     0.25   0.804    -.8331169    1.074581
            HYUNDAI  |   .0553614   .3342317     0.17   0.868    -.5997207    .7104435
                KIA  |   .4083106    .332837     1.23   0.220     -.244038    1.060659
              MAZDA  |   .1687051   .4451964     0.38   0.705    -.7038639    1.041274
         MITSUBISHI  |   .3980309   .4238251     0.94   0.348     -.432651    1.228713
             NISSAN  |    .234335   .3222035     0.73   0.467    -.3971722    .8658423
            RENAULT  |   .2075322   .5476415     0.38   0.705    -.8658254     1.28089
             SUZUKI  |   .1430628    .463807     0.31   0.758    -.7659822    1.052108
             TOYOTA  |    .350519   .3227219     1.09   0.277    -.2820043    .9830424
         VOLKSWAGEN  |    .386635   .3717838     1.04   0.298    -.3420478    1.115318
        Other brand  |    .220389   .3429627     0.64   0.520    -.4518055    .8925836
            No data  |  -.0312564   .8226859    -0.04   0.970    -1.643691    1.581178
                     |
           taxi_anho |
               2010  |   .0305074   .1750613     0.17   0.862    -.3126065    .3736213
               2011  |    .015601   .1286668     0.12   0.903    -.2365814    .2677833
               2012  |   .0912868   .1358665     0.67   0.502    -.1750066    .3575802
               2013  |  -.0502736   .1344472    -0.37   0.708    -.3137853    .2132382
               2014  |   .2185206    .137482     1.59   0.112    -.0509391    .4879804
               2015  |   .0845313   .1311105     0.64   0.519    -.1724407    .3415032
               2016  |   .2886914   .1345329     2.15   0.032     .0250117    .5523711
                     |
     c_dÍadeobservac |
             Sunday  |   -.094907   .0746395    -1.27   0.204    -.2411977    .0513838
                     |
         c_horariogr |
Mid-afternoon (12-16)  |   .2703832   .0988898     2.73   0.006     .0765627    .4642036
    Afternoon (16-20)  |    .182691   .1037254     1.76   0.078     -.020607     .385989
---------------------+----------------------------------------------------------------
               /cut1 |  -3.487428   .3896347    -8.95   0.000    -4.251098   -2.723758
               /cut2 |   .6604647   .3769788     1.75   0.080    -.0784002     1.39933
               /cut3 |   2.815657   .3808577     7.39   0.000      2.06919    3.562125
---------------------+----------------------------------------------------------------
c_observador         |
           var(_cons)|   .3220203   .1217465                          .1534859    .6756128
--------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 214.27        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 31.74 | degrees of freedom: 34 | p-value: 0.58

```
.          lrtest m_all_L1 m_null

Likelihood-ratio test                                 LR chi2(34) =      31.93
(Assumption: m_null nested in m_all_L1)               Prob > chi2 =     0.5693
```

Appendix H: Results of multilevel ordinal logistic regressions using variables only at level 2

Estimation using level-2 covariates

Level 2 predictor variable, no controls: c_fsexo Sex

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3527.5892
Iteration 2:   log likelihood = -3527.5878
Iteration 3:   log likelihood = -3527.5878

Refining starting values:

Grid node 0:   log likelihood = -3433.9281

Fitting full model:

Iteration 0:   log likelihood = -3433.9281  (not concave)
Iteration 1:   log likelihood = -3431.5364  (not concave)
Iteration 2:   log likelihood = -3429.5729
Iteration 3:   log likelihood = -3427.0302
Iteration 4:   log likelihood = -3427.0087
Iteration 5:   log likelihood = -3427.0086

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(1)      =       1.27
Log likelihood = -3427.0086                     Prob > chi2       =     0.2589
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
     c_fsexo |
     Female  |  -.3097231   .2743564    -1.13   0.259    -.8474518    .2280056
-------------+----------------------------------------------------------------
       /cut1 |  -3.981325   .2203613   -18.07   0.000    -4.413225   -3.549425
       /cut2 |     .14471    .194291     0.74   0.456    -.2360933    .5255133
       /cut3 |   2.286587   .2001403    11.42   0.000     1.894319    2.678855
-------------+----------------------------------------------------------------
c_observador |
   var(_cons)|   .2820197   .1067502                      .1343019    .5922111
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 201.16        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 1.27 | degrees of freedom: 1 | p-value: 0.26

```
Likelihood-ratio test                                 LR chi2(1)  =      1.23
(Assumption: m_null nested in m_c_fsexo)              Prob > chi2 =    0.2674
variable m_c_fsexo not found
```

Level 2 predictor variable, controlling by level 1 variables: m_c_fsexo

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood =  -3511.099
Iteration 2:   log likelihood = -3511.0738
Iteration 3:   log likelihood = -3511.0738

Refining starting values:

Grid node 0:   log likelihood = -3422.0004

Fitting full model:

Iteration 0:   log likelihood = -3422.0004  (not concave)
Iteration 1:   log likelihood = -3419.6202  (not concave)
Iteration 2:   log likelihood = -3417.5259
Iteration 3:   log likelihood = -3411.1754
Iteration 4:   log likelihood =  -3411.014
Iteration 5:   log likelihood = -3411.0119
Iteration 6:   log likelihood = -3411.0119

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(35)     =      33.00
Log likelihood = -3411.0119                     Prob > chi2       =     0.5652
```

precio_cat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
c_fsexo						
Female	-.3258649	.2812686	-1.16	0.247	-.8771411	.2254114
taxi_color						
Blue	-.107064	.196988	-0.54	0.587	-.4931534	.2790255
Beige	-.0590718	.2588603	-0.23	0.819	-.5664285	.448285
White	-.158629	.1396599	-1.14	0.256	-.4323573	.1150993
Gray	-.141532	.1576203	-0.90	0.369	-.4504621	.1673981
Black	-.0718337	.1464881	-0.49	0.624	-.3589451	.2152777
Silver	.0353007	.1461678	0.24	0.809	-.2511829	.3217843
Red	-.0257604	.1695397	-0.15	0.879	-.3580521	.3065314
Green	-.1016819	.2637627	-0.39	0.700	-.6186473	.4152835
Other color	.1702383	.2494527	0.68	0.495	-.31868	.6591566
No data	.2250386	.7586325	0.30	0.767	-1.261854	1.711931
taxi_marca						
CHEVROLET	.2915743	.3388771	0.86	0.390	-.3726127	.9557613
DAEWOO	.0439736	.4249094	0.10	0.918	-.7888336	.8767807
HONDA	.1212505	.486664	0.25	0.803	-.8325934	1.075095
HYUNDAI	.0557368	.3342055	0.17	0.868	-.5992939	.7107676
KIA	.4089955	.3328077	1.23	0.219	-.2432955	1.061286
MAZDA	.1692473	.4452374	0.38	0.704	-.703402	1.041897
MITSUBISHI	.3993735	.423804	0.94	0.346	-.431267	1.230014
NISSAN	.2347716	.3221723	0.73	0.466	-.3966745	.8662178
RENAULT	.2110258	.5476308	0.39	0.700	-.8623109	1.284363
SUZUKI	.1438279	.4637641	0.31	0.756	-.7651331	1.052789
TOYOTA	.3510495	.3226898	1.09	0.277	-.2814108	.9835099
VOLKSWAGEN	.3873755	.3717604	1.04	0.297	-.3412614	1.116012
Other brand	.2212463	.3429296	0.65	0.519	-.4508834	.8933761
No data	-.031404	.822679	-0.04	0.970	-1.643825	1.581017
taxi_anho						
2010	.0337351	.1750837	0.19	0.847	-.3094226	.3768929
2011	.0154699	.1286618	0.12	0.904	-.2367027	.2676424
2012	.0916049	.1358666	0.67	0.500	-.1746886	.3578985
2013	-.0504091	.1344427	-0.37	0.708	-.313912	.2130938
2014	.2176759	.1374847	1.58	0.113	-.0517892	.4871411
2015	.0835516	.1311081	0.64	0.524	-.1734156	.3405189
2016	.2885609	.1345241	2.15	0.032	.0248985	.5522234

```
       c_díadeobservac |
                Sunday  |  -.1001302   .0747856    -1.34   0.181    -.2467072    .0464469
                        |
            c_horariogr |
 Mid-afternoon (12-16)  |   .2670953   .0988833     2.70   0.007     .0732875    .4609031
     Afternoon (16-20)  |   .1813593   .1036277     1.75   0.080    -.0217473    .3844659
------------------------+----------------------------------------------------------------
                  /cut1 |  -3.652972   .4130671    -8.84   0.000    -4.462569   -2.843375
                  /cut2 |   .4945419   .4008209     1.23   0.217    -.2910527    1.280136
                  /cut3 |   2.649828   .4042176     6.56   0.000     1.857576     3.44208
------------------------+----------------------------------------------------------------
c_observador            |
              var(_cons)|   .2960367   .1123009                      .1407486    .6226546
-----------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 200.12        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 33.00 | degrees of freedom: 35 | p-value: 0.57

```
Likelihood-ratio test                                 LR chi2(34) =     31.99
(Assumption: m_c_fsexo nested in m_c_fsexo_L1)        Prob > chi2 =    0.5663
```

Level 2 predictor variable, no controls: c_fedad Age

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3513.5356
Iteration 2:   log likelihood = -3513.5156
Iteration 3:   log likelihood = -3513.5156

Refining starting values:

Grid node 0:   log likelihood = -3433.2379

Fitting full model:

Iteration 0:   log likelihood = -3433.2379  (not concave)
Iteration 1:   log likelihood = -3430.8335  (not concave)
Iteration 2:   log likelihood = -3428.7622
Iteration 3:   log likelihood = -3426.7956
Iteration 4:   log likelihood = -3426.6431
Iteration 5:   log likelihood = -3426.6423
Iteration 6:   log likelihood = -3426.6423

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(2)      =       2.09
Log likelihood = -3426.6423                     Prob > chi2       =     0.3520
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
     c_fedad |
      19-64  |  -.4611942   .3286851    -1.40   0.161    -1.105405    .1830167
        65+  |  -.2589914   .3275939    -0.79   0.429    -.9010637    .3830809
```

```
-------------+----------------------------------------------------------------
       /cut1 |  -4.006233   .2163681   -18.52   0.000    -4.430306   -3.582159
       /cut2 |   .1208225   .1889861     0.64   0.523    -.2495834    .4912284
       /cut3 |   2.262366   .1951212    11.59   0.000     1.879936    2.644797
-------------+----------------------------------------------------------------
c_observador |
   var(_cons)|   .2674803    .101974                      .1267011    .5646809
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 173.75        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 2.09 | degrees of freedom: 2 | p-value: 0.35

```
Likelihood-ratio test                                 LR chi2(2)  =      1.96
(Assumption: m_null nested in m_c_fedad)              Prob > chi2 =    0.3748
variable m_c_fedad not found
```

Level 2 predictor variable, controlling by level 1 variables: m_c_fedad

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3500.0697
Iteration 2:   log likelihood = -3500.0102
Iteration 3:   log likelihood = -3500.0102

Refining starting values:

Grid node 0:   log likelihood = -3420.0373

Fitting full model:

Iteration 0:   log likelihood = -3420.0373  (not concave)
Iteration 1:   log likelihood =  -3417.648  (not concave)
Iteration 2:   log likelihood = -3415.5369
Iteration 3:   log likelihood = -3410.9904
Iteration 4:   log likelihood = -3410.7198
Iteration 5:   log likelihood = -3410.7157
Iteration 6:   log likelihood = -3410.7157

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(36)     =      33.62
Log likelihood = -3410.7157                     Prob > chi2       =     0.5825
-------------------------------------------------------------------------------------
            precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+-------------------------------------------------------------
                       |
               c_fedad |
                19-64  |  -.4666156   .3383647    -1.38   0.168    -1.129798     .196567
                  65+  |  -.2542155   .3373993    -0.75   0.451     -.915506     .407075
                       |
            taxi_color |
                 Blue  |   -.108833   .1969859    -0.55   0.581    -.4949181    .2772522
                Beige  |  -.0600528   .2588877    -0.23   0.817    -.5674633    .4473578
                White  |  -.1575022   .1396567    -1.13   0.259    -.4312242    .1162199
```

```
                Gray |  -.1420909   .1576125    -0.90   0.367    -.4510057    .1668239
               Black |  -.0714948   .1464851    -0.49   0.626    -.3586002    .2156106
              Silver |   .0352641   .1461594     0.24   0.809     -.251203    .3217312
                 Red |  -.0257455   .1695202    -0.15   0.879    -.3579989    .3065079
               Green |  -.1009832   .2637034    -0.38   0.702    -.6178324    .4158661
         Other color |   .1684833   .2494393     0.68   0.499    -.3204086    .6573753
             No data |    .231611   .7586967     0.31   0.760    -1.255407    1.718629
                     |
          taxi_marca |
           CHEVROLET |   .2889695   .3389303     0.85   0.394    -.3753217    .9532608
              DAEWOO |   .0399809   .4248655     0.09   0.925    -.7927401    .8727019
               HONDA |   .1211722   .4866684     0.25   0.803    -.8326804    1.075025
             HYUNDAI |   .0526192   .3342505     0.16   0.875    -.6024997    .7077382
                 KIA |   .4060015   .3328558     1.22   0.223    -.2463839    1.058387
               MAZDA |   .1657892   .4452218     0.37   0.710    -.7068295    1.038408
          MITSUBISHI |   .3945443   .4238511     0.93   0.352    -.4361885    1.225277
              NISSAN |   .2327847   .3222195     0.72   0.470     -.398754    .8643234
             RENAULT |   .2037758   .5476892     0.37   0.710    -.8696753    1.277227
              SUZUKI |   .1438651    .463839     0.31   0.756    -.7652425    1.052973
              TOYOTA |   .3482022   .3227373     1.08   0.281    -.2843514    .9807557
          VOLKSWAGEN |   .3860531   .3717761     1.04   0.299    -.3426146    1.114721
         Other brand |   .2197846   .3429714     0.64   0.522     -.452427    .8919962
             No data |  -.0397053   .8227473    -0.05   0.962     -1.65226     1.57285
                     |
           taxi_anho |
                2010 |   .0291513   .1750653     0.17   0.868    -.3139703    .3722729
                2011 |   .0160905   .1286638     0.13   0.900    -.2360859    .2682669
                2012 |   .0896244   .1358789     0.66   0.510    -.1766934    .3559423
                2013 |  -.0515951   .1344462    -0.38   0.701    -.3151048    .2119145
                2014 |    .218162    .137485     1.59   0.113    -.0513037    .4876277
                2015 |    .085344   .1311237     0.65   0.515    -.1716538    .3423418
                2016 |   .2887248   .1345372     2.15   0.032     .0250367    .5524129
                     |
    c_díadeobservac |
              Sunday |  -.0914809   .0746443    -1.23   0.220    -.2377811    .0548193
                     |
         c_horariogr |
 Mid-afternoon (12-16)  |   .2731502   .0987998     2.76   0.006     .0795061    .4667942
   Afternoon (16-20)  |     .17894   .1036408     1.73   0.084    -.0241922    .3820722
---------------------+----------------------------------------------------------------
               /cut1 |   -3.66772    .410928    -8.93   0.000    -4.473124   -2.862316
               /cut2 |    .480802   .3982506     1.21   0.227    -.2997549    1.261359
               /cut3 |   2.635777   .4017456     6.56   0.000     1.848371    3.423184
---------------------+----------------------------------------------------------------
c_observador         |
           var(_cons)|   .2837253   .1082203                          .1343468    .5991957
--------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 178.59        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 33.62 | degrees of freedom: 36 | p-value: 0.58

```
Likelihood-ratio test                                 LR chi2(34) =      31.85
(Assumption: m_c_fedad nested in m_c_fedad_L1)        Prob > chi2 =     0.5732
```

Level 2 predictor variable, no controls: c_ftez Phenotype

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3502.8895
Iteration 2:   log likelihood = -3502.8458
Iteration 3:   log likelihood = -3502.8458

Refining starting values:

Grid node 0:   log likelihood = -3432.6229

Fitting full model:
```

```
Iteration 0:   log likelihood = -3432.6229  (not concave)
Iteration 1:   log likelihood = -3427.4263
Iteration 2:   log likelihood = -3425.8406
Iteration 3:   log likelihood = -3425.7192
Iteration 4:   log likelihood = -3425.7188
Iteration 5:   log likelihood = -3425.7188

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(1)      =       4.31
Log likelihood = -3425.7188                     Prob > chi2       =     0.0378
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
      c_ftez |
      White  |   .5239237   .2522914     2.08   0.038     .0294416    1.018406
-------------+----------------------------------------------------------------
       /cut1 |  -3.566099   .2044529   -17.44   0.000    -3.966819   -3.165379
       /cut2 |   .5599797   .1782411     3.14   0.002     .2106336    .9093258
       /cut3 |   2.702041   .1859758    14.53   0.000     2.337535    3.066547
-------------+----------------------------------------------------------------
c_observador |
   var(_cons)|   .2354375   .0908112                      .1105494    .5014121
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 154.25        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 4.31 | degrees of freedom: 1 | p-value: 0.04

```
Likelihood-ratio test                                 LR chi2(1)  =      3.81
(Assumption: m_null nested in m_c_ftez)               Prob > chi2 =    0.0510
variable m_c_ftez not found
```

Level 2 predictor variable, controlling by level 1 variables: m_c_ftez

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3488.5487
Iteration 2:   log likelihood = -3488.4429
Iteration 3:   log likelihood = -3488.4429

Refining starting values:

Grid node 0:   log likelihood =  -3421.401

Fitting full model:

Iteration 0:   log likelihood =  -3421.401  (not concave)
Iteration 1:   log likelihood = -3416.1756
Iteration 2:   log likelihood = -3410.7485
Iteration 3:   log likelihood = -3409.9877
Iteration 4:   log likelihood = -3409.9811
```

```
Iteration 5:   log likelihood = -3409.9811

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(35)     =      35.23
Log likelihood = -3409.9811                     Prob > chi2       =     0.4572
------------------------------------------------------------------------------
           precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------+-------------------------------------------------------
                      |
               c_ftez |
                White |   .5087603   .2629431     1.93   0.053    -.0065987    1.024119
                      |
          taxi_color  |
                 Blue |  -.1118056    .196982    -0.57   0.570    -.4978831     .274272
                Beige |  -.0609607   .2588769    -0.24   0.814      -.56835    .4464287
                White |  -.1612554   .1396773    -1.15   0.248    -.4350178     .112507
                 Gray |  -.1437951    .157624    -0.91   0.362    -.4527325    .1651422
                Black |  -.0734868   .1464975    -0.50   0.616    -.3606166     .213643
               Silver |   .0337952   .1461694     0.23   0.817    -.2526916     .320282
                  Red |  -.0278159    .169545    -0.16   0.870     -.360118    .3044862
                Green |  -.1034863   .2637235    -0.39   0.695    -.6203749    .4134024
          Other color |   .1702936   .2494276     0.68   0.495    -.3185754    .6591627
              No data |   .2106933   .7584615     0.28   0.781    -1.275864     1.69725
                      |
          taxi_marca  |
            CHEVROLET |   .2883198   .3389273     0.85   0.395    -.3759655     .952605
               DAEWOO |   .0341834   .4248728     0.08   0.936    -.7985519    .8669187
                HONDA |   .1236284   .4866424     0.25   0.799    -.8301731     1.07743
              HYUNDAI |   .0527213    .334254     0.16   0.875    -.6024045    .7078471
                  KIA |   .4060167   .3328601     1.22   0.223     -.246377     1.05841
                MAZDA |   .1722593   .4451887     0.39   0.699    -.7002946    1.044813
           MITSUBISHI |   .3913355   .4238112     0.92   0.356    -.4393191     1.22199
               NISSAN |   .2328063   .3222265     0.72   0.470    -.3987461    .8643587
              RENAULT |   .2043549   .5474795     0.37   0.709    -.8686852    1.277395
               SUZUKI |   .1387498    .463788     0.30   0.765     -.770258    1.047758
               TOYOTA |   .3484115   .3227437     1.08   0.280    -.2841546    .9809776
           VOLKSWAGEN |    .381954   .3717885     1.03   0.304    -.3467381    1.110646
          Other brand |   .2191255   .3429817     0.64   0.523    -.4531063    .8913573
              No data |  -.0241175   .8225004    -0.03   0.977    -1.636189    1.587954
                      |
           taxi_anho  |
                 2010 |   .0298616    .175026     0.17   0.865    -.3131831    .3729063
                 2011 |   .0169974    .128674     0.13   0.895     -.235199    .2691939
                 2012 |   .0917812   .1358758     0.68   0.499    -.1745305    .3580929
                 2013 |  -.0490767   .1344242    -0.37   0.715    -.3125434      .21439
                 2014 |   .2202845   .1374562     1.60   0.109    -.0491247    .4896937
                 2015 |   .0853256    .131111     0.65   0.515    -.1716473    .3422986
                 2016 |   .2870505   .1345146     2.13   0.033     .0234067    .5506942
                      |
     c_díadeobservac  |
               Sunday |  -.0976114   .0746024    -1.31   0.191    -.2438294    .0486066
                      |
          c_horariogr |
 Mid-afternoon (12-16)|   .2598974   .0990298     2.62   0.009     .0658027    .4539922
     Afternoon (16-20)|   .1823827   .1033681     1.76   0.078     -.020215    .3849804
----------------------+-------------------------------------------------------
                /cut1 |  -3.242377     .40454    -8.01   0.000    -4.035261   -2.449493
                /cut2 |   .9051405    .393037     2.30   0.021     .1348023    1.675479
                /cut3 |   3.060596    .397163     7.71   0.000     2.282171    3.839022
----------------------+-------------------------------------------------------
c_observador          |
            var(_cons)|   .2559396   .0986851                       .1202074    .5449336
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 156.92        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 35.23 | degrees of freedom: 35 | p-value: 0.46

```
Likelihood-ratio test                                 LR chi2(34) =     31.48
(Assumption: m_c_ftez nested in m_c_ftez_L1)          Prob > chi2 =    0.5920
```

Level 2 predictor variable, no controls: c_facento Accent

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3524.9138
Iteration 2:   log likelihood = -3524.9107
Iteration 3:   log likelihood = -3524.9107

Refining starting values:

Grid node 0:   log likelihood = -3433.7904

Fitting full model:

Iteration 0:   log likelihood = -3433.7904  (not concave)
Iteration 1:   log likelihood = -3431.3946  (not concave)
Iteration 2:   log likelihood = -3429.3929
Iteration 3:   log likelihood = -3426.9467
Iteration 4:   log likelihood = -3426.9143
Iteration 5:   log likelihood = -3426.9142

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(1)      =       1.48
Log likelihood = -3426.9142                     Prob > chi2       =     0.2238
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
   c_facento |
  Foreigner  |    .331679   .2726338     1.22   0.224    -.2026735    .8660315
-------------+----------------------------------------------------------------
       /cut1 |  -3.662052   .2173854   -16.85   0.000     -4.08812   -3.235985
       /cut2 |   .4645032   .1925893     2.41   0.016     .0870351    .8419712
       /cut3 |   2.606304   .1994245    13.07   0.000     2.215439    2.997169
-------------+----------------------------------------------------------------
c_observador |
   var(_cons)|   .2782338   .1054921                      .1323359    .5849816
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 195.99        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 1.48 | degrees of freedom: 1 | p-value: 0.22

```
Likelihood-ratio test                                 LR chi2(1)  =      1.42
(Assumption: m_null nested in m_c_facento)            Prob > chi2 =    0.2336
variable m_c_facento not found
```

Level 2 predictor variable, controlling by level 1 variables: m_c_facento

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3510.0059
Iteration 2:   log likelihood = -3509.9772
Iteration 3:   log likelihood = -3509.9772

Refining starting values:

Grid node 0:   log likelihood = -3420.1505

Fitting full model:

Iteration 0:   log likelihood = -3420.1505  (not concave)
Iteration 1:   log likelihood = -3417.7664  (not concave)
Iteration 2:   log likelihood = -3415.7039
Iteration 3:   log likelihood = -3410.9284
Iteration 4:   log likelihood = -3410.8426
Iteration 5:   log likelihood = -3410.8414
Iteration 6:   log likelihood = -3410.8414

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(35)     =      33.36
Log likelihood = -3410.8414                     Prob > chi2       =     0.5476
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
   c_facento |
   Foreigner |   .3645158   .2784166     1.31   0.190    -.1811708    .9102024
             |
  taxi_color |
        Blue |   -.109532   .1969619    -0.56   0.578    -.4955703    .2765062
       Beige |  -.0591378    .258856    -0.23   0.819    -.5664863    .4482107
       White |  -.1567838   .1396555    -1.12   0.262    -.4305036     .116936
        Gray |  -.1418615   .1576145    -0.90   0.368    -.4507803    .1670572
       Black |  -.0721955   .1464799    -0.49   0.622    -.3592909    .2148999
      Silver |   .0357619    .146155     0.24   0.807    -.2506966    .3222204
         Red |  -.0261624   .1695127    -0.15   0.877    -.3584011    .3060764
       Green |  -.1017856   .2637241    -0.39   0.700    -.6186754    .4151042
 Other color |     .16977   .2494775     0.68   0.496    -.3191969     .658737
     No data |   .2265989   .7586627     0.30   0.765    -1.260353    1.713551
             |
  taxi_marca |
   CHEVROLET |    .291801   .3388804     0.86   0.389    -.3723924    .9559944
     DAEWOO  |   .0416661     .42483     0.10   0.922    -.7909854    .8743175
      HONDA  |   .1188468   .4866629     0.24   0.807     -.834995    1.072689
    HYUNDAI  |   .0557018   .3342041     0.17   0.868    -.5993262    .7107297
        KIA  |   .4089163   .3328133     1.23   0.219    -.2433857    1.061218
      MAZDA  |    .166861    .445175     0.37   0.708    -.7056658    1.039388
```

```
       MITSUBISHI  |   .3997839   .4238082     0.94   0.346     -.430865    1.230433
           NISSAN  |   .2336623   .3221774     0.73   0.468    -.3977937    .8651184
          RENAULT  |   .2059331   .5476504     0.38   0.707     -.867442    1.279308
           SUZUKI  |   .1444976   .4637957     0.31   0.755    -.7645253     1.05352
           TOYOTA  |   .3507772   .3226977     1.09   0.277    -.2816987     .983253
       VOLKSWAGEN  |   .3864188   .3717606     1.04   0.299    -.3422186    1.115056
      Other brand  |   .2193162   .3429383     0.64   0.522    -.4528306     .891463
          No data  |  -.0344347   .8227153    -0.04   0.967    -1.646927    1.578058
                   |
        taxi_anho  |
             2010  |   .0301388   .1750365     0.17   0.863    -.3129265    .3732041
             2011  |   .0158345   .1286661     0.12   0.902    -.2363465    .2680155
             2012  |   .0916908   .1358559     0.67   0.500    -.1745819    .3579634
             2013  |  -.0512043   .1344367    -0.38   0.703    -.3146954    .2122868
             2014  |    .218062   .1374789     1.59   0.113    -.0513917    .4875157
             2015  |   .0849346   .1311104     0.65   0.517     -.172037    .3419063
             2016  |   .2894139   .1345265     2.15   0.031     .0257468    .5530811
                   |
  c_dÍadeobservac  |
           Sunday  |  -.0924156   .0746144    -1.24   0.216    -.2386571    .0538259
                   |
      c_horariogr  |
Mid-afternoon (12-16) |  .2757414  .0988685   2.79   0.005     .0819627      .46952
 Afternoon (16-20) |   .1808223   .1035806     1.75   0.081    -.0221918    .3838365
-------------------+----------------------------------------------------------------
            /cut1  |  -3.304017   .4114891    -8.03   0.000    -4.110521   -2.497513
            /cut2  |   .8441483    .400106     2.11   0.035     .0599549    1.628342
            /cut3  |   2.999289   .4040007     7.42   0.000     2.207463    3.791116
-------------------+----------------------------------------------------------------
c_observador       |
        var(_cons) |   .2894913   .1099239                       .1375379    .6093246
------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 198.27        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 33.36 | degrees of freedom: 35 | p-value: 0.55

```
Likelihood-ratio test                                 LR chi2(34) =     32.15
(Assumption: m_c_facento nested in m_c_facento_L1)    Prob > chi2 =    0.5587
```

Level 2 predictor variable, no controls: c_fimagen Appearance

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3529.3919
Iteration 2:   log likelihood = -3529.3912
Iteration 3:   log likelihood = -3529.3912

Refining starting values:

Grid node 0:   log likelihood = -3434.0908

Fitting full model:

Iteration 0:   log likelihood = -3434.0908  (not concave)
Iteration 1:   log likelihood = -3431.7128  (not concave)
Iteration 2:   log likelihood =  -3429.929
Iteration 3:   log likelihood = -3427.4271
Iteration 4:   log likelihood = -3427.4184
Iteration 5:   log likelihood = -3427.4184

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
```

```
                                                   avg =      221.1
                                                   max =        316

Integration method: mvaghermite                Integration pts.  =          7

                                               Wald chi2(1)      =       0.42
Log likelihood = -3427.4184                    Prob > chi2       =     0.5191
------------------------------------------------------------------------------
  precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             |
   c_fimagen |
      Tacky  |   .1812066   .2810545     0.64   0.519      -.36965    .7320633
-------------+----------------------------------------------------------------
       /cut1 |  -3.735884   .2238916   -16.69   0.000    -4.174704   -3.297065
       /cut2 |   .3904631   .1993544     1.96   0.050    -.0002643    .7811905
       /cut3 |   2.532239   .2057365    12.31   0.000     2.129003    2.935475
-------------+----------------------------------------------------------------
c_observador |
   var(_cons)|   .2969584   .1123603                      .1414575    .6233978
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 203.95        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 0.42 | degrees of freedom: 1 | p-value: 0.52

```
Likelihood-ratio test                                 LR chi2(1)  =      0.41
(Assumption: m_null nested in m_c_fimagen)            Prob > chi2 =    0.5217
variable m_c_fimagen not found
```

Level 2 predictor variable, controlling by level 1 variables: m_c_fimagen

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood =  -3515.019
Iteration 2:   log likelihood = -3515.0008
Iteration 3:   log likelihood = -3515.0008

Refining starting values:

Grid node 0:   log likelihood = -3420.0484

Fitting full model:

Iteration 0:   log likelihood = -3420.0484  (not concave)
Iteration 1:   log likelihood =  -3417.685  (not concave)
Iteration 2:   log likelihood = -3415.8668
Iteration 3:   log likelihood = -3411.3803
Iteration 4:   log likelihood = -3411.3672
Iteration 5:   log likelihood = -3411.3672

Mixed-effects ologit regression                Number of obs     =      3,537
Group variable:    c_observador                Number of groups  =         16

                                               Obs per group:
                                                   min =        120
                                                   avg =      221.1
                                                   max =        316

Integration method: mvaghermite                Integration pts.  =          7

                                               Wald chi2(35)     =      32.30
```

```
Log likelihood = -3411.3672                       Prob > chi2       =     0.5992
-----------------------------------------------------------------------------------------
             precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------------+----------------------------------------------------------------
                        |
              c_fimagen |
                  Tacky |   .2211196   .2876011     0.77   0.442    -.3425682    .7848073
                        |
             taxi_color |
                   Blue |  -.1093159   .1969911    -0.55   0.579    -.4954114    .2767796
                  Beige |  -.0590784   .2588811    -0.23   0.819    -.5664761    .4483192
                  White |  -.1583724   .1396578    -1.13   0.257    -.4320967     .115352
                   Gray |  -.1431697   .1576272    -0.91   0.364    -.4521133     .165774
                  Black |  -.0726244   .1464894    -0.50   0.620    -.3597383    .2144896
                 Silver |   .0353874   .1461621     0.24   0.809    -.2510849    .3218598
                    Red |  -.0255048   .1695294    -0.15   0.880    -.3577762    .3067666
                  Green |  -.1039617   .2637583    -0.39   0.693    -.6209185    .4129952
            Other color |    .170417   .2495013     0.68   0.495    -.3185966    .6594307
                No data |   .2261816   .7586309     0.30   0.766    -1.260708    1.713071
                        |
             taxi_marca |
              CHEVROLET |   .2925179   .3388746     0.86   0.388    -.3716641    .9566999
                 DAEWOO |   .0441836   .4248878     0.10   0.917    -.7885813    .8769485
                  HONDA |   .1201763    .486644     0.25   0.805    -.8336285    1.073981
                HYUNDAI |   .0560386   .3341981     0.17   0.867    -.5989776    .7110548
                    KIA |   .4095803    .332806     1.23   0.218    -.2427075    1.061868
                  MAZDA |   .1700406   .4451681     0.38   0.702    -.7024728    1.042554
             MITSUBISHI |   .3990796   .4237981     0.94   0.346    -.4315493    1.229709
                 NISSAN |   .2348461   .3221665     0.73   0.466    -.3965887    .8662809
                RENAULT |   .2084217   .5476372     0.38   0.704    -.8649275    1.281771
                 SUZUKI |   .1440319   .4637989     0.31   0.756    -.7649973    1.053061
                 TOYOTA |    .351318   .3226864     1.09   0.276    -.2811357    .9837717
             VOLKSWAGEN |   .3881424   .3717584     1.04   0.296    -.3404906    1.116775
            Other brand |   .2207425   .3429263     0.64   0.520    -.4513807    .8928657
                No data |  -.0341993   .8226796    -0.04   0.967    -1.646622    1.578223
                        |
              taxi_anho |
                   2010 |   .0302061   .1750714     0.17   0.863    -.3129276    .3733398
                   2011 |   .0152293   .1286651     0.12   0.906    -.2369496    .2674083
                   2012 |   .0915124   .1358617     0.67   0.501    -.1747717    .3577964
                   2013 |  -.0502219   .1344489    -0.37   0.709    -.3137369     .213293
                   2014 |   .2191389   .1374846     1.59   0.111    -.0503259    .4886038
                   2015 |   .0845977   .1311125     0.65   0.519    -.1723782    .3415735
                   2016 |   .2889059   .1345392     2.15   0.032     .0252139    .5525978
                        |
       c_díadeobservac |
                 Sunday |  -.0926656   .0746747    -1.24   0.215    -.2390253    .0536941
                        |
             c_horariogr |
  Mid-afternoon (12-16) |   .2740482   .0989433     2.77   0.006      .080123    .4679735
      Afternoon (16-20) |   .1852114   .1036748     1.79   0.074    -.0179874    .3884103
------------------------+----------------------------------------------------------------
                  /cut1 |  -3.372939   .4161496    -8.11   0.000    -4.188577   -2.557301
                  /cut2 |   .7749563   .4046273     1.92   0.055    -.0180986    1.568011
                  /cut3 |   2.930108   .4083729     7.18   0.000     2.129712    3.730505
------------------------+----------------------------------------------------------------
c_observador            |
              var(_cons)|   .3101481   .1174065                       .1476886    .6513155
-----------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 207.27        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 32.30 | degrees of freedom: 35 | p-value: 0.60

```
Likelihood-ratio test                                 LR chi2(34) =      32.10
(Assumption: m_c_fimagen nested in m_c_fimagen_L1)    Prob > chi2 =     0.5609
```

Level 2 predictor variable, no controls: c_fvestimenta Attire

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3532.1045
Iteration 2:   log likelihood = -3532.1045

Refining starting values:

Grid node 0:   log likelihood = -3434.2174

Fitting full model:

Iteration 0:   log likelihood = -3434.2174  (not concave)
Iteration 1:   log likelihood =  -3431.842  (not concave)
Iteration 2:   log likelihood = -3430.0913
Iteration 3:   log likelihood = -3427.4629
Iteration 4:   log likelihood =  -3427.454
Iteration 5:   log likelihood =  -3427.454

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(1)      =       0.34
Log likelihood =  -3427.454                     Prob > chi2       =     0.5583
------------------------------------------------------------------------------
   precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
              |
c_fvestimenta |
       Casual |  -.1649644   .2818179    -0.59   0.558    -.7173173    .3873885
-------------+----------------------------------------------------------------
        /cut1 |  -3.909525   .2251379   -17.37   0.000    -4.350788   -3.468263
        /cut2 |   .2166722   .2001028     1.08   0.279    -.1755222    .6088665
        /cut3 |   2.358584   .2058945    11.46   0.000     1.955038     2.76213
-------------+----------------------------------------------------------------
c_observador  |
    var(_cons)|    .298663   .1128584                      .1424067    .6263721
------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 209.30        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 0.34 | degrees of freedom: 1 | p-value: 0.56

```
Likelihood-ratio test                                 LR chi2(1)  =      0.34
(Assumption: m_null nested in m_c_fvestime~a)         Prob > chi2 =    0.5602
variable m_c_fvestimenta not found
```

Level 2 predictor variable, controlling by level 1 variables: m_c_fvestimenta

Fitting fixed-effects model:

```
Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3518.3949
Iteration 2:   log likelihood = -3518.3828
```

```
Iteration 3:   log likelihood = -3518.3828

Refining starting values:

Grid node 0:   log likelihood = -3421.1546

Fitting full model:

Iteration 0:   log likelihood = -3421.1546  (not concave)
Iteration 1:   log likelihood = -3418.7952  (not concave)
Iteration 2:   log likelihood = -3417.0029
Iteration 3:   log likelihood =  -3411.571
Iteration 4:   log likelihood = -3411.5548
Iteration 5:   log likelihood = -3411.5548

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(35)     =      31.92
Log likelihood = -3411.5548                     Prob > chi2       =     0.6174
------------------------------------------------------------------------------
          precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
                     |
       c_fvestimenta |
             Casual  |  -.1322944   .2906879    -0.46   0.649    -.7020322    .4374434
                     |
          taxi_color |
               Blue  |  -.1080163   .1969918    -0.55   0.583     -.494113    .2780805
              Beige  |  -.0596116   .2588657    -0.23   0.818    -.5669791    .4477559
              White  |  -.1583665   .1396624    -1.13   0.257    -.4320998    .1153669
               Gray  |  -.1421379   .1576251    -0.90   0.367    -.4510775    .1668016
              Black  |  -.0721468    .146493    -0.49   0.622    -.3592678    .2149742
             Silver  |   .0353734   .1461677     0.24   0.809    -.2511101    .3218569
                Red  |  -.0259177   .1695333    -0.15   0.878    -.3581969    .3063614
              Green  |  -.1019844   .2637443    -0.39   0.699    -.6189138    .4149449
        Other color  |   .1706129   .2494713     0.68   0.494    -.3183419    .6595677
            No data  |   .2230826   .7586028     0.29   0.769    -1.263751    1.709917
                     |
          taxi_marca |
          CHEVROLET  |    .290716   .3389143     0.86   0.391    -.3735438    .9549759
             DAEWOO  |    .041703   .4249134     0.10   0.922    -.7911118    .8745179
              HONDA  |   .1220421   .4866954     0.25   0.802    -.8318633    1.075948
            HYUNDAI  |   .0553328   .3342348     0.17   0.869    -.5997554    .7104211
                KIA  |   .4082394   .3328405     1.23   0.220     -.244116    1.060595
              MAZDA  |   .1676181   .4452102     0.38   0.707    -.7049779    1.040214
         MITSUBISHI  |   .3985629   .4238263     0.94   0.347    -.4321213    1.229247
             NISSAN  |   .2341375   .3222082     0.73   0.467     -.397379    .8656539
            RENAULT  |   .2066461   .5476217     0.38   0.706    -.8666727    1.279965
             SUZUKI  |   .1432638    .463793     0.31   0.757    -.7657537    1.052281
             TOYOTA  |   .3502247   .3227268     1.09   0.278    -.2823082    .9827576
         VOLKSWAGEN  |   .3866706   .3717921     1.04   0.298    -.3420285     1.11537
        Other brand  |   .2200705   .3429676     0.64   0.521    -.4521337    .8922747
            No data  |  -.0304598   .8226696    -0.04   0.970    -1.642863    1.581943
                     |
           taxi_anho |
               2010  |   .0303534   .1750541     0.17   0.862    -.3127464    .3734532
               2011  |   .0161447   .1286721     0.13   0.900    -.2360479    .2683374
               2012  |   .0913194   .1358624     0.67   0.501     -.174966    .3576048
               2013  |   -.049791    .134454    -0.37   0.711    -.3133161     .213734
               2014  |    .218353   .1374857     1.59   0.112     -.051114    .4878201
               2015  |   .0843021   .1311108     0.64   0.520    -.1726704    .3412745
               2016  |   .2886725   .1345343     2.15   0.032     .0249901     .552355
                     |
     c_dÍadeobservac |
             Sunday  |  -.0946599   .0746385    -1.27   0.205    -.2409488    .0516289
                     |
         c_horariogr |
Mid-afternoon (12-16)  |    .269039   .0989349     2.72   0.007       .07513    .4629479
    Afternoon (16-20)  |   .1832995   .1037138     1.77   0.077    -.0199759    .3865748
---------------------+--------------------------------------------------------------
               /cut1 |   -3.55388   .4158078    -8.55   0.000    -4.368848   -2.738911
               /cut2 |   .5938423   .4039279     1.47   0.142    -.1978419    1.385526
               /cut3 |   2.749128   .4073855     6.75   0.000     1.950668    3.547589
---------------------+--------------------------------------------------------------
c_observador         |
           var(_cons)|   .3178628   .1201868                          .1514931      .66694
```

```
-----------------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 213.66        Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 31.92 | degrees of freedom: 35 | p-value: 0.62

```
Likelihood-ratio test                                  LR chi2(34) =     31.80
(Assumption: m_c_fvestime~a nested in m_c_fvestime~1) Prob > chi2 =    0.5760
```

Appendix I: Results of multilevel ordinal logistic regressions using variables at levels 1 and 2

Full L2 model

Level 2 predictor variables, no level 1 controls (full L2)

Level 2 all predictor variables: c_fsexo c_fedad c_ftez c_facento c_fimagen c_fvestimenta .

. meologit precio_cat $L2_list || c_observador:,

```
Fitting fixed-effects model:

Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3462.4461
Iteration 2:   log likelihood = -3462.1783
Iteration 3:   log likelihood = -3462.1782

Refining starting values:

Grid node 0:   log likelihood = -3430.7241

Fitting full model:

Iteration 0:   log likelihood = -3430.7241  (not concave)
Iteration 1:   log likelihood = -3424.8129  (not concave)
Iteration 2:   log likelihood = -3422.6231
Iteration 3:   log likelihood = -3421.7989
Iteration 4:   log likelihood = -3421.5447
Iteration 5:   log likelihood = -3421.5403
Iteration 6:   log likelihood = -3421.5403

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
                                                              avg =      221.1
                                                              max =        316

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(7)      =      18.07
Log likelihood = -3421.5403                     Prob > chi2       =     0.0117
--------------------------------------------------------------------------------
   precio_cat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+-----------------------------------------------------------------
              |
      c_fsexo |
      Female  |  -.3092238   .1948995    -1.59   0.113    -.6912197    .0727722
              |
      c_fedad |
       19-64  |  -.4697757    .239927    -1.96   0.050    -.9400239    .0004725
         65+  |  -.2683313   .2384493    -1.13   0.260    -.7356835    .1990208
              |
       c_ftez |
       White  |   .5178298   .1949954     2.66   0.008     .1356458    .9000137
              |
    c_facento |
   Foreigner  |   .3374812   .1949202     1.73   0.083    -.0445553    .7195177
              |
    c_fimagen |
       Tacky  |   .1810937   .1948801     0.93   0.353    -.2008642    .5630517
              |
c_fvestimenta |
```

```
      Casual  |  -.1629161    .194864    -0.84   0.403    -.5448426    .2190104
--------------+----------------------------------------------------------------
        /cut1 |  -3.731554   .2761893   -13.51   0.000    -4.272875   -3.190233
        /cut2 |   .3946438   .2568939     1.54   0.124    -.1088591    .8981467
        /cut3 |   2.536761    .261917     9.69   0.000     2.023413    3.050109
--------------+----------------------------------------------------------------
c_observador  |
    var(_cons)|   .1328933   .0538501                        .06006    .2940499
-------------------------------------------------------------------------------
LR test vs. ologit model: chibar2(01) = 81.28         Prob >= chibar2 = 0.0000
```

Test if parameters of equation are equal to zero:

chi2: 18.07 | degrees of freedom: 7 | p-value: 0.01

```
.         lrtest m_all_L2 m_null

Likelihood-ratio test                                 LR chi2(7)  =     12.17
(Assumption: m_null nested in m_all_L2)               Prob > chi2 =    0.0952
```

Full L1 and L2 model

Level 2 predictor variables with level 1 controls full L1 and L2)

__Level 1 & 2 all predictor variables: taxi_color taxi_marca taxi_anho c_dÍadeobservac c _horariogr & c_fsexo c_fedad c_ftez c_facento c_fimagen c_fvestimenta __

. meologit precio_cat $L1_list $L2_list || c_observador:,

```
Fitting fixed-effects model:

Iteration 0:   log likelihood = -3533.0037
Iteration 1:   log likelihood = -3449.3762
Iteration 2:   log likelihood = -3448.9782
Iteration 3:   log likelihood = -3448.9781

Refining starting values:

Grid node 0:   log likelihood =  -3416.435

Fitting full model:

Iteration 0:   log likelihood =  -3416.435  (not concave)
Iteration 1:   log likelihood = -3410.6079  (not concave)
Iteration 2:   log likelihood = -3408.4139
Iteration 3:   log likelihood = -3406.4287
Iteration 4:   log likelihood = -3405.7705
Iteration 5:   log likelihood = -3405.7623
Iteration 6:   log likelihood = -3405.7623

Mixed-effects ologit regression                 Number of obs     =      3,537
Group variable:    c_observador                 Number of groups  =         16

                                                Obs per group:
                                                              min =        120
```

	avg =	221.1
	max =	316
Integration method: mvaghermite	Integration pts. =	7
	Wald chi2(41) =	47.99
Log likelihood = -3405.7623	Prob > chi2 =	0.2106

precio_cat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
taxi_color						
Blue	-.1108916	.1969381	-0.56	0.573	-.4968832	.2750999
Beige	-.0599888	.258878	-0.23	0.817	-.5673803	.4474027
White	-.1593696	.1396548	-1.14	0.254	-.433088	.1143489
Gray	-.1442718	.1576034	-0.92	0.360	-.4531688	.1646252
Black	-.0728195	.146461	-0.50	0.619	-.3598777	.2142387
Silver	.0327723	.1461453	0.22	0.823	-.2536673	.3192119
Red	-.0290614	.1695155	-0.17	0.864	-.3613057	.3031829
Green	-.1054162	.2636154	-0.40	0.689	-.622093	.4112605
Other color	.1638511	.2493781	0.66	0.511	-.3249211	.6526232
No data	.2325494	.7584762	0.31	0.759	-1.254036	1.719135
taxi_marca						
CHEVROLET	.2870713	.3388103	0.85	0.397	-.3769847	.9511274
DAEWOO	.0344627	.4247331	0.08	0.935	-.797999	.8669243
HONDA	.1253713	.486628	0.26	0.797	-.828402	1.079145
HYUNDAI	.0485181	.3341415	0.15	0.885	-.6063871	.7034233
KIA	.4050375	.3327482	1.22	0.224	-.247137	1.057212
MAZDA	.1672014	.4452328	0.38	0.707	-.7054388	1.039842
MITSUBISHI	.3895476	.4237213	0.92	0.358	-.4409309	1.220026
NISSAN	.2291004	.3220981	0.71	0.477	-.4022003	.8604012
RENAULT	.1988176	.5474032	0.36	0.716	-.8740731	1.271708
SUZUKI	.1440167	.46368	0.31	0.756	-.7647793	1.052813
TOYOTA	.3452292	.3226164	1.07	0.285	-.2870874	.9775457
VOLKSWAGEN	.3819278	.3716508	1.03	0.304	-.3464944	1.11035
Other brand	.2168799	.3428366	0.63	0.527	-.4550675	.8888273
No data	-.0453888	.8224363	-0.06	0.956	-1.657334	1.566557
taxi_anho						
2010	.0314828	.1750049	0.18	0.857	-.3115205	.3744862
2011	.0193968	.1286735	0.15	0.880	-.2327986	.2715922
2012	.0909676	.1358647	0.67	0.503	-.1753223	.3572575
2013	-.0516814	.1343939	-0.38	0.701	-.3150886	.2117258
2014	.2191373	.1374532	1.59	0.111	-.050266	.4885406
2015	.0859443	.1311351	0.66	0.512	-.1710757	.3429643
2016	.2874936	.1344955	2.14	0.033	.0238873	.5510999
c_díadeobservac						
Sunday	-.0933224	.0748034	-1.25	0.212	-.2399344	.0532896
c_horariogr						
Mid-afternoon (12-16)	.266493	.0989388	2.69	0.007	.0725766	.4604095
Afternoon (16-20)	.1754652	.1024135	1.71	0.087	-.0252616	.3761921
c_fsexo						
Female	-.3240642	.2029094	-1.60	0.110	-.7217593	.0736309
c_fedad						
19-64	-.4747281	.2494493	-1.90	0.057	-.9636397	.0141836
65+	-.2626006	.2481166	-1.06	0.290	-.7489003	.223699
c_ftez						
White	.4994238	.203093	2.46	0.014	.1013687	.8974788
c_facento						
Foreigner	.3688159	.203042	1.82	0.069	-.029139	.7667709
c_fimagen						
Tacky	.2200524	.2031334	1.08	0.279	-.1780819	.6181866
c_fvestimenta						
Casual	-.1300543	.2027549	-0.64	0.521	-.5274466	.267338
/cut1	-3.365935	.4511409	-7.46	0.000	-4.250155	-2.481715
/cut2	.781952	.4407442	1.77	0.076	-.0818907	1.645795
/cut3	2.937418	.444202	6.61	0.000	2.066798	3.808038
c_observador						
var(_cons)	.1443103	.0582182			.0654489	.318194

LR test vs. ologit model: chibar2(01) = 86.43 Prob >= chibar2 = 0.0000

Test if parameters of equation are equal to zero:

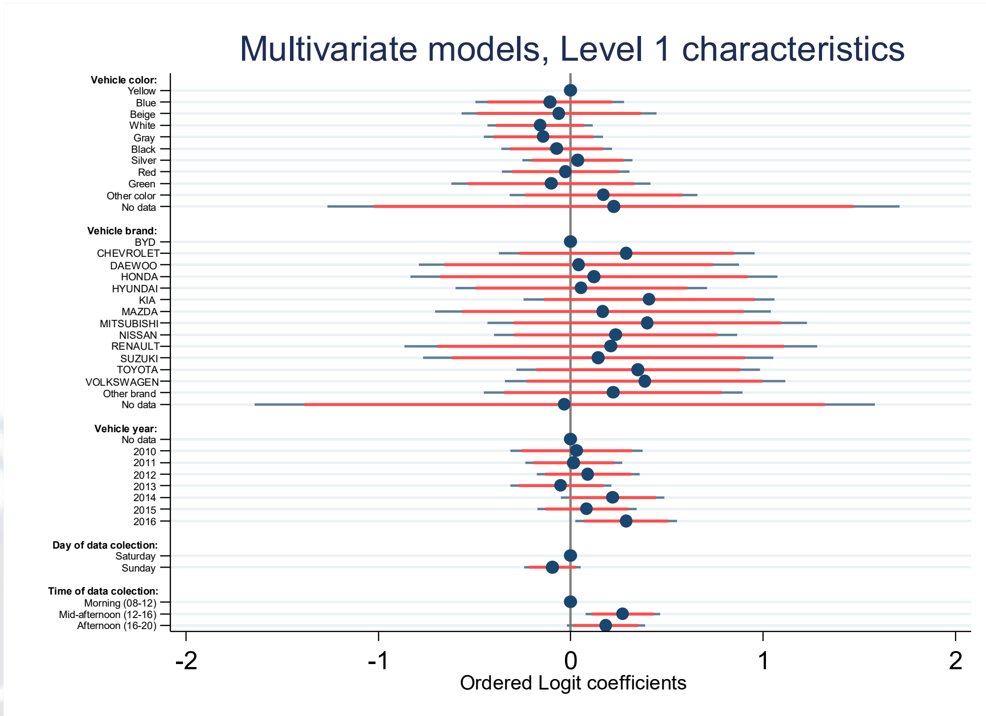
chi2: 47.99 | degrees of freedom: 41 | p-value: 0.21

```
.         lrtest m_all_L1_L2 m_all_L2

Likelihood-ratio test                                 LR chi2(34) =     31.56
(Assumption: m_all_L2 nested in m_all_L1_L2)          Prob > chi2 =    0.5880


-------------------------------------------------------
        name | command      depvar       npar  title
-------------+-----------------------------------------
      m_null | meologit     precio_cat      4
m_taxi_color | meologit     precio_cat     15
m_taxi_marca | meologit     precio_cat     19
 m_taxi_anho | meologit     precio_cat     12
m_c_dÍadeo~c | meologit     precio_cat      6
m_c_horari~r | meologit     precio_cat      7
    m_all_L1 | meologit     precio_cat     43
   m_c_fsexo | meologit     precio_cat      6
m_c_fsexo_L1 | meologit     precio_cat     45
   m_c_fedad | meologit     precio_cat      7
m_c_fedad_L1 | meologit     precio_cat     46
    m_c_ftez | meologit     precio_cat      6
 m_c_ftez_L1 | meologit     precio_cat     45
 m_c_facento | meologit     precio_cat      6
m_c_facent~1 | meologit     precio_cat     45
 m_c_fimagen | meologit     precio_cat      6
m_c_fimage~1 | meologit     precio_cat     45
m_c_fvesti~a | meologit     precio_cat      6
m_c_fvesti~1 | meologit     precio_cat     45
    m_all_L2 | meologit     precio_cat     17
 m_all_L1_L2 | meologit     precio_cat     56
-------------------------------------------------------
```

Appendix J1: Coefficients associated to characteristic at level 1 (vehicle and data collection characteristics)



Effect of vehicle characteristics

- No significant effect of color brand and year of the vehicle

Data collection (experiment) characteristics:

- Prices on Sunday (Domingo) are lower than prices offer to the client on Saturdays
- Prices offers are significantly higher on mid-afternoon and afternoon shifts than in the morning shift.