Information Frictions on a Marketplace for Services

Preliminary

February 28, 2018

*Milena J Petrova**

Toulouse School of Economics, JJ Laffont Chair on the Digital Economy

JEL classification: L81, D82, D83

Abstract

I study the effect of information frictions in an online services market, where capacity constraints and differentiation play an important role. I construct a structural model of buyer search and matching with sellers who have an unobserved willingness to provide the service, which in my model is represented by a private reservation price. I recover the distribution of the reservation price, the effects of match-related characteristics on the match output, the buyer search cost bounds and their coefficients. This paper contributes to the young and growing empirical literature on online service marketplaces by proposing for the first time a structural model that studies the effects of information frictions on match formation at the individual interaction level. I evaluate the efficiency loss due to information frictions, both in terms of missed opportunities to match and inefficient assignment of match partners, by using the estimated primitives to simulate a number of counterfactuals and to calculate the total search cost relative to the generated surplus.

*I thank Marc Ivaldi, Jacques Cremer, Shruti Sinha, Christian Bontemps, Andrey Fradkin, Jorge Flores Acosta and Laura Lasio for their constructive feedback and support. I am also thankful to the participants of the 2016 UPF Student Seminar, the 2016 TSE Applied Micro Seminar, 2017 TSE IO Seminar, the 2017 TSE Student Workshop, and the UCL and UC3M ENTER Seminars for their insightful comments. Lastly, I am grateful to Stoyan Irinkov and the MaistorPlus team for the generously provided time and data.

1 Introduction

Although e-commerce is traditionally dominated by goods marketplaces like Amazon and eBay, in recent years service platforms have gained prominence, facilitating the interaction between service clients and professionals. The two most well-known such platforms are AirBnB and Uber, pioneers of the so-called peer-to-peer or asset sharing economy. Among others, Upwork (formally, Elance and oDesk, for IT services), Thumbtack (any type of local services, for example music lessons or party DJ-ing), and TaskRabbit (low skill domestic services) represent the new wave of online service marketplaces.

Both online and offline, two important features set service markets apart form goods markets. Firstly, physical capacity plays a very important role in service markets. As an example, consider a home owner who would like to hire a roof repairman to clean the gutters on his roof top. He may already know someone, or he may search for a contact in the the yellow pages and by asking around. The home owner would have to verify if the the repairman is available, as service workers have limited physical capacity and if they are working on one job that precludes them from working on another. The repairman being available does not guarantee that he will take the job at any price, because prevents him form taking another job during the same working hours. Hence, he must be compensated for the potential missed opportunities. Capacity issues are not as relevant in goods markets, where sellers can produce multiple units and supply multiple buyers at the same time. The second important difference between goods and service markets is that in service markets both buyers and sellers are differentiated and can be ranked relative to others on the same side of the market. Continuing the example from before, the roof repairmen could come with different levels of experience, but also the properties in question could be in a better or worse state, thus requiring different amounts of effort and time, and therefore having different profitability. While seller/product differentiation is common in goods markets, the sellers are not concerned with the features of the buyers other than their willingness to pay. Thus, it is important to consider these two issues when I think about how buyers and sellers meet and transact in service markets.

The growth in online service platforms (and the data they generate) has sparked the interest of economists and the associated work is concerned with how buyers and sellers interact on these platforms, given the above described features of service markets. Two recent applied studies are Fradkin (2017) for AirBnB and Horton (2017) for oDesk. The main interest of this work is studying buyer-seller interactions in what is a very frictional environment. As the online service markets successfully solve the coordination friction - allowing interested parties to learn about each other and maintaining public profiles - the main question is the persistence of asymmetric information with respect to the seller's availability and willingness to provide the service. On oDesk, no prior information on availability is provided to the buyer and he must contact sellers one by one with the job offer, being rejected 45 percent of the cases. A self-reported availability feature has limited success with sellers overstating their true capacity and unwilling to modify their availability continuously. On AirBnB, the buyers browse only properties that are available but their requests are still rejected in 42 percent of cases, because sellers have different willingness to provide the given service. AirBnB has attempted to solve this issue by an instant booking feature, but many hosts have resisted it because it allows "bad" buyers to book their property. Both Fradkin (2017) and Horton (2017) show that the information friction affects the number of matches formed on the platform negatively: the buyers are likely to give up searching for a service provider after a rejection.

The focus of this work is the information friction created by private seller availability and willingness to transact with application the marketplace *MaistorPlus* for home improvement projects.[1,2] The platform is relatively young as it was founded in 2012. Between January 2013 and June 2015, we have a total of 4,542 clients that posted jobs, or on average 120 jobs posted each month. The total proposed budget of these jobs is 12.6 million Euro and the number of active subscribing professionals is 823.

In a market with many small service providers, the platform solves the coordination friction by bringing the interested parties together and maintaining seller profiles. Furthermore, all subscribing sellers are notified of the job post, and only those who are available message the client. The seller willingness to provide the service remains private information, as home service professionals typically have job prospects from a number sources such as newspapers, yellow pages, and past client referrals. In this framework, the outside potential employment determines the seller *reservation price* for a given project on the marketplace. The friction arises because this information is private and the buyer has to contact or *search* the sellers in order to discuss and negotiate offers which are based on the reservation price. Because these contacts are costly, the buyer must optimize his search activity. The goal of my work is to develop a structural model that describes this interaction. The estimated primitives allow us to construct counterfactual scenarios and to quantify the effect of the different information frictions in this market, those that have already been resolved by the marketplace and the remaining unobserved reservation price.

I analyze the buyer-seller interaction on MaistorPlus in the framework of two-sided matching: differentiated agents with limited capacity compete with each other to form a stable match with an agent on the other side of the market. I construct, identify and estimate a two-stage search and matching model. To show identification, I rely on the results of Lewbel (2010) and Bontemps, Magnac and Maurin (2012). In the first stage, the buyer observes all available sellers and their characteristics, but not exactly how attractive the project is relative to the sellers' outside employment opportunities. Asymmetric information with respect to the seller willingness

[1] Other similar platforms in the US are Thumbtack, Angie's List, Houzz, Fixr. In the UK, there are RatedPeople, MyBuilder, and Home Jane. In Germany, there is MyHammer, Blau Arbeit, and Haus Helden. In France, there is Travaux. In each of these countries the market is still very decentralized, with multiple platforms of different design.

[2] Through out the paper, I use interchangeably the following terms: (service) buyer and client; (service) seller and professional; project and job.

to provide the service is modeled as an ex-ante random *reservation price* that enters additively in the match surplus. The buyer decides which sellers to contact in a directed, simultaneous sample search manner, paying a constant search cost for each contact. In the *Second stage* of the game, the contacted sellers make competing offers to form a match with the buyer, where their offers are a function of the match output and reservation price. The model allows us to recover the distribution of the reservation price, to estimate the effect of match-related characteristics on the match surplus, and to recover the search cost bounds and their coefficients.

My work is relevant to understanding and potentially alleviating the market failure arising from information frictions in online service markets, with implications for market design. I am the first to take a structural approach to describe search and matching at the individual interaction level in the literature on online service marketplaces, and this approach allows me to construct and evaluate counterfactual settings.[3] First, I take a step back and consider the information frictions that the platform is already resolving: making multiple sellers' information available to the buyer, eliciting the seller availability, and maintaining informative profiles of the sellers. Next, I consider the remaining information friction - the seller private reservation price - by simulating a full information outcome. This can be achieved either by lowering search costs to zero, or by incentivising sellers to report their reservation price truthfully. I quantify two measures of efficiency: the probability of a match and the efficiency of the match assignment. The model also allows us to quantify the amount of the search cost relative to the generated match output.

The results from the counterfactual analysis are the following. In a completely *random* environment, the buyer does not observe the seller characteristics nor their availability. As a baseline, I consider the buyer contacting a single seller this way: *random one contact*. This represents the real-life situation of the client asking around for a plumber and being given the contact information of only one such professional. The match probability predicted by my model is only 2 percent. If the buyer is able to get the information on multiple sellers, what I call *random search*, the match probability goes up to 6 percent.[4] In the *random available search* scenario, the buyer knows which sellers are available but does not observe the public profile. He contacts multiple available sellers and the match probability increases drastically - it is now 42 percent. Next, I consider the *directed available search* environment, where the client can benefit from the marketplace maintaining a public profile of the sellers which would direct his search effort. This scenario represents how the platform is actually structured at the moment, and the predicted match probability is 53 percent. Lastly, under the *frictionless* scenario, where either the client observes the seller reservation price or searching is costless, the match probability is 59 percent. Further improvement in the match outcome is not possible as the service

[3] Cullen and Farronato (2014) also prose a stuctural model of buyer-seller interactions on the services marketplace TaskRabbit , however their is an aggregate frictional labor market model.

[4] To make comparison easier, under *random search* the client contacts the same number of sellers that the client would contact in the scenario that represents how the platform actually works (described below, *directed available search*). This also applies to the next scenario, *random directed search*.

providers remain reliant on employment opportunities from outside of the platform.

The average match surplus generated in each of the described scenarios grows as more information is made available, except under the frictionless scenario where it is lower because the extra matches that are created have lower surplus and bring the average down. Looking only at those matches that are formed under the frictionless and the directed available search scenarios, the matches formed under the frictionless scenario may be more efficient (higher surplus) because the buyer may be matched with a better seller when he perfectly observes the seller reservation prices. I can see this for only 3 percent of the matches formed under both scenarios, and the increase in surplus is only 1 percent. I conclude that removing the last information friction affects the match probability in a more substantial way than the match efficiency. Lastly, search costs are estimated at 30 percent of generated match surplus.

The paper has the following structure. The next subsection puts my work in the context of the relevant economic literature. In Section 2, I describe how the platform works, and present the data and reduced form evidence that supports the modeling choices. In Section 3, I describe and solve for the equilibrium of the two-stage search and matching game. Section 4 demonstrates the identification of the primitives of interest - the reservation price distribution, parameters of the match output, and search cost bounds. Section 5 details the steps I take to estimate the model. Section 6 contains the results of the estimation, and Section 7 contains the counterfactual analysis. I conclude in Section 8.

Related literature

In terms of theory and empirical strategy, my work is related to the literature on two-sided matching with transfers, although I take a different approach to modeling search. I also see similar a concept - unobserved willingness to transact - discussed in the macroeconomics labor literature, where search and match formation are modeled at a more aggregate level. As this paper is applied, it is most closely related to other work that studies information frictions in online service environments. The two closest such studies are Horton (2017) and Fradkin (2017), both using proprietary data to perform reduced form exploration of their respective service marketplace.

The theoretical and empirical literature on two-sided matching with transfers is concerned with studying markets where the goods (or services) to be allocated are heterogeneous and indivisible (Roth and Sotomayor (1992)). The two main empirical frameworks for estimating the fundamentals of two-sided matching markets are Choo and Siow (2006) and Fox, Hsu and Yang (2015), but they do not consider information frictions on the part of the market participants.[5] The framework of Choo and Siow (2006) imposes a structure on the unobserved

[5] For an excellent survey of the empirical literature on matching, see Chiappori and Selanie (2016).

error term in order to identify coefficients on the match surplus. Fox et al (2015), on the other hand, identify the unobserved error term by assuming a certain observable characteristic that enters the match surplus, a method similar to the special regressor of Lewbel (2012). The identification of the *Second stage* of the model is very close to Fox et al (2015), the main difference being that the random component of the match surplus is private information in the *First stage* of the game. Two conceptual, but not technical, differences between this literature and the model are the following. Firstly, traditional matching models treat the market as centralized: agents who do not match on it are assumed to simply remain unmatched rather than saving themselves to match on another market. Secondly, the random component error term is treated as a taste heterogeneity unobserved to the econometrician while in the specific setting I believe this term is a private seller reservation price due to the sellers multi-homing.

In more recent work, the empirical two-sided matching literature has incorporated search in the presence of incomplete and costly information regarding potential match partners. Chade, Eeckhout and Smith (2017) define *search frictions* to arise when agents do not observe all potential partners and *incomplete (asymmetric) information* when agents do not observe the characteristics of these partners.[6] This is slightly different from the micro search-theoretic literature, where search is typically defined as the costly gathering of information on a characteristic (price, quality) of an agent or a good. Simultaneous search for unobserved prices was first introduced by Stigler (1961), while Chade and Smith (2006) extend the paradigm to allow for ex-ante heterogeneous, stochastic rewards. This is the approach that I take in the model, with the following modification: the expected outcome from the search process, the transaction price, is determined endogenously through a competitive process in the matching stage. Modeling search friction in this simultaneous way appears in other settings with heterogeneous options such as Chade, Lewis and Smith (2014) for college admissions and Kircher (2009) and Galeniakos and Kircher (2009) for labor markets.

The seminal paper of McCall (1970) lays the foundation for models of sequential search, where the searcher's optimal strategy is fully summarized by a reservation wage above which he should stop searching. This framework has become the fundamental building block for macroeconomic models of the labor market and is also extensively used in the microeconomic two-sided matching models with search. Assortative matching and random search was first introduced by Shimer and Smith (2000) in a continuous-time model, where potential partners of unknown quality arrive randomly and the search cost is impatience. Postel-Vinay and Robin (2002) are the first to explore the structural identification and estimation of this framework. While my work shares the preoccupation with incomplete and costly information, I am motivated by a setting where the set of potential partners and some characteristics are ex-ante observable, therefore search is directed by these characteristics, rather than random. Random search is more suitable for settings where the search takes place over a long time horizon and there is uncertainty about candidate arrival, for example in labor markets.

[6]Chade, Eeckhout and Smith (2017) review search and matching theory, as well as the recent contributions combining these literatures.

Another branch of the theoretical search and matching literature postulates directed, sequential search (Eeckhoot and Kircher (2010), Shi (2001)). The searched side of the match sets prices and the searching side is directed in its efforts by these prices, a mechanism is also known as competitive search. The crucial difference with my model is that in my setting search is directed by characteristics which are pre-determined at the start of the game (the agents are ex-ante heterogeneous), and the outcome (the transaction price) is determined in the *Second stage* of the interaction.[7] This is more suitable in my setting, as the agents are unwilling to commit to a price before discussing the details of the job with the buyer, negotiating offers and getting an idea for the level of competition.

In the macroeconomic labor literature, search frictions and matching are modeled in the Diamond-Mortnesen-Pissarides (DMP) framework: the agents are matched by an aggregate matching function, they are identical ex-ante, search frictions are not explicitly modeled although they may be informational, unobserved heterogeneity, congestion, messaging/application costs, or else. A number of empirical studies explore the lack of information about agent availability. Arnosti, Johari and Kanoria (2014) show congestion externalities on both sides of the market arise when agents spend resources to be matched with others that are already unavailable. In a continuous-time equilibrium search unemployment model, Cheron and Decreuse (2016) study information persistence with phantom agents that are already matched but their status is not updated quickly enough.

The lack of information on availability and willingness to transact is a common theme in studies of online matching markets for services. Compared to goods providers, service providers are physically constrained by their fixed capacity or working hours. Fradkin (2017) and Horton (2017) tackle this subject, using proprietary data from the online matching platforms AirBnB and oDesk respectively. A main difference between my model and their work is that I am the first to propose a structural model of the buyer-seller interaction, allows us to construct and evaluate counterfactual scenarios and to quantify the different information frictions that arise in the specific environment.

Horton (2017) studies oDesk, an IT task platform where the buyer looks up and invites sellers to submit offers on a job. The sellers may be unavailable, or not willing to transact at the price that the buyer is proposing. In the main empirical analysis of the paper, Horton (2017) demonstrates that rejection leads to a decrease in the probability that a job is eventually filled, which suggests that finding ways to reveal more information about seller availability would improve the matching outcome and welfare. The marketplace designer could direct buyers to sellers who are more likely to have free capacity, a market intervention that would be simple to implement as a feature on the online marketplace. Indeed, the platform introduces a signaling feature that makes it possible for the sellers to publicly state their availability. He finds that when workers state they are free, they

[7] There is also a small literature on pre-matching investment in characteristics. For references, refer to Chade, Eeckhood and Smith (2017).

receive more invitations, are more likely to apply to those jobs, quote lower hourly rates, and are more likely to be hired. At the same time, the signaling feature was only partially successful in eliciting the true availability of the sellers: most workers never change their status after it was set once, typically to the highest availability. It is unclear weather this is because there is no change in their availability or because they are reluctant to reveal it.[8]

The difference between the matching technology of oDesk and MaistorPlus is mainly in the rules of buyer-seller interaction. On oDesk, the buyers look up and invite sellers, with no information on whether they are available. While Horton (2017) does not model explicitly buyer search costs, they are presumably associated with the time and effort spent on looking up, messaging the sellers, and waiting for the offer of those available. As the first step in the process on MaistorPlus, the platform notifies all sellers of the job post, and the available sellers write a message to the buyer. My counterfactual analysis indicates that this feature contributes greatly to improving the match probability on the platform. The main empirical analysis in Horton (2017) focuses on buyers sending a single (early) invitation, which represents 30 percent of the data. The analysis does not cover buyers sending multiple invitations to counter rejections due to unobserved availability or willingness to transact, which is fully incorporated in my structural model. Another missing piece in his analysis is modeling how the market clears and how prices are set. From the description of the interaction, I learn that the clients propose prices (fixed or hourly). Whether a seller will respond to the client depends on the price, but also the sellers may respond with a price that they find to be more appropriate. These exchanges, however, do not constitute commitment: these is still the possibility for rejection, and the process of making, comparing and accepting offers not considered in more detail.

Despite the design differences, on oDesk there are information frictions very similar to those I observe on MaistorPlus. About half of all first invitations on oDesk are rejected, with the most common reasons being that ”too busy” (48 percent) or ”not interested in this project” (29 percent), demonstrating the importance of limited capacity of sellers and the heterogeneity of projects. Lower availability sellers apply with a higher price, a sign that they look to be compensated for missed employment opportunities rather than some physical cost. Indeed, there is substantial heterogeneity in hours worked on oDesk and in messages indicating availability on MaistorPlus, an activity pattern which suggests outside employment is an important source of occupation for sellers.

Using data from AirBnB, Fradkin (2017) studies transaction costs, the role of marketplace design in reducing them, and the potential for further improvement in search and matching. Buyer-seller interaction on AirBnB is set-up in the following way: the guests search for properties using the platform search tool, often using filters, and the displayed results show only properties that are available for that time period. Upon being contacted by

[8] More specifically, 45 percent of workers set their signal to being available full time. Most workers never changed the signal, and the majority of those who never changed their signal set it to being available full time (Horton (2017)).

the guest, the property host may reject the request despite the property being available. Fradkin (2017) classifies rejections as stale vacancies (15 percent), congestion (8 percent), and selection due to characteristics of the seller or the trip (19 percent). Sellers may e *de facto* unavailable, in the case of stale vacancies and congestion, or hosts may preffer to wait for another alternative with better characteristics down the line. Fradkin (2017) defines rejection due to host preferences as the main information friction in this set-up: communication is costly and leads to delay and uncertainty, and rejection makes clients less likely to message another host and more likely to leave the platform. In my model, the search cost is similarly modeled as the client's effort on discussing over the phone, meeting the sellers at this home, negotiating and comparing offers between the contacted sellers.

In the main part of his analysis, Fradkin (2017) constructs and estimates a searcher choice and host rejection model, which could also be interpreted as structural under a number of assumptions with different plausibility. The search behavior on the platform is complex and Fradkin (2017) finds it difficult to reconcile with a specific theoretical model. Instead, he estimates a discrete choice model to predict the searcher's decision of whether and whom to contact. A second model of host screening and rejection accounts for trip and guest characteristics, but does not consider how prices are set and the host's expectations over the flow and quality of future guests arriving on the platform. Fradkin (2017) does not clarify how prices are set on AirBnB, and potentially the hosts could be waiting for higher valuation clients to arrive closer to the date of the service, or for trips with better characteristics. It would be interesting to consider whether sellers would still reject less attractive stays if a higher price were offered. The MaistorPlus environment is very flexible in how prices are set, for better or for worse, as it allows for prices to be determined with respect to job characteristics and the seller reservation price.

A main difference between these studies and my paper is that I take a structural approach that allows us to construct and evaluate counterfactual scenarios in order to evaluate the effect of the unobserved seller reservation price. Even so, both Horton (2017) and Fradkin (2017) are able to quantify some welfare effects related to the information friction in their respective settings. Horton (2017) does a "back-of-the-envelope" evaluation of the signaling feature introduced by oDesk on buyer behavior. A structural model of this environment could also look at how seller price offers and signal are co-determined with their reservation prices. Fradkin (2017) uses his predictive model to evaluate two scenarios. In a pre-AirBnB scenario with no indication on seller availability, no filtering options, and no search ranking algorithm, there is a clear increase in seller rejections and buyer exit. In an improved AirBnB scenario, the ranking of search results also incorporates the likelihood of acceptance. Of course, this is not a general equilibrium result as hosts that are more likely to accept buyer requests cannot accommodate all buyers.

2 Data

In this section, I introduce MaistorPlus marketplace and how the different agents, buyers and sellers, interact. I describe the data set and provide some reduced form supporting evidence for the assumptions of my model.

2.1 The marketplace

I work with company data from the MaistorPlus online services marketplace, who are based in Bulgaria and started operating in 2012.[9] The marketplace connects clients to home service professionals. The clients do not pay to post jobs on the marketplace, and the professionals pay a 3-month subscription fee. The fee design affects only the extensive margin (how many professionals have subscribed) and not the intensive margin (the activity of the subscribed professionals). This allows us to consider the seller and buyer usage (sellers sending messages of availability, buyers contacting the sellers) as decisions which are not affected by the marketplace design.[10] The marketplace is additionally financed by advertising.

In the complete sample, I have 4,542 jobs posted on the platform between January 2013 and June 2015. About 7.7 percent of the jobs receive no messages from available sellers, and I believe these to be mistaken/incomplete job posts. The total suggested budget by the client for the remaining 4,192 jobs is 12.6 million euro. On average, each job was sent to 80.6 professionals who were registered in that category of activity. These activity on the job is the following: the avreage number of seller availability messages that the buyer receives is 5.4 and the average number of contacts that the buyer makes is 1.4. This information is available in Table 1.

Within this subsample of 4,192 jobs that have received at least one message indicating availability, the client does not contact any professional in 34.5 percent of the jobs (1,448 jobs). This is very high in comparison to the number of cases in which the client does not receive any message indicating availability (350 jobs). Buyers may decide not to contact anyone for two reasons: search costs and a high reservation price for the buyer. A buyer reservation price cannot be identified separately from other common costs at the level of the job, unfortunately, but I am able to control for them by including a job fixed effect in the estimations. Conditional on contacting at least one seller, the hiring probability increases from 0.27 percent to 0.41 percent.

These descriptive statistics are in line with what I know about search and matching on other service provision platforms. For example, Fradkin (2017)'s data for activity on AirBnB shows that clients who send a contact

[9]http://maistorplus.com/

[10]I do not go into issues pertaining to two-sided markets, such as pricing structure and how it affects membership and usage, as well as the corresponding externalities, because my main interest is match formation with heterogeneous and capacity constrained agents. More information on two-sided markets and the issues pertaining to them can be found in Rochet and Tirole (2006).

view 5.5 percent of all available listings (73 listings), and contact 2.4 listings on average. Overall, 42 percent of all contacts are rejected. On oDesk, clients invite 2 professionals on average to apply to the job they post, the invitation acceptance rate is 55 percent, and the probability that the job opening is eventually filled is 55 percent (Horton (2017)). Cullen and Ferronato (2014) report similar results for TaskRabbit: of all posted tasks, 78 percent receive at least one offer, the average being 2.8 offers per job; 49 percent of tasks result in a match.

Tab. 1: Outcomes at the level of the job

Outcomes	Obs.	Mean	St. Dev.	Min	Max
Jobs with at least one message indicating availability					
Notified sellers	4,192	80.6	46.7	1	401
Available sellers	4,192	5.4	5.5	1	61
Contacted sellers	4,192	1.4	1.7	0	10
Probability of hiring	4,192	0.27	0.44	0	1
Jobs with at least one contact					
Notified sellers	2,744	82.1	50.0	1	401
Available sellers	2,744	5.4	5.4	1	61
Contacted sellers	2,744	2.2	1.6	1	10
Probability of hiring	2,744	0.41	0.50	0	1

In the sample of 4,192 jobs, where the seller receives at least one message indicating availability, I have a total of 22,547 observations of sellers messaging the buyer to indicate availability. The clients contacted 5,911 professionals and hired someone in 1,126 cases. In the sample of 2,744 jobs where the the seller has contacted at least one professional, I have 14,900 observations of sellers messaging the client, the client has contacted 5,911 professionals, and hired someone in 1,126 cases.

When the client posts the job, he provides a textual description, indicates the job category (one of 38 categories such as carpentry, roof repairs, construction, etc), expected start date (one of 8 categories), and proposed budget (one of 14 categories). The frequencies of these different categories can be found in Appendix 1.

There are a total of 823 active professionals in the sample who have sent a message indicating availability at least once. I have the following information about the seller's profile: categories of activity, profile description, references from previous clients and pictures from past projects. Summary statistics are provided in Table 2.

Some characteristics of the professionals are measured at the moment that a particular job is posted. I observe the following two sets of variables at this level of variation: the seller experience and variables related

Tab. 2: Seller fixed characteristics.

	Obs.	Mean	St. Dev.	Min	Max
Active categories	823	4.78	4.57	0	28
References	823	0.14	0.48	0	3
Profile description (chars.)	824	547	496	0	3,645
Profile pictures	823	10.8	24.9	0	490

to the seller's message indicating availability. The seller experience variables are the seller tenure on the marketplace (in months), the total times the seller was hired up to the month the job was posted, and the seller's percent positive reviews, defined as number of positive reviews on all jobs for which the seller was hired.[11] The message-related variables are message length (measured in characters) and the time of the message (measured in hours since the job was posted on the online marketplace). The summary statistics for these variables are presented in Table 3.

2.2 Reduced form evidence

In this section, I present reduced form evidence which supports the modeling choices. More specifically, I show that sellers do have capacity constraints and that the clients perceive them as differentiated, which means that the matching framework is appropriate for modeling interactions on the marketplace; I show that the seller's decision to send a message indicating availability is not exclusively explained by seller or job characteristics, therefore variation of outside demand must play an important role; that there is uncertainty about seller willingness to undertake a project even if he has indicated availability by sending a message; and that buyers experience search costs. I discuss these issues one by one.

2.2.1 Capacity constraints

The seller's availability to provide the service is significantly constrained by the physical time needed to perform the service: the data demonstrates that sellers are not always available and that the probability to form a match goes down in high demand periods. The summary statistics presented in Table 1 indicate that even if about 81 sellers receive the notification about any given job, only about 5.4 of them indicate that they are available by

[11]The percent positive reviews is a reputation or quality measure is defined as in Tadelis (2016). I count missing reviews of completed jobs as non-positive reviews, as it has been demonstrated that clients are reluctant to leave negative feedback.

Tab. 3: Summary statistics for seller characteristics measured at the time of the job posting.

	Obs.	Mean	St. Dev.	Min	Max
Jobs with at least one offer, sellers who made an offer					
Experience					
Percent positive reviews	22,547	0.32	0.40	0	1
Marketplace tenure (months)	22,547	7.41	7.07	0	36
Total times hired	22,547	3.80	7.09	0	46
Message-related					
Message length (chars.)	22,547	239	290	0	11,932
Time of message (hours)	22,547	3.58	14.6	0	576
Jobs with at least one contact, sellers who were contacted					
Experience					
Percent positive reviews	5,911	0.43	0.40	0	1
Marketplace tenure (months)	5,911	11.6	9.92	0	46
Total times hired	5,911	5.55	8.37	0	46
Message-related					
Message length (chars.)	5,911	251	279	0	3,514
Time of message (hours)	5,911	2.12	11.0	0	503

sending the client a message.[12]

For the individual job, higher overall demand leads to a lower number of available sellers and a lower probability to hire someone, even after controlling for the outcomes on the particular job and for the characteristics of the sellers available for that job. I perform two linear regressions to demonstrate this. In the first regression, the dependent variable is the log of *N. available* of availability that the buyer receives. In the second regression, the dependent variable is an indicator of whether someone was hired for a given job, so this is a linear probability model. The main covariate of interest is *Demand activity*, which measures the number of jobs posted on the platform during the month that the particular job was posted. Demand fluctuations on the marketplace also indicate fluctuations outside of the marketplace, as the seasonality dynamics are exogenous and driven by the weather. I also include the number of *Sellers notified*, *Available sellers*, and *Contacts made* for the given job. *Demand activity* and *Sellers notified* can be considered as truly exogenous, while *Available sellers*, and *Contacts made* are outcomes. I also control for job characteristics (category, budget and expected start) and for the date. In the second regression, I add variables measuring the average quality of the professionals: their profile characteristics (pictures, references, categories in which they are active, and the length of their profile description), their experience (number of times hired, marketplace tenure, and percent positive reviews) and their message characteristics (message length and time of offer).

The results are presented in Table 4 below. Both regressions indicate a very significant and high in magnitude effect of the *Demand activity* on the outcome variables. I interpret this effect in the following way: a 100 percent increase in the *Demand activity* leads to a 78 percent decrease in *N. available* and 26 percent decrease in *Pr(Hire)*. Even if the professionals can accommodate higher demand to some extent, they are ultimately constrained by their physical capacity.

2.2.2 Seller availability difficult to predict

With this analysis, I demonstrate that the seller's decision to send a message indicating availability for a given job depends on seller and job characteristics (observed or unobserved), however a large part of the variation remains unexplained. I also show that seller demand and activity in the immediate period prior to the job being posted on the platform are strong predictors, demonstrating that sellers have activity spells: the more messages

[12] Potentially, there are two other reasons why sellers may be unwilling to send a message for each job posted in their categories of work. This may be a result of coordination, similar to a bidding ring. However, there are numerous sellers on the marketplace and they have limited contact opportunities, which significantly lowers their chances for coordination. In addition, sellers do not see the identities of other sellers who message the client, making monitoring difficult. A second reason may be seller messaging costs. A seller messaging cost does not contradict the set-up of the model, except that cut-off for availability of the seller must be at least high enough to rationalize the cost of the message as well. However, discussion with the marketplace owners suggests that sellers use similar message templates, which reduces this cost.

Tab. 4: Available sellers and hiring outcome for a given job.

	N. available	$Pr(Hire)$
Demand activity	-0.778***	-0.263***
	(0.133)	(0.065)
Notified sellers	0.433***	0.019
	(0.025)	(0.016)
Available sellers		-0.016
		(0.020)
Contacts made		0.131*
		(0.025)
Job characteristics fixed effects	Yes	Yes
Date fixed effect	Yes	Yes
Average seller characteristics	No	Yes
R^2	0.63	0.44
N.	2,744	2,744

Significant at: $p < 0.1$: *; $p < 0.05$: **; $p < 0.01$: ***. Regressions estimated on subsample with at least one contact. Robust standard errors in parenthesis. Continuous variables are transformed by taking the natural logarithm. More detailed results are available upon request.

that have sent recently, the more likely they will send a message again.

I work with the full sample of professionals who were notified for job (225,224). In the first specification, I regress the indicator for seller availability on a given job on the following characteristics: seller demand and experience (percent positive reviews, total times hired, platform tenure), date, job and seller fixed effects. I also include variables measuring the seller's recent activity on the platform: demand, availability messages sent, and contacts in the last 7 and 30 days. The results are presented in Table 5.

In the second specification, I regress the seller indicator for a seller being hired for a job given that he is available on the same set of covariates plus the available seller and competitor message and experience characteristics. This regression serves to demonstrate that given that the seller is available, his hiring outcome is not a function of his recent activity on the platform.

Tab. 5: Linear probability models of availability and hiring outcome for a given seller-job.

	$Pr(Available)$	$Pr(Hired \mid Avbl.)$
Demand last 7 days	-0.035***	-0.008
	(0.002)	(0.006)
Availability msgs. last 7 days	0.120***	-0.003
	(0.005)	(0.006)
Contacts last 7 days	-0.003	0.005
	(0.007)	(0.006)
Demand last 30 days	-0.052***	0.005
	(0.002)	(0.006)
Availability msgs. last 30 days	0.079***	0.001
	(0.003)	(0.005)
Contacts last 30 days	0.010***	-0.007
	(0.005)	(0.005)
Seller and competitor experience		
and message characteristics	N/A	Yes
Date, seller, job chars. fixed effects	Yes	Yes
R^2	0.51	0.52
N	225, 224	14,480

Firstly, the seller's probability of indicating availability for a given job decreases with demand, which is once again evidence for the fixed seller capacity. The recent seller activity on the platform - availability messages and contacts - also have significant coefficients. Lastly, I note the amount of unexplained variation in seller availability is quite high - 49 percent - which leads us to conclude that unobserved seller demand from outside of the platform is responsible for this variation.

In the second regression, the recent demand conditions and seller activity on the platform do not appear to be significantly correlated with the seller probability to be hired on the given job. While the platform may use information on recent seller activity to predict which seller is available, to some limited extent, this same information would not be useful to predict how likely that seller would be eventually hired.

2.2.3 Differentiated players

Matching markets are characterized by differentiation and competition among the players who are on the same side of the market. On the side of the sellers, this can be easily demonstrated. Firstly, I examine the likelihood that any available seller is contacted by the buyer. I show that the probability that a seller is contacted depends both on that seller's characteristics and on those of his competitors. Again, I opt for a linear regression to ease the interpretation. The seller's competitors are those other sellers who have also indicated availability for the respective job. I have the following sets of seller characteristics: profile, experience, and message-related. *Profile* characteristics are fixed at the level of the professional, while *experience* and *message* characteristics are measured at the job-seller interaction level. In Table 6 I present the initial regression $C1$ which includes seller and competitor profile characteristics. $C2$ is the same regression with seller fixed effects, and $C3$ has both seller and competitor fixed effects.

$C1$ demonstrates that competitor characteristics are important factors in the client's decision to contact any professional. For example, increasing the number of active categories for the seller's competitors by 10 percent lowers their chance of being hired by 0.5 percent. Adding unobserved seller fixed effects in $C2$ improves the explanatory power of the regression much more than the competitor fixed effects in $C3$. The competitor's message-related variables remain of high magnitude and significance in all regression. The competitor's experience related variables have the correct sign in all specifications but they are not always significant.

I am unable to perform a similar analysis to explicitly demonstrate project differentiation because there is considerable activity off the marketplace that I do not observe. I do not know what projects rival each other at the level of a given available seller, at any point in time.

2.2.4 Private seller willingness to transact

When sellers indicate availability, this does not imply that they have zero outside job opportunities. In this set-up, the seller's willingness to transact is modeled as a random reservation price that enters the match surplus.

Tab. 6: Linear probability model of seller being contacted for a given job.

Pr(Contact)	C1	C2	C3
Profile			
Seller active categories	-0.014**		
	(0.007)		
Competitor active categories	-0.004	-0.002	
	(0.015)	(0.016)	
Seller references	0.015		
	(0.011)		
Competitor references	-0.012	-0.004	
	(0.020)	(0.021)	
Seller profile descr. length	0.003*		
	(0.002)		
Competitor profile descr. length	-0.014***	-0.014**	
	(0.006)	(0.006)	
Seller profile pictures	0.006**		
	(0.003)		
Competitor profile pictures	-0.002	0.002	
	(0.006)	(0.006)	
Experience			
Seller percent positive reviews	0.076***	0.101***	0.102***
	(0.017)	(0.026)	(0.027)
Competitor percent positive reviews	-0.040	-0.038	-0.065
	(0.032)	(0.032)	(0.044)
Seller total times hired	0.033***	-0.006	-0.009
	(0.006)	(0.010)	(0.011)
Competitor total times hired	-0.030***	-0.035***	-0.015
	(0.009)	(0.009)	(0.012)
Seller marketplace tenure	0.006	0.018	0.016
	(0.004)	(0.012)	(0.012)
Competitor marketplace tenure	-0.004	-0.007	-0.028*
	(0.010)	(0.011)	(0.017)
Message			
Seller message length	0.021***	0.022***	0.023***
	(0.003)	(0.004)	(0.004)
Competitor message length	-0.036***	-0.036***	-0.030***
	(0.005)	(0.005)	(0.006)
Seller time of message	-0.086***	-0.079***	-0.082***
	(0.004)	(0.005)	(0.006)
Competitor time of message	0.069***	0.065***	0.059***
	(0.011)	(0.011)	(0.015)
Job characteristics fixed effects	Yes	Yes	Yes
Date fixed effects	Yes	Yes	Yes
Seller fixed effects	No	Yes	Yes
Competitor fixed effects	No	No	Yes
R^2	0.34	0.40	0.43
N	14, 480	14, 480	14, 480

Significant at: $p < 0.1$: *; $p < 0.05$: **; $p < 0.01$: ***. Robust standard errors in parenthesis. Regression estimated on subsample with at least one seller contacted. Continuous variables are transformed by the natural logarithm.

This assumption is supported by the following observations. Firstly, Table 4 shows that the *Pr(Hire)* goes down in high demand periods, even when I control for the number and characteristics of contacted sellers. This is strong evidence for the seller's reservation prices being higher when there are more jobs also outside of the platform.[13]

Secondly, the reservation price is an important component of the match surplus. Were that not the case, I would see the buyer always hiring the most *ex-ante* attractive seller, where sellers are ranked based on their observable characteristics. I use the regression $C3$ from Table 6 to predict the ex-ante ranking of each seller based on the attractiveness of their features, with 1 being the highest rank and corresponding to the most attractive seller. Table 7 shows the average the rank of the hired seller, and the highest and lowest rank sellers that were contacted. While on average the most attractive seller is contacted by the buyer (rank 1.25), usually a lower ranked seller is hired (rank 1.71), and clients hedge their bets by contacting multiple sellers (up to rank 2.92).

Tab. 7: Seller ranks: hired, maximum and minimum contacted

	Obs.	Mean	St. dev.	Min	Max
Highest rank contacted	1,126	1.25	1.10	1	18
Rank of hired seller	1,126	1.71	1.91	1	24
Lowest rank contacted	1,126	2.92	3.40	2	48

2.2.5 Costly, directed and fixed sample search

I argue that the buyer does experience non-trivial search costs associated with contacting the sellers. This step normally involves discussions over the phone, finding time to arrange visits, negotiating and comparing offers from multiple sellers back and forth. If search were costless, I would see the buyer contacting all available sellers. Table 1 demonstrates this not the case: on average, the buyer contacts half of the available sellers. Furthermore, the buyer sees the sellers as differentiated, which directs his search. Table 6 shows that seller characteristics are important: they are informative and create an order of attractiveness among the sellers and direct the buyer's search process.

The assumption of simultaneous search is supported by the following observations on the marketplace. The buyers are allowed and advised to contact multiple sellers at the same time rather than sequentially as this is in their interest for the following two reasons. Firstly, sellers often require to visit the site but their opportunity to do so may differ, therefore it is not a good use of the buyer's time to wait for one seller to visit before arranging

[13] It is unlikely that this effect is due to fluctuating time or material costs needed to perform any given job. If anything, sellers would try to be more efficient, rather than less efficient, in high demand periods so that they can take on a higher number of projects.

a visit with another one. Secondly, early seller availability may expire, sellers may take on another project or their reservation prices may change, which would make it more difficult for the buyer to put sellers in direct competition with each other. Lastly, there are no posted prices in this setting: in their messages, sellers may state hourly wages but they do not commit to them.[14] The transaction price is set in the matching stage of the game, when the buyer communicates with the available sellers and compares their offers. There is a strong incentive for the buyer to search simultaneously and let more sellers compete, as this increases the buyer's expected utility.

3 Model

In this section, I model the interaction between clients (buyers) and providers of services (sellers) on the home services marketplace MaistorPlus. I consider a model of one-to-one matching with transfers (prices) between one buyer and multiple sellers. Consider one buyer indexed by i and N_i available, differentiated sellers indexed by j. N_i is random, depends on individual seller availability from outside of the platform as seen from the regressions in Table 5. Seller availability and reservation prices are defined in the following way. When the seller is not available, he is physically not capable of taking the job because he has already committed to another job during the time period in question. In the case that he is available, he places a value on the potential jobs from clients outside of the marketplace, which is summarized by the reservation price r_{ij}.

For the moment, I consider a single buyer and drop the buyer index i. When the buyer is matched with seller j, the pair create match output f_j and the seller must be compensated for his reservation price r_j, thus the match surplus is $s_j = f_j - r_j$.[15] In the *First stage* of the game, the buyer observes the match output f_j but not the match surplus s_j, because the seller reservation price r_j is private information of the seller. The buyer decides which sellers to search in a directed, simultaneous manner and at a positive and constant marginal cost c.[16]

In the *Second stage*, contacted sellers compete by making offers in order to form a match with the buyer. The equilibrium concept is stability: the match must satisfy individual rationality for each party and must assign the buyer to the seller whose match surplus is highest. Because sellers are differentiated by the match surplus s_j, the buyer would not be willing to pay the same price for a higher surplus seller as for a lower surplus seller. Hence, the assignment mechanism - the manner in which sellers compete to form a match with the buyer - must

[14]This is likely due to common costs being revealed in the search stage, which is something I consider in the **Model** section.

[15]Common costs, such as materials, are assumed to be constant across sellers. I do not consider individual seller uncertainty about common costs because from discussions with the marketplace managers, the buyers procure their materials separately, and these costs do not typically fluctuate.

[16]Even if the sellers are the ones to incur the search cost, Ye (2007) demonstrates that it is passed through to the buyer, similarly to how common costs are passed on in the English auction or Bertrand competition.

take this into account. I assume that the mechanism by which the stable match is reached and which dictates how the surplus is split though the price is the English auction in utility space.

This assumption is suitable for the following reasons. The English auction guarantees that the transaction utility and the respective transaction price constitute an ex-post equilibrium, in the sense that no participant would be willing to change their offer after observing the offers of the others. This is important because the online marketplace does not restrict the interaction between the parties, hence an outcome that is not an ex-post equilibrium (for example, the outcome of a first price auction) would not be realistic in this setting. Furthermore, the English auction format is especially practical because the equilibrium is in dominant strategies. As a result, I do not make assumptions about how much information the sellers have about each other because it is only important that they know their own reservation price.

The search mechanism is similar to the directed search models of Chade and Smith (2006) for fixed sample and Weitzman (1979) for sequential search, where the searching party is directed by the observed characteristics of the searched party. The main difference is that in those models the characteristics to be revealed, the prize, is fixed ex-ante (Weitzman (1979)) or stochastic (Chade and Smith (2006)). In this setting, the prize is the expected outcome of the *Second stage* matching game.

3.1 Second stage

Let's assume that buyer contacts n sellers, and order them by match surplus: $s_1 = f_1 - r_1 \geq ... \geq s_n = f_n - r_n$. Note that the the match surplus s_j, the reservation price r_j and the match output f_j may not follow the same ranking.

The strategies of the players are defined in terms of the utility they are willing to offer to the buyer, similarly to Laffont-Tirole (1993): $u_j = s_j - p$. The individual rationality constraint for the buyer is $IR_j^b : u_j = s_j - p \geq 0$ when transacting with seller j, and for seller j it is $IR_j^s : v_j = p - r_j \geq 0$. The English auction works in the following way. The auctioneer starts from an utility offer of zero and raises it. The sellers remain in the auction while they agree to the offer, and the game ends when only one seller remains. The transaction utility is that at which the second-last seller drops out of the game. The players' weekly dominant strategies are to remain in the game up to the point they are indifferent (Vickrey (1961)). In other words, player j with match surplus $s_j = f_j - r_j$ remains up to the utility offer $u_j = s_j$ and drops out afterwards.

The game can be summarized by the following three cases and their respective outcomes:[17]

[17] In the case that $n_i = 1$, I only have the two last cases.

1. $0 < s_2 \leq s_1$: 1 wins, gives the buyer utility $u_1 = s_2$
   - transaction price is determined by $f_1 - p = f_2 - r_2$ so $p = f_1 - f_2 + r_2$
   - $IR_1^s : v_1 = p - r_1 \geq 0$ is satisfied because $s_1 = f_1 - r_1 \geq s_2 = f_2 - r_2$
2. $s_2 \leq 0 \leq s_1$: 1 wins, gives the buyer utility $u_1 = 0$
   - transaction price is determined by $f_1 - p = 0$ so $p = f_1$
   - $IR_1^s : v_1 = p - r_1 \geq 0$ is satisfied because $s_1 = f_1 - r_1 \geq 0$
3. $s_2 \leq s_1 < 0$: no transaction

The transaction takes place only in the first two cases, and in only the first case the buyer gets positive utility.[18]

3.2 First stage

In the *First stage* of the game, the buyer must choose among N differentiated stochastic options, which is a combinatorial optimization problem, a set-up similar to the simultaneous search of stochastically dominated prizes of Chade and Smith (2006). I demonstrate that the buyer contacts sellers in order of decreasing match output f_j, and that the marginal benefit of each additional contacted seller decreases. Since the marginal cost c stays the same, the buyer will keep adding sellers to the search set up to the point where the marginal benefit is less than c.[19]

In the *First stage* of the game, the buyers observe a seller-specific match output f_j but not the seller reservation price r_j. The reservation price is a private value, in the language of auction models, because it has private

[18] Search without hiring is very common in the data, which in the model is represented by the third case. At the same time, one wonders why sellers with $s \leq 0$ would indicate that they are available. I resolve this issue by conjecturing that the sellers do not observe a common component of f that is then revealed in the search stage by the buyer. Let us call this common component $\tilde{r}$, and hence $s = f - r = \tilde{f} - \tilde{r} - r$. For example, $\tilde{r}$ could be the buyer's reservation price or some other common cost of the project. As long as this common component is not correlated with r, this does not cause selection on the part of the sellers who indicate availability in the first stage. The augmented set-up does not alter any of my derivations. A common component of f that is initially unobserved to the sellers would also explain why they are adamant about not committing to a price before discussing with the buyer or visiting the property, i.e. before learning $\tilde{r}$.

[19] The first search that the buyer performs is normalized to zero, similarly to Hortacsu and Syverson (2004) and Dubois and Perrone (2015). In my model, when the buyer contacts only one seller, that seller is a monopolist with respect to the buyer's demand and the expected utility of the buyer is zero. Ex-ante no buyer would be willing to contact only a single seller at a positive search cost, but I do see many such instances in the data. The assumption of zero search cost for the first search makes buyers indifferent between contacting 1 and 0 sellers, and in the data I see that the same number of buyers pick these options. One could think of the decision to search at least one professional as incorporated into the decision to post the job on the marketplace.

relevance: discovering your competitor's reservation price does not make you re-evaluate your own reservation price. From the buyer's perspective, the surplus of the match is a random variable $S_j = f_j - R$, where R is the ex-ante random reservation price with continuous CDF $G_R(r)$. I assume that the distribution of the reservation price is independent from the seller-specific match surplus: $G_R(r|f_j) = G_R(r)$.

Let us consider what the last assumption implies, in particular the relation between observed seller quality and the private reservation price. An available seller of higher quality generates higher surplus on the marketplace as well as out of the marketplace, and this is common information. A seller of higher quality will be compensated more compared to a seller of a lower quality, which in the model is represented by a higher f and in the estimation by a seller fixed effect. I assume that the random reservation price is not associated with the observed seller quality. Rather, it represents the irregular demand - the incidence, the size and the profitability of projects from outside of the platform - after controlling for the observed higher surplus generated by higher quality sellers.[20] In other words, I assume that seller types are perfectly observed and accounted for by the parties, and the asymmetric information arises from random demand conditions that are not withing the control of the seller and not visible to the buyer. Mean independence is necessary for deriving optimal search in the *First stage* because it establishes stochastic dominance among the distributions of the match surplus S of sellers with different match output f. The distribution of the match surplus S_j, which I denote for simplicity $G_j(s)$, is derived from from the distribution of the reservation price R:

$$G_j(s) = Pr(S_j \leq s|f_j) = Pr(f_j - R \leq s) = Pr(f_j - s \leq R) = 1 - Pr(R \leq f_j - s) = 1 - G_R(f_j - s)$$

Let the random variables S^1 and S^2 be the highest and second-highest expected realizations of match surplus. The buyer anticipates the three potential outcomes of the second stage. Only in the first case he receives positive expected utility, which is equal to $E[U] = E[S^2|S^2 > 0]Pr(S^1 \geq S^2 > 0)$. To decide which sellers to contact in the *First stage* (his *search set*), the buyer maximizes his expected utility net of search costs.[21]

Let the buyer search a random set of L sellers. The second highest draw from this set S^2 has cumulative distribution $G^{S^2:L}(s)$. Because S^2 is an order statistic, I know this distribution is:

$$G^{S^2:L}(s) = Pr(S^2 \leq s|L) = \sum_{j=1}^{L}[1 - G_j(s)] \prod_{k \neq j} G_k(s) + \prod_{j=1}^{L} G_j(s)$$

[20] Full independence between r and f is strong as it excludes that the distribution of r depend on the seller quality in other ways, for example through the second moment, the variance. As I do not consider buyers who are risk averse, mean independence of r and f suffices. However, r being IID across sellers and jobs is essential for the identification and estimation if the model.

[21] Note, the search costs are sunk in the *First stage* hence they do not affect the outcome of the *Second stage* in any other way except for determining the search set. The sellers do not commit to compensating the buyer for this cost, especially in the case that no seller is hired.

The expected utility of the buyer given that he contacts the sellers in the set L is the following:

$$E[U|L] = E[S^2|S^2 \geq 0]Pr(S^2 \geq 0) = \frac{\int_0^\infty s\frac{d}{ds}G^{S^2:L}(s)ds}{1 - G^{S^2:L}(0)}.(1 - G^{S^2:L}(0)) = \int_0^\infty s\frac{d}{ds}G^{S^2:L}(s)ds$$

I want to demonstrate that the buyer adds sellers to his search set in the order of decreasing match output f_j and that the marginal benefit of each additional seller decreases. First, I show that if $f_l > f_{l'}$, the buyer prefers the set $L + \{l\}$ to the set $L + \{l'\}$. By induction, this holds for sets of any size and composition. To compare the expected utilities from different searched sets, I will compare the CDF of the respective S^2 by stochastic dominance. The distribution of S^2 for the set $L + \{l\}$ is:

$$G^{S^2:L+\{l\}}(s) = G_l(s)\Big(\sum_{j=1}^{L}[1 - G_j(s)]\prod_{k\neq j}G_k(s)\Big) + \prod_{j=1}^{L}G_j(s)$$

The distribution $G^{S^2:L+\{l'\}}(s)$ is analogous.

Whenever the difference between $G^{S^2:L+\{l\}}(s)$ and $G^{S^2:L+\{l'\}}(s)$ is negative, by the property of first order stochastic dominance the random variable distributed by $G^{S^2:L+\{l\}}(s)$ has a higher expected value.

$$G^{S^2:L+\{l\}}(s) - G^{S^2:L+\{l'\}}(s) = [G_R(f_{l'} - s) - G_R(f_l - s)]\Big(\sum_{j=1}^{L}[1 - G_j(s)]\prod_{k\neq j}G_k(s)\Big)$$

Since $f_l > f_{l'}$, I know $G_R(f_l - s) > G_R(f_{l'} - s)$ because $G_R(r)$ is an increasing function. This makes the first part of the expression negative. Thus, adding a seller with higher f_j to any set L is optimal as it leads to a higher expected value of S^2.

To show that the buyer will eventually stop adding sellers to his search set, I show that the marginal benefit of doing so decreases (while the marginal cost c stays the same). Let D_l be the difference in the distribution of S^2 from an additional seller l:

$$D_l = G^{S^2:L+\{l\}}(s) - G^{S^2:L}(s) = (G_l(s) - 1)\Big(\sum_{j=1}^{L}[1 - G_j(s)]\prod_{k\neq j}G_k(s)\Big)$$

Let $D_{l'}$ be the difference in the distribution of S^2 from further adding l' such that $f_l \geq f_{l'}$:

$$D_{l'} = G^{S^2:L+\{l\}+\{l'\}}(s) - G^{S^2:L+\{l\}}(s) = (G_{l'}(s) - 1)\Big(\sum_{j=1}^{L+\{l\}}[1 - G_j(s)]\prod_{k\neq j}G_k(s)\Big)$$

To prove that the marginal benefit of additional sellers decreases, I show that $D_{l'} \leq D_l$. This can be expressed as:

$$D_{l'}-D_l = [G_{l'}(s)-1][1-G_l(s)]\prod_{j=1}^{L}G_j(s)+\Big(G_R(f_{l'}-s)[1-G_R(f_l-s)]-G_R(f_l-s)\Big)\sum_{j=1}^{L}[1-G_j(s)]\prod_{k\neq j}G_k(s)$$

I have two terms on the right had side of the expression, and I can show that both are negative. The first term is negative because $[G_{l'}(s) - 1] \leq 0$. Inside the brackets of the second term, I have $G_R(f_{l'} - s) \leq G_R(f_l - s)$. Multiplying $G_R(f_{l'} - s)$ by $[1 - G_R(f_l - s)] \leq 1$ makes this difference even larger, which guarantees that the second term is also negative. Thus, the marginal benefit of an additional seller decreases, which satisfies the second order condition of the problem.

4 Identification

Identification of the *Second stage* matching game is achieved using the *special regressor* method developed by Lewbel (2000, 2012).[22] A special regressor is an observed covariate with properties that facilitate identification and estimation. This method is used whenever the researcher's main object of interest is the distribution of the error term, which in my model is the seller reservation price. It is applied in a variety of settings, and more recently in matching games by Fox, Yang and Hsu (2017). Following Lewbel (2000, 2012), I demonstrate the semi-parametric identification of the *Second stage*: non-parametric identification of the reservation price distribution and parametric identification of the coefficients of the match output.

Using the objects identified from the *Second stage*, I can construct the bounds on the individual buyer search cost. These bounds come from the equilibrium inequalities that the search cost must satisfy in the *First stage*. I turn to the literature on *partial identification* to claim set identification of the search cost parameters. I apply Bontemps, Magnac and Maurin (BMM) (2012)'s results on set identification due to incomplete linear moment restrictions.

[22] See Lewbel (2012) for the background on this method, as well as interesting examples of its use.

4.1 Observables and primitives

In the data, I observe buyers $i \in \{1, ..., M\}$ posting jobs on the marketplace. I take each buyer (job) to represent a separate market. In each market, I observe the set of differentiated sellers N_i who have indicated their availability by sending the buyer a message, each seller indexed by j where $j \in \{1, ..., N_i\}$. The available sellers are ordered by decreasing match output f_{ij}. I observe the identities of the sellers contacted by the buyer $n_i \subset N_i$. Let $A_{ij} \in \{0, 1\}$ be the assignment, and it equals 1 when buyer i hires seller j.

For each market i I observe the characteristics of the job, the sellers who are available, and the seller message and experience characteristics at the time of the job post. I group these in matrix $X_i = (X_{i1}, ..., X_{iN_i})'$. There is a sub-set of covariates that affect the buyer's search cost c_i, X_i, and they are not dependent on the seller's identity. Lastly, there exists one regressor $z_i = (z_{i1}, ..., z_{iN_i})'$ that varies across seller-job pairs and satisfies the special regressor conditions.

When I discuss the identification and estimation of the model primitives, I single out the special regressor z_{ij} by separating it from the remaining match surplus. I reconcile this with my previous notation in the following way:

$$s_{ij} = f_{ij} - r_{ij} = z_{ij} + X'_{ij}\beta - r_{ij}$$

For each buyer i, the true search cost c_i cannot be identified but I can identify its bounds: $c_i \in [\underline{c}_i, \overline{c}_i]$. I also postulate a linear specification for the seller search cost c_i:

$$c_i = X_i^{'}\gamma + \epsilon_i$$

The main primitive of interest is the distribution of the reservation price $G_R(r)$. The parameters of interest are the coefficient vectors β of X_{ij} for the match surplus, the individual search cost bounds$[\underline{c}_i, \overline{c}_i]$ and coefficient vectors γ coefficients of X_i for the search cost.

4.2 Second stage

The arguments here follow the binary choice identification strategy of Lewbel (2000, 2012). Consider the binary variable $y \in \{0, 1\}$ which indicates whether the match surplus is greater or equal to zero:

$$y = I[s \geq 0] = I[z + X'\beta - r \geq 0]$$

The data does not allow me to construct y_{ij} for all potential matches between buyer i and sellers j. For example, if $A_{i1} = 1$, I know that seller 1's match surplus must be greater or equal to zero, therefore $y_{i1} = 1$. However, the outcomes $y_{ij'}$ for sellers $2, ..., J_i$ could be anything. It could be that they are all $y_{ij'} = 1$ in the case all $2, ..., J_i$ match surpluses are above zero but seller 1 has the greatest match surplus among the J_i sellers. Or, it could be that all are $y_{ij'} = 0$ when all $2, ..., J_i$ match values are below zero. Because of this, I base the identification on the following two observable cases in my data. Firstly, I use all observations when no seller is hired, where I know for sure that $y_{ij} = 0$ for all j. Secondly, when there is a seller hired and $A_{ij} = 1$, I use the observation for that seller j because $y_{ij} = 1$.

I start by identifying the distribution of the variable w defined as the rest of the surplus: $s = z + X'\beta - r = z + w$. I observe z but not w because I do not observe r. I have that:

$$y = I[z + w \geq 0]$$

I assume that the special regressor z has the following properties:

Assumption 1.

1. $z \perp w|X$
2. *z is additive in the match surplus with coefficient* 1
3. *$-z$ varies continuously over the support of w*

These assumptions allow us to recover the distribution of w at the values $-z$ in the following way. In the data, I observe the expected value of y given z, X. I can re-write this in the following way, using the assumed properties:

$$1 - E[y|z, X] = 1 - Pr(y = 1|z, X) = Pr(y = 0|z, X) = Pr(z + W \leq 0|X) = Pr(W \leq -z|X) = G_{W|X}(-z)$$

Assumption $A1.1$ allows us to use the mean of outcomes y conditional on z, X to construct the marginal distribution of w conditional on X. Assumption $A1.2$ allows us to express this conditional mean as the CDF of the random variable w at the value $-z$. Scaling the coefficient of z to 1 is a scale normalization.[23] Assumption $A1.3$ allow us to trace out the CDF of w over its full support: the variation in $1 - E[y|z, X]$ over all possible values of z allows us to trace out the distribution of w over all its possible values.

Now I let the additional covariates X determine w: $w = X'\beta - r$. The linear structure allows to identify and estimate the coefficients β in a manner similar to OLS. This step also allows us to eventually recover the distribution of the reservation price, G_R, which is the main primitive of interest to my work. I make the following assumptions:

Assumption 2.

1. $z \perp r|X$
2. $E(X'r) = 0$
3. $E(X'X)$ *non-singular*

The first assumption follows from $A1.1$. The last two are standard OLS assumptions and imply that $\beta = E(XX)^{-1}E(X'w)$. I apply the Law of Total Expectations and plug in $G_{W|X}$ into the expression of β:

$$\beta = E(X'X)^{-1}E(X'w) = E(X'X)^{-1}E(X'E[w|X])$$

As the right hand of the expression is identified, so is β. Knowing β, I also know $f = z + X'\beta$. The variation in $E[y|z, X]$ over all possible values of f allows us to trace out the distribution of r over its complete domain. More specifically, I recover CDF of r by the following expression:

$$E[y|z, X] = Pr(y = 1|z, X) = Pr(f - R \geq 0) = Pr(-R \geq -f) = Pr(R \leq f) = G_R(f)$$

4.3 First stage

Observing f and knowing $G_R(r)$, I (and the buyer) know the distribution of match surplus that any partifular f implies $G(s) = 1 - G_R(f - s)$. As I showed in the **Model** section, the buyer contacts sellers in order of decreasing f. The expected utility when the n-highest f sellers are contacted $E[U|n]$ can be constructed when

[23] Models like probit normalize the error term's variance to be 1, but this is observationally equivalent to normalizing the positive coefficient of a regressor, here the special regressor, to one (Dong and Lewbel (2012)).

I know the $G(s)$. The optimal search set n of the buyer implies the following equilibrium inequalities on the search cost c:

$$\underline{c} = E[U|n+1] - E[U|n] \leq c \leq E[U|n] - E[U|n-1] = \overline{c}$$

For each buyer these bounds will be different because the sellers are differentiated and different sellers are in N. Hence, this allows us to identify the search cost bounds $[\underline{c}, \overline{c}]$ for each buyer. Additionally, the characteristics X affect the buyer search cost through parameters γ:

$$c = X\gamma + \epsilon$$

Since I do not observe c but only $\underline{c}$ and $\overline{c}$, I turn to the literature on partial identification, which demonstrates identification of the parameter set Γ that contains all possible values of γ satisfying the constructed inequalities. BMM show that the set Γ is non-empty, bounded and convex, which allows them to identify the set Γ though its support function and to derive an estimation procedure.

5 Estimation

In this section, I discuss in more detail how I use the data and identification results to estimate the fundamentals of the structural model.

5.1 Second stage

I use the seller *Percent positive reviews* variable as the special regressor z. It is a continuous variable that measures the number of positive reviews that the seller has received relative to the total times he was hired up to the time the job was posted. This variable is a proxy for seller revealed quality and commitment to the marketplace.

To satisfy the special regressor properties, I must have that $z \perp r|X$. This assumption would be violated if the random reservation price is correlated with *Percent positive reviews*. It is likely that z is correlated with the average quality of the seller, assuming the quality of his work on and off the online marketplace is similar. However, the set of controls X does include seller reputation measures and seller fixed effects. The reservation price r measures the seller outside demand at the particular moment that the job was posted and in comparison to that specific job, which is random and not observable to the buyer. The buyer's decision to post a job on the marketplace, and when to do that, may be guided by the general quality and availability of sellers but not by an individual seller's outside demand relative to the buyer's job at a particular point in time.

Another of the special regressor properties, that it varies continuously over the full support of r thus allowing us to trace out its CDF, can be verified when I recover the distribution of r from the nonparametric regression of y on the fitted match surplus f. Figure 3 in the **Results** section demonstrates that the probability of hiring is always 1 for very high values of f and 0 for very low values of f.[24]

To estimate β following the OLS equation from the **Identification** section, I would have to start by a non-parametric estimation of $G_{W|X}$. This can be especially challenging because of the large dimension of X and the relatively small sample size. Instead, I prefer a computationally simpler method proposed by Lewbel (2000). He proves that $E[w|X] = E[w^*|X]$ where

$$w^* = \frac{y - I[z \geq 0]}{g_{Z|X}}$$

Constructing w^* is a two-step procedure, where the first step requires the estimation of $g_{Z|X}$. To avoid the curse of dimensionality due to the large dimension of X, I employ a semi-parametric procedure that follows Dong and Lewbel (2012). Let $z = X'\alpha + u$. If $u \perp X$, then $g_{Z|X} = g_U$. Define w^{**} by

$$w^{**} = \frac{y - I[z \geq 0]}{g_u}$$

and correspondingly construct in the data

$$\hat{w}_{ij}^{**} = \frac{y_{ij} - I[z_{ij} \geq 0]}{\hat{g}_U(u_{ij})}.$$

The special regressor conditional independence assumption will be satisfied if $u \perp w|X$, and therefore if $u \perp r|X$. The advantages of this construction is that each u will be estimated as the residuals from an OLS regression of z on X, and g_U can be estimated by a kernel density estimator applied on the set of residuals.

On a final note regarding the construction of w^{**}, g_U may have a large support and so it may be very close to zero for very high and very low values of u. As I am dividing by the probability density, the corresponding values of $\hat{w}^{**}$ then may be extreme in magnitude. I therefore trim 5 percent of the data where $\hat{w}^{**}$ is most extreme.

[24]This property is called *Complete variation* in Magnac and Maurin (2008).

Convergence of the estimator of $\hat{\beta}$ depends on the properties of the density g_U in the denominator. Parametric convergence rate can be obtained in the case that r and z have finite support, or the density of z (therefore of u) has very thick tails, or when r satisfies a tail symmetry condition as defined by Magnac and Maurin (2007). See Lewbel (2012) for references on more detailed discussions on the general limiting distribution theory regarding estimators with an estimated density in the denominator.

Lastly, the estimation of the distribution of r, $G_R(r)$, is performed in the following way. The variable $\hat{f} = z + X'\hat{\beta}$ can be constructed using my estimated $\hat{\beta}$. I perform a non-parametric regression (Nadaraya-Watson local constant) of the sample equivalent of $\hat{E}[y|z, X]$ on $\hat{f}$ to estimate the function $G_R(r)$:

$$\hat{E}[y|z_{ij}, X_{ij}] = \hat{G}_R(\hat{f}_{ij})$$

The limiting distribution of this function will be the same as if $\hat{\beta}$ were replaced by the true β whenever the parameter vector converges to its limit at the parametric rate of convergence (Lewbel (2014)).

5.2 First stage

Once I have $\hat{G}_R(r)$ and $\hat{f_{ij}}$, I can construct the CDF of the match surplus S for any job-seller pair ij:

$$\hat{G}_{ij}(s) = 1 - \hat{G}_R(\hat{f}_{ij} - s)$$

This allows us to construct the difference in distribution of S^2 from an additional seller in the search set, where the sellers are added in order of decreasing $\hat{f}_{ij}$. I go back to my notation from the **Identification** section, where I expressed the differences in distributions of S^2 in the following way. Let $\hat{D}_{n_i}$ denote the difference between the distributions of the second highest surplus S^2 from the sets of n_i and $n_i - 1$ sellers:

$$\hat{D}_{n_i} = \hat{G}^{S^2:n_i}(s) - \hat{G}^{S^2:n_i-1}(s)$$

$\hat{D}_{n_i+1}$ is defined similarly:

$$\hat{D}_{n_i+1} = \hat{G}^{S^2:n_i+1}(s) - \hat{G}^{S^2:n_i}(s)$$

Hence, I can construct the upper and lower bound on the search cost as follows:

$$\hat{\bar{c}}_i = \hat{E}[U|n_i] - \hat{E}[U|n_i - 1] = \sum_0^{\bar{s}} s\left(\frac{\Delta\hat{D}_{n_i}}{\Delta s}\right)$$

$$\hat{\underline{c}}_i = \hat{E}[U|n_i + 1] - \hat{E}[U|n_i] = \sum_0^{\bar{s}} s\left(\frac{\Delta\hat{D}_{n_i+1}}{\Delta s}\right)$$

Finally, I have the bounds on the individual seller search costs c_i:

$$\hat{\underline{c}}_i \leq c_i = X_i\gamma + \epsilon_i \leq \hat{\bar{c}}_i$$

Because the variables X_i are discrete, I use a simplified version of the estimation procedure developed by BMM. I apply their result on the variables one by one, focusing only on the dimension of that variable. The estimation of $\Gamma_k = [\underline{\gamma}_k, \overline{\gamma}_k]$, the identified parameter set of the kth variable, is achieved by the following four steps. To estimate $\overline{\gamma}_k$, the upper bound on the k'th coefficients, I:

1. Construct the vector $q_k = (0, 0, ..., 0, 1, 0, .., 0)$ where the k'th component of the vector is 1
2. Construct the variable q_k^* defined as: $q_k^* = X(X'X)^{-1}q_k$
3. Constructing a modified cost: $\hat{c}_i^* = \hat{\underline{c}}_i + I[q_{ik}^* \geq 0](\hat{\bar{c}}_i - \hat{\underline{c}}_i)$
4. Perform a linear regression of the modified cost c_i^* on X_i

The kth coefficient of the last regression is the estimate of $\overline{\gamma}_k$ To estimate $\underline{\gamma}_k$, I perform the same steps but replace 1 with -1 when constructing q_k in the first step. The kth coefficient is the estimate of $\underline{\gamma}_k$.

BMM show that their estimates converge at a parametric rate to a sum of a Gaussian process and a process they characterize and whose support comprises the points of non-differentiability of the support function of the identified parameter set. However, the bounds on the search cost in my application are estimates rather than the true bounds and this result may not hold.

5.3 Data and selection

Traditionally, buyers and sellers in the home service sector meet though recommendations and adds in newspaper or trade magazine publications. I consider briefly the possibility of selection of the MaistorPlus marketplace users here. For clients who do not use the platform, it is likely that they prefer to hire someone they know from before or is recommended by a friend, or that they are simply unaware of the existence of the website. Such projects would not be any different from the ones I see on the marketplace. Selection issues might arise if the projects posted on the marketplace have been rejected by outside sellers, who were contacted initially. One reason would be that the project is of low quality, but it is also likely that the outside seller was simply unavailable. Regarding the first possibility, in the man regression I am able to control for unobserved project/client quality features so this should not affect the estimation of the coefficients and the reservation price distribution. Regarding the second possibility, I see no reason for selection. The random availability of sellers (both on and off the platform) is difficult to predict but also what makes the platform especially useful because it can bring together parties which are willing to transact but do not necessarily know each other.

Turning my attention to the subscribing sellers, there is a high variability in activity: after controlling for seller, job and date fixed effects, I am only able to explain only 52 percent of seller availability. This suggests that sellers are using the platform to supplement demand coming from outside sources, as job arrivals are random while sellers may need a more constant stream of revenue to pay expenses and staff. It is possible that some inexperienced or lower quality sellers use the platform more intensely because they have fewer referrals from past clients, which would suggest a negative selection on seller characteristics. At the same time, the reputation system may induce a positive selection due to either adverse selection or moral hazard: sellers subscribing to the platform agree to a public reputation so they may be of higher integrity, but also public reviews incentivize sellers to make more effort. My main theoretical and identifying assumption is that the private seller reservation price is a function of the randomness of demand and independent from observed seller quality and characteristics, thus selection should not affect its estimation.

At the level of the individual job i, I have 4,192 jobs overall, 2,744 with at least one contacted seller and 1,417 with at least two contacted sellers. At the level of the job-seller interaction ij, I have two types of samples. The available seller-job data consists of 22,547 observations of sellers sending messages of availability, 14,900 of those are to jobs where the buyer eventually contacts at least one seller and 9,745 are to jobs where the buyer contacts at least two sellers. The contacted seller-job data consists of 5,911 contacted sellers overall, 5,911 contacted for jobs where at least one seller was contacted, and 4,584 contacted where at least two sellers were contacted.

There are two important points that require us to work with the sample of jobs i and buyer-seller interactions (available or contacted) ij where the buyer has contacted at least two sellers. These are the following: buyer reservation prices and other characteristics at the level of the job are not observed; the search cost bound is not identified for jobs where less than two sellers are contacted or for jobs where the number of contacted sellers equals the number of available sellers; I discuss them in more detail below.

The very high number of buyers who decide not to hire anyone indicates that it is also very likely that buyers have reservation prices. Including this in the theoretical model is straightforward: I express the generated surplus net of the buyer reservation price. Although I am not able to identify the buyer reservation price, I can control for the potential bias that its omission may cause including a fixed effect at the level of the individual job in the estimations. The job fixed effects also control for other job-level unobserved variation coming from common costs or unobserved (to the econometrician) job quality. I am not able to separately identify these sources of unobserved variation, but controlling for them is enough to make sure my estimated primitives, the coefficients $\hat{\beta}$ and $\hat{G}_R(r)$, are unbiased.

In the estimation of the search cost bounds, currently I exclude jobs where the sellers contact 0 or 1 buyer and where the sellers contact all available buyers. Given the theoretical model and assumptions, the buyer is indifferent between contacting 0 or 1 sellers as in either case he gets zero expected utility. Such job observations give us a lower bound on the search cost: the cost to contact 2 sellers is higher than the expected benefit of doing so. There is a total of 2,774 such jobs where 0 or 1 sellers are contacted. Similarly, when the buyer contacts all available sellers, I am only able to estimate an upper bound on his search cost. This is the case with 381 jobs. Currently, jobs with a single inequality on the search cost are not included in the estimation. The resulting number of jobs that I use to estimate the search cost bounds is 1,036.

There is a reason why the selection of jobs described above would be less problematic in my setting in comparison to other search models. In models where the sellers are identical ex-ante, as in Hong and Shum (2006), the marginal benefit of search does not differ between jobs or by searched seller, which creates a one-to-one correspondence between search costs and the number of searched sellers. However, this is not the case in my setting. As I have mentioned previously in the **Data** section, the sellers are heterogeneous and different sellers are available for each job. This creates a variation in the marginal benefit of search. It could be the case that no sellers is contacted because the only available sellers are of low quality, even when the buyer has a low search cost. Thus, when I work with jobs where at least two sellers are contacted and less sellers are contacted than available, the selection is not as problematic as in a game with ex-ante identical sellers (and marginal benefit of search).

Lastly, I perform the *Second stage* estimation using the sample of 1,417 jobs where at least two sellers were contacted, where the clients have sent 9,745 messages of availability and the buyers have made 4,584 contacts. To apply the Lewbel (2010) method that lets us estimate β and $G_R(r)$, I restrict this sample in two further ways. Firstly, some of the job-seller contact observations cannot be used because I cannot say if $y_{ij} = Pr(s_{ij} \geq 0)$ or not. For more information on this, see the **Identification** of the *Second stage*. I go down to 3,394 observations, or 74 percent of the sample. Secondly, I trim 5 percent of observations with extreme values of $\hat{g}_u$, leaving us with a sample of 3,230 contacts. I do not anticipate any selection issues arising from this.

6 Results

In this section, I present my estimations of the model primitives: the coefficients β on the match output, the distribution $G_R(r)$ of the reservation price, the individual search cost bounds $[\underline{c}_i, \overline{c}_i]$, and the coefficient sets Γ on the search cost.

6.1 Second stage

I start by demonstrating the monotonic relationship between the outcome variable y and the candidate for a special regressor z, the *Percent positive reviews*. Because both are correlated with covariates X, excluding X from the analysis could lead to a biased relationship so I take a partial regression approach. I perform two separate regressions of y on X and z on X and then take the residuals. Let's call these r_y and r_z respectively. Then, I non-parametrically regress r_y on r_z using an local linear constant estimator. As you can see from Figure 1, there is a positive and monotonic relationship between the residuals r_y and r_z, which indicates that the *Percent positive reviews* is suitable for the special regressor method.

The next step is the semi-parametric estimation of the equation $z = X'\gamma + u$. I perform an OLS regression of z on X, and take the residuals $\hat{u}$. Then, I estimate $\hat{g}_U$ non-parametrically with an Epanechnikov kernel and an optimal bandwidth. Figure 2 displays the resulting density, which is fairly symmetric.

I perform a simple linear regression of $\hat{w}^{**}$ on X to get the coefficients $\hat{\beta}$. The results can be found in Table 8. The β coefficients represent the effects of the covariates on the match output f. Due to the scaling assumption which states that z should have coefficient 1, all other coefficients are scaled by its marginal effect on the match output. The interpretation of the coefficients is therefore more focused on their relative magnitude and sing. The results indicate that both the marketplace tenure and the total times a professional is hired have a significant and negative effect on the match output. This is surprising because in the reduced form regressions in Table 6, where I study the probability that a particular seller is contacted, these variables have positive coefficients (even if not always significant). These negative coefficients are likely the result of scaling all estimated coefficients by the coefficient of the special regressor. The coefficient on the message length is positive and significant, indicating that sellers willing to write longer messages are more suitable for the job. Lastly, the time of the message does not have a significant effect on the match output. All other variables are absorbed by the date, seller and job fixed effects.

The last step in the estimation of the *Second stage* is deriving $\hat{G}_R(r)$. I start by constructing the fitted match output $\hat{f} = z + X'\hat{\beta}$. I get $\hat{E}[y|\hat{f}]$ as the fitted values of a non-parametric local constant regression of y on $\hat{f}$.

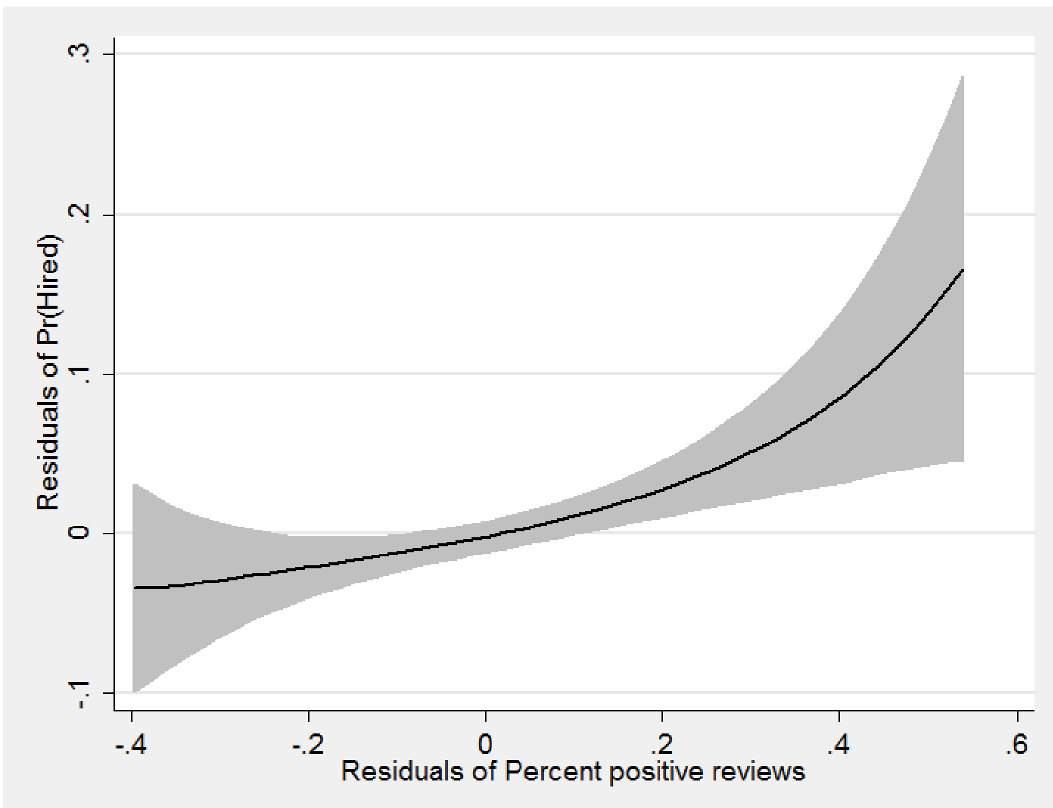

Fig. 1: Non-parametric regression of r_y on r_z.

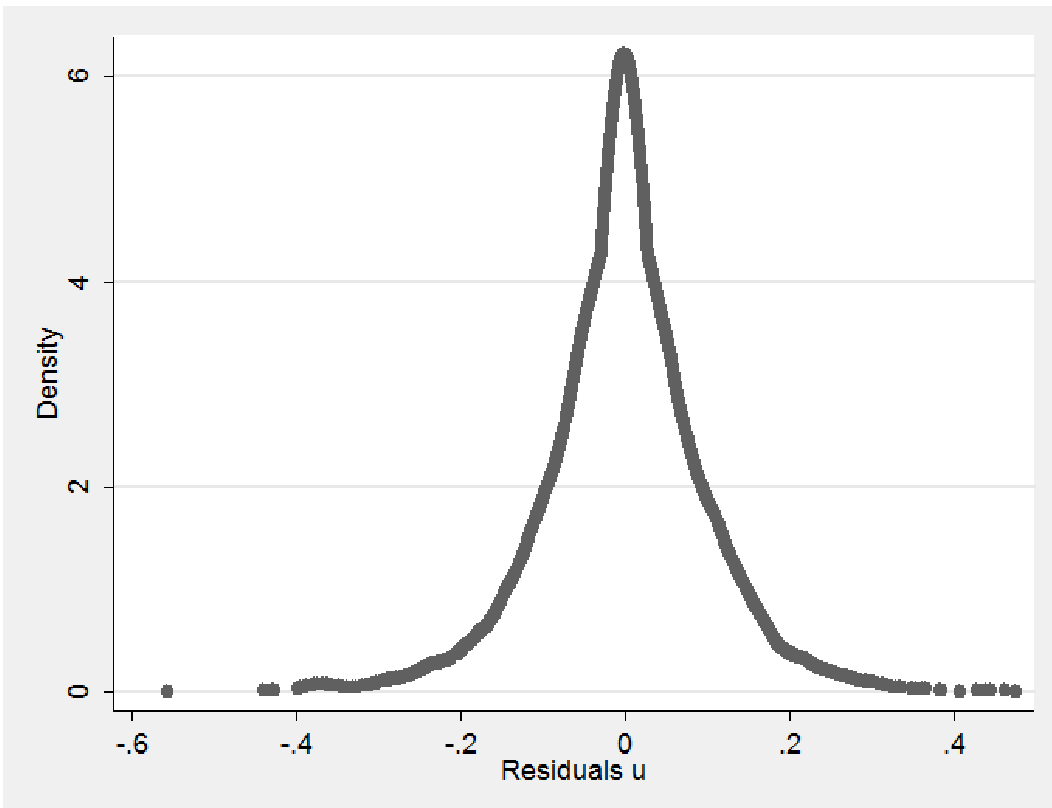

Fig. 2: Non-parametric estimation of $\hat{f}_U$.

Tab. 8: OLS regression of $\hat{w}^{**}$ on the covariates X.

Dependent variable	$\hat{w}^{**}$
Regressors	Coefficient (St. error)
Marketplace tenure	-0.096***
	(0.020)
Total times hired	-0.148***
	(0.014)
Message length	0.023***
	(0.007)
Time of message	0.006
	(0.008)
Date fixed effects	Yes
Seller fixed effects	Yes
Job fixed effects	Yes
R^2	0.80
N	3,236

Significant at: $p < 0.1$: *; $p < 0.05$: **; $p < 0.01$: ***. Robust standard errors in parenthesis. All continuous variables are transformed by taking their natural logarithm and their coefficients are interpreted as elasticities.

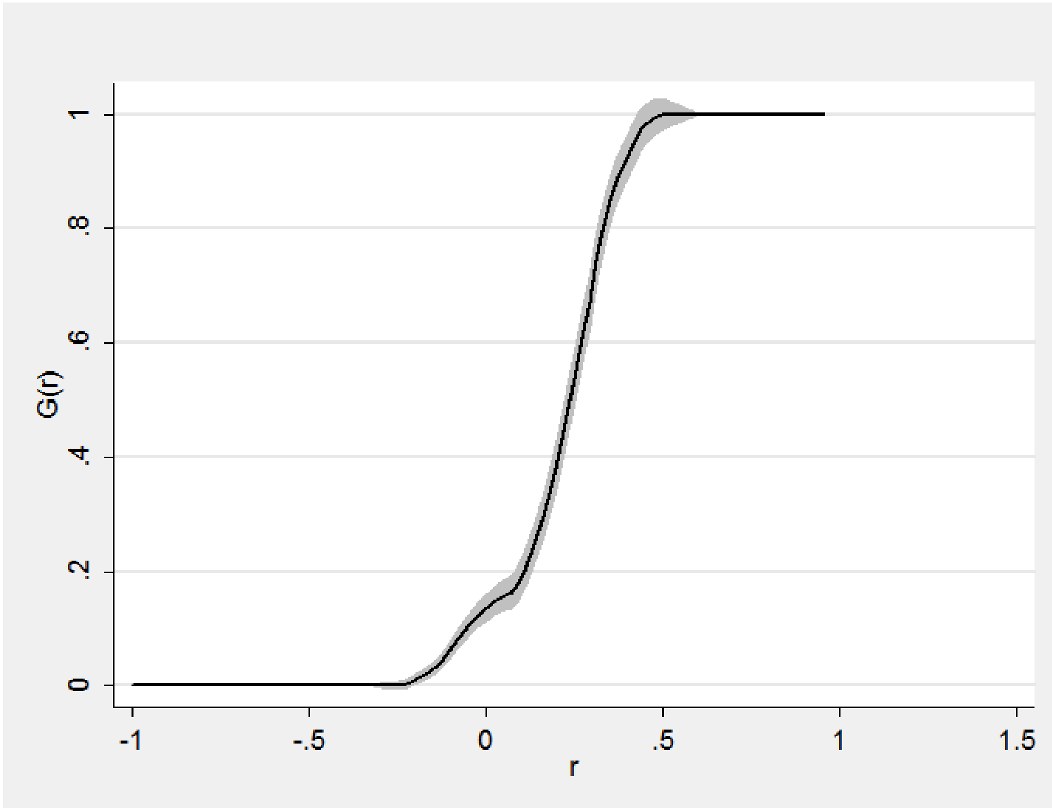

Fig. 3: Non-parametric estimation of the cumulative distribution of the reservation price $G_R(r)$.

By the identification argument, $\hat{E}[y|\hat{f}] = \hat{G}_R(\hat{f})$. The resulting cumulative distribution can be found in Figure 3.

I see from this graph that the reservation price may take negative values, which corresponds to cases where the seller is urgently in need of working on a project. However, the cumulative distribution has value 0.1 at zero, and the majority of mass in the distribution of the reservation price is positive as I would expect. The range of $\hat{f}$, which corresponds to the values on the x-axis, appears to contain the range of r and is sufficient to identify the distribution $G_R(r)$ over its full support. The estimates suggest that the reservation price has finite support, therefore the estimated parameters and distribution converge to the true primitives at parametric rates.

6.2 First stage

The fundamentals estimated for the *First stage* of the model are the individual search cost bounds $[\underline{c}_i, \overline{c}_i]$. I also set identify the coefficients on X_i, the 77 dummy variables which affect the search cost. These are the job's category, expected start, proposed budget, and the date when the job was posted.

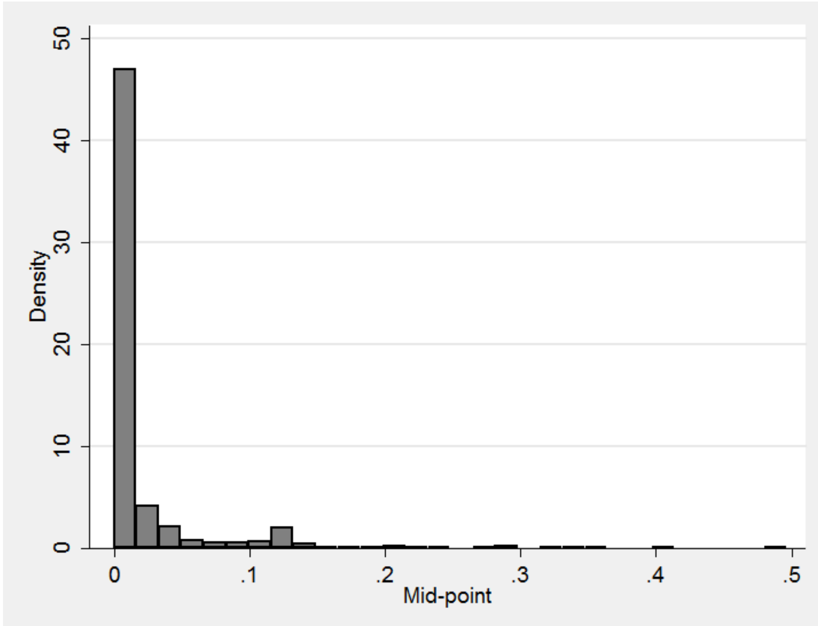

Fig. 4: Distribution of the mid-point of the estimated search cost set.

To construct the job-specific search cost bounds $\underline{\hat{c}}_i$ and $\hat{\bar{c}}_i$ for each observation i. Figure 4 presents a plot of the mid-point of these intervals, $\tilde{c}_i$. As you can see, the graph is highly skewed and the majority of observations are close to zero.

To estimate the identified coefficients set Γ_k of the kth variable, I follow the method proposed by BMM that I describe in the **Estimation** section. This method allows is to recover the identified interval for each coefficient: $\hat{\Gamma}_k = [\underline{\hat{\gamma}}_k, \hat{\bar{\gamma}}_k]$. For brevity, I only present the coefficient sets for the indicators for job category, job start and job budget. These results can be found in Table 9, Table 10 and Table 11 respectively.[25] The reference category - the omitted category against which these effects are measured - is the first variable in each table.[26]

7 Counterfactuals

In this section, I use the estimated model primitives to simulate counterfactual scenarios of the buyer-seller interaction. I start by exploring how the matching outcome would change if the client had less information about the sellers, their availability and their characteristics. These scenarios allow us to quantify how the existence of MaistorPlus improves efficiency by bringing buyers and sellers together, eliciting seller availability and by maintaining verified seller profiles. Secondly, I simulate a complete information environment and quantify the

[25] The estimated parameters for the date indicator variables are available upon request.

[26] **Hopefully I can discuss these results in January.**

Tab. 9: Estimated boundaries for the job category coefficient sets Γ_k.

Job category	$\underline{\hat{\gamma}}_k$	$\hat{\bar{\gamma}}_k$
Architecture and design	-	-
Bathroom repair	-0.099	0.152
Building restoration and insulation	-0.074	0.070
Carpentry	-0.078	0.043
Chimney and fireplace repairs	-0.081	0.041
Cleaning services	-0.097	0.060
Construction	-0.097	0.083
Demolish, clean and transport	-0.068	0.036
Doors and barriers	-0.076	0.064
Dry construction	-0.077	0.059
Electrical repairs	-0.076	0.077
Energy efficiency	-0.070	0.126
Equipment repairs	-0.086	0.059
Floors: parquet and tiles	-0.101	0.079
Furniture	-0.069	0.078
Heating and air conditioning	-0.074	0.065
Kitchen repairs	-0.090	0.062
Landscaping	-0.073	0.068
Masonry	-0.086	0.055
Painting and decoration	-0.182	0.096
Railings	-0.068	0.068
Road construction	-0.135	0.062
Roof repairs	-0.073	0.042
Sewage and sanitation	-0.082	0.044
Smithery services	-0.082	0.086
Textile and upholstery	-0.088	0.043
Transport services	-0.112	0.076
Welding	-0.101	0.079
Window pane and glass repairs	-0.117	0.057

Tab. 10: Estimated boundaries for the job start coefficient sets Γ_k.

Job start	$\underline{\hat{\gamma}}_k$	$\hat{\bar{\gamma}}_k$
After reviewing offers	-	-
Immediately	-0.054	0.060
In 2 days	-0.042	0.053
In 2 months	-0.046	0.045
In 2 weeks	-0.052	0.058
Just checking offers	-0.053	0.044
More than 2 months	-0.038	0.058
Unknown	-0.068	0.065

Tab. 11: Estimated boundaries for the job budget coefficient sets Γ_k.

Job budget	$\underline{\hat{\gamma}}_k$	$\hat{\bar{\gamma}}_k$
Not indicated	-	-
25	-0.103	0.085
50	-0.081	0.097
100	-0.074	0.076
250	-0.067	0.050
1,000	-0.067	0.060
2,000	-0.068	0.045
2,500	-0.072	0.058
4,000	-0.073	0.036
8,000	-0.073	0.028
15,000	-0.063	0.049
30,000	-0.077	0.032
Above 30,000	-0.080	0.042

extent to which the remaining information friction, generated by a positive search cost in combination with an unobserved seller reservation price, affects the outcomes on the platform. I look at two measures of efficiency: the probability of a match and the generated surplus.

I believe that the buyers come to the MaistorPlus marketplace for the following reasons: they do not know a trusted seller, their trusted seller is not available, or has a very high reservation price. MaistorPlus, on the other hand, provides the buyers with the following benefits. Firstly, it has the contact information of the multiple sellers who subscribe to the platform. Secondly, the platform notifies all these sellers simultaneously about the client's project and those available send a message, thus saving the client from searching unavalable sellers. Lastly, the website contains verified information on the experience and work of each seller. I evaluate the contributions of the platform to the market efficiency by simulating the following three counterfactual scenarios. In the *Random single contact* scenario, I have the client randomly contacting one seller, who may or may not be available. There is a transaction only when the seller is available and the surplus generated by the potential match is greater than zero. In the *Random search* scenario, I now assume that the buyer can contact multiple sellers but does not know if they are available or how they are differentiated. The difference in match probability and generated surplus between *Random single contact* and *Random search* is due to simply informing the buyer of the existence of multiple potentially available sellers. Lastly, in the *Random available search* scenario, I allow the buyer to randomly search among the sellers who have indicated to be available. Compared to the *Random search*, I measure the effect of the marketplace taking the first step in the search process: informing the sellers of the job, and allowing the available sellers to message the client.

Denote as the *Directed available search* scenario what I have in reality. There is a crucial difference between *Directed available search* and *Random available search*: in *Random available search*, the sellers are ex-ante identical, which means that the buyer does differentiate between sellers with different experience or reputation (which enter the match output f). In addition to generating information on seller availability, the MaistorPlus marketplace verifies the information that sellers put on their profiles and maintains a public reputation system. For simplicity and to make comparison easier between the counterfactual scenarios, I assume that the same number of sellers that are contacted in *Directed available search*, are also contacted in *Random available search* and *Random search*.

Despite MaistorPlus being able to bring interested parties together, to facilitate their interaction and to provide credible differentiating information, there is still incomplete information because the buyers cannot observe costlessly the private reservation prices of the sellers. Let's consider the game once again, and what would change if I remove the information friction. I call this the *Frictionless* scenario. If search were costless, the client can contact all the sellers ($n_i = N_i$) and invite them to compete. The client will match with the highest surplus seller among N_i in the case that $s_1 \geq 0$, and he would get s_2 in the case that $s_2 \geq 0$. Alternatively, if search were costly but the reservation prices of the sellers were visible, the client would contact the sellers with surpluses s_1 (for free) and s_2 (at a cost c) whenever $s_2 \geq 0$. In either case, the client matches with the highest surplus seller among all available sellers N_i whenever $s_1 \geq c$. Ex-ante, the probability of a match is maximized and the maximum surplus possible is created. In the *Directed available search* scenario, the buyer the client matches with the best seller among the searched set n_i, which may not be the best seller among the full set of available sellers N_i. The probability of a match is also lower because it is more likely that the highest surplus seller in a smaller set (n_i versus N_i) has a non-positive surplus.

7.1 Method

To simulate the five scenarios of *Random single contact*, *Random search*, *Random available search*, *Directed available search* and *Frictionless*, I use the following primitives that are identified and estimated by my model: the distribution of the seller random reservation price r, the distribution of the seller-buyer match output f, and the distribution of the mid-point of the search cost $\tilde{c}$. Using the data directly, I can approximate the distributions of the number of notified sellers $\tilde{N}$ and the distribution of the number of available sellers N, that is a function of $\tilde{N}$.

I simulate the outcome of each scenario for 4,000 jobs. I start by jointly simulating by bootstrap the number of notified sellers $\tilde{N}_i$ and the number of interested sellers N_i for each of the 4,000 jobs. Now, only N_i of the $\tilde{N}_i$ sellers are actually available. For the available sellers for each job j, I draw a match output f_{ij} by bootstrapping from the empirical distribution of $\hat{f}$. To avoid selection, this distribution is constructed using the fitted match output for all sellers who have ever sent a message for any job, rather than just contacted sellers: the full sample of 22,547 messages sent by any seller to any job. Lastly, for each available seller I draw the reservation price r_{ij} from $\hat{G}_R(r)$ and I construct $s_{ij} = f_{ij} - r_{ij}$.

I do not identify and estimate the distribution of the search cost c, but rather the bounds on the individual search cost for each buyer as a function of the marginal utility generated by the differentiated sellers. Thus, to simulate search costs I choose to work with the distribution of the mid-point of the search cost interval $\tilde{c}$. I construct a sample of estimated mid-points of the search cost $\tilde{c}_i$, and again use bootstrap to draw a search cost for each job i.

Scenarios *Random single contact* and *Random search* involve the buyer not observing whether the contacted sellers belong to $\tilde{N}_i$ or N_i. Whenever a randomly contacted seller is not available, meaning he is in $\tilde{N}_i$ but not N_i, the buyer cannot transact with this seller and this seller thus does not participate in the English auction of the *Second stage*. To simulate the buyer contacting 1 or n_i sellers randomly, in *Random single contact* and *Random search* respectively, I draw from a uniform distribution on $U[0, 1]$ for each seller $\tilde{N}_i$ - job i pair. In the *Random single contact* scenario, the buyer contacts the seller with the highest realization of this random variable. A match is formed when the contacted seller j is available (belongs to N_i) and has a surplus $s_{ij} \geq 0$. In the *Random search* scenario, he contacts the sellers with the n_i highest realizations of this random variable. A match is formed with the highest surplus seller of the contacted set whenever he is available and $s_{i1} \geq 0$.

In the *Random available search* and *Directed available search* scenarios, the buyer contacts sellers who are available: they belong to the set N_i. In the first case, the buyer does not observe the characteristics of the sellers, f_{ij}, while in the second case these characteristics direct his search. For the *Random available search* scenario, I derive the search set by taking the sellers with the n_i highest realizations of a random variable $U[0, 1]$. For the *Directed available search* scenario, I derive the search set n_i using the equilibrium conditions of the *First stage* described in the **Model** section. I rank the sellers in N_i by f_{ij}, then calculate the expected benefit to the buyer of an additional seller added to his search set n_i when sellers are added in order of decreasing f_{ij}. I add sellers to the search set n_i up to the point where the expected benefit is lower than the expected cost $\tilde{c}_i$. To make comparison easier between the counterfactual scenarios, the buyer contacts the same *number* of sellers in the *Random available search* and *Random search* as the optimal number of sellers in *Directed available search* (but not the same *set* of sellers). For the *Directed available search* scenario, I calculate the amount of the total search cost $(n_i - 1) * c_i$ relative to the surplus that is generated by the match.

Lastly, in scenario *Frictionless*, the buyer observes the match surplus s_{ij} of all available sellers and his search set is *de facto* the set N_i. He is matched with the best seller in N_i whenever this seller generates a positive surplus: $s_{i1} \geq 0$.

7.2 Results

The results of the counterfactual analysis can be found in Table 10: the probability of a match and the average surplus generated in each of the five alternative scenarios: *Random single contact*, *Random search*, *Random available search*, *Directed available search*, and *Frictionless*. The baseline scenario, *Random single contact*, achieves a match in only about 2 percent of the jobs, which is extremely low. Allowing the buyers to contact multiple sellers but with no information about availability brings that up to 6 percent. The biggest improvement in the probability of the match is achieved when the seller availability is visible to the buyer: in the *Random available search* scenario, the probability of a match is 41 percent. About 12 percentage points are added whenever the buyers can also observe seller differentiation, bringing the match probability up to 53 percent. Lastly, the full information outcome, *Frictionless*, achieves a 59 percent of match probability. I keep in mind that a reduction in the information friction could also reduce the buyer search costs, which are estimated to be on average 30 percent of the generated surplus in *Directed available search*.

In terms of the generated surplus, the average increases as I go from one scenario to the next: 0.18 for *Random single contact*, 0.21 for *Random contact*, 0.25 for *Random available contact*, and 0.30 for *Directed available contact*. The *Frictionless* scenario average surplus is 0.29 because compared to the *Directed random search*, about 244 more matches take place, but the generated surplus is on average lower for these new matches. For matches that happen under both scenarios, the *Frictionless* counterfactual generates 1 percent higher surplus. This is due to 3 percent of the matches where an assignment of higher surplus is achieved when there are no information frictions. The counterfactuals show that the information friction affects match formation to a greater extent than match assignment.

Tab. 12: Match probability and surplus under the different counterfactual scenarios

	Observations	Mean	St. Dev.	Min.	Max.
Match probability					
Random one contact	4,000	0.02	0.14	0	1
Random search	4,000	0.06	0.23	0	1
Random available search	4,000	0.41	0.49	0	1
Directed available search	4,000	0.53	0.50	0	1
Frictionless	4,000	0.59	0.40	0	1
Match surplus					
Random one contact	81	0.18	0.16	0.00	0.93
Random search	224	0.21	0.19	0.00	0.99
Random available search	1,652	0.25	0.20	0.00	1.64
Directed available search	2,123	0.30	0.22	0.00	1.96
Frictionless	2,367	0.29	0.22	0.00	1.96

8 Conclusion

I see more and more service markets, where the agents are differentiated and capacity constrained, moving their activity to online platforms. The availability of data and the scope for improving market design in such settings opens up many interesting research questions. In this paper, my goal is to model how a service seller-buyer match is formed when the marketplace is not centralized, the sellers have unobserved availability and willingness to transact and the buyers must search for this information to be revealed. I am motivated to understand this interaction because it brings us closer to how agents operate in many other service markets, where imperfect and costly information is the norm.

I construct, identify and estimate a structural model of search and matching using data from the online home services marketplace MaistorPlus. The estimated primitives of interest - the distribution of the unobserved seller reservation price, the bounds on the buyer search cost, and parameters of the match output - allow us to perform counterfactual analysis of this market. I estimate the substantial efficiency contribution of MaistorPlus to the market: the platform facilitates the meeting of buyers with multiple sellers, lets buyers know if the sellers are available, and verifies seller profile information regarding their experience and reputation. I also estimate the potential gain in efficiency from removing the remaining information friction, the private seller reservation prince in combination with a positive seller search cost. I show that if this information friction can be removed, the probability of match formation will increase by 6 percentage points but match assignment (a potentially higher surplus in already existing matches) would not improve substantially. Additionally, this could save the buyer the effort of search, which is estimated at 30 percent of the generated surplus on average.

9 References

Arnosti, Nick, Ramesh Johari, and Yash Kanoria. ”Managing congestion in decentralized matching markets.” Proceedings of the fifteenth ACM conference on Economics and computation. ACM, 2014.

Bontemps, Christian, Thierry Magnac, and Eric Maurin. ”Set identified linear models.” Econometrica 80.3 (2012): 1129-1155.

Chade, Hector, and Lones Smith. ”Simultaneous search.” Econometrica 74.5 (2006): 1293-1307.

Chade, Hector, Jan Eeckhout, and Lones Smith. ”Sorting through search and matching models in economics.” Journal of Economic Literature 55.2 (2017): 493-544.

Cheron, Arnaud, and Bruno Decreuse. ”Matching with phantoms.” The Review of Economic Studies (2016): rdw032.

Chiappori, Pierre-Andre, and Bernard Salanie. ”The econometrics of matching models.” Journal of Economic Literature 54.3 (2016): 832-861.

Choo, Eugene, and Aloysius Siow. ”Who marries whom and why.” Journal of political Economy 114.1 (2006): 175-201.

Dong, Yingying, and Arthur Lewbel. ”A simple estimator for binary choice models with endogenous regressors.” Econometric Reviews 34.1-2 (2015): 82-105.

Dubois, Pierre, and Helena Perrone. ”Price Dispersion and informational frictions: Evidence from supermarket purchases.” (2015).

Eeckhout, Jan, and Philipp Kircher. ”Sorting and decentralized price competition.” Econometrica 78.2 (2010): 539-574.

Cullen, Zo, and Chiara Farronato. ”Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms.” Job Market Paper (2014).

Fox, Jeremy T., David H. Hsu, and Chenyu Yang. ”Unobserved Heterogeneity in Matching Games with an Application to Venture Capital.” No. w18168. National Bureau of Economic Research, 2017.

Fradkin, Andrey. ”Search, matching, and the role of digital marketplace design in enabling trade: Evidence from Airbnb.” (2017).

Galenianos, Manolis, and Philipp Kircher. ”Directed search with multiple job applications.” Journal of Economic Theory 144.2 (2009): 445-471.

Hortasu, Ali, and Chad Syverson. ”Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds.” The Quarterly Journal of Economics 119.2 (2004): 403-456.

Horton, John. Buyer Uncertainty about Seller Capacity: Causes, Consequences, and Partial Solution. Working Paper, 2016.

Kircher, Philipp. ”Efficiency of simultaneous search.” Journal of Political Economy 117.5 (2009): 861-913.

Laffont, Jean-Jacques, and Jean Tirole. A theory of incentives in procurement and regulation. MIT press, 1993.

Lewbel, Arthur. ”Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables.” Journal of Econometrics 97.1 (2000): 145-177.

Lewbel, Arthur. ”An overview of the special regressor method.” Manuscript (2012).

Magnac, Thierry, and Eric Maurin. ”Identification and information in monotone binary models.” Journal of Econometrics 139.1 (2007): 76-104.

Magnac, Thierry, and Eric Maurin. ”Partial identification in monotone binary models: discrete regressors and interval data.” The Review of Economic Studies 75.3 (2008): 835-864.

McCall, John Joseph. ”Economics of information and job search.” The Quarterly Journal of Economics (1970): 113-126.

PostelVinay, Fabien, and JeanMarc Robin. ”The distribution of earnings in an equilibrium search model with statedependent offers and counteroffers.” International Economic Review 43.4 (2002): 989-1016.

Rochet, JeanCharles, and Jean Tirole. ”Twosided markets: a progress report.” The RAND journal of economics 37.3 (2006): 645-667.

Roth, Alvin E., and Marilda Sotomayor. ”Two-sided matching.” Handbook of game theory with economic applications 1 (1992): 485-541.

Shapley, L. S., and Martin Shubik. Game Theory in Economics: Chapter 3: the rules of the Games”. Rand, 1972.

Shi, Shouyong. ”Frictional assignment. i. efficiency.” Journal of Economic Theory 98.2 (2001): 232-260.

Shimer, Robert, and Lones Smith. ”Assortative matching and search.” Econometrica 68.2 (2000): 343-369.

Stigler, George J. ”The economics of information.” Journal of political economy 69.3 (1961): 213-225.

Tadelis, Steven. ”Reputation and feedback systems in online platform markets.” Annual Review of Economics 8 (2016): 321-340.

Weitzman, Martin L. ”Optimal search for the best alternative.” Econometrica: Journal of the Econometric Society (1979): 641-654.

Ye, Lixin. ”Indicative bidding and a theory of two-stage auctions.” Games and Economic Behavior 58.1 (2007): 181-207.

A Appendix

A.1 Descriptive statistics

Tab. 13: Job expected start in full jobs sample.

	Freq.	Percent	Cum.
After reviewing offers	184	4.05	4.05
Immediately	638	14.05	18.1
In 2 days	910	20.04	38.13
In 2 months	429	9.45	47.58
In 2 weeks	1,369	30.14	77.72
Depending on the offers	480	10.57	88.29
More than 2 months	97	2.14	90.42
Unknown	205	4.51	94.94
Missing	230	5.06	100
Total	4,542	100	

Tab. 14: Jobs proposed budget in full jobs sample.

Job budget	Freq.	Percent	Cum.
Missing	536	11.8	11.8
25	322	7.09	18.89
50	380	8.37	27.26
100	529	11.65	38.9
250	515	11.34	50.24
500	503	11.07	61.32
1000	423	9.31	70.63
2000	271	5.97	76.6
2500	154	3.39	79.99
4000	111	2.44	82.43
8000	55	1.21	83.64
15000	25	0.55	84.19
25000	1	0.02	84.21
30000	20	0.44	84.65
Above 31000	697	15.35	100
Total	4,542	100	

Tab. 15: Job categories in the full jobs sample.

Category	Freq.	Percent	Cum.
Architecture and design	58	1.28	1.28
Bathroom repair	488	10.74	12.02
Building restoration and insulation	232	5.11	17.13
Car repairs	53	1.17	18.3
Carpentry	212	4.67	22.96
Chimney and fireplace repairs	41	0.9	23.87
Cleaning services	44	0.97	24.83
Constr. equipment for rent	4	0.09	24.92
Construction	158	3.48	28.4
Control and access	9	0.2	28.6
Cutting and engraving	7	0.15	28.75
Demolish, clean and transport	61	1.34	30.1
Doors and barriers	129	2.84	32.94
Dry construction	140	3.08	36.02
Electrical repairs	235	5.17	41.19
Energy efficiency	9	0.2	41.39
Equipment repair	164	3.61	45
Floors: parquet and tiles	318	7	52
Furniture	252	5.55	57.55
Heating and airconditioning	69	1.52	59.07
Kitchen repair	108	2.38	61.45
Landscaping	38	0.84	62.29
Lathe services	3	0.07	62.35
Locksmith	22	0.48	62.84
Masonry	10	0.22	63.06
Metalworking	13	0.29	63.34
Painting and decoration	492	10.83	74.17
Pest control	25	0.55	74.72
Railings	17	0.37	75.1
Road construction	36	0.79	75.89
Roof repairs	284	6.25	82.14
Sewage and sanitation	408	8.98	91.13
Smithery services	101	2.22	93.35
Surveying services	3	0.07	93.42
Textile and upholstery	17	0.37	93.79
Transport services	15	0.33	94.12
Welding	33	0.73	94.85
Window pane and glass repairs	234	5.15	100
Total	4,542	100	