Search and Matching with Outside Options on an Online Marketplace for Services

Milena Petrova[*]

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

JEL classification: L81, D82, D83

Abstract

Buyer-seller interactions in certain service markets can be accurately described by two-sided matching theory, which emphasizes the differentiation and limited capacity. When the marketplace is not centralized, incomplete information about availability plays an important role. To this end, we construct a structural model of search and matching where seller have an unobserved outside option that enters the match surplus as a random reservation price. We estimate the model using proprietary data from an online services marketplace; we recover the distribution of the reservation price, and the effects of match-related characteristics on the match output and search cost. Our paper contributes to the young and growing empirical literature on online service marketplaces, which studies the determinants of match formation and information frictions due to unobserved availability.

1 Introduction

In service markets, limited capacity and differentiation on both sides of the interaction are important factors. For example, a client booking a roof repair worker to clean the gutters on his roof would prevent that worker from taking a better paid roof repair job during that same week. Thus, the service provider must be compensated

[*]I thank Marc Ivaldi, Jacques Cremer, Shruti Sinha, 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.

appropriately because being hired on one project causes him to miss out on other opportunities, whenever these are present. Another important aspect of such interactions is that both clients (and the respective service they request) and sellers are vertically differentiated and ranked relative to others on the same side of the market. Continuing our example from before, the roof repairmen could come with different levels of experience and the properties in question could be in a better or worse state. Thus, we turn to two-sided matching theory to understand how differentiated agents compete with those on the same side of the market to form a stable match with an agent on the other side.

The goal of our work is to model and estimate the role of unobserved outside options in a online matching market for services. We are motivated by the numerous online and offline matching markets where interactions are not centralized, and partners have private information about potential matches on alternative platforms. In the literature of two-sided markets, this practice is known as multi-homing: the agents are not able to observe their potential competitors - agents that their potential partner is considering on another platform - and this information is important for the formation when contacting - or searching - partners is costly. Our work is especially relevant to understanding and potentially alleviating the market failure arising from unobserved availability in online matching markets, as they are more amenable to being designed.

We construct and identify a two-stage search and matching model, which is then estimated using data from a Bulgarian online home services marketplace (*MaistorPlus*).[1] In our empirical application, service providers (sellers) that have subscribed to MaistorPlus also look for clients in a number of ways: advertising on online forums, past client referrals, newspaper adds. The employment opportunities from these alternative marketplaces can be considered as an outside option with respect to jobs posted on MaistorPlus. When a client (buyer) who has posted a job on MaistorPlus considers a potential match with a seller, the merit of that buyer's project relative to the outside option of the seller must be established. Indeed, the MaistorPlus data on the interactions between buyers and available sellers indicates that initially the sellers never commit to working on a job or to a price offer because there is uncertainty about the project's value to them. The uncertainty is resolved by a process of *search*: the buyer exchanges contacts with the seller, arranges a site visit or discusses the project in detail over the phone. This allows the seller to ascertain how this project compares relative to his outside option. When the search process is costly for the buyer, the buyer must optimize his search activity.

Our model describes the interaction between service sellers and service providers in two stages. In the first stage of the game, the buyer observes all available sellers and their characteristics, but not exactly how attractive the project is relative to the sellers' outside option. The uncertainty is modeled as an ex-ante random reservation

[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 there countries the market is still very decentralized, with multiple platforms of different design.

price that enters additively in the match surplus. The buyer decides which sellers to contact in a directed, fixed sample search manner, paying a constant search cost for each contact. In the second stage of the game, the contacted sellers reveal their reservation price and compete to form a match with the buyer. Our methodological contribution is constructing, identifying and estimating a matching model where outside options are important but initially unobserved, and search is performed in a simultaneous manner. The model allows us to recover the distribution of the reservation price, and to estimate the effects match-related characteristics on the match surplus and search costs.

This paper is related to the theoretical and empirical literature on two-sided matching with transfers, which is concerned with studying markets where the goods (or services) to be allocated are heterogeneous and indivisible (see Roth and Sotomayor (1992)). In the beginning days of this literature, information frictions were not typically studied. The agents were assumed to observe all agents on both sides of the market, as well as the potential match surplus and the utility of remaining unmatched. 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), and they do not consider information frictions on the part of the market participants.[2] The framework of Choo and Siow (2006) imposes a structure on the unobserved error term in order to identify coefficients on the match surplus. Fox et al (2015), on the other hand, develop a method which identifies the unobserved error term by assuming we observe a certain observable characteristics enters the match surplus, a method similar to the special regressor of Lewbel (2012). The identification of the second stage of our model is very close to Fox et al (2015), the main difference being that the random component of the match surplus is initially unobserved by some of the agents in the first stage of the game. The two conceptual, but not technical, differences between this literature and our model are motivated by our detailed match-level interaction data and the high decentralization of the market. In the Firstly, traditionally the market is treated as centralized, and agents who do not match on it are assumed to simply remain unmatched rather as saving themselves to match on another market. Secondly, the unobserved error term is interpreted as a taste heterogeneity while in our specific setting we believe this term is more appropriately interpreted as a reservation price.

More recently, the empirical two-sided matching literature has incorporated search in the presence of imperfect and costly information regarding potential match partners.[3] Microeconomic search theory explores how option value governs choices: where to search, how long or how much to search. Simultaneous search was first introduced by Stigler (1961), while Chade and Smith (2006) extend the paradigm to allow for ex-ante heterogeneous options. This framework is very similar to the first stage in our model, except for the fact that in their setting each option generates a stochastic reward while in our setting the searched options compete to determine the resulting reward. Modeling search friction in this simultaneous way appears in models with multiple appli-

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

[3] Chade, Eeckhout and Smith (2017) review search and matching theory, as well as the recent contributions marrying both these theories.

cations and 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) introduces sequential search to economics, where the searcher's optimal strategy is fully summarized by a reservation wage above which the 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. Lise, Meghir and Robin (2016), Lise and Robin (2017), and Jacquemet and Robin (2013) model agents who meet match partners through time intensive random search. This is a continuous-time model, where potential partners of random quality arrive at a certain rate. While our work shares the preoccupation with imperfect and costly information, we are motivated by a setting where some characteristics of the potential partners are observable ex-ante and therefore search is directed by these characteristics, rather than random. We also claim that fixed sample, rather than sequential, search is more suited for our application, resulting in a two-stage rather a continuous time model.

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 typically not explicitly modeled although they may be informational, heterogeneities, congestion, messaging/application, or else. A number of empirical studies in this literature explore the lack of information about agent availability, which is similar to our interest in agents' outside options. 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. Cheron and Decreuse (2016) study information persistence modeled as phantom agents that are already matched but their status has not updated quickly enough in a continuous-time equilibrium search unemployment model.

Lack of information on availability is a common problem in a wide range of online matching markets for services. Fradkin (2017) and Horton (2016) tackle this subject, using proprietary data from the online matching platforms AirBnB and oDesk. While their work is reduced form, we consider these two papers especially close to ours both because of the application and the main issue of interest, unobserved seller availability. Horton (2016) studies oDesk, an IT task platform, where the buyer can invite sellers to submit offers on a job but does not have information on which of these sellers are potentially available at the current moment. The professionals on oDesk have a wide distribution of worked hours per week, which makes it difficult to predict if they would be available or not at any point in time. This activity pattern suggest a random outside occupation, similar to the outside option for the sellers on MaistorPlus. Horton (2016) demonstrates that rejection leads to a decrease in the probability that a job is filled, which suggests that finding ways to reveal more information about seller availability would contribute to the platform's financial success.

The difference between the matching technology of oDesk and MaistorPlus is mainly due to how the marketplace rules of buyer-seller interaction. While on oDesk the buyers look up and contact sellers, in our setting sellers indicate availability by messaging the buyer themselves. However, availability is not necessarily at all costs as the benefit of taking on the job is measured against an unobserved outside option. In fact, what happens on MaistorPlus is in some sense closer to what Fradkin (2017) describes happening on AirBnB: property managers that have indicated availability often reject clients who want to stay in their property. He classifies rejections into stale vacancies (15 percent of the time), congestion (8 percent of the time), and due to characteristics of the seller or the trip (77 percent of the time). The first two are due to de facto unavailability, while the last reason can be attributed to a high reservation price: the property is available but the manager prefers to wait for another offer with better characteristics.

Our paper has the following structure. In Section 2, we solve for the equilibrium of the two-stage search and matching. In the following section, Section 3, we demonstrate that the primitives of interest - the reservation price distribution and parameters of the match output and search cost - are identified. Section 4 details the steps we take to estimate the model. In Section 5, we present the data and reduced form evidence that supports our modeling choices. Section 6 contains the results of our estimation. We conclude our work in Section 7.

2 Search and matching game

We model the interaction between clients (buyers) and providers of services (sellers) on the home services marketplace MaistorPlus. We consider a model of one-to-one matching with transfers (prices). We have one buyer indexed by i and N available, differentiated sellers indexed by j.[4] For the moment, we 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 $s_j = f_j - r_j$.[5,6]

In the first stage of the game, the buyer and sellers know f_j as they observe the job and seller characteristics. The buyer does not fully observe the match surplus s_j, because the seller reservation price r_j is not revealed until after buyer *searches* the seller. The buyer decides which sellers to search in a directed, simultaneous

[4]We see seller availability as a binary state. 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 compares the job to his outside option.

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

[6]When estimating the model, buyer outside options are implicitly incorporated into the additive job-level fixed effect that enters the match surplus.

manner and at a positive and constant marginal cost c.[7,8] Once the seller is contacted, he is able to evaluate the attractiveness of the job relative to his outside options, thus discovering his reservation price r_j. The mechanism to uncover the reservation price is similar to the directed search models of Chade and Smith (2006) for fixed sample and Weitzman (1979) for sequential search. In the first stage, the buyer maximizes his expected net utility of the second stage with respect to the set of sellers he contacts.

In the second stage, contacted sellers know the match output s_j and are available 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 take this into account. We 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.

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, we 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 and match surplus.

2.1 Second stage

In the second stage of the game, the contacted sellers perfectly observe the match surplus s_j. 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

[7] Even if the sellers are the ones to incur the search cost, Ye (2005) demonstrates that it is passed through to the buyer, similarly to how common costs are passed on in the English auction or Bertrand competition.

[8] The assumption of simultaneous search is supported by the following features of the marketplace. The buyers are allowed and advised to contact multiple sellers at the same time as this is in the client's interest for the following two reasons. Firstly, seller availability to visit the site may differ, therefore it is not a good use of the buyer's time to wait for one seller to visit before arranging a visit with another one. Secondly, early seller availability may expire or outside options may change, which would make it more difficult for the buyer to put sellers in direct competition with each other unless he collects their offers at the same time.

to Laffont-Tirole (1993): $u_j = s_j - p$. The individual rationality constraint for the buyer is $IR^b_j : u_j = s_j - p \geq 0$ when transacting with seller j, and for seller j it is $IR^s_j : v_j = p - r_j \geq 0$. 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. In the first case, we have that $0 < s_2 \leq s_1$, seller 1 wins, and gives the buyer utility $u_1 = s_2 \geq 0$. The transaction price is determined by $u_1 = f_1 - p = s_2 = f_2 - r_2$. IR^s_1 is satisfied because $s_1 \geq s_2$. In the second case, we have that $s_2 \leq 0 \leq s_1$, seller 1 wins again, and gives the buyer utility $u_1 = 0$. The transaction price is determined by $u_1 = f_1 - p = 0$. IR^s_1 is satisfied because $s_1 \geq 0$. Lastly, if $s_2 \leq s_1 < 0$ there is no match as neither rationality constraint can be satisfied.

2.2 First stage

In the first stage of the game, the buyers observe a seller-specific match output f_j but neither the buyer nor the sellers observe the reservation price r_j. The reservation price is a private value, in the language of auction models, because it has private relevance: discovering your competitor's outside option does not make you re-evaluate your own outside option. 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 $G_R(r)$. We assume that the distribution of the reservation price is independent from the seller-specific match surplus: $G_R(r|f_j) = G_R(r)$. We derive the distribution of the match surplus S_j, which we denote for simplicity $G_j(s)$, from the distribution of the reservation price R:

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

The buyer must choose among N differentiated stochastic options, which is a combinatorial optimization problem, a set-up is similar to the simultaneous search of stochastically dominated prizes of Chade and Smith (2006). The main difference is that in their model the prize is drawn from the distribution of the winner, while in our model the prize is the equilibrium transaction price of the second stage.

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.

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, we know this distribution is:

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

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

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

We 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, we 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, we 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.

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

Since $f_l > f_{l'}$, we 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, we 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, we show that $D_{l'} \leq D_l$. This can be expressed as:

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

The first term is negative because $[G_{l'}(s) - 1] \leq 0$. We also know that $G_R(f_{l'} - s) \leq G_R(f_l - s)$, and multiplying the left side of the inequality by $[1 - G_R(f_l - s)] \leq 1$ makes this difference even larger. Thus, the marginal benefit of an additional seller decreases, which satisfies the second order condition of our problem.

3 Identification

Identification of the second stage matching game is achieved using the *special regressor* method developed by Lewbel (2000, 2012).[9] 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 our model is the 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), we 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 surplus.

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

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

3.1 Data generation, observables and primitives

In the data, we observe buyers $i \in \{1, ..., M\}$ posting jobs on the marketplace. We take each buyer (job) to represent a separate market. In each market, we 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 $j \in \{0, 1, ..., N_i\}$. The available sellers are ordered by decreasing match output f_{ij}. We 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 we observe the characteristics of the job, the sellers who are available, and the seller-job pair. We 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 variable $z_i = (z_{i1}, ..., z_{iN_i})'$ that varies across seller-job pairs and satisfies the special regressor conditions.

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

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

We 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, and coefficient vectors γ of X_i for the buyer search cost.

3.2 Second stage identification

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:

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

Our data does not allow us to construct y_{ij} for all potential matches between buyer i and sellers j. For example, if $A_{i1} = 1$, we 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, we base the identification on the following two observable cases in our data. Firstly, we use all observations when no seller is hired, where we know for sure that $y_{ij} = 0$ for all j. Secondly, when there is a seller hired and $A_{ij} = 1$, we use the observation for that seller j because $y_{ij} = 1$. This corresponds to keeping 75 percent if the observations of buyer-seller contacts in our sample.

We start by identifying the distribution of the variable w defined as $w = X'\beta - r$. We observe z but not w, which is sometimes referred to as a latent variable in the special regressor literature. We have that:

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

We 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*

The the expected value of y given z, X allows to trace out the distribution of the random variable w in the following way:

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

Assumption $A2.1$ allows us to use the mean of outcomes y conditional on z, X to construct the marginal distribution of w conditional on X. Assumption $A2.2$ 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 makes that easier and it is a scale normalization.[10] Assumption $A2.3$ allow us to trace out the full support of w, otherwise the distribution of w will only be identified at the values z takes.

Now we let the additional covariates X determine w: $w = X'\beta - r$. The linear structure of w allows us to identify and estimate the coefficients β in a manner similar to OLS. We 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 $A2.1$. The last two assumptions imply that $\beta = E(XX)^{-1}E(X'w)$. We 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, we also know $f = z + X'\beta$ and the CDF of r can be recovered easily. :

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

3.3 First stage identification

Observing f and knowing $G_R(r)$, we (and the buyer) know the distribution of match surplus that any partifular f implies $G(s) = 1 - G_R(f - s)$. As we 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

[10] 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)).

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

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

Additionally, the characteristics X_i affect the buyer search cost through parameters γ:

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

We do not observe c but only $\underline{c}$ and $\overline{c}$. We turn to the literature on partial identification to identify the parameter set Γ that contains all possible values of γ satisfying the constructed inequalities. Bontemps, Magnac and Maurin (BMM) (2012) 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.

This identification approach cannot be applied to 27 percent of the jobs in our sample, where the buyer has contacted all available sellers. Thus, for these jobs we are unable to construct $\underline{c}_i$ and we disregard them. .

4 Estimation

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

4.1 Second stage estimation

We 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 at the time the job was posted on the marketplace.

To satisfy the special regressor properties, we 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 outside option 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 fixed effects, which should account for such average effects. The reservation price r measures the seller outside option 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 of sellers but not by an individual seller's outside demand relative to the buyer's job at a particular point in time.

To estimate β following the OLS equation derived in the **Identification** section, we 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, we prefer a computationally simpler method proposed by Lewbel (2000). He proves that 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, we 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 we are dividing by the probability density, the corresponding values of $\hat{w}^{**}$ then may be extreme in magnitude. We therefore trim 5 percent of the data where $\hat{w}^{**}$ is most extreme.

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 $\hat{f}_{ij} = z_{ij} + X'_{ij}\hat{\beta}$ can be constructed using our estimated $\hat{\beta}$. We perform a non-parametric regression (for example, Nadaraya-Watson local constant) of the sample equivalent of $E[y|z_{ij}, X_{ij}]$ on $\hat{f}_{ij}$ to estimate the function G_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.

4.2 First stage estimation

Once we have $\hat{G}_R(r)$ and $\hat{f}_{ij}$, we can construct the CDF of the match surplus 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, where the sellers are added in order of decreasing $\hat{f}_{ij}$:

$$\hat{D}_{n_i} = \hat{G}^{S^2:n_i}(s) - \hat{G}^{S^2:n_i-1}(s) = (\hat{G}_{i,n_i}(s) - 1)\Big(\sum_{j=1}^{n_i-1}[1 - \hat{G}_{ij}(s)]\prod_{k\neq j}\hat{G}_{ik}(s)\Big)$$

$$\hat{D}_{n_i+1} = \hat{G}^{S^2:n_i+1}(s) - \hat{G}^{S^2:n_i}(s) = (\hat{G}_{i,n_i+1}(s) - 1)\Big(\sum_{j=1}^{n_i}[1 - \hat{G}_{ij}(s)]\prod_{k\neq j}\hat{G}_{ik}(s)\Big)$$

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

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

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

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

$$\hat{\underline{c}}_i \leq c_i \leq \hat{\overline{c}}_i$$

Because the variables X_i are discrete, as it is in our application, we use a simplified version of the estimation procedure for Γ developed by BMM. This simplified procedure allows us to apply their result on the variables one by one, focusing only on the dimension of that variable. To estimate the $\Gamma_k = [\underline{\gamma}, \overline{\gamma}_k]$, the set of parameters for the kth variable, we perform the following four steps. Firstly, we construct the vector $q_k = (0, 0, ..., 0, 1, 0, .., 0)$ where the k'th component of the vector is 1. Next, we construct the variable q_k^* defined as:

$$q_k^* = X(X'X)^{-1}q_k$$

The third step is constructing a modified cost:

$$c_i^* = \underline{c}_i + I[q_{ik}^* \geq 0](\overline{c}_i - \underline{c}_i)$$

.

The last step is a linear regression of the modified cost c_i^* on X_i, where the kth coefficient estimates $\overline{\gamma}_k$. To estimate $\underline{\gamma}_k$, we perform the same steps but replacing 1 with -1 in q_k. The negative of the kth coefficient is the estimate of $\underline{\gamma}_k$.

BMM show that the 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 our application are estimates rather than the true bounds and this result may not hold.

5 Data

In this section, we introduce the data set and provide some reduced form supporting evidence for the assumptions of our model.

5.1 Description

We work with company data form the MaistorPlus online services marketplace, who are based in Bulgaria and started operating in 2012.[11] The marketplace connects clients to subscribing home service professionals, and is financed by professionals' 3-month subscription fees and by advertising.

The client signs up freely and posts a job: the project for which they want to hire a professional. At the level of the job, we observe the following characteristics X_i job category (one of 38 categories such as carpentry, roof repairs, construction, etc), expected start date (one of 8 options), and proposed budget (one of 14 options). According to our preliminary robustness checks, the available job-level covariates X_i are not able to control for important, unobserved job-level characteristics in the match output. For example, common costs and buyer outside options are not observable and cannot be identified otherwise.

We therefore limit our sample to jobs where the buyer has contacted at least two sellers because this allows us to include a job fixed effect in the estimation of the match output.[12] We work with a total of 1,417 jobs, the sellers were notified 126,950 times, and they indicated interest via a message 9,746 times. The clients contacted 4,585 professionals and hired someone in 539 cases. The summary statistics for the activity at the level of the individual job can be found in Table 1.

There are a total of 717 active professionals in our sample. At the level of the available professionals j, we observe the following variables X_j that describe the seller's profile: categories of activity, profile description, references from previous clients and pictures from past projects. Summary statistics are provided in Table 1. In the main estimations, these variables are absorbed by the seller fixed effects.

We define the job-seller interactions ij as observations where seller j has indicated availability for the job i . We observe the following two sets of variables at this level: the seller experience and variables related to the seller's message indicating availability. The seller experience variables are the seller tenure on the marketplace (in months) and the total times the seller was hired up to the month the job was posted. The seller's *Percent positive reviews*, the special regressor, is defined as number of positive reviews on all jobs for which the seller was hired. 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

[11] http://maistorplus.com/

[12] We consider potential selection issues stemming from this choice. We do not believe that the buyer's decision to contact more than two sellers is related to the individual sellers' outside options, hence the estimated reservation price should not suffer from selection bias. Selection on the buyer search cost is more likely: buyers with lower search costs contact more sellers. However, because sellers are differentiated in our model, there is no one-to-one correspondence between search costs and the number of contacted sellers as there is in models where the sellers are symmetric (see Hong and Shum (2006)). Thus, any selection effect is mitigated by our results indicating that the identity of the available sellers (the match output they generate) is very important for the buyer's search decision.

these variables are presented in Table 1.

Tab. 1: Summary statistics for the continuous variables

Activity at the job level					
Sellers notified for the job	1, 417	89.59	50.74	2	401
Messages received for the job	1, 417	6.87	5.69	2	61
Contacts made for the job	1, 417	3.24	1.66	2	10
Probability of hiring	1, 417	0.38	0.48	0	1
Seller profile characteristics					
Active categories	717	5.26	4.67	0	28
References	717	0.15	0.49	0	3
Profile description (chars.)	717	558.02	497.63	0	3,645
Profile pictures	717	11.56	26.35	0	490
Variables at the job-seller level					
Seller experience					
Percent positive reviews	9,746	0.32	0.40	0	1
Marketplace tenure (months)	9,746	8.30	7.44	0	36
Total times hired	9,746	3.72	7.21	0	46
Message-related					
Message length (chars.)	9,746	250.85	308.17	0	6,153
Time of message (hours)	9,746	3.77	14.19	0	433

Cullen and Ferronato (2014) report similar results for TaskRabbit: 49 percent of tasks remain unmatched; of all posted tasks, 78 percent receive offers and that is an average of 2.8 offers per job; 63 percent of tasks are completed.

5.2 Reduced form evidence

In this section, we present reduced form evidence which supports our modeling choices. More specifically, we show that sellers do have capacity constraints and they are viewed as differentiated by the buyers, which means that the matching framework is appropriate for modeling interactions on the marketplace; that there is uncertainty about seller willingness to undertake a project; and that buyers experience search costs. We discuss these issues one by one.

5.2.1 Capacity constraints

We believe the seller's availability to provide the service is significantly constrained by the physical time needed to perform the service, which makes the matching framework - where agents are constrained in the number of interactions - suitable. Firstly, the data demonstrates that sellers are not always available and that their activity goes down in high demand periods. The job-level activity presented in Table 1 indicates that even if about 90 sellers receive the notification about any given job, only about 7 of them indicate that they are available by sending the client a message.[13]

At the level of the individual job, higher overall demand leads to fewer available professionals and a lower probability to hire someone. We define two outcome variables: the number of messages indicating interest (*Messages received*) and the hiring outcome (*Pr(Hire)* $\in$ $\{0, 1\}$) of a given job. We regress these on the *Demand activity* on the marketplace for that time period, measured as the number of jobs posted during the time period the 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. We also control for the overall level of activity on the marketplace at that time: total number of messages (*Message activity*), contacts (*Contact activity*) and professionals that were hired(*Hired activity*) on the marketplace during that time. Of the marketplace activity variables, only the *Demand activity* can be considered as truly exogenous. We include the rest to control for omitted variable bias as they are correlated with *Demand activity*. The results are presented in Table 2 below.

High *Demand activity* lowers the *Messages received* and *Pr(Hired)* for the individual job, suggesting that the sellers are not available to accommodate demand as their capacity is fixed. For a 10% increase in *Demand activity* in that period, the number of messages received for any given job decreases by 4% and the probability of hiring decreases by 3%.

[13] 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 that seller messaging costs may be significant. A fixed seller messaging cost does not contradict the set-up of our model, as now the cut-off for availability of the seller must be at least high enough to rationalize the cost of the message as well. Discussion with the marketplace owners suggests that sellers use similar message templates, which reduces this cost.

Tab. 2: Activity at the level of the individual job as affected by overall demand and marketplace activity.

	Messages received	Pr(Hire)
Marketplace activity		
Demand activity	-0.410***	-0.310***
	(0.108)	(0.119)
Message activity	0.439***	0.008
	(0.076)	(0.073)
Contact activity	-0.163***	-0.013
	(0.065)	(0.075)
Hired activity	0.030	0.298***
	(0.077)	(0.098)
R^2	0.69	0.38
N.	1,417	1,417

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

Other controls which are not presented here for brevity are job characteristics, fixed and job-level characteristics of the average available seller, and activity at the level of the job (sellers notified, messages, contacts). More detailed results are available upon request.

5.2.2 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 by examining the likelihood that any available seller is contacted by the buyer. We show that *Pr(Contact)*$\in \{0, 1\}$ depends both on that seller's characteristics and on those of his competitors. Again, we 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. We have the following sets of seller characteristics: profile, experience, and message-related. *Profile* characteristics are fixed at the level of the professional i, while *experience* and *message* characteristics are measured at the job-seller interaction level ij.[14] In Table 3 we present the initial regression $R1$ which includes seller and competitor profile characteristics. $R2$ is the same regression with seller fixed effects, and $R3$ has both seller and competitor fixed effects.

All else equal, $R1$ 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. Comparing $R2$ to $R3$, we see that including competitor fixed effects does improve the explanatory power of the regression.

We are unable to perform a similar analysis to explicitly demonstrate project differentiation because there is considerable activity off the marketplace that we do not observe. Thus, we do not know what projects rival each other at the level of a given available seller, at any point in time. However, Table 2 demonstrates that projects are in competition with each other for the seller's time as higher demand leads to less activity at the individual job level.

5.2.3 Unobserved seller outside options

When sellers indicate availability, this does not imply that they have no outside options. In our set-up, the seller's outside option is modeled as a random reservation price that enters the match surplus. We believe the reservation price is not known ex-ante by the sellers, that it is an important component of the match surplus, and that it is a reservation price rather than any other form of cost. Our assumptions are supported by the following observations.

Firstly, Table 2 shows that the *Pr(Hire)* goes down in high demand periods, even after controlling for all buyer-seller interactions, project and seller characteristics. We believe this is strong evidence for the seller's

[14] We also control for time fixed effects and jobs characteristics, and the more detailed results are available upon request. We were unable to include job fixed effects because of the small sample size.

Tab. 3: Seller probability of being contacted by the client: *Pr(Contact)*.

	R1	R2	R2
Profile			
Seller active categories	-0.008		
	(0.009)		
Competitor active categories	-0.048**	-0.039*	
	(0.020)	(0.022)	
Seller references	0.016		
	(0.013)		
Competitor references	-0.057**	-0.046*	
	(0.025)	(0.026)	
Seller profile descr. length	0.003		
	(0.002)		
Competitor profile descr. length	-0.033***	-0.032***	
	(0.006)	(0.007)	
Seller profile pictures	0.004		
	(0.004)		
Competitor profile pictures	-0.024***	-0.021**	
	(0.008)	(0.008)	
Experience			
Seller percent positive reviews	0.097***	0.124***	0.125***
	(0.023)	(0.035)	(0.036)
Competitor percent positive reviews	0.061	0.073*	0.101*
	(0.042)	(0.044)	(0.060)
Seller marketplace tenure	0.009	-0.006	-0.014
	(0.007)	(0.019)	(0.020)
Competitor marketplace tenure	0.008	0.003	-0.101***
	(0.015)	(0.016)	(0.026)
Seller total times hired	0.031***	0.005	0.010
	(0.007)	(0.014)	(0.015)
Competitor total times hired	-0.030**	-0.039***	0.006
	(0.012)	(0.013)	(0.016)
Message			
Seller message length	0.015***	0.017***	0.018***
	(0.004)	(0.005)	(0.005)
Competitor message length	-0.039***	-0.040***	-0.030***
	(0.006)	(0.006)	(0.007)
Seller time of offer	-0.088***	-0.084***	-0.097***
	(0.005)	(0.006)	(0.008)
Competitor time of offer	0.107***	0.100***	0.038*
	(0.015)	(0.015)	(0.022)
Date FE	No	Yes	Yes
Seller FE	No	Yes	Yes
Competitor FE	No	No	Yes
R^2	0.29	0.36	0.43
N	9,745	9,745	9,745

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

reservation prices being higher in high demand periods due to more outside options. It is unlikely that this effect is due to fluctuating physical time or material costs needed to perform the 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.

Secondly, the sellers themselves are uncertain about the attractiveness of any given job as inspection of the messages they send demonstrates that they insist on visiting or a more detailed discussion before making an offer or committing in any way. If the sellers did observe their outside option, the messaging stage would be *de facto* the matching stage and we would see the buyer contacting only the ”winning” seller. Table 1 demonstrates that this is not the case on average as sellers contact more than one buyer, and we believe this is because there is uncertainty about that buyer’s outside options.

Lastly, the outside option is an important component of the match surplus. Were that not the case, we would see the buyer always hiring the most *ex-ante* attractive seller, where sellers are ranked based on their observable characteristics. We use the regression $R3$ from Table 3 to predict the ex-ante ranking of each seller, with 1 being the highest rank. Using the predicted rankings, Table 4 shows on average the rank of the hired seller, and the highest and lowest ranked sellers that were contacted. While on average the most attractive seller is contacted by the buyer (rank 1.12), usually a lower ranked seller is hired (rank 2.35).

Tab. 4: Seller ranks: hired, maximum and minimum contacted for jobs where at least two sellers are contacted

	Obs.	Mean	St. dev.	Min	Max
Rank of hired seller	539	2.35	2.41	1	23
Highest rank contacted	539	1.12	0.90	1	12
Lowest rank contacted	539	4.63	3.71	1	32

5.2.4 Costly, directed search

We argue that the buyer does experience non-trivial search costs associated with contacting the sellers, because this usually involves lenghty discussions over the phone and finding time to arrange a visit. If search were costless, we would see the buyer contacting all available sellers so that there is tougher competition in the matching stage. Table 1 demonstrates this not the case: on average, the buyers contacts half of the available sellers. Furthermore, the buyer sees the sellers as differentiated, which directs his search. Table 3 shows that seller characteristics are important: they create an order of attractiveness among the sellers and direct the buyer’s search process.

6 Results

In this section, we present our estimations of the model primitives: the coefficients β, the distribution $G_R(r)$ and the coefficient sets Γ.

6.1 Second stage estimations

We observe 4,585 interactions where the buyer has contacted the seller. To apply our estimation technique, we construct a sample of outcomes y as defined in the **Identification** section. This brings us down to 3,395 observations, losing about 25 percent of the data. Our next step is to demonstrate that there exists a 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 we take a partial regression approach. We perform two separate regressions of y and z on X and then take the residuals. Let's call these r_y and r_z respectively. Then, we non-parametrically regress r_y on r_z using an Epanechnikov kernel of second degree and an optimal (rule-of-thumb) bandwidth calculated by Stata. As you can see from Figure 1, there is a positive and relatively 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 $z = X'\gamma + u$. We perform an OLS regression of z on X, and take the residuals $\hat{u}$. Then, we estimate $\hat{f}_U$ non-parametrically with an Epanechnikov kernel and an optimal bandwidth. Figure 2 displays the resulting density, which is fairly symmetric.

After constructing $\hat{w}^{**}$, we trim 5 percent of observations to avoid extreme values due to dividing by densities very close to zero. The resulting sample has a total of 3,230 observations. We perform a simple linear regression of $\hat{w}^{**}$ on X to get the coefficients $\hat{\beta}$. The results can be found in Table 5 below.

These coefficients represent the effects of the covariates on the match output. Due to our scaling assumption which states that z should have coefficient 1, all other coefficients are scaled by its marginal effect on the match output. Our results indicate that both the marketplace tenure and the total times a professional is hired have a negative effect on the output. This is surprising because professionals who these variables measure experience and commitment, but possibly there are reputation building (and using up) effects that we do not model. The coefficient on the message length is positive and significant, indicating that sellers willing to write longer messages are more suitable or more trustworthy. Lastly, the time of the message does not appear to matter for the match output. All other variables are absorbed by the date, seller and job fixed effects.

The last step in the estimation of the first stage is deriving $\hat{G}_R(r)$. We start by constructing $\hat{f} = z + X'\hat{\beta}$.

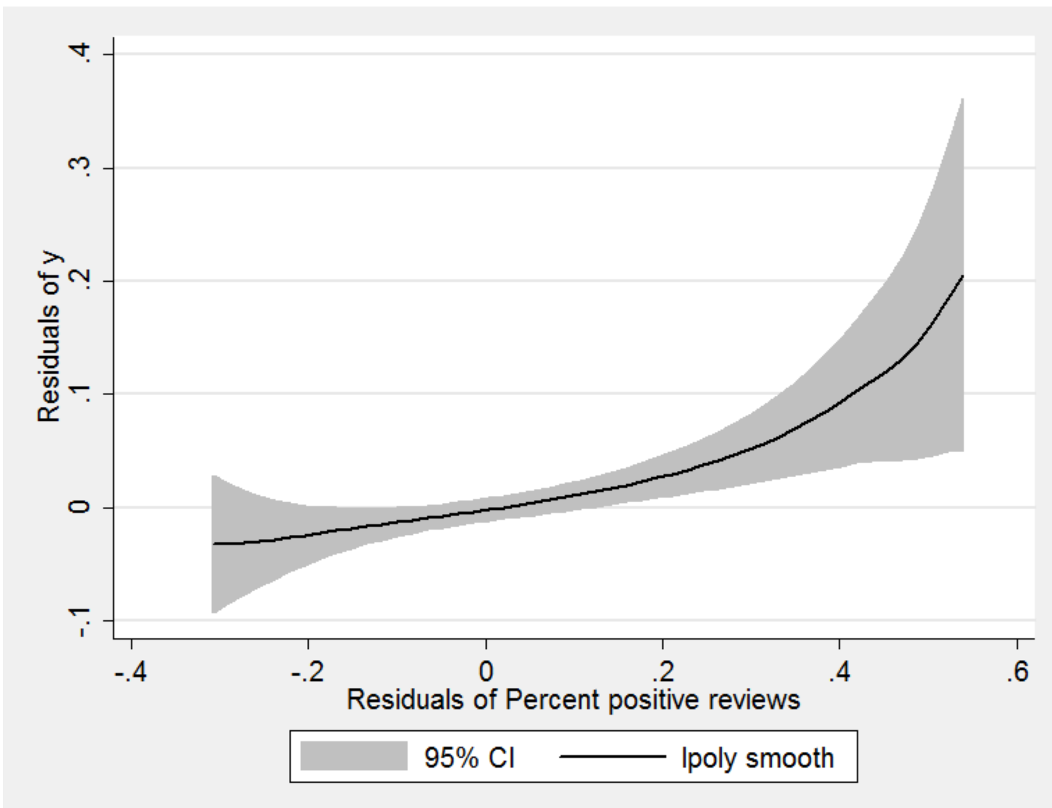

Fig. 1: Non-parametric regression of r_y on r_z.

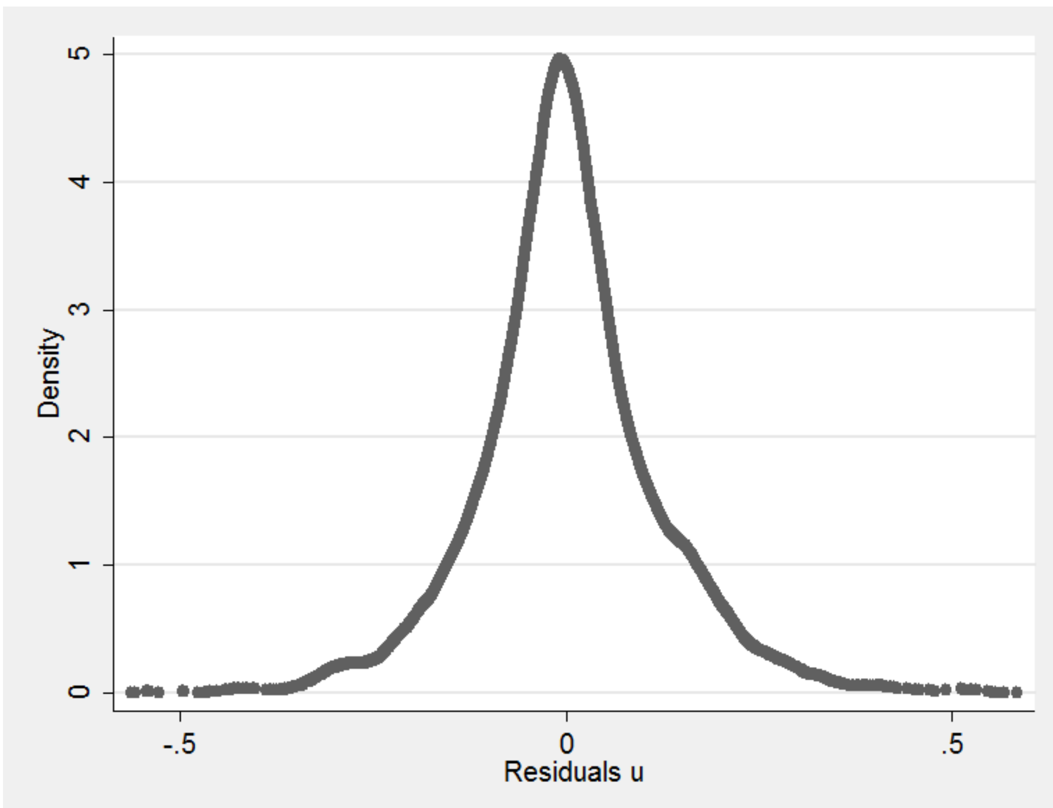

Fig. 2: Non-parametric estimation of $\hat{f}_U$.

Tab. 5: OLS regression of $\hat{w}^{**}$ on the covariates X.

Dependent variable	$\hat{w}^{**}$
Regressors	Coefficient (St. error)
Marketplace tenure	-0.142***
	(0.030)
Total times hired	-0.126***
	(0.018)
Message length	0.038***
	(0.010)
Time of message	-0.0167
	(0.011)
Constant	0.0470
	(0.394)
Date FE	Yes
Seller FE	Yes
Job FE	Yes
R^2	0.83
N	3,230

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

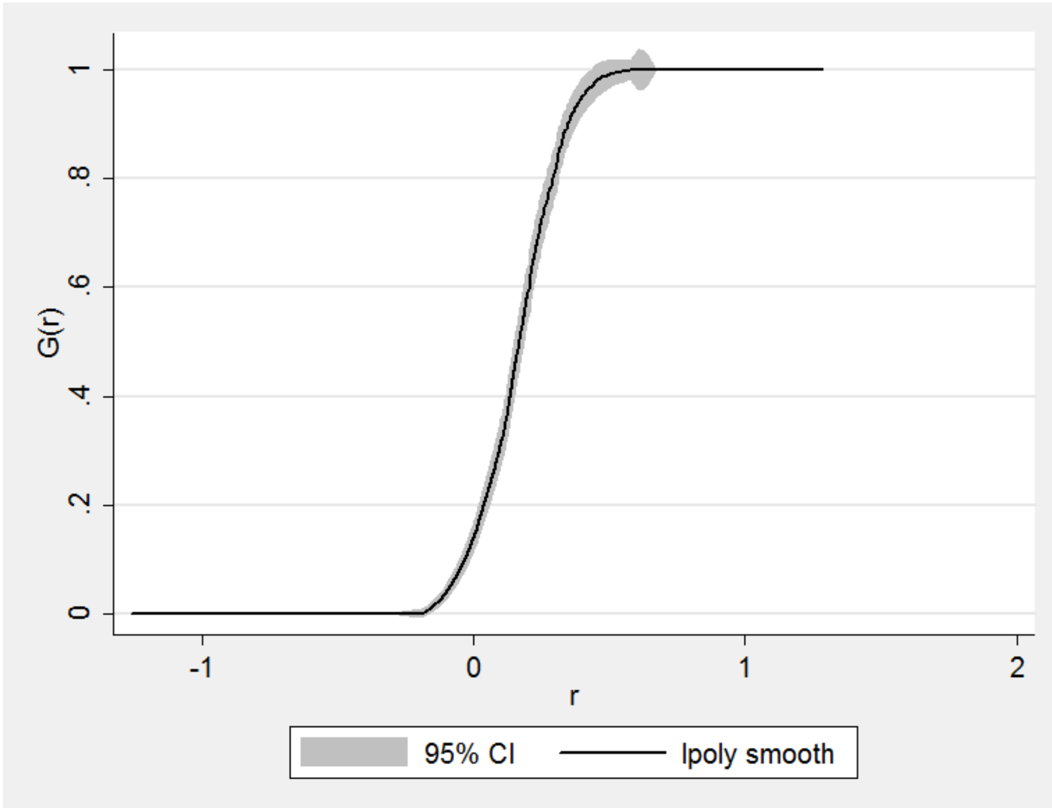

Fig. 3: Non-parametric estimation of the cumulative distribution of the reservation price $G_R(r)$.

We get $\hat{E}[y|\hat{f}]$ as the fitted values of a non-parametric regression of y on $\hat{f}$. By our identification argument, $\hat{E}[y|\hat{f}] = \hat{G}_R(\hat{f})$. The resulting cumulative distribution can be found in Figure 3.

We 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 0.2 at zero, which means that on average the reservation price is positive as we 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 $G_R(r)$. Our estimates suggest that the reservation price has finite support, therefore the estimated parameters and distribution converge to the true values at parametric rates.

6.2 First stage estimation

We have 81 dummy variables which determine the search cost of the seller, X_i, and they indicate the job's category, expected start, proposed budget, and the date when the job was posted. Our first step is to construct the job-specific search cost bounds $\underline{\hat{c}}_i$ and $\hat{\bar{c}}_i$. This identification approach cannot be applied to 382 of the 1,417 jobs in our sample, where the buyer has contacted all available sellers. Thus, for these jobs we are unable to construct $\underline{c}_i$. and we work with the remaining 1,036 jobs. To estimate the identified coefficients set Γ_k of the kth variable, we follow the method proposed by BMM that we describe in the **Estimation** section.

In Table 6, we present the estimated parameter sets $\hat{\Gamma}_k$ for the indicators of the job category.[15] The reference category - the omitted category against which these effects are measured - is *Architecture and design*. Some of the estimated coefficient sets contain zero, for example those of *Control and access*, *Furniture*, and *Landscaping*. For these parameters, we can not reject the hypothesis that the true coefficient is zero. Some of the sets are quite narrow (*Energy efficiency*) compared to others (*Doors and barriers*).

7 Conclusion

Online marketplaces where the agents are differentiated and capacity constrained, be it accommodation, hope repairs, or dating, are more popular than ever, and with that there is more data available which opens up many interesting research questions. In this paper, our goal is to model how a service seller-buyer match is formed when the marketplace is not centralized and the sellers have unobserved outside options. We are motivated to understand this interaction because such a model bring us closer to how agents make decisions in the real world, where imperfect and costly information is the norm. For this, we construct, show identification for, and estimate a structural model of search and matching using data from the online home services marketplace MaistorPlus. The model has the following primitives of interest: the the distribution of the seller reservation price and parameters of the match output and search cost.

[15] The estimated parameters for the other indicator variables are available upon request.

Tab. 6: Estimated boundaries for the job category coefficient sets Γ_k.

Job category	$\hat{\bar{\gamma}}_k$	$\hat{\underline{\gamma}}_k$
Architecture and design	-	-
Bathroom repair	0.031	0.019
Building restoration and insulation	0.022	0.006
Car repairs	0.039	0.023
Carpentry	0.026	0.011
Chimney and fireplace repairs	0.019	-0.009
Cleaning services	0.019	0.010
Construction	0.027	0.018
Control and access	0.074	-0.032
Demolish, clean and transport	0.023	0.016
Doors and barriers	0.042	0.006
Dry construction	0.028	0.013
Electrical repairs	0.030	0.009
Energy efficiency	0.034	0.026
Equipment repair	0.030	-0.001
Floors: parquet and tiles	0.031	0.007
Furniture	0.030	-0.008
Heating and conditioning	0.019	0.018
Kitchen repair	0.029	0.009
Landscaping	0.059	-0.003
Masonry	0.052	-0.051
Metalworking	0.053	-0.046
Painting and decoration	0.023	0.013
Railings	0.009	0.008
Road construction	0.045	0.001
Roof repairs	0.024	0.021
Sewage and sanitation	0.010	-0.002
Smithery services	0.017	-0.008
Surveying services	0.013	0.002
Textile and upholstery	0.052	-0.047
Welding	0.039	-0.030
Window pane and glass repair	0.003	0.002

8 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.

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.

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

Jacquemet, Nicolas, and Jean-Marc Robin. ”Assortative matching and search with labor supply and home production.” Centre for Microdata Methods and Practice, Institute for Fiscal Studies (2012).

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).

Lise, Jeremy, and Jean-Marc Robin. ”The Macrodynamics of Sorting between Workers and Firms.” The American Economic Review 107.4 (2017): 1104-1135.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. ”Matching, sorting and wages.” Review of Economic Dynamics 19 (2016): 63-87.

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

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

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.

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

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