Reputation in an online programmer’s market
Snijders, Chris; Weessie, Jeroen

Published in:
Default journal

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
March 2008

Reputation in an online programmer’s market.
A peach for the price of a lemon

Chris Snijders
Dept of Technology Management
Eindhoven University of Technology
PO Box 513
5600 MB Eindhoven
The Netherlands
Tel 040 247 5596
Fax 084 830 6374
c.c.p.snijders@tue.nl

Jeroen Weesie
Dept of Sociology
Utrecht University
Heidelberglaan 2
3584 CS Utrecht
The Netherlands
Tel 030 253 1922
Fax 030 2534405
j.weesie@fss.uu.nl

Abstract

The dominance of eBay as an auction site and the availability of eBay data for research purposes have created the situation in which much of the empirical work about reputation mechanisms takes eBay as its focus. This has disadvantages because the available data are not as close to the actual decision making of the buyer as one might think. In particular, characteristics of all the goods a person could have but did not buy do not play a role in typical eBay data. We show that this may very well be the reason why researchers find general inconsistency in the empirical results, in particular, a lack of consistent support for a positive effect of reputation on both probability of sale and end price. One of the issues that has consistently been overlooked is that the theoretical arguments about the effects of reputation in most eBay-papers concern how much buyers would be willing to pay extra for a product from a seller with a good reputation. However, eBay data typically allow the researcher to measure how much they have to pay extra, which is not the same. Using discrete choice models and data that include foregone options, we can adequately estimate the willingness to pay for reputation from choices that buyers make among alternatives. Our analyses use data from the online programmer’s site RentACoder.com (about n=22,000 auctions with more than 250,000 bids). When analyzed using such data, we do find strong support for a positive effect of a positive rating on the willingness to pay. We outline the implications of our findings for research into the effects of trust and reputation in an online setting.

Acknowledgement

We gratefully acknowledge support by Ian Ippolito from RentACoder.com, who generously supplied us with the data for this project.
The cement of a virtual society

Many of the mechanisms that exist offline and ensure that an interaction between people runs smoothly are not available in online interaction. For instance, in online interactions a large “shadow of the future” cannot easily be guaranteed: who knows whether you are going to deal with the people in this online help forum again, and who knows whether you will be buying from the same online reseller again? In addition, because the interacting partners can be geographically dispersed it is often impossible to sanction a partner who has misbehaved towards you, even if you manage to discover the real world identity of the virtual-world partner you were dealing with. In the absence of such mechanisms, interactions online tend to reduce to single-shot social dilemmas with high risk of opportunistic behavior and, as a result, a high risk of a possible break-down in such potentially profitable interactions.

Reputation systems, used by online auction sites such as eBay but also by other sites such as Amazon or Yahoo, provide one online mechanism that can potentially increase trust between interacting partners online. The lack of trust between partners that can arise because of the online setting of the interaction might be compensated for when both partners can judge the trustworthiness and capabilities of the partner through their reputation. Many of us will recognize such a reputation system from eBay. When you have bought or sold something on eBay, both buyer and seller get the opportunity to leave a “feedback score” (on eBay this is either a +1, 0, or –1) and a brief comment about the business partner. One’s total score and the comments are made public and are available to all other buyers and sellers to see. The underlying idea is, of course, that buyers and sellers value their reputations, find it worthwhile to invest in them, and abstain from the opportunistic behavior (such as selling junk or refusing to pay) that they would otherwise be tempted to engage in. Although sites differ in the details of the reputation system they use, most of the larger sites which use reputation systems tend to follow a procedure that is similar to the one used by eBay.

Several studies have tried to assess whether or not reputation systems might work, and under which circumstances. Virtually all empirical studies focus on two dependent variables: the probability of sale and the selling price. Strangely enough, the empirical evidence in favor of reputation systems as a safeguard against opportunistic behavior is not that overwhelming. Some researchers find no positive effects of a good reputation on probability of sale and selling price at all and others find statistically significant but substantially insignificant effects. Only a handful of researchers find substantial effects in the expected direction, for instance a positive correlation between the reputation score of a seller on eBay and the selling price of the product (we provide a brief overview in the next section). This general lack of empirical support, combined with the current boom in online interaction, poses a tantalizing problem, part of which we hope to resolve in this chapter. As we will argue below, eBay data pose several problems to adequately test effects of reputation when one considers probability of sale and selling price as the dependent variable. Moreover, what is crucial but generally neglected is that arguments about the effects of reputation concern how much buyers would be willing to pay extra for a product from a seller with a good reputation, rather than how much they have to pay extra for that good. The latter depends on many, often unmeasured, factors such as the market conditions, which makes using the actual selling price as the dependent variable a poor proxy of what we actually should measure, the willingness to pay extra for an additional unit of reputation. We show that when you look in an online context where these problems of data quality and availability are less severe, the positive effect of reputation on the willingness to pay does indeed surface.
Some issues regarding reputation scores on eBay

Most of the empirical work on the effects of reputation systems on sales in online interaction is based on eBay data. One reason for this is that eBay is by far the biggest auction site, representing 203 million users and a gross volume of 12.9 billion US dollars in 2006 (eBay, 2006). This suggests that whoever can understand eBay, understands a large part of the financial online consumer world. Another reason is that eBay auction data are relatively easy to collect, either by hand or automatically through software that ‘spiders’ the web site. For the few who have never heard about eBay auctions: on eBay, the seller puts a product online for a specific amount of time (often 7 days) and buyers can bid against one another. The one with the highest bid gets the product and pays according to his bid. In fact auctions allow several possibilities that seriously complicate the matter, such as private reserved prices (a minimum amount the seller wants to make), starting prices, and “Buy it now” prices (prices for which the seller is willing to end the auction and sell immediately), but let us not consider those for now although they may have strong effects on probability of sale and selling price. As mentioned before, after each transaction, buyer and seller get the possibility to leave a “feedback score” and a brief comment about the business partner. The number of (unique) positive evaluations minus the number of (unique) negative evaluations forms a person’s public reputation score.

In most empirical applications, the data collected from eBay include as the dependent variable whether a seller sold the product and, if yes, at what price. Researchers vary considerably with respect to which operationalization of a seller’s reputation they use (number of positive comments, number of negative comments, percentage of negative comments, etc.) and with respect to which other characteristics they control for in their analyses (duration of auction, shipping costs, product characteristics, etc). Most of the empirical papers use some form of regression analysis to analyze the data.

The most straightforward effects researchers aim to find, a better reputation leads to a higher probability of sale and a higher selling price, are not always found. Some researchers even come up with highly unexpected results, such as a negative reputation having a positive effect on an auction’s end price (Eaton 2002; Kauffman & Wood 2000). Dellarocas (2003; Table 2, page 1412) provides a recently published overview; many working papers on the topic are available from the Reputations Research Network (http://databases.si.umich.edu/reputations). In the majority of cases there is a positive effect of the reputation score on selling price, but the size of the effects is generally small. Several studies show no effect of the reputation score on price. Similar inconsistencies hold for negative reputation scores. They sometimes have a negative effect on the probability of sale and selling price, and when they do they are often larger than the effect of a positive reputation, but just as often we do not find any significant effects of negative reputation. It seems that the effects are stronger when the goods are less standardized (for instance when they are second hand instead of new, or complex rather than simple), which makes sense intuitively. More recent studies have reported relatively strong effects, for instance Grund and Gürtler (2006) find in an empirical analysis of DVD sales on eBay that there are substantial negative effects of the percentage of negative ratings on the selling price. McDonald and Slawson Jr. (2002) find a positive effect of reputation on selling price in the online sales of Barbie dolls.

There are several reasons for the variation in results found in previous studies (as mentioned in Lee, Im, and Lee, 2000 and Snijders and Zijdeman, 2004). First, there are differences with
respect to the kind of product that is being sold in the auction data. Some studies consider only new products others only consider used products. And, some studies analyze several different products at the same time, whereas others analyze data about one kind of product. As argued and shown in Snijders and Zijdeman (2004) the latter can have an impact on the results of the analyses in the sense that it is more likely to find statistically significant results when only one kind of product is analyzed. In addition, it is likely that the larger the prices of products are and the more abundant the supply of sellers, the more a buyer will be inclined to use every bit of information that he can get, including the reputation score of the seller. Another important factor that might lead to results that are not consistent across studies lies in the way in which reputation is operationalized. Some studies consider the difference of (unique) positive minus (unique) negative ratings – which is what eBay reports as a person’s reputation score – whereas others use the absolute number of negative ratings or the percentage of negative ratings as well, in some cases log- or otherwise transformed. Third, the way in which data were collected could contribute to the mixed results. While some researchers control for variables concerning product condition and others control for transaction related variables like shipment costs and means of payment, still others control for few, if any, other possible intervening variables. Finally, as Resnick, Zeckhauser, Swanson, and Lockwood (2002) have already pointed out, researchers have been using different statistical methods of analysis.

Actually, another reason why using eBay data (or online data in general) is considered an attractive way of doing research is that, in principle, the researcher has pretty much the same data available as the eBay-user. One could therefore hope that, whatever the reasons are for the behavior of the buyer or seller, a large part of the information the buyer or seller can use as a basis for this behavior, is in fact available to the researcher. In that sense, using online data has a natural place in between a standard lab-experiment (where treatments are completely under the control of the experimenter) and the survey (where the researcher tries to measure anything that can be thought to cause or correlate with the behavior of the respondent). However, the analysis of eBay data, or data from an auction site with a similar setup, is not that straightforward, and the data are not as close to the underlying theory as one might expect from looking at research reports that use them.

The first thing to note about auction data as can be found on eBay is that they do not fully encompass the underlying arguments that ‘should’ be involved according to auction theories. A typical person bidding on an item will choose his bid based on the value of the good to himself and on the expected maximum bid of other potential buyers. That is, to really statistically model the situation as close to the actual decision as possible, one needs to include how many other potential buyers the focal buyer expects in this particular auction, or in any case some kind of measure of the expected maximum of the other buyers’ bids as experienced by the focal buyer. What makes the matter even more complicated, is that it may very well be the case that a potential buyer can simultaneously consider similar auctions (that is, auctions for the same or a similar product), or can even decide to postpone buying until a similar auction comes along. Usually, such additional data are not available or hard to estimate on the basis of the data that are available online.

Other incompleteness issues concern the fact that giving feedback is voluntary, so that an analysis of reputation data should consider selectivity of the sample or assume it away. It is also possible to change identities or create a high reputation score by engaging in a lot of relatively small transactions. But even when we accept the incompleteness of the data, and simply forget about such strategic and other arguments playing a role, there remains the issue
the most appropriate mode of statistical analysis. First, because not all products get sold, there are reasons to model auction data using Heckman selection models: some factors influence whether or not the good gets sold, other and partly the same factors determine the selling price given that the good gets sold. Finding factors that determine one but not the other, are difficult to find, which complicates estimating these models reliably. Since the selling price typically underestimates the value of the good to the buyer – the buyer might have been willing to pay more – there are also reasons to use tobit analysis or interval regression when the selling price is the dependent variable. However, in analyses with such ‘truncated samples’, the consequences of heteroscedasticity are severe (Hurd, 1979), which implies for instance that one needs to be careful not to analyze loads of auction data regarding many different products in a single analysis (cf. Snijders & Zijdeman, 2004).

Thus, there are several reasons why simply running OLS analyses on auction data – even when you have thousands of data points – might not immediately reveal the ‘real’ underlying effects of reputation on auctions. Several attempts are being made to connect theoretical auction models and empirical data by making use of structural modeling. For example, Wang, Jank & Shmueli (2004) use the bidding data to infer the distribution of other bids as assumed by the bidders and Gonzalez, Hasker & Sickles (2004) estimate and control for the number of potential bidders for an auction.¹

Our attempt is different. At its core, the issue is whether a buyer values a positive reputation of a seller. That is, would a buyer be willing to pay more for the product when the seller has a better reputation score. This willingness to pay might depend on other factors, such as the condition of the product, shipping costs, or on transaction costs such as whether payment through PayPal is possible. However, the selling price (and the probability of sale) depends on many other unmeasured factors, such as other buyers’ willingness to pay for the product, the number of suppliers of similar goods, and other market conditions such as the degree to which buyer and sellers have access to information about the market itself. While we do not deny that probability of sale and selling price are interesting in their own right – what a good reputation is worth in the market place is certainly an important question – such analyses should not be confused with what the theory behind alleged effects of reputation is actually about. Of course it would be possible to measure willingness to pay in an experimental setting but this is throwing out the baby with the bath-water because then one of the main advantages of eBay data, the fact that the data concern real rather than artificial transactions, is lost. To estimate willingness to pay from behavioral data, data should include the options between which an agent can choose and the attributes of these options. In our case, this implies that our data must at least contain the price and the reputation of all the options between which the agent can choose. If one has data of this kind, the effect of reputation on willingness to pay can be estimated using the discrete choice approach (e.g., McFadden, 1974; Louviere, Hensher and Swait, 2000). Collecting such data is not feasible with eBay data, for instance because the set of alternative options is not well-defined and the information about which options have been considered by buyers is difficult to collect.² We therefore tried to find and analyze an online reputation system where the typical disadvantages of eBay data do not play a role, or at least to a much lesser extent. We found RentACoder.com. Our hypothesis is that

¹ Wang et al. do not consider effects of seller’s reputation score on price. Gonzalez et al. do find a non-linear effect of feedback rating on price (without distinguishing between positive and negative feedback).
² With the cooperation of eBay one could try to assess such information about buyers’ search behavior to determine the set of options, but even then the strategic arguments with regard to, for instance, the number of potential other buyers, is unavailable.
because the data provide a more seamless fit between data and underlying model, we will find a more pronounced positive effect of reputation there.

RentACoder.com

RentACoder.com is an internet site where people can offer their programming jobs to programmers. The procedure runs as follows. First a buyer puts a “bidrequest” online, describing the programming job in as much detail as possible, including a deadline for the project as a whole. This can range from designing a website, enhancing an algorithm, or creating a web-store, to coding scripts or parts of scripts in Perl, PHP or some other programming or scripting language. The buyer also determines a rough price estimate: either “very small business project” (<100 dollars), “small business project” (100-500), “medium business project” (500-5000), or “large business project” (>5000). Other options are “open for fair offer” and “unsure of project price”. For a fixed period of time, usually one or two weeks, programmers—called “coders” on the site—can respond to the bidrequest and mention the amount of money for which they would be willing and able to complete the job. All bids are visible to the buyer, but coders do not know the bids of other coders. During this phase, buyer and seller can communicate, for instance about specific details of the coding job, but in principle they can also haggle about the price. Coders can offer some general information about themselves; usually they only provide information about their capabilities and experience in programming (“I am an experienced programmer and am fluent in VisualBasic, C, and C++”). At any time, the buyer can decide to accept the bid of a particular coder and end the initial phase. As soon as the buyer has accepted the bid of a particular coder, the amount of money agreed upon is transferred to RentACoder. During the coding phase, the coder has the obligation to provide weekly status reports about the progress of the coding job. When the coding has been completed, the source code and other deliverables are transferred to the buyer who then acknowledges that the job has been completed, after which RentACoder transfers the money to the coder. RentACoder profits because it keeps 15% of the total amount agreed upon. RentACoder has been in business since 2001 and in the summer of 2006 it hosted about 60,000 buyers and 155,000 coders from all over the world.

Choosing between coders

As is the case on eBay, coders on RentACoder have a reputation score, but the scoring mechanism is slightly different. At RentACoder, after a coding job has been completed, buyer and coder rate each other on a scale from 1 (horrible) to 10 (excellent). The name of the coder always shows his or her average rating directly beneath it, so that buyers get an idea about how satisfied previous buyers have been about this coder. In addition, information about all previous coding jobs of the coder is available, and a page with personal information about the coder. Figure 1 shows an example of the kind of information available per coder.
Figure 1: Information about a coder as displayed on RentACoder.com.

As can be seen from Figure 1, the information about the coder includes the average rating (for PSergei this is 9.86), a quantile score comparing the coder with all other coders, sign up date, number of jobs completed thus far, number of jobs in progress and several other characteristics. Two elements of the information deserve specific attention. The first one is the number of arbitrations the coder has been involved in. Whenever buyer or coder are dissatisfied with the way in which the interaction proceeds, they can ask the website to intervene. This typically occurs when a coder has delivered the code to the buyer, but the buyer is for some reason not satisfied with it. Because all communication between buyer and coder can (and should) run through the website, an arbitrator can make a reasonable estimate as to who is right and what should be done about it. As it appears, PSergei has been involved in 10 such arbitrations and has lost none of them. The second factor is the country of residence. This can be an important characteristic for a buyer. For many programming jobs, communicating about what is needed has to be detailed and precise. It certainly helps a lot when a coder is fluent in English: the buyer can be more certain that the coder has understood the bidrequest and that when nevertheless trouble arises along the way it will be solved more easily. It is also important because it can be inconvenient to have to deal with a coder who is in a time zone that makes rapid email communication impossible (for instance when the coder is typically sleeping when the buyer is awake and vice versa). The kind of choice a buyer needs to make can be seen in Figure 2.
Table 1: Example of coders bidding on a programming job.

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Username</th>
<th>Location</th>
<th>Rating</th>
<th>Bid</th>
<th>View 1 message(s)</th>
<th>Reply</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/8/2006 8:45:48 AM</td>
<td>oztsov</td>
<td>Semenov, Russian Federation</td>
<td>Not rated yet.</td>
<td>Ranked #65,522 out of 255,649 (higher than 55.00% of their peers)</td>
<td><img src="#" alt="View" /></td>
<td><img src="#" alt="Reply" /></td>
<td>$50.00</td>
</tr>
<tr>
<td>8/8/2006 8:57:19 AM</td>
<td>Tirzer</td>
<td>Unterfohring, Germany</td>
<td>🥇 Excellent out of 10 ratings</td>
<td>Ranked #85,821 out of 155,592 (higher than 94.45% of their peers)</td>
<td><img src="#" alt="View" /></td>
<td><img src="#" alt="Reply" /></td>
<td>$50.00</td>
</tr>
<tr>
<td>8/8/2006 9:04:11 AM</td>
<td>firejump</td>
<td>Odessa, Ukraine</td>
<td>🥇 7.34 (Good) out of 47 ratings</td>
<td>Ranked #135,047 out of 155,592 (higher than 95.00% of their peers)</td>
<td><img src="#" alt="View" /></td>
<td><img src="#" alt="Reply" /></td>
<td>$15.00</td>
</tr>
<tr>
<td>8/8/2006 9:10:11 AM</td>
<td>source</td>
<td>City of Glasgow, United Kingdom</td>
<td>🥇 9.93 (Excellent) out of 16 ratings</td>
<td>Ranked #95,099 out of 155,983 (higher than 96.21% of their peers)</td>
<td><img src="#" alt="View" /></td>
<td><img src="#" alt="Reply" /></td>
<td>$50.00</td>
</tr>
<tr>
<td>8/8/2006 9:12:13 AM</td>
<td>Net#Team</td>
<td>Timisoara, Romania</td>
<td>🥇 8.62 (Excellent) out of 11 ratings</td>
<td>Ranked #7,621 out of 155,592 (higher than 95.00% of their peers)</td>
<td><img src="#" alt="View" /></td>
<td><img src="#" alt="Reply" /></td>
<td>$20.00</td>
</tr>
<tr>
<td>8/8/2006 9:40:30 AM</td>
<td>Artis</td>
<td>Donetsk, Ukraine</td>
<td>🥇 9.78 (Excellent) out of 19 ratings</td>
<td>Ranked #2,245 out of 155,592 (higher than 98.55% of their peers)</td>
<td><img src="#" alt="View" /></td>
<td><img src="#" alt="Reply" /></td>
<td>$50.00</td>
</tr>
</tbody>
</table>

Figure 2: Example of coders bidding on a programming job.

The programming job in Figure 2 was a relatively small task. Still, the bids varied between 15 and 59 dollars. In this case the lowest bidder ("firejump") did not get the job. Perhaps his relatively low rating made the buyer decide to choose an offer that was 5 dollars higher (from “Net#Team”, see Figure 2).

A mechanism as on RentACoder is not an auction in the same sense of the word as on eBay: the coder with the lowest bid need not get the job. However, the same underlying problems as on eBay apply. The buyer is not sure about whether or not the coder will deliver, and cannot be sure about the quality of the work. Detailed demands that are mentioned in the bidrequest can be checked upon delivery, of course: whether a program can do what it is supposed to do is relatively easy to find out. However, if the coder is not able to change his code in a satisfactory way, this leaves the buyer empty-handed even though he will get his money back. In other cases it might not be easy to check whether all demands of the bidrequest are met. For instance, how do you decide whether the coder has indeed designed a website that is “smooth and flashy” as suggested in the bidrequest. Often, several issues in a bidrequest must remain somewhat subjective and whether the coder will be responsive to comments during and after the coding process remains to be seen. In addition, in programming it is often important that the coding is transparent so that the buyer can understand how it works and can adapt it if the need arises. Again, the competencies and willingness of the coder to provide clear coding are typically hard to establish beforehand. That puts buyers at RentACoder at a
similar risk as buyers on eBay: you want to buy something but are not sure whether it will be
delivered at all and if yes in what condition the product will be. One way to decrease the risk
as a buyer is to stick to using coders with higher reputation scores.

Before describing the data in more detail it is useful to realize that Figure 2 shows that indeed
the data from RentACoder allow statistical modeling that is closer to the actual decisions
buyers are making. In the RentACoder transactions we basically see all other options that
would have been possible but did not materialize, which is what we need to estimate the
willingness to pay for good reputation.

Descriptive results from the RentACoder data

We analyze all public\textsuperscript{3} bidrequests between May 14, 2001 and January 15, 2005. During this
time a total of 30,742 completed transactions took place, 10,710 buyers and 40,697 coders
were active on the site. Table 1 shows the transactions per year, separately for all winning
bids and for only those winning bids larger than 20 US dollars. About 25% of the transactions
ended with a winning bid smaller than 20 US dollars (see Table 1). We exclude these from the
analyses because it is likely that these bidrequests deal with coding jobs that are so small in
size that a serious comparison of bidders is not likely to occur and because the differences in
bids will tend to be small.

<table>
<thead>
<tr>
<th>Year</th>
<th>Transactions</th>
<th>Transactions &gt; 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;number&gt;</td>
<td>&lt;%&gt;</td>
</tr>
<tr>
<td>2001</td>
<td>1,426</td>
<td>4.6</td>
</tr>
<tr>
<td>2002</td>
<td>5,102</td>
<td>16.6</td>
</tr>
<tr>
<td>2003</td>
<td>10,071</td>
<td>32.8</td>
</tr>
<tr>
<td>2004</td>
<td>13,877</td>
<td>45.2</td>
</tr>
<tr>
<td>2005</td>
<td>266</td>
<td>0.8</td>
</tr>
<tr>
<td>Total</td>
<td>30,742</td>
<td>100.0</td>
</tr>
</tbody>
</table>

We do not know how many bidrequests there have been on RentACoder that did not receive
any bid (the data contains only those bidrequests which actually received bids). Closer
inspection of the bids showed 922 excessive bids (amounts larger than 50,000 dollars). We
believe these are bogus bids and they were excluded from the analysis. From those
bidrequests that did receive at least a single bid, the average number of received bids is 11,
with a median of 7. Five percent of the bidrequests received 35 bids or more (max=908). Of
the 10,721 buyers who rented a coder at least once, 55% rented a coder once, 55% rented a coder once, and 11% rented coders at least 5 times. On average, buyers have rented a coder (not necessarily the same one)
about three times. A striking result is that 82% of the coders who have ever placed a bid, have
never been chosen, which suggests that this is a buyer’s market. Actually, there is a small
minority of about 750 coders (about 2% of all coders) who have completed 10 or more jobs.
Together these 2% of coders performed about 35% of the coding jobs. Table 2 gives an idea

\textsuperscript{3} It is also possible to arrange a private auction. That is, a buyer invites only a specific coder to bid on a
bidrequest. Such cases are excluded from the analyses.
of the value of the winning bids. A large number of the winning bids are smaller than 500 US dollars.

Table 2: Percentages of winning bids for different price categories

<table>
<thead>
<tr>
<th>Winning bid</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;20</td>
<td>27% (excluded from further analysis)</td>
</tr>
<tr>
<td>&gt;=20 but &lt;100</td>
<td>44%</td>
</tr>
<tr>
<td>&gt;=100 but &lt;500</td>
<td>25%</td>
</tr>
<tr>
<td>&gt;=500 but &lt;5000</td>
<td>4%</td>
</tr>
<tr>
<td>&gt;=5000</td>
<td>0% (26 cases)</td>
</tr>
</tbody>
</table>

One might think that the coder with the lowest bid is always chosen, but this is not the case. This can be seen by considering the cases in which the eventual winning bid is larger than 20 dollars, there are at least two bids (otherwise there is nothing to choose), and a unique minimum bid exists. We then find that minimum bids are chosen 28% of the time, whereas on average non-minimum bids are chosen only 8% of the time. We can conclude that to get the job it helps to have the lowest price, but it is certainly no guarantee. The strongest effect we find in these raw comparisons is when we compare a coder who has not done business with the buyer before with a coder who has done business with the buyer before. A coder who has done business with the buyer before has a 57% probability of being chosen, a coder who has not has a 3% probability.

Finally, we show the reputation ratings. As mentioned above, on RentACoder a reputation is a real number between 1 and 10, where 10 is best. The distribution of the reputation scores for coders when we consider all bidrequests can be seen in Figure 3a. When we exclude the cases with winning bids smaller than 20 dollars, we get the distribution as in Figure 3b.

Figures 3a and b: Distribution of the reputation scores

As Figures 3a and 3b show, the bulk of the coders have a reputation of 6.5 or larger, so that differences are really between 6.5 and 10. We also see that by restricting our analyses to winning bids larger than 20 dollars removes many of the coders with dubious reputations.

Analytical strategy

We analyze the data using conditional logit modeling. That is, the dependent variable is whether a certain bid is chosen or not, but we make use of the fact that we know that for each
bidrequest there are a certain number of bids of which only one is chosen. The underlying model is equivalent to McFadden’s choice model. That is, we assume that the latent attractiveness \( A_{ij} \) to buyer \( i \) of option \( j \) (in our case: the attractiveness to buyer \( i \) of a bid by coder \( j \)) can be written as

\[
A_{ij} = c_{0i} + c_{i}X_{ij} + \ldots + c_{pj}X_{pj} + \epsilon_{ij}, \quad j = 1, \ldots, n_j \text{ and } i = 1, \ldots, n
\]

where the attributes \( X_{hij} \) (\( h=1, \ldots, p \)) are characteristics of the bid (such as a bidder’s reputation score or the asking price), \( \epsilon_{ij} \) has an extreme-value distribution of the first type, and the attribute weights \( c \) are unknown coefficients. The buyer-specific intercept \( c_{0i} \) does not affect choice (it is constant within buyers) and cannot be estimated from choice data. Note that the number \( n_i \) of alternatives of buyer \( i \) is allowed to vary with \( i \). We then estimate the values \( c \) from the data by maximum likelihood under the assumption that the buyer chooses the option with the highest value of \( A \) (McFadden, 1974). As our predictors \( X_{ij} \) we use the variables in Table 3. We distinguish between characteristics of the coder and characteristics of the coder–buyer combination. By definition of the choice model, buyer characteristics are not included (the analysis is “within buyers”, so any characteristic that is constant within buyers cannot add to the model as a main effect, but interactions of characteristics of the buyer with attributes of the alternatives are possible).

**Table 3: Overview of variables used in the conditional logit analyses**

**Dependent variable**

| <Bid was chosen> | Whether bid was chosen (=1) or not (=0) |

**Coder characteristics**

| <nr. of past transactions> | Number of past transactions by coder. |
| <has a reputation> | Coder has a reputation (1=yes). |
| <reputation score> | Public reputation score on RentACoder (0-10, equal to 0 if no score) |
| <ln(amount)> | The natural log of the amount that was bid (in US dollars). |
| <length of bio> | Length of biographical information (measured in number of characters) |
| <coder jpg> | Coder has uploaded a photo or logo (1=yes). |
| <country of residence> | Country dummies for the US, Romania, Canada, UK, Pakistan, Russia, Ukraine, Australia (the countries of coders that occur most often) |
| <corruption index> | Corruption index 2004 of coder home country as indicated on Transparency International (www.transparency.org). Minimum Value in the data is 1.5, maximum is 9.7, with a mean of 4.7. Higher is less corrupt. |

**Characteristics of the (buyer, coder) pair**

| <past business> | Whether or not buyer and coder have done business before (1=yes). |
| <past business (n)> | How often buyer and coder have done business before (number). |
| <same country> | Buyer and coder are from the same country (1=yes) |
As an example, suppose we only consider \( X_1 = \text{amount} \), \( X_2 = \text{nr. of past transactions} \), and \( X_3 = \text{reputation score} \) as attributes. Suppose we fitted the model and found estimates \( c_1 = -1 \), \( c_2 = 2 \), and \( c_3 = 0.03 \) for the attribute weights. The willingness to pay of buyers – the exchange rate of money, so to speak – can then conveniently be written as:

\[
WTP_i = c_0 + \frac{2}{1} <\text{#past transactions}> + \frac{0.03}{1} <\text{reputation score}>
\]

so that then, for each unit of reputation, the \( WTP \) would increase by 3 dollar cents (irrespective of the unknown value \( c_0 \)). Note that in this example we used \( <\text{amount}> \) rather than \( <\ln(\text{amount})> \) for ease of exposition. With \( <\ln(\text{amount})> \) the WTP-expression takes the form:

\[
WTP_i = \exp(c_0 + \frac{2}{1} <\text{#past transactions}> + \frac{0.03}{1} <\text{reputation score}>)
\]

and then for each unit of reputation the \( WTP \) would increase by \( \exp(0.03)-1\times100 = 3\% \), again irrespective of the value \( c_0 \).

**Analyses**

You see the results of our analyses in Table 4. The models we use differ in the extent to which they use predictor variables, and with respect to the kinds of cases we include.
Table 4: Conditional logit analysis on the probability of being chosen for different models. Standard errors are adjusted for clustering on buyer, ignoring clustering on coders.

<table>
<thead>
<tr>
<th>Model</th>
<th>Ln(amount)</th>
<th>Past business (dummy)</th>
<th>Past business (n)</th>
<th>Number of past transactions</th>
<th>Coder has reputation score</th>
<th>Mean rating of reputation (0 if no reputation)</th>
<th>Buyer and coder are from same country</th>
<th>Corruption Index (higher=less corrupt) for coder country</th>
<th>Length of bio coder</th>
<th>Coder uploaded jpg</th>
<th>Coder from US</th>
<th>Coder from India</th>
<th>Coder from Romania</th>
<th>Coder from Canada</th>
<th>Coder from UK</th>
<th>Coder from Pakistan</th>
<th>Coder from Russia</th>
<th>Coder from Ukraine</th>
<th>Coder from Australia</th>
<th>Number of buyers</th>
<th>N</th>
<th>Pseudo-R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-0.7535***</td>
<td>1.7203***</td>
<td>0.1692**</td>
<td>0.0153***</td>
<td>-3.5882***</td>
<td>0.4825***</td>
<td>0.4687***</td>
<td>0.0485***</td>
<td>0.0232***</td>
<td>0.2181***</td>
<td>-0.1536***</td>
<td>0.0864**</td>
<td>0.2394***</td>
<td>0.1637***</td>
<td>0.0413</td>
<td>-0.4074***</td>
<td>0.0638</td>
<td>0.2262***</td>
<td>0.1753*</td>
<td>8,865</td>
<td>250,178</td>
<td>0.13</td>
</tr>
<tr>
<td>B</td>
<td>-0.7626***</td>
<td>1.7378***</td>
<td>0.1496*</td>
<td>0.0147***</td>
<td>-3.1337***</td>
<td>0.4340***</td>
<td>0.4038***</td>
<td>0.0515***</td>
<td>0.0121</td>
<td>0.2230***</td>
<td>-0.1375</td>
<td>-0.1425**</td>
<td>0.1924***</td>
<td>0.1222</td>
<td>0.0895</td>
<td>0.0415***</td>
<td>0.0502</td>
<td>0.0317***</td>
<td>0.2964*</td>
<td>8,850</td>
<td>249,118</td>
<td>0.14</td>
</tr>
<tr>
<td>C</td>
<td>-0.3566***</td>
<td>1.8814***</td>
<td>0.0857</td>
<td>0.0125***</td>
<td>-2.7130***</td>
<td>0.3882***</td>
<td>0.5029***</td>
<td>0.0453***</td>
<td>0.0317**</td>
<td>0.2230***</td>
<td>-0.1941*</td>
<td>-0.1356*</td>
<td>0.2697***</td>
<td>0.0891</td>
<td>0.0895</td>
<td>0.0475***</td>
<td>0.0493</td>
<td>0.0891</td>
<td>0.0839</td>
<td>3,857</td>
<td>70,116</td>
<td>0.12</td>
</tr>
<tr>
<td>D</td>
<td>-0.8612***</td>
<td>1.4004***</td>
<td>0.4819*</td>
<td>0.0159***</td>
<td>-2.9998***</td>
<td>0.4261***</td>
<td>0.7002*</td>
<td>0.0515***</td>
<td>0.0121</td>
<td>0.2230***</td>
<td>-0.1941*</td>
<td>-0.1356*</td>
<td>0.2697***</td>
<td>0.0891</td>
<td>0.0895</td>
<td>0.0475***</td>
<td>0.0493</td>
<td>0.0891</td>
<td>0.0839</td>
<td>3,555</td>
<td>76,947</td>
<td>0.13</td>
</tr>
<tr>
<td>E</td>
<td>-1.3467***</td>
<td>6.8926**</td>
<td>2.3176</td>
<td>0.0257***</td>
<td>-6.6255**</td>
<td>0.7996***</td>
<td>0.7996***</td>
<td>0.0453***</td>
<td>0.0317**</td>
<td>0.2230***</td>
<td>-0.1941*</td>
<td>-0.1356*</td>
<td>0.2697***</td>
<td>0.0891</td>
<td>0.0895</td>
<td>0.0475***</td>
<td>0.0493</td>
<td>0.0891</td>
<td>0.0839</td>
<td>285</td>
<td>7,117</td>
<td>0.13</td>
</tr>
</tbody>
</table>

* p<0.05; ** p<0.01; *** p<0.001

Model A = all cases (and basic predictors)
Model B = all cases (and extended list of predictors)
Model C = cases with 20 < maximum bid ≤ 100
Model D = cases with 100 < maximum bid ≤ 1000
Model E = cases with maximum bid > 1000

Consider Models A and B in Table 4. We see that the probability of being chosen as a coder is higher for coders who posted low bids, for coders who have done business with the coder before (the more often the better), for coders with more experience, and coders who uploaded a photo or logo on their personal information page. We also sometimes see a small effect of the length of the biographical information. In the model, we represent a coder’s reputation by two attributes: <has a reputation> and <reputation score>. Clearly, coders with higher reputations are preferred. In terms of the willingness to pay, as explained in the previous section, the results imply that buyers are willing to pay a whopping 40% to 50% more for a
coder with a reputation score that is one unit higher (note that the bulk of the reputation scores for the coders who have one, runs between 7.6 and 10).

Models C, D, and E in Table 4 are included to see how robust our results are for different kinds of coding jobs. Model C considers relatively cheap cases with a maximum bid as defined by the buyer between 20 and 100 dollars. Model D looks at the cases with a maximum bid between 100 and 1000 dollars, and Model E looks at the most expensive cases with a maximum bid of more than 1000 dollars. We find that the results with respect to reputation are consistent across the models in the sense that the estimated effect is positive, significantly different from zero, and substantially similar. It seems that buyers engage in relatively high stakes interactions mainly with coders with which they have done business before: the effect of this variable is stronger in Model E. Interestingly, when we added a dummy-variable <bid is the lowest bid> to the analysis of Model B, it showed no statistically significant effect.

The results concerning country differences are interesting in their own right. Across all models except Model E, we can rank order the countries of the coders by attractiveness for the average buyer from most to least attractive: Romania, Ukraine, Australia, Canada, Russia, UK, the reference category with all others, India, US, and Pakistan. These results strike us as somewhat counterintuitive or at least as not easy to interpret, given that the analyses control for effects of the bidding amount, the reputation of the coder, and so on. Additional analyses trying to unravel what is going on here are necessary.

We ran separate analyses to further assess the stability of the results under different implementations of the analyses (not reported in detail in this chapter). The results appear to be extremely robust. For instance, we restricted the analysis to only those cases where there are more than 5 (or 10, or 20) alternatives. We also estimated the model separately for bidrequests where all the bidding coders had not previously been engaged in business with the seller. None of these affect the substantial gist of our results: reputation has a positive impact on the probability of being chosen as the winning bidder.

In the above analyses, the effect of reputation on the willingness to pay is positive, significantly different from zero, and substantial. Now let us compare these results with the results when we analyze these data as is usual in most eBay-research, by only considering the eventual sales (disregarding the potential sales that could have occurred but did not). Table 5 shows the result of such analyses.
Our focus is, once again, on the effect of having a positive reputation. Table 5 shows that, although there still is a positive effect of reputation, its size has diminished dramatically to about 3.5% (it was over 40% in Table 4). Apparently, even though buyers value reputation highly, the structure of the market – for instance because there is an abundant supply of coders – is such that the net value of a high reputation is much smaller than what buyers would be willing to pay for it. In fact, this may very well be one of the main reasons why the effects of reputation that are found in the empirical literature differ across analyses and are generally small. Such analyses can only consider what reputation is worth in the market (the amount you have to pay extra for a seller with a high reputation), which is not the same as the extent to which buyers value reputation (the amount a buyer would be willing to pay extra for a seller with a high reputation).

Table 5: Regression analysis on the natural log of the amount paid for the coding job. Standard errors are adjusted for clustering on buyer, ignoring clustering on coders.

<table>
<thead>
<tr>
<th></th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past business (dummy)</td>
<td>-0.1603***</td>
<td>-0.1607***</td>
<td>-0.1635***</td>
</tr>
<tr>
<td>Past business (n)</td>
<td>-0.0410</td>
<td>-0.0411</td>
<td>-0.0399</td>
</tr>
<tr>
<td>Number of past transactions</td>
<td>-0.0022***</td>
<td>-0.0027***</td>
<td>-0.0027***</td>
</tr>
<tr>
<td>Coder has reputation score</td>
<td>-0.2985*</td>
<td>-0.3018*</td>
<td>n.a.</td>
</tr>
<tr>
<td>Mean rating of reputation (0 if no reputation)</td>
<td>0.0342**</td>
<td>0.0342**</td>
<td>0.0367**</td>
</tr>
<tr>
<td>Buyer and coder are from same country</td>
<td>-0.0036</td>
<td>-0.0407</td>
<td></td>
</tr>
<tr>
<td>Corruption Index (higher=less corrupt) for coder country</td>
<td>-0.0087</td>
<td>-0.0092</td>
<td></td>
</tr>
<tr>
<td>Length of bio coder</td>
<td>0.0001***</td>
<td>0.0001***</td>
<td></td>
</tr>
<tr>
<td>Coder uploaded jpg</td>
<td>0.0875***</td>
<td>0.0889***</td>
<td></td>
</tr>
<tr>
<td>Coder from US</td>
<td>-0.0103</td>
<td>-0.0327</td>
<td></td>
</tr>
<tr>
<td>Coder from India</td>
<td>0.0440</td>
<td>0.0491</td>
<td></td>
</tr>
<tr>
<td>Coder from Romania</td>
<td>-0.0774**</td>
<td>-0.0826**</td>
<td></td>
</tr>
<tr>
<td>Coder from Canada</td>
<td>0.0007</td>
<td>-0.0061</td>
<td></td>
</tr>
<tr>
<td>Coder from UK</td>
<td>0.0035</td>
<td>-0.0055</td>
<td></td>
</tr>
<tr>
<td>Coder from Pakistan</td>
<td>-0.1007*</td>
<td>-0.0712</td>
<td></td>
</tr>
<tr>
<td>Coder from Russia</td>
<td>-0.0195</td>
<td>-0.0538</td>
<td></td>
</tr>
<tr>
<td>Coder from Ukraine</td>
<td>0.1929***</td>
<td>0.2121***</td>
<td></td>
</tr>
<tr>
<td>Coder from Australia</td>
<td>-0.1030</td>
<td>-0.0901</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>4.2621***</td>
<td>4.2173***</td>
<td>3.9060***</td>
</tr>
</tbody>
</table>

Model A = all cases (and basic predictors)
Model B = all cases (and extended list of predictors)
Model C = all cases where winning coder has a rating
Conclusion and discussion

The dominance of eBay as an auction site and the relatively easy availability of eBay data for research purposes have created the situation in which most of the empirical work about reputation mechanisms takes eBay as its focus. This has several disadvantages: the available data are not as close to the actual decision making of the buyer as one might think, which makes estimating the effect of reputation on probability of sale or on auction end price problematic. Our idea was that this may very well be the reason why researchers find general inconsistencies in the empirical results, in particular, the lack of consistent support for a positive effect of reputation on both probability of sale and end price. We examined whether a positive effect of reputation exists in online data where the data are much closer to the actual decision making, by making use data from the online programmer’s site RentACoder.com. In this case, we do find strong support for a positive effect of a positive rating on the probability of being chosen as the preferred coder. That is, buyers value a positive reputation in this market.

An important issue that needs to be highlighted is that the analyses as presented here crucially differ from the kinds of analyses used when the data come from eBay (or other auction platform) in terms of what it is they estimate. In our analyses, we assume the underlying logic of McFadden’s choice model and hence model the implicit attractiveness of a bid by a coder as perceived by the buyer. In this model, what we estimate is the effect of reputation (and price, etc) on the perceived attractiveness of the coder’s bid. That is, we directly estimate a seller’s preference: how much would a seller be willing to pay extra for a partner with a better reputation. This is fundamentally different from what is being calculated in most other publications where one uses the prices of the completed eBay-auctions and correlates them with the reputation score. In such cases the researcher estimates what (on average) a buyer must pay extra to find a partner with a better reputation. Arguments about the positive effects of reputation are in essence about the preferences of individuals over the reputation scores of others. This makes using data from RentACoder a more appropriate test of the reputation mechanism. In any case, we show that the results strongly differ. Although buyers are willing to pay up to 40% extra for coders with a good reputation, the market conditions are such that they only have to pay about 3 to 4% extra for a coder with a good reputation. Although our data are not necessarily representative for interactions on eBay, our results clearly show two important issues. First, one should clearly distinguish theoretically between the effect of reputation on the willingness to buy on the one hand and the effect of reputation on the selling price on the other. These are different entities, and although reputation arguments are basically about the former, they are usually tested with data on the latter. Second, this discrepancy could be one of the main reasons why the effects of reputation on selling price (or probability of sale) are so erratic and generally small. We find something similar in the RentACoder data: a high willingness to pay for reputation, but given the market there is no need to do so.

Although our data allow estimating models that are closer to the real life decision-making of people, they are still far from perfect and partly suffer from the same problems as eBay-data do. For instance, we completely disregard strategic bidding by coders, who might anticipate how many other coders will participate. Coders might also offer extremely low bids to buyers with whom they have never done business before, in the hope that the buyer will offer them more business after a successful initial coding job. Nevertheless, our results do suggest that thus far the effect of reputation on trust between buyer and coder in an online environment
may have been underestimated, and that incorporating market conditions in the analyses is likely to make the effect of reputation on trust more salient, as was the case in our data.
References


