

Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors*

Gregory Lewis[†]
Department of Economics
Harvard University

February 3, 2009

Abstract

Since Akerlof (1970), economists have understood the adverse selection problem that information asymmetries can create in used goods markets. The remarkable growth in *online* auctions of used goods, where buyers generally purchase sight unseen, therefore poses a puzzle. I argue that part of the solution is that sellers voluntarily disclose their private information to buyers through photos, text and graphics on the auction webpage. In so doing they define a precise contract between buyer and seller — to deliver the car shown — and this helps protect the buyer from adverse selection. Extending previous theoretical work by Jovanovic (1982), I model the impact of contractible disclosure and changes in disclosure costs on performance and adverse selection on online auction platforms. To test this theory, I examine data from eBay Motors. I find first that sellers selectively disclose information; second that this reduces information asymmetry; and finally that disclosure costs impact both the level of disclosure and the prices obtained by sellers, and consequently incentives for seller participation.

*This is a heavily revised version of my job market paper. I would especially like to thank Pat Bajari for his support and advice on this paper. I am also grateful to Dan Akerberg, Jim Adams, Susan Athey, Matt Backus, Tilman Börgers, Phil Haile, Oliver Hart, Han Hong, Hugo Hopenhayn, Ali Hortaçsu, Guido Imbens, Jakub Kastl, Kai-Uwe Kühn, Jim Levinsohn, Julie Mortimer, Serena Ng, Scott Page, Ariel Pakes, Nicola Persico, Rob Porter, Dan Silverman, Jagadeesh Sivadasan, Doug Smith, Lones Smith and Charles Taragin; and for excellent research assistance from Chris Sullivan.

[†]Email: glewis@fas.harvard.edu

1 Introduction

The rise of the internet has seen a huge rise in the volume of used goods traded online. Online auction sites such as eBay and Yahoo! Auctions compete worldwide with specialized listing sites such as usedcomputer.com and cars.com in the retail trade of consumer goods. Meanwhile, business to business transactions totaling billions of dollars take place through online auctions in industries as diverse as aviation and mining.¹ At first glance, this growth is somewhat surprising. Since Akerlof's classic paper, economists have been aware of the potential for adverse selection in markets with information asymmetries, such as used good markets. Information asymmetries are exacerbated in online transactions, where the buyer typically does not see the good in person. Why then has the volume of trade in these markets proved so robust to adverse selection?

One potential explanation is that sellers themselves endogenously limit information asymmetries, by voluntary online disclosure of their private information. Where disclosure is public, such as through the listing webpage, it helps define the contract between buyer and seller, which is that the seller will deliver the item described in the listing. Indeed, given sufficiently rich and enforceable contracts, the initial information asymmetry should play no role in determining the performance of the market. Two factors prevent this from happening. First, not all of the seller's private information can be verified ex-post, and therefore contractually enforced. Second, disclosure may be costly to the seller, so that disclosure is limited and contracts are coarse. The presence of disclosure costs may therefore impact market outcomes.

In this paper, I examine the role of disclosure and disclosure costs in online used goods markets, taking as a case study eBay Motors, the largest used car marketplace in the United States. In this market, the stakes are high for both sides, the information asymmetries are substantial, and yet there is a high volume of trade with approximately 36000 cars sold each month.² By developing a new model of adverse selection under costly disclosure, I show that higher disclosure costs adversely affect the distribution of seller types participating in the market. Turning to the data, I

¹For example, DoveBid.com holds auctions for aviation and mining equipment, with over \$5 billion in sales thus far (source: <http://www.dovebid.com/company/introduction.asp>).

²Source: eBay Press Releases (<http://investor.ebay.com/releasedetail.cfm?ReleaseID=206868>, <http://investor.ebay.com/releasedetail.cfm?ReleaseID=306677>).

find evidence first that photos and text posted by the seller on the auction webpage influence demand; second that disclosure costs affect how much information the seller decides to post; and finally that higher disclosure costs lead to lower average prices. I conclude that disclosure costs — whether caused by technology, bandwidth, or time costs — are an important determinant of the success of online goods marketplaces.

The paper is in three parts. First, I model the relationship between disclosure, disclosure costs and adverse selection on an online auction platform. Sellers have two-dimensional type, knowing both their outside (offline) option for the vehicle and the quality of the vehicle. Disclosing this quality is costly. The model nests both the canonical Akerlof (1970) model of adverse selection, and the case of information “unravelling” (Grossman and Hart (1980), Grossman (1981), Milgrom (1981)) as special cases. I characterize the equilibrium listing, disclosure and secret reserve setting strategies of different seller-types, and analyze bidding in the presence of a subtle Winner’s curse effect for bidders who meet the secret reserve. I also deduce an intuitive comparative static: that as disclosure costs increase, the expected quality of vehicles listed falls, both because marginal high quality sellers select out, and because marginal low quality sellers select in.

Second, I test for a relationship between disclosed information, and auction outcomes such as prices, reserves and sales rates. I make use of a new dataset with over 90000 eBay Motors listings, including variables relating to item, seller and auction characteristics. I measure the quantity of disclosure by counting the number of photographs provided by the seller, and also by looking for key phrases in the text. By running a series of hedonic regressions, I show that these measures are significantly correlated with price, reserves and sales rates. The estimated coefficients are large and significant, and are robust to a wide variety of specifications. The results suggest that the webpage content affects demand, and moreover that seller disclosures are selective, so that higher quality cars are associated with more information.

Third, I examine the role of disclosure costs. To do this, I compare dealers who upgrade their auction listing software during the sample period to those who do not. Improved listing software greatly reduces the cost of putting together an auction webpage, and is a good cost shifter. I find that relative to non-upgraders, those who upgrade their software put up far more photos after upgrading than they did before, showing that disclosure costs affect the level of disclosure. Next, I use software as

an instrument for the number of photos, and verify that the number of photos has a causal effect on price. This makes sense since with better software, sellers of high quality cars can better demonstrate this quality and receive higher bids; and those with low quality cars can behave as before, putting up few photos. A natural concern is that the decision to upgrade software is endogenous, and therefore may be correlated with the quality of future listings. I run a number of robustness tests to account for this possibility, and find no such effects. I conclude that differences in disclosure costs do result in differences in equilibrium outcomes, as the theory predicts.

This paper is closely related to both the theoretical and empirical literature on disclosure. On the theoretical side, the most similar paper is Jovanovic (1982), which also emphasizes the relationship between disclosure costs and welfare. In that paper, seller types are characterized by a single dimensional private signal that indicates their value to owning the vehicle. Sellers must make inferences about how buyers will value the vehicle based on their own signal. This paper differs in that I allow sellers to have two dimensional type, knowing both their outside option and the “quality” of the vehicle, where “quality” is defined by how the average informed buyer perceives the value of the car. The model therefore allows for the possibility that the value of the outside option (e.g. a local sale) may be quite different from value assessed by online buyers. Other related theory papers are Akerlof (1970), Grossman and Hart (1980), Grossman (1981) and Milgrom (1981).

On the empirical side, there is a diverse literature on both mandatory and voluntary disclosure (see e.g. Mathios (2000), Jin and Leslie (2003) and Jin (2005)). A closely related paper is Jin and Kato (2006) who examine the sale of baseball cards on eBay, and conclude that seller claims about baseball card quality are often not truthful. We will see that in the used car market the stakes are higher, and so buyers generally verify the information before payment, limiting incentives for sellers to lie. Relative to this disclosure literature the contribution of the paper is to document a case in which voluntary disclosures partially unravel the information asymmetry, and also to show empirically that disclosure costs determine how much unravelling occurs.

With the notable exceptions of Jin and Kato (2006) and Yin (2006), the online auctions literature has ignored seller disclosure and instead focused on the role of the seller feedback mechanism (e.g. Resnick and Zeckhauser (2002), Houser and Wooders (2006)). In this market, I find that seller feedback ratings play relatively little role in

influencing prices. While this certainly does not undermine the well-established results on the importance of seller feedback in other markets, it does suggest that other market design features, such as the disclosure mechanism, can also be important determinants of the success of online markets with information asymmetry. A last set of related papers look directly for evidence of adverse selection in used good markets (e.g. Bond (1982), Genesove (1993)). In particular, Adams, Hosken and Newberry (2006) looks directly for adverse selection in the eBay market for used Corvettes, and finds little evidence of it. This paper sheds little direct light on this issue, emphasizing instead how disclosure may help to limit adverse selection.

The paper proceeds as follows. Section 2 describes the market, while section 3 introduces the theory. Section 4 provides a description of the data, the estimation strategy and the estimation results. Section 5 concludes.³

2 eBay Motors

eBay Motors is the automobile arm of online auctions giant eBay. It is the largest automotive site on the Internet, attracting 11 million unique visitors to its site each month, and commanding 34% of all minutes spent online on automotive sites.⁴ Every month, approximately 36000 vehicles are sold, a rate just slightly slower than a car a minute. This trading volume dwarfs those of its online competitors, the classified services cars.com, autobytel.com and Autotrader.com. In contrast to these sites, most of the sellers on eBay Motors are private individuals, although dealers still account for around 30% of the listings. Another big difference is that a large proportion (75%) of vehicles are sold to out-of-state buyers. Because of this, bidders can typically neither examine the car in person nor rely entirely on the seller's reputation, and must rely on the information on the auction webpage.⁵

Listing a car on eBay Motors is straightforward. For a fixed insertion fee of \$40, a seller may post a webpage with photos, a standardized description of the car, and a more detailed description that can include text and graphics. An additional transaction

³Proofs and tables of regression results are to be found in the appendix.

⁴Source: Nielsen//NetRatings.

⁵Source: Auction123 (<http://www.auction123.com/ebayadvantages.aspx>).

Item Specifics - Cars & Trucks

Chevrolet : Corvette LINGENFELTER

01 CORVETTE Z06 LINGENFELTER TWIN TURBO 500+HP VIDEO!

Miles: 13879
 Transmission: Manual
 Warranty: No
 Title: Clear
 Condition: Used
 Engine: 8 CYLINDER 5.7 LITER MFI
 Fuel Type: Gasoline

Doors: 2 doors or more
 Interior: Red
 Year: 2001
 VIN Number: [1g1yv12sx15114360](#)
 Get the Vehicle History Report
 Exterior: Red
 Inspection: -

Standard Equipment:
 CASSETTE
 CD
 Z06 HARDTOP
 8 CYLINDER 5.7 LITER MFI

POWER WINDOWS
 POWER DOOR LOCKS
 2 WHEEL DRIVE

CRUISE CONTROL
 ABS
 POWER SEAT DRIVER

LEATHER
 AM/FM RADIO
 DUAL AIR BAGS

FEATURES AND OPTIONS

500 HP TWIN TURBO LS6 ENGINE
 Two Garrett true ball bearing turbochargers
 Two high efficiency air to air charge coolers
 Custom molded silicone air ducts
 4 into 1 custom exhaust manifolds
 Two cast turbo outlet adapters
 Turbocharger scavenge pump
 Turbo to drain reservoir
 Two K&N 360 degree conical air filters
 Heat shields and stainless clamps
 Professional installation, testing and tuning
 Turbocharger water feed and return hoses
 Turbocharger oil feed and return hoses
 Gaskets, spark plugs and fluids
 388 ci.
 CNC Porting and polishing of 6 liter cylinder heads
 Port matched LS6 intake manifold
 Lanard custom crank shaft
 JE forged custom aluminum pistons
 Lingenfelter custom hydraulic roller camshaft
 High performance valve springs & lightweight retainers
 Stainless steel one piece heavy duty intake valves

Inconel one piece heavy duty exhaust valves
 Three angle valve job, checking of spring tensions & heights
 Computer balanced rotating assembly
 Reassembly and blueprinting of engine
 50 lbs gm fuel injectors and high capacity fuel system with 18 volt adjustable fuel regulator
 B&B stainless steel exhaust system
 Professional installation, testing and PCM recalibration by Lingenfelter

The clutch was also upgraded to a Spec Stage III Clutch
 177 M/T Radiator

18 Inc Michelin Tires

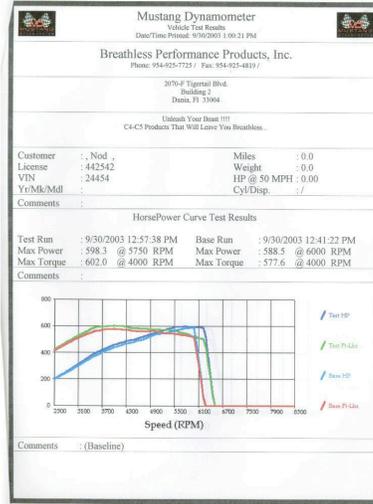


Figure 1: Information disclosed in an eBay Motors auction On this auction webpage, the seller has provided many different forms of information about the Corvette he is selling. These include the standardized eBay description (top panel), his own full description of the car's options (middle panel), and many photos (two examples are given in the bottom panel). The right photo is of the results of an independent car performance analysis done on this vehicle.

fee of \$40 is charged in the event of a sale and \$5 for the use of a secret reserve.⁶ The direct costs of posting photo, graphics and text-based content are negligible as text and graphics are free, while each additional photo costs \$0.15. Yet the opportunity costs are higher, as it is time-consuming to take, select and upload photos, write the description, generate graphics, and fill in the forms required to post all of these to the auction webpage. While these opportunity costs may seem small, the fact that professional car dealers typically invest in advanced listing management software to limit these costs suggests that they are not insignificant. Such software allows easier photo uploading and maintenance, graphics production and listing management, and is offered by companies such as CARad, eBizAutos and Auction123 at costs ranging from \$10 a listing to a flat \$300 a month fee.

For example, consider Figure 1, which shows screenshots from an eBay Motors auction for a Corvette Lingenfelter. The seller has provided a detailed text description of the features of the car and taken many photos of both the interior and the exterior. He has in addition taken photographs of original documentation relating to the car, including a series of invoices for vehicle modifications and an independent analysis of the car's performance. This level of detail appears exceptional, but it is in fact typical in most eBay car auctions for sellers to post many photos, a full text description of the car's history and features, and sometimes graphics showing the car's condition.

Once the seller has put the vehicle up for sale, people may bid on it. Some sales take place at a fixed price, but the vast majority of cars are sold in an English auction format with a proxy bidding system.⁷ Potential buyers can communicate with the seller throughout the auction process, either through e-mail or by phone (if the seller has provided a phone number). This allows bidders to query the seller on particular details. The vast majority of winning bids are placed towards the very end of the auction, as shown in Figure 2. At the close of the auction, the highest bidder receives the car, provided that the secret reserve (if any) has been met. The use of secret reserve prices is widespread on eBay Motors.⁸ As noted by eBay, "most buyers opt to pickup the vehicle in person."⁹ The result is that much of the information provided

⁶This has changed since the time my data was collected: now listings are free, but sellers are charged \$125 for each successful sale, and \$7 if they use a secret reserve.

⁷The fixed price formats are either "best offer" or "buy-it-now" auctions.

⁸At least 55% of the auctions in my data have secret reserves: since I do not observe the existence of a secret reserve when it has been met, this is a lower bound.

⁹Source: eBay Motors Seller's Guide, <http://pages.motors.ebay.com/howto/selling/closeB.html>

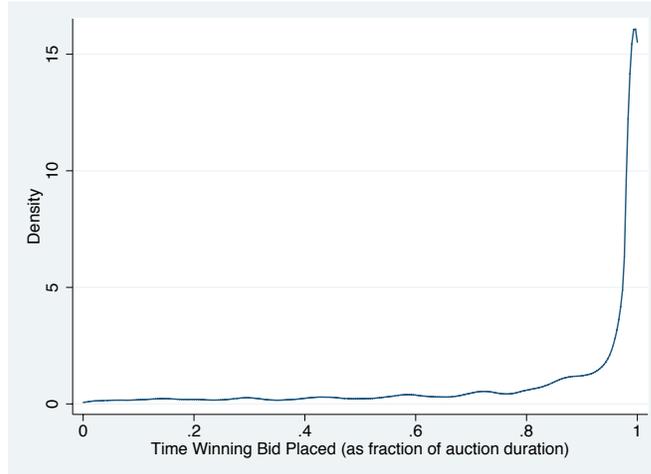


Figure 2: **Timing of Bids.** This shows the estimated density of the time the winning bid was placed, as a fraction of auction duration. 75% of winning bids are placed in the last day of the auction, and 25% in the last hour.

by the seller is often verifiable before payment is made.¹⁰

The webpage defines the contract between the buyer and the seller, which is that the buyer agrees to purchase the vehicle described by the seller at the final closing price of the auction. The principle of *caveat emptor* applies: it is the buyer's responsibility to ask questions about details not listed on the webpage before bidding.¹¹ On the other hand, any material misrepresentations by the seller constitute fraud, so sellers have weak incentives to lie. In contrast to private car sales where it may be difficult to establish exactly what the seller did or did not promise, the webpage is stored by eBay for at least 20 days after the sale, so that these online transactions have a clearly defined contract in the event of a dispute. Rich media such as photos and videos may define the contract terms more precisely than text. Together, these act to limit the asymmetric information problem faced by bidders. In the next section, I model how this occurs, and how it affects the quality distribution of listed cars.

¹⁰Even when the seller ships the vehicle to the buyer, payment is often held in escrow (e.g. through paypal) until the buyer has had a chance to examine the vehicle.

¹¹Experienced sellers often explicitly include a boilerplate disclaimer of this form.

3 Theory

In this section, I develop a theory of adverse selection under costly information disclosure, adapted to the eBay auction framework. There are two key elements of the model. First, I introduce information asymmetry and costly disclosure. I assume that the seller privately observes an ex-post verifiable signal of the value of the car — the car “quality” — but the buyers do not. The seller can publicly disclose the quality of the car, by posting information on the auction webpage, but doing so is costly. For example, the seller may take and upload photos of the car, invoices for parts, service history etc, but this is time consuming. Such disclosures are assumed to be truthful, given that they can be verified ex-post and that therefore any misrepresentation may lead to fraud charges. The buyer’s valuations of the object are influenced by the information posted, resulting in an auction with interdependent values.

Second, I model eBay auctions as second-price sealed bid auctions with a secret reserve price and an unobserved number of participants. This sealed-bid abstraction is motivated in Bajari and Hortacısu (2003), where the authors argue that although eBay operates as an ascending (English) auction with re-entry, it is an equilibrium for all bidders to wait until the end of the auction and bid only then, a practice called “sniping”. The two auction formats become strategically equivalent if this is the case. As shown earlier, last minute bidding is prevalent in this market, so this seems like a reasonable modeling choice. I allow the seller to set a secret reserve price in advance of the auction — this is also motivated by the data. As on eBay, I assume that the number of participants is unknown by either the seller or the buyers, but is instead drawn from a known distribution.¹² One key simplification is that I consider the market to be a series of separate eBay auctions, rather than modeling them all at once (e.g. modeling the matching of participants to individual auctions). With that assumption, I can deduce how disclosure costs affect market performance by looking at the change in expected outcomes in individual auctions. It would obviously be better to account for potential interdependencies, but this would take us far beyond the scope of this paper.

Before diving into the full model, let’s look at a simple example that shows how costly

¹²This imposes that the number of potential participants is independent of any information disclosed by the seller. Endogenizing participation would require a richer model of the auction market; but in the empirical section I try to control for endogenous participation.

disclosure falls “in between” the benchmark cases of adverse selection and information unravelling. Suppose there are three car types, the peach (high), apple (middle) and lemon (low). Each is equally likely ex-ante. The peach is worth \$2500 to buyers, and \$2000 to sellers; the apple is worth \$1800 to buyers and \$1500 to sellers; and the lemon is worth \$1100 to buyers and \$1000 to sellers. Assume that there are “many” buyers, so that sellers have all the bargaining power. Then if verifiable disclosure is costless, all cars will trade at the buyer valuations. If no disclosure is possible, then quick calculations show that only lemons will trade. Now suppose that the car type can be disclosed at a cost of \$400. In that case peaches will disclose, and trade. But apples find it neither profitable to disclose nor to pool with the lemons so they select out. Lemons will not disclose, and will trade. So the outcome is “in-between” in terms of welfare, since two of the three types trade with costly disclosure.

Having gained some intuition from the example, now consider a formal model. A seller S has a car to sell, and has some idea of the quality of the car, captured in a private signal Q . This signal should be thought of as a quality index, encapsulating characteristics such as repair history, exterior condition and so on that are ex-post verifiable. She also has an outside option for selling the car, V_S , privately known to her. The seller’s type is the pair (Q, V_S) , and has joint distribution F_{Q, V_S} with bounded support. The variables Q and V_S are affiliated, so that on average, the better the car is, the better the seller’s outside option.

The seller must decide whether to list the car on eBay. If so, N buyers will participate in an a second-price sealed bid auction for the car, where the number of bidders N is unknown to all participants, but has some known distribution F_N . The seller sets a secret reserve r , and so the car will sell if and only if the secret reserve is met, and if it is, the transaction price p will be the maximum of r and the second highest bid.

Buyers each have a private signal X_i about the car, independently drawn from a distribution F_X with strictly increasing and continuously differentiable hazard rate. These signals X_i are assumed to be independent of Q , and so these should be interpreted as idiosyncratic tastes for the car, rather than additional information.¹³ Buyers are assumed to be risk neutral with quasi-linear payoffs of form $v(Q, X_i) - p$ if they win, and zero otherwise. Their valuation $v(Q, X_i)$ is assumed to be strictly increasing and

¹³This is not an important assumption: the model can be extended to allow (Q, X_i) pairs to be affiliated.

continuously differentiable in both arguments.¹⁴

Notice that if Q were observed by all the bidders, this would be a conditionally independent private values (CIPV) auction model, since conditional on Q , the bidders differ only in their idiosyncratic tastes X_i . But instead the buyers observe the seller's public report ξ , equal to Q if she chooses to disclose, and equal to the null report \emptyset otherwise. Disclosure is costly, and the seller must pay some cost $c \geq 0$ to disclose signal Q . In addition, the seller pays a fixed cost F to list the car. Overall, the seller's payoff is given by:

$$\pi_S = \begin{cases} p - F - cd & \text{lists, object sells} \\ v_S - F - cd & \text{lists, object doesn't sell} \\ v_S & \text{doesn't list online} \end{cases}$$

where d is a binary variable equal to 1 if the seller discloses, and v_S is the realized outside option defined earlier.

The game takes place in two stages. In the first stage, the seller simultaneously decides whether to list the car, whether to disclose, and what secret reserve to set (if any). In the second stage, bidders observe the disclosed information ξ , and simultaneously bid according to strategies $\beta(\xi, x_i)$. Because the reserve r is secret it does not directly enter the bidding function. We look for a sequential equilibrium of the game in symmetric bidding strategies.¹⁵

So to start the analysis, let us fix equilibrium bidding and disclosure policies, and consider how the seller should set his secret reserve. By symmetry of the bidding strategies, ex-ante the seller knows that the bids will be N independent draws of $\beta(\xi, x_i)$, where the X_i are i.i.d F_X . Letting the distribution of $\beta(\xi, x_i)$ conditional on disclosed information ξ be $G|\xi$, standard results then show that the optimal secret

¹⁴The model can be extended to allow the buyer valuation to be $v(Q, V_S, X_i)$, as in the case where the seller knows something about the quality of the car that is not ex-post verifiable — and therefore not in Q — but impacts his outside option V_S and therefore V_S is of interest to the buyer. The bidding function will be more complex, to account for an additional Winner's curse effect, but nothing substantive about the equilibrium characterization will change.

¹⁵As Milgrom and Roberts (1986) note, the sequential equilibrium restriction rules out implausible Nash equilibria in which the bidders ignore disclosed information, and therefore sellers never disclose. It seems reasonable to focus on symmetric equilibria, given that bidders are not aware of the identities of other participants in the auction until they bid, potentially at the very end of the auction.

reserve price $r^*(\xi, v_S)$ is uniquely defined by the implicit equation $r = v_S + (1 - G(r|\xi))/g(r|\xi)$.¹⁶ As usual, the optimal reserve does not depend on the distribution of potential bidders F_N (since the secret reserve comes into play only when a single bidder has valuation above the reserve).

Next, consider the listing decision. Define the equilibrium surplus from listing given disclosure ξ , reserve r and outside option v_S , and a particular realization of the buyer's private signals, ordered from highest to lowest as $x^{(1)}, x^{(2)} \dots x^{(n)}$:

$$\tilde{\pi}_S(\xi, r, v_S) = \begin{cases} \beta(\xi, x^{(2)}) - v_S - cd & \beta(\xi, x^{(2)}) \geq r \\ r - v_S - cd & \beta(\xi, x^{(1)}) > r > \beta(\xi, x^{(2)}) \\ -cd & \text{otherwise} \end{cases}$$

The seller should list iff it the surplus is greater than the fixed listing cost F :

$$E[\tilde{\pi}_S(\xi, r, v_S)] \geq F \tag{1}$$

By inspection, the value of listing $\tilde{\pi}_S$ is decreasing in v_S for fixed disclosure ξ . It follows that the seller should list iff the outside option is less than some critical value $v_S^*(\xi)$ defined so that (1) holds with equality.

At this point, it is worth pausing to consider what happens under the extreme cases of costless disclosure — which leads to full disclosure, as in Grossman and Hart (1980), Grossman (1981) and Milgrom (1981) — and infinitely costly disclosure, which leads to adverse selection. In the left panel of Figure 3 we show the case with full disclosure. Types that list with disclosure are shown as triangles, those that list without disclosure as stars, and those that don't list at all are squares. Bidders are fully informed and bid more for better quality cars, and consequently sellers of good cars will typically list them even though they have a high outside option for the vehicle. This is shown in the figure by the number of high quality types that choose to list. In the right panel, we show the case with no disclosure ($c = \infty$). Here, bidders are completely uninformed and on average bid the same amount for all quality levels. Consequently, most sellers with good cars have better outside options and choose not

¹⁶See e.g. Krishna (2002). The solution to the implicit equation is unique because for $\beta(\xi, x_i)$ strictly increasing in x_i , $G|\xi$ inherits the increasing hazard rate property from F_X .

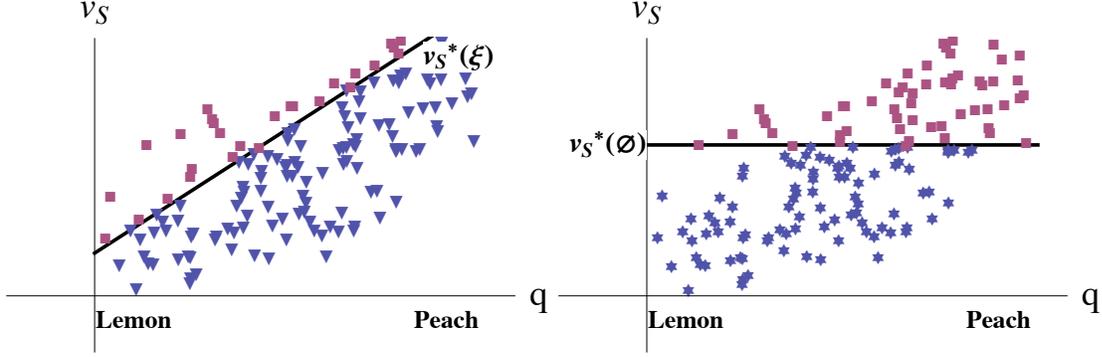


Figure 3: **Listings under extreme costs.** Cars listed with disclosure (triangles), listed without disclosure (stars) and not listed (squares) for a random set of (correlated bivariate normal) seller types. The left panel depicts the case with costless disclosure. Sellers fully disclose, and expect the auction price to be increasing in quality, so the cutoff $v_S^*(\xi)$ increases with q . The right panel depicts the case with infinitely costly disclosure. Expected auction prices are independent of quality, so the listing cutoff $v_S(\emptyset)$ is constant in q .

to list. This leads to adverse selection, as shown in the figure by the high fraction of lemons listing and low fraction of peaches. Next, we characterize the intermediate case of costly disclosure in a formal proposition:

Proposition 1 (Equilibrium with Costly Disclosure) *In all symmetric sequential equilibria:*

- (a) *Disclosure is characterized by a threshold $q^*(v_S)$ with $d = 1$ iff $q \geq q^*(v_S)$.*
- (b) *Bidders bid according to:*

$$\beta(\xi, x_i) = \begin{cases} v(q, x_i) & , \xi = q \\ E[v(Q, x_i) | Q < q^*(v_S), V_S \leq v_S^*(\emptyset)] & , \xi = \emptyset \text{ and } x_i \geq x^* \\ E_V[E[v(Q, x_i) | Q < q^*(v_S), v_S = V]] & , \xi = \emptyset \text{ and } x_i < x^* \end{cases}$$

where $V \sim H_{x_i}$ and both H_{x_i} and x^* are defined in the appendix.

- (c) *The seller sets optimal reserve $r^*(\xi, v_S)$ increasing in v_S , and lists iff $v_S \leq v_S^*(\xi)$.*
- (d) *If $v(Q, X_i)$ is additively separable in its arguments, $q^*(v_S)$ is increasing in v_S and $r^*(\xi, v_S)$ is increasing in ξ (under the ordering $\emptyset < q \forall q$).*

The intuition for the bidding is as follows. If the seller discloses, bidders have all the

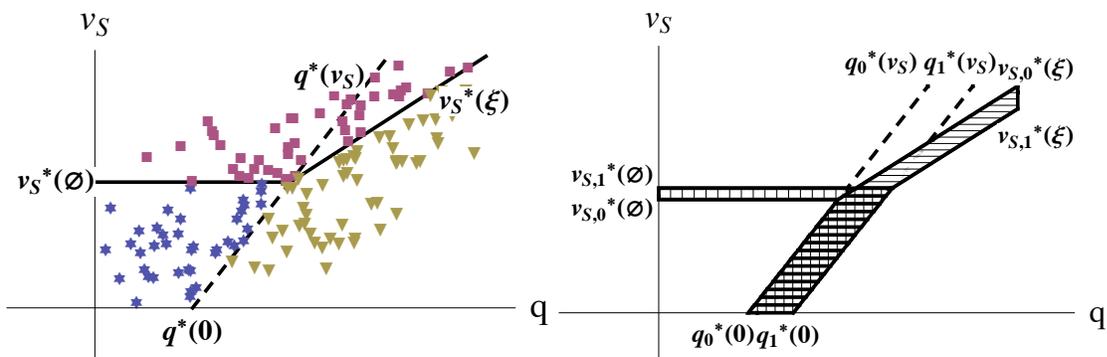


Figure 4: **Equilibrium and Comparative Statics.** The left panel shows cars listed with disclosure (triangles), listed without disclosure (stars) and not listed (squares) for a random set of (correlated bivariate normal) seller types. The right panel depicts the effect of an increase in disclosure costs. Types in the horizontally shaded region (peaches) no longer list; types in the crosshatched region list but no longer disclose; and type in the vertically shaded region (lemons) enter the market and list.

information they need to arrive at a private valuation of the vehicle, and since it is a second price auction, it is weakly dominant to bid their valuation. This is case 1 above ($\xi = q$). On the other hand, if the seller does not disclose, the buyers must form some expectation of the quality of the vehicle. There are three main effects. First, they must rationally expect that non-disclosure signals low quality, so they condition on $Q < q^*(v_S)$. Second, they should be concerned about adverse selection: any car that is listed must have a low outside option, so it may be a lemon. This leads them to condition on $v_S \leq v_S^*(\emptyset)$, the threshold outside option for a seller who does not disclose. These two effects describe case 2 in the bidding function (i.e. $\xi = \emptyset$ and $x_i \geq x^*$). But for low bids — those where the bid may not meet the secret reserve unless the outside option is low — there is also a Winner’s curse effect. On meeting the secret reserve and winning the car, the bidder learns that the seller’s outside option was poor, which is bad news since v_S and Q are positively correlated. Sophisticated bidders will account for this by conditioning on the expected value of v_S in the event that they win. This is case 3 ($\xi = \emptyset$ and $x_i < x^*$).¹⁷

The left panel of Figure 4 shows the equilibrium in the seller type-space, for the case with additively separable $v(q, x_i)$. The disclosure threshold $q^*(v_S)$ is shown as a dotted line, and types to the right of the line disclose. It is increasing in v_S since part of the gain to disclosure is that it increases the buyer’s willingness to pay and thus the

¹⁷A full derivation is provided in the appendix.

probability of sale, and as the outside option rises, this increased sales probability is of less value.¹⁸ The listing threshold $v_S^*(\xi)$ is constant in q for $\xi = \emptyset$ (non-disclosure), since bidders are not informed about quality and on average bid the same. It is then increasing in q for $\xi = q$ (disclosure), since expected payoff to listing rises and the indifferent type v_S^* is therefore also higher. In equilibrium, the type-space is divided into three regions. Types with outside option above $v_S^*(\xi)$ don't list. Of the remaining types, those with $q < q^*(v_S)$ list but don't disclose; and those with $q \geq q^*(v_S)$ list and disclose.

In the right panel of Figure 4 we show how the equilibrium shifts with an increase in disclosure costs. There are three effects. First, and most importantly, types who optimally disclose (high quality types) now find it less profitable to list because of the increased disclosure costs. This manifests itself as a shift down in $v_S^*(\xi)$. Second, types who were indifferent between disclosure and non-disclosure now strictly prefer not to disclose, so $q^*(v_S)$ shifts right. Finally, given that the pool of types who don't disclose has improved in quality — since $q^*(v_S)$ has shifted right — the payoff to non-disclosure also improves. Thus some low-quality types who were indifferent about listing now prefer to list, and $v_S^*(\emptyset)$ shifts up.

Examining the changes, it is easy to see how important disclosure costs can be for incentives to participate in the market. As disclosure costs increase, sellers of high-quality cars select out (the horizontally shaded region) and low quality-cars select in (the vertically shaded region). This is adverse selection in action! Worse still, if outside options and quality are tightly correlated, the type density over the vertically shaded region will be small relative to that over the horizontally shaded region, so that more types will select out than in, and the market will contract in size (possibly dramatically, as in Akerlof (1970)).

Increased disclosure costs may limit buyer incentives for participation too. Non-disclosure makes purchases more risky, in that there is some chance that the buyer's valuation is less than the price paid. Thus the probability of ex-post regret is strictly increasing in the disclosure cost. Another way in which disclosure costs may hurt buyer participation is through decreasing market size, as with fewer vehicles to be sold, buyers expect more competition for the vehicles and lower surplus, and thus given any

¹⁸The extreme case is where v_S is infinite, so that the reserve is also infinite and the car will never sell; in that case, it is never worthwhile to disclose.

participation costs may select out. Overall then, it is clear that disclosure costs are theoretically an important determinant of market participation, and therefore of the success of online auction platforms. We now turn to the data to test that hypothesis.

4 Empirical Analysis

The theory developed above makes quite specific predictions about the behavior of buyers and sellers of used goods in auctions markets. At a basic level, it argues that sellers with high quality goods should take the time to document the characteristics of these goods that make them high quality; whereas those with lower quality goods should not disclose their private information. Buyers should respond only to disclosures that can be verified ex-post and enforced as contractual claims, but otherwise should treat disclosures as truthful and condition their bidding behavior on the disclosed information. Moreover, buyers should interpret the absence of information as a bad signal about the quality of the good. Ex-ante information asymmetry is reduced through public information disclosure of ex-post verifiable information.

At a higher level, the theory claims that there is an important relationship between disclosure costs — whether opportunity costs for bidders or technological constraints imposed by the auction platform — and market performance. Specifically, in markets with high disclosure costs, bidders cannot distinguish low and high quality cars, and consequently bid lower amounts on high quality cars than they would in a perfect information environment. This results in lower sales rates, and potentially in high-quality sellers adversely selecting out of the market, as listing their car brings a low expected surplus.

In the empirical analysis that follows, I test the theory on each of these levels. First I test whether bidders perceptions of car quality are influenced systematically by information on the auction webpage. Second, I test whether there is a causal relationship between disclosure costs and the level of disclosure, and thus between disclosure costs, buyer expectations and auction outcomes.

The main challenge for the analysis is that the key variable — the information content of the auction webpage — is latent. Although I as the econometrician observe the same webpage as all the buyers, I am obliged to coarsen this content by encoding

into variables like the number of photos on the auction webpage. This is clearly less rich than the photos themselves. Because of this latent variable problem, it will be important to be explicit about differences between what the econometrician is conditioning on and what bidders condition on. This necessitates some care in specification and interpretation of the regressions.

4.1 The Data

The main data source is a collection of auction webpages from completed used car auctions on eBay Motors. This data was obtained by downloading the auction webpages for certain car models over an 8 month period, and then implementing a pattern matching algorithm to pull variables of interest from the webpage html code. I drop observations with nonstandard or missing data; new or certified pre-owned cars, and cars under salvage title.¹⁹ I also drop auctions in which the webpage was not created using either the basic eBay listing tools, or one of the most commonly used proprietary listing platforms, CARad, Auction123 or eBizAutos (11% of the remaining listings). The resulting dataset consists of 82538 observations of 18 models of vehicle. The models of vehicle are grouped into three main types: those which are high volume Japanese cars (e.g. Honda Accord, Toyota Corolla), a group of vintage and newer “muscle” cars (e.g. Corvette, Mustang), and most major models of pickup truck (e.g. Ford F-series, Dodge Ram).

Table 1 summarizes the variables in the dataset. For each auction, I observe a number of item characteristics including model, year, mileage and transmission and the number of options/accessories such as car radio etc listed by the seller. I also observe whether the vehicle sold is currently under manufacturer warranty. As a measure of reputation, I have the seller’s eBay feedback. All of this information is standardized and mandatory, in that the seller must provide it when listing the vehicle. My focus here is on the information *voluntarily* disclosed by the seller in the item description. I have two simple measures of this content: first, the number of photos posted on the auction webpage (my primary measure) and second, dummies for whether key text phrases (such as “dent” or “rust”) are used. As is clear from the summary statistics

¹⁹I drop cars under salvage title because they attract a completely different set of buyers, and are arguably in a different market.

in Table 1 and Table 4 (which has the text analysis), there is substantial variation in the information content of these webpages.

I supplement this main data source with data on private party book values publicly available at [edmunds.com](http://www.edmunds.com).²⁰ For model-years dated 1990 or later, I obtained the typical dealer retail value for each model-year of the models in my data set, and then matched this with each observation in the main data set, matching on trim where possible. This gives me book value data for nearly 55000 observations.

Last, I collected additional data on the buyers of middle-aged Honda Accords (model-years 1996-2000). eBay does not make information on its members easily available, and so I collected data on buyers and their state of residence manually.²¹ This allowed me to evaluate the distance between the buyer and the seller (measuring the buyer's location as the population centroid of their home state). Finally, for an even smaller subsample of 100 accords where the auction was successful in selling the car, I pulled the full Carfax reports.

4.2 Descriptive Evidence

The descriptive statistics are interesting in their own right. On average, the cars are quite old (nearly 16 years on average), well-travelled (about 90000 miles on the odometer) and there are 17 photos of them on the webpage. Sellers are typically experienced, with average feedback scores of 148. The minimum bid is usually set well below the book value of the vehicle, and thus most (85%) auctions receive at least 1 bid, with the highest bid averaging just over \$11 000. But only 28% of the cars actually sell, because of the widespread use of secret reserves.

Dealers and non-dealers differ quite markedly. Dealers list newer cars (3.5 years newer), with lower mileage (17000 miles less) and these cars are more than twice as likely to be under warranty. They also behave quite differently, using professional listing software for 47.5% of listings, versus 3.7% for non-dealers; and put up many more photos (21.4 versus 12.7). They use lower minimum bids, but higher secret reserves, so that average dealer sales rates are around 6% lower.

²⁰I used the "used car appraiser" at <http://www.edmunds.com/tmv/used/index.html>.

²¹I used Amazon's "Mechanical Turk" service to parcel the job out to a number of individuals.

Turning to the subsample of data on Accords, we see that on average the buyer and seller are nearly 500 miles apart. Figure 6 in the appendix plots the locations of sellers where the buyer is in the state of Illinois. Of the 22 observations, exactly half are out of state, some as far away as New York and California. Given that there are substantial transportation costs in transacting over this distance, the online transaction must yield relatively high surplus relative to more local alternatives. For the sample of 100 Accords, the Carfax data shows that a change of title occurred within a couple of months of the auction in 85% of the cases. Since one concern with looking only at the auction data is that it doesn't reflect actual ex-post outcomes, it is reassuring to see that most transactions do indeed go through.

4.3 Prices and Information

To begin the regression analysis, I derive the relationship between prices and observable covariates from the theory. Let bidders be indexed by i and let auctions be indexed by t . In each auction, the seller is obliged to provide certain standardized information, such as make, model, year, mileage etc. All of this information is observable, and we denote this by z_t . The seller has also publicly disclosed additional information through text and photos, but this information is not standardized and cannot be easily "coded up". Instead I observe a coarser measure of the amount of information, the number of photos posted, which I denote by I_t .

I specify a log-linear form for the bidder's valuation functions $v(q, x_i)$:²²

$$\log v_{it} = z_t\beta + q_t + x_{it} \tag{2}$$

where q_t is, as before, a quality index. It represents the common assessment of all bidders as to whether the car is better than average (given z_t), in which case $q_t > 0$, or is a lemon, in which case $q_t < 0$. Now, bidders do not observe q_t unless the seller fully discloses. Instead, they observe the disclosed information ξ_t and must make

²²The choice of a log-linear specification is for efficiency reasons, since the distribution of prices (the eventual dependent variable) is approximately log-normal conditional on the covariates.

inferences about q_t . This leads to the following specification of the bids:

$$\log b_{it} = z_t\beta + f(\xi_t) + \epsilon_{it} \quad (3)$$

where $f(\xi_t)$ is the equilibrium perception of a bidder with average private information of the quality of the vehicle taking into account adverse selection and winner’s curse effects, and ϵ_{it} is a mean zero shock that combines the idiosyncratic tastes of the buyer with her individual specific winner’s curse effects.²³

These bids should be interpreted as the *intended* bids of the bidders. In practice, not every bidder will be able to make their intended bid, and so their intended bids will not be observed. To see this, consider an auction where all bidders bid at a random time in the last 60 seconds of the auction. If, randomly, the first two bids happen to be the highest two bids of the set of bids, then after these bids have been entered, the standing price will jump to the second-highest bid. Every bid thereafter — i.e. 3rd highest bid, 4th highest etc, in random order — will be rejected by the proxy bidding system as being too low, as they are lower than the standing price. Only two bids will be recorded in this hypothetical auction. Henceforth, we refer to all bidders who intended to bid as the actual number of bidders n ; and to the number whose bids were actually recorded as the observed number of bidders n^* .

Fortunately for our purposes, the final price in the auction is always observed, and is equal to the intended bid of the bidder with the second highest private signal $x^{2:n}$.²⁴ The final price in the auction is given by:

$$\log p_t = z_t\beta + f(\xi_t) + \epsilon_t^{2:n} \quad (4)$$

Now, neither ξ_t nor $\epsilon_t^{2:n}$ are observed. We will use the number of photos I_t as a proxy for the disclosed information ξ_t ; and the observed number of bidders n_t^* as a proxy

²³Recall from Proposition 1 that the Winner’s Curse is more severe the less favorable the bidder’s private information.

²⁴This is not strictly correct: the final price is between $b^{2:n}$ and $b^{2:n} + \Delta$, where $\Delta > 0$ is the minimum bid increment (around \$50). We ignore this complication here, as doing so amounts to ignoring mis-measurement of the dependent variable, which does not affect the consistency of our estimates.

for the actual number of bidders n_t . Formally, we assume that:

$$\begin{aligned} E[f(\xi_t)|z_t, I_t, n_t^*] &= \gamma I_t + g_1(n_t^*) \\ E[\epsilon_t^{2:n}|z_t, I_t, n_t^*] &= g_2(n_t^*) \end{aligned}$$

so that average disclosure is linear in I_t ; and the observed number of bidders is potentially related to both the level of disclosure and the idiosyncratic taste of the second highest bidder. It is natural to expect that the number of photos is positively correlated with the level of disclosure ξ_t . Similarly, when more bidders are observed, it seems reasonable to expect that there were more people participating in the auction and seriously contemplating bidding (or bidding but not having their bids recorded). It is less clear that the level of disclosure should be correlated with the number of observed bidders — a car that looks great may attract bidders, but it might also repel bargain hunters who anticipate high bids — but in any case, this possibility is controlled for. This leads to an estimating equation:

$$\log p_t = z_t\beta + \gamma I_t + g(n_t^*) + u_t \tag{5}$$

where $g(n_t^*) = g_1(n_t^*) + g_2(n_t^*)$ and the mean-zero error term u_t is equal to $f(\xi) + \epsilon_t^{2:n} - \gamma I_t - g(n_t^*)$, independent of (z_t, I_t) under the assumptions above. Using dummy variables to flexibly model $g(n_t^*)$, OLS estimation will be consistent for β and γ .²⁵

This equation is the workhorse of the first part of the empirical analysis. The theory suggests that γ should be positive, since if sellers of high quality vehicles disclose more, we should on average see higher prices for vehicles listed with many photos.

I report the results of a wide variety of specifications in table 2. In the base specification (1), the vector of covariates includes car characteristics (mileage, number of options, model, year and transmission fixed effects), the number of photos and it squared, a fixed effect for the week of listing (to control for seasonal demand fluctuations), and a pair of seller characteristics (log feedback and percentage negative feedback). In this first specification I omit the number of bidders fixed effects, allowing for a possible effect of disclosure in increasing participation.

The coefficients generally have the expected sign and all are highly significant. The

²⁵Note that the constant term in β is not identified, since it is conflated with $g(n_t^*)$.

relationship between the number of photos and the price is estimated to be concave, which suggests decreasing returns to disclosure. Of particular interest is the sheer magnitude of the positive coefficients on the number of photos. A change from 9 to 10 photos is associated with a selling price that is approximately 1.75% higher, which for the average car in the dataset is around \$190 more. To put the magnitude of the coefficient in context, the value of a used car of a given model-year and mileage can vary by thousands of dollars depending on factors such as vehicle condition, maintenance history and documentation, all of which can be shown in photos. What this result suggests then is that bidders do rely heavily on photos to form perceptions of quality, and that the market is operating as expected. Sellers of high quality cars contract to provide high quality cars by carefully describing them on the webpage; those selling low quality cars provide weakly specified contracts through minimally descriptive webpages, and duly receive lower bids.

Notice also that the effects of the seller feedback ratings are quite small. The percentage negative feedback has a very small negative effect on price, while the coefficient on total log feedback is negative, which is the opposite of what one would expect. A possible reason for these results is that for the used car market, the volume of transactions for any particular seller is small and this makes seller feedback a weak measure of seller reputation, particularly as it conflates transactions in cars with other items.

The coefficients on photos get slightly smaller when we control for observed participation in specification (2), which suggests that increased disclosure leads to increased participation. In column (3) I interact photos with age and warranty status, expecting that photos have a greater impact on prices for older cars (due to greater heterogeneity) and a lower impact for cars under warranty (since the buyer is partially insured by the warranty). This is significantly the case. In column (4) I add the log book value as an additional control, allowing for a possibly non-linear relationship by model-year, as well as accounting for possible differences due to trim. The coefficients on photos remain significant and positive, though smaller, since this is for the subsample of newer cars (model-years after 1990).

In specifications (5) and (6), I consider two other dependent variables. Recall from the theory that sellers who disclose more favorable information should expect higher bids (in the sense of FOSD) and therefore should set higher secret reserves. Now I don't observe the secret reserves, nor do I observe whether a secret reserve was even

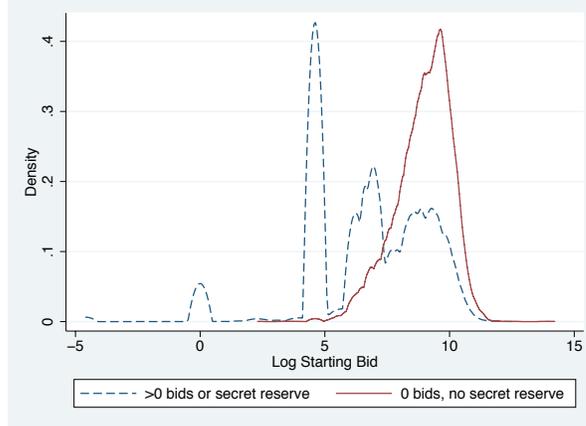


Figure 5: **Minimum Bids.** Density of Minimum Bids, for auctions in which no bids were made and no secret reserve was used, versus minimum bids where there were either a positive number of bids or a secret reserve. Notice that the former graph (solid) is quite reasonable; whereas the latter case (dashed) has huge spikes at 0 (\$1), 4.6 (\$100) and 6.9 (\$1000) dollars respectively.

set in the case where the item does in fact sell. This makes it difficult to analyze the secret reserves empirically. Instead, I look at the minimum bid set by the seller, which is effectively a standard reserve price. I want to focus on cases in which the seller used only a standard reserve, rather than both a standard and secret reserve, in which case the minimum bid may be meaningless (like \$1, which is very common). These auctions can be identified as auctions in which no bids were received and the phrase "the reserve is not yet met" does not appear.²⁶ Figure 5 shows the densities of minimum bids for this subsample of auctions, and the rest.

Regressing the minimum bid on covariates for this subsample gives the results in (5). The coefficient on the number of photos is positive and similar to that in (1). Ignoring the issue of selection bias resulting from picking auctions with zero bidders (i.e. unrealistically high minimum bids), we see that sellers set higher reserves when they have more favorable information to disclose, consistent with the theory.

Finally, in (6) I regress the binary variable indicating a successful sale on the covariates. I exclude the number of observed bidders as a regressor, as this is not consistent with the linear specification.²⁷ Here we see that cars with a higher number of photos — presumably better quality cars — are more likely to sell. The theory makes no

²⁶Where "the reserve is not yet met", there is a definitely a secret reserve; where the number of bids is positive, there may have been a secret reserve that was met

²⁷If no bidders are observed, the car certainly does not sell.

prediction here, since better cars also have better outside options, and therefore attract higher reserves. Yet taken together with the price results, we see that demand is higher for cars with a higher number of photos, consistent with buyers learning from the photos that these are cars of higher quality.

Seller Heterogeneity: One might be concerned that the results are driven by seller heterogeneity. Frequent sellers such as car dealerships may have lower disclosure costs and put up more photos. Then if buyers prefer to buy from professional car dealers, I may be picking up this preference rather than the effects of information disclosure. To examine this, I split my dataset into those sellers who list multiple different cars during the 9-month period, and call those dealers, and classify the remainder as private sellers. The first two columns in table 3 give the results of separate hedonic regressions for those two groups. The coefficients on photos remain positive, significant and large for both groups. In the final column, I consider the sample of dealers and include a seller-specific fixed effect for each of them (dropping the seller characteristic measures). The results show that even after controlling for seller identity, there is a large and significant relationship between price and the number of bidders and photos. This suggests both that dealers vary the amount of photos for each individual listing (i.e. the information posted is vehicle specific), and that such information variation positively co-varies with prices. This is consistent with a selective disclosure policy by dealers.

Text Analysis: Another measure of webpage content is the text of the car description. So I look for key phrases associated with car value in the item description, and see if price responds to the presence of these phrases on the auction webpage. In table 4, I present the proportions of webpages with affirmative (e.g. “has dent”) and negative (e.g. “does not have dent”) versions of these phrases. I then create variables for each phrase by coding affirmative statements as 1 and negative statements as -1 . The third column presents the estimated coefficient on these variables in a hedonic regression of the form given in equation (4). The proportion of affirmative phrases is probably biased upwards and the negatives downward as I pick up a phrase like “rust free” as an affirmative because it contains “rust” without a qualifier like “no” or “not” in front of it. Nonetheless, there is evidence that the text-based descriptions do contain valuable information about the car to be sold. In the regression, the coefficients on the phrases have the expected sign, with cars with “broken” features

selling for considerably less (13.7% less), while those that are “garage kept” and that come with documentation and receipts sell on average for considerably more (13.4% more in the case with “receipts”). Interestingly, the coefficient on photos remains of similar magnitude to before. This may be because even with these text phrases controlled for, photos still play an important role in defining exactly what a car in “mint” condition is for contracting purposes.

4.4 Disclosure Costs, Disclosure and Equilibrium Outcomes

In the section above, I have shown that the amount of information is positively correlated with buyer’s perceptions of the quality of the car; and that these perceptions seem to be informed by ex-post verifiable information. In this section, I look directly at the relationship between disclosure costs, disclosure, prices and sales rates. To get at this, I need a disclosure cost-shifter. A natural candidate is the listing software ($SOFT_t$) used by the seller to create the webpage. In the data I have sellers who use the standard eBay software, and those who use the professional listing platforms provided by CARad, Auction123 and eBizAutos. These technologies promise users that they will simplify and streamline the process of creating a listing, through simple user interfaces, templates and free photo hosting and management services. It seems reasonable then that they should lower disclosure costs.

The downside with using this as a cost-shifter is that it is potentially endogenous. As we know from the summary statistics reported in Table 1, dealers are overwhelmingly more likely to use the professional platforms. There are a couple of reasons for this. First, there is a large initial fixed cost associated with setting up the templates properly (e.g. most dealerships include an “about us” part of the template, which private sellers would not bother with). Second, the platforms have a menu of prices, where one-off listings are relatively expensive (\$10 for CARad, \$15 for Auction123, not available for eBizAutos), but unlimited monthly listing plans may be cost effective for high volume sellers (they range from \$200-\$300 a month).

Fortunately, for dealers I have a panel of observations. So rather than identifying off (potentially endogenous) differences in software choice across sellers, I instead look at switches in software choice for a given dealer. Provided the decision to switch software is uncorrelated with the unobservables, the effect of software switches will

be consistently identified. It is essentially differences-in-differences, comparing those who switch to those who don't.

Formally, the number of photos posted should depend on both car characteristics (including quality), and seller characteristics (including software and possibly experience). Letting sellers be indexed by j , we have:

$$I_{jt} = z_{jt}\beta_I + \alpha_I SOFT_{jt} + \bar{q}_j + \Delta q_{jt} - \gamma_j - \tilde{\epsilon}_{jt} \quad (6)$$

where the error term $\bar{q}_j + \Delta q_{jt} - \gamma_j - \tilde{\epsilon}_{jt}$ comprises the average quality of cars sold by j , \bar{q}_j ; the deviation for this particular car's quality Δq_{jt} ; the average cost of disclosure for j , γ_j ; and an idiosyncratic disclosure cost draw $\tilde{\epsilon}_{jt}$. Including a seller fixed effect soaks up $\bar{q}_j - \gamma_j$, and then the required orthogonality condition is that $SOFT_{jt}$ is uncorrelated with $\Delta q_{jt} + \tilde{\epsilon}_{jt}$. This amounts to assuming that sellers don't switch software precisely when they have a particularly high quality car to sell, or when they are facing particularly low disclosure costs.

The results of regressing photos on characteristics and software are reported in the first column of table 5. They indicate that dealers who switch to professional listing software subsequently put up significantly more photos than those that don't, around 10 more. Those selling cars with more options, or lower mileage, also tend to put up more photos. Overall, there is a causal relationship between disclosure costs and car quality, and the number of photos.

Now, I would like to use software as an instrument for the number of photos. I have just shown that it is relevant. For it to be exogenous, I need to assume both that it is uncorrelated with seller-specific deviations in quality — as I did before — and that it has no direct effect on equilibrium prices. The concern here is an “advertising effect”, whereby better software produces slicker webpages and buyers are persuaded to bid more. Given that this is a major purchase, if any such effects exist, they should be small, and I assume that software has no direct effect on buyer valuations.²⁸ Under these assumptions, software is a valid instrument for photos.

So in column (2) we rerun the hedonic regressions from earlier, instrumenting for photos with software. The first-stage F is a respectable 34.2, and the estimated

²⁸Notice that the reputation of the seller, if it matters, is controlled for with the fixed effect; a direct effect would imply a change in the perceived reputation of the seller with a change in software.

coefficient on photos is significant and positive. Compared to the OLS results from earlier, the coefficient is smaller, as one would expect. This result suggests that disclosure costs have a causal effect on equilibrium prices, through affecting the level of disclosure.

Next, we look at the impact on sales rates. Column (3) reports the IV results. Photos has a statistically zero coefficient, implying that disclosure costs do not have an equilibrium effect on sales rates. The theory predicts that in the case of additively separable valuations, lower disclosure costs will lead to higher sales rates because bids will increase relatively more than reserves. So this result could be rationalized as either a failure of the additively separable valuations assumption; or as an indication that I don't have enough switches in the data to pick up a significantly positive effect. Overall though, the results back up the theory, in showing that disclosure costs affect disclosure, and through disclosure, prices.

Robustness: The key assumption in the above analysis was that software choice was unrelated to seller-specific quality deviations Δq_{jt} . A natural concern is that the results might have been driven by dealers who actively managed their software choice, paying the \$10-\$15 to list using CARad or Auction123 when they had a high-quality car to sell, and using the native eBay software when the car was of lower quality. I conduct two kinds of robustness checks. First, I rerun the regressions of columns (1)—(3) on a subsample of dealers who switch software at most once, and list at least 5 cars. This bounds the potential for active management, while at the same time unfortunately limiting the identifying power of the approach (fewer switches). I find that in the first stage, all the coefficients are smaller, but still significant — in other words, software switches do cause more disclosure, although there is some evidence of active switching. In the IV regressions, the coefficient on photos continues to be significant and positive for prices, and now significant and positive for sales as well. This set of regressions is re-assuring.

Second, I try to see if changes in observable car characteristics are correlated with switches in software, since that might indicate that unobservable quality differences are as well. So I code up a variable "upgrade", coded as a 1 when dealers switch from basic to any of the professional software; coded as -1 when dealers downgrade to the basic software; and 0 otherwise. Then for every listing, I compute the difference between the average mileage of cars sold after that date (including that listing) and

before that date (excluding that listing). I repeat this construction for other characteristics. I then regress upgrade on changes in characteristics. The idea is that if sellers adopt the software when they anticipate getting younger, lower mileage or generally better cars, this regression should pick that up. The results are reported in Table 6. Basically, all the coefficients are statistically insignificant from zero, for both the full sample, and one in which book values are recorded. An F-test fails to reject the null of zero coefficients. This again helps to allay concerns that changes in software are driven by upward trends in car quality, which would bias our results.

5 Conclusion

Given the increasing growth of online transactions in used goods markets, it is important to understand what makes these markets work. This paper shows that information asymmetries in these markets can be endogenously resolved, so that adverse selection need not occur. The required institutional feature is a means for credible disclosure. With this in place, sellers have both the opportunity and the incentives to remedy information asymmetries between themselves and potential buyers.

Disclosure costs are important in determining how effective this remedy is. Where the bandwidth and technology is available to tightly define the contract between buyer and seller through rich media such as photos and videos, adverse selection problems can be mitigated. Looking at one determinant of disclosure costs — the software used to create the listings — we see that better software does lead to more disclosure, higher prices for the seller, and therefore greater incentives for participation in this market.

One avenue for future research would be to analyze how disclosure interacts with dynamic incentives such as reputation in markets with repeated interaction, such as those characterizing many business to business transactions.

References

- Adams, Christopher P., Laura Hosken, and Peter Newberry**, “Vettes and Lemons on EBay,” 2006. Unpublished.
- Akerlof, George A.**, “The Market for Lemons: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics*, 1970, *84* (3), 488–500.
- Bajari, Patrick and Ali Hortaçsu**, “The winner’s curse, reserve prices and endogenous entry: empirical insights from eBay auctions,” *Rand Journal of Economics*, Summer 2003, *34* (2), 329–355.
- Bond, Eric W.**, “A Direct Test of the “Lemons” Model: The Market for Used Pickup Trucks,” *The American Economic Review*, September 1982, *72* (4), 836–840.
- Genesove, David**, “Adverse Selection in the Wholesale Used Car Market,” *The Journal of Political Economy*, August 1993, *101* (4), 644–665.
- Grossman, Sanford J.**, “The Informational Role of Warranties and Private Disclosures,” *Journal of Law and Economics*, 1981, *24* (3), 461–483.
- and **Oliver D. Hart**, “Disclosure Laws and Takeover Bids,” *Journal of Finance*, 1980, *35* (2).
- Houser, Daniel and John Wooders**, “Reputation in Auctions: Theory and Evidence from eBay,” *Journal of Economics and Management Strategy*, 2006, *15* (2), 353–370.
- Jin, Ginger Zhe**, “Competition and Disclosure Incentives: An Empirical Study of HMOs,” *Rand Journal of Economics*, 2005, *36* (1), 93–112.
- and **Andrew Kato**, “Price, Quality and Reputation: Evidence from an Online Field Experiment,” *Rand Journal of Economics*, 2006, *37* (4).
- and **Phillip Leslie**, “The Effect of Information on Product Quality: Evidence From Restaurant Hygiene Grade Cards,” *Quarterly Journal of Economics*, 2003, *118* (2), 409–451.
- Jovanovic, Boyan**, “Truthful Disclosure of Information,” *Bell Journal of Economics*, 1982, *13* (1), 36–44.

- Krishna, Vijay**, *Auction Theory*, Academic Press, 2002.
- Mathios, Alan D.**, “The Impact of mandatory disclosure laws on product choice: An analysis of the salad dressing market,” *Journal of Law and Economics*, 2000, 43 (2), 651–677.
- Milgrom, Paul**, “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics*, 1981, 12 (2), 380–391.
- and **John Roberts**, “Relying on the information of interested parties,” *Rand Journal of Economics*, 1986, 17 (1).
- Resnick, Paul and Richard Zeckhauser**, *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay’s Reputation System*, Vol. 11 of *Advances in Applied Microeconomics*, Elsevier Science, 2002.
- Yin, Pai-Ling**, “Information Dispersion and Auction Prices,” 2006. Harvard Business School Working Paper, No. 06-011.

Appendix

Proof of Proposition 1:

Fix any sequential equilibrium in symmetric bidding strategies. First note that by the monotonicity of $\tilde{\pi}_S$ in v_S , in equilibrium the seller optimally lists iff $v_S \leq v_S^*(\xi)$. We proceed to prove the remainder of the proposition in three steps:

Step 1: Bidding Strategies. If the seller discloses, standard Vickrey auction results establish that bidders bid $v(q, x_i)$. So consider the case with non-disclosure. Fix an arbitrary disclosure policy $\xi(q, v_S)$ and let $\tilde{v}(x_i, v_S) = E_Q[v(Q, x_i) | \xi(Q, v_S) = \emptyset]$. Let $Y = \max_{j \neq i} X_j$ be the highest opposing signal, having distribution $\tilde{F}(y) = E_N[F_X(y)^{n-1} | N \geq 1]$. From above, the listing policy with non-disclosure is characterized by a cutoff $v_S^*(\emptyset)$. Let the distribution of v_S conditional on $v_S \leq v_S^*(\emptyset)$ be H . Then suppressing dependence on ξ , a bidder of type x_i solves:

$$\max_b \int_{-\infty}^{r^{-1}(b)} \int_{-\infty}^{\beta^{-1}(r(v_S))} (\tilde{v}(x_i, v_S) - r(v_S)) d\tilde{F}(y) dH(v_S) + \int_{-\infty}^{r^{-1}(b)} \int_{\beta^{-1}(r(v_S))}^{\beta^{-1}(b)} (\tilde{v}(x_i, v_S) - \beta(y)) d\tilde{F}(y) dH(v_S)$$

where we assume for the moment that $r^{-1}(b) < v_S^*$. Then taking the first order condition and simplifying we get:

$$(\tilde{v}(x_i, \hat{v}) - b) \frac{\beta'(x_i) \tilde{F}(x_i)}{r'(\hat{v})} + \int_{-\infty}^{\hat{v}} (\tilde{v}(x_i, v_S) - b) dH(v_S) = 0$$

where \hat{v} is equal to $r^{-1}(b)$. Define $H_{x_i}(v_S) = H(v_S)$ for $v_S < \hat{v}$ and $H_{x_i}(v_S) = H(v_S) + \frac{\beta'(x_i) \tilde{F}(x_i)}{r'(\hat{v})}$ otherwise. Then re-arranging the first order condition into a single integral, taking b outside of the integral and dividing both sides by the constant $\int_{-\infty}^{\hat{v}} dH_{x_i}(v_S)$, we get:

$$b^* = E_{V_S} [\tilde{v}(x_i, V_S)] = E_{V_S} [E_Q[v(Q, x_i) | \xi(Q, V_S) = \emptyset, V_S]]$$

Now provided $r^{-1}(b^*) < v_S^*$, this is the optimal bid. That is, $\beta(\xi, x_i) = b^*$ for all $b^* < r(\xi, v_S^*)$, or equivalently for $x_i < x^* = \beta^{-1}(\xi, r(\xi, v_S^*))$. On the other hand, if $x_i \geq x^*$, meeting the reserve is uninformative (since for all seller types who list, the reserve is below the bid), and in that case it is weakly dominant for the bid-

der to bid their valuation conditional on the seller's disclosure and listing policies, $E[v(Q, x_i)|\xi(Q, V_S) = \emptyset, V_S \leq v_S^*(\emptyset)]$.

Step 2: Disclosure. We now show $\xi(q, v_S)$ must take the threshold form, disclose iff $q \geq q^*(v_S)$. Towards a contradiction, suppose not, so that for some fixed v_S there exist $q < q'$ with $\xi(q, v_S) = q$, $r(q, v_S) = r$, $\xi(q', v_S) = \emptyset$ and $r(q', v_S) = r'$. Since deviation is not profitable for type (q, v_S) , $E[\tilde{\pi}(q, r, v_S)] \geq E[\tilde{\pi}(\emptyset, r', v_S)]$. But then from step 1 above, the bid distribution increases in q (in the sense of FOSD) conditional on disclosure, so $E[\tilde{\pi}(q', r, v_S)] > E[\tilde{\pi}(q, r, v_S)]$. This implies a profitable deviation for type (q', v_S) by $E[\tilde{\pi}(q', r, v_S)] > E[\tilde{\pi}(\emptyset, r', v_S)]$, yielding a contradiction.

Step 3: Slopes of Equilibrium Strategies. The optimal reserve satisfies $r = v_S + (1 - G(r|\xi))/g(r|\xi)$. Let the hazard ratio of the bid distribution be $\lambda_G(y|\xi) = g(y|\xi)/(1 - G(y|\xi))$. Then since the bidding strategies are increasing in x_i for any ξ , $\beta(\xi, x_i)$ is invertible, and $\lambda_G(y) = \lambda_{F_X}(\beta^{-1}(\xi, y))$. The RHS is increasing in y since F_X has the increasing hazard rate property, which implies the LHS is increasing in y also. Then, applying the implicit function theorem to the optimal reserve, you get that $\partial r/\partial v_S = \lambda_G(r|\xi)^2 / (\lambda_G(r|\xi)^2 + \lambda'_G(r|\xi)) > 0$.

Next, given $v(Q, X_i)$ additively separable in Q and X_i , the bidding function is piecewise additively separable in ξ and x_i , since x_i can be pulled outside the expectation in Q i.e. $\beta(\xi, x_i) = \beta_1(\xi) + \beta_2(x_i)$ for $x_i > x^*$ and $\beta(\xi, x_i) = \beta_3(\xi) + \beta_2(x_i)$ for $x_i < x^*$. Then for $b \geq \beta_1(\xi) + \beta_2(x^*)$, $G(b|\xi)$ is equal to $G(\beta_2^{-1}(b - \beta_1(\xi)))$. Taking derivatives, we get that $\partial \lambda_G/\partial \xi > 0$, and then by the IFT $\partial r/\partial \xi > 0$. The case on the other region is similar.

Finally, to show $q^*(v_S)$ increasing in v_S , define the marginal benefit to disclosure $\Delta(q, v_S) = E[\tilde{\pi}_S(q, r^*(q, v_S), v_S)] - E[\tilde{\pi}_S(\emptyset, r^*(\emptyset, v_S), v_S)]$. In equilibrium, $q^*(v_S)$ must solve $\Delta(q, v_S) = 0$. Now, $\partial \Delta/\partial q > 0$, so by the IFT we will obtain our result if $\partial \Delta/\partial v_S < 0$. Considering $\partial \Delta/\partial v_S$, interchange integral and derivative, and then notice that terms of the form $\partial \tilde{\pi}_S/\partial r^*$ are zero since the FOC must hold at r^* . Then it is easy to simplify the expression and get $\partial \Delta/\partial v_S = \mathbb{P}(\beta(\emptyset, x^{(1)}) \geq r^*(\emptyset, v_S)) - \mathbb{P}(\beta(q, x^{(1)}) \geq r^*(q, v_S))$. To show this negative, it suffices that $\beta(\emptyset, x^{(1)}) \geq r^*(\emptyset, v_S) \Rightarrow \beta(q, x^{(1)}) > r^*(q, v_S)$, which by piecewise additivity of the bids reduces to showing $r^*(q, v_S) < r^*(\emptyset, v_S) + \beta_1(q) - \beta_1(\emptyset)$ (for $x^{(1)} > x^*$, other case similar). This holds since the FOC for the reserves implies $r^*(q, v_S) - r^*(\emptyset, v_S) < \beta_1(q) - \beta_1(\emptyset)$.

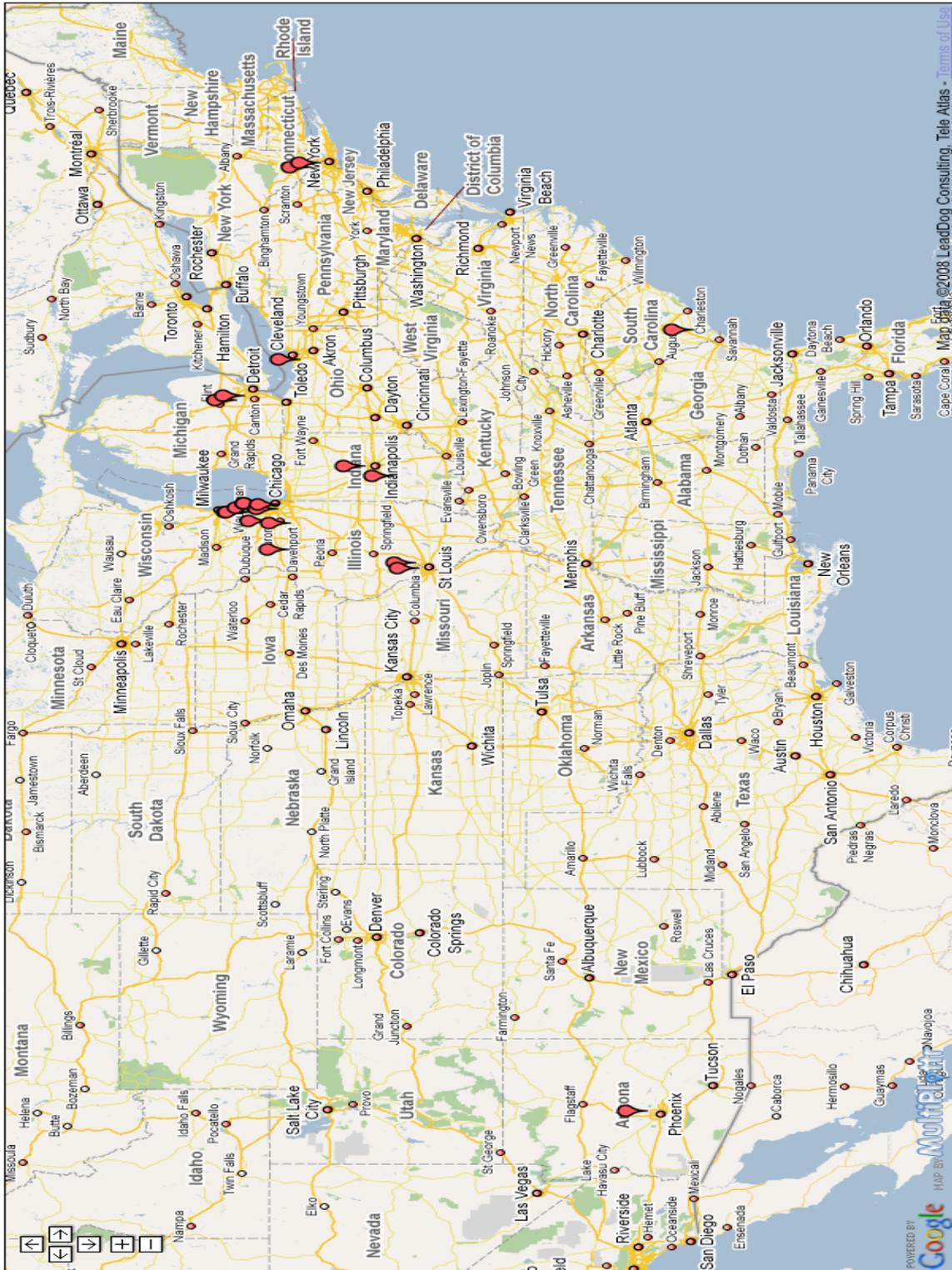


Figure 6: **Location of Sellers for Illinois Buyers** The map shows the locations of cars bought by buyers residing in Illinois, for a subsample of 1995-2000 Honda Accords. Of the 22 observations, 50% are out-of-state purchases.

Table 1: Summary Statistics

	Full Sample		Non-Dealers	Dealers
	Mean	Std. Dev	Mean	Mean
Car Characteristics				
Miles	90181	90663	98320	81217
Age (in years)	15.8	13.6	17.5	14.0
% Manual Transmission	30.4	—	33.6	26.8
% Warranty	18.6	—	12.2	25.5
# of Options	5.2	5.2	5.4	5.0
# of Photos	17.0	10.8	12.7	21.4
% “Reliable” Cars	18.7	—	16.5	21.1
% Pickups	30.3	—	28.9	31.9
% CARad	14.7	—	2.9	27.7
% Auction123	3.8	—	0.5	7.4
% eBizAutos	6.1	—	0.3	12.4
Seller Characteristics				
Seller Feedback Score	148.0	556.7	115.0	184.5
% negative feedback	1.60	6.00	1.42	1.78
Auction Characteristics and Outcomes				
Minimum Bid* (% of book value)	52.4	78.5	62.6	43.0
% auctions with ≥ 1 bid	85.2	—	82.5	87.9
% sold	28.4	—	31.4	25.2
Highest Bid	11110	13018	9173	13113
Distance between buyer and seller*	448	562	437	474
% change of title observed*	85	—	—	—

This table provides summary statistics for the covariates used in the analysis. Standard deviations for categorical variables are not reported. Asterixes denote that the statistic is from a subsample, as described in the text.

Table 2: Hedonic Regressions

	Log Price				Log MinBid	Sell
	(1)	(2)	(3)	(4)	(5)	(6)
Log Miles	-0.130263 (0.004520)	-0.130999 (0.004521)	-0.127750 (0.004537)	-0.184320 (0.007045)	-0.140871 (0.011358)	0.022123 (0.001537)
Number of Photos	0.019752 (0.001114)	0.018934 (0.001099)	0.000942 (0.000693)	0.008555 (0.000994)	0.017181 (0.002702)	0.002462 (0.000779)
Photos Squared	-0.000229 (0.000020)	-0.000227 (0.000020)		-0.000096 (0.000019)	-0.000202 (0.000055)	-0.000056 (0.000015)
Number of Options	0.015373 (0.000755)	0.014947 (0.000743)	0.014152 (0.000745)	0.007747 (0.000622)	0.013135 (0.002057)	0.000898 (0.000579)
Log Feedback	-0.009155 (0.002284)	-0.011351 (0.002220)	-0.011775 (0.002295)	-0.014267 (0.002036)	-0.004590 (0.005843)	0.012460 (0.002098)
% Negative Feedback	-0.003701 (0.000994)	-0.003696 (0.000956)	-0.003696 (0.000926)	-0.003271 (0.000941)	-0.000217 (0.002680)	0.001330 (0.000617)
AgeXPhoto			0.000602 (0.000060)			
Warranty			0.076367 (0.017731)			
WarrantyXPhoto			-0.000868 (0.000781)			
Log Book Value				0.579769 (0.015615)		
Model/Year/Week FE	yes	yes	yes	yes	yes	yes
Number of Bidders FE	no	yes	yes	yes	yes	no
R-squared	0.6945	0.6998	0.7036	0.7863	0.6968	0.1156
N	71292	71292	71292	47148	6303	82538

Standard errors (clustered by seller) are given in parentheses. In all specifications, fixed effects for the car model, car year and week of sale are included, though their coefficients are not reported. In (1)-(4), the dependent variable is log price, and the model fits well, with the R^2 ranging from 0.68 in (1) to 0.78 in (3). Specifications (1)-(3) are for the full sample, with specification (2) adding number of bidders fixed effects, and (3) adding interaction terms. Specification (4) is for the sub-sample where I have book value data. Specification (5) is on the subsample with minimum bids, no secret reserve and no recorded bids. The dependent variable is log minimum bid. Specification (6) is on the full sample, with the sale outcome as the dependent variable.

Table 3: Hedonic Regressions - Dealer Status

	Log Price		
	Non-Dealers	Dealers	Dealer Fixed Effects
Log Miles	-0.132320 (0.005472)	-0.124631 (0.007114)	-0.087396 (0.006305)
Number of Photos	0.026677 (0.001313)	0.012514 (0.001670)	0.017170 (0.001701)
Photos Squared	-0.000351 (0.000034)	-0.000133 (0.000027)	-0.000112 (0.000026)
Number of Options	0.019038 (0.000832)	0.010837 (0.001082)	0.011804 (0.000944)
Log Feedback	-0.010653 (0.002251)	-0.013246 (0.003573)	
% Negative Feedback	-0.002816 (0.001123)	-0.004647 (0.001423)	
Model/Year/Week FE	yes	yes	yes
Number of Bidders FE	yes	yes	yes
Seller Fixed Effects	no	no	yes
R-squared	0.6812	0.7056	0.5691
N	33232	38060	40316

The results are from regressions of log price on car and seller covariates, separately for dealers and non-dealers. In the last specification, seller fixed effects are included and the feedback measures are dropped. Standard errors (clustered by seller) are given in parentheses.

Table 4: Key Phrases

	% Affirmative (e.g. has dent)	% Negative (e.g. is not dented)	Estimated Coefficient	Standard Error
"Repainted"	2.76	0.01	0.1358	0.0174
"Missing"	4.13	0.05	-0.174	0.0200
"Accident"	2.18	0.36	-0.103	0.0229
"Dent"	9.37	1.80	-0.0563	0.0088
"Worn"	2.12	0.10	-0.0563	0.0218
"Broken"	2.71	0.11	-0.137	0.0205
"Crack"	7.81	1.69	-0.0736	0.0093
"Dirty"	0.57	0.01	-0.1538	0.0373
"Faded"	1.29	0.09	-0.1656	0.0515
"Rust"	13.05	4.69	-0.1272	0.0075
"Receipt"	3.05	0.01	0.1340	0.0180
"Documentation"	4.16	0.04	0.0427	0.0186
"Service"	24.82	0.02	-0.0502	0.0104
"Aftermarket"	4.02	0.10	0.0556	0.0137
"Paperwork"	3.42	0.03	0.0490	0.0182
"Detail"	14.55	0.05	0.1038	0.0109
"Inspection"	15.47	0.08	-0.0028	0.0110
"Record"	3.75	0.21	0.0601	0.0131
"Invoice"	0.45	0	0.3361	0.0540
"Garage Kept"	3.57	0.01	0.1672	0.0143
"Original Part"	0.64	0	0.0867	0.0535
"Receipt"	0.54	0.01	0.1516	0.0392
"Mint"	3.50	0.03	0.0907	0.0177
Number of Photos	—	—	0.0184	0.0011
Photos Squared	—	—	-0.00022	0.00002

The above table shows the frequencies of affirmative and negative versions of key phrases used in the text descriptions of cars being auctioned (columns 1 and 2), and the estimated coefficients and associated standard errors in a hedonic regression of log price on dummies for the phrases and all the covariates used in specification (2) of table 2 above.

Table 5: Cost and Equilibrium Outcomes

	All Dealers			≥ 5 Cars Listed, ≤ 1 Switch		
Log Miles	-0.171291 (0.049181)	-0.087796 (0.002628)	0.007496 (0.001824)	-0.201655 (0.074163)	-0.090804 (0.003613)	0.007864 (0.002399)
Number of Options	0.073878 (0.011205)	0.012010 (0.000783)	-0.002428 (0.000554)	0.097482 (0.015963)	0.010060 (0.001106)	-0.002414 (0.000773)
Number of Photos		0.008137 (0.002969)	0.000048 (0.002046)		0.017772 (0.006818)	0.009892 (0.004966)
CARad	9.126304 (1.117055)			7.760723 (2.096538)		
Auction123	13.453836 (1.442224)			10.975200 (2.257123)		
eBizAutos	10.373457 (1.560884)			8.523578 (2.505598)		
Model/Year/Week FE	yes	yes	yes	yes	yes	yes
Number of Bidders FE	no	yes	no	no	yes	no
Seller FE	yes	yes	yes	yes	yes	yes
First Stage F	—	34.23	33.38	—	7.15	8.01
N	24028	21297	24028			

Regressions are for the subsample of dealers (columns 1-3) and for those dealers who list at least 5 different cars during the sample period, and switch listing software at most once (columns 4-6). In each set of three, the dependent variable in the first column is photos; in the second and third column it is the log price and a dummy for sale respectively. In the columns marked IV, photos is instrumented for using the software dummies, and the first-stage F is reported. Robust standard errors are given in parentheses.

Table 6: Software Upgrades and Observables

	Dealers Only	Book Value Subsample
	Upgrade	Upgrade
Δ Avg. Age	-0.000050 (0.000084)	-0.000111 (0.000145)
Δ Avg. Miles	-0.000005 (0.000008)	-0.000006 (0.000014)
Δ Avg. Options	0.000099 (0.000168)	0.000085 (0.000190)
Δ Listings per Day	0.000216 (0.000306)	0.000106 (0.000434)
Δ Avg. Book Value		-0.000035 (0.000143)
R-squared	0.0000	0.0000
N	40224	28728
P-value (H_0 : all coefficients are zero)	0.6929	0.4751

The dependent variable is upgrade, equal to 1 when a seller upgrades from basic eBay software to any professional software; -1 in the event of a downgrade; and zero otherwise. The regressors are difference in average car characteristics (and listing volume) before and after the listing. The results indicate that a change in characteristics (e.g. book value) is not significantly correlated with software upgrades and downgrades. Standard errors are clustered by seller.