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

Gregory Lewis†
Department of Economics
Harvard University

September 9, 2007

Abstract

Since the pioneering work of Akerlof (1970), economists have been aware of the adverse selection problem that information asymmetries can create in used goods markets. The remarkable growth in *online* auctions of used goods therefore poses a puzzle. I argue that given a means for credible information disclosure, sellers will voluntarily disclose their private information to buyers through online media. This limits information asymmetries and adverse selection. To test this theory, I examine the role of information in a large online used car market, eBay Motors. I find that sellers selectively disclose information; that information asymmetries are reduced by these disclosures; and that online media such as photos, text and graphics provide a rich environment for information disclosure.

*This is a 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 Jim Levinsohn for his comments, as well as those of Dan Ackerberg, Jim Adams, Susan Athey, Tilman Börgers, Han Hong, Hugo Hopenhayn, Guido Imbens, Kai-Uwe Kühn, Serena Ng, Scott Page, Ariel Pakes, Nicola Persico, Dan Silverman, Jagadeesh Sivadasan, Doug Smith, Lones Smith and Charles Taragin. All remaining errors are my own.

†Email: glewis@fas.harvard.edu

1 Introduction

The rise of the internet has also seen a huge rise in the volume of used goods traded online. Online auction sites such as eBay and Yahoo! Auctions compete with specialized listing sites such as usedcomputer.com and cars.com in the retail trade of consumer goods. Meanwhile, business to business transactions are conducted 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 voluntarily disclosing their private information to potential buyers. In this paper, I undertake an empirical analysis of this "unravelling" explanation, taking as a case study eBay Motors, the largest used car market in the United States. In this $12 billion market, the stakes are high for both sides, the information asymmetries are substantial, and yet there is a high volume of trade with approximately 20000 cars sold each month.[2] I find here that although sellers disclose information, they do so selectively; that information asymmetries do indeed "unravel"; and that online media such as photos, text and graphics provide a rich environment for information disclosure. I conclude that in online markets where sellers can make credible disclosures through online media, information asymmetries and thus adverse selection will be limited.

There are two important qualifications in the above statement. The first is that there be a sufficiently rich medium for information transmission. On eBay Motors, sellers have a variety of ways to communicate the features and history of their car to potential buyers. 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.

[1]For 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).

[2]Sources: eBay Annual Report 2006 (investor.ebay.com/common/dar/dar.cfm?documentid=1649&companyid=ebay); eBay Press Release (http://investor.ebay.com/releasedetail.cfm?ReleaseID=206868).

Item Specifics - Cars & Trucks

Chevrolet : Corvette LINGENFELTER

01 CORVETTE Z06 LINGENFELTER TWIN TURBO 500+HP VIDEO!

Miles:	13879	Doors:	2 doors or more
Transmission:	Manual	Interior:	Red
		Year:	2001
Warranty:	No	VIN Number:	1g1yy12sx15114360 Get the Vehicle History Report
Title:	Clear	Exterior:	Red
Condition:	Used	Inspection:	–
Engine:	8 CYLINDER 5.7 LITER MFI		
Fuel Type:	Gasoline		

Standard Equipment:

CASSETTE	POWER WINDOWS	CRUISE CONTROL	LEATHER
CD	POWER DOOR LOCKS	ABS	AM/FM RADIO
Z06 HARDTOP	2 WHEEL DRIVE	POWER SEAT DRIVER	DUAL AIR BAGS
8 CYLINDER 5.7 LITER MFI			

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 oil 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
Lanardi 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
The clutch was also upgraded to a Spec Stage III Clutch
97y M/T Radiator

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
Comp-Cam
Oil cooler with braided lines
Ron Davis racing radiator
Stainless steel headers
Greedy spec 2 boost gauge
World challenge brake set-up
14"Eradi speed slotted front rotors with Alcon Bear calipers and Hawk HT14 pads
12.9 inch rotors with large corvette calipers on rear with Hawk ht10 pads
Stainless steal brake lines
T1 gm racing sway bars
Coil overs with front HR springs (800 lbs)
HRE wheels model 546
Michelin pilots fronts are 265 and backs are 335
5 point sparco safety harness both seats with safety bar
Most of the original equipment that was up graded is available
18 inc Michelin Tires

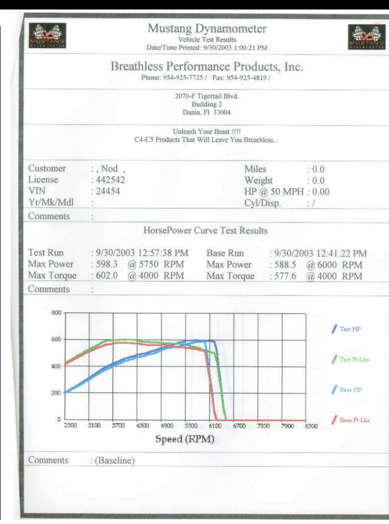

Figure 1: **Information disclosed in an eBay Motors auction** On this auction webpage, the seller has provided many different forms of information about the Corvette he is selling. These include the standardized eBay description (top panel), his own full description of the car's options (middle panel), 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.

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. Given the rapidly falling cost of storage space and server time, rich media are increasingly accessible for online transactions in other markets as well.

The second qualification is that the information disclosures should be credible. In the case of eBay Motors, buyers typically see the car before final payment and thus can verify the information provided. In addition to this, eBay helps to ensure that sellers are actively prosecuted for material misrepresentation, misrepresentation that itself is made difficult by the nature of the medium, since it requires actively doctoring or substituting photos. In general, this credibility requirement is more strenuous, since good technology may not be enough without institutional participation and some thought as to the appropriate payment system. Yet where these requirements are satisfied, one can expect to see voluntary information disclosure by sellers in online markets.

My empirical analysis of this particular online market is in three parts. Initially, the focus is on testing under what circumstances sellers disclose information, and how buyers respond to the information disclosed. I run an extensive set of hedonic regressions to demonstrate that there is a large, significant and positive relationship between auction prices and the number of photos and bytes of text on the auction webpage. But since auction prices are determined by strategic interaction among the bidders, a precise test of the selective disclosure hypothesis requires structural estimation of an auction model. In this second part of the paper, I argue that a common value auction model is appropriate for this setting, and develop a new pseudo-maximum likelihood estimation approach to recover the structural parameters. I conclude by using these parameters to simulate a counterfactual with a coarse information transmission mechanism, to quantify the importance of rich media in reducing the potential for adverse selection.

I make use of a new dataset with over 40000 eBay Motors listings over a 6 month period, containing variables relating to item, seller and auction characteristics. I consider two measures of the quantity of information on an auction webpage: the number of photographs, and the number of bytes of text. Through running a series of hedonic regressions, I show that these measures are significantly and positively correlated with price. The estimated coefficients are extremely large, and this result proves robust to controls for marketing effects, seller size and seller feedback. This suggests that it is the webpage content itself rather than an outside factor that affects prices, and that the text and photos are therefore genuinely informative. Moreover, it seems that information is selectively disclosed by sellers, so that higher quality cars are associated with more information.

Yet it is possible that the positive relationship between prices and information could arise from strategic considerations related to the winner's curse effect. Since it is difficult to distinguish these strategic effects from the selective disclosure explanation in the hedonic regressions, I structurally estimate a common value auction model. In estimating the model, I propose a new pseudo-maximum likelihood estimation approach that addresses two of the common practical concerns associated with such estimation. One of the concerns is that in an eBay context it is difficult to determine which bids actually reflect the bidder's underlying valuations, and which bids are spurious. In equilibrium though, the final auction price will always reflect the valuation of the bidder with second most favorable private information. I thus obtain a robust estimator by maximizing a function based on the observed distribution of prices, rather than the full set of observed bids. The second concern relates to the computational cost of computing the moments implied by the structural model, and then maximizing the resulting objective function. I show that by choosing a pseudo-maximum likelihood approach, one is able to use a nested loop procedure to maximize the objective function that limits the curse of dimensionality.

The results suggest that sellers selectively disclose information based on the quality of their vehicle. One natural question that arises is how much better the current eBay system is at limiting information asymmetries and adverse selection than a system with more coarse information transmission mechanism would be. In the final part of the paper, I investigate this with the aid of a counterfactual simulation. As I have a structural model, I am able to simulate the bidding functions and compute expected prices for the observed regime and a counterfactual in which sellers can only reveal standardized and basic characteristic data. The estimated impact of this regime change is large, driving a wedge between the value of the car and the expected price paid by buyers. This suggests that the richness of the disclosure mechanism plays an important role in reducing adverse selection in this market.

This paper is closely related to the diverse empirical literature on disclosure (recent papers in economics include Jin and Leslie (2003) and Jin (2005)). A closely related paper is that of Jin and Kato (2005) who examine the sale of baseball cards on eBay, and conclude that seller claims about about baseball card quality are often not truthful. We will see that in the used car market where the stakes are higher, the disclosed information tends to be more detailed, and that the incentives for sellers to lie are weakened because the information may be verified before payment. This paper contributes to this literature by meticulously documenting a case in which there is selective disclosure, rather than no disclosure (as in the adverse selection setting of Akerlof (1970)) or complete unravelling (as in the theories of Grossman (1981) and Milgrom (1981)). This is also the first paper to use a structural model to quantify the impact of these disclosures in reducing information asymmetry. Consistent with my results, Adams, Hosken, and Newberry (2006) looks directly for adverse selection in the eBay market for used Corvettes, and finds little evidence of it.

The online auctions literature has tended to ignore seller disclosure and instead focus on the role of the seller reputation mechanism. Various authors have found that prices respond to feedback ratings, and have suggested that this is due to the effect of these ratings in reducing uncertainty about seller type (e.g. Resnick and Zeckhauser (2002), Houser and Wooders (2006)). I find that on eBay Motors the strength of the relationship between price and information measures is far larger than that between price and feedback ratings. This suggests that reputation may be only part of the explanation for the success of online markets with item heterogeneity and information asymmetry. A notable exception to this literature is the paper by Yin (2006), which examines the link between information dispersion and auction prices in used computer sales on eBay. She finds a negative relationship between information dispersion and auction prices, and takes this as evidence of a winner's curse effect. I examine this hypothesis directly, and find that although there is evidence of information-related winner's curse effects, in this market they are small.

I also make a methodological contribution with regard to common value auction estimation. I provide a pseudo maximum likelihood estimation strategy that has computational advantages over the Bayesian approach used in Bajari and Hortaçsu

(2003) and the quantile regression approach of Hong and Shum (2002). In arguing for a common values auction mode, I provide an implementation of the test for common values suggested by Athey and Haile (2002) in the supplementary material, using the subsampling approach of Haile, Hong, and Shum (2003).

The paper proceeds as follows. Section 2 outlines the market, and section 3 presents an analysis of the role of information in the market. Section 4 introduces the auction model and presents the estimation approach. The estimation results follow in section 5. Section 6 details the counterfactual simulation and section 7 concludes.[3]

2 Description of the Market and Data

2.1 eBay Motors

eBay Motors is the automobile arm of online auctions giant eBay. It is the largest automotive site on the Internet, with over $12 billion worth of vehicles traded in 2006 alone. Every month, approximately 95000 passenger vehicles are listed on eBay, and about 15% of these will be sold on their first listing.[4] 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, so that bidders typically cannot examine the car in person and must rely on the information on the auction webpage.[5]

Listing a car on eBay Motors is straightforward. For a fixed 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. 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

[3]Proofs of all propositions, as well as most estimation results, are to be found in the appendix.

[4]The remainder are either re-listed (35%), sold elsewhere or not traded. For the car models in my data, sales are more prevalent, with 27.94% sold on their first listing.

[5]Source: Auction123 (http://www.auction123.com/ebayadvantages.aspx).

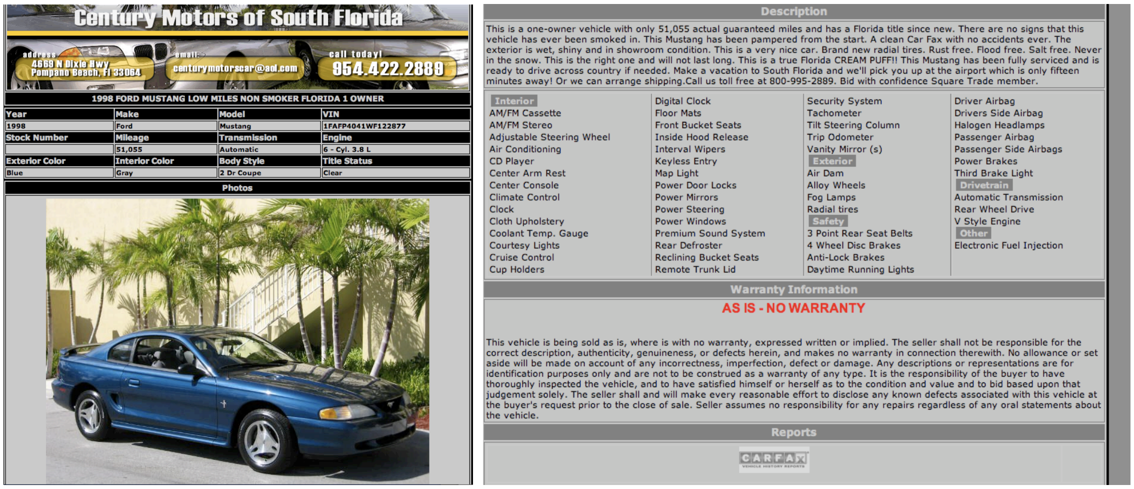

Figure 2: **Part of an auction webpage** The dealer selling this vehicle has used proprietary software to create a professional and detailed listing. The item description seen on the right contains information on the car, a list of all the options included, warranty information, and a free CARFAX report on the vehicle history. The dealer also posted 28 photos of the car.

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.[6]

An example webpage is given in Figure 2, which shows part of the listing by a dealer. His customized auction webpage contains some useful information for buyers: a text based description of the history of the car, a full itemization of the car features in a table, a free vehicle history report through CARFAX, and a description of the seller's dealership. In some listings sellers document the parts used in repairs of the vehicle and modifications that they may have made (often with photographic evidence of the original receipts, as seen in Figure 1). In the left panel of Figure 3, I show a graphic displaying the condition of a car's exterior, while in the right I show the Kelley Blue Book report on the retail value of a vehicle that was provided by a seller. Overall, eBay Motors differs markedly from the rest of eBay in the level of detail provided by sellers in listing their items for sale. This will be clear in the data analysis.

Once the seller has put the vehicle up for sale, people may bid on it. Some sales

[6]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.

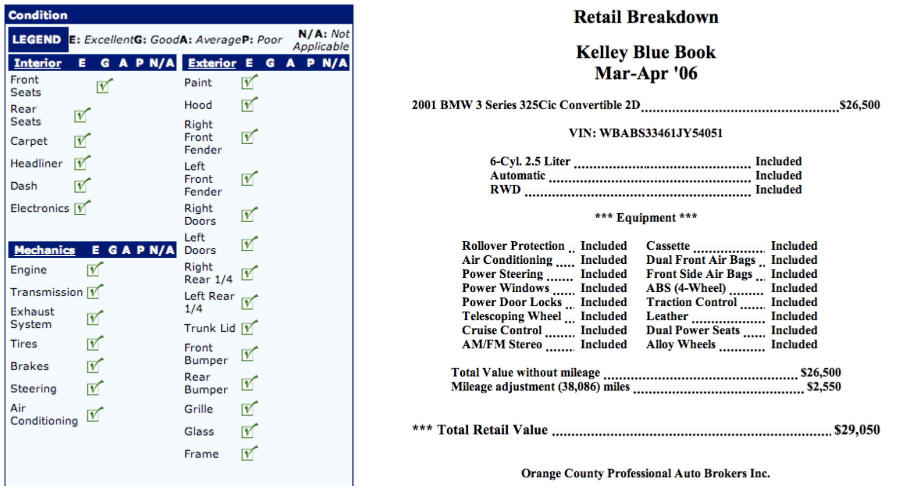

Figure 3: **Other forms of information** The left panel shows a graphic that details the exterior condition of the vehicle. The right panel shows the Kelley Blue Book information for the model-year of vehicle being auctioned.

take place at a fixed price, but the vast majority of cars are sold in an English auction format.[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 through direct communication via e-mail or telephone. At the close of the auction, the highest bidder receives the car. As noted by eBay, "most buyers opt to pickup the vehicle in person."[8] The result is that much of the information provided by the seller is often *ex-post verifiable*. Material misrepresentations by the seller are grounds for the buyer to break the contract and not purchase the car. In fact, eBay will act to have sellers prosecuted for fraud in cases of blatant misrepresentation. Sellers thus have limited incentive to mislead buyers, and so buyers can therefore be reasonably confident that the (verifiable) information provided by sellers during the auction is accurate. This information acts to limit the asymmetric information problem faced by bidders. Quantifying the impact of these disclosures on information asymmetries is an empirical question, and thus it is to the data I now turn.

[7]The fixed price formats are either "best offer" or "buy-it-now" auctions.
[8]Source: eBay Motors Seller's Guide, http://pages.motors.ebay.com/howto/selling/closeB.html

2.2 Data and Variables

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 models of car 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, cars with salvage title, re-listings, and those auctions in which no one entered a bid.[9] I also drop auctions in which the webpage was created using proprietary software.[10] The resulting dataset consists of over 50000 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).

In each auction, I observe a number of item characteristics including model, year, mileage and transmission. I also observe whether the vehicle is currently under 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 quantitative measures of this content: the number of bytes of text in the vehicle descriptions provided by the seller, and the number of photos posted on the auction webpage. In addition, I have encoded dummy variables based on text searches for key phrases such as "rust" or "no rust", focusing on information a potential buyer might typically be interested in when buying a car. I construct one more additional variable, "competition index", which is defined as the ratio of the number of models of similar age and identical model also being auctioned that week to the weekly average.[11] This captures how competitive the eBay Motors market for similar vehicles is in that particular week.

[9] I drop cars under salvage title because they are in a different market from the remaining cars. I drop re-listings both because the information set held by bidders bidding on the re-listed cars includes the unmodeled past bidding history, and to avoid oversampling identical cars. I drop auctions without bids because I have no dependent variable.

[10] Webpages created using proprietary software are often based on standardized templates, and some of the text of the item description is not specific to the item. Text-based comparisons of standard eBay listings with those generated by advanced software are not meaningful.

[11] I am implicitly defining the cars in competition as identical models of similar age in the same week. I have experimented with broader market definitions, with little change in the results.

I supplement this main data source with data on book values publicly available at edmunds.com.[12] 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 28000 observations.

3 The Role of Information

I wish to analyze when sellers disclose information, and how buyers respond to this information. To test buyer response, I run a series of hedonic regressions and note the relationship between price and information measures, controlling for a wide range of possible confounders. Testing for a monotone seller disclosure policy is more difficult. A direct test would be to sample a set of cars, hire a mechanic, assess the quality of each car and check that on average sellers disclose more information when selling higher quality vehicles. Unfortunately, I do not have access to direct quality measures, and so I must rely on an indirect test. Under the maintained assumptions that sellers prefer to purchase high quality cars, and that information is truthfully disclosed by sellers, I can infer a monotone disclosure policy from a monotone relationship between price and the information measures. This is because if buyers are obtaining genuine information from the webpage, and are willing to pay more on average when there is more information, it must be because on average the cars they are obtaining are better. In a later subsection, I examine carefully the assumption that the webpage content is informative in order to support my selective disclosure interpretation of the data.

3.1 Hedonic Regressions

I begin with a number of hedonic regressions. The specification is log-linear:

$$\log(p_j) = z_j\beta + \varepsilon \tag{1}$$

[12] I used the "used car appraiser" at http://www.edmunds.com/tmv/used/index.html.

where z_j is a vector of covariates.[13] I estimate the relationship via ordinary least squares (OLS), including model and year fixed effects in all specifications. I report the results in table 1, suppressing the fixed effects.

In the first specification, the vector of covariates includes mileage and transmission, as well as the log number of bytes of text and the number of photos. The coefficients generally have the expected sign and all are highly significant.[14] Of particular interest is the sheer magnitude of the coefficients on the amount of text and photos. Each additional photo on the webpage is associated with a selling price that is 1.75% higher, which for the average car in the dataset is around $175 more. Likewise, a 12% increase in the amount of text (one standard deviation in the data) is associated with a 0.75% increase in the price, or $75. A picture is indeed worth many words! This is because pictures are generally far more informative about vehicle condition and quality than the text in vehicle descriptions, although in later regressions we will see that certain key phrases are also associated with large effects.

I do not assert a causal relationship between the amount of text and photos and the auction price. If this were the case, you could make a lot of money by selling extremely well photographed wrecks. My claim is that sellers use a monotone disclosure policy, whereby they disclose only sufficiently favorable information. Buyers then use this information to make an informed decision as to the car's value. The information that sellers may disclose could relate to the options on the car, the condition of the vehicle, maintenance history and documentation, and vehicle history and usage, all of which are strong determinants of car value. To put the magnitude of the coefficients in context, the value of a used car of a given model-year and mileage can vary by thousands of dollars depending on these factors. Given that sellers with favorable information will disclose it, rational buyers will view the absence of information as a bad signal about vehicle quality, and adjust their bids accordingly. This combination of direct information revelation (about good features) and signaling through absence of information (bad features) explains the observed positive correlation in the data between price and the information measures.

In the next three specifications, I explore alternate explanations. It could be that this

[13]Prices are always positive, so a log-linear specification is appropriate.

[14]One might have expected manual transmission to enter with a negative coefficient, but I have a large number of convertible cars and pickups in my dataset, and for these, a manual transmission may be preferable.

Table 1: Hedonic Regressions

	Log Price			
	(1)	(2)	(3)	(4)
Log Miles	-0.1248	-0.1251	-0.1316	-0.1304
	(0.0024)	(0.0023)	(0.0025)	(0.0025)
Log Text Size (in bytes)	0.0629	0.0560	0.0819	0.0860
	(0.0037)	(0.0036)	(0.0041)	(0.0043)
Number of Photos	0.0175	0.0166	0.0168	0.0182
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Manual Transmission	0.0904	0.0920	0.0904	0.0897
	(0.0069)	(0.0068)	(0.0072)	(0.0072)
Log Feedback			-0.0225	-0.0223
			(0.0019)	(0.0019)
Percentage Negative Feedback			-0.0035	-0.0035
			(0.0009)	(0.0009)
Competition Index			-0.0098	-0.0094
			(0.0129)	(0.0129)
Warranty				0.5780
				(0.0840)
Warranty*Logtext				-0.0459
				(0.0119)
Warranty*Photos				-0.0128
Model Fixed Effects	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes
Number of Bidders Fixed Effects	no	yes	yes	yes

Estimated standard errors are given in parentheses. The model fits well, with the R^2 ranging from 0.649 in (1) to 0.66 in (4). The nested specifications (2)-(4) include fixed effects for the observed number of bidders. Specification (3) includes controls for seller feedback and competition from similar listings. Specification (4) includes a warranty dummy and its interactions with the information measures.

is an entry effect, whereby slick webpages with a lot of text and photos attract more bidders and thus the cars sell for higher prices. In table 5 in the appendix, I present results from a regression of the observed number of bidders on car characteristics and the text and photo measures. From those results it is clear that though there is a significant and positive relationship between the observed number of bidders and the information measures, it is relatively weak, with each additional photo associated with an extra 0.04 bidders, and an increase of 10% in the text associated with 0.01 extra bidders. In column (2) of table 1 I control for the entry effect in the hedonic regressions by adding a fixed effect for each number of observed bidders. I find that even after controlling for entry, the coefficients on text size and the number of photos are extremely large and highly significant.

Another potential explanation is that the amount of text and photos are correlated with seller feedback ratings, which are often significant determinants of prices on eBay. In column (3), I find not only that including these controls has little effect on the information coefficients, but that for this specific market the effects of seller reputation are extremely weak. 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. I also include my constructed index of market competition in the specification in column (3), but although the coefficient has the correct sign, it is not significant. This may be because many of these cars are not close substitutes, particularly the vintage cars. Later on in the paper we will see that the index shows up significantly in the market for reliable cars.

In column (4), I try to test my hypothesis that text and photos matter because they inform potential buyers about car attributes that allow them to infer the quality of the vehicle. To do this, I include a dummy for whether the car is under warranty and interact it with the information measures. One would expect that if the positive relationship between price and text and photos stems from the information content of the webpage, the coefficient would be smaller on a car under warranty since the information content is less valuable when the buyer is partially insured from risk by the warranty. The results in column (4) indicate that the interaction terms are significantly negative as expected, while the warranty significantly raises the expected price of the vehicle. For cars that are not under warranty, the coefficients on text and photos increase.

3.2 Controlling for Seller Heterogeneity

One might also be concerned that the results are driven by seller heterogeneity. Frequent sellers such as car dealerships tend to produce webpages with more content, since they have stronger incentives to develop a good template than less active sellers. 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 6, to be found in the appendix, give the results of separate hedonic regressions for those two groups. The coefficients on the information measures are similar in size across both groups, and are also similar to those for the full sample. In the final column, I consider the sample of dealers who list at least four cars during the sample period, and include a seller-specific fixed effect for each of them. 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 text and 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.

3.3 Controlling for Book Values

In tables 7 in the appendix, I examine whether these results are robust when book values are included rather than model and year fixed effects. This allows for a possibly non-linear relationship by model-year, as well as accounting for possible differences due to trim. Unfortunately, I only have book value data for model years from 1990 onwards. So for this subset of the data, I regress price on the dealer retail book value as well as the information measures.[15] I find that cars on eBay sell on average for 85% of the retail book value, which is consistent with the gap between private party prices and retail prices found in the Kelly Blue Book. The coefficients on the text measures are still large and significant, but smaller than in regressions on the full dataset because these are newer cars and hence differences in condition, repair and other factors are smaller. Comparing the book value results in the first column with the corresponding fixed effect results, we see that using fixed effects instead of book values barely impacts the coefficient on photos, while it increases that on text slightly. This suggests that estimates of the relationship between price and the number of photos are reasonably robust to specification changes, while that of text is less so.

Table 2: Key Phrases

	% Affirmative (e.g. has dent)	% Negative (e.g. is not dented)	Estimated Coefficient	Standard Error
"Dent"	9.9	1.73	-0.1013	0.0094
"Accident"	2.06	0.26	-0.1572	0.0205
"Crack"	9.10	2.04	-0.0829	0.0094
"Broken"	2.62	0.14	-0.1997	0.0185
"Rust"	15.61	5.90	-0.1786	0.0069
"Factory Part"	0.09	0	0.2147	0.1042
"Original Part"	0.74	0	0.1223	0.0359
"Garage Kept"	3.74	0	0.1932	0.0172
"Invoice"	0.48	0	0.2947	0.0440
"Documentation"	2.24	0.06	0.1263	0.0205

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 (3) of table 1 above.

3.4 Testing for Information

One of the maintained assumptions in this analysis is that the webpage content is genuine information about the car. While this certainly seems reasonable, it is nonethe-

[15] Since the book values are for a typical mileage, I also include interactions between year and log mileage to capture differences in price due to mileage.

less key to the analysis and deserves some attention. In this section, I perform two tests on the information content of the auction webpage. The first is to 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 2, I present the proportions of webpages with affirmative (e.g. "has dent") and negative (e.g. "does not have dent") versions of these phrases, and the estimated coefficient on a dummy for this phrase in a hedonic regression of the form given in equation 1. 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 (20% less), while those that are "garage kept" and that come with documentation and receipts sell on average for considerably more (19% more).

Second, I fix a particular model of vehicle (the mustang, chosen because it is the model with the most observations) and consider the year-by-year relationship between price and photos and text from 1984-2006. If the webpage content conveys information, then such information is probably most valuable for assessing the quality of older cars, since there is considerable variance in the driving history and condition of these cars. On the other hand, if webpage content is uninformative, there is no reason to expect the relationship between price and information measures to vary with the age of the car. To test this, I run year-by-year hedonic regressions for the mustang, and recover the coefficients on log of text and the number of photos for each year. By fixing a specific model of vehicle I am able to abstract away from potential non-linearities across model-year. [16] I plot the estimated coefficients against age in Figure 4. Under the hypothesis that webpage content is informative, one would expect the coefficients to increase with the age of the vehicle. This is indeed what I find. The solid line plots the quadratic best fit to the estimated coefficients, and is clearly monotone increasing. In particular, the relationship between price and the number of photos is extremely strong for older mustangs.

[16] An alternative is to run the year-by-year estimation on the full sample with model fixed effects. The results obtained are similar.

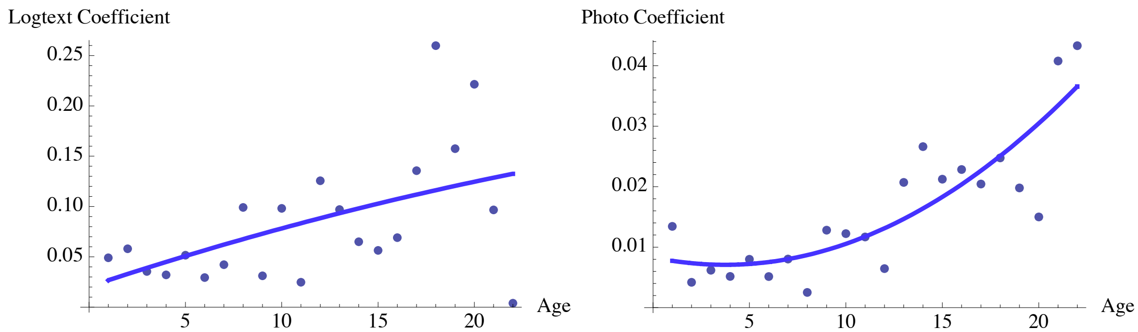

Figure 4: **Age and Information** The left panel plots car age against the estimated coefficients on logtext obtained from year-by-year hedonic regressions. The right panel is an identical plot for age against photo coefficients. The solid lines are quadratic best fits to the generated data points.

3.5 Discussion

There are two points that emerge from the above analysis. First, the final price is higher in auctions where there is a lot of webpage content in the form of text and photos, even after controlling for feedback, number of bidders, seller identity, book value and measures of competition. Second, this content appears to be genuine information relating to the vehicle. It follows that sellers selectively disclose information, posting more photos and text when selling a high quality vehicle than when selling a "lemon." Bidders pay more for high quality vehicles, and this in turn induces the positive correlation between information measures and prices we observe in the data.

It is perhaps not surprising that bidders should respond to the text and photos posted by sellers. This information is ex-post verifiable, and there are penalties for misrepresentation. It is interesting though how big the effects are. I think of this as a joint measure of the richness of the information transmission mechanism and of the underlying heterogeneity of the vehicles.

By contrast, the result that information disclosure by sellers is selective, is somewhat surprising. The astute reader will wonder why an unraveling of the information asymmetry does not obtain, whereby sellers with cars of middling bad quality disclose their private information to buyers so as not to be confused with those who are selling cars of terrible quality. There are a number of reasons to believe that disclosure here will be selective and unravelling will not be complete. First, there are disclosure costs for the seller in the form of the opportunity costs of taking photos, writing text and finding documents to support their claims. As Jovanovic (1982) points out, this leads immediately to selective disclosure, since only those with favorable information find it profitable to incur the disclosure costs. Second, there is uncertainty on the bidder side about what the seller knows. As a result, bidders cannot infer with certainty that the absence of information is bad news, and this in turn dilutes the benefits of disclosure (Shin 1994). Third, information is often binary ("this car has/has not only had original parts"), and sellers with unfavorable news are at best indifferent about revealing bad news, and strictly prefer not to given any disclosure costs. For all of these reasons, sellers will generally reveal information about a car only when it is positive (or obvious from any photos they might take).

Overall then, this analysis of the data provides strong evidence that it is selective disclosure that generates the relationship between price and seller provided information such as text and photos. Yet in an auction setting, bids are strategically chosen and are not in general equal to the underlying valuations held by bidders. In particular, where there is uncertainty about a common component of value and bidders have access to different signals, they will be concerned that in the event that they win, they have been the most optimistic in their valuation, and thus overpaid. Bidders account for this Winner's Curse effect by shading their bids down, and the extent to which they do so depends on the level of common uncertainty. Since text and photos reduce uncertainty, it could be that bidders are more aggressive in auctions with a high level of information, and it is this strategic interaction that explains the above results, rather than selective disclosure. Put another way, two identical cars could sell for different prices because one had more text and photos - that is, that there is a strategic causal effect of information provision. I argue that although this may be the case, causation typically runs the other way, so that sellers with high quality cars document this meticulously through text and photos, and receive higher prices because buyers are made aware of the quality of the car.

Teasing these stories apart requires explicit examination of the strategic interactions, and thus a model of the auction equilibrium. This will allow me to recover the relationship between the latent valuations held by bidders and the information measures, which are uncontaminated by Winner's Curse effects. I will also be able to simulate a counterfactual scenario in which the information transmission mechanism is less rich, and examine the adverse selection implications of this coarsening.

4 Auction Model and Estimation

Modeling demand on eBay is not a trivial task. The auction framework has a bewildering array of institutional features, such as proxy bidding, secret reserves and sniping software. With this in mind, it is critical to focus on the motivating question: how does disclosed information affect bidder valuations, and what is its impact on bidding strategies and prices?

I make three important modeling choices. The first is to model an auction on eBay Motors as a symmetric pure common value auction. Realistically, there are both common and private components of value, and bidders are asymmetric in their access to information. Local bidders are both likely to privately value the car more (their shipping costs are lower), and to have different information about common components (they can arrange to view the car in person). The correct model is certainly one of affiliated values. But estimation of these models is extremely difficult, and so it is necessary to choose either the affiliated private values (APV) framework, or the pure common values (CV) framework.

I believe that the common value framework is better suited to my purposes than the private values framework for a number of reasons. It is consistent with standard models of adverse selection, where all buyers value the car equally, but face common uncertainty as to the car's quality. The information disclosed by the seller on the auction webpage is commonly observed by all bidders, and thus can be modeled as the source of the common prior that bidders have for the object's value. There are genuine differences in the information possessed by agents, depending for example on whether they have privately inspected the car, or hired a mechanic to do a check, or accessed title information on Carfax.com. This will lead bidders to be cautious of the Winner's Curse and overpaying. Indeed I implement the Athey and Haile (2002) test of symmetric private values against symmetric common values in the supplementary material, and reject the private value null hypothesis at the 5% level. This rejection stems from evidence of the Winner's Curse in the data.

Insisting that the model is one of private values ignores the possibility of a Winner's Curse effect and thus the potential for an explanation of the earlier results in which causality flows directly from information through to prices, rather than from underlying quality differences. If I specified a private values model, I would in some sense impose a selective disclosure explanation on the data by not allowing one based on strategic effects. The converse is not true. Ignoring the private value elements in the data creates biases in the estimates, but the bias can be signed, and as I will show in the results section, such biases are in favor of my conclusions.

The second modeling choice is to model an eBay auction as a second price sealed bid auction with unobserved bids. This is motivated by the Bajari and Hortaçsu (2003) model of late bidding on eBay, which argues that the presence of "sniping" software and bids in the final moments of an auction make bids essentially simultaneous, since players do not have time to observe and best respond to each other's play, as would be typical in an English auction. Indeed, even on eBay Motors, there is a large degree of sniping, with 5% of all bids on Corvettes over a two year period coming in the last minute of the auction (Adams, Hosken, and Newberry (2006)). It is also important to note that many bids will be unobserved, since a bid is only recorded on eBay if it is above the current price. Then, as has been argued by Song (2004), the only bids that will be recorded with certainty are those of the first and second highest bidders. The bidding history does not reveal the bid of the highest bidder, so essentially, the only bid that you can be sure represents the true valuation of a bidder in the auction is that of the bidder with the second highest valuation. Because of eBay's proxy bidding system, this turns out to be the price, which is why the price will be the dependent variable in the estimation strategy that follows.

The final modeling choice is to fully parameterize the auction model. This is motivated both by practical and theoretical concerns. On the theoretical side, there are a number of strong negative identification results for nonparametric common value auction models (Athey and Haile (2002)). I defer a discussion of this issue to the section on identification below. From a practical perspective, I show that the parametric pseudo-maximum likelihood estimator I present below has a number of computational advantages, and this allows the model to be quickly estimated, permitting many different specifications to be examined. In the results below I do some out of sample simulations and show that chosen parametric model performs reasonably well in making out of sample predictions. I believe this should allay concerns that a parametric model is unreasonable here.

4.1 Model

Consider a symmetric common values auction model, as in Wilson (1977). There are N symmetric bidders, bidding for an object of unknown common value V. Bidders have a common prior distribution for V, denoted F. Each bidder i is endowed with a private signal X_i, and these signals are conditionally i.i.d., with common conditional distribution $G|v$. The variables V and $X = X_1, X_2 \cdots X_n$ are affiliated. The realization of the common value is denoted v, and the realization of each private signal is denoted x_i. The realized number of bidders n is common knowledge.[17] One may think of the bidders' common prior as being based on the content of the auction webpage. Their private signals may come from a variety of sources such as private communications with the seller through e-mail or phone and the results of inspections they have contracted for or performed in person. I assume that the seller does not set a reserve price, as on eBay the vast majority of auctions have low minimum bids and *secret* reserve prices, which in equilibrium should not affect entry and bidder behavior. Then in a second price sealed bid auction, the equilibrium bid strategy for a bidder with private signal x is given by:

$$v(x, x; n) = E[V|X_i = x, Y = x, N = n]$$

where Y is the maximum of the other bidder's signals. Letting $x^{(1:n)}, x^{(2:n)} \cdots x^{(n-1:n)}, x^{(n:n)}$ denote the ordered signals (from lowest to highest), the final price in the auction is the bid of the bidder with the second highest signal. Thus the final price is given by:

$$p = v(x^{(n-1:n)}, x^{(n-1:n)}; n) \tag{2}$$

4.2 Specification

Auction covariates such as item characteristics, photos and text are observed by all bidders, and thus influence their common prior. Thus I let covariates enter the model

[17]This assumption is somewhat unrealistic, as bidders in an eBay auction can never be sure of how many other bidders they are competing with. The alternative is to let auction participants form beliefs about n, where those beliefs are based on a first stage entry game, as in Levin and Smith (1994). I choose this assumption for simplicity.

in influencing both the mean and variance of the prior distribution. Formally, let $F|z$ be the log-normal distribution, and let $G|v$ be the normal distribution centered at $\tilde{v} = \log v$. Then we have:

$$\begin{aligned} \tilde{v} &= \mu + \sigma\varepsilon_v \sim N(\mu, \sigma^2) \\ x_i|\tilde{v} &= \tilde{v} + r\varepsilon_i \sim N(\tilde{v}, r^2) \end{aligned}$$

where $\varepsilon_v, \varepsilon_i$ are i.i.d standard normal random variables for all i.[18] Three variables form the primitives of the auction model: the prior mean μ, the prior standard deviation σ, and the signal standard deviation r.

Next I link the primitives (μ, σ, r) to the auction covariates. I allow the mean and standard deviation of F, μ and σ, to vary with auction covariates z. I include in z my empirical measures of information content, the number of photos and bytes of text. The signal variance parameter r is treated as a constant, as the private information is by definition orthogonal to the publicly available information captured in the covariates. I choose a specification in index form:

$$\begin{aligned} \mu_j &= \alpha z_j \\ \sigma_j &= \kappa(\beta z_j) \end{aligned}$$

where κ is an increasing function with strictly positive range, so that $\kappa(x) > 0 \; \forall x$.[19] This specification yields admissible values for σ_j.

Through identifying the parameter vector α, I pin down how prior valuations respond to the photos and text on the auction webpage, and thus determine what bidders learn from the webpage. The prior valuations, unlike the final prices, are uncontaminated by strategic considerations. The parameter vector β captures the relationship between the prior variance and the information measures, and thus determines to what extent

[18] The choice of the log normal distribution for v is motivated by two considerations. First, all valuations should be positive, so it is important to choose a distribution with strictly positive support. Second, the log prices observed in the data are approximately normally distributed, and since the price distribution is generated by the value distribution, this seems a sensible choice (even though there is no formal reason that log normal values would generate log normal prices). The choice of a normal signal distribution is for convenience, and seems reasonable given that the signals are latent and unobserved.

[19] For computational reasons, I choose to base $\kappa(\cdot)$ on the arctan function - specifics are given in the supplementary material.

seller provided information really tightens the valuation distribution and limits the potential for adverse selection. It is also through this reduction in uncertainty that the seller may limit the Winner's Curse and induce more aggressive bidding and higher prices. This Winner's Curse effect is mediated by the signal variance r and can be identified from the full model.

4.3 Estimation Approach

The above specification governs the relationship between the parameter vector $\theta = (\alpha, \beta, r)$, bidder prior valuations, and bidder signals. Prices are a complex function of these parameters, since they are equal to the equilibrium bid of the bidder with the second highest signal, an order statistic. Thus even though I have a fully specified parametric model for valuations and signals, the distribution of prices is hard to determine. This prevents a straightforward maximum likelihood approach. I can however compute the mean $E[\log p|n, z_j, \theta]$ and variance $\Omega[\log p|n, z_j, \theta]$ of log prices implied by the structural model through equation 2 for any number of bidders n, given parameter vector θ and covariates z_j.[20] To do so, I take the bidding function given in equation 2 and integrate out the latent ordered private signal $x^{(n-1:n)}$ under the distributional assumptions parameterized by θ and z_j.

Notice that these predicted moments vary with the number of bidders n. In practice the true number of bidders is unobserved, precisely because, as noted above, some bids are unobserved. The number of bidders observed in the data is a lower bound for the true number of bidders. I make the simplifying assumption that the true number of bidders n is exactly equal to the number of bidders that I observe bidding in my data.[21] I perform robustness checks later in the paper to see how the results would be affected by changing this assumption.

Computing these moments allows me to estimate the model by pseudo-maximum likelihood (Gourieroux, Monfort, and Trognon 1984).[22] In this approach, I maximize

[20] Explicit formulae for these moments are to be found in the supplementary material.

[21] Where there is only one bidder, I compute the moments as though there were two, as if the bidder believed he was competing against the seller.

[22] One may be concerned about parameter support dependence, a problem noted in Donald and Paarsch (1993). It can be shown that under this specification the equilibrium bid distribution has full support on $[0, \infty)$, so this problem does not arise here.

a pseudo-likelihood function derived from a quadratic exponential family. Since the observed distribution of log prices is approximately normal, and the normal distribution is quadratic exponential, I choose it as the basis of my pseudo likelihood function and maximize the criterion function:

$$\mathcal{L}(\theta) = (-1/2) \sum_{j=1}^{J} \left(\log(\Omega[\log p|n, z_j, \theta]) + \frac{(\log p_j - E[\log p|n, z_j, \theta])^2}{\Omega[\log p|n, z_j, \theta]} \right) \tag{3}$$

This is not the only estimation approach available. Estimation by generalized method of moments (GMM) with explicit moment computation or by simulated method of moments (SMM) with simulation of the moments $E[\log p|n, z_j, \theta]$ and $\Omega[\log p|n, z_j, \theta]$ would work just as well. But in all of these approaches the large number of covariates implies a high dimensional parameter space, and the resulting maximization problem is extremely challenging. In the case of the pseudo maximum likelihood approach, however, there is a result which considerably eases the computational burden:

Proposition 1 (Properties of the Bidding Function) *Let $v(x, x; n, \mu, \sigma, r)$ be the bidding function under model primitives (μ, σ, r), and let $v(x-a, x-a; n, z, \mu-a, \sigma, r)$ be the bidding function when the signals of all bidders and the prior mean are decreased by a. Then $\log v(x, x; n, \mu, \sigma, r) = a + \log v(x-a, x-a; n, \mu-a, \sigma, r)$ and consequently:*

$$E[\log p|n, z_j, \alpha, \beta, r] = \alpha z_j + E[\log p|n, z, \alpha = 0, \beta, r] \tag{4}$$

$$\Omega[\log p|n, z_j, \alpha, \beta, r] = \Omega[\log p|n, z_j, \alpha = 0, \beta, r] \tag{5}$$

This result is extremely useful. To see this, fix the parameters β and r, and consider finding the parameter estimate $\widehat{\alpha}(\beta, r)$ that maximizes the criterion function (3). Using the result of the above proposition, we can write this parameter estimate as:

$$\widehat{\alpha}(\beta, r) = \underset{\alpha}{\text{Argmin}} \sum_{j=1}^{J} \frac{1}{\Omega[\log p|n, z_j, \alpha = 0, \beta, r]} (\log p_j - E[\log p|n, z_j, \alpha = 0, \beta, r] - \alpha z)^2$$

which for fixed β and r can be solved by weighted least squares (WLS) with the dependent variable $y_j = \log p_j - E[\log p|n, z_j, \alpha = 0, \beta, r]$ and weights $w_j = 1/\Omega[\log p|n, z_j, \alpha =$

$0, \beta, r]$. This allows a nested estimation procedure in which I search over the parameter space of (β, r), using WLS to calculate $\widehat{\alpha}(\beta, r)$ at each step. Since I choose a specification in which most of the covariates affect the prior mean and not the variance, this space is considerably smaller than the full space of (α, β, r), and the computation is much simpler.[23]

4.4 Identification

Under the parametric specification, the model is identified. To see this, consider the case where there are no covariates apart from n, and the econometrician observes the prices from repeatedly auctioning objects with values drawn from a common prior. The aim is to identify the three model primitives, which are the prior mean μ and standard deviation σ and the signal standard deviation r. Intuitively, for a fixed n I can observe two relevant moments of the log price distribution, its mean and variance; and then variation in n generates Winner's Curse effects that allow me to pin down the third primitive. Formally, the global identification condition is that

$$E[\log p | n, z_j, \theta_0] = E[\log p | n, z_j, \theta_1]$$
$$\Omega[\log p | n, z_j, \theta_0] = \Omega[\log p | n, z_j, \theta_1]$$

for all (n, z_j) implies $\theta_0 = \theta_1$. Numerical analysis on a wide range of parameter values $\theta_0 \neq \theta_1$ shows that if the moment predictions are equal for some n_0, then they diverge for all other n. This identifies the parametric model.

One would like, however, for the model to be nonparametrically identified. Whether this is the case is unclear. The signal distribution is not identified except up to a normalization of signals, since any monotone function of the signals defines a new signal distribution and induces the same bidding behavior (Laffont and Vuong 1996). Even with such a normalization, with no variation in the number of bidders n, the model remains unidentified (Athey and Haile 2002). But under the normalization that signals are unbiased, and with variation in the number of bidders, the model may be identified. Certainly, it can be shown that the conditional mean valuation μ is identified by information aggregation as the average price in a sequence of auctions

[23]Further computational details are to be found in the supplementary material.

as the number of bidders n tends towards infinity (Kremer 2002). It follows that with sufficient variation in the covariates z, the parameter vector α is also identified. Given the ambiguity, it seems important to test the goodness of fit of the parametric model. I do this by examining the out-of-sample performance of the model in predicting the third highest bid in the auction, in the subsection on goodness of fit in the results section below.

4.5 Discussion

Two different approaches to estimating common value auctions have been tried in recent papers. Bajari and Hortaçsu (2003) employ a Bayesian approach with normal priors and signals, and estimate the posterior distribution of the parameters by Markov Chain Monte Carlo (MCMC) methods. Hong and Shum (2002) use a quantile estimation approach with log normal priors and signals, and obtain parameter estimates by maximizing the resulting quantile objective function. This pseudo maximum estimation approach has a couple of advantages relative to these other types of estimators. The main advantage lies in the computational simplicity of my approach.

As shown above, the criterion function may be maximized through a nested loop, in which the inner loop runs weighted least squares (WLS) and the outer loop is a standard quasi-Newton maximization algorithm. For most applications, the outer loop will search over a low dimensional space, while the inner loop may be of almost arbitrarily high dimension since the maximizer may be computed analytically. This avoids the curse of dimensionality, and allows me to include over thirty regressors in the structural model while still maximizing the objective function in under an hour. By contrast, the existing estimation techniques would require an MCMC implementation in a thirty dimensional parameter space, and it would typically take days of computation to get convergence of the chain to the stationary distribution of the parameters.

A second advantage is that this estimation approach permits fairly standard classical approaches for dealing with endogeneity. In previous versions of this paper I showed that if the relationship between the endogenous regressor and the source of the endogeneity is monotone, a semiparametric control function approach can be used to control for endogeneity (Lewis 2007).

5 Results

The results of the auction model estimation follow in table 3.[24] In that table, I estimate the model both for the full sample and for a number of subgroups, including model fixed effects as covariates in every specification. In interpreting the results, it is helpful to remember the goal of the estimation procedure. The idea is to test for selective disclosure by looking for a positive relationship between the common valuation of the participants and the information measures. Since bidder valuations, unlike price, are fundamental and are not an outcome of strategic interactions, this tests whether the beliefs of the bidders about vehicle quality are on average positively influenced by information revealed by sellers. At the same time, estimation of the model allows me to quantify the effect of information measures on prior uncertainty, and thus on prices through the Winner's Curse.

5.1 Valuations and Information

Consider first the top part of table 3, which gives the estimate of the parameter vector α that describes the relationship between the prior mean valuation held by bidders and the covariates. In column (1) we see that the relationship between the text and photos and the prior mean is significant and of large magnitude. Each additional percentage of text implies a percentage increase of 0.0693% in the prior mean, or for a typical car in the dataset, nearly $7. Similarly, each additional photo is associated with an increase of 1.4% in the prior mean, or for a typical car, $140. Comparing the magnitude of these results to those in the hedonic regressions, we see that the effects are smaller. This is because the structural model is of the relationship between the prior mean and the information measures, and as such is free of the strategic interactions that may link price to information measures. From the results, it is clear that the prior mean valuation held by bidders is positively related to the number of photos and bytes of text. This is consistent with selective disclosure on the part of sellers, as if sellers only put up favorable text or photos, then those webpages with many photos or a lot of text will typically be those of "peaches" and the prior valuation will reflect that.

[24]In all cases, standard errors are obtained from estimates of the asymptotic variance-covariance matrix, supplemented by the delta method where necessary.

Table 3: Auction Model Results

	Full	Classic	Reliable	Pickups	Non-Dealer	Dealer
Prior Mean						
Age	-0.1894	-0.1709	-0.1468	-0.1641	-0.1915	-0.1856
	(0.0015)	(0.0018)	(0.0093)	(0.0028)	(0.0021)	(0.0022)
Age Squared	0.0040	0.0037	-0.0004	0.0026	0.0039	0.0040
	(0.0000)	(0.0000)	(0.0004)	(0.0001)	(0.0000)	(0.0000)
Log Miles	-0.1452	-0.1484	-0.1471	-0.1361	-0.1418	-0.1473
	(0.0039)	(0.0044)	(0.0152)	(0.0087)	(0.0056)	(0.0053)
Log Text Size	0.0693	0.0828	0.0371	0.0536	0.0731	0.0693
	(0.0044)	(0.0063)	(0.0079)	(0.0067)	(0.0063)	(0.0063)
Number of Photos	0.0140	0.0167	0.0051	0.0115	0.0138	0.0136
	(0.0005)	(0.0007)	(0.0014)	(0.0009)	(0.0007)	(0.0007)
Manual Transmission	0.0682	0.2299	0.0374	-0.1491	0.0631	0.0736
	(0.0068)	(0.0087)	(0.0136)	(0.0139)	(0.0094)	(0.0100)
Log Feedback	-0.0228	-0.0173	-0.0168	-0.0274	-0.0235	-0.0234
	(0.0018)	(0.0025)	(0.0033)	(0.0031)	(0.0024)	(0.0026)
% Negative Feedback	-0.0038	-0.0040	-0.0005	-0.0027	-0.0035	-0.0043
	(0.0009)	(0.0013)	(0.0013)	(0.0015)	(0.0012)	(0.0013)
Competition Index	-0.0036	0.0094	-0.0347	-0.0061	0.0008	-0.0106
	(0.0142)	(0.0244)	(0.0274)	(0.0168)	(0.0143)	(0.0252)
Prior Standard Deviation (index)						
Age	0.0377	0.0301	0.2089	0.0443	0.0376	0.0373
	(0.0013)	(0.0025)	(0.0212)	(0.0028)	(0.0018)	(0.0017)
Text	-0.0465	-0.0192	0.0003	-0.0768	-0.0203	-0.0610
	(0.0193)	(0.0238)	(0.0883)	(0.0305)	(0.0295)	(0.0262)
Photos	-0.0174	-0.0140	-0.0204	-0.0180	-0.0147	-0.0195
	(0.0026)	(0.0032)	(0.0210)	(0.0045)	(0.0035)	(0.0039)
Marginal Effects for Prior Standard Deviation						
Age	0.0119	0.0103	0.0185	0.0112	0.0117	0.0118
Text	-0.0146	-0.0066	0.0000	-0.0194	-0.0064	-0.0193
Photos	-0.0055	-0.0048	-0.0018	-0.0045	-0.0046	-0.0062
Mean Prior Deviation	0.7904	0.8023	0.5470	0.7324	0.7845	0.7939
Signal Standard Deviation	0.9943	0.8932	0.3965	0.9087	0.9946	0.9947
	(0.0103)	(0.0154)	(0.0472)	(0.0185)	(0.0161)	(0.0142)

Standard errors are in parentheses. Each column reports estimation results for a different subset of the data. This top part of the table reports the estimated coefficients relating to the relationship between the prior mean valuation μ and the auction covariates. The second part gives the coefficients for the prior standard deviation σ. The final part reports mean values and marginal effects of the covariates on σ, and the estimated value of the signal standard deviation r.

5.2 Winner's Curse and Information

The second half of the table on the following page contains the coefficients on σ and the constant bidder signal standard deviation r. Consider the full sample analysis in column (1), and note that all coefficients for σ are significant. To interpret the coefficients, I turn to the marginal effects. A extra year of age is associated with a 0.0377 increase in the standard deviation of the prior, as one would expect given the greater uncertainty about the condition of older cars. An additional percentage of text is associated with a small 0.0465% decrease in the deviation. Photos do far more to reduce uncertainty, with each additional photo associated with a 0.0174 decrease in the prior standard deviation. This is not surprising, as photos are "harder" information than text.

To quantify the magnitude of the change in the winner's curse effect induced by changes in information I need the estimated marginal effect of σ on price. From the moment computations, I find that on average an increase of one standard deviation in the prior uncertainty is associated with approximately a 10% decrease in price. Thus each additional photo, through decreasing the prior uncertainty and promoting aggressive bidding, is expected to lead to a 0.174% increase in price, or about \$17 for a typical car. This effect is substantial, though considerably smaller than the direct effect of the photo content in changing mean valuations.

The coefficient estimates, though consistent for the relationship between the prior moments and the covariates, will typically underestimate the full effect of information disclosure on prior uncertainty and subsequent bidding behavior. This is because as the econometrician I am using a weak quantitative measure of information - the number of photos and bytes of text - whereas bidders will actually observe the true information content. Bidders will learn more from the auction webpage than I can capture in my measures, and more information is being disclosed than is suggested by these results. It follows that the role of information disclosure in reducing information asymmetries and therefore the potential for adverse selection is in some sense "bounded below" by these results.

5.3 Variation across Models and Sellers

In the second, third and fourth columns of table 3, I examine how these effects vary for different models of car. I group cars into three groups, one consisting of "classic" cars such as the Mustang and Corvette, one consisting of "reliable cars" such as the Honda Accord and Civic, and one consisting solely of pickup trucks (such as the Ford F Series and GM Silverado). The vehicles in the "classic" group are considerably older on average than those in the other two groups, at an average of 24 years old versus 10 years for the reliable cars and 15 for the pickup trucks. A number of differences arise. The predicted relationship between mileage, age and price varies for the three groups. In the case of classic cars, mileage has a stronger negative coefficient than for other models. This may be because vintage cars in near mint condition command much higher premiums. I also find differences in the coefficient on transmission, which presumably reflects a preference for manual transmission when driving a Corvette or Mustang, but not when driving a pickup. In the case of "reliable" cars, the competition index coefficient is much larger than for the other groups, and though not significant, suggests that competition on eBay Motors plays more of a role in determining their prices. This makes sense since the reliable cars are more close substitutes for each other.

I find that the coefficients on the information measures vary considerably. For classic cars, where vehicle history, repair and condition are of great importance, the text size and photo coefficients are both large. On the other hand, for the reliable and largely standardized cars such as the Accord and Toyota, these effects are much less. The coefficients for pickups fall somewhere in between, perhaps because pickups have more options (long bed, engine size) for a given model, and more variation in wear-and–tear than do the reliable cars.

I also find that the prior standard deviation varies considerably among groups, even after accounting for age, text and photos. There is considerable prior uncertainty about the value of classic cars, and far less for reliable cars, which makes sense. As before, text and photos do decrease prior uncertainty, although text is only significant in the case of pickups, while photos are significant for both classic cars and pickups.

In the final two columns of 3, I consider private sellers and dealers separately. The coefficients are extremely similar across the two regressions, suggesting that the disclosure policies of private sellers and dealers are not all that different. One major difference occurs in the coefficients relating text and photos to σ. It appears that the text and photos of dealers are generally more informative, decreasing prior uncertainty by markedly more than in the case of private sellers. Given the greater experience of dealers in selling cars, this is not surprising.

5.4 Robustness to Number of Bidders

Identification of this model makes use of exogenous variation in the number of participating bidders. Yet as noted earlier, I do not observe the number of bidders who participate, but only those who are actually observed making bids. In this section, I examine the robustness of the results to the assumption that the number of participating bidders is the same as the number observed. In table 4 I present the results of three regressions on the full sample. In the first, I maintain the assumption that the participating number is equal to n, where n is the number of observed bidders. In columns 2 and 3, I instead assume that the participants are equal to $1.5n$ and $2n$ respectively. The results on the prior mean μ are almost identical across all specifications, with the coefficients on logtext and photos relatively stable and always significant and positive. This confirms that the selective disclosure finding is robust to alternate participation assumptions.

By contrast, the estimates of the relationship between information measures and the valuation standard deviation are less stable. In particular, the coefficients on text vary considerably, and in the last two specifications are not significant. Given that this was also the case for some of the subsets analysed in table 3, one should be concerned that the effect of text in reducing uncertainty is not robust to changes in participation assumptions. Photos continue to be significant in reducing uncertainty in all specifications, but the magnitude of their effect in reducing the Winner's Curse is less well identified, with estimates ranging from \$11 to \$17.

5.5 Robustness to Private Value Components

The other key assumption of the model was that it was pure common values. So suppose instead that the underlying model was an affiliated values model. It follows

Table 4: Robustness: Observed Bidders

	n	$1.5n$	$2n$
Prior Mean			
Age	-0.1894	-0.1888	-0.1883
	(0.0015)	(0.0016)	(0.0015)
Age Squared	0.0040	0.0040	0.0040
	(0.0000)	(0.0000)	(0.0000)
Log Miles	-0.1452	-0.1457	-0.1456
	(0.0039)	(0.0040)	(0.0040)
Log Text Size	0.0693	0.0716	0.0689
	(0.0044)	(0.0060)	(0.0039)
Number of Photos	0.0140	0.0146	0.0147
	(0.0005)	(0.0006)	(0.0005)
Manual Transmission	0.0682	0.0678	0.0700
	(0.0068)	(0.0069)	(0.0068)
Log Feedback	-0.0228	-0.0224	-0.0212
	(0.0018)	(0.0018)	(0.0017)
% Negative Feedback	-0.0038	-0.0038	-0.0036
	(0.0009)	(0.0009)	(0.0008)
Competition Index	-0.0036	-0.0029	-0.0008
	(0.0142)	(0.0138)	(0.0130)
Prior Standard Deviation (index)			
Age	0.0377	0.0348	0.0381
	(0.0013)	(0.0022)	(0.0014)
Text	-0.0465	-0.0599	-0.0304
	(0.0193)	(0.0424)	(0.0217)
Photos	-0.0174	-0.0116	-0.0142
	(0.0026)	(0.0045)	(0.0030)
Marginal Effects for Prior Standard Deviation			
Age	0.0119	0.0119	0.0142
Text	-0.0146	-0.0204	-0.0113
Photos	-0.0055	-0.0040	-0.0053
Mean Prior Deviation	0.7904	0.8116	0.8531
Signal Standard Deviation	0.9943	1.0013	1.1004
	(0.0103)	(0.0064)	(0.0031)

Standard errors are in parentheses. Each of the columns reports estimated coefficients under different assumptions on how the actual number of bidders relates to the observed number of bidders n. This top part of the table examines the relationship between the prior mean valuation μ and the auction covariates. The second part gives coefficients for the prior standard deviation σ. The final part reports mean values and marginal effects of the covariates on σ, and the estimated value of r.

that some of the underlying variation in prices is a result of idiosyncratic private value differences, rather than variation in the common value component of the valuations. But then the estimate of the prior variance term σ here is an overestimate, and consequently the Winner's Curse effects for an average car are also overestimated. Bidders are in fact less afraid of the Winner's Curse than my estimates would suggest, since their prior valuations are more precise, and thus the strategic role of information in affecting prices is smaller than estimated. This suggests that the non-strategic role of information in affecting bidder priors through seller disclosure is in fact underestimated. I conclude that the selective disclosure finding is robust to introducing private value components into the model.

5.6 Goodness of Fit

One might be concerned that with the restrictive parametric assumptions, the model fitted may bear little resemblance to the underlying data generating process. In this section, I examine the goodness of fit of the model. To do this, I take the estimated coefficients for each of the three groups of cars (classic, reliable, and pickups), and use the structural model to simulate prices and the third highest bids. I compare the simulated price and bid distributions to the observed price and 3rd highest bid distributions. Notice that the 3rd highest bids were not used in the estimation of the model, and hence the simulated bids constitute an out-of-sample prediction by the model.

The results are shown in figure 5, which presents kernel density estimates of the observed and simulated log bid distributions for each of the three car groups. The fit of the model is exceptionally good. For classic cars, the simulated log price distribution matches the observed price distribution well, although in the tails it is somewhat less good. The simulated 3rd highest bids also fit well, although in general they are a slight underestimate. For reliable cars, which are more homogenous, the model performs better, matching both the observed prices and 3rd highest bid distributions extremely closely. In the case of pickup trucks, the observed price and bid distributions have an unusual and left-skewed shape. The model is able to match this to some extent, producing a simulated price distribution with a flat density function, but ultimately cannot fully account for the observed price distribution. These results should limit

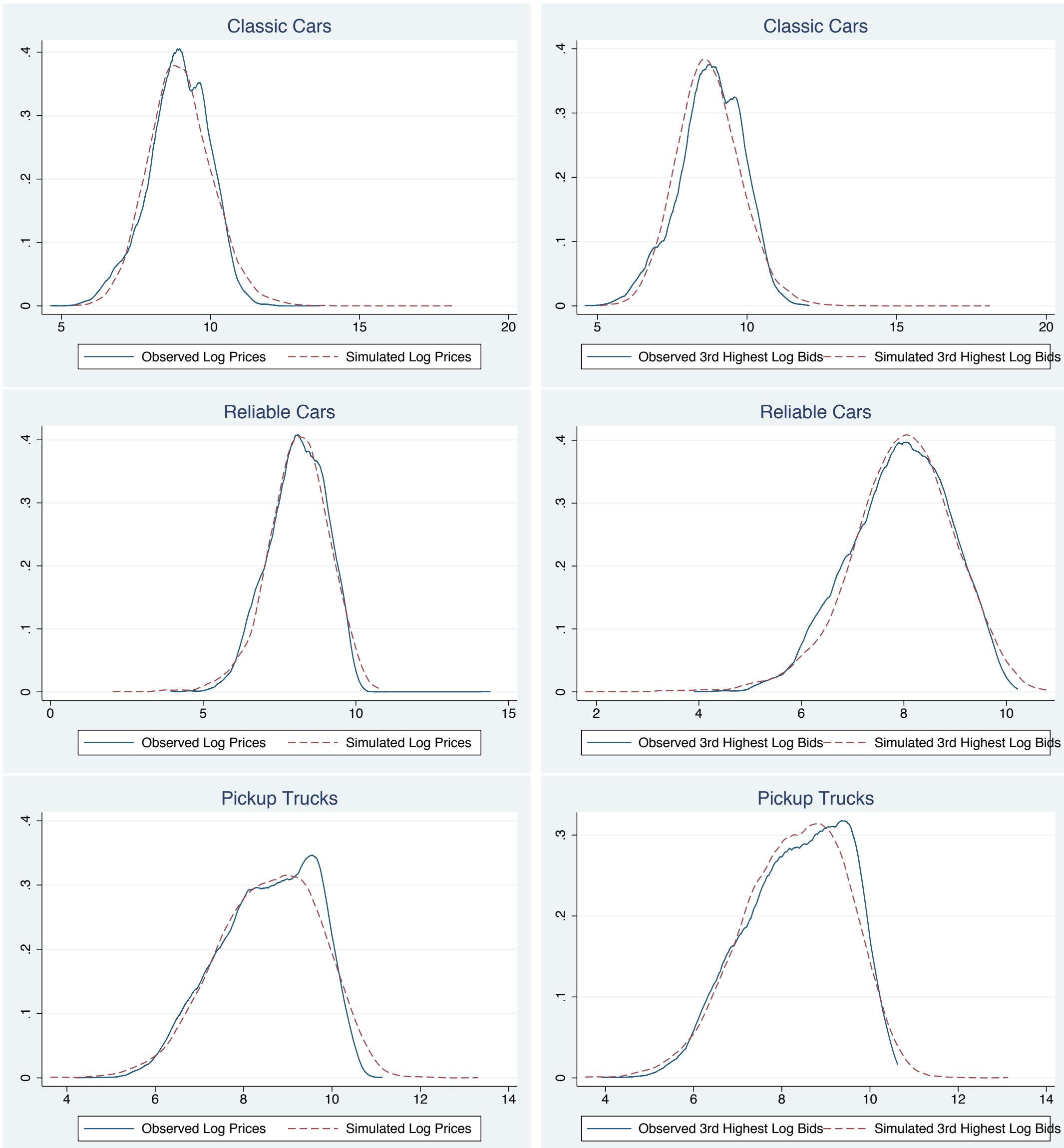

Figure 5: **Goodness of Fit** This figure presents a comparison of observed bids with those simulated under the estimated parametric model. The left column presents the observed price distribution and the simulated price distribution for each of the three groups of cars, classic, reliable and pickups. The right column presents the observed 3rd highest bids and simulated 3rd highest bids for each group of cars. The latter is an out of sample prediction.

concerns about the use of a fully parametric estimation procedure. The parametric model is able to simulate both observed prices and 3rd highest bids quite effectively, even though the latter are not used in estimation.

Overall then, the main results are as follows. First, there is a significant and large positive relationship between prior valuations and text and photos, in all specifications. Given that this cannot be generated by strategic considerations, this indicates that sellers are selectively disclosing information and bidders update their priors on the car's value accordingly. Second, text and photos play a role in decreasing prior uncertainty, due to their role in decreasing information asymmetry. Photos seem to have a far stronger effect, consistent with the fact that they are "hard" information. It follows that there is a role for information in decreasing the winner's curse effect, but the magnitude of this effect is considerably smaller than the direct effect through disclosure and prior updating.

6 Simulation: Coarse Disclosure

In this section I return to the idea that motivated the paper, and quantify the magnitude by which seller disclosures through the auction webpage help to reduce the potential for adverse selection. My tool in this endeavor is a stark counterfactual: a world in which eBay stops sellers from posting additional text and photos on the auction webpage, so that only basic and standardized car characteristic data is publicly available. This is hardly an idle comparison. Local classified advertisements typically provide only basic information, while even some professional websites do not allow sellers to post their own photos.[25] In the case of local purchases, of course, potential buyers will view the cars in person and so their private signals will be much more precise. This is typically not the case on eBay Motors. I will show that even when private signals reflect the true value of the vehicle, the less precise prior valuations held by bidders in the counterfactual world will lead them to systematically underbid on high quality vehicles and overbid on "lemons". Given that sellers have many other channels through which to sell their vehicles, those with high quality cars may select

[25]The South African version of Autotrader, autotrader.co.za, is such an example. Even on large online websites such as cars.com and autobytel.com, there are many listings with only a single stock photo.

out of the eBay Motors market. Relative to the counterfactual then, seller disclosures through the auction webpage help to limit differences between prices and valuations, and reduce adverse selection.

I maintain a number of important assumptions throughout the counterfactual. First, I assume that adverse selection does not occur, so that the quality distribution of cars sold in the counterfactual is the same as that observed in the market. This is a necessary assumption, since I have no data that would allow me to model the listing decisions of sellers. The goal of this exercise is to obtain estimates of the prices that would obtain under the counterfactual and thus judge the probability that adverse selection would occur in the absence of seller disclosures on the auction webpage. The other maintained assumption is that sellers would not adjust their offline information disclosure policies under the counterfactual. One could easily imagine that if eBay did not allow photos and text, sellers would make more of an effort to interact with bidders privately through e-mail and phone calls, so that the private signals held by bidders would become more precise. I do not allow for this. It should be noted though that the two forms of information are not perfect substitutes, since public disclosures on the webpage are enforceable, in the sense that the buyer must get what the seller advertised or can revoke the contract. This is not true of private disclosures. So with more precise private signals, the magnitudes of the results would change, but the general implications probably would not.

Formally, let the data generating process be as specified in the structural model presented earlier. Let I be the vector of information measures (photos and text), and let z' be a vector of standardized characteristics. Recall that values are log normally distributed, so that $\tilde{v} = \log\, v$ has normal distribution with mean $\mu = E[\tilde{v}|I, z']$ and standard deviation $\sigma = \sqrt{Var[\tilde{v}|I, z']}$. Bidders independently observe unbiased and normally distributed private signals of the log value $\tilde{v}$, with signal standard deviation r. Now consider two regimes. In the true regime, bidders observe the auction webpage and form a prior valuation based on the standardized characteristic data z' and the amount of information I. Their prior is correct in equilibrium, and thus parameterized by μ and σ. In the counterfactual regime, the auction webpage contains only the standardized data and so bidders have a less informative prior. This prior is log normal, and parameterized by $\mu^c = E[\tilde{v}|z']$ and $\sigma^c = \sqrt{Var[\tilde{v}|z']}$. So for example, someone selling a "peach" with extensive photos and text to prove it would

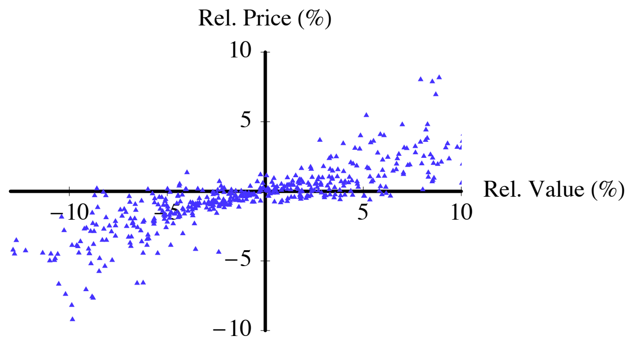

Figure 6: **Simulation Results** The figure plots relative value, defined as the ratio of predicted values to those predicted under the counterfactual $E[V|I,z']/E[V|z']$, against relative prices, defined as the ratio of predicted prices to those predicted under the counterfactual $E[p|I,z']/E[p^c|I,z']$. Cars with a high relative value are "peaches" while those with low relative value are "lemons". The observed positive relationship between relative prices and relative values indicates "peaches" earn a price premium when sellers can disclose information, and "lemons" are similarly penalized.

have $\mu > \mu^c$ and $\sigma < \sigma^c$. Bidders still obtain independent and unbiased private signals in both regimes.

I simulate the bidding function and compute expected prices under both regimes for a random sample of 500 car auctions in my dataset.[26] The results are graphically depicted in Figure 6. On the x-axis, I have the relative value of the car, which I define as the percentage by which the informed prior valuation $E[V|I,z']$ exceeds the uninformed prior $E[V|z']$. A simple characterization of this is that those cars with relative value greater than zero are "peaches", while those with relative value less than zero are "lemons". On the y-axis, I have the relative price of the car, which I define as the percentage by which the expected price under the current regime $E[p|I,z']$ exceeds the expected counterfactual price $E[p^c|I,z']$. As you would expect, there is a general upward slope, with "peaches" expected to fetch higher prices in the current regime in which sellers can publicly disclose information to potential buyers, than in the counterfactual in which such disclosures are impossible. On average, bidders would underbid on "peaches" and overbid on "lemons".

To get a sense of the magnitude of the results, I fit a linear model of the relationship between relative value and relative price. The estimated slope is 0.357, indicating that the expected price of a peach worth 10% more than the average car of its char-

[26]Computational details about the counterfactual simulation are given in the supplementary material.

acteristics is 3.57% higher with disclosure. For a typical car in the dataset the final transaction price is $10000, indicating that the premium to selling a peach with surplus value of 10% when disclosure is possible is around $357. The implication is that in the counterfactual sellers of high quality vehicles will realize significantly less revenue than they currently do, and this may induce them to choose a different selling channel or even consider a trade-in at a dealer. Certainly some sellers of high quality vehicles would choose not to sell on eBay Motors, and adverse selection problems would increase. In fact, since bidders in fact learn more from the auction webpage than I capture in my coarse information measures, the simulation underestimates the difference between the counterfactual and the current regime.

These simulation results make clear the role that seller disclosures can play in reducing adverse selection in this market. Whenever sellers with high quality cars are unable to extract a commensurate price premium from buyers because of information asymmetries, there is the potential for adverse selection. In the counterfactual regime without disclosures, the gap between relative value and relative price is around 35%, and this potential is high. In reducing this gap, seller disclosures reduce the potential for adverse selection significantly.

7 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. One avenue for future research is to examine the barriers to full unravelling, and determine whether disclosure costs, the coarseness of the disclosure technology or some other mechanism prevents sellers from fully revealing their private information. Another interesting line of analysis 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, C. P., L. Hosken, and P. Newberry (2006): "'Vettes and Lemons on EBay," Unpublished.

Akerlof, G. A. (1970): "The Market for Lemons: Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Athey, S., and P. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107–2140.

Bajari, P., and A. Hortaçsu (2003): "The winner's curse, reserve prices and endogenous entry: empirical insights from eBay auctions," *Rand Journal of Economics*, 34(2), 329–355.

Donald, S. G., and H. J. Paarsch (1993): "Piecewise Pseudo Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34(1), 121–148.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52(3), 681–700.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosures," *Journal of Law and Economics*, 24(3), 461–483.

Haile, P., H. Hong, and M. Shum (2003): "Nonparametric tests for common values at first-price sealed bid auctions," NBER Working Paper 10105.

Hong, H., and M. Shum (2002): "Increasing Competition and the Winner's Curse: Evidence from Procurement," *Review of Economic Studies*, 69, 871–898.

Houser, D., and J. Wooders (2006): "Reputation in Auctions: Theory and Evidence from eBay," *Journal of Economics and Management Strategy*, 15(2), 353–370.

Jin, G., and A. Kato (2005): "Price, Quality and Reputation: Evidence from an Online Field Experiment," Forthcoming in The Rand Journal of Economics.

Jin, G. Z. (2005): "Competition and Disclosure Incentives: An Empirical Study of HMOs," *Rand Journal of Economics*, 36(1), 93–112.

Jin, G. Z., and P. Leslie (2003): "The Effect of Information on Product Quality: Evidence From Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics*, 118(2), 409–451.

Jovanovic, B. (1982): "Truthful Disclosure of Information," *Bell Journal of Economics*, 13(1), 36–44.

Kremer, I. (2002): "Information Aggregation in Common Value Auctions," *Econometrica*, 70(4), 1675–1682.

Laffont, J.-J., and Q. Vuong (1996): "Strutural Analysis of Auction Data," *American Economic Review*, 86(2), 414–420, Papers and Proceedings of the 108th Meeting of the American Economic Association.

Levin, D., and J. L. Smith (1994): "Equilibrium in Auctions with Entry," *American Economic Review*, 84(3), 585–99.

Lewis, G. (2007): "Essays on Choice and Matching under Information Asymmetry," Ph.D. thesis, University of Michigan.

Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2), 380–391.

Resnick, P., and R. Zeckhauser (2002): *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System*vol. 11 of *Advances in Applied Microeconomics*, chap. 5. Elsevier Science.

Shin, H. (1994): "News Management and the Value of Firms," *Rand Journal of Economics*, 25(1), 58–71.

Song, U. (2004): "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders," Working Paper, University of British Columbia.

Wilson, R. (1977): "A Bidding Model of Perfect Competition," *Review of Economic Studies*, 4, 511–518.

Yin, P.-L. (2006): "Information Dispersion and Auction Prices," Harvard Business School Working Paper, No. 06-011.

8 Appendix

A. Proof of Proposition 1

By definition, $v(x, x; n, \mu, \sigma, r) = E[V|X_i = x, Y = x, N = n]$. Condition on all other signals, so that we have:

$$
\begin{aligned}
v(x, x; n, \mu, \sigma, r) &= E_{X_{-i}|Y<x}\left[V|X_i = x, X_{-i} = x_{-i}, N = n\right] \\
&= E_{X_{-i}|Y<x}\left[\exp\left\{\frac{r\mu + \sigma\sum_{i=1}^{n} x_i}{r + n\sigma} + \frac{1}{2}\left(\frac{r\sigma}{r + n\sigma}\right)^2\right\}\middle| X_i = x, X_{-i} = x_{-i}, N = n\right] \\
&= E_{X_{-i}|Y<x}\left[\exp\left\{a + \frac{r(\mu - a) + \sigma\sum_{i=1}^{n}(x_i - a)}{r + n\sigma} + \frac{1}{2}\left(\frac{r\sigma}{r + n\sigma}\right)^2\right\}\middle| X_i = x, X_{-i} = x_{-i}, N = n\right] \\
&= e^a E_{X_{-i}|Y<x}\left[\exp\left\{\frac{r(\mu - a) + \sigma\sum_{i=1}^{n}(x_i - a)}{r + n\sigma} + \frac{1}{2}\left(\frac{r\sigma}{r + n\sigma}\right)^2\right\}\middle| X_i = x, X_{-i} = x_{-i}, N = n\right] \\
&= e^a v(x - a, x - a; n, \mu - a, \sigma, r)
\end{aligned}
$$

where the second line follows from the posterior mean and variance of log v given normal unbiased signals $\{x_i\}$, and the fourth line follows on noting that e^a is a constant and thus can be taken outside of the expectation. Taking logs on both sides, we get $\log v(x, x; n, \mu, \sigma, r) = a + \log v(x - a, x - a; n, \mu - a, \sigma, r)$.

To get equation (4), note that $E[\log p(n, \mu, \sigma, r)] = E[\log v(x^{(n-1:n)}, x^{(n-1:n)}, n, \mu, \sigma, r)]$. It follows that $E[\log v(x^{(n-1:n)}, x^{(n-1:n)}, n, \mu, \sigma, r)] = a + E[\log v(x^{(n-1:n)}, x^{(n-1:n)}; n, \mu - a, \sigma, r)]$ where the signals in the RHS expectation have the same conditional distribution $G|v$ as those on the LHS, but the latent v now has a prior with mean $\mu - a$ on the RHS. Now take $a = \alpha z_j$ for any observation j. Then by specification $\mu - a = 0$, and

$$E[\log p(n, z_j, \alpha, \beta, r)] = \alpha z_j + E[\log p(n, z_j, 0, \sigma, r)]$$

The variance result follows in similar fashion.

Table 5: Effect of Information on Entry

	Log Price
Log of Miles	0.0342
	(0.0152)
Log Text Size (in bytes)	0.1078
	(0.0247)
Number of Photos	0.0426
	(0.0029)
Log Feedback	-0.0040
	(0.0114)
Manual Transmission	-0.0024
	(0.0435)
Percentage Negative Feedback	0.0020
	(0.0057)
Ends Sunday	0.1707
	(0.0462)
Ends afternoon/evening	0.3315
	(0.0386)
Model Fixed Effects	yes
Year Fixed Effects	yes

Estimated standard errors are given in parentheses. The results show that although there is a statistically significant positive relationship between the observed number of bidders and the information measures, the estimated coefficient is small in magnitude, and thus is not a satisfactory explanation of the observed link between price and information measures.

Table 6: Hedonic Regressions - Dealer Status

	Log Price		
	Non-Dealers	Dealers	Dealer Fixed Effects
Log Miles	-0.1302	-0.1315	-0.0798
	(0.0036)	(0.0035)	(0.0056)
Log Text Size (in bytes)	0.0843	0.0835	0.1786
	(0.0061)	(0.0056)	(0.0168)
Number of Photos	0.0162	0.0168	0.02644
	(0.0007)	(0.0007)	(0.0018)
Manual Transmission	0.0717	0.1114	0.0888
	(0.0099)	(0.0104)	(0.01450)
Log Feedback	-0.0230	-0.0232	-0.0955
	(0.0026)	(0.0028)	(0.0238)
Percentage Negative Feedback	-0.0020	-0.0055	0.0076
	(0.0013)	(0.0014)	(0.04912)
Competition Index	-0.0102	-0.0106	0.0121
	(0.0172)	(0.0196)	(0.0230)
Model Fixed Effects	yes	yes	yes
Year Fixed Effects	yes	yes	yes
Number of Bidders Fixed Effects	yes	yes	yes
Seller Fixed Effects	no	no	yes

Estimated standard errors are given in parentheses. The results in the first two columns show that the estimated relationship between price and the information measures is relatively similar across dealers and non-dealers. The specification in the final column uses only the sub-sample of dealers who list at least four cars, and includes seller fixed effects. The results show that controlling for seller identity, variation in information is positively associated with variation in price.

Table 7: Hedonic Regressions - Book Value

	Log Price	
	Book Value	Fixed Effects
Log Book Value	0.8497	
	(0.0083)	
Log Text Size (in bytes)	0.0369	0.0464
	(0.0043)	(0.0041)
Number of Photos	0.0102	0.0103
	(0.0005)	(0.0005)
Manual Transmission	0.1362	0.0594
	(0.0076)	(0.0080)
Log Feedback	-0.0224	-0.0206
	(0.0021)	(0.0020)
Percentage Negative Feedback	-0.0031	-0.0041
	(0.0010)	(0.0010)
Competition Index	-0.0599	-0.0046
	(0.0107)	(0.0111)
Model Fixed Effects	no	yes
Year Fixed Effects	no	yes
Number of Bidders Fixed Effects	yes	yes

Estimated standard errors are given in parentheses. This is estimated on a sub-sample of 22331 data points of cars with model-year of 1990 or later. The results show that the estimated relationship between price and the information measures remains strong and positive even after controlling for book value, although the coefficient on the text measure is smaller with the book value controls.