A Survey-Based Procedure for Measuring Uncertainty or Heterogeneous Preferences in Markets

Pai-Ling Yin
Harvard Business School

April 22, 2005 (1st draft: November 25, 2002)

Abstract

This paper shows how surveys can be used to generate a measure of the amount of information and/or heterogeneity of preferences within a market. This measure can be employed as a regressor in empirical work where variance in the dependent variable (e.g., auction prices, retail price dispersion, or investment choices in stocks, R&D, or education) might be explained by uncertainty about the value of the item being sold or the returns to investment choice and/or heterogeneous preferences in the market. The effects of incomplete information and heterogeneous preferences are usually relegated to the error term, which a) confounds these effects with other drivers of the error term and b) could lead to heteroskedasticity at best or omitted variable bias at worst. Furthermore, by specifically modeling the effect of this uncertainty or dispersed taste, one can estimate policy implications such as the effect of publicly introducing information into the market or selecting the pool of agents to change the distribution of preferences.

I demonstrate the validity and usefulness of my survey-based procedure by using it to measure the mean and dispersion of private information signals in eBay online auctions for personal computers. I exploit a mixture of respondents with and without experience on eBay. The use of inexperienced respondents permits the survey to be implemented more quickly and with a larger number of respondents than if it were restricted to experienced respondents. The use of experienced respondents allows me to correct for potential bias from using more noisy, inexperienced responses.

[0]I would like to thank Pat Bajari; Paul Hartke for scripting; Tiffany Chow, Yi-Xuan Huynh, Steve Yuan and the many people who filled out my survey for their research assistance; and The John M. Olin Foundation through the Stanford Institute of Economic Policy Research and Stanford School of Humanities and Sciences Graduate Research Opportunities Grant for their financial support. Any mistakes in the paper are completely my own; comments and suggestions are very welcome. pyin@hbs.edu

1 Introduction

Traditionally, surveys have been used to elicit unobservable information about people's valuations of goods when markets and prices for those goods are absent. They can also be a valuable source of information when markets exist. People outside a market can assess descriptions about the items being sold in that market, and this paper shows how surveys can be used to exploit their assessments to generate a measure of the amount of information and/or heterogeneity of preferences within that market. Specifically, in markets where participants receive different signals about an item's value due to noise and/or due to different costs and preferences, the survey can be used to estimate the characteristics (mean, variance) of the distribution of those signals.

This measure can be employed as a regressor in empirical work where variance in the dependent variable (e.g., auction prices, retail price dispersion, or investment choices in stocks, R&D, or education) might be explained by uncertainty about the value of the item being sold or the returns to investment choice and/or heterogeneous preferences in the market. The effects of incomplete information and heterogeneous preferences are usually relegated to the error term, which a) confounds these effects with other drivers of the error term and b) could lead to heteroskedasticity at best or omitted variable bias at worst. Furthermore, by specifically modeling the effect of this uncertainty or dispersed taste, one can estimate policy implications such as the effect of publicly introducing information into the market or selecting the pool of agents to change the distribution of preferences (Yin 2005).

Consider an empirical setting that examines the difference between prices and choices over a variety of goods. Examples include the dispersion of retail prices over different types of drugs (Sorensen 2000), the choice (or level) of investment in different types of financial instruments, the R&D contributions in different industries, the choice to attain different levels of education, or the prices of different items being auctioned. Survey responses could be used to produce a measure of the dispersion of information regarding the efficacy of the drug or the returns to different types of investments or levels of education. In the auction setting, including the variance of survey responses in a price regression allows one to control for the dispersion of information in the market if the item being auctioned has a common value or control for the dispersion of preferences in the market if the item being auctioned has a private value. In all these examples, the regression model is misspecified if one fails to account for differing preferences or information. At best, this leads to heteroskedasticity, since the different goods all have different error distributions arising from variance in information or preferences. At worst, this omitted variable will bias coefficient estimates of the other regressors if those regressors are correlated with the realization of preferences or information. In an auction setting, survey data has the added benefit of being measured independently from data generated during the auction itself. By definition, data from the auction is a function of bidder behavior. This makes the external survey data useful for testing hypotheses about bidding behavior in the auction that otherwise could not be conducted without making assumptions about the nature of the private information signals.

Researchers often avoid using surveys due to the time and effort involved in conducting them. However, use of online surveys reduces some of the cost. This paper suggests a survey design technique and econometric tool to deal with a general population of survey respondents, including those who participate in the market of interest and those who do not. The use of inexperienced respondents permits the survey to be implemented more quickly and with a larger number of respondents than if the researcher had to restrict her search for an equivalent number of experienced respondents. The use of experienced respondents allows me to correct for potential bias from using more noisy, inexperienced responses.

To the extent that the survey is still more costly to conduct than gathering observable data, this paper argues that the survey data is more valuable because it exploits the human ability to assess complex information sets in a way that cannot be accomplished by hedonic evaluation. Often, hedonics are used to control for the value of the good. However, hedonic methods suffer from the need to define a good into a limited set of characteristics, and it does not provide any means for taking into account anomalies in products that may not fit any category. Survey data exploits human assessment of information to collapse many dimensions into a single numerical value. This does not preclude the econometrician from also employing hedonic measures along side the survey data.

The particular application used here is for eBay online auctions for personal computers (PCs). In all auctions, private information signals (not directly observable to the econometrician) about the value of the item being sold is dispersed among the auction participants. I used a survey to measure the mean and dispersion of those information signals in computer auctions.

In a common values (CV) auction setting, each auction participant's private signal contains information that is relevant to the other participants' assessments of the value of the item. In this setting, the average of these survey responses provides a potential measure of the common value of the item being auctioned. The standard deviation of responses provides a potential measure of the dispersion of information in the auction. An auction where more information is publicly available to all the bidders will be reflected in less dispersed signals.

In a private values (PV) setting, a private signal only concerns the recipient's own value for the item. In this case, the survey measures the average private value and dispersion of private values among bidders. One can use these averages and standard deviations to test between PV and CV settings (or the dominant component if the setting is mixed) while also testing for Bayesian-Nash equilibrium bidding behavior (see Yin 2005).

Analysis of the survey results confirms that the survey is able to successfully generate estimates of information dispersion and average item values. Auction descriptions matched results generated in the survey. Auctions which my survey respondents designated to be of equal value contained equivalent hardware specifications. The auction description that provided more details (i.e., revealed more information to all auction participants) had a lower standard deviation of survey respondent's valuations. The price attained in that auction was higher than that attained for the item with a less informative auction description. This last finding is consistent with the auction theory prediction that prices decline with more information dispersion in CV settings.

I collect background data on the survey respondents during the survey to determine which respondents are experienced with eBay computer auctions (and thus similar to the auction participants) and which respondents are inexperienced. Because survey measures are prone to bias, I exploit a mixture of respondents, some experienced with eBay computer auctions and some not, to correct for any bias between the mean and standard deviation of the survey responses and the true common value and dispersion of information facing the auction participants. The use of inexperienced respondents increases the pool of potential survey respondents and permits the survey to be implemented quickly. The use of experienced respondents allows me to correct for potential bias from using more noisy inexperienced responses.

Section 2 reviews the motivation for a survey based measure of information dispersion in auctions. Section 3 presents the auction data employed and the survey design. Section 4 analyzes the success of the survey as a correlated measure. Section 5 presents the background data collected in the survey and its implications for correcting for survey bias. Section 6 presents the bias correction procedure. Section 7 examines the difference between results from the survey-based measures and alternative hedonic regression methods. Section 8 concludes this paper.

2 Motivation for Survey Data

There are several reasons why a researcher might want to collect survey data to augment data from commercial markets. A researcher must often control for the value of the item when determining the effect of other regressors on price. Empirical work in general has employed hedonic regression of price on product characteristics to control for the value of the item. A large number of hedonic characteristics will demand a large number of observations for identification. A survey allows respondents to flexibly assess the value of a large number of characteristics even in a small sample of items. Alternatively, empirical work has restricted itself to examining identical items to control for item values. However, using identical items may result in a sample that is either too small or exhibits too little variation in the regressors of interest. Survey respondents can handle differences in item characteristics, allowing the researcher to include more heterogeneous items in order to ensure a sufficiently large sample and sufficient variation in the regressors of interest. The survey measure of value can be constructed to be independent of the price, as long as survey respondents are not shown price information. Thus, the survey data provides an exogenous regressor that controls for the value of the item.

The survey measures can be designed to be independent of other regressors as well. For example, in my survey design, details about the seller, the bidders, and the bids in the auction are omitted. This creates several advantages for using survey data over methods that recover the signals from the observed distribution of bids. My survey responses are functions of the product description only. Reserve prices and opening bids that appear in many online auctions would truncate the observed distribution of bids. My survey responses are not influenced by the number of bidders in the auction nor by the reputation of the seller. By construction, the survey data is not a function of bidding behavior in the actual auction. The independence of my survey data from the auction data allows me to test between different types of bidding behavior and separately identify the effect of reputation from the dispersion of information or preferences and from other determinants of price.

Empirical work has also proxied for the common value using blue book values. However, blue book values and hedonic methods cannot take into account any anomalies in the products. For example, a computer that was being sold on eBay was described as working but locked: the password had been lost, so there was no way to logon to the computer. Hedonic estimation or the use of a blue book value would treat this anomaly as unobservable to the econometrician, but such anomalies may be important determinants of the price of the auctioned item. They might drive the number of bidders that enter the auction. The number of bidders is often included in the price model as a regressor. This presents an endogeneity problem for estimation, since the number of bidders is now correlated with the error term. In contrast, the human readers' estimates do reflect values that are more closely tied to the semantics of the product description than any hedonics-based measure or book value. By having people read the auction descriptions and respond with their value for the item, I am able to capture the idiosyncrasies of each item as well as its hedonic characteristics.

Variation in the survey responses also generates information that does not exist in one-dimensional measures from hedonic analysis or book values. The standard deviations over the responses in each auction serve as a measures of the dispersion of private information signals in the auctions. They provide a measure of the survey respondents' certainty about their valuations.

This extra information about the unobservable private signals is particularly useful in testing auction theory. Often, the only information available from auctions is the number of bidders, observed bids, and product characteristics. In a limited number of cases, ex post values of the auctioned item are available. The literature on nonparametric identification has shown that given this observable data, the distribution of private signals is just identified assuming a private values setting but underidentified in a common values setting without further parametric assumptions.[1] As a result, tests of information structure in an auction (whether auctions are private value or common value) and bidding behavior (whether or not bidders play Nash equilibrium strategies) are rarely conducted jointly. By measuring dispersion, identifying power is not expended on recovering the underlying distribution of information signals, so a joint test of an auction's information structure and bidding behavior is possible.

In sum, the ability to design the independence of survey data from other regressors, the ability to exploit human assessment of information, and information provided by the second moment of survey data make it an appealing source of information to complement market data, in particular for auctions.

[1] Laffont and Vuong 1996; Li, Perrigne, and Voung 2000; Guerre Perrigne, and Vuong 2000; Athey and Haile 2002; Li Perrigne and Vuong 2003.

3 Survey Design

Over 5000 new and used computers are listed daily in the eBay PC desktop category by both individuals and businesses. Auction participants may perceive these computers as a mixture of common and private values. To obtain an estimate of the mean and dispersion of private signals received by the auction participants, I created a web-based survey. The survey encouraged survey respondents to focus on the CV component, since other evidence in this market suggested that the CV component dominates the PV component in this market (see Yin 2005).

The content for the survey came from the auction descriptions for 222 eBay PC auctions held between June 24 and July 12, 2002. eBay works as follows: A seller posts an item for auction. She creates a product description through text and pictures and any other media that can be displayed on the eBay website. Auction participants can observe this auction description along with other information about the seller, the current price, and the number of bids submitted up to that point in the auction.

Anyone could respond to my survey, except for the actual bidders in my sample of auctions. The survey was distributed to acquaintances by word of mouth. I asked people to read computer auction descriptions and then answer the following question: "If a friend wanted to buy the computer described below, what is the most she should pay for it?" (see Appendix A) My descriptions contained only the information provided by the seller in the "descriptions" section. Information listed by eBay about the bids, reservation values, number of bidders, and the seller's identity and reputation were removed. I also collected background data on survey respondents, asking them about their experience working with computers, purchasing computers, and buying computers in online auctions.

I proposed incentives to encourage survey respondents to think about the computers' values and to consider the CV component of the computers. Respondents viewing the "prize details" webpage were told that they would receive an prize of $60 for being closest to all the other valuations or for being closest to a panel of computer experts. The advantage of these incentives is that they focus participant attention on the CV component rather than on idiosyncratic private tastes. I expected respondents who were unfamiliar with computer values to gravitate towards the first incentive and try to think of the CV component that other non-experts perceived. I expected respondents with more familiarity with computers to think about the CV component they shared with other experts. The incentive also provided some incentive to discourage respondents from merely typing in random numbers. The disadvantage of these incentives is that they may have introduced some bias that would be correlated with value of the item. However, I propose a correction method for this bias in Section 6. Only a small percentage (6%) of respondents actually looked at the webpage on "prize details," so these issues only affected a small number of my respondents.

The survey was designed to employ as wide a pool of respondents as possible. This made implementation of the survey easier and faster. However, one must account for the potential bias from using survey data to estimate information possessed by the auction participants. Fortunately, one can collect background data on the respondents during the survey. This information can then been exploited in order to identify the types of adjustments that need to be made to correct for that bias.

3.1 Issues in Survey Methodology

The literature on the contingent valuation survey method, where people are asked to state their willingness-to-pay for a good, relates most closely to my survey method. Several critiques have been made about the validity of the contingent valuation method (c.f. Hausman 1993) as well as of the entire field of survey data. However, in several ways my survey either escapes those critiques or employs methods of responding to those critiques which were suggested in the literature.

One problem with contingent valuation surveys is that respondents must often estimate the value of a vaguely defined item for which they have no previous market or pricing experience (e.g., "How much do you value clean water?"). In my case, this criticism is not as relevant; the presence of a retail market helps to create realistic bounds for my participants' valuations. Empirical studies comparing contingent valuation surveys to actual revealed preference data show that the estimates correspond very closely to the market prices. (Bjornstad & Kahn 1996) In addition, my survey respondents see everything that the bidder sees in that particular auction, so my survey reflects the appropriate informational context. The literature often refers to the differential effects that starting points can make in valuations. (Aadland & Caplan 1999; Bateman & Willis 1999) However,both the actual auction participants and my respondents would be influenced by the same types of anchor prices appearing in auction descriptions or in the retail market.

Another criticism of survey data is the lack of realistic incentives. Respondents' valuations may be inflated because they are not dealing with their own money, and have no incentive to be conservative. On the other hand, their valuations may be deflated since they have no incentive to think carefully about their true maximum willingness-to-pay. As a result, respondents may have different dispersion of valuations than the actual bidders. This is an important criticism, and I suggest a procedure for correcting for such potential differences between bidders and respondents in Section 6.

4 Survey Performance

On average, I collected 46 responses per auction. For each respondent i and for each auction t, I denote the respondent's valuation of the item by $X_{i,t}$. I denote the average of the responses in each auction by V_t. I denote the standard deviation of responses in each auction by SD_t. Summary statistics are presented in Table 1.

In Figure 1, auctions are ordered along the horizontal axis by increasing eBay price, P_t. The corresponding averages of survey responses, V_t, are plotted for each auction as well. The plot shows that V_t is highly correlated with P_t. If prices are correlated with the value of the item, then this plot suggests that V_t is likely to be correlated with the value of the

Table 1: Summary statistics for survey on 222 auctions

Variable (831 respondents)	**mean**	**st. dev.**	**min**	**max**
no. of responses/auction	46	6	25	65
average: V_t	\$666.43	\$317.28	\$101.48	\$1,816.98
standard deviation: SD_t	472.38	153.94	163.57	980.50

item. The correlation between P_t and V_t suggests that survey data can be used to measure unobservable item values.

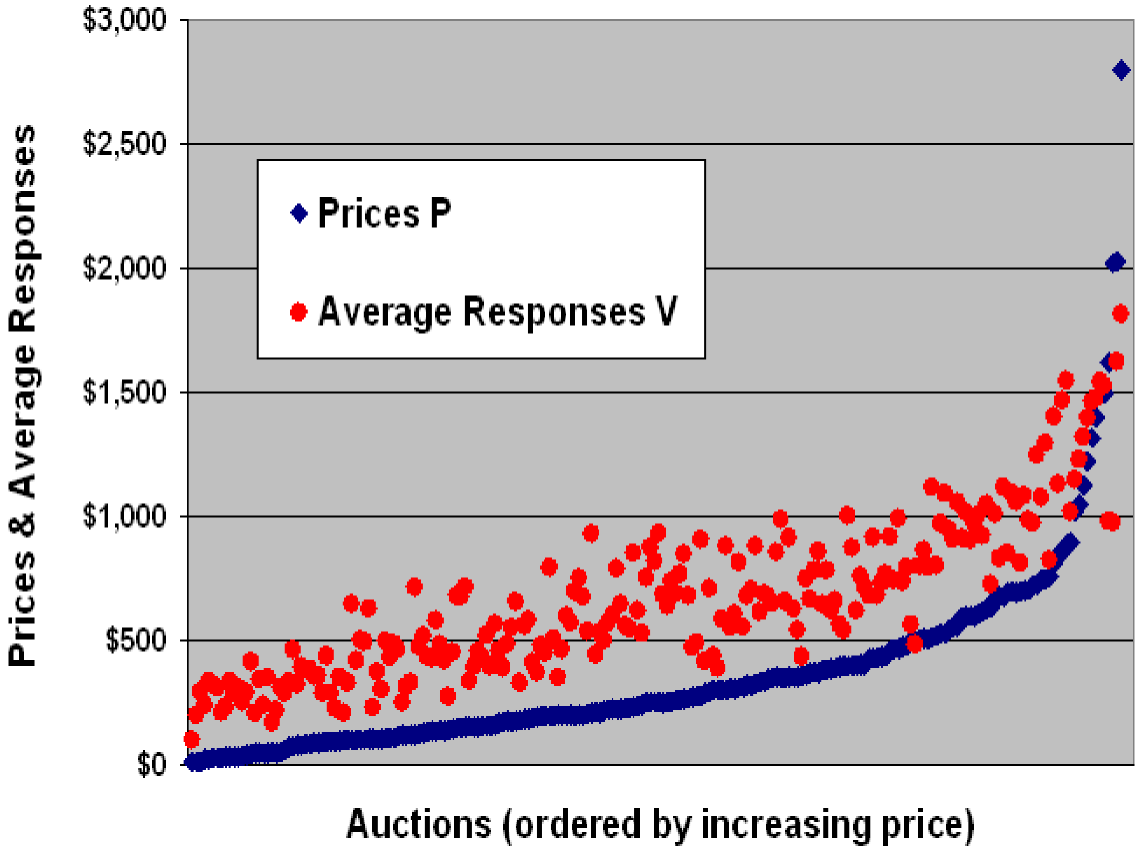
Possession of more or better information should lead auction participants to be more certain about the value of the item, decreasing the standard deviation in their signals. I examine items with similar V_t to see what my survey respondents considered to be high and low dispersion items. Figures 2 and 3 show the complete auction description from an item with $V_{highsd} = \$313.81$ and the auction description excluding picture for an item with a similar V_t of $V_{lowsd} = \$290.23$. The technical specifications (speed of processor, RAM, CD-ROM, and hard drive capacity) of these computers are approximately the same, indicating that the respondents valuations seem to account for hedonic characteristics. The first item had $SD_{highsd} = 505.23$, while the second item had $SD_{lowsd} = 304.74$ (the coefficients of variation are 1.61 and 1.05, respectively).

The survey seems to correctly distinguish between the informative and less informative product descriptions. Note that the high-dispersion item lacks the level of detail of the low-dispersion item. Both descriptions show pictures, but the low-dispersion description includes a picture of the actual computer for sale, while the high-dispersion picture only shows a similar computer. Presumably the high-dispersion computer does not include monitor, keyboard, mouse, etc., but what is meant by "System" is not made explicit. The low-dispersion seller states exactly what is still required.

The information that is dispersed with respect to the high-dispersion computer may take the form of different knowledge among auction participants about the similarity between the computer for sale and the picture in the ad, or the quality of reclaimed computers generally. The low-dispersion seller describes exactly how the computer does *not* work. Although this flaw may lower the auction participants' estimates of the value of the computer, participants are more certain about that valuation. If the seller merely said "This computer does not work" or didn't mention the flaw at all, information would be dispersed between those who were familiar with the types of failures encountered with Hewlett-Packard computers and those who were not. By revealing exactly what type of problem the computer possesses, the seller was able to lower the dispersion of that information.

In a CV setting, rational bidders respond to information dispersion when constructing their bids. Based on evidence that eBay PC auctions are CV and exhibit rational bidding behavior (Yin 2005), we would expect P_{highsd} to be lower than P_{lowsd}. Indeed, the high-

Figure 1: eBay Prices P_t vs. Average of Survey Responses V_t
Auctions are ordered along the horizontal axis by increasing eBay price, P_t. The corresponding averages of survey responses, V_t, are plotted for each auction as well. The correlation between P_t and V_t suggests that survey data can be used to measure unobservable values of items.



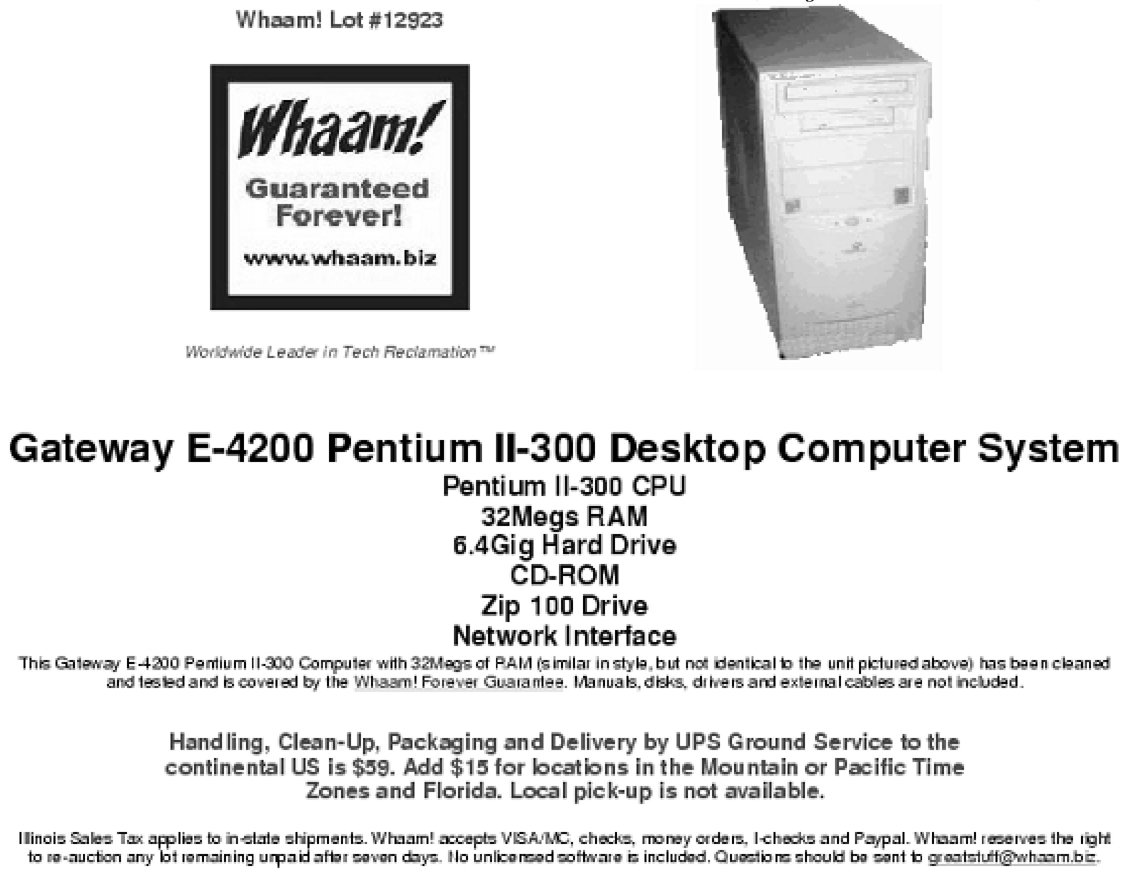
dispersion item sold for \$55.00; the low-dispersion item sold for \$96.50.

We have seen that surveys can provide measures of unobservable information. Both V_t and the SD_t generate results that are consistent with what we would expect from the relation between item values, information dispersion in the auctions, and prices in a CV setting. The next section describes the data used to correct for potential bias in these survey measures.

5 Participant Background

The survey asked about background characteristics of the respondents. Their responses are summarized in Table 5. The first set of responses in the table shows the number of yes and no responses for each auction in my sample, so those who responded to multiple auctions

Figure 2: High dispersion item description ($V_{highsd} = \$318.81$)



were counted multiple times. Half of all responses were from those who were familiar with computers. Half of all responses were also from those who had been shopping for computers and those who had looked at an online computer auction before. Of those that had recently bought a computer, most had bought their computer through a retail outlet. The next set of responses in the table show the number of people who recently bought 0, 1, or 2 or more computers, respectively. The survey respondents were then asked how many online auctions they had entered (0, 1, 2 through 5, and 6 or more were the respondents' possible choices). Those who had entered auctions were then asked whether all, none, or some of those auctions were on eBay, and whether they had won all, some, or none of those auctions. The majority of respondents had not bought a computer in the last six months. Most people had not entered an online auction, including most of those who had looked at online auctions. Those who had entered auctions tended to have done so more than once, favored eBay auctions, and had won some of those auctions.

I ran a least squares regression of individual responses $X_{i,t}$ on the background characteristics to determine how valuations differed between different types of respondents. The

Table 2: Summary statistics of survey respondents' backgrounds

Background questions	**Responses**			
(10,350 observations)	**no**	**yes**		
familiar w/computers	5140	5202	-	-
shopped for computer in last 6 months	4923	5427	-	-
bought computer via auction	-	403	-	-
bought computer via retail	-	2988	-	-
bought computer via wholesale	-	967	-	-
looked at online computer auction	5762	4588	-	-
looked at eBay computer auction	6449	3881	-	-
	0	**1**	**2+**	-
# computers bought last 6 months	6836	2471	1033	-
	0	**1**	**2-5**	**6+**
# online computer auctions entered	7428	750	1271	828
	none	**some**	**all**	-
... on eBay	4751	960	1509	-
... won on eBay	5338	1558	324	-

results are summarized in Table 5. The largest differences in valuations were correlated with differences in the respondents' familiarity with computers, recent purchases, and their familiarity with eBay auctions. Participants who were less experienced on these dimensions tended to value items more highly. Although a number of the coefficients are statistically significant, their magnitude relative to the average of V_t is low, and the overall explanatory power reflected by the R-squared statistic is low. Averaging over the responses of different types of respondents will probably not result in large differences from adjusting the mean for the different types, but we will allow for this possibility in the bias correction process in Section 6.

6 Bias Correction

V_t and SD_t may be biased measures of the true CV (or average PV), denoted v_t, and the dispersion of information facing auction participants, denoted $\sigma_{x|v,t}$. This section proposes a bias correction method which exploits the background data collected on the survey respondents.

On average, 20% of the responses for each auction in my sample came from respondents who had won all or some of the eBay online computer auctions in which they had entered. I designate their responses as "experienced" responses (subscripted by e), and designate the rest of the responses as "inexperienced" responses (subscripted by a).

I model and estimate the potential bias as follows. I treat the valuations $X_{i,t}$ from my survey respondents as potentially biased draws of signals $x_{i,t}$ that the auction participants draw about v_t. Thus, $X_{i,t}$ are drawn from a potentially different distribution than the one that the auction participants face. I model the responses from my inexperienced respondents, denoted $X_{a,i,t}$, as draws from a distribution whose mean may differ from v_t by a shift factor γ_0 and a scale factor γ_1 and whose variance may be different as well: $X_{a,i,t} \backsim (\gamma_0 + \gamma_1 v_t, \sigma^2_{x|v,a,t})$. I assume that the experienced survey respondents more closely resemble the auction participants. I model their responses as being drawn from a distribution whose mean only differs from v_t by a shift factor θ_0 and whose variance may be different: $X_{e,i,t} \backsim (\theta_0 + v_t, \sigma^2_{x|v,e,t})$. An unbiased estimate of v_t can then be written as

$$\hat{v}_t = \frac{J_{e,t}}{J_t}(V_{e,t} - \theta_0) + \frac{J_{a,t}}{J_t}\left(\frac{V_{a,t} - \gamma_0}{\gamma_1}\right), \tag{1}$$

where $J_{e,t}$ is the number of experienced survey responses in each auction, $J_{a,t}$ is the number of inexperienced survey responses in each auction, and J_t is the total number of survey responses to each auction. The average of the survey responses $X_{e,t,i}$ and $X_{a,t,i}$ are denoted $V_{e,t}$ and $V_{a,t}$, respectively. The parameters to be estimated are θ_0, γ_0, and γ_1. They capture the amount of bias in the responses.

I use the same process to model the potential bias in SD_t as a measure of $\sigma_{x|v,t}$. I assume that my experienced respondents draw from a distribution with variance $\sigma^2_{x|v,e,t} = \eta_0 + \sigma^2_{x|v,t}$, whereas my inexperienced respondents draw from a distribution with variance

Table 3: Effects of respondent's background on survey results

Variable	Coefficient
Constant	\$736.78‡
	(10.6199)
Familiarity w/computers	-\$74.37‡
	(13.2572)
Recently shopped	\$23.76
	(15.4142)
No. bought	-\$42.74‡
	(11.4023)
Venue of purchase	-\$2.01
	(4.1568)
Looked @ auctions	\$18.17
	(22.6089)
Looked on eBay	-\$62.30‡
	(22.3294)
No. auctions entered	-\$26.78‡
	(9.2624)
On eBay	\$1.21
	(11.8184)
eBay auctions won	\$26.60†
	(14.2579)

‡significant at 5%, †significant at 10%, $R^2 = 0.01$

$\sigma^2_{x|v,a,t} = \delta_0 + \delta_1 \sigma^2_{x|v,t}$. The resulting unbiased estimate of the information dispersion faced by the auction participants is as follows:

$$\hat{\sigma}_{x|v,t} = \sqrt{\frac{J_{e,t}}{J_t}(SD^2_{e,t} - \eta_0) + \frac{J_{a,t}}{J_t}\left(\frac{SD^2_{a,t} - \delta_0}{\delta_1}\right)}. \tag{2}$$

The variance of the signals $X_{e,t,i}$ and $X_{a,t,i}$ are denoted $SD^2_{e,t}$ and $SD^2_{a,t}$, respectively. The parameters to be estimated are η_0, δ_0, and δ_1. They capture the amount of bias in the dispersion of responses.

I can use a moment condition to identify θ_0, γ_0, and γ_1. I set the standard deviation of the experienced survey responses equal to the definition of the sample standard deviation, replacing $V_{e,t}$ with $\hat{v}_t + \theta_0$. The following moment condition is then estimated simultaneously with a price equation which includes $\hat{v}_t$ as a regressor:

$$SD_{e,t} = \sqrt{\frac{\sum^{J_{e,t}}(X_{e,i,t} - (\hat{v}_t + \theta_0))^2}{J_{e,t} - 1}}. \tag{3}$$

Results from estimation of a price equation in Yin (2005) show that the measurement bias in the common value is not severe relative to the average V_t of \$666.43: $\gamma_0 = \$83.61$, $\gamma_1 = 1.03$, and $\theta_0 = \$27.04$. On average, both the experienced and inexperienced respondents underestimate the value of the items, although the experienced respondents underestimate by less. The inexperienced respondents capture the scale of v_t almost perfectly.

The bias on dispersion for the experienced respondents is $\eta_0 = -60222.6$, whereas $\delta_0 = 76381.0$ and $\delta_1 = 1.83$ for the inexperienced respondents. To place these parameter estimates approximately in the context of standard deviations, the experienced responses underestimate $\sigma_{x|v,t}$ by 245.40 ($= \sqrt{60222.6}$). The inexperienced responses are approximately 1.35 ($= \sqrt{1.83}$) times larger than $\sigma_{x|v,t}$ and overestimate $\sigma_{x|v,t}$ by 276.37 ($= \sqrt{76381.0}$). The measurement bias in dispersion is relatively large compared to an average SD_t of 472.38. The bias is consistent with the expectation one might have that information is more dispersed among my survey respondents compared to the auction respondents, even after viewing the same auction description. This difference could be due to different interpretation of the information by the two groups or to differing initial levels of information dispersion between these groups prior to viewing the auction description.

The need for bias correction could be avoided by simply restricting the survey respondents to the experienced types. However, the rapid depreciation in values of the computer necessitated quick execution of the entire survey. By broadening the participant pool, I could achieve more responses per auction per day. Bias would be incurred whether I used a more restricted respondent pool or adjusted for depreciation for a more lengthy survey window; the use of a larger respondent pool had the advantage of avoiding the time and costs of a screening process for survey respondents.

In sum, although survey measures are prone to bias, the use of survey data also allows the collection of covariates on the survey respondents which can be exploited to correct for potential bias. The ability to conduct such a correction makes implementation of the survey much easier and faster. A researcher need only find a small sample of survey respondents who are just like the target population. The rest of the respondents can be drawn from the general population and generate information with noise, as long as that information is correlated with the true values.

7 Performance of Hedonic Regression Alternative

The convenience of using hedonic regression to control for the item value or for the dispersion of information signals may outweigh the benefits of using a survey the difference between each method's estimates of v_t and $\sigma_{x|v,t}$ are not substantial. To examine the difference between estimates from the survey method versus estimates from hedonic regression, I first generate the bias-corrected estimates of v_t and $\sigma_{x|v,t}$ from my survey. I plug the estimated survey bias parameters back into $\hat{v}_t$ and $\hat{\sigma}_{x|v,t}$ to generate a $\tilde{v}_t$ and $\tilde{\sigma}_{x|v,t}$ for each auction. I then regress these variables on observable covariates from the auctions which one might employ in a hedonic regression. The difference between the fitted and dependent variables from these regressions will reveal the extent to which hedonic characteristics can explain the variation in information captured by my survey procedure.

The characteristics chosen to model the common value in each auction are presented in Table 7 of Appendix B. The characteristics chosen to model the dispersion of information in each auction are presented in Table 4. These characteristics included dummies for whether the seller neglected to include information on various computer components (RAM memory $RAMNI_t$, operating system $OSNI_t$, floppydrive $FLOPPYNI_t$, keyboard $KEYBDNI_t$, CD/DVD drive $CDNI_t$, mouse $MOUSENI_t$) and the number of pictures and words included in the auction description ($PICS_t$, $WORDS_t$). I also included $BRAND_t$ and $PROCESSOR_t$ regressors described in Appendix B, since they may reflect differences in popular knowledge of the performance and quality of different types of computers and processors.

The ordinary least squares results are reported in Tables 6 and 5. Table 6 presents the regression of $\tilde{v}_t$ on covariates describing the item's value, and Table 5 presents the regression of $\tilde{\sigma}_{x|v,t}$ on covariates describing the information dispersion in the auction. The R^2 statistics are 0.71 for the common value and 0.27, respectively. The hedonic measures are unable to explain a third of the variation in common value and two-thirds of the variation in information dispersion that is captured by my survey methods. These results suggest that employing my survey procedure can provide information significantly different from that provided by hedonic regression.

This regression of bias-corrected survey measures on hedonic characteristics suggests a possible means of extending the survey results to a different sample. The coefficients estimated in Tables 6 and 5 can be used to generate a prediction of v_t and $\sigma_{x|v,t}$ for auctions outside of my current sample if repeating the survey procedure is too costly. In general, if the difference between the fitted values and dependent variable is acceptable, a researcher could employ survey methods for a small sample, and then extend the survey results to

Table 4: Summary statistics for information dispersion covariates

Variable (222 auctions)	**mean**	**st. dev.**	**min**	**max**
memory not indicated RAMNI_t	0.03	-	0	1
OS not indicated OSNI_t	0.45	-	0	1
Floppy drive not indicated $\mathrm{FLOPPYNI}_t$	0.33	-	0	1
Keyboard not indicated $\mathrm{KEYBDNI}_t$	0.38	-	0	1
CD/DVD drive not indicated CDNI_t	0.16	-	0	1
Mouse not indicated $\mathrm{MOUSENI}_t$	0.28	-	0	1
No. words in description WORDS_t	449.4	460.8	23	2727
No. pictures in description PICS_t	4.05	5.02	0	25

a larger sample by using the survey results to determine the relationship between hedonic characteristics and the survey measures.

8 Conclusion

This paper has presented several reasons why the use of survey data to augment auction data is valuable and feasible. The survey method could be beneficial in other research involving dispersed private information accompanying market data. Survey based measures that are used to augment market data will avoid problems which traditional surveys have when applied to non-market valuations; by referring to a market that already exists, framing problems are less severe. In addition, background data on survey participants can be collected and used to correct for biases between the survey and the information in the actual market.

In this paper, the survey is used to generate a measure of average values and the dispersion of private information in eBay personal computer auctions. This method has several advantages over the traditional method of estimating private information from auction observables alone. The richer measure of the common value avoids problems of endogeneity with the number of bidders in modeling price. The use of a survey method also generates a second moment that can be interpreted as a measure of information dispersion. This external information permits hypothesis testing that cannot be conducted otherwise. Even if the setting is private values, the survey data can be interpreted as the average and dispersion of private values.

An analysis of my survey results suggests that it is successful at accounting for technical characteristics that would determine the value of the computer, as well as the semantics of the auction description that would determine the dispersion of information. Estimates of the actual bias and scale differences between bidders and my survey respondents were either small or in the expected direction. Estimates of the difference between information gathered via the survey process described here and via alternative hedonic measures are

Table 5: OLS regression of $\tilde{v}_t$ on hedonic characteristics

Variable	**Coefficient**
Constant	$260.912^{\ddagger}$
	(39.333)
PRICE	$0.707^{\ddagger}$
	(0.032)
BRAND	-15.864
	(27.664)
PROCESSOR	-2.949
	(11.061)
SPEED	$0.049^{\dagger}$
	(0.029)
RAM	0.123
	(0.102)
HARDDRIVE	-1.180E-03
	(0.001)
MODEM	-8.787
	(15.032)
MONITOR	32.545
	(31.450)
MOUSE	$80.421^{\dagger}$
	(49.750)
KEYBOARD	$-94.556^{\dagger}$
	(49.910)
ZIP	-14.737
	(57.688)
FLOPPY	38.305
	(26.002)
APPLICATION	-36.003
	(32.900)
OS	$52.949^{\dagger}$
	(26.968)

‡significant at 5%, †significant at 10%, $R^2 = 0.71$

Table 6: OLS regression of $\tilde{\sigma}_{x|v,t}$ on hedonic characteristics

Variable	**Coefficient**
Constant	$309.508^{\ddagger}$
	(31.793)
PRICE	$0.161^{\ddagger}$
	(0.022)
WORDS	-0.011
	(0.022)
PICS	-0.772
	(1.973)
SPEEDNI	11.452
	(79.522)
RAMNI	-41.403
	(73.793)
HDNI	-41.584
	(82.167)
MODEMNI	-14.826
	(44.100)
MONTITORNI	4.853
	(17.987)
MOUSENI	40.627
	(25.859)
FLOPPYNI	-2.052
	(17.051)
KEYBOARDNI	-56.329
	(24.402)
APPNI	$-4.76949^{\ddagger}$
	(20.322)
OSNI	20.867
	(17.233)
BRAND	-17.308
	(18.531)
PROCESSOR	$-13.735^{\dagger}$
	(7.362)

‡significant at 5%, †significant at 10%, $R^2 = 0.27$

large, indicating that the survey method does capture significantly more information.

This method has applications in any setting where hedonic estimation may ignore important idiosyncratic differences. Models which include expectations over privately held information may find surveys to be a useful way of simulating the distribution of that information. Since the researcher can exploit background characteristics of the survey respondents to correct for bias, the convenience and speed of implementing the survey is improved through including the general population amongst the survey respondents. Even if a survey can only be conducted for a part of the sample, the survey results can be combined with hedonic regression to extend those results to the rest of the sample. The advantages of the extra information gathered and the ability to correct for errors reduce the relative cost of administering a survey.

A Survey Description

Auction descriptions were edited to remove all bids and identities (other than seller identification within the auction description itself) and reputations involved. A CGI script was developed by Paul Hartke to translate PostScript graphics of these auctions into web-viewable formats and to automate a process to assign unique ID numbers to survey respondents and record which auctions were viewed and respondents' values. A separate Formage script was written to solicit background information on the respondents. The following solicitation was sent to friends of the author and posted to relevant newsgroups:

"Could you please help my friend Pai-Ling Yin, http://www.stanford.edu/~pyin, in her PhD economics research project to determine the distribution of commonly held values for products? Just fill out a short survey asking you to look at the descriptions of 10 computers and giving your estimate of how much they are worth. Even if you are not familiar with computers and their prices, your best guess will still be useful to Pai. So send this on to your grandparents, parents, siblings, cousins, friends, and co-workers for extra chances at winning!

"All completed surveys will be entered in a drawing for two $1,000.00 prizes and thirty $60 prizes. For each friend you get to do the survey, you get an extra chance to win. Deadline for all submissions is 11:59pm, July 20, 2002. E-mail pyin@stanford.edu if you can't make the deadline but still want to participate.

"Thanks very much! Email pyin@stanford.edu if you have questions. Privacy will be honored; no names or emails will be released except for the winners (posted at the survey site after 1/1/03)."

PRIZE DETAILS:

"As a reward for participating, a drawing will take place on January 1, 2003, over all completed surveys and referrals. Two people will win checks for $1,000.00. Odds of winning depend on the number of times you participate and the total number of surveys completed.

"As an incentive to think sincerely about your estimates, fifteen $60.00 prizes will be awarded to the people whose estimates are closest to the average of all other estimates in the same auction, and fifteen $60 prizes will be awarded to the people whose estimates are closest to a set of estimates provided by a panel of computer sales people. This allows both computer experts and non-experts to have a chance at winning."

BACKGROUND:

1. Please enter your e-mail address. This will be used only to contact you if you win. Please use the same e-mail if you participate more than once.

2. Did someone refer you to this survey? Please enter his/her e-mail address:

3. Are you involved in work or hobbies that cause you to be very familiar with the prices of computers and computer components? YES/NO

4. Have you been shopping for a computer in the last 6 months? YES/NO

5. How many computers have you bought in the past 6 months (either for personal use or for work)? 0/1/2+

   If you bought a computer, did you buy it/them through (check all that apply):

   an auction process (does not include using “Buy It Now”)?

   a retailer (includes using “Buy It Now” to buy the computer at a set price rather than at the winning auction price)?

   a wholesaler (someone who normally sells computers to stores, not directly to consumers)?

6. Have you ever looked at computers on an online auction website? YES/NO

   On eBay? YES/NO

7. In how many online computer auctions have you participated in your life?

   0/1/2-5/6+

   How many were on eBay? NONE/SOME/ALL

   How many of the computer auctions did you win? NONE/SOME/ALL

“After you hit the submit button, you will be given descriptions to evaluate, one at a time. You will be given 1 chance to win prizes for every 10 auctions you complete.

“You may want to copy your answers for each auction on some paper so that you can compare auctions.

“You can use the ‘Back’ and ‘Forward buttons on your browser to compare descriptions; if you want to change answers, you can use the back button as well, but make sure to click ”Send” to register the change. Then click ‘Send’ on the subsequent pages to return to the auction you left off with.

“Send e-mail to pyin@stanford.edu if you have any problems, want to change an answer after exiting, or want to confirm your entries. Please make sure the above answers are correct before you click ‘Send’, so that you don’t have to backtrack to this page to change any answers.

“Please wait a few seconds while the computer description loads...

“Assume that your friend is interested in buying the computer described below. Taking into account all information that you see (including shipping and insurance costs), what is the MOST she should be willing to pay for this computer (NOT how much she should bid!)? Even if you don’t understand some of the description, please do your best to be consistent (better computers cost more). Feel free to look at ads or websites to help you make better recommendations, but please DO NOT look at online auction sites to get a sense of prices. Scroll ALL the way down to enter your value at the bottom of the description.”

B Regressors for $\mu_{v,t}$

I constructed a set of hedonic characteristics of the computers to be used as regressors for determining $\mu_{v,t}$. The expected value of a computer satisfying certain criteria before a bidder

Table 7: Summary statistics for regressors for a priori value

Variable (222 auctions)	**mean**	**s.d.**	**min**	**max**
recognizable computer $BRAND_t$	0.27	-	0	1
quality of brand of $PROCESSOR_t$	2.14	1.11	0	3
processor $SPEED_t$	1088	684.88	0	2530
RAM_t memory capacity	210.77	196.04	0	1100
$HARDDRIVE_t$ memory capacity	27724	27755	0	160000
device for $INTERNET_t$ access	1.31	0.83	0	2
includes $MONITOR_t$	0.21	-	0	1
includes CD_t or DVD drive	0.83	-	0	1
includes $FLOPPY_t$ drive	0.66	-	0	1

has seen the auction description is captured in $\mu_{v,t}$, while v_t measures the value of a computer after having seen the auction description. Summary statistics are presented in Table 7.

The dummy variable $BRAND_t$ indicates whether the computer had a recognizable brand name (Toshiba, Dell, Hewlett-Packard, IBM, Compaq) or not. A ranking of the processor brands in $PROCESSOR_t$ ranged from no mention of processor brand (= 0) to Pentium (= 3). The processor's speed was denoted as $SPEED_t$. The amount of memory included was characterized by the RAM and harddrive capacity ($RAM_t, HARDDRIVE_t$). I ranked the presence of a communications device in $INTERNET_t$ (0 for no device, 1 for modem, 2 for other). Dummies were created for whether a monitor, CD/DVD drive, and floppy drive was included or not ($MONITOR_t, CD_t, FLOPPY_t$). If the auction description did not provide any information about a characteristic, the value was coded as 0.

References

Aadland, David and Arthur Caplan, "Household Valuation of Curbside Recycling," *Journal of Environmental Planning and Management*, 1999, *42*, 781–800.

Athey, Susan and Philip A. Haile, "Identification of Standard Auction Models," *Econometrica*, 2002, *70(6)*, 2107–2140.

Bateman, I.J. and Ken G. Willis, in I.J. Bateman and Ken G. Willis, eds., *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford: Oxford University Press, 1999.

Bjornstad, David J. and James R. Kahn, in David J. Bjornstad and James R. Kahn, eds., *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs*, Brookfield, VT: Edward Elgar, 1996.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong, "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 2000, *68*, 525–574.

Hausman, Jerry A., in Jerry A. Hausman, ed., *Contingent Valuation: A Critical Assessment*, New York: North Holland, 1993.

Laffont, Jean-Jacques and Quang Vuong, "Structural Analysis of Auction Data," *The American Economic Review, Papers and Proceedings*, 1996, *86(2)*, 414–420.

Li, Tong, Isabelle Perrigne, and Quang Vuong, "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 2000, *98(1)*, 129–161.

—, —, and —, "Structural Estimation of the Affiliated Private Values Auction Model," *Rand Journal of Economics*, 2002, *33(2)*, 171–193.

Yin, Pai-Ling, "A Survey-Based Procedure for Measuring Unobservable Information," March 2005. Working Paper.

Figure 3: Low dispersion item description (V_{low} = \$290.23)

Hewlett-Packard Pavilion 6330 Computer

Includes:
Tower
Keyboard
Mouse
Power cord
Phone cord
Mousepad
Norton Utilities
America Online 7.0 installation disk
Prodigy installation disk
Mindspring installation disk
CompuServe Wow! installation disk
HP Pavilion recovery disk
Corresponding owner's manuals and documents
Extra parallel port(installed in tower)

Requires:
Monitor
Printer
Speakers

Hewlett-Packard Pavilion 6330 Specs:
Windows 98 Operating System
AMD-K6 - 2/300 processor with 3DNow! technology
Ultra expandable and upgradable with 6 bays and 5 slots(1 taken
______by extra parallel port)
48 MB SDRAM shared memory architecture – up to 4 MB video
______memory
Spacious 4GB hard drive
24x max CD-ROM drive
High velocity V.90,K56flex Data/fax modem
One-touch keyboard
2 USB ports for easy plug and play
Year 2000 compliant

This computer works fine when hooked up only to a monitor, printer,
and speakers with no other additional hardware options. Whenever

I connect my Iomega Zip Drive and scanner however, this
computer starts to have problems. As long as you don't connect
any unnecessary external hardware devices other than the printer
and speaker you should be alright.

Winning bidder to pay shipping & handling.
I reserve the right to refuse bidders with negative feedback.
I accept EBay online payments.
Please be prompt in your post-auction correspondence
Reserve price: $50.00.

The eBay Way to Pay - Enjoy Full Purchase Protection!