Research Report – UCD-ITS-RR-12-17

Internet and the Efficiency of Decentralized Markets: Evidence from Automobiles

September 2012

David Rapson

Institute of Transportation Studies ◦ University of California, Davis
One Shields Avenue ◦ Davis, California 95616
PHONE (530) 752-6548 ◦ FAX (530) 752-6572
www.its.ucdavis.edu

UC DAVIS
SUSTAINABLE TRANSPORTATION CENTER
of the Institute of Transportation Studies

STC Final Research Report R03-5

Internet and the Efficiency of Decentralized Markets: Evidence from Automobiles

September 2012

David Rapson

Department of Economics
University of California, Davis
One Shields Ave., Davis, CA 95616

Internet and the Efficiency of Decentralized Markets: Evidence from Automobiles*

David Rapson & Pasquale Schiraldi

University of California at Davis, London School of Economics

September, 2012

Abstract

We estimate the causal effect of Internet on volume of trade in a major distributed marketplace. From 1997 to 2007, a period during which residential Internet use tripled, Internet increased the volume of used cars traded in California by 7.2 percent. This implies a substantial welfare gain due to reductions in market frictions of approximately $43 million per year relative to 1997 Internet penetration levels. These gains are likely attributable to search and matching cost reductions and increased symmetry of information availability across buyers and sellers. Intriguingly, the effect is stronger in thick markets and urban areas, implying that Internet may help to diminish choice overload. Our results suggest that policies promoting broadband Internet deployment may enhance efficiency.

Keywords: Decentralized markets, transaction costs, information, Internet. automobiles

JEL classification: D12, D83, L62, L68

*Special thanks to participants in several seminars for comments and suggestions. We are grateful to STICERD at LSE for financial support. The research was also supported by a grant from the Sustainable Transportation Center at the University of California Davis, which receives funding from the U.S. Department of Transportation and Caltrans, the California Department of Transportation, through the University Transportation Centers program. Any errors are our own. Email: dsrapson@ucdavis.edu, p.schiraldi@lse.ac.uk

1 Introduction

The goal of this paper is to estimate the efficiency gains from the availability of Internet in a large and active decentralized trading market: the market for used automobiles in California. Access to the Internet drastically reduces the cost of acquiring information, and this potentially improves several aspects of the market. Buyers and sellers can search for each other and be matched with unprecedented ease. Internet may also reduce asymmetries in information between buyers and sellers, helping to overcome the well-known problems associated with adverse selection. Overall, improved access to information can allow more of the potential gains from trade to be realized. These possibilities are especially relevant today in light of the extraordinary growth in online marketplaces (e.g. eBay and Amazon.com)[1], as well as online resources that support all forms of trade (e.g. CarFax, Autobytel, Angie's List).

Much of the existing empirical literature on the impact of Internet focuses on changes in the level and distribution of prices.[2] Such analyses are generally concerned with consumer welfare rather than joint surplus (though some also derive implications about the nature of competition). Our aim in this study is to shed light on the magnitude of total welfare gains, making volume-of-trade the relevant economic outcome variable of interest. The empirical literature examining Internet's effect on transaction volume is sparse, and a main contribution of this study is add a credible data point from an important market.[3]

The used car market has many features that make it an appealing setting for our analysis. In addition to being very large and active, it is also a market where transaction costs are high and

[1] An ever-increasing volume of goods is now traded in decentralized markets. In 2008, nearly $60 billion of goods were transacted on eBay, and Amazon.com had $20 billion in total revenue.

[2] Examples include Brown and Goolsbee (2002) on life insurance, Orlov (2011) on airline fares, and Zettelmeyer et al (2006) on car transactions.

[3] Dana and Orlov (2009) examine the Internet effect on airline capacity utilization, and Ghose et al (2006) study the market for used books.

information is costly to acquire[4] and often asymmetric, with potential buyers trying to assess unobserved quality that is known to the seller. Over the last 15 years, several Internet-based resources have emerged in the industry. Some retailers use the Internet as a their primary customer acquisition channel (e.g. carmax.com, autotrader.com, carsdirect.com). Referral services such as autobytel.com help to direct consumers to low-cost offline retailers.[5] The Internet is also a platform for several purely information-based resources. The most relevant is carfax.com, which provides vehicle history reports on used cars and is has all but eliminated the information gap between buyers and sellers. Moreover, the possibility of gathering more information and reaching a broader number of buyers/sellers has substantially increases the surplus generated upon trade.

The mechanism at work is the following: the market is characterized by prospective buyers and sellers that enter into a costly, two-sided search and matching process. Consumers have heterogeneous preferences over product attributes, in our case automobile characteristics. Hence, a given car has different appeal to different consumers (as emphasized by the empirical literature starting with Berry at al. 1995). Therefore the availability of more information and a broader market reduces search costs and favors better matches (i.e. generates matches with higher joint surplus).[6] Each of the aforementioned potential mechanisms by which Internet would effect a decentralized market–decreased search costs, improved match quality, and less adverse selection[7]–would be similar in at least one important way: they would each increase volume of trade.

In order to test this hypothesis empirically, we assemble and analyze a detailed dataset that includes county-level Internet penetration and used car registration and transaction counts during the period 1997-2007 in California. These data allow us to exploit three main sources of variation:

[4]Porter and Sattler (1999) and Gavazza, Lizzeri and Roketskiy (2012) reports transaction costs that range between 15-50% of the trade-in value of the used cars in US.

[5]Zettelmeyer et al (2006).

[6]A similar framework has been use in Ngail and Tenreyro (2010) to study the housing market.

[7]Lewis (2011) shows that used-car sellers are able to partially contract on the quality of their goods by disclosing private information on the webpage, i.e. by posting text, photos, etc. He shows that woluntarily disclosure of private information on the web helps to protect the buyer from adverse selection.

(1) differences in Internet penetration over time; (2) differences in Internet penetration across geographical locations; (3) differences in the counts of registered and sold cars across counties and over time. This variation allows us to estimate the marginal effects of Internet penetration on trade volume using a continuous difference-in-differences approach. We account for the possibility of within-county correlation by clustering standard errors at the county level.

We also explicitly address concerns about the endogeneity of the vehicle stock as a right-hand-side variable in the trade volume regressions. Our estimates would be biased if, for example, leakage of used cars across county boundaries is correlated with changes in Internet penetration. We use an array of instrumental variables to address this concern, and discuss our approach in detail in the section on empirical evidence.

Our results show evidence of substantial efficiency gains in the used car market due to the Internet. In California, Internet penetration increased by 28 percentage points (from 14 to 42 percent) from 1997 to 2007. Our estimates imply that over 120,000 additional used car transactions occurred state-wide in 2007 than would have occurred had Internet stayed at 1997 levels. This is an increase in the resale ratio of 7.2 percent. Further, the specifications that we employ allow price to vary along the equilibrium path. Thus, these results should be interpreted as general equilibrium changes in quantity. The magnitude of Internet's impact on volume of trade in this market is meaningful. The 7.2 percent increase in trade volume implies a lower-bound welfare gain from transaction and search cost reductions of $25 per unit traded (relative to the 1997 Internet levels). Across the entire state of California, this translates into an annual welfare gain of $43 million.

Previous studies suggest that market thickness–the density of potential buyers and sellers–will be an important feature of our setting.[8] We confirm both that transaction volume increases in the

[8] For example, Gavazza (2010) studies a decentralized market for used aircraft (which is also characterized by two-sided search), and shows that an increasing returns-to-scale matching function increases the probability of a meeting and trade when the number of potential buyers and sellers is high. In his empirical results, he finds that thicker markets result in higher transaction frequencies.

size of the vehicle fleet, and that important differences exist between rural and urban areas. Urban areas exhibit more used car transactions, even after conditioning on market thickness. Further, the Internet increases trade more in urban areas than rural. Together, these results imply that subsidies for Internet will yield higher benefit in urban areas than in rural areas (at least with respect to the functioning of decentralized markets).

The present paper builds on a growing literature that examines the role of transaction costs in the second-hand market for durable goods. Several studies develop theoretical frameworks to investigate patterns in asset allocation and trade in secondary markets.[9] Others choose a calibration approach in which certain key parameters (e.g. transaction costs) may be perturbed, and predicted model counterfactuals observed along the equilibrium path.[10] Two empirical contributions are particularly relevant. Porter and Sattler (1999) estimate the role of asset depreciation on trade volume, with a particular focus on the role of age of the durable good. They find that patterns of trade in the used car market are consistent with an important role for transaction costs, but fit the adverse selection model less well. Schiraldi (2011) develops structural model that can be estimated to retrieve the actual magnitude of transaction costs, and confirms that they must be very large in order to rationalize the observed persistence of car ownership. These papers support the hypothesis that higher transaction costs and frictions in the secondary market lead to lower volume of trade.

A separate economic literature exists on the effect of Internet on the operation of markets.[11] These studies mainly focus on the price of (new) goods, both in terms of level and dispersion. A few papers also show that Internet facilitates better matching between products and consumer tastes, as in the present work.[12] In light of these two full and important strands of literature, our contribution may be viewed as a purely empirical effort that exploits exogenous variation in transaction costs to

[9]See, for example, Anderson and Ginsburgh (1994), Hendel and Lizzeri (1999), Schiraldi and Nava (2011).

[10]These papers include Stolyarov (2002), Chen et al (2012), and Gavazza et al (2012).

[11]See Brown and Goolsbee (2002), Brynjolfsson and Smith (2000), Ellison and Ellison (2005), Orlov (2011), Zettelmeyer et al (2006).

[12]Examples include Scott Morton (2005), Brynjolfsson et al (2003) and Ghose et al (2006).

examine market outcomes. Underpinning our empirical strategy is the exogeneity of the availability of Internet technology with respect to the individual used car buyer and seller.

The paper proceeds as follows: in Section 2 we describe a theoretical framework that captures some of the important features of our setting, including search costs and match-specific quality. We then test hypotheses against data from the California used car market, and those data are presented in Section 3. We discuss the results in Section 4 and Section 5 concludes.

2 Data

The empirical focus is on the state of California between the years 1997 and 2007. California is the most populous state in the US, with a population of nearly 32 million in the year 2000. It was also the frontier of the high technology boom that yielded many of the computing and Internet technologies that are now commonplace in the US. As we will show, the decade spanned by our data is characterized by rapid consumer adoption of Internet at home. Annual automobile purchases over the same period increased from nearly 1 million in 1997 to 2 million in 2007, or roughly one car for every 18 people. To investigate the relationship between Internet penetration and automobile transactions, we combine data from several sources.

The cross-sectional unit of interest in this study is the county, and each variable ultimately appears at the county-year level. Car registration and transaction counts were acquired from R.L. Polk, which gathers these data from the California Department of Motor Vehicles. An observation is the total number of cars for each make-model-year that are registered or transacted in a given year, by county. These include both new and used cars, the latter of which may be either dealer or private transactions. We augment these data with car characteristics, which were acquired from Wards Automotive's Automotive Yearbook. New and used car prices were available at the state level by Ward's and the National Automobile Dealers Association (NADA), respectively.[13] From the R.L.

[13]Theoretical predictions are clearer for volume of trade than for price, and in any case these are not terribly useful

Polk dataset, for each make-model-year we observe the stock and transaction of all cars of each type in each time period t. The stock of cars available of each type k in period t provides a measure of the "thickness", or size, of the market for each model in a given year.

In this study, we postulate that services provided via the Internet serve to lower transaction costs in the used car market. As such, residential Internet penetration is the explanatory variable of interest in this study. We hypothesize that widespread use of the Internet by both buyers and sellers decreases the costs associated with buying and selling cars, in particular used cars. Figure (1) displays graphs of Internet penetration for each county in the sample.[14] We obtained county-by-year Internet penetration from the Current Population Survey (CPS) Supplement Files. We consider a household to be an Internet user if they respond "yes" to the question "Use Internet at Home?" and any of the similar Internet access questions. These questions appear in the CPS in a subset of years from 1997 to 2007.[15]

Table (1) provides summary statistics of the variables used in the empirical analysis.[16] We present statistics for three years of data: 1997, 2003, and 2007. These data reveal a clear increase in the annual number of used cars owned and sold between 1997 and 2007. While it is not shown directly here, the 70 percent increase in the stock of registered used cars, along with a near doubling of the number of distinct models, reflect an increase in the overall quantity of new cars purchased. This explains why, despite the number of used car transactions increasing over these years, the volume of

when implementing a county-level analysis.

[14]While one might expect Internet penetration to increase monotonically over time, our data show some non-monotonicity. These may reflect changes in household perceptions about what "Internet use" refers to in the survey (e.g. dial-up versus broadband). There may also have been sampling error that affected smaller counties disproportionately. Classical measurement error would impose attenuation bias on our results, biasing them towards no effect.

[15]The supplementary Internet questions appear in the CPS in 1997, 1998, 2000, 2001, 2003, and 2007.

[16]The automobile and Internet data sources are described above. The demographic data are obtained from the U.S. Bureau Census and the Employment Development Department of the State of California. Finally, county-level gasoline price are from the Oil Price Information Service (OPIS).

trade as a fraction of total used cars registered declined from 13.5 percent to just below 10 percent over time.

Importantly for our purposes, the residential penetration of Internet use increased three-fold over the decade. The most Internet-intense county in 1997 (Orange County) had only 25 percent penetration, far below its 57 percent level in 2007.[17] By the end of our sample, 17 counties had achieved over 45 percent Internet penetration, and only eleven counties were below 35 percent. This decade also witnessed significant economic growth, population aging, and a doubling in the price of unleaded gasoline. Median incomes rose from $37,000 in 1997 to $53,000 (in nominal terms), while unemployment fell from 9.2 percent to 6.7 percent. The presence of extensive heterogeneity across counties leads us to include these time-varying county demographics as regressors in our empirical analysis (in addition to an array of fixed effects).

3 Empirical Approach and Evidence

In the automobile market, replacement decisions are based on depreciation in quality of goods over time. Each period, consumers assess the quality of the durable good they own. If the gain in utility from updating their holdings, net of prices, exceeds transaction costs, consumers sell their used goods in the second-hand market and replace them with durable goods of the preferred quality. It follows that, in the absence of any market frictions, consumers have no incentive to hold their durable goods across multiple quality levels. If goods depreciate every period, consumers will never hold onto their durables for more than one period but will always update to their preferred car quality.[18] Hence, lowering market frictions should induce a high turnover rate. The resale ratio (i.e. the ratio between number of traded cars and stock for each make-model-year) is a direct approximation of the magnitude of frictions that prevent consumers from replacing their car every

[17]The county with the highest Internet penetration in 2007 was Marin at 64 percent.

[18]See, for example, Hideo and Sandfort, 2002; Rust, 1985; Schiraldi 2010, among others)

period. In a world without frictions, this ratio should be equal to one.

Conceptually, several potential measures of activity in this setting would shed light on the magnitude of transaction costs or search frictions. One could glean a measure of the size of frictions by looking at how long a car stays on the market before selling. Alternately, if one knew how many potential buyer and seller meetings occur before closing a sale, one could observe how these quantities change over time, all else equal, as Internet penetration increases. We pursue a somewhat less direct approach, but one that is feasible given the available data.

We investigate the role of Internet under the assertion that it alters the search environment in a meaningful way. We hypothesize that it either reduces cost of search, improves match-specific quality, improves the symmetry of information, or induces some combination of these effects. In each case the gains from trade would increase, resulting in higher volume of trade. The potential mechanisms via which Internet may have this effect are many. Internet may directly match buyer and seller, and perform the function of market maker. It may also facilitate consumers' search for a seller/dealer, of the latter's search for the former. In this case the transaction does not necessarily occur online.

The Internet my also help to inform consumers and by so doing potentially reducing information asymmetry.[19] Adverse selection is widely viewed to be an important feature of the used car market, so it is quite plausible that technological advance in the dissemination of relevant (and previously one-sided) information would have significant efficiency implications. It has been argued that the magnitude of search costs in a market is likely to be a function of the market's "thickness", and we will investigate this hypothesis in this study. The idea is that markets with a higher number of potential buyers and sellers are likely to exhibit higher meeting *rates* and, correspondingly, a

[19]Lewis (2011) shows that used-car sellers are able to partially contract on the quality of their goods by disclosing private information on the webpage, i.e. by posting text, photos, etc. He shows that woluntarily disclosure of private information on the web helps to protect the buyer from adverse selection.

higher rate of transactions.[20] In our setting, the most direct interpretation would be to view market thickness in terms of registered cars of a given type. The relationship to market thickness and Internet use is potentially important for interpreting the mechanism by which Internet improves market outcomes. For example, were Internet to facilitate an increasing trade ratio of cars at lower registration counts, that is suggestive of either a decrease in search costs or an increase in the match-specific quality achieved in the market. At higher registration counts, though, finding potential matches is likely easier, and thus search costs are likely not as important as match-specific quality.

In this section we describe several specifications that explore the relationship between Internet use and volume of trade in the used car market. The first set of specifications examines the overall effect on volume of trade and resale ratio, and the second set sheds light on the mechanisms by allowing for effects that differ by market thickness. Finally, we explore the role of level of urbanization and how the effect of Internet on trade operates differentially in this context. The population density and level of urbanization analysis lends additional insight into the mechanisms at work, and is also of direct policy interest.

Note that the predicted effects are general equilibrium in nature, and that supply and demand curves need not be separately estimated to test the predictions. We concentrate the analysis on equilibrium volume of trade and resale ratio. Thus, in all specifications below we choose to omit price as a right-hand-side variable, allowing it to vary on the transition path.

3.1 Volume of Trade

We first test the hypothesis that, *ceteris paribus*, increases in Internet penetration lead to increases in volume of trade. Support for the hypothesis of interest will be found if higher equilibrium volume of trade is observed within a county during years of high Internet penetration. The non-random

[20] See, for example, Gavazza's (2011) investigation of the commercial aircraft market.

diffusion of Internet requires us to control for some important potential confounders. Locations with higher levels of income may have a higher Internet adoption rate, since Internet was a new and expensive technology during our period of study. Further, the costs of installing and accessing the network have a high fixed component, which leads us to expect more densely-populated regions to be serviced first. If income or population density are correlated with used car transaction volume (and both seem plausible, if not likely), we must control for these variables. It is also possible that there are other, unobserved variables correlated with both Internet penetration and the number of used cars traded, that are driving the relationship. Systematic differences in characteristics of the locations with different levels of Internet penetration and volume of trade will bias the estimates.

In order to deal with these endogeneity concerns, we exploit the panel structure of the data by employing a difference-in-differences specification. As regressors, we include observable demographics that vary at the county-year level to address many of the aforementioned endogeneity concerns. These demographic controls include the average income level in each county, the unemployment rate, the age distribution in the population, and population level itself. We also include the average price of gasoline for each county-year combination, thus controlling for the possibility that changes in fuel price affect the volume of used cars traded and its allocation. The set of control variables is completed via inclusion of county and time fixed-effects. These control for any residual time-invariant county unobservables and aggregate shocks that are shared by all locations in a given year. Identification of the Internet effect is thus achieved from within county variation in transaction volume and Internet penetration over time.

The baseline relationship between Internet and transaction volume is estimated by the following equation:

$$y_{jct} = \gamma_1 Internet_{ct} + \beta D_{ct} + \alpha_c + \alpha_t + \alpha_j + \epsilon_{jct} \tag{1}$$

Subscript c denotes county, t denotes year, and j denotes model type. The dependent variable, y_{jct}, is the log of the resale ratio: the number of used cars of model j traded at the county level in each year, divided by the stock of model j cars in that county-year. The coefficient of interest is γ, which conveys the conditional effect of Internet on the resale ratio. Depending on the specification, a variety of model-based fixed effects (α_j) are used. Additionally, we include time-varying county demographic variables (D_{ct}) to control for potentially important confounding variables. In the specifications shown, we control flexibly for car make, model, and model-year, though both the qualitative and quantitative results are robust to alternate specifications of car-level controls.[21]

Traditionally, inference is based on the assumption that the data are independent across c and t. Each observation is treated as a random draw from the same population, uncorrelated with the observation before or after. However, this is not always the case, and the most important form of dependence arises in data with a group structure. For example, time series observations within a group may exhibit serial correlation. Failure to account for this will cause standard error estimates to be biased downwards, leading to incorrect inference. To account for these considerations, we take advantage of the high number of counties (57) which allows us to implement a simple and widely used approach. In every specification, standard errors are clustered at the county level.

We estimate several OLS specifications of equation (2), using the resale ratio as well as count of used cars sold by model as the outcome variable of interest.[22] Table (2) displays the results, with the addition of incremental control variables when moving left to right between columns. It is illuminating to understand how the results change with the addition of these controls. Column 1 shows a baseline regression of used car resale ratios on county Internet penetration, with only car characteristic controls (which are present throughout). This sparse group of covariates captures very little of the variation in resale ratio. The effect of interest is not statistically different from zero, but nonetheless it indicates a positive qualitative relationship between Internet and resale ratio. For these specifications, the coefficient on Internet should be interpreted as a percentage point change in the average fraction of used cars transacted in a given year.

Columns 2 through 5 display results that show the importance of including county demographics

[21]Recall that our identifying variation occurs at the county-year level, so this is not surprising.

[22]"Model" here is specific to a make and model of a particular vintage (model-year) each year.

and the cross-section and time fixed effects. They also help to confirm some basic priors about what affects trade volume. In column 2 we include demographic covariates, but no county or time fixed effects. The estimated treatment effect increases relative to column 1, but is still insignificant. The coefficient on income implies that wealthy households hold their cars longer, and that fuel price increases are associated with fewer transactions.

Columns 3 and 4, which include either year or county fixed effects, should be interpreted relative to column 5, the specification for which includes both. Column 3 indicates that failing to control for time-invariant county unobservables biases the treatment effect downwards. Time-invariant county characteristics may include such variables as sales tax, the rate of which varies by county in California. Failure to account for this variation biases the estimated effect downwards, which is the expected direction of bias for a variable such as sales tax (which is negatively correlated with transaction volume).

Column 4 swaps year fixed effects for county fixed effects, and thus uses across-county variation within each time period to strip out any unobserved aggregate shocks. Again, comparing it to the estimates from column 5 provides a natural explanation for the change in treatment estimate. Using across variation instead of within variation takes out the spurious correlation between Internet penetration and volume of trade that is due to unobservable aggregate shocks over time. Aggregate shocks will include macroeconomic shifts, which are likely an important determinant of used car demand. The Internet coefficient in this specification increases to 0.063, indicating that the inclusion of aggregate time effects absorbs variation that is positively correlated with transaction volume (e.g. economic growth). Thus, it appears important to include both of the controls discussed above.

Column 5 is thus our preferred specification. It displays results that control for county, year, and car fixed effects, in addition to the full array of demographic characteristics. In this specification, when a county goes from no residential Internet penetration to ubiquitous usage, the resale ratio increases by 4.4 percentage points, an effect that is significant with 95 percent confidence. We interpret these results supporting the hypothesis that the Internet improves the search and matching technology in California's used car market.

To put the magnitude of this effect into perspective, let us examine the used car market in 2007 relative to 1997. The weighted mean county Internet penetration is 43.6 percentage points in 2007, implying an increase of 28.6 percentage points over this period. The coefficient from column 5 translates into a 1.9 percentage point rise in the resale ratio (which was 0.174 in 2007) due to Internet. These estimates suggest a causal effect of Internet which increases the resale ratio by 7.2 percent relative to 1997 levels. State-wide, this translates into over 120,000 additional transactions per year relative to what would have occurred in 2007 at 1997-level Internet penetration. As discussed above a high level of frictions on the secondary market (transaction and search costs, asymmetric information etc.) reduces the frequency of replacement. Hence the resale ratio approximates the size of this frictions preventing consumers from replacing their cars in every period.

State-level automobile price data prices offer an estimate of the average transaction costs. As in Porter and Sattler (1999) and Gavazza et al (2012), the difference between the trade-in value and the retail price offers an intuitive measure of the cost of two-sided search and matching. In our data, this difference ranges from $1,607 to $1,837. Under the assumption of a linear relationship between transaction costs and resale ratio, the shared surplus gain in 2007 due to a 28.6 percentage point increase in Internet penetration is $25 per car traded. When compared to the counterfactual of zero Internet, this gain increases to $38. State-wide, the efficiency gain from transaction cost reduction alone is $43 million annually, or roughly a 2 percent reduction in transaction costs for 1.7 million used cars traded. There are two reasons why this amount represents a lower-bound of the total welfare gain due to Internet penetration in this market. First, it does not include surplus generated by the realization of additional transactions (over 120,000 annually). Second, it is likely that Internet facilitates better sorting and matching between buyers and vehicles, which would lead to further increases in joint surplus for the entire population of traded vehicles.

3.2 Market Thickness and Internet effect

A market with many more used cars for sale will increase the probability for a buyer to find a better match and, conditional on the matching technology, will increase the number of transactions. We test for this market thickness effect. In addition, one of the main contributions of this paper is to explore whether improvements in matching technology have different effects in markets of varying thickness. That is, does Internet enhance market efficiency more in thin markets or thick markets? There are competing hypotheses here. In thin markets it may be more difficult for buyers to find a car that matches their tastes, and the presence of Internet may provide a more efficient search technology. On the other hand, thick markets may present buyers with choice overload.[23] Were this to be the case, Internet may provide a superior matching technology that enables consumers to distill a vast number of choices, both increasing the probability of a transaction occurring as well as improving the quality of the match.

We revisit the specification in equation (2) to examine the role of market thickness. In our context, market thickness is measured as the number of cars of a given make/model/model-year (e.g. 1997 Honda Accord) registered in a county in a given year. Conditioning on market thickness thus requires adding the this measure of the vehicle stock as a right-hand side variable. However, we cannot both do this and use resale ratio as the dependent variable without embedding a structural endogeneity problem: the stock variable would then appear both in the denominator on the left hand side as well as on the right hand side as an explanatory variable. There is no valid instrument to correct such a problem. So, instead we use a slightly different specification that uses the log of vehicles sold as the dependent variable:

$$y_{jct} = \gamma_0 \ln S_{jct} + \gamma_1 Internet_{ct} + \gamma_2 (lnS_{jct} * Internet_{ct}) + \beta D_{ct} + \alpha_c + \alpha_t + \alpha_j + \epsilon_{jct} \qquad (2)$$

[23]Consumers facing a large number of alternatives may in fact become less likely to make a purchase than if their choice set is more limited (Iyengar and Lepper 2000).

This is similar to equation (2), but with additional regressors to account for the log of stock (S_{jct}), and its interaction with Internet penetration rate. The variable S_{jct} measures the log of vehicle registration counts for each make/model/model-year available at the beginning of period j in county c. Its inclusion reflects the fact that, in the used car market, the available stock of each car type is predetermined.

A somewhat subtle endogeneity concern leads us to pursue an instrumental variable approach to estimation. Because cars can be traded across the borders, and the stock variable includes the flows of cars traded, stock is endogenous if time-varying unobservables are correlated over time. For example, an unobservable demand shock that increases demand for a particular type of car in county c may lead that car to be imported from nearby counties. This increases the stock and, in turn, sales of that car in the importing county (c) next period. To address this issue, we use two different sets of instruments. The first exploits the fact that in aggregate the stock of used cars cannot increase. With that in mind, we instrument for the stock of model j in county c at time t using the sum of all type-j cars in California at the time the car was released as a new vehicle multiplied by the share of the population in county c in 1990.

Supply shifters provide the second set of instruments. These are available due to the fact that important material inputs to production are traded on large global markets, and the stock of vehicles today is influenced by production decisions in the past. Hence, we use the interaction of the price of aluminum in the production year, multiplied by the weight of the car and by the size of the car, as a plausibly-exogenous supply shifter. Notice that while the input prices do not themselves vary at the county-year level, their interaction with car characteristics generates the necessary variation (due to county-year variation in fleet composition). Finally, we use the exchange rate between the dollar and the currency of the origin countries for the imported cars (which, again, will vary at the county-year level based on heterogeneity in the composition of the vehicle fleet between counties). These instruments are jointly significant, and have strength in the first stage. Table (3) reports the results from a first stage regression and the relative F-test.

The OLS and 2SLS estimates of Internet on market thickness are displayed In Table (4). There one can see that (after using IV to correct for endogeneity of the registration count variable) the elasticity of quantity sold with respect to registration count is greater than one. This implies that an increase in the vehicle stock more than proportionally increases sales, which is consistent with the results from Gavazza (2010) even in a market where transactions of used goods are due to quality depreciation rather than productivity shocks. The significant coefficient on the square of log-registration also reveals a nonlinear relationship between market thickness and sales.

Our parameters of interest (Internet and Internet interacted with log-registration) reveal a pattern that supports the hypothesis that consumers in thick markets may be hindered by choice overload. When Internet enters linearly and without interaction, its coefficient implies a sales elasticity of approximately 0.23. Thus, a doubling of Internet penetration ought to increase sales by 23 percent. However, most of this effect is transmitted through an interaction with market thickness. Internet has an increasing effect on transaction volume at higher levels of registration counts. Notice that while the coefficient on Internet level term is negative, the effect of Internet at the lowest level of stock (10 units) is still positive.

3.3 Internet Effects by Population Density and Level of Urbanization

Finally, we investigate the channels through which Internet penetration operate, specifically looking at the differential effect of Internet across counties with different levels of population density and urbanization. The idea is that the new technology can have heterogeneous effects in different environments.

To investigate whether the Internet penetration rate effects urban and rural locations differently, we use Urban Influence Codes (UIC) created by the Economic Research Service, a division of the U.S. Department of Agriculture. The urban influence codes form a classification scheme that distinguishes metropolitan and non-metropolitan counties by the size of their largest city or town, level of urbanization and proximity to other metropolitan and micropolitan counties. Metropolitan counties are further classified as "large" if they have at least 1 million residents and "small" metropolitan counties have fewer than 1 million (but more than 50,000). "Micropolitan" counties are an additional classification of urban areas. These are based around a core city or town with a population of 10,000 to 49,999. California is a state with many highly populated regions, as well as large areas of agricultural and relatively low-density land. This heterogeneity makes California an appealing setting in which to examine the differential effect of Internet across areas of differing population density.

We divide the counties into 2 groups using an aggregated version of UIC: (1) metropolitan counties and surrounding areas (2) all other counties. We treat 38 counties in our sample as "urban". These consist of large and small metropolitan counties and micropolitan adjacent to large metropolitan counties. The remaining 19 counties are either micropolitan, or small metropolitan counties adjacent to a micropolitan county (but not adjacent to a large metropolitan county). Table (5) presents the results of specifications similar to equation (2), but with an additional urban dummy variable entering as a control.[24] The less-urbanized counties are the reference group in these specifications.

From column 1 we see that urban counties exhibit higher volume of trade than less urban counties, even after conditioning on all previously-mentioned covariates. Column 2 allows us to quantify any incremental effect that urban has on the effect of Internet as a facilitator of trade. The coefficient on Internet interacted with the urban indicator is positive and different from zero with 99 percent confidence. This result suggests that Internet improves the efficiency of the second-hand market more in urban and urban-influenced locations than in the more isolated locations. This effect was previously being subsumed in the Internet-Registration interaction term in Table (4) column 4, the coefficient on which is two full standard deviations smaller when the urban variables are included.

To put these effects in context, over the decade spanned by our data (and assuming the mean

[24]In all specifications in Table (5), we once again instrument for the registration count variable as in the specifications above.

changes in Internet penetration over those years), urban counties would in 2007 would exhibit 12 percent more used car transactions that their counterparts.[25] These findings seems to suggest that the area's geographic context has a significant role in amplifying the effect of Internet in improving market efficiency. Craigslist is an example of this process. In metropolitan areas, the socio-economic environment allowed for early development of platforms that took advantage of the availability of Internet technology which facilitates the search process. Craigslist sites were gradually added in different markets. It was first launched in San Francisco Bay Area in 1995, and then rolled out to other major cities such as Los Angeles, Sacramento and San Diego before diffusing more broadly across the State.

4 Conclusion

In this study, we present estimates that quantify the effect that Internet proliferation has had on the efficiency of the used car market in California. Counties exhibiting large increases in Internet penetration also show significant increase in volume of trade in the used car market. After controlling flexibly for an array of potential confounders including aggregate time-specific effects, time-invariant county characteristics, and several time-varying county characteristics, a strong, causal Internet effect remains. On average, Internet increased the resale ratio of used cars in California by 7.2 percent (roughly 120,000 cars annually) in 2007 relative to levels of Internet penetration from a decade earlier.

We remain intentionally agnostic about the mechanism driving these effects. However, three channels stand out as plausible explanations: search costs, match quality, and asymmetric information. It is likely that the cumulative effect that we measure is comprised of a combination of improvements on these dimensions. Given the large degree of heterogeneity in preferences and product

[25]The change in mean Internet penetration was 27.9 percentage points. Multiplying this by 0.432, the Internet*Urban coefficient from Table (5), implies the 12 percent urban differential.

attributes, the search and matching process between buyers and sellers is costly. Internet has the potential to improve market efficiency by reducing the cost of search and by improving the quality of matches. That Internet provides more benefit (all else equal) in urban areas than rural is intriguing. This is consistent with buyers in densely-populated counties facing a very large menu of choices, where Internet may aid as a screening device.

Another plausible explanation is that Internet helps to eliminate asymmetries in information. Websites such as Carfax.com provide vehicle-specific histories that inform buyers of payoff relevant past events that have traditionally only been available to the sellers (e.g. accident reports and maintenance records). Sellers also voluntary disclose private information online (for example, via text or photos, as described in Lewis 2011). Whether due to the presence of third-parties or direct sharing of information between buyers and sellers, the Internet has virtually equalized access to information. It is thus likely that a substantial portion of our estimated effects are from a decrease in the adverse selection problem.

The magnitude of the estimated efficiency gains due to decreased frictions are substantial. Joint surplus increases in 2007 by $25 per unit traded as compared to 1997, and the state-wide gain is on the order of $ 43 million annually. If similar efficiency benefits were to accrue to other sectors of the economy, the overall efficiency gains from Internet would be even larger.

The benefits described above are likely to be largely external to the households that purchase Internet access, and thus a case for policy intervention naturally follows. In the case of search and matching, it is well-known that many of the benefits are general equilibrium in nature – largely in the form of inducing a higher number of participants to enter as buyers and sellers. Already, subsidies for broadband Internet deployment are hotly debated. This is particularly true of subsidies for rural areas. To the extent that external benefits are incorporated into the cost-benefit calculation, our study implies that the marginal subsidy may produce larger efficiency gains in urban areas.

References

[1] Anderson, S. and V. Ginsburgh, “Price discrimination via Second-Hand Markets”, *European Economics Review*, Vol. 38 (1994), pp 23-44.

[2] Berry S., J. Levinsohn and A. Pakes. “Automobile Prices in Market Equilibrium.” *Econometrica*, Vol. 63 (1995), pp. 841–890.

[3] Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman, “What’s Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment”, *The Quarterly Journal of Economics*, Vol. 125 (2010), pp. 263-306.

[4] Boatwright, P., and J. Nunes, “Reducing Assortment: An Attribute-Based Approach,” *Journal of Marketing*, Vol. 65 (2001), 50-63.

[5] Brown, J., and A. Goolsbee, “Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry,” *Journal of Political Economy*, Vol. 110 (2002), 481–507.

[6] Brynjolfsson, E., Y. Hu, and M. Smith, “Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety,” *Management Science*, Vol. 49 (2003), 1580–1596.

[7] Chen, J., S. Esteban, and M. Shum, “How Much Competition is a Secondary Market?” Working Paper (2011).

[8] Dana, J. and E. Orlov. “Internet Penetration and Capacity Utilization in the US Airline Industry,”Working Paper (2009).

[9] Ellison, G., and S. Ellison, “Lessons about Markets from the Internet,” *Journal of Economic Perspectives*, Vol. 19 (2005), 139–158.

[10] Gavazza, A. “The Role of Trading Frictions in Real Asset Markets”, *American Economic Review*, Vol. 101 (2011), pp. 1106-43.

[11] Gavazza, A., A. Lizzeri, and N. Roketskiy,“A Quantitative Analysis of the Used Car Market”, Working Paper (2012)

[12] Ghose, A., M. Smith, and R. Telang, "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research*, Vol. 17 (2006), 3–19.

[13] Hendel, I. and A. Lizzeri. "Adverse Selection in Durable Goods Markets." *American Economic Review*, Vol. 89 (1999), pp. 1097-1115.

[14] Hideo, K. and Sandfort, M. T. "Existence of stationary equilibrium in the markets for new and used durable goods." *Journal of Economics Dynamics & Control*, Vol. 26 (2002), pp. 1029-1052.

[15] Lewis, G. "Adverse Selection in Durable Goods Markets." *American Economic Review*, Vol. 101 (2011), pp. 1535–1546.

[16] Orlov E., "How Does the Internet Influence Price Dispersion? Evidence from the Airline Industry", *The Journal of Industrial Economics*, Volume 59 (2011), pp. 21-37.

[17] Porter, R. and P. Sattler "Patterns of Trade in the Market for Used Durables: Theory and Evidence", mimeo

[18] Rust, J, "Stationary Equilibrium in a Market for Durable Assets." *Econometrica*, Vol. 53 (1985), pp. 783-805.

[19] Schiraldi, P. and F. Nava, "Resale and Collusion in a Dynamic Market for Semidurable Goods", *Journal of Industrial Economics*,Volume 60 (2012), pp. 187–345.

[20] Schiraldi, P. "Automobile Replacement: A Dynamic Structural Approach", *Rand Journal of Economics*, Vol. 42 (2011), pp. 266-291.

[21] Scott Morton, F., "Consumer Benefit from Use of the Internet," *NBER Innovation Policy and the Economy*, 6 (2005), 67–90.

[22] Stolyarov, D. "Turnover of Used Durables in a Stationary Equilibrium: Are Older Goods Traded More?" *Journal of Political Economy*, Vol. 110 (2002), pp. 1390-1413.

[23] Zettelmeyer F., F. Scott Morton, and J. Silva-Risso, “How the Internet Lowers Prices: Evidence from Matched Survey and Automobile Transaction Data”, *Journal of Marketing Research*, Vol. 43 (2006), pp. 168-181.

Table 1: Summary Statistics of County Means

	1997				2003				2007			
	Mean	StDev	Min	Max	Mean	StDev	Min	Max	Mean	StDev	Min	Max
Registration Count	99,404	239,674	95	1,651,633	139,189	313,614	161	2,145,228	170,791	388,698	216	2,676,614
Volume of Trade	14,970	34,359	54	235,058	27,524	62,128	115	429,200	29,704	66,824	117	465,193
Trade Volume Ratio	0.25	0.18	0.04	1.00	0.22	0.13	0.11	0.99	0.20	0.11	0.12	0.81
Number of Car Models	643	292	49	1,052	1,050	422	98	1,583	1,237	536	108	1,960
Internet Penetration	0.14	0.05	0.04	0.25	0.38	0.07	0.22	0.55	0.42	0.08	0.24	0.64
Income	37,283	8,662	23,359	60,967	43,449	10,383	28,533	68,223	52,961	13,314	33,576	84,265
Unemployment Rate	0.09	0.05	0.03	0.27	0.08	0.03	0.05	0.16	0.07	0.02	0.04	0.18
Median Age	37.0	3.8	29.0	50.0	38.2	4.2	27.3	48.5	39.7	4.0	32.0	51.0
Fuel Price ($/gallon unleaded)	1.46	0.09	1.28	1.63	1.93	0.08	1.81	2.23	3.13	0.10	2.97	3.44

Notes: This table displays means of county-level variables for the state of California. All automobile statistics refer to used cars only (no new cars or light/heavy-duty trucks are included). Number of car models refers to the make/model/model-year level. Sources: RL Polk, Current Population Survey, Oil Price Information Service.

Table 2: Internet on Resale Ratio

	Resale Ratio				
	1	2	3	4	5
Internet	0.024	0.036	0.027	0.063**	0.044**
	(0.030)	(0.033)	(0.040)	(0.026)	(0.018)
ln(Population)		-0.003	-0.008*	-0.105**	-0.109**
		(0.003)	(0.004)	(0.050)	(0.049)
Median Income		-0.063**	-0.047*	-0.012	-0.021
		(0.025)	(0.026)	(0.027)	(0.042)
Mean Unemployment		0.000	-0.001	-0.001*	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Average Age Head-of-HH		-0.002	-0.002	-0.001	-0.001
		(0.001)	(0.001)	0.000	(0.001)
Fuel Price		-0.042***	-0.165***	-0.033***	-0.09
		(0.006)	(0.058)	(0.005)	(0.081)
County F.E.	No	No	No	Yes	Yes
Year F.E.	No	No	Yes	No	Yes
Make/Model/Age/Modelyr F.E.	Yes	Yes	Yes	Yes	Yes
N	213,700	213,700	213,700	213,700	213,700
R-Square	0.001	0.159	0.167	0.231	0.232

Standard errors in parentheses, clustered at the household level. * Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level.

Table 3: 1st Stage of 2SLS IV

	Registration Counts	(Registration Counts)^2
Log Cumulative Stock of each make/model x Share of Population of Counties (1990)	0.613***	-0.257***
	(0.002)	(0.010)
Price Aluminium *Weight	0.006	0.200***
	(0.004)	(0.023)
Exchange Rate	-0.063***	-1.256***
	(0.014)	(0.088)
Suare of (Log Cumulative Stock of each make/model x Share of Population of Counties	0.041***	0.986***
	(0.000)	(0.001)
Square of (Price Aluminium *Weight)	-0.042***	0.070**
	(0.005)	(0.031)
Square of (Exchange Rate)	-0.002***	-0.033***
	(0.000)	(0.003)
N	478,940	478,940
R-Square	0.866	0.917

Standard errors in parentheses. * Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level.

Table 4: Internet on Market Thickness

	ln(Quantity Sold)			
	1	2	3	4
	OLS	IV	OLS	IV
ln(Registration)	0.855***	1.122***	0.824***	1.073***
	(0.020)	(0.031)	(0.025)	(0.032)
ln(Registration)^2	0.009***	-0.007**	0.009***	-0.007**
	(0.002)	(0.003)	(0.002)	(0.003)
Internet	0.227**	0.236**	-0.158	-0.379***
	(0.094)	(0.099)	(0.105)	(0.112)
Internet*ln(Registration)			0.110***	0.172***
			(0.027)	(0.024)
ln(Population)	-0.323*	-0.438**	-0.303	-0.406**
	(0.187)	(0.201)	(0.187)	(0.200)
Income	-0.328	-0.314	-0.404**	-0.435**
	(0.205)	(0.214)	(0.203)	(0.210)
Unemployment	-0.008	-0.005	-0.011**	-0.010*
	(0.005)	(0.005)	(0.005)	(0.005)
Average Age Head-of-HH	-0.003	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Fuel Price	-0.454*	-0.458*	-0.652**	-0.769***
	(0.266)	(0.263)	(0.260)	(0.243)
County F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Make/Model/Age/Modelyr F.E.	Yes	Yes	Yes	Yes
N	180,376	180,376	180,376	180,376
R-Square	0.870	0.867	0.870	0.867

Standard errors in parentheses, clustered at the household level. * Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level.

Table 5: Internet on Urban/Rural

	ln(Quantity Sold) w/ Urban Split	
	1	2
	IV	IV
ln(Registration)	1.073***	1.081***
	(0.032)	(0.033)
ln(Registration)^2	-0.007**	-0.007**
	(0.003)	(0.003)
Internet	-0.379***	-0.650***
	(0.112)	(0.108)
Internet*ln(Registration)	0.151***	0.126***
	(0.027)	(0.026)
Urban	1.400**	1.461**
	(0.625)	(0.640)
Internet*Urban		0.432***
		(0.125)
ln(Population)	-0.406**	-0.458**
	(0.200)	(0.211)
Income	-0.435**	-0.539***
	(0.210)	(0.201)
Unemployment	-0.010*	-0.011**
	(0.005)	(0.005)
Average Age Head-of-HH	-0.003	-0.003
	(0.002)	(0.002)
Fuel Price	-0.367	-0.352
	(0.283)	(0.278)
County F.E.	Yes	Yes
Year F.E.	Yes	Yes
Make/Model/Age/Modelyr F.E.	Yes	Yes
N	180,376	180,376
R-Square	0.867	0.867

Standard errors in parentheses, clustered at the household level. * Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level.

Table 6: Implied Increase in Joint Surplus

Demand Elasticity	Supply Elasticity				
	5	4	3	2	1
-5	213	240	284	373	640
-4	240	267	311	400	666
-3	284	311	355	444	711
-2	373	400	444	533	800
-1	640	666	711	800	1,066

Increase in joint surplus per used car traded associated with an increase in volume of trade of 100,000. Calculated at the 2007 California average used car price of $7,508 for a range of demand and supply elasticities.

Table 7: Annual Welfare Gain

Demand Elasticity	Supply Elasticity				
	5	4	3	2	1
-5	10,661,360	11,994,030	14,215,147	18,657,380	31,984,080
-4	11,994,030	13,326,700	15,547,817	19,990,050	33,316,750
-3	14,215,147	15,547,817	17,768,933	22,211,167	35,537,867
-2	18,657,380	19,990,050	22,211,167	26,653,400	39,980,100
-1	31,984,080	33,316,750	35,537,867	39,980,100	53,306,800

Annual welfare gain associated with an increase in volume of trade of 100,000. Calculated at the 2007 California average used car price of $7,508 and a range of demand and supply elasticities.

Figure 1: Internet Penetration by County, 1997-2007