Internet and the Efficiency of Decentralized Markets: Evidence from Automobiles[1]

David Rapson[2] and Pasquale Schiraldi[3]

April 14, 2013

Abstract

We estimate the 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. Additional gains may accrue when externalities are associated with the rate of vehicle fleet turnover (e.g. with respect to greenhouse gas emissions from fuel economy standards).

JEL: D1, D8, L9

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

[2]Department of Economics, University of California at Davis, One Shields Ave, Davis CA 95616; Email: dsrapson@ucdavis.edu

[3]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE; Email: 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)[4], 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.[5] 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.[6]

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 information is costly to acquire and often asymmetric, with potential buyers trying to assess unobserved quality that is known to the seller.[7] Over the last 15 years, several Internet-based resources

---

[4]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.

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

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

[7]Porter and Sattler (1999) and Gavazza et al (2012) report transaction costs in the range of 15-50% of

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 (Zettelmeyer et al 2006). 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 has all but eliminated the information gap between buyers and sellers. Moreover, the possibility of gathering more information and reaching a broader number of counterparties has substantially increased the surplus generated upon trade.

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. The availability of more information and a broader market reduces search costs and favors better matches (i.e. generates matches with higher joint surplus). Each of the potential mechanisms by which Internet would effect a decentralized market–decreased search costs, improved match quality, and less adverse selection–would be similar in at least one important way: they would each increase volume of trade.

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: (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.

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

_____

the trade-in value of used cars.

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.

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. From the R.L. Polk dataset, for each make-model-year we observe the stock and transaction of all cars of each type in each time period t.

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. 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.[8] CPS is weighted to be representative at the state (not county) level, raising the possibility of sampling error affecting smaller counties. We implement robustness specifications (described below) to check for this, and find no evidence that sampling error affects our numerical results.

Table (1) provides summary statistics of the variables used in the empirical analysis.[9] 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 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. 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. Our empirical challenge is to control for these

---

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

[9]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).

potential confounders to separately identify the treatment effect of interest.

3   Empirical 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. If goods depreciate every period, consumers will never hold onto their durables for more than one period, instead updating to their preferred car quality.[10] Hence, lowering market frictions should induce a higher 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 period. In a world without frictions, this ratio should be equal to one. 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, allowing price to vary on the transition path (Ie. omitting it as a right-hand-side variable).

We 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 diffusion of Internet requires us to control for some important potential confounders. We exploit the panel structure of the data by employing a difference-in-differences specification and non-parametric controls for county and time fixed-effects. Regressors also include observable demographics that vary at the county-year level, such as average income, the unemployment rate, age distribution in the population, and the population level itself. We also include the average price of gasoline to control for the possibility that changes in fuel price affect the volume of trade. The fixed effects control for any residual time-invariant county unobservables and aggregate shocks shared by all

---

[10]See Rust 1985, Hideo and Sandfort 2002, and Schiraldi 2011.

locations in a given year. Identification is achieved from within-county variation in transaction volume and Internet penetration over time, net of controls.

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

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

Subscript c denotes county, t denotes year, and j denotes model type.[11] 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 the Internet penetration rate, Z_{ct}, 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.[12] In every specification, standard errors are clustered at the county level.

We estimate several OLS specifications of equation (1), using the resale ratio as well as count of used cars sold by model as the outcome variable of interest. 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

---

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

[12] To address concerns about CPS sampling variation at the county level, we re-estimate eq. (1) on a sample which excludes the smallest ten percent of counties. Point estimates of this robustness check are identical to those reported in Table (2), indicating that sampline error is not biasing our results.

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 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.

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. 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 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.

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 and diminished 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. This amount represents a lower-bound of the total welfare gain because it does not include surplus generated by the realization of additional transactions or superior sorting and matching between buyers and vehicles. Further, some benefits are likely to be external to the market. For example, more rapid turnover of the vehicle fleet provides benefits in the form greenhouse gas emissions abatement from fuel economy standards, which apply only to new vehicles.

References

[1] 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.

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

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

[4] 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.

[5] 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.

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

[7] 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.

[8] Porter, R. and P. Sattler "Patterns of Trade in the Market for Used Durables: Theory and Evidence", NBER Working Paper 7149 (1999).

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

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

[11] 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.