



# **MIT Sloan School of Management**

**Working Paper 4271-02  
December 2002**

## **Home Buyer Search Duration and the Internet**

Victoria T. D'Urso

© 2002 by Victoria T. D'Urso. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full credit including © notice is given to the source.

This paper also can be downloaded without charge from the  
Social Science Research Network Electronic Paper Collection:

[http://ssrn.com/abstract\\_id=360460](http://ssrn.com/abstract_id=360460)

# HOME BUYER SEARCH DURATION AND THE INTERNET

VICTORIA T. D'URSO

**ABSTRACT.** In this study we examine the impact of internet use on the duration of search in the housing market. We develop a model of partial equilibrium in the housing market which suggests an ambiguous effect on the search duration when internet resources are employed. In this model, the impact of using the internet can be viewed as increasing the search efficiency, or as altering the distribution of potential matches from which the home buyer can choose. We use data from the 2000 Home Buyer and Seller Survey collected by the National Association of Realtors. While theory suggests there might be an increase or a decrease in search times when using on-line resources in the search, in this data using an Instrumental Quantile Regression approach we find a tendency for internet use to increase the duration of home search relative to employing more conventional search methods.

**KEY WORDS:** Sequential Search, Housing Market, Duration, Instrumental Quantile Regression

---

*Date:* December 5, 2002.

©2002 Victoria T. D'Urso. Contact [vickyd@mit.edu](mailto:vickyd@mit.edu). The author would like to thank William Wheaton for useful suggestions and help with the theoretical part of this work. Special thanks to Victor Chernozhukov for many helpful comments and suggestions and to him and Christian Hansen for sharing computer programs contributing to the IQR estimation and inference in this work. Thanks to Mark Calabria at the National Association of Realtors for his generous help and support in obtaining the Buyer and Seller Survey data used in this work. This work was supported in part by NSF grant number IIS-0085725.

## 1. INTRODUCTION

The internet, what we today accept as the implementation of a shared body of information, available at the fingertips of those equipped with a computer and a connection, has been operating since 1979. The influence of the internet has increased tremendously since then and has become a valuable, if not an essential component of life in the US today. While the exact amount of importance attributable to the internet may be open for debate, one thing is certain: in recent years the internet has become an increasingly useful tool and source of information for buyers in a variety of markets. The addition of the internet as a resource for comparison of goods and services and as a medium for business transactions has prompted researchers to examine the internet's impact on traditional markets [1]. The internet's far reaching, and unprecedented impact is quickly, and justifiably so, becoming the focus of an ever increasing body of economic research.

The change in the economic landscape brought on by the impact of the internet's presence on the way business is conducted has led to work such as Brown and Goolsbee [2]. In their study, the authors investigate the presence of internet markets as they relate to non-sequential search in the insurance market. As a counterpart to their line of investigation we ask the question about the relationship between the internet and markets where consumers choose offers in a sequential fashion.

In this study we examine the impact of on-line resources for search in the housing market. The housing market is a natural choice for this investigation as it is one of the largest markets in the US where the search occurs in a sequential manner. A home buyer seeking to purchase a home must decide, as offers arrive, whether to take the current offer or leave it expecting a subsequent, better match to arrive. In the latter case, the previous offer cannot be held on to, while the home buyer keeps looking for a better match.

Brown and Goolsbee [2] concentrate their work on the impact of the internet on insurance prices. In our study, we focus instead on the impact of on-line search on the duration until a home (which is eventually purchased) is located by the home buyer. In particular, in this study we ask the question: does the use of internet resources in home search generally increase or decrease the time it takes to find a house to purchase.

Building on the standard models of sequential search [3, 4] found in the literature, we develop a model of sequential search in the housing market that incorporates internet use. Pre-existing search models dictate that the use of on-line resources

in addition to traditional methods of search reduce the costs of search. When one searches on-line, it takes less time to learn about the choices offered, their location, features and amenities. In our model, rather than changing the costs of search, using the internet as part of the search effort acts to increase the arrival rate of offers or to increase the number of available choices to the home buyer. The internet brings a wider selection of houses to the home buyer to be viewed, bid on, and ultimately purchased. Our model suggests that the use of internet resources as part of the search process in this market has an ambiguous effect on the duration of search.

In our empirical analysis of the relationship between home search durations and the use of the internet in search for a new home, we use data from the 2000 Home Buyer and Seller Survey conducted by the National Association of Realtors (NAR). The survey includes data on duration of home search, various ways of using the internet as part of the search and some demographic characteristics of the home buyers. Since the survey data we use in this study only includes information about individuals involved in home search and their particular level of internet use for the purposes of this search, we are concerned about individual heterogeneity which may be driving internet use and influencing the speed with which individuals in our data locate a home. In order to control for the possible endogeneity in the data we use a simulated instrument for internet use. Data from the 2000 Current Population Survey (CPS) Supplement on Computer Ownership and Internet Use, conducted by the Bureau of Labor Statistics and The US Census Bureau, was used to construct a predicted internet use level by computing the mean internet use in the CPS sample in each age and income group available in the main Home Buyer and Seller Survey data.

We estimate the effect of on-line resources using instrumented quantile regression approach, recently developed by V. Chernozhukov and C. Hansen, [5]. This framework allows for a flexible specification and a detailed analysis of the treatment effect of internet use at different points in the distribution of search durations. After instrumenting for internet, the quantile treatment effect in the data from the NAR survey used here suggests that search durations are likely to be longer when employing the internet as part of the home search relative to search with conventional methods. The treatment effect of internet use in the housing market search is largest near the median of the distribution and close to zero in both tails.

## 2. THEORETICAL DISCUSSION

**2.1. Search in the Housing Market.** There are  $H$  households and a fixed housing stock with  $N$  units in the market. We assume there are enough units to house all

the households and there is a vacancy rate  $V$ , as some of the units are not occupied. There are three states in which these households can be located. Matched ( $M$ ) in which the household is satisfied with its current housing choice and is not looking for a home to buy and move to. A matched household can become mismatched ( $S$ ) and search for a new home until it finds a suitable match, at which point the household buys the second home and becomes matched but owning two homes ( $D$ ). When the previous home of a household in state  $D$  is sold, the household returns to state  $M$ . Thus, the total number of households is simply the sum of households located in each state,  $H = H_M + H_S + H_D$ . Households experience a (yearly) match shock probability of  $\beta$  which changes a household in the matched state into a household which is mismatched, and corresponds to a transition rate from a matched to a mismatched state. The magnitude of this shock is  $\alpha$ . The number of households owning two houses is simply the number of units multiplied by the vacancy rate.  $H_D = VN$ . We assume perfect credit markets.

There is a match probability function  $F(X)$ . It corresponds to the quality of offers a mismatched household considers as part of the search. If we assume that the quality of offers a household looks at during the search is a normally distributed random variable  $X$  with mean  $\mu$  and variance  $\sigma$ , then  $F(X)$  is the cumulative distribution function of the above normal. Here  $X$  indexes how closely a household's preferences are matched to each particular home. Households have a reservation level  $R$ , below which the household would not accept a given housing choice. The magnitude of the transition shock  $\alpha$  moves the household from utility level  $U_M = \mathcal{U}(R)$  to  $U_S = \mathcal{U}(R - \alpha)$  where  $\mathcal{U}(\cdot)$  is a suitable utility function. While the reservation utility is endogenous to the model, the utility level of a mismatched state is predetermined and does not adjust endogenously, hence the partial equilibrium nature of the model.

Equating the flows in and out of search, in equilibrium we get

$$H_M\beta = \lambda(1 - F(R))H_S,$$

so that the fraction of matched households who experience the transition into mismatched state, that is the flow into search, equals the accept rate of offers  $(1 - F(R))$  multiplied by  $\lambda$ , the search efficiency and  $H_S$  the number of searching households. Another equivalent interpretation of  $\lambda$  is the arrival rate of offers per given period of time. In this model we will decompose the arrival rate of offers into a baseline arrival rate due to search by conventional methods and an arrival rate due to the use of the internet in searching for a suitable match <sup>1</sup>,  $\lambda = li$ .

---

<sup>1</sup>Whether the internet effect  $i$  is modeled as a multiplicative or an additive effect to the baseline arrival rate of offers does not change the results of the model in an important way.

Let

$$q = \lambda(1 - F(R))$$

so that  $q$  is the probability of finding a suitable housing unit in a unit of time and  $1/q$  is the duration of search. In the data used in the empirical estimation of the effect of the internet on housing market search, we observe a search duration equivalent to  $1/q$  in this model. Let

$$z = \frac{\lambda(1 - F(R))H_S}{V}$$

be the probability of sale.

The present discounted value of being in each of the three states, is governed by the standard <sup>2</sup> flow equations:

$$\begin{aligned} rV_M &= U_M - \beta(V_M - V_S), \\ rV_S &= U_S + q(V_D - V_S - P), \\ rV_D &= U_M + z(V_M - V_D + P) \end{aligned}$$

Here,  $V_M$ ,  $V_S$  and  $V_D$  are the present values of each state,  $U_M$  and  $U_S$  are the utility flows of being matched and mismatched, respectively,  $P$  is the market price of a matched house, and  $r$  is the discount rate. The above equations together with the condition that  $V_D - V_S - P = V_M - V_D + P$  allow us to solve for the the price and the present values of being in each state in terms of the utility flows and the parameters of the model. Thus,

$$\begin{aligned} P &= \frac{(U_M - U_S)(2\beta + r + z)}{r(2\beta + 2r + q)}, & V_M &= \frac{(2r + q)U_M + 2\beta U_S}{r(2\beta + 2r + q)} \\ V_D &= \frac{(2\beta + 2r + q + z)U_M - zU_S}{r(2\beta + 2r + q)}, \text{ and } & V_S &= \frac{qU_M + 2(\beta + r)U_S}{r(2\beta + 2r + q)}. \end{aligned}$$

Each household chooses  $R$  to maximize the value of being mismatched. After recalling the definition of  $q$  as a function of  $R$  and imposing a functional form for the utility of a matched state as a function of the reservation  $R$  as well, together with values for the parameters of the model, we can numerically solve for the maximum value of being in a mismatched state. This maximum occurs at  $R^*$ , the value of  $R$  corresponding to the peak of the value of being mismatched.

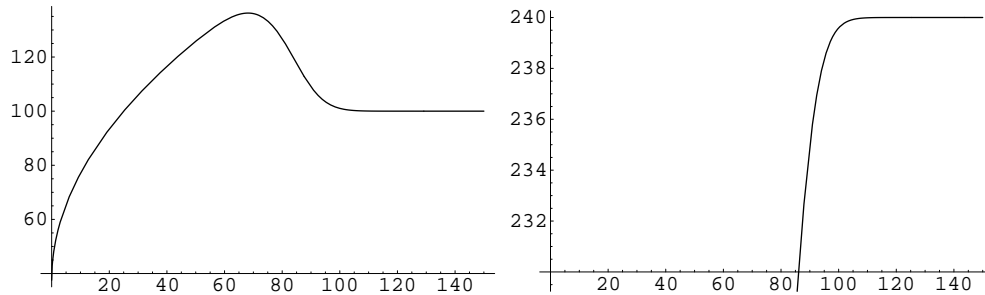
For example, with  $U_M = \sqrt{R}$ , match quality distributed  $N(75, 10)$ , a discount rate of 5%, transition rate  $\beta$  of 10%, search efficiency of 50%, and  $U_S = 5$ , we obtain

---

<sup>2</sup>While a richer model of search in the housing market (see [3]) needs to include the probability of (demographic) transition back to a matched state from a mismatched state, trivially ending search, here adding such a term to the  $rV_S$  equation does not meaningfully alter the results and has been omitted for computational simplicity.

$R^* = 68.1624$  as shown in Figure 1 a). For the above parameter values the value of being mismatched achieves a well defined, unique maximum at  $R^*$ . However, when the magnitude of the transition shock is small, so the drop in utility from a matched to a mismatched state is small, the home buyer is indifferent between housing choices above a certain level (see part b) of the figure).

With a small drop in utility, the cost of remaining mismatched is not sufficient to cause the household to search and move to a new home. Rather, the household will hold out indefinitely for the perfect match. This situation is equivalent in this model's framework to an infinitesimally small accept rate of offers. When the probability of finding a suitable new match in a given period of time,  $q$ , is 0, the value of being mismatched reduces to  $V_S = \frac{U_S}{r}$ . In all further discussion we will assume that the drop in utility is large enough, so that being mismatched is bad enough to require an adjustment of the reservation level to a new, well defined  $R^*$ . In either case, for sufficiently large  $R$ ,  $F(R)$  is 1, and  $V_S$  levels out to  $\frac{U_S}{r}$ . In order to have a well defined, unique maximum for  $V_S$ , we need  $V_S$  evaluated at  $R^*$  to exceed  $\frac{U_S}{r}$ . This condition reduces to  $\mathcal{U}(R^*) > U_S$ .



(a) Well defined  $R^*$ , with large drop in utility when mismatched.

(b)  $V_S$  levels out at  $\frac{U_S}{r}$  for a small drop in utility when mismatched before reaching a maximum.

FIGURE 1. Plot of  $V_S$  (y-axis) vs.  $R$  (x-axis)

**2.2. Internet Use in the Framework of the Model.** The use of on-line resources as part of the search in the housing market enters into this model through two separate channels. First, using the internet as part of the search could simply speed up the arrival of offers, so that one can view the set of available choices in a shorter amount of time, or view a larger number of offers in any given time period. Speeding up the arrival of offers, internet use enters into our model through the parameter  $\lambda$ . However, looking at potential housing choices on-line carries more information than

simply delivering these choices faster. If using the internet in the search delivers a larger set of options, the actual distribution of match qualities might be affected. The additional information about each house available on-line allows the home buyer to rule out unsuitable choices more easily and concentrate the search efforts only on highly suitable choices. Rather than having to spend time and resources driving out to each potential house location to visit, the home buyer is able to substitute visiting the house with viewing it over the internet. Both the mean and the variance of the distribution of seriously considered choices would increase when the internet is used in the search as a substitute for actual visiting of some houses. A larger variety of choices in terms of the match quality can be viewed on-line, increasing the variance of the distribution of choices. In addition, one could choose to visit houses that are much better matches than he or she would have visited had the search been conducted through traditional search methods. Dismissing choices after viewing them on-line that would have been ruled out only after visiting when searching through traditional methods increases the mean of the distribution of potential housing matches. Thus, a second way in which the internet affects search is through increasing the mean and/or the variance of the distribution of choices.

Consider the effect of a change in the parameters of the model on the change in  $R^*$ . The optimum reservation can be written as an implicit function of the parameters,  $\beta$ ,  $r$ ,  $l$ ,  $i$ ,  $\mu$ ,  $\sigma$  and  $U_S$  as a solution to the equation  $\frac{dV_S}{dR} = 0$ . We verify that in fact  $\frac{d^2V_S}{dR^2} < 0$  here, and decompose  $\lambda = li$  to distinguish an internet specific increase in arrival rate. The optimum reservation value decreases with an increase in the transition rate. The more likely a household is to experience the adverse mismatching shock, the less the household holds out for a better match, and thus the lower the optimum reservation. Likewise, with a higher interest rate, the optimum reservation drops, as is to be expected. An increase in  $\lambda$ , and more specifically an increase in the internet portion,  $i$ , leads to an increase in  $R^*$ . Similarly, the optimum reservation level increases with an increase in  $\mu$ ,  $\sigma$  and  $U_S$ . For example, using parameter values as those in the numerical example used above, we see that fixing all but one parameter at a time produces a change in  $R^*$  as shown in Figure 2.

How does an increase in internet use during the search affect the duration of search? First, let's examine the search efficiency effect. Since  $\frac{\partial i}{\partial R^*} > 0$ , when  $i$  increases  $R^*$  adjusts up as well. Recall the definition of the probability of finding a suitable match in a given time period,  $q = (li)(1 - F(R))$ . The first term,  $li$ , increases with  $i$ , but the second term decreases since  $R^*$  adjusts up in response to a higher internet use. The overall effect on  $q$  and therefore on the duration of search  $\frac{1}{q}$  is at least ambiguous. However, looking at the numerical example above, while  $i$  doubles,  $R^*$  increases from



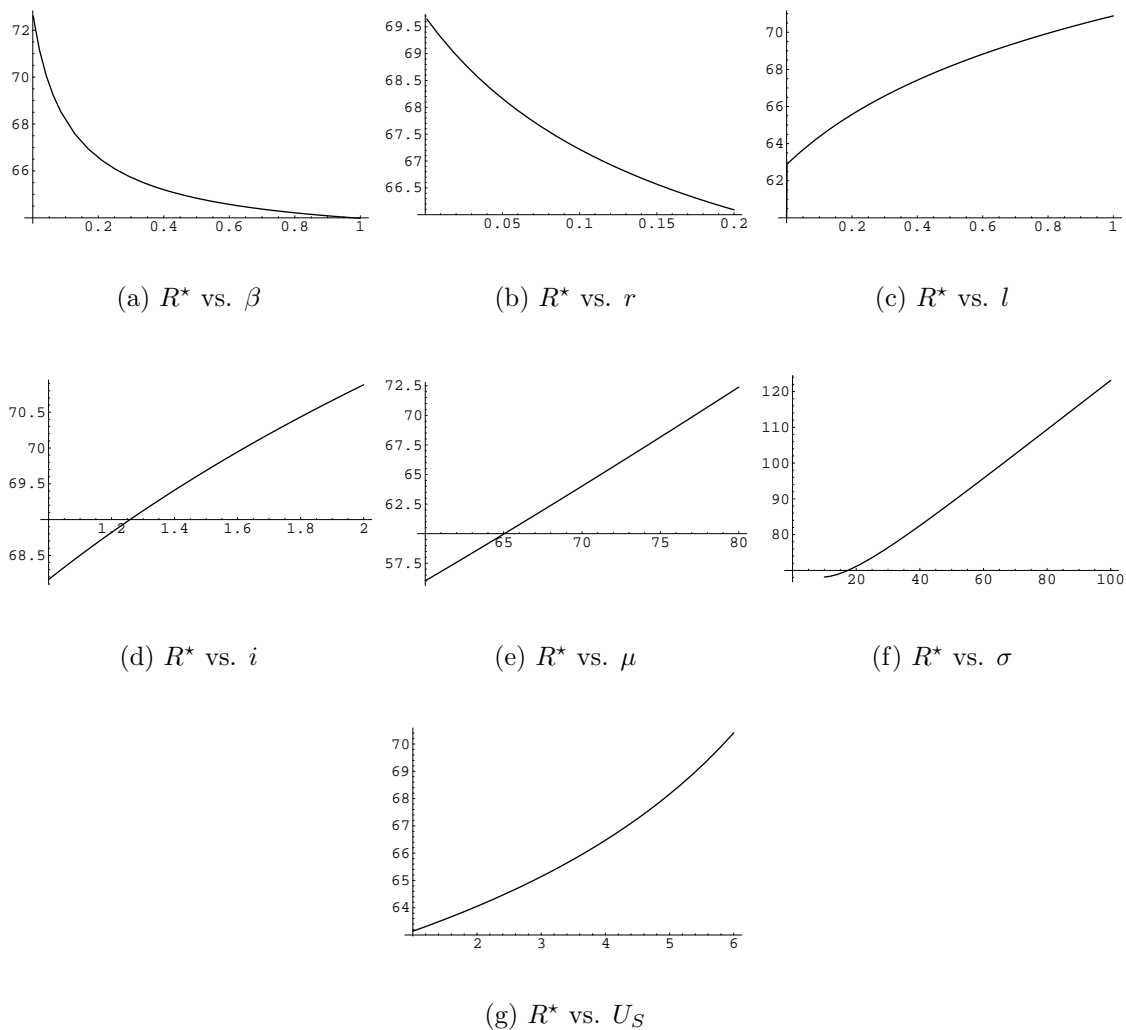


FIGURE 2. Sensitivity of  $R^*$  to model parameters

about 67 to 71, which translates to an increase of about 0.13 in terms of the CDF of  $N(75, 10)$ . Thus, the overall effect on  $q$  is positive, the effect on  $\frac{1}{q}$  is negative, and the increase in internet use, when the internet acts through the search efficiency, should result in a decrease in search times. With this reasonable choice of parameters it is then plausible to conclude that if the internet only acts to increase the arrival rate of offers, search duration is likely to decrease as a result of increased internet use in the housing market search. In the current model any costs associated with search, in terms for example of effort exerted by the potential home buyer in the process enter through this search efficiency parameter.

An increase in the arrival rate of offers is not the only possible channel through which employing the internet in the search process can affect the duration of search. As discussed by T. Malone and his co-authors in [6], as a larger amount of information becomes readily available to the to the buyer through the internet, the structure of the market undergoes a fundamental change. In their work, Malone et. al., do acknowledge the increase of the arrival rate of offers when using the internet through that they call an *electronic communication effect*. It increases the amount of information that can be exchanged between parties in a given amount of time and acts to decrease the costs associated with, in our case, search. This corresponds to the parameter  $\lambda$  in our model.

A different effect of the internet discussed in their work is what they have termed a *brokerage effect*. The internet serves to create an electronic market that “allow[s] a buyer to screen out obviously inappropriate suppliers and to compare the the offerings of many different suppliers”.<sup>3</sup> In the setup of search in the housing market, this effect can be interpreted as the internet acting as a filtering mechanism for offers. Since home buyers are able to input specific characteristics or ranges of features they desire in a home, using the internet can quickly and easily personalize the range of offers available to suit each home buyer. In addition, with the availability of virtual tours, and multiple angle views of the house offers available on the internet, home buyers can immediately rule out choices that they would have at least driven by to look at when searching through conventional methods.

Thus, for any amount of time spent in search, using the internet provides the home buyer with a set of offers that are better suited to the individual home buyer than conventional search methods could provide. This effect translates in our model to a higher mean,  $\mu$ , in the distribution of offers available to each home buyer. In addition, by increasing the number of suppliers the internet acts to increase the overall variety of offers available and thus increase the spread,  $\sigma$  of the distribution of offers available when using the internet in the search.

If the internet acts to change the distribution of the available choices by increasing the mean of the choices or by increasing the variance of the available houses to consider during the search, without increasing the arrival rate of offers, this model predicts an increase of search duration. The logic here is straight forward: when  $\mu$  or  $\sigma$  increase,  $F(R^*)$  increases, and with the absence of change in other parameters, this leads to a decrease in  $q$ , and an increase in  $\frac{1}{q}$ . The search duration unambiguously increases.

---

<sup>3</sup>See [6], p. 488.

In reality, it is likely to expect that the role of the internet is a combination of an arrival rate increase and a shift/spread of the distribution of the potential matches' quality. Whether one effect or the other dominates, is impossible to distinguish through theory. Thus, the remaining of this study focuses on the empirical effect of internet use on search durations. By empirically determining whether the increase in internet use leads to shorter or longer search times, we can then distinguish whether the internet mostly functions to increase the arrival rate of offers, or to mostly change the underlying distribution of offer qualities available to a home buyer in the housing market.

### 3. ESTIMATION STRATEGY

A standard framework for analysis of duration data has typically been a hazard model specification, frequently in a strict parametric form, see [7]. A parametric regression framework assumes that

$$z(T_i) = x_i' \beta + \sigma \epsilon_i,$$

where  $T$  is the duration of search and  $x_i$  is the vector of covariates for the  $i$ th observation,  $z(\cdot)$  is a transformation function, and  $\beta$  and  $\sigma$  are unknown parameters. The random variable  $\epsilon$  is zero mean, unit variance, and its density function,  $f$  is independent of  $x$ . Here,  $f$  can be any of a number of standard density functions (Gaussian, log-normal, Weibull, exponential, etc.) giving rise to a variety of proportional hazard models initially proposed by Cox [8]. For example the Weibull or the exponential density function give rise to the Accelerated Time Failure model with the specification

$$\ln(T_i) = x_i' \beta + \sigma \epsilon_i.$$

There is a major problem associated with the proportional hazard models for estimation relevant here. The above assumes that only the conditional mean  $z(T)$  depends on the covariates. This implies a constant effect of treatment across all points in the distribution of durations. In other words, treatment produces a simple locational shift in the distribution of search times. This assumption is not likely to hold here: it is unlikely that the internet would act to shift each search time by the exact same fixed amount. It is possible, and is not to be overlooked, that using the internet in the home search results has a bigger impact on the search duration at some lengths of search and a smaller, or perhaps even the opposite effect at different points in the distribution of search times. To allow of a more general change in this distribution, a different estimation model is appropriate here. In addition, as  $f$  is explicitly specified, the proportional hazard model can be estimated using Maximum

Likelihood. This produces optimal estimators if the model is specified correctly, but is vulnerable to specification error<sup>4</sup>. As addressed by [7], the Quantile Regression (QR) model is robust to misspecification of the underlying hazard.

**3.1. A Quantile Regression Approach.** A quantile regression estimates a conditional quantile function. The idea behind this technique is analogous to the traditional ordinary least squares regression where one solves

$$\min_{\mu \in \mathfrak{R}} \sum_{i=1}^n (y_i - \mu(x_i, \beta))^2$$

as an estimate of the conditional sample mean,  $E(Y|x)$ . The median regression obtains an estimate of the conditional sample median by minimizing the sum of absolute values of the residuals. This minimization problem can be generalized to estimate conditional quantiles other than the median. That is, solve

$$\min_{\mu \in \mathfrak{R}} \sum_{i=1}^n \rho_{\tau}(y_i - \xi(x_i, \beta)),$$

where  $\rho_{\tau}$  is the absolute value function for  $\tau = .5$  and is a “tilted absolute value” function for other values of the quantile index  $\tau \in (0, 1)$  as illustrated in [9]. The more general  $\rho_{\tau}$  allows estimation of conditional quantiles other than the median and generalizes the median regression to a quantile regression for any quantile index  $\tau$ . This technique of estimating a conditional quantile function is different than subsetting the sample and estimating each section of the unconditional distribution, as such truncation on the dependent variable would yield incorrect results. Again see [9]. Here, all observations are used in determining the regression fitting of each quantile.

A quantile regression approach allows for a greater flexibility in the underlying distribution of search durations as they get affected by internet use. By performing a median rather than a mean regression and then for each quintile, decile, or in general terms for each quantile of observations in the data we can map out the effect of the internet for each portion of the search duration distribution. An instrumental variable technique may still be warranted in the quantile regression analysis setting to correct for any individual heterogeneity present when one uses the internet in the home search. We present ordinary QR results and then employ Instrumental Variable Quantile Regression (IQR) technique developed in a recent work by V. Chernozhukov and C. Hansen, [5].

---

<sup>4</sup>Results from an estimation of the conditional mean using a Cox proportional hazard model showed no statistical significance in our data

Following their work, let search duration outcome be denoted

$$Y_d = q_d(X, U_d)$$

in the two states of the world with  $d \in \{0, 1\}$  where  $d$  is an indicator for internet use as part of the search,  $X$  is a vector of observable covariates, and  $U_d$  is unobservable individual heterogeneity such as quality concern or pickiness when choosing a house. The individual decision to use the internet (or not) in the search is in general

$$D = 1(\varphi(Z, X, V) \geq 0)$$

so the unobserved vector  $V$  could depend on unobservables such as the pickiness  $U_d$  producing endogeneity in the model.

This model requires the assumption that conditional on  $(Z, X, V)$ ,  $U_0$  and  $U_1$  are equal in distribution, that is, that people decide to use the internet (or not) in their search without knowing how picky they are in their housing choice relative to other, observationally same home buyers. This is less restrictive than the usual assumption of identical  $U_0$  and  $U_1$ . Another relevant relaxation of a usual assumption afforded by this model is that it allows for an arbitrary correlation between the instrument  $Z$ , and the error  $V$ . Such a correlation is not allowed in other settings such as 2SLS. However, as our instrument for internet use when buying a home is a measure of predicted general internet use, that is mean internet use in each home buyer's demographic group as defined by age group, income category, number of children, race, state of residence and so on, and it is very likely to expect that  $Z$  in this analysis is correlated with the error.

V. Chernozhukov and C. Hansen, [5] devise an Instrumented Quantile Regression (IQR) estimator that accounts for quantile treatment effects by solving the following problem<sup>5</sup>: find a function  $q(x, d, \tau)$  such that 0 is the solution to the quantile regression problem, in which one regresses  $Y - q(x, d, \tau)$  on some function of  $X$  and  $Z$ .

In the style of the Instrumented Quantile Regression, we estimate the log-linear model

$$Q_{\ln(Y_d)|X}(\tau) = \alpha(\tau)d + X\beta(\tau),$$

where  $d$  indicates a dichotomous “treatment” status of internet use in the home search, the outcomes  $Y_d$  is duration of search, and  $X$  is a matrix of covariates including variables such as age categories, income ranges, race indicators, and distance of the move. The coefficient  $\alpha$  has the interpretation of an elasticity of search duration with respect to internet use, and is the causal treatment effect of internet use on the duration of search. The coefficient on internet use in the standard Quantile Regressions has a

---

<sup>5</sup>See [5] p. 10

different interpretation. It estimates the statistical effect of internet use on the duration of search through conditional quantiles. Therefore, the comparison between the QR and the IQR results is analogous for example to a comparison between results from OLS and 2SLS models.

**3.2. Instrument Selection.** The dependent variable in in this study is a continuous variable representing the number of weeks a home buyer spent actively searching until finding a home which is eventually purchased by this home buyer. We regress the logarithm of this duration on a dichotomous measure of internet use while searching for a home and a number of demographic and geographic controls. These variables come from the NAR survey data used in this study. Unfortunately, the above may not be enough to correctly identify the effect of the internet use on the duration of home buying. There is a potentially serious endogeneity of internet use influencing the duration of home search. If individual home buyer heterogeneity exists, in terms, for example of how picky the home buyer is, how quality concerned or prone to lengthy search, which on one hand is correlated with internet use while buying a house, and on the other hand affects the duration of the search, the results would be biased.

In order to correct this potential endogeneity in the system, a technique of instrumental variables is warranted. The NAR Survey data itself does not contain any potential instruments for internet use. However, through the use of an auxiliary sample, in the form of the CPS Supplement on Internet Use and Computer Ownership, we can construct a simulated instrument for internet use in our main sample. From the CPS data we construct mean internet use in the CPS sample, which is representative of the US population at large by demographic categories such as age, race, household composition, residence location, and income. We then match this predicted internet use to the corresponding demographic cell in the NAR Survey main sample.

It is reasonable to expect that general internet use varies by age and income level, with younger and higher income households having a higher degree of internet use since to a great extent income proxies for educational attainment. A variable describing the level of education is not part of the NAR Survey. Both because home internet access is costly and because education and age can discern individuals who are part of the information age generation, income and age play an important role in deciding to use the internet as a tool for gathering information. Other demographic characteristics such as the number of children and the level of urbanization of the neighborhood are also important in deciding to have internet access at home. Urban and to some extent suburban areas have higher home internet access than rural areas, as internet providers offer more local dial-up services and high speed connections in

cities. At the time the data for this study is collected, certain states (because of economic or population density conditions) have a higher rate of internet availability than other states. For example, states such as California where information intensive industries are located have in place an infra-structure with more internet connections than agricultural states have. The presence of children in the household may also influence the decision to have internet access at home: new parents may find information and parenting help on-line, and parents decide to provide their school aged children with access to new technology and internet resources as part of enhancing their children's education. Thus, the average level of general internet use in one's particular demographic and geographic group is likely to be highly correlated with his or her internet use when searching for a home to purchase.

On the other hand we find it reasonable to conclude that average level of general internet use by these categories is uncorrelated with the speed with which home buyers find a suitable housing match. Home buyers' preferences for housing vary greatly within each category whether they use the internet at home as a general tool or not. Therefore we use the mean of general in-home internet use within age, income, number of children in the household, race, and state categories from the CPS as an instrument for internet use when buying a home in our NAR Survey data sample. Using combinations of these demographic and geographic characteristics we devise an instrument for internet use when searching for a home which is finely matched to a particular demographic and geographic group of home buyers.

#### 4. DATA

The National Association of Realtors conducts surveys on a regular basis of home buyers and home sellers in order to gather information about their home buying or selling experience and to assess the role of real estate professionals in these transactions. At the beginning of the year 2000, the NAR mailed questionnaires to 20,000 consumers who purchased or sold a home in 1999. The address database was ultimately derived from courthouse records of recent home buyers in the United States. This survey resulted in 1,778 usable observations. The 2000 NAR Survey is of particular interest to this study as it is the first NAR questionnaire to include detailed questions about the home buyers use of the internet from the onset of the search to the actual purchase. [10]

From this NAR Survey we use 1,746 observations which include information on home buying (as opposed to home selling). The weeks of home search variable used in this study comes from answers to the question: How long did you actively search

before you located the home you recently purchased? This response provided a number of weeks and was used as a continuous duration of search variable. While the 2000 NAR Home buyer and Seller Survey asks whether the internet was used as a source of information in the home search, we consider that answer not to be highly relevant to the degree of internet use while locating a home to be purchased. While 37% of the survey respondents indicated that they used the internet as an information source, there is no indication here about whether the internet was used specifically as a source for locating homes. We use answers to the question: What actions have you taken as a result of accessing real estate information from the internet? that include making an offer on a home found on-line, visiting a home found on-line or purchasing a home found on line as the relevant internet use in home buying durations. Using this information we created a dichotomous zero - one internet use variable that takes on a value of one when any of the above actions were taken by the home buyer in the home search. Our definition of internet use ensures that the those indicating internet use in the home search are serious about finding a new home and not simply casual lookers at houses with little intent of an actual purchase<sup>6</sup>.

The demographic characteristics of the home buyers in this survey include age, income, race, Hispanic ethnicity, number of children, household composition, number of earners, and primary language spoken. This demographic information in the data is by no means extensive; highest level of education completed would have been very useful in this study but is unfortunately unavailable. The geographic information in the data available includes state where the previous home is located, and the home search is most likely conducted from this state, and the metropolitan, suburban or rural nature of the previous home. These geographic characteristics are also available with relation to the consequently purchased home.

In order to correct for the possible endogeneity between internet use and home buyers in our main data we use age and income category means in a more general sample of US residents as an instrument for internet use. In order to construct this simulated instrument we use data from the Current Population Survey, Internet and Computer Use Supplement. The CPS is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The questionnaires are conducted either by telephone, or by an interviewer who visits the sample unit. The sample provides estimates for the nation as a whole and includes a number of different topical supplements each month. The Internet and Computer

---

<sup>6</sup>The results calculated using the less restrictive measure of internet use as answer to the question: Was the internet used as an information source in the home search? produced no significant effect on duration of search.



Use Supplement questionnaire has been conducted in December 1998 and again in August 2000.<sup>7</sup> Computer Ownership data exists for years before 1998, but collection of internet use data in the CPS starts in 1998.

We use answers to the questions: Did the respondent use the internet in the home? and: Did the respondent use the internet outside the home? as a measure of internet use. After creating mean internet usage by age and income categories in the CPS sample, we generate age, income category, and other demographics, groups in the main NAR survey sample and merge the internet use means for each age and income, etc. group to use as a simulated instrument for internet use in our NAR survey data.

Only 15% of the households in our data moved across state borders from their previous home to a new one. State areas are large enough that internet use affects the search in terms of cost and in terms of the choices available, whether a household moves within the state or to a different one. We assume that whether the move is within the state or out of state is exogenous to the model and proceed with an estimation of the effect of internet use on the duration of home search for within state movers only.

The descriptive statistics of the key variables from the NAR survey data are presented in Table 2, first for the entire sample and then for within-state movers only, which are used in our analysis. In the NAR survey data, 85% of the respondents moved to a new home in the same state.<sup>8</sup> The ages of the respondents are recorded in 5 year ranges, and the mean age of home buyers is between 35 and 39 years old in both the full sample and in the within-state movers group. Income is similarly divided in categories in our data, with the mean income falling between \$40,000 and \$50,000. Demographic characteristics available for the respondents and their families include number of children, race and Hispanic origin indicators, marital status and number of income earners in the household. The mean number of children in the home buyers' families is just under two, and again there are no significant differences in the number of children for within-state versus cross state movers. Over 85% of

---

<sup>7</sup>We also calculated our results using the December 1998 CPS sample. There were no significant differences in the results when using the December 1998 CPS data in the calculation of the instrument for internet use. While the overall amount of internet use increased in the interim, this indicates a proportional shift in internet use by demographic categories used here, and no spatial change in the type of people using the internet relevant to our study between the two dates when the CPS data was collected.

<sup>8</sup>When either the previous or the new state of residence was missing from the data, households moving within 50 miles of their previous home were assumed to have moved within state. When both the previous and new state of residence was missing the observation was not used in the analysis. We expect that those who did not report either state of residence did so at random in this sample.

the sample is White while only 6% is Black. The prevalent household type includes a married couple, at 63% of the within-state movers group, with single female households following at 19% and single male households, at 9%. The unmarried couple households account for 7% of the within-state movers sub-sample.

Of the within-state movers, 22% lived within city neighborhoods before their move, 19% of the home buyer households' previous home is in a suburb, and only 5% searched for a new home from rural areas. The CPS sample metropolitan area inhabitants are 27% of that sample, relatively similar to the percent of home buyers, searching for a home from within a metropolitan, or city area. The percent of suburban households engaged in home search is much smaller, relative to the percent of general suburb dwellers. This is to be expected as in the US there is not only a large, but a growing suburban population. The situation is similar for households searching for a home from non-urban, rural areas. Only 5% of home buyers search from a rural area, while the percent of non-metropolitan area dwellers in general, from the CPS sample, is as high as 28%. These numbers are also not surprising considering the lower mobility rates in non-urban areas.

Internet use, in terms of the actions taken as a result of using the internet in the home search, including visiting a home found on-line, making an offer on a home found on line, or purchasing a home found on-line is at 30% in the entire NAR Survey sample and 28% among within-state movers. The main reason for making the move was the the desire to own a home with 34% of the respondents pointing this as a reason for buying a home, followed by the need for more space, with 18% of the respondents giving this as a reason for the move, and 12% listing a relocation or a new job as a reason for the move.

The CPS data used in this study is summarized in the third column of Table 2. Overall internet use at home in the CPS data is at 35%, somewhat higher than internet use as part of the home search in the NAR Survey sample of home buyers. Those sampled in the CPS are slightly less affluent, which is to be expected in the general population relative to those households active in home buying. There is a slightly higher percent of Hispanic ethnicity observations in the CPS, as well as female-head households. The male-head households also account for a higher percent of the CPS sample. It is possible that there are differences in the manner in which single versus other household type is reported in the two surveys that accounts for this difference. It is also possible that the higher percent of lower income households in the CPS sample accounts for the presence of more single household heads and Hispanic respondents in the CPS relative to the households surveyed by the NAR.

The average number of weeks of search for a home is 15 for the entire NAR sample, and 16 weeks for the within state movers group. Even though all home search in this data ended with a successful location and purchase of a suitable home, the variation in the durations of home search is enormous. There is as little as less than a weeks' time of search in the data until the home, which was eventually purchased by the home buyer, was found, and up to as much as 465 weeks of search until success. In the within-state movers sub-sample there are 383 households which used the internet in the home search and 967 which did not. Among the internet users group, the most successes in finding a home occurred at 12 weeks of search, and the height of this peak in the distribution involves 51 households. The largest number of home buyers among the non-internet group were successful at only 4 weeks of search, and since this is a more numerous group in our data, this peak involves 116 home buyers.

## 5. RESULTS

Figure 4 describes the distribution of search times until success in finding the home eventually purchased. The distribution of search times for those observations where the internet was used is in part a) of the figure, and for those where internet resources were not used in the home search is in part b). Kolmogorov–Smirnov tests for the equality of the two distributions reject at the 8% level, and even though the two distributions appear somewhat similar, we are confident that there are two distinct distributions. There is a rather anomalous peak in the distribution of search durations, both in the case of internet use, and in the case of no internet exactly at 52 weeks of search, and then again at 104 weeks of search. This presents an interesting point that needs to be addressed here. It is possible that these peaks are due to misreporting in the NAR data sample. It is rather unusual to suppose that there is a valid reason such that those who have searched for almost a year should find their match in the housing market at exactly 52 weeks. It is likely that the spikes in successes of search occurring at precisely 1 year and 2 years of search are due to observations in the NAR survey where respondents erroneously remembered that it took them a year to find a house and reported search of 52 weeks, while in reality it may have taken them close to 52 weeks, but not exactly.

Fortunately, this possible misreporting does not present a problem for the quantile regression analysis performed here, since 52 weeks of search (and also 104 weeks of search) are located well in the tail of the distribution of search times. In our data, 90% of the respondents find a match after 36 weeks of search, and both peaks above are located past the 90th percentile of search durations. Any misreporting of the number of weeks as 52, or 104 is likely to reflect actual search duration close to the

reported 52 weeks or 104 weeks. Misreporting within the 90th quantile does not affect the quantile regression results concerning the rest of the distribution of search times.

The question as to whether a simple locational shift of the distribution has occurred or whether a more complicated change in the shape results from internet use in the search, can be addressed through quantile regression analysis as well. If the coefficients on internet use are the same across all quantiles of search times, then the change is a pure locational shift. If there is any difference in the effect of the internet on the duration of search, then the evidence points to a more complicated change of search times due to the use of on-line resources in the search, and justifies the choice of quantile regression analysis over a proportional hazards model. Our evidence points to the latter and rules out a simple locational shift.

**5.1. Quantile Regression Model.** While the overall effect on the distribution of search times may be a shift out, we need to use quantile regression analysis in order to find out if the internet has a different effect across quantiles. The median time to find a suitable home in our data is 8 weeks, so that 50% of those searching for a house in our data find a suitable match at  $\tau = .5$ , after 8 weeks. The first quantile,  $\tau = .1$  represents in our sample search duration of one week,  $\tau = .2$  represents search lasting three weeks,  $\tau = .3$  is at 4 weeks, and  $\tau = .4$  is at six weeks. The sixth quantile,  $\tau = .6$  represents search of 12 weeks before finding a suitable house to purchase,  $\tau = .7$  represents search of 14 weeks,  $\tau = .8$  is at 22 weeks, and after 36 weeks of search 90% of our sample have found a suitable match. There is a considerable right tail in the distribution of search times extending to over 200 weeks of search<sup>9</sup>.

The results from the standard quantile regression analysis are graphically represented in figure 5. Each panel of the figure tracks the effect of the variable on the y-axis, with the quantile index represented on the x-axis. Figure 3 a) tracks the impact of internet use when searching for a house on the search outcome in logarithmic terms. This is the direct impact of internet use without accounting for any possible endogeneity. While in the low quantiles using the internet acts to prolong the search duration, in the very last quantile, for those searching for 36 weeks or more, the use of the internet actually speeds up the time until a suitable match is found. The results for this last quantile in the right tail of the distribution include observations of search duration ranging from 36 to 456 weeks of search. As discussed in [11], there are theoretical reasons why results concerning the outliers in the right tail of the distribution

---

<sup>9</sup>There are 25 observations of search over 104 weeks of search and even one report of searching for 456 weeks before finding the house that was then purchased.

of search times may be inaccurate and spurious. We will therefore refrain from relying heavily on results about the 90th quantile of search times in the present analysis.

The results for most of the distribution of search times in the housing market are consistent with the notion that the internet changes the type of houses available to choose from for each home buyer as outlined in the search model presented above: a higher mean in the distribution of housing choices results in a longer search. When the choices one searches through are easy to examine in detail, it is feasible to visit each option and look through it in detail, making sure that more subtle details such as the direction certain rooms face, the size and relationships between the rooms, closets, staircases, and the condition of the structure match the home buyer's preferences. The internet brings each housing choice closer to every home buyer through virtual tours. One can examine the details, and choose among a distribution of houses that is overall better suited to himself or herself over the internet, independent of distance. In the absence of on-line resources (or their use), if a house is far from the home buyer, one drives by to make sure that the structure is standing, and if it simply has the right number of rooms and bathrooms it is considered among the potential matches.

The results show that the effectiveness of the internet to provide a better distribution of housing matches declines with the duration of search. Thus, at first, the internet acts to provide better housing choices to the home buyer, but as the search goes on it's role to provide better suited choices declines. It seems that the home buyer using the internet slowly learns which houses are the most highly suited to his or her preferences, and the distribution of choices available to search through does not keep improving indefinitely. This result should not come as a surprise since there is a limit to the improvement in the mean of the distribution of housing choices the internet can offer to each individual home buyer, as perhaps the ideal choice for each home buyer does not even exist in the housing market.

The demographic characteristics such as age income, and race do not vary significantly across quantiles and in each quantile in the distribution does not have a significant effect on the duration of search. Similarly, geographic characteristics such as the type of neighborhood and state do not affect the search times (differentially) across quantiles. The main significant effects on duration of search involve the internet use. These results indicate that by and large the internet acts to improve the types of choices available to each home buyer by increasing the mean, and also perhaps by increasing the spread of matches available to view, but this improvement has a limit, as the mean of the distribution of matches either reaches the perfect match or stops short for lack of a perfect match for the home owner's preferences. At longer

durations of search, the internet has a smaller impact on the distribution of choices, resulting in a smaller increase in search durations when the internet is used relative to the shorter search durations.

Inference on the quantile regression for the effects on internet use on the distribution of search durations was performed using tests developed by V. Chernozhukov in [12]. Namely, we are interested in testing for three possibilities.

- the effect of internet use is a pure location shift for most of the distribution,  $\alpha_\tau = \alpha$  for all quantiles in  $\tau \in [.1, .9]$ ,
- the effect of internet use affects the location and scale only of the outcome distribution,
- the effect of using the internet is unambiguously positive, that is testing the null hypothesis of  $\alpha_\tau \geq 0$  for all quantiles in  $\tau \in [.1, .9]$ .

The results of the tests of the three hypothesis are presented in table 1. The subsample size<sup>10</sup> for the bootstrap technique in the re-sampling technique used in the tests was 3000.

TABLE 1. Tests Results for the Internet Use in Home Search up to 36 Weeks.

Hypothesis	Null	Alternative	Smirnov Statistic	Critical Value (5%)	Decision
Pure Location					
Shift	$\alpha_\tau = \alpha$	$\alpha_\tau \neq \alpha$	1.67	1.31	Reject
Location-Scale					
Shift	$\alpha_\tau = \alpha + \gamma\alpha_\tau$	$\alpha_\tau \neq \alpha + \gamma\alpha_\tau$	0.93	1.42	Can't Reject
Stochastic					
Dominance	$\alpha_\tau \geq 0$	$e\tau : \alpha_\tau < 0$	0.48	3.14	Can't Reject

Note: b=3000, sub-sampling with replacement. Quantile index,  $\tau \in [.1, .9]$

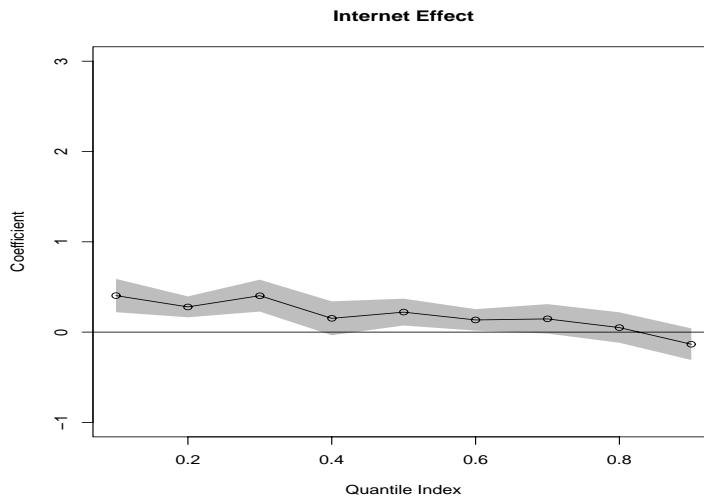
The most important hypothesis is the first one, and it is clearly rejected. The coefficient on internet use is not constant across quantiles so that the use of on-line resources has a differential effect on different parts of the distribution of search times. We cannot reject the hypothesis that only the mean and the scale of the distribution of are affected as a result of internet use. It is likely that the first part of the distributions shifts out, prolonging the duration of search in the low quantiles, and the last part of the distribution of search times shifts in, shortening the search for those that search

<sup>10</sup>Smaller size sometimes yielded singular results as some of the dichotomous covariates, such as some race indicators attained a value of 1 in very small percent of the observations.

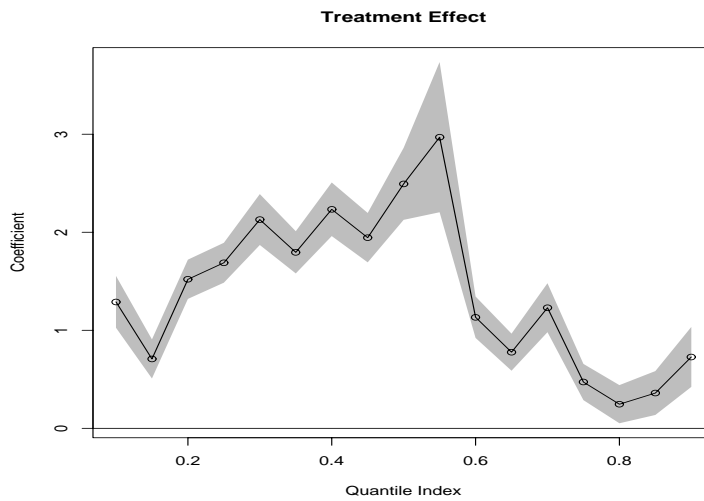
the longest, together with an increase in the mean time of search when using the internet. However, the precise form of the change in the distribution of search times when using the internet is not of particular economic interest, as long as the change is not constant across quantiles. We cannot reject the hypothesis that internet use slows down the search in all quantiles ( $\tau \in [.1, .9]$ ) quantiles. Specifically, even though the results late in the distribution suggest a possible decrease in the search duration as a result of internet use, the stochastic dominance tests suggests that it is likely there is an overall increase in search times. Therefore, we conclude that the effect of the internet is to prolong search duration relative to using conventional methods of search, especially for search duration lasting no more than 36 weeks.

**5.2. Instrumental Quantile Treatment Effects.** What about the possible endogeneity of internet use as part of the home search? The quantile regression analysis, together with an instrument for internet use constructed by age, income, number of children, race, Hispanic ethnicity, and type of urban/rural location, categories is presented in Figure 3 b) shows the results for the internet use treatment effect on the search duration outcome (again in terms of log of weeks of search). We control for similar demographics as in the proportional hazards model and in the un-instrumented version of the quantile regression analysis, such as income, age and distance of the move from the old home to the new one.

Relative to the IQR the QR estimates appear to be approximately constant. The IQR analysis shows that there is an increase in search durations in every quantile where the results are significant. Moreover, in each quantile examined the coefficient on internet use is higher after instrumenting for a possible endogeneity of internet use while searching for a house in the model than in the standard quantile regressions. At the median search duration of 8 weeks, the elasticity of search in the IQR regression is 2.5, so that at the median the increase in search duration is 25% when the internet was used in the search relative to using conventional search method only. This translates to an increase of search by about two weeks at the median search duration when the internet is used in the search. The analysis shows that internet use has a significant positive causal effect on search duration. The increase in search durations is more pronounced after controlling for any possible individual heterogeneity. Even without instrumenting, in the standard quantile regressions we see an increase in the search durations as a result of internet use. After accounting for a possible endogeneity in the model, the search times increase even further as a result of internet use in every quantile where the results are significant. The effect for different points in the distribution of search times is quite varied. The largest effect occurs near the middle of the distribution, and the impact is smallest in both tails. That is, internet use



(a) Effect of Internet Use on Search (QR)



(b) Effect of Internet Use on Search (IQR)

FIGURE 3. The Effect of Internet Use on the Log of Weeks of Search Across Different Regression Quantiles. The shaded region is the 95% confidence band. Errors estimated with kernel density methods. See appendix A for estimates' values.



strongly affects the duration of search for those searching for 3 to 14 weeks and less so for those with very long or very short search durations. Overall, the internet acts to increase the length of search for most of the distribution of search times in the housing market.

After controlling for individual heterogeneity such as prone to search, picky about the housing choice or house quality concerns among the home buyers in our data, internet use further increased the search duration. Thus, we are confident in the results (from the standard quantile regressions) pointing to an increase in search duration when using the internet. However, the endogeneity problem is largest in quantiles near the median, further justifying the use of an instrumental variable technique here.

The empirical analysis of this data suggests an overall increase in search durations for reasonable lengths of search (lasting between 3 and 14 weeks) of one to two weeks when employing the internet. This allows us to distinguish between two likely hypothesis about the role of the internet as part of search in the housing market. The above empirical evidence, together with the theoretical model developed here<sup>11</sup>, suggest that the internet plays a role in the search that goes beyond a change in the arrival rate of offers. Given the results above, we can conclude that the internet carries additional information about the potential housing choices available, and not just adds to the volume of choices the home buyer can have access to in a given period of time.

## 6. CONCLUSION

The influence of the internet and on-line resources on many aspects of life today is currently of interest to economists and other social scientists alike. This comes without a surprise, as the internet has changed the way we do business, search for information and purchase goods and services. As users of the internet, we experience its power to deliver information and services quickly. Our “fingers do the walking” and get to their objective in virtual space much more easily than ever before. For consumers, the information and services available through the internet are available conveniently and with fast, uninterrupted access twenty-four hours a day, which creates the feeling that the internet speeds up the execution of tasks which used to take longer before the wide-spread use of the internet.

---

<sup>11</sup>Further work by the author suggests that the listing a home on the internet increases the duration until purchase by about two weeks as well, justifying the use of a housing market search model where all buyers are sellers [13].

This notion that the internet speeds up certain tasks is not always correct. In the case of sequential search in the housing market the theoretical prediction of the effect of using internet resources as part of a home search is, first of all, ambiguous. As discussed above, since the use of the internet increases the search efficiency, a home buyer who uses the internet has an increased number of choices and can find a suitable home faster, but at the same time there is an adjustment of the reservation level so that rather than experiencing an effect that speeds up the search time, the home buyer benefits from using the internet in the search beyond simply being able to look at more choices. The internet is able to deliver specific information about the features of the particular house, and allows the home buyer to browse through choices that are better matched to his or her household than conventional methods of search can provide. During the on-line search the home buyer might also be able to look at a larger variety of types of choices, some that would not be available through conventional methods, because perhaps of geographic location, or because of a more narrow choice of offers available in a newspaper advertisement, or a real estate agency. These extra choices may be very well or very badly suited to the home buyer relative to choices available through conventional methods. If in fact a Realtor presents a specific type of houses to a home buyer because of commission ranges or other geographical concerns, using the internet in the search would increase the variance of the distribution of choices. We find evidence in this study that there is a change in the distribution of choices either through increasing the mean or through increasing the mean and the variance of the underlying distribution.

Here, Using data from the National Association of Realtors 2000 Home Buyer and Seller Survey, and an auxiliary sample from the Current Population Survey, August 2000 Supplement on Computer Ownership and Internet Use, we find that employing internet resources as part of the home search in the US housing market tends to increase search durations. We conclude that since the internet increases search durations, the more important aspect of internet use in the search is not the ability to look at choices faster, but the ability to explore choices that are better tailored to each home buyer by increasing the variety of choices available to consider. This is an important finding as it relates to search durations in the housing market and the use of the internet for this purpose, but it also has broad implications for the relation between the role of the internet in markets and sequential search in general.

This study presents important and interesting findings and sheds light on the workings of sequential search in the housing market. However, it poses a number of interesting questions suitable for further investigations. The housing market in the US is one of the larger markets in the US where buyers perform sequential search, but there

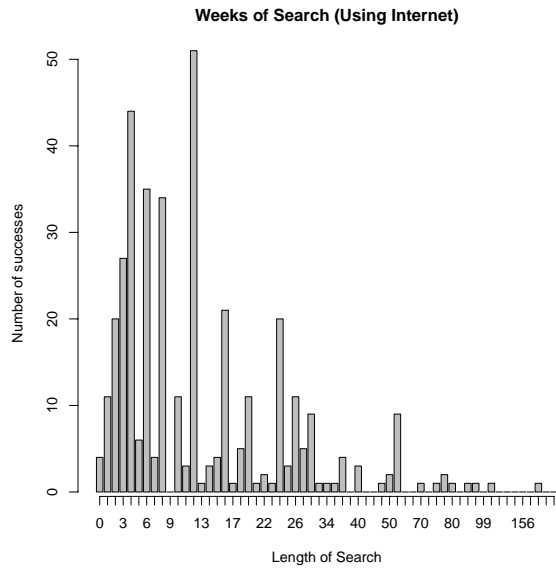
are other important and extensive markets, such as the job market or dating/marriage searches. It would be of great interest to find out whether the implications of increased search efficiency and improvement to the distribution of choices for the duration of search carry over to other markets. Perhaps the opposite effect dominates in different situations. Clearly, the sequential manner of the search is an important feature of this market which affects the theory and the empirical results. The effect of using the internet on the amount of search effort exerted is another topic of further research. We hope to address these and other related issues in further studies.

TABLE 2. Descriptive Statistics of Key Variables

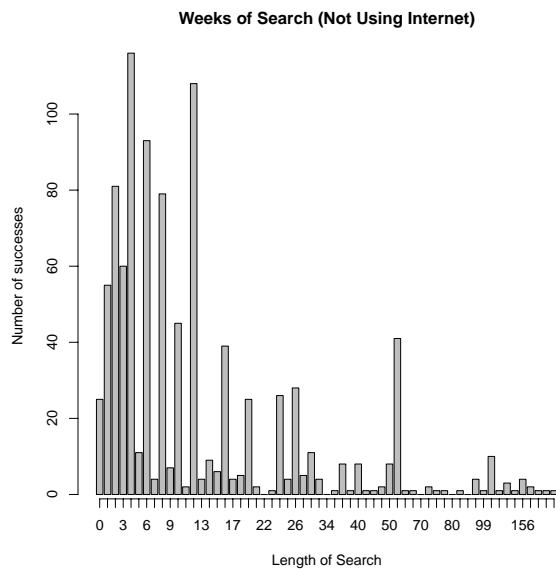
	NAR Full Sample (1)	NAR Within-state Movers Only (2)	CPS Sample (3)
Within State Movers	.850 (.358) [1630]	1 (0) [1385]	-
Weeks of Search	14.98 (25.94) [1746]	15.97 (26.99) [1350]	-
Internet Used in Search	.298 (.457) [1787]	.279 (.449) [1385]	-
Internet Use at Home	-	-	.345 (.475) [121745]
Age Group	4.641 (2.248) [1751]	4.458 (2.226) [1361]	4.173 (2.999) [121745]
Household Income Category	5.966 (2.174) [1690]	5.856 (2.190) [1319]	4.462 (2.753) [103750]
Number of Children	1.738 (1.026) [1771]	1.743 (1.039) [1376]	1.428 (.895) [121745]
White	.880 (.325) [1705]	.866 (.340) [1321]	.838 (.368) [121745]
Hispanic	.060 (.238) [1643]	.070 (.255) [1279]	.106 (.308) [121745]
Married Couple	.656 (.475) [1770]	.627 (.484) [1374]	.664 (.472) [121745]
Single Female Head of Household	.178 (.383) [1770]	.186 (.389) [1374]	.219 (.418) [121745]
Single Male Head of Household	.091 (.287) [1770]	.103 (.305) [1374]	.116 (.321) [121745]
Unmarried Couple	.656 (.475) [1770]	.627 (.484) [1374]	-
Number of Earners in Household	1.559 (.528) [1693]	1.587 (.525) [1323]	-
(Previous) Home Location: Metropolitan Area	.235 (.424) [1787]	.224 (.417) [1385]	.269 (.455) [103273]
(Previous) Home Location: Suburb	.223	.194	.455

	(.415)	(.396)	(.498)
(Previous) Home Location:	[1787]	[1385]	[103273]
	.063	.0533	.276
Non-metropolitan / Rural	(.243)	(.225)	(.447)
	[1787]	[1385]	[103273]

*Note:* Data in columns (1) and (2) from the National Association of Realtors 2000 Home Buyer and Seller Survey. Data in column (3) from August 2000 Current Population Survey Supplement on Computer Ownership and Internet Use. For each variable the mean value, the standard error (in parenthesis), and the number of observations [in brackets] are presented. Age group definitions: (1) less than 25 years old, (2) 25-29 years old, (3) 30-34 years old, (4) 35-39 years old, (5) 40-44 years old, (6) 45-49 years, (7) 50-54 years old, (8) 55-64 years, and (9) 65 years or older. Income category definition: (1) under \$25,000, (2) \$25,000 - \$29,999, (3) \$30,000 - \$34,999, (4) \$35,000 - \$39,999, (5) \$40,000 - \$49,999, (6) \$50,000 - \$59,999, (7) \$60,000 - \$69,999, and (8) \$70,000 or more.



(a) Internet Used



(b) No Internet Used

FIGURE 4. Distributions of Home Search Duration

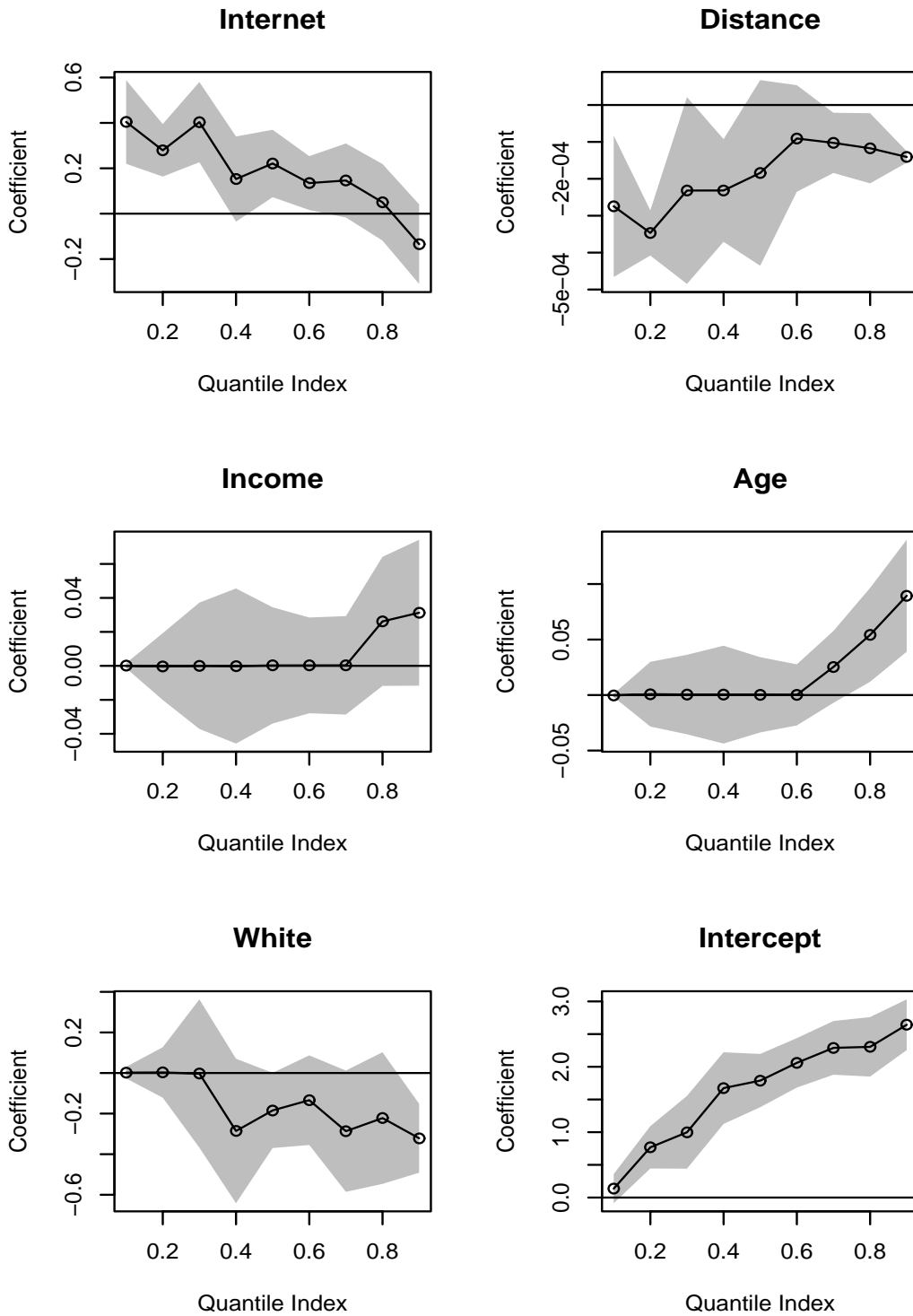


FIGURE 5. Standard Quantile Regression Results: Effects of Covariates on the Log of Weeks of Search

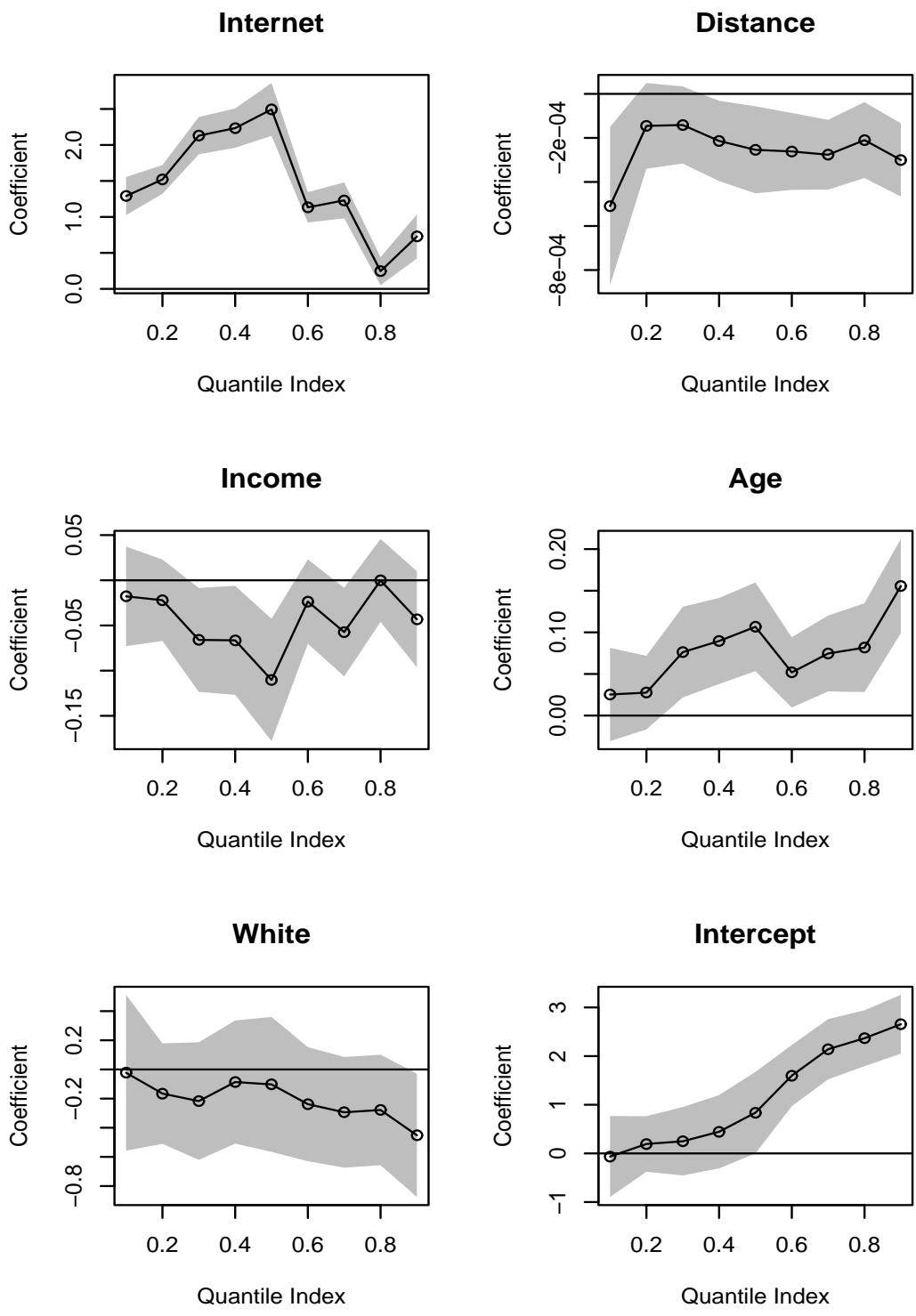


FIGURE 6. Instrumented Quantile Regression Results: Effects of Co-variates on the Log of Weeks of Search



## APPENDIX A

### Standard and Instrumented Quantile Regressions Comparison

Table 3 presents the coefficients for the quantile regression models before and after instrumenting for a possible individual heterogeneity in the model.

TABLE 3. Effect of Internet Use: Comparison of Quantile Regression Results

Search Duration Lasting	Quantile Index (1)	QR Internet Use Coefficient (2)	IQR Internet Use Coefficient (3)
One week	$\tau = .1$	0.404 (0.094)**	1.290 (0.135)**
3 weeks	$\tau = .2$	0.279 (0.059)**	1.520 (0.102)**
4 weeks	$\tau = .3$	0.403 (0.090)**	2.129 (0.132)**
6 weeks	$\tau = .4$	0.153 (0.095)	2.234 (0.139)**
8 weeks	$\tau = .5$	0.221 (0.076)*	2.494 (0.187)**
12 weeks	$\tau = .6$	0.135 (0.060)	1.133 (0.107)**
14 weeks	$\tau = .7$	0.146 (0.083)	1.230 (0.127)**
22 weeks	$\tau = .8$	0.050 (0.089)	0.246 (0.099)*
36 weeks	$\tau = .9$	-0.134 (0.089)	0.728 (0.156)**

*Note:* N=1171, standard errors in (parenthesis), (.)\* indicates significance at the 5% level, (.)\*\* indicates significance at the 1% level.

Each coefficient reflects the effect of internet use in the search in a separate regression at the given quantile index and also either under the standard QR or the IQR model. If those who are prone to search more through unobservable characteristics such as pickiness, or in extreme terms perfectionism or obsessiveness in their personalities are also likely to use the internet in their search for a house the resulting endogeneity in the model will bias the results. In order to control for this possibility, we use a simulated instrument for overall likelihood to use the internet in other activities. The instrument is thus the mean internet use of each demographic group in the general population divided by age, income, number of children, race, Hispanic origin, and geographic characteristics such as type of urban/rural location and state of residence from which the search was conducted. Covariates include age, income

categories, the distance of the move, race, and an indicator for the primary reason for the move: corporate relocation/new job versus a less need to move.

## REFERENCES

- [1] James McQuivey, Kate Delhagen, Kip Levin, and Maria LaTour Kadison. Retail's growth spiral. *Forrester Report*, 1(8), 1998.
- [2] Jeffrey R. Brown and Austan Goolsbee. Does the internet make markets more competitive? evidence from the life insurance market. National Bureau of Economic Research, Working Paper 7996, November 2000.
- [3] William C. Wheaton. Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, 98(6):1270–1292, December 1986.
- [4] Christopher A. Pissarides. *Equilibrium Unemployment Theory*. Cambridge, MA: The MIT Press, 2000.
- [5] Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. MIT Department of Economics Working Paper, December 2001.
- [6] Thomas W. Malone, Joanne Yates, and Robert I. Benjamin. Electronic markets and electronic hierarchies. *The Communications of the ACM*, 30(6), 1987.
- [7] Machado Jose A. F and Pedro Portugal. Quantile regression analysis of transition data. Preliminary version, Faculdade de Economia, Universidade Nova de Lisbon, June 2001.
- [8] D. R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society B*, 34, 1972.
- [9] Roger Koenker and Kevin F. Hallock. Quantile regression an introduction. University of Illinois at Urbana-Champaign, Prepared for the Journal of Economic Perspectives “Symposium on Econometric Tools”, December 2000.
- [10] Kevin A. Roth et. al. The 2000 national association of realtors profile of home buyers and sellers. National Association of Realtors, Economic Research Group, 700 11th Street, NW, Washington, DC, 20001.
- [11] Victor Chernozhukov. Conditional extremes and near extremes: Concepts, asymptotic theory and economic applications. Ph.D. Thesis, Stanford University, August 2000.
- [12] Victor Chernozhukov. Inference on quantile regression process, an alternative. MIT Department of Economics Working Paper, January 2002.
- [13] Victoria T. D'Urso. Internet listing and time on the market for residential housing: A simultaneous equations approach. Massachusetts Institute of Technology, PhD Thesis, Chapter 2, 2002.