Internet Use and the Duration of Buying and Selling in the Residential Housing Market, Economic Incentives and Voting

by

Victoria Tanusheva D'Urso

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2002

Author . . . .
Department of Economics
June 25, 2002

Certified by . . . .
William Wheaton
Professor, Urban Economics
Thesis Supervisor

Certified by . . . .
James Snyder
Professor, Political Economy, American Politics, Formal Political Theory
Thesis Supervisor

Accepted by . . . .
Peter Temin
Elisha Gray II Professor, Economic History and Applied Economics
Chairman, Department Committee on Graduate Studies

Internet Use and the Duration of Buying and Selling in the Residential Housing Market, Economic Incentives and Voting

by

Victoria Tanusheva D'Urso

Submitted to the Department of Economics
on June 25, 2002, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In this study we examine the impact of internet use on the duration of non-sequential 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 the search. 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 we find a tendency for internet use to increase the duration of home search relative to employing more conventional search methods.

We use a simultaneous equations approach for the analysis of the impact of internet listing on the duration until sale in the residential housing market. In this model, the time on the market and the selling price are jointly determined, once asking price and the method used for the listing of the property is chosen by the home seller or agent. We use data from the 2000 Home Buyer and Seller Survey collected by the National Association of Realtors. We find that using the internet to list a house increases its time on the market. The results presented here are consistent our with previous findings pertaining to the use of the internet and the duration of search until a buyer locates a home to purchase. These results, together with the findings of the present study show evidence for a model of the housing market where all buyers are sellers.

We investigate the differential propensity of voters in the US to participate in national only versus national and local elections. We use data from the 1987 US General Social Survey to asses the importance of demographic and local community attachment characteristics of voters for this differential voting decision. We find that local community attachment and civic duty play an important role for this voting decision while personal monetary gains and redistributions do not appear to factor into the decision. In particular, education, age of respondent and length lived in community act to lower the costs of voting locally, and influence the voters' decision to participate in local elections as well as in national ones. However, economic incentives such as real estate capital values, local taxes and Social Security allocations do not appear to drive the differential voting decision for participating in local and national elections versus participating in national level elections only.

Thesis Supervisor: William Wheaton
Title: Professor, Urban Economics

Thesis Supervisor: James Snyder
Title: Professor, Political Economy, American Politics, Formal Political Theory

To Brian,

without whose encouragement and love

none of this work would have been possible.

Acknowledgments

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 their computer programs used for the inverse quantile regression 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 Home Buyer and Seller Survey data. Thanks to James Snyder for helpful discussions and for providing data from the American National Election Studies.

Contents

List of Figures

List of Tables

Chapter 1

Home Buyer Search Duration and the Internet

1.1 Introduction

The internet, what we today accept as the initial implementation of a shared body of information, available at the fingertips of those equipped with a computer and an internet connection, has been around 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 important 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 [9]. This far reaching, and unprecedented impact is quickly, and justifiably so, becoming the focus of an ever increasing body of 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 lead to work such as Brown and Goolsbee [1]. In their study, the authors investigate the presence of internet markets as they relate to 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 non-sequential fashion.

In this study we examine the impact of on-line resources for the non-sequential 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 non-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, which gives rise to the non-sequential nature of the search in the housing market.

Brown and Goolsbee [1] concentrate their work on the impact of the internet on insurance prices. In our study we do not examine any impact of the on-line search on housing prices as there is not a meaningful way of controlling for the match quality in our data, as prices in this market vary greatly with the existing house amenities and exact location. Instead, we focus on the impact of on-line search on home buying duration. 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.

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. As a first step in this analysis, we employ a hazard model regression to distinguish the effect of the internet on the duration of home search. In addition, we use an instrumental variables technique to control for a particular endogeneity which may exist in this data. Because of the non-linear nature of the estimation methods employed in combination with the instrumental variables technique in our setup, there could be bias in the results which is impossible to eliminate. The proportional hazards model we employ provides a framework in which the instrumental variables technique required here can be easily implemented. It is therefore a useful first estimation approach in this study. However, the (Cox) proportional hazards model places unnecessary restrictions on the changes the distribution of search times undergoes as a result of home buyers using the internet. Therefore, we further estimate the effects of on-line resources using a different, quantile regression approach. Of course, the use of an instrumental variables technique is still warranted in the quantile regression setting. In a recent work, V. Chernozhukov and C. Hansen [4] define a technique for the use of an instrumental variable in quantile regressions, which we employ in this study.

Since the survey data we use in this study only includes information about people 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 particular we are concerned that it is possible that only home searchers with a large amount of time on their hands go on-line as part of their home search, or perhaps alternatively, that only those in a hurry to purchase a home and move quickly employ internet resources in their search. In order to control for this endogeneity 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. Other similar mean levels of internet use such as those constructed by age, income and the metropolitan status of the previous home location were also considered.

Search models dictate that the ability to use on-line resource in addition to traditional methods of search reduces the cost of search as it takes less time and money to learn about the choices offered, their location and features. Building on the standard models of non-sequential search [11, 10] found in the literature, we develop a simple model of non-sequential search in the housing market. In this 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 as it brings a wider selection of houses 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. After instrumenting for internet use as mentioned above, the data from this NAR survey used here suggests that search durations are likely to be longer when employing the internet as part of the home search.

1.2 Theoretical Discussion

Theoretical developments in the literature pertinent to this study include Wheaton's [11] model of general equilibrium in the housing market and Pissarides's [10] unemployment equilibrium model adapted to the housing market situation. Wheaton's model is the more simple and straight forward model of the workings of the the housing market, as it treats the turnover rate, the rate at which households become dissatisfied with their current housing choice and consequently search for a new home, as exogenously determined [1]. Pissarides, on the other hand, incorporates the turnover rate into his model and treats it as endogenous.

1.2.1 Simple Model of Non-Sequential Search

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

[1] The model presented here is a revision of Wheaton's model of matching in the housing market as outlined in [11].

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 β 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 α. 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 μ and variance σ, then $F(X)$ is the cumulative distribution function of the above normal. Households have a reservation level R, below which the household would not accept a given offer. The magnitude of the transition shock α moves the household from utility level $U_M = \mathcal{U}(R)$ to $U_S = \mathcal{U}(R - \alpha)$ where $\mathcal{U}(.)$ 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 λ, the search efficiency and H_S the number of searching households. Another equivalent interpretation of λ 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 [2], $\lambda = li$.

[2]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 [3] flow equations:

$$rV_M = U_M - \beta(V_M - V_S),$$

$$rV_S = U_S + q(V_D - V_S - P),$$

$$rV_D = U_M + z(V_M - V_D + P)$$

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,

$$P = \frac{(U_M - U_S)(2\beta + r + z)}{r(2\beta + 2r + q)}, \qquad V_M = \frac{(2r + q)U_M + 2\beta U_S}{r(2\beta + 2r + q)}$$

$$V_D = \frac{(2\beta + 2r + q + z)U_M - zU_S}{r(2\beta + 2r + q)}, \text{and} \qquad V_S = \frac{qU_M + 2(\beta + r)U_S}{r(2\beta + 2r + q)}.$$

[3] While a richer model of search in the housing market (see [11]) 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.

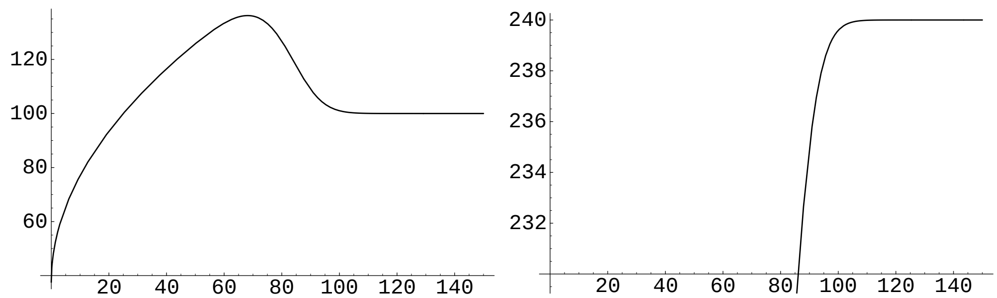
Each household chooses R to maximize the value of being mismatched. After recalling the definition of q as a function of R and imposing a functional form for the utility of a matched state as a function of the reservation R as well, together with values for the parameters of the model, we can numerically solve for the maximum value of being in a mismatched state. This maximum occurs at $R^\star$, the value of R corresponding to the peek 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 β of 10%, search efficiency of 50%, and $U_S = 5$, we obtain $R^\star = 68.1624$ as shown in Figure 1-1 a). For the above parameter values the value of being mismatched achieves a well defined, unique maximum at $R^\star$. 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 $R^\star$. 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 $R^\star$ to exceed $\frac{U_S}{r}$. This condition reduces to $\mathcal{U}(R^\star) > U_S$.

1.2.2 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

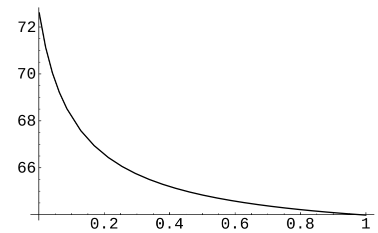


(a) Well defined $R^{\star}$, with large drop in utility when mismatched.

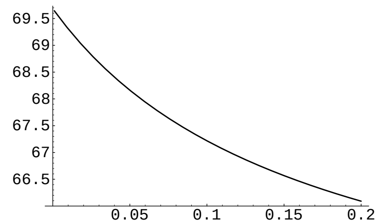
(b) V_S levels out at $\frac{U_S}{r}$ for a small drop in utility when mismatched before reaching a maximum.

Figure 1-1: Plot of V_S (y-axis) vs. R (x-axis)

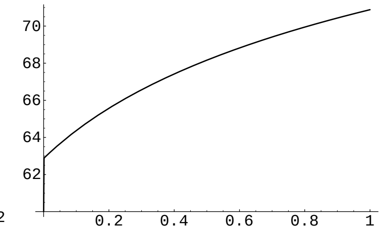
any given time period. Speeding up the arrival of offers, internet use enters into our model through the parameter λ. However, looking at potential housing choices on-line carries more information than simply delivering these choices faster. If using the internet in the search delivers a larger set of 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.



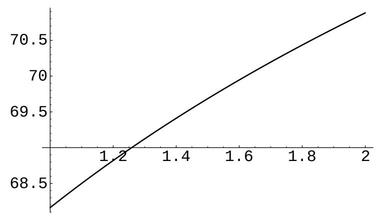
(a) $R^{\star}$ vs. β



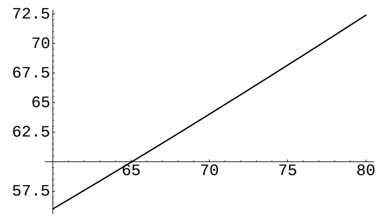
(b) $R^{\star}$ vs. r



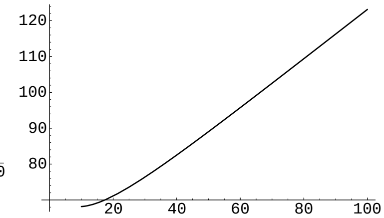
(c) $R^{\star}$ vs. l



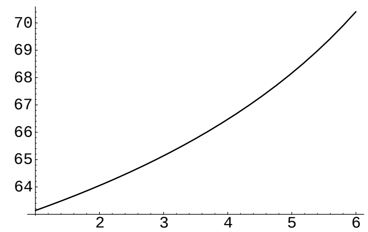
(d) $R^{\star}$ vs. i



(e) $R^{\star}$ vs. μ



(f) $R^{\star}$ vs. σ



(g) $R^{\star}$ vs. U_S

Figure 1-2: Sensitivity of $R^{\star}$ to model parameters

Consider the effect of a change in the parameters of the model on the change in $R^{\star}$. The optimum reservation can be written as an implicit function of the parameters, β, r, l, i, μ, σ 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 a 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 λ, and more specifically an increase in the internet portion, i, leads to an increase in $R^\star$. Similarly, the optimum reservation level increases with an increase in μ, σ 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 $R^\star$ as shown in Figure 1-2.

How does an increase in internet use during the search affect the duration of search? First, let's examine the search efficiency effect. Since $\frac{\partial i}{\partial R^\star} > 0$, when i increases $R^\star$ 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 $R^\star$ 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, $R^\star$ increases from 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 [8], 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 at. 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 λ 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". [4] 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, μ, 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, σ 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 μ or σ increase, $\mathrm{F}(R^{\star})$ 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.

It 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

[4] See [8], p. 488.

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.

1.3 Estimation Strategy

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, 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 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 may play an important role in deciding to use the internet as a tool for gathering information. We expect that internet use varies with income and age. On the other hand we find it reasonable to conclude that income – age categories are uncorrelated with the speed with which households find a suitable housing match. Therefore we use the means by income–age categories of general internet use as an instrument for internet use when buying a home in our NAR Survey data sample. Other demographic characteristics such as the number of children and the level of urbanization of the previous neighborhood are also reasonably uncorrelated with the home search duration, yet influence the level of internet use as part of the search. 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. 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. Using combinations of these demographic and geographic characteristics we devise a instrument for internet use which is more finely matched to a particular demographic and geographic group of home buyers.

The geographical location of the home to be purchased as well as the location of the previous address, from which the search would be conducted is available at the state level. Thus there are two relevant state variables for each observation: the previous state and the new one. In our analysis, we need to control for the previous state in order to remove any possible differences of internet availability that might exist across states. It is possible that highly urbanized states such as California, where information technology businesses and internet savvy individuals are clustered, might have higher internet use rates than rural states, such as some of the states located in the Midwest, where the internet connectivity, which is somewhat less important for agriculture, lags behind. We likewise need to control for the particular new state to which a household id moving to in order to account for different economic conditions across states. This requires a large number of state indicator variables to be included in out regressions. This large number of controls to be included as right hand side variables makes the analysis impractical. We restrict our investigation to those households in the NAR Survey sample which moved within state. 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 reduces the cost of search 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. This restriction of the data used allows us to reduce the number of state indicators to be included in our regressions, as now there is simply a state that needs to be controlled for.

A simple estimation strategy that could be employed, would consist of a probit model, regressing internet use when buying a home on the predicted internet use derived from the CPS and a matrix of demographic characteristics as a first stage and then using the predicted values from this first stage in the second stage Cox model regression. A bootstrapping [6] technique is then needed to obtain unbiased standard errors in the second stage regression. There are, however, two major problems associated with this technique. The first is a possible source of bias in the results due to the non-linear nature of the probit model used as a first stage. Because of the non-linear nature of the regression in the second stage Cox model, in a combination with the probit first stage as an instrumental variable technique, the bias in the results would be impossible to eliminate in this estimation framework. A linear probability model rather than the probit model first stage regression would eliminate the problem, however, since the split of internet users versus non-internet users in our sample is 30% to 70%, the linear probability model produces results that are quite different than the probit estimates and is therefore not quite appropriate here.

A different shortcoming of the proportional hazards estimation comes in through the restriction placed on the underlying distribution of search durations in this model. The Cox model assumes the internet effect 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 more likely that using the internet in the home search results in 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.

1.3.1 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

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

as an estimate of the conditional sample mean, $E(Y|x)$. The quantile 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

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

where ρ_τ 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 [7]. The more general ρ_τ allows estimation of conditional quantiles other than the median and generalizes the median regression to a quantile regression for any quantile index τ. 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 [7]. 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 instrument 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 quantile regression results and then employ an instrumented quantile regression technique as outlined in a recent work by V. Chernozhukov and C. Hansen, [4].

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 nd 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 the same home buyers. This is less restrictive that 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 absolutely 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, [4] devise an Inverse Quantile Regression (IQR) estimator that accounts for quantile treatment effects by solving the following problem[5]: 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, Z).

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

$$Q_{\ln(Y_d)|X}(\tau) = d\alpha_\tau + X'\beta_\tau,$$

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 α 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 (QR) has a 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.

1.4 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 a questionnaire to 20,000

[5]See [4] p. 10

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 used 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. [5]

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[6].

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

[6]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.

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 Use Supplement questionnaire has been conducted in December 1998 and again in August 2000. [7] 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.

The descriptive statistics of the key variables from the NAR survey data are presented in Table 1.2, first for the entire sample and then for within-state movers

[7] 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.

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.[8] The ages of the respondents are recorded in 5 years 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 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

[8] 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 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 1.2. 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 my 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 peek 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 peek involves 116 home buyers.

1.5 Results

Figure 1-4 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 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.

1.5.1 Quantile Regression Model

A proportional hazards model does not allow for a detailed look at the changes in the distribution of search times as a result of internet use. While the overall effect on this 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[9].

The results from the standard quantile regression analysis are graphically represented in figure 1-5. 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 1-3 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

[9] 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.

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 [2], there are theoretical reasons why results concerning the outliers in the right tail of the distribution 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 [3]. 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 1.1. The subsample size[10] for the bootstrap technique in the resampling technique used in the

[10] Smaller size sometimes yielded singular results as some of the dichotomous covariates, such as

tests was 3000. 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 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 internet, 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 this result is likely to be spurious. 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.

Table 1.1: Tests Results for the Internet Use in Home Search up to 36 Weeks.

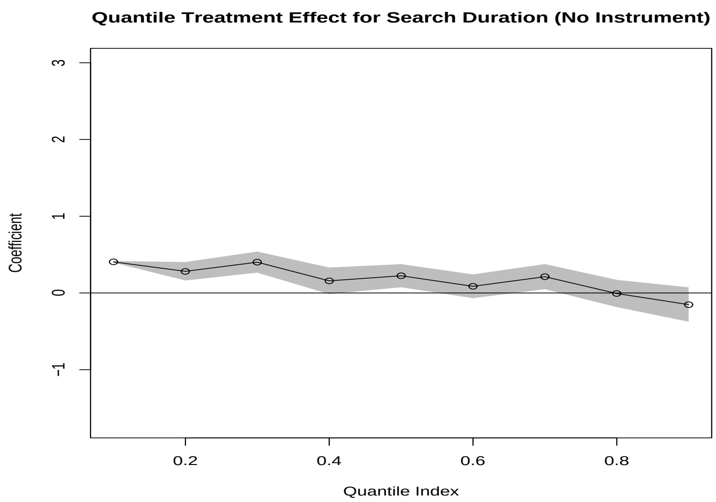
Hypothesis	Null	Alternative	Smirnov Statistic	Critical Value (5%)	Decision
Pure Location Shift	$\alpha_\tau = \alpha$	$\alpha_\tau \neq \alpha$	1.67	1.31	Reject
Location-Scale Shift	$\alpha_\tau = \alpha + \gamma\alpha_\tau$	$\alpha_\tau \neq \alpha + \gamma\alpha_\tau$	0.93	1.42	Can't Reject
Stochastic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

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

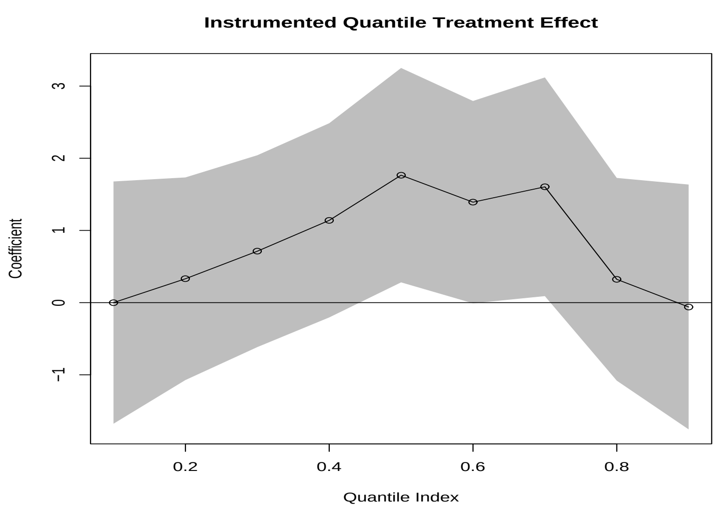
1.5.2 Instrumental Quantile Treatment Effects

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

some race indicators, attained a value of 1 in very small percent of the observations.



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



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

Figure 1-3: The Effect of Internet Use on the Log of Weeks of Search Across Different Regression Quantiles by age, income, number of children, race, Hispanic ethnicity, and type of urban/rural location, categories is presented in Figure 1-6. The corresponding objective to be minimized is plotted against α for some of the regression quantiles in figure 1-7. Figure 1-3 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.

The IQR analysis shows that there is an increase in search durations in every quantile where the results are significant. Moreover, in each quantile where the results are significant 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 1.8, which translates to an increase of search by about two weeks when the internet is used in the search. This analysis shows that the internet does have a significant positive effect for durations of search around the median of the distribution. So, as the overall effect of the internet is to increase the search times, 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. Thus, we are confident that the internet acts to slow 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 slowed the search even more. Thus, we are confident in the results (from the standard quantile regressions) pointing to an increase in search duration when using the internet. Furthermore, the causal effect of internet use in the housing market search is that of an increase in search duration around median of the distribution of search times. The quantiles in both tails of the distribution, indicating very short and very long search durations, do not exhibit a causal increase of search duration due to internet use. In these quantiles there is no significant bias in the results due to individual heterogeneity among the home buyers on our sample. However, around the median of the distribution, the above results show that if a home buyer were forced to use the internet in the home search the result would be a longer search. The difference between the QR and IQR results around median lengths of search shows that picky or quality concerned home buyers tend to avoid the use of the internet in their search more than other home buyers do.

The empirical analysis of this data suggests a likely overall increase in search durations for reasonable lengths of search (lasting between 4 and 14 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, 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.

1.6 Conclusion

The influence of the internet and on-line resources on many aspects of life today is currently of interest to economists and 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.

This notion that the internet speeds up certain tasks is not always correct. In the case of non-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 match qualities either through increasing the mean or through increasing the mean and the variance of the distribution of choices.

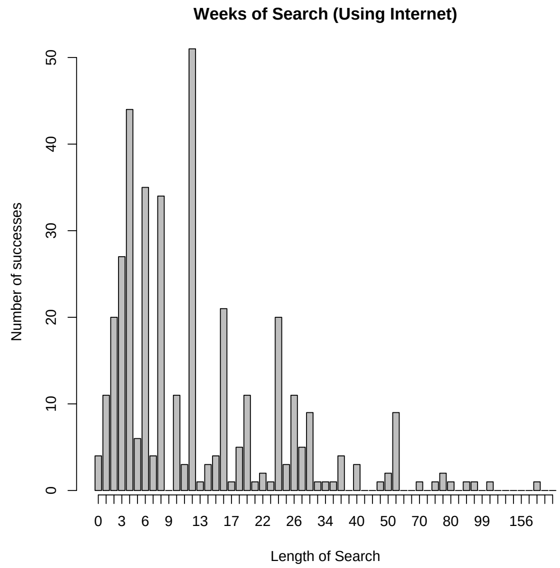
Here, Using data from the National Association of Realtors 2000 Home Buyer and Seller Survey, and an auxiliary sample from the Current Population Survey, August 2000 Supplement on Computer Ownership and Internet Use, we find that employing internet resources as part of the home search in the US housing 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 non-sequential search in general.

This study presents important and interesting findings and sheds light on the workings of non-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 were buyers perform non-sequential search, but there are other important and extensive markets, such as the job market, where job seekers search for suitable employment. It would be of great interest to find out whether the implications of search cost reductions for the duration of search carry over to other markets. Clearly, the non-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.

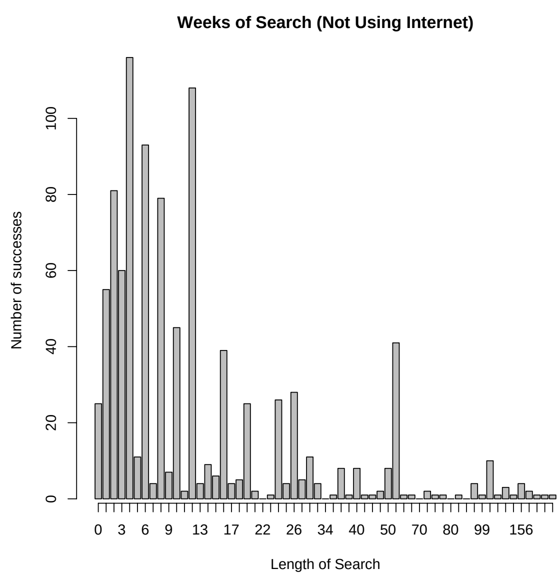
Table 1.2: Descriptive Statistics of Key Variables from the NAR and CPS Samples

	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 (.415) [1787]	.194 (.396) [1385]	.455 (.498) [103273]
(Previous) Home Location: Non-metropolitan / Rural	.063 (.243) [1787]	.0533 (.225) [1385]	.276 (.447) [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 1-4: Distributions of Home Search Duration

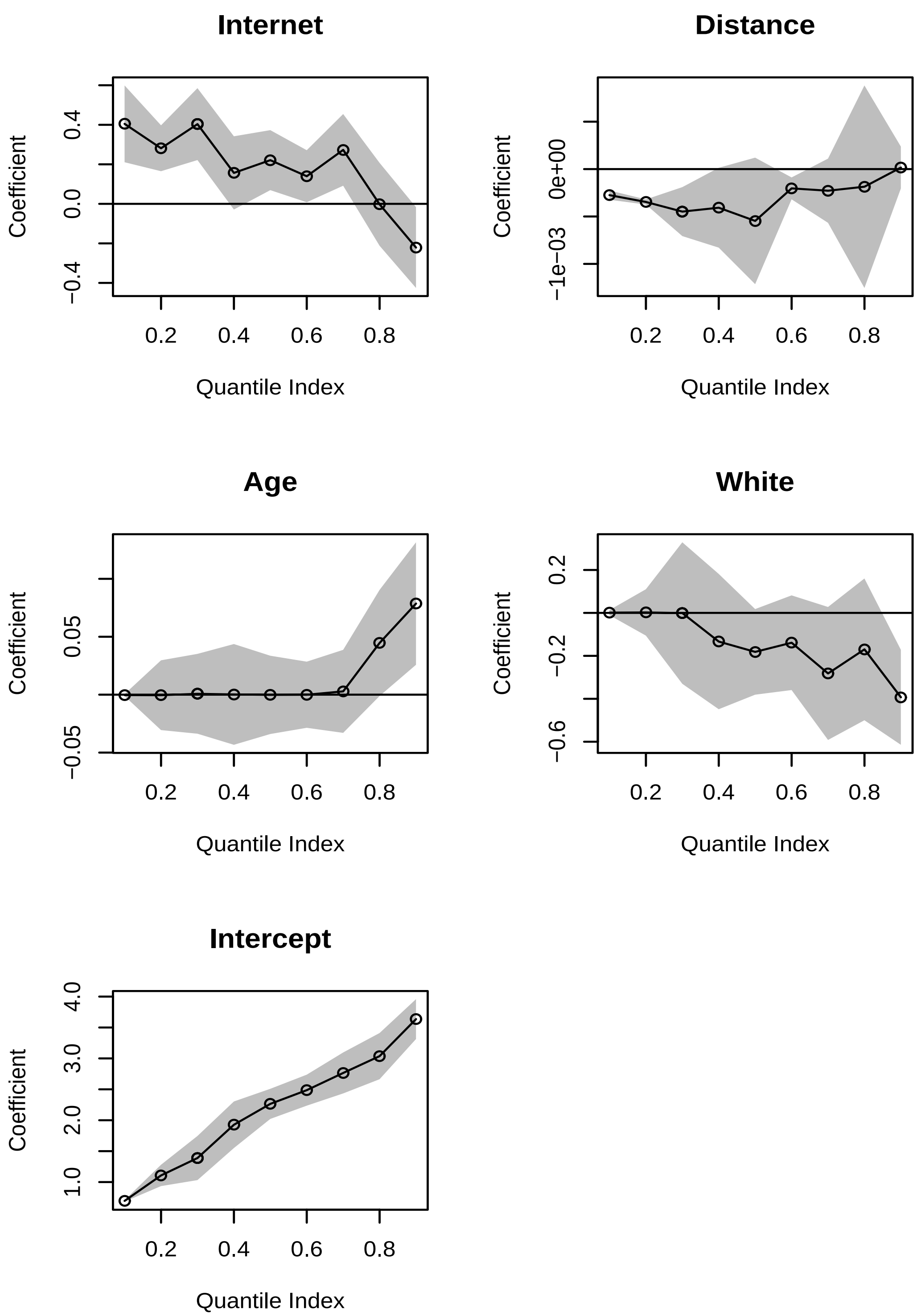


Figure 1-5: Standard Quantile Regression Results: Effects of Covariates on the Log of Weeks of Search

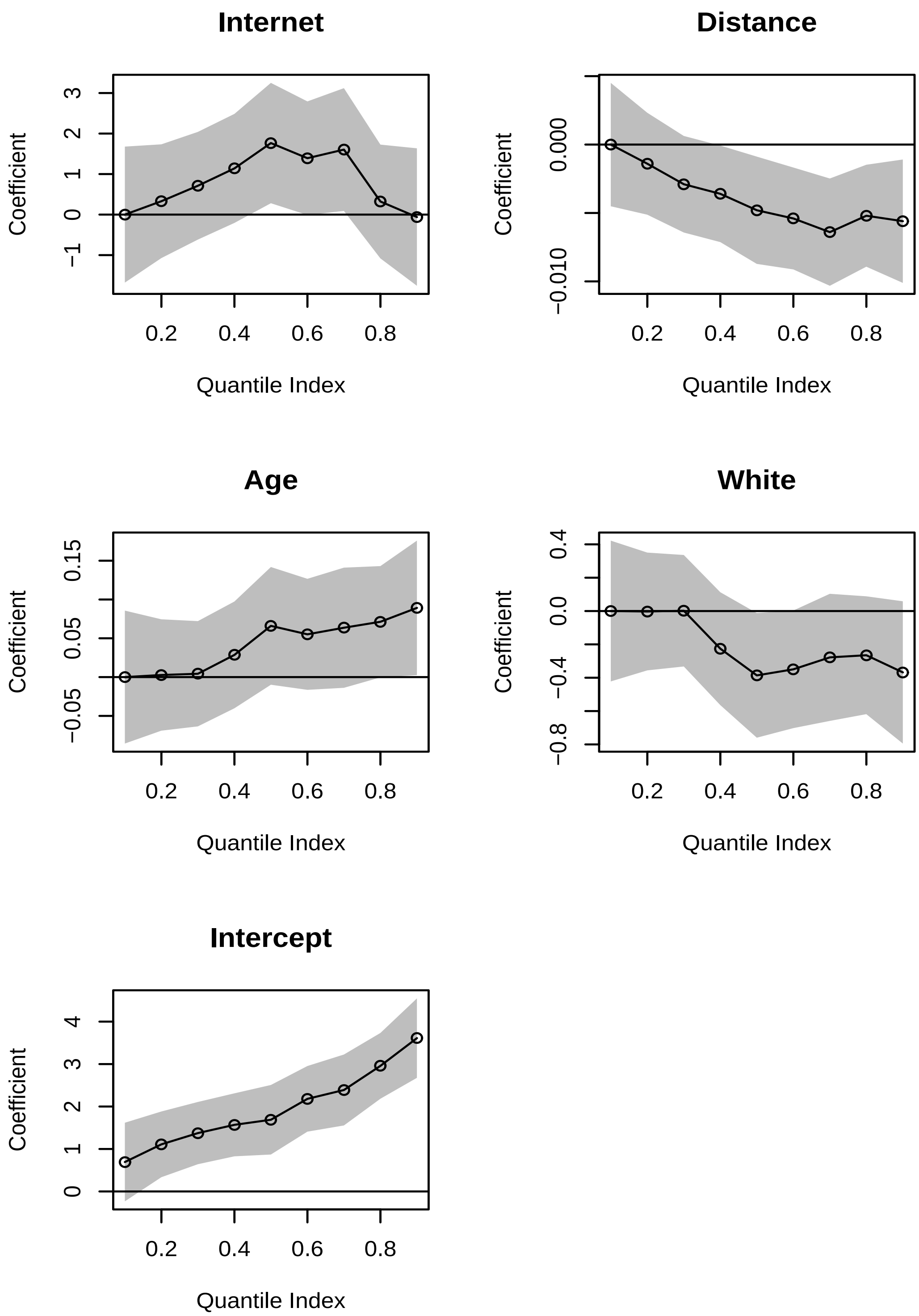
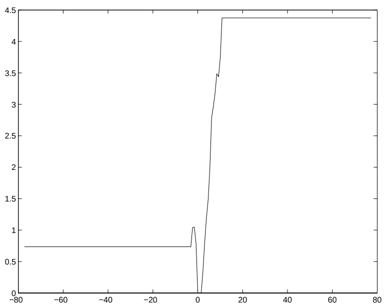
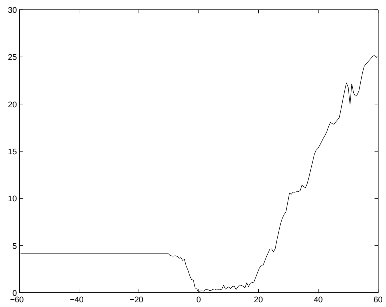


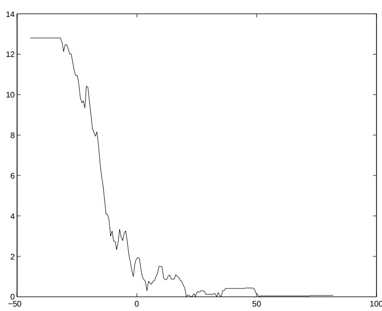
Figure 1-6: Instrumented Quantile Regression Results: Effects of Covariates on the Log of Weeks of Search



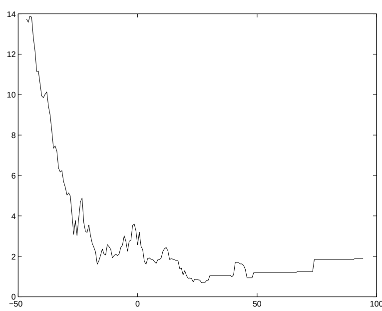
(a) Quantile index q=.1 (weeks of search = 1)



(b) Quantile index q=.3 (weeks of search = 4)



(c) Quantile index q=.5 (weeks of search = 8)



(d) Quantile index q=.7 (weeks of search = 14)

Figure 1-7: IQR Objective Functions (y-axis) vs. α (x-axis)

Appendix A

Standard and Instrumented Quantile Regressions Comparison

Table 1.3 presents the coefficients for the quantile regression models before and after instrumenting for a possible individual heterogeneity in the model. 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.

Table 1.3: 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	$\tau = .1$	0.405	0.000
		(0.005)*	(0.992)
3 weeks	$\tau = .2$	0.282	0.330
		(0.061)*	(0.830)
4 weeks	$\tau = .3$	0.401	0.712
		(0.070)*	(0.786)
6 weeks	$\tau = .4$	0.158	1.139
		(0.088)*	(0.796)*
8 weeks	$\tau = .5$	0.224	1.765
		(0.077)*	(0.879)*
12 weeks	$\tau = .6$	0.087	1.392
		(0.079)*	(0.829)*
14 weeks	$\tau = .7$	0.212	1.605
		(0.084)*	(0.896)*
22 weeks	$\tau = .8$	-0.007	0.831
		(0.091)	(0.896)
36 weeks	$\tau = .9$	-0.151	-0.061
		(0.114)*	(1.004)

Note: N=1183, standard errors in (parenthesis), (.)* indicates significance at the 90% level.

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.

Appendix B

Cox Proportional Hazards Model Results

Table 1.4 presents the results from the regression analysis using the Cox proportional hazards model.

Table 1.4: Effect of Internet Use on weeks of Search - Within State Movers Only

Instrument Mean Home Internet Use By:	First Stage Coefficient on Instrument	Second Stage Coefficient on Predicted Internet Use
NO INSTRUMENT		-0.034
		[0.47]
Age, Income	1.156	-1.453
	[1.76]	[1.79]
Age, Income, Children	0.870	-1.603
	[1.38]	[1.85]
Age, Income,	0.230	-1.448
Metro/Non-metro	[1.55]	[1.72]
Age, Income,	0.947	-1.345
Metro/Suburb/Non-metro	[1.64]	[1.62]

Note: t-statistics in brackets

In our regressions we control for differences in age, income, number of children in the household, household composition, number of income earners, previous neighborhood metropolitan status, state, and for various reasons for the move among home buyers in the NAR data. Without instrumenting for internet use while searching for a home with a predicted level of internet use by demographic categories, the effect of internet use is slightly negative, close to zero, and statistically insignificant. Once any possible individual heterogeneity among home buyers in their internet use is accounted for with an instrument such as the mean home internet use by age groups and income categories from the CPS sample, the effect of employing on-line resources On the duration of home search becomes quite more negative and considerably more significant. Because of the non-linear nature of the empirical model used here, the magnitude of the coefficient of predicted internet use does not easily lend itself to interpretation. The meaning of its sign however, is clear: the use of internet resources in home search tends to increase the duration of that search. Because of a negative sign in the Cox model specification, a more negative coefficient coefficient on internet use in the second stage implies a longer search duration as a result of internet use. This analysis suggests that overall, the internet affects the search by changing the types of houses a home buyer considers as part of the search rather than by simply speeding the arrival rate of offers as discussed in the theory section above.

Bibliography

[1] 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.

[2] Victor Chernozhukov. Conditional extremes and near extremes: Concepts, asymptotic theory and economic applications. Ph.D. Thesis, Stanford University, August 2000.

[3] Victor Chernozhukov. Inference on quantile regression process, an alternative. MIT Department of Economics Working Paper, January 2002.

[4] Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. MIT Department of Economics Working Paper, December 2001.

[5] 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.

[6] Joel L. Horowitz. The bootstrap. Department of Economics, University of Iowa, Iowa City, IA 52242, Prepared for the Handbook of Econometrics, Vol. 5., July 1998.

[7] 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.

[8] Thomas W. Malone, Joanne Yates, and Robert I. Benjamin. Electronic markets and electronic hierarchies. *The Communications of the ACM*, 30(6), 1987.

[9] James McQuivey, Kate Delhagen, Kip Levin, and Maria LaTour Kadison. Retail's growth spiral. *Forrester Report*, 1(8), 1998.

[10] Christopher A. Pissarides. *Equilibrium Unemployment Theory*. Cambridge, MA: The MIT Press, 2000.

[11] William C. Wheaton. Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, 98(6):1270–1292, December 1986.

Chapter 2

Internet Listing and Time on the Market for Residential Housing: A Simultaneous Equations Approach

2.1 Introduction

This study explores the impact of listing a residential property for sale using on-line resources on the time it takes until an offer to sell the home is accepted. Time on the market until a property is sold is a frequently explored topic in the real estate literature. A large number of researchers have studied the impact of various factors, circumstances and conditions on the length of time it takes a particular property to sell. Most of this substantial body of literature, starting with Cubbin [2] and Miller [11], has focused on the relationship between selling price and the duration until the house is sold. This duration is commonly referred to as time on the market. Cubbin argues that the homes' quality is assessed by its listing price. He concludes that higher price decreases the time on the market for homes. Miller, on the other hand, argues a positive relationship between sales price and time on the market, but the results obtained by Miller are somewhat inconclusive. These authors have focused on the impact of time on the market (an independent variable) on the selling price.

Others have attempted to explain the time on the market (the dependent variable) as a function of the selling price. See Belkin, Hempel and McLeavey [1] as just one example. The authors of this study estimate the time on the market using the spread between the selling and the listing price of the home. While the authors recognize the importance of the list price in determining the time until sale, they treat the list price (and the selling price) as exogenous. Further studies in this literature, such as Kang and Gardner [10] estimate the impact of market conditions such as mortgage rates, volatility and financing, or house attributes such as age of the home and its amenities on time on the market.

In many of the studies conducted, the importance of all three variables: original listing price, selling price and time until sale has been recognized. Yet by and large, researchers have focused either on estimating time on the market or on predicting the selling price, while treating the remaining two variables as exogenously determined. Only recently have some authors used simultaneous models for the determination of time on the market and sale price. Huang and Palmquist [8] use a two equation structural model for the impact of environmental conditions on these two variables. They acknowledge that the choice of an original list price affects the outcome of both the selling price and the time on the market. Yet, list price is included neither in their model nor in their estimation approach. Work by Green and Vandell [5] considers the choice of an asking price by the seller. The authors develop an optimal asking price and time on the market model based on maximization of a net present value for the home. A reservation price for the seller is based on the stream of bid arrivals. In their model, the asking price is allowed to adjust while the home is on the market.

In continuation of the trends in this literature, in the present study we consider a simultaneous equations model for the determination of the actual sale price and the duration of time on the market. We refine the empirical model connecting listing price, selling price and time on the market. Unlike previous research on this topic, in our model we include the choice of a sale strategy by the seller or his/her agent which affects the above three variables. The list price, and the decisions leading to the method used for listing the home are pre-determined, endogenous variables that affect the simultaneously determined selling price and time on the market. The listing method includes the choice of the seller to use an agent or to sell the property oneself, and the choice whether to use the internet as a means of listing the property for sale.

In this study we aim to answer the question: how does internet listing influence the time on the market. Unlike the large number of studies in the pre-existing literature, we are not concerned with the interplay between list, or sales price and time on the market. By estimating a five equation model involving the choice to use an agent, the choice to list using the internet, the original asking price, the selling price and the time on the market, we account for the endogeneity among these variables, even though we are interested only in the relationship between internet use and time on the market. Since the relationship between selling price and duration until sale is not of interest here, we are able to use a somewhat reduced form of the model which is more robust to the particular specification than previous models in the literature.

Wheaton [12] has developed a model of search and matching in the housing market in which all home buyers are sellers as well, as opposed to models where new construction or passing homes down to the next generation dominate the activities in the housing market. Results from previous work by the current author show the duration of search, until a home to purchase is found, increases when the internet is used by the buyer in the home search. If the housing market behaves according to Wheaton's model, these results suggest we should expect to find an increase in the duration until a home is sold when the internet is used to list it [1]. Since in this model

[1] According to Wheaton's model, revized to include the effect of internet use in the housing market by the current author, the time until sale is $z = \frac{\lambda(1-F(R))H_s}{V}$. (See previous chapter.) When the pool

the buyers are also sellers, we expect the amount of time by which the sale is slowed down to correspond to the amount of time by which the home purchase is slowed down as a result of internet use in the market.

We use data from the 2000 Home Buyer and Seller Survey, a US national level survey collected by the the National Association of Realtors. In this data, we find an increase in the time on the market for residential homes in the US in response to listing the property on-line. Furthermore, the amount of time by which the duration until sale increases is almost exactly the same as the amount of time by which search duration increases on the buyer's side of the market. The results of this study, together with our previous results involving home buyer behavior and internet use show evidence for a housing market where all buyers are sellers as well.

2.2 Theoretical Discussion

We consider the seller's choices leading to the sale of the home in two steps. The first step is to determine the sale strategy. This involves three simultaneous decisions: use an agent or sell as owner, use the internet to list or no on-line listing, and select the initial listing price. These decisions temporally precede the accepting or rejecting of subsequent offers that determine the actual sale price and the duration of time on the market. The maximization of the objective (the net present value of the home) comes next. Figure 2-1 describes the time-line of the seller's decisions involved.

Many of the studies in the literature, including Huang and Palmquist [8], and Green and Vandell [5], use a hedonic reservation price model to describe the relationship between the observed sale price and the unobserved reservation price of the seller. The reservation price determines the probability of accepting an offer and thus

of potential buyers becomes more diverse as result of listing on-line, or when a set of better suited to the home potential buyers view this home on-line, $F(R)$ increses while the rest of te parameters stay constant. As a result, z increases by the same amount as the duration until a home buyer finds a home to purchase increases.

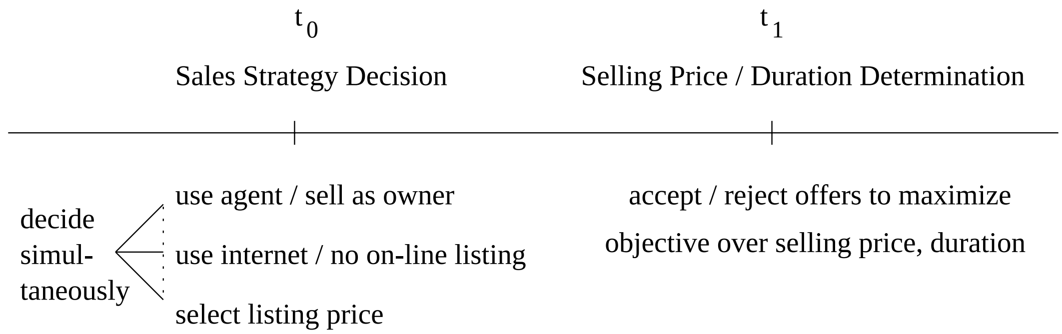


Figure 2-1: Temporal Aspects of the Selling Decision Process

the duration until the house is sold. Furthermore the seller's reservation price can adjust as more information is gathered during the time the house is on the market through the distribution and the frequency of arriving offers. It can influence not only the duration until sale, but the actual sale price by affecting which particular offer is accepted. This gives rise to the simultaneous determination of the observed selling price and time on the market.

The list price depends on the (initial) reservation price of the seller and therefore it affects both the sale price and the duration until sale. In this model, the list price itself is pre-determined with respect to both the actual selling price and the duration until sale. Because of the temporal separation between selecting a list price and accepting an offer with a particular sale price, the list price cannot be directly affected by either the selling price or the on the market duration. Another difference of our model with previous models is the addition of the seller's decisions about the method used to sell the home. This involves the set of choices to use an agent or to sell as owner, and whether or not to list the property for sale on the internet. Again, the temporal separation between the two steps in the decision process, t_0 and t_1 in figure 2-1, forces the sales method to be pre-determined with respect to the actual selling price and the time on the market. The decision to use an agent or not, to list on the internet or not and the original list price are simultaneously determined and endogenous with respect to each other.

2.2.1 Sales Strategy Decision

The decision to use on-line listing is affected by the choice to use an agent and the selection of a particular list price. As agents might have more knowledge and better access regarding listing a home for sale on-line than individual home sellers have, the decision to use the internet and to use an agent are co-determined. This is particularly relevant for the time-frame of our data[2]. If a home seller is intent on using the new internet technology, perhaps to increase the pool of potential offers, he or she might decide to use an agent to assist with listing the home on-line. The costs associated with listing a home for sale on the internet are higher than the costs associated with conventional listings. These costs include not only the know-how needed, but also equipment such as digital cameras to provide the virtual tour photos, the computer hardware and software and the internet connections needed, and exceed the costs of a conventional listing methods. Consequently, only home sellers with a high reservation price might be using the internet in their sale strategy, so that only homes in a certain range of initial asking prices would be listed on-line.

Thus, the probability of using internet listing in the sale strategy is given by the function $g_1(.)$:

$$\Pr(I_j \neq 0 | A_j, L_j, X_j) = g_1(A_j, L_j, X_j; \beta_1) + \epsilon_1,$$

where I_j is a 0, 1 indicator for internet use in listing the home, A_j, is a 0, 1 indicator of whether an agent's services are used to assist with the sale, L_j, is the original list price, and X_j includes demographic characteristics of the seller, and geographic characteristics of the home.

Similarly, the decision to use an agent in the sale is determined by the function $g_2(.)$:

$$\Pr(A_j \neq 0 | I_j, L_j, X_j) = g_2(I_j, L_j, X_j; \beta_2) + \epsilon_2.$$

[2] In 1999, when our data was collected, the technology needed to list a home for sale on-line would have been more readily available to real estate professionals than to individual home sellers.

Agent use is affected by the choice to use the internet and the selection of an initial list price. A higher list price may prompt the seller to employ a real estate agent. The match between a seller with a high listing price for the home and a buyer with an appropriate level of housing consumption is more difficult to achieve when the home is offered for sale by owner. A home with a high listing price would require a larger, or more specific pool of potential bids. This can be obtained more easily with the help of a real estate agent. In addition, if a home seller is intent on using on-line resources to list the home, an agent's knowledge may be needed to help list the home on the internet.

The asking price is affected by the decisions to use an agent, and to use the internet. An agent can recommend a range of asking prices. This asking price range might vary depending on the attributes of the house and the urgency of the seller to sell the home.

It might also vary depending on whether the internet is used to list the property, especially if the home seller or agent is looking to recover some of the costs of listing on-line.

The third equation in the joint decision of the selling strategy defines the asking price,

$$L = g_3(I_j, A_j, X_j; \beta_3) + \epsilon_3,$$

as a function, $g_3(.)$, of the decision to use the internet to list the home as opposed to using only conventional ways to list it (such as a newspaper advertisement), (I), the decision to use an agent to assist with the sale (A), and the above demographic characteristics of the seller and geographic characteristics of the home, (X_j). In each of the above equations, β_i is a vector of parameters, and ϵ_i is a stochastic component.

We adopt a linear specification, and write these three equations in reduced form with respect to each other as:

$$A = X'\alpha_1 + \epsilon_1 = \hat{A} + \hat{\epsilon}_1,$$

$$I = X'\alpha_2 + \epsilon_2 = \hat{I} + \hat{\epsilon}_2,$$

and

$$\log(L) = X'\alpha_3 + \epsilon_3 = \widehat{\log(L)} + \hat{\epsilon}_3.$$

where A is the agent use indicator, I is the internet use indicator, and L is the original list price which enters in logarithmic terms here. The vectors α_i denote the reduced form coefficients on the matrix of covariates that enters in each equation.

2.2.2 Hedonic Reservation Price and Time on the Market

The difference between the list price and the actual sales price has been described using a reservation price model. Following Huang and Palmquist [8], sale occurs when the offer is higher than the seller's reservation price. The observed selling price (P) must then be a lower bound for the seller's reservation price ($P^\star$). It is common in hedonic price models and in stochastic frontier models in general, see Aigner at. al., [3] to assume a truncated error structure. Here, the truncation of the error is from below. As $P = P^\star + u$, the random component, $u = P - P^\star > 0$.

The reservation price of the seller is determined by the attributes of the home, such as the type of structure and geographic location as well as the seller's urgency to sell, the number of times the home has been previously placed on the market by this seller and the duration the house has been on the market and remained unsold. Thus,

$$P = P^\star + u = h(T, Y; \gamma) + v + u,$$

where $h(T, Y; \gamma)$ is the expected reservation price of the home seller. Here, T is the duration of time on the market and Y is a vector of exogenous parameters such as market conditions and geographic, locational and structure type characteristics of the home. There are two random variables: $-\infty < v < \infty$ and $0 \leq u < \infty$. One reflects the usual error and the other reflects the systematic error between the sales price and the reservation price. Huang and Palmquist [8] estimate this equation, together with the time on the market equation in their model, using a maximum likelihood technique which reflects the truncated nature of the error in the sales price equation[3]. Since the authors above do not find an empirical advantage to the more complicated error structure in this model, for computational reasons we shall assume a bivariate normal distribution for v and u here and ignore any truncation in the error.

The time on the market until the house is sold is governed by a hazards model with:

$$log(T) = k(P^{\star}, Y; \delta) + \nu,$$

The parameters of the function $k(.)$ are designated by the vector δ, and ν is an additive error term. We will assume that ν is normally distributed. The functional form of the duration equation we use in our estimation is:

$$\log(T) = -\log(\delta) + \delta' \log(P^{\star}) + \delta' Y + \nu.$$

In a regression setting, the above would simply become:

$$\log(T) = -\log(\delta) + \delta' \log(P) + \delta' Y + \nu.$$

In this specification, the probability of receiving an offer increases at first and the decreases as time goes by.

[3] Huang and Palmquist [8] calculate their results using a FIML technique assuming normally distributed errors as well. They do not find a significant difference between the results of the model accounting for a truncated error structure and the model assuming normal errors in the sales price equation.

There is a number of functional forms used in the survival analysis literature for describing duration, as well as different methods of estimation. The functional form above allows for a relatively straightforward estimation of the joint model. The above model for the time on the market does not account for any possible truncation in the data for duration until sale. It is possible that some homes remain unsold after a given period of time and are taken off the market by the seller. We do not include a truncation correction in our model because of the specific nature of the data used here. In our sample, after removing observations with other missing data, no homes remained unsold.

Since we are not interested in the impact of sales price on the duration on the market, the above two equations can be written in reduced form with respect to each other, so that:

$$\log(P) = \hat{A}'a_1 + \hat{I}'a_2 + \widehat{\log(L)}'a_3 + Y'a_4 + \nu_4,$$

and

$$\log(T) = \hat{A}'b_1 + \hat{I}'b_2 + \widehat{\log(L)}'b_3 + Y'b_4 + \nu_5.$$

The above five equations form a model involving the vector of endogenous, left hand side variables (I, A, L, P, T). We solve the system of these five equations by simultaneously estimating the sales strategy equations as a first step, and using the predicted values ($\hat{I}$, $\hat{A}$, and $\widehat{\log(L)}$) in the second step, simultaneous estimation of the sales price and duration equations.

The existence of a correlation between ϵ_i in any of the sales strategy equations, and ν_j in the sales price or duration equation in our setup is quite possible. The reservation price of the seller affects both the sales strategy decision and the acceptance of potential offers. A home with a high initial asking price is likely to have many amenities such as large rooms, fireplaces and so on. The seller of such a home is likely to use the services of an agent and perhaps also list it on-line, in order to gain exposure to a wider pool of buyers including those who prefer high levels of housing consumption. The high reservation price would also induce the seller to be choosy about the offers he or she receives, affecting the sale price and duration. In our data, reservation prices and house amenities are unobservable, and the errors terms between any of the sales strategy equations and the sale price/duration equations are likely to be correlated.

2.3 Estimation Strategy

We employ a Three Stage Least Squares (3SLS) technique in the estimation of this model. Since we assume normally distributed errors, 3SLS is asymptotically equivalent to the Full Information Maximum Likelihood (FIML) estimation of this linear, five equation system, and of course, much easier to compute than FIML. Since we are only interested in the impact of internet use in the sale strategy for the duration of time on the market, we can use a somewhat reduced form system rather than the structural form of the equations we present in the theoretical discussion.

Many of the studies in the housing market literature have focused on the relationship between selling price and time on the market. In order to estimate this joint relationship, researchers such as Huang and Palmquist [8] estimate the structural form simultaneous equation system. A similar approach here is not not necessary. In addition, any misspecification in any one equation can propagate throughout the structural form system and affect the results in all equations[4]. Yet, as we are interested in the impact of one left hand side variable, internet use (I), on another left hand side variable, duration on the market (T), in another equation, we cannot use a completely reduced form estimation of the five equation system.

The estimation technique we employ is therefore similar to an equation by equation estimation of a recursive system of linear equations. In the triangular, or fully

[4]See Greene [6], p.760

recursive system, each equation contains a left hand side variable that is predetermined with respect to those in the following equation. A consistent estimation of all equations of the system can be obtained. by performing Ordinary Least Squares (OLS)) on the first equation, using the estimates form it in the OLS regression of the second equation, and so on. Since this technique results in consistent, but not efficient estimates, robust errors for the estimates need to be computed as well.

We can use a similar technique in the estimation of the simultaneous system at hand. This five equation system is not strictly triangular, but in a sense, block triangular. The first three equations are pre-determined with respect to the last two. We write the first three equations which involve the sales strategy decision in a reduced form with respect to each other:

$$A = X'\alpha_1 + \epsilon_1 = \hat{A} + \hat{\epsilon}_1,$$

$$I = X'\alpha_2 + \epsilon_2 = \hat{I} + \hat{\epsilon}_2,$$

and

$$\log(L) = X'\alpha_3 + \epsilon_3 = \hat{\log(L)} + \hat{\epsilon}_3.$$

where A is the agent use indicator, I is the internet use indicator, and L is the original list price which enters in logarithmic terms here. The vectors α_i denote the reduced form coefficients on the matrix of covariates that enters in each equation. The matrix X includes demographic characteristics of the sellers such as income, age and number of children, the urgency of the seller to sell the primary reasons for the move, a measure of the number of internet listings of homes for sale in the seller's state, and the type of property.

We use a linear probability model for the first two equations. While we acknowledge the shortcomings of using this regression form for the 0, 1 indicators in this model, the inclusion of the third simultaneous equation for the original asking price makes this estimation technique the most computationally feasible one. In addition, results obtained by a bivariate probit estimation on the first two equations give very similar results to the ones obtained with a joint, linear probability estimation on these two equations. As long as there is no cross correlation between ϵ_i in one equation and an estimate of the respective left hand side variable in another, this technique produces consistent estimates for the internet use, the agent use and the sales price. There is no reason to expect that such correlation exists here. Thus, we can proceed with the 3SLS estimation of the reduced form system for the sales strategy. This will provide a first step to the estimation of the effect of internet use on the duration until sale in the housing market.

We use estimates for I, A, and L from the 3SLS estimation of the above three equations in the joint regressions of the estimation of P and T according to:

$$\log(P) = \hat{A}'a_1 + \hat{I}'a_2 + \widehat{\log(L)}'a_3 + Y'a_4 + \nu_4,$$

and

$$\log(T) = \hat{A}'b_1 + \hat{I}'b_2 + \widehat{\log(L)}'b_3 + Y'b_4 + \nu_5.$$

Here, Y is the matrix of covariates including market conditions, such as seasonal and regional indicators, the number of times the home has been on the market, variables describing the type of home, the primary reason for the seller's move, and the urgency of the seller to sell the home. We use a 3SLS technique for the joint estimation of price (P) and time on the market (T) as well.

The coefficient b_2 is of interest in this study. It links internet use in the listing strategy with the time on the market for the home. As the time on the market is in logarithms, and internet use (I) is a variable that takes only the values

0 or 1, b_2 has the interpretation of elasticity. Exponentiating this coefficient (e^{b_2}) gives the change in duration until sale in terms of T when on-line listing is used in

the selling strategy. The robust standard errors in the sales price and in the duration equations are calculated using a parametric bootstrap technique [7].

2.4 Data

This study uses a cross-sectional survey data collected by the Research Division of the National Association of Realtors (NAR). The data is summarized in the 2000 NAR Profile of Home Buyers and Sellers [4]. The survey includes recent home buyers from all over the US. At the beginning of the year 2000, the NAR mailed a questionnaire 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. From these observations we use answers to the question asked by the survey of home buyers: What did you do to your previous home? To this there were around 1,000 non-blank answers that include selling the home, attempting to sell unsuccessfully, or hold it as investment property or as second home.

For those respondents who attempted to sell their previous home, the survey also contains information about the ways the home was listed, including listing using the internet. Whether an agent was used in the sale and the type of agent was given. The survey contains information about the original asking price, the selling price, the duration until an offer was accepted, as well as some geographic and demographic information about the seller. The survey does not contain data about the house features, such as size, age or specific location and amenities of the property.

Table 2.1 lists the descriptive statistics of the data used here. The average respondent to the home seller part of the NAR survey is between 44 and 49 years old. The average income is between \$50,000 and \$59,999 in 1999 dollars. The mean number of children for the respondents here is 1.79 and 43.4% of the respondents have any children. There are 74.4% married respondents and only 2.9% are Black. In terms of demographics, the average home seller in the NAR sample has a higher income and has slightly more children than the average person in the US as recorded by the 2000 Current Population Survey (CPS). The CPS, conducted biennially by the US Census Bureau is representative of the population in the US at large. There are substantially fewer Blacks and Hispanic Origin respondents in the NAR survey used here than there are in the general US population in 2000. This is to be unexpected, as the respondents here are homeowners, who tend to be more affluent than the average US resident. Credit markets might discriminate against potential homeowners on the basis of race and Hispanic origin. This could account for the substantially lower number of Black and Hispanic origin home sellers recorded here than residents in the general US population. Homeowners tend to be more frequently married as well. The average number of married respondents in the 2000 CPS is 65.6%, and in our NAR sample there are 74.4% married respondents. Perhaps this is due to credit markets favoring married couples over single home buyers as well.

The average number of years the home sellers in this survey have owned the property before placing it on the market is just under 10 years. About 48% are in a rush to sell it and the primary reason for making the move (and perhaps selling the previous home) most frequently mentioned is the desire for more space, followed by corporate relocation or a new job. When an attempt to sell the previous home was made by the respondent, only 3.34% tried to sell unsuccessfully and consequently still own the home. Respondents in this survey who tried to sell, placed the home on the market anywhere between 1 and 5 times, the mean being 1.19 times, and only 14% tried more than once. The average number of weeks the home was on the market is 12.8 for the entire sample and 9.53 weeks for the homes that sold on the first try. The average asking price is \$168.3 thousand and the average selling price is \$161.9 thousand. The average selling price is lower than the average asking price as to be expected from the reservation price model. However, there are 66 observations in our sample where the sale price exceeded the asking price.

In the NAR sample of home sellers, 68.6% used an agent to sell the home and 13.9% listed the home for sale by owner. Whether an agent was used or not, 27.9% of the home sellers indicated that the internet was used as a method to list the home. This overall internet use number can be broken down into internet use when and agent was employed in the home sale process and when the home was listed for sale by owner. Among the agent assisted homes on the market, 37.5% used the internet as a listing method for the property. In contrast, only 6.87% percent of the for sale by owner homes used the internet to list the home. In 1999 more agents had access and knowledge about listing a property using on-line resources. At that particular time, not many individual homeowners had the resources and the knowledge necessary to list their home on the internet. This is reflected in the much smaller number of properties for sale by owner in our sample listed using the internet. This shows there is a differential advantage between homeowners and agents relating to listing a home for sale using the internet. The choice to list on-line or not depends on the choice to sell with or without the services of an agent. This difference in the skills and resources needed to list a home for sale on the internet may be decreasing as time progresses. In the fast changing environment of internet use in the recent past, even a couple of years bring a substantial change in the categories of internet savvy individuals. However, in 1999 the use of digital technology and environments is just becoming popular. At this time, home sales professionals are already able to use this technology to list a home for sale, while the average private individual by and large, is not equipped to do so. In 1999, many private individuals had sufficient knowledge to search for home to buy using the internet, but as listing a home for sale on the internet requires a lot more internet specific skills and new equipment, such as digital cameras, the usual homeowner in 1999 would not have been able to list a home for sale using the internet by himself or herself.

The homes offered for sale using an agent have a higher mean initial asking price and a higher mean selling price than the respective mean prices for the homes offered for sale by owner. On average, higher priced properties are sold when internet listing was used than when the home was not listed using the internet. This points to a decision or a strategy about how to sell a home involving whether to use an internet listing, whether to use an agent to assist with the sale and at the same time what initial asking price to select. These joint decisions then determine the selling price and the duration until the house is sold. The data points to a possibility that this simultaneous decision process is in fact happening here.

2.5 Results

The results from the first stage in the estimation are presented in table 2.2. Higher income increases the initial asking price. This is consistent with the possibility that a higher income corresponds to a higher reservation price of the seller. Higher income could also be correlated with owning a larger, better equipped home. The attributes of the home for sale, such as living area size, number of rooms, fireplaces and so on are not available in our data. The observed effects on income could be largely due to home characteristics that are unobservable here.

The number of total internet listings in the seller's state of residence has a somewhat negative impact on the initial list price. The larger the fraction of homes that are offered on the internet for sale relative to the total number of listings in the state, the lower the initial asking price. This effect could signal that there is less friction in a housing market in which buyers are able to locate a suitable choice to purchase faster an easier through the internet. In this situation, the housing market becomes more competitive and the price sellers can expect to get for their home is reduced.

Homes within cities are listed with a lower initial asking price that other homes and so are homes in areas where the primary reason for the move indicated by the seller is decline of the neighborhood. This relationship between location an reservation price, reflected in the original asking price does not come as a surprise, and is consistent with the stylized facts in the housing market literature about the desirability of home location.

Higher income also increases the probability that the internet is used as part of the selling strategy. An increase in the age of the seller corresponds to a higher asking price and a lower probability of using the internet. Age itself may be correlated with owning a larger, or more expensive home. The effect is similar to the relationship between income and original asking price with relation to (unobservable here) house characteristics. Older sellers are less likely to employ the internet as a means to list their home. Older individuals may be less aware of the possibility of on-line listing for a home on the market. They are therefore less likely to employ the internet when they are listing a home for sale by owner. They are also not likely to consider the possibility of an agent using on-line resources to list the home, and they would not necessarily seek out and employ a particularly internet savvy agent. Both of these factors contribute to the relationship between the seller's age and internet use for listing the home for sale. These considerations are particularly relevant to the time of the data sample used in this study. In 1999 the internet had just started influencing economic activities. Current Population Survey data from this time period suggests that age is by far one of the biggest factors, together with education, that influence internet use for various purposes. As income is a reasonable proxy for educational attainment, the effect of income on internet use could be due to the effect of education on internet use in this data. The effect of income and age on internet use when listing a home, or employing an agent to do so are consistent with established facts about an individual's demographics and internet use in various activities.

Home sellers in the Western region of the US are more likely to use the internet as means to list a home relative to all other geographic regions. The asking price is also significantly higher in the Western US than in all other regions. It is possible that this effect is due to a housing market in the Western region of the US that sustains both a higher asking and a higher sales price compared to the rest of the country. This possibility is supported here by the significantly higher selling prices we find in the Western US in our regressions. In addition, it is likely that a higher proportion of internet savvy home sellers live in the Western US, and thus are more likely to use the internet to list a home for sale. As shown by our sales price regression, a higher list price is not associated with more internet listings. Thus, internet savvy, rather than an expensive home to sell may be driving the decision to use the internet in the selling strategy. None of the variables examined here have a significant impact on the decision to use an agent in the listing and sale of the home.

The estimation of simultaneously determined selling price and time on the market is presented in table 2.3. The table shows the coefficients from the 3SLS regression for the effect of the covariates on each variable: logarithm of the selling price and logarithm of the weeks until an offer for the sale of the home was accepted[5]. The internet used, agent used and the logarithm of the asking price variables in table 2.3 are the residuals from the joint estimation of internet use, agent use and the logarithm of the asking price from table 2.2.

The most important result to note is that using the internet to list a home significantly increases the duration of the home on the market compared to using conventional listing methods. The elasticity of time on the market with respect to internet use is 0.347 This translates in an increase of in an increase of about 1.4 weeks added

[5]There are some observations in the data were sale that did not occur, and homeowner still owns the property after placing it unsuccessfully on the market. However, after excluding observations with missing data for any of the variables used none of the observations resulting in no sale were usable in the regressions performed. Assuming (as it is customary) that data is missing at random there is no need to correct for a truncation problem in the data.

to the time on the market at the mean of the distribution of sale times. The mean selling time for this sample is 12.8 weeks until an offer for sale is accepted. This result is consistent with our previous findings pointing to an increase in the duration until a buyer in this market finds a home to purchase. The above two results are consistent with Wheaton's model of equilibrium in the housing market where all buyers are sellers. If the internet use as a medium for buying and selling homes increases the duration until a home is bought, in an equilibrium situation it is natural to expect that the internet increases the duration of sale by the same amount. Previous findings about the role of the internet in the duration of search point to an increase of about 1.6 weeks at the mean duration of search. The mean duration of home search for the sample of home buyers who previously owned a home is about 14 weeks of search until finding a home that is eventually purchased by this home buyer.

We find that the more times on the market a home has been placed, the longer it takes for it to sell. This is not surprising: a home that does not sell the first time a seller attempts to do so, is likely to stay in the market longer in subsequent attempts. This would happen when the home has undesirable features either in terms of amenities or location. Subsequent attempts to sell this undesirable property result in longer times on the market even when it does sell. Alternatively, the home may be incorrectly priced the first time an attempt was made to sell it. As the seller gains information about the market, he or she can make revisions to his or her reservation price in subsequent attempts to sell it. However, there is no significant impact either on the initial asking price or on the resulting sales price of the number of times the home was offered for sale by this homeowner. Therefore, mis-pricing and subsequent correction is not likely to be driving the number of times a home is placed for sale. Rural homes take longer to sell than other properties, and so do ones in declining neighborhoods. Sellers for whom the primary reason to move is decline of the neighborhood, sell the home for a lower average price than when the primary reason for the move is not associated with neighborhood decline.

The sale price is not significantly affected by the decision to use the internet in the sale strategy. The variables that have the largest impact on the selling price are income and original asking price. The higher the seller's income the higher the selling price. It is likely that larger, more desirable and consequently more expensive homes are owned and sold by individuals with higher incomes. Since in this data we cannot control for the amenities of the home, this effect is expressed here through income. The higher the asking price, the higher the selling price as well. If a reservation price is driving the seller's selection of an initial asking price, and also influencing the acceptance of offers, we expect to find a positive effect of asking price on the selling price. The location of the home with relation to the type of neighborhood influences the selling price. Relative to resort properties and other types of homes, city neighborhood and suburban homes sell for less, by about the same amount. Rural properties on average are less expensive than resort properties and other homes, but are closer in price to them than city and suburban homes are.

Using an agent to assist with the sale has no effect on selling price. This result is consistent with previous studies in the literature. Jud [9] finds that real estate brokers do not influence the prices of the houses they sell relative to homes sold by the owner. He finds that the brokers instead influence the level of housing consumed. One interesting result that comes out of this analysis is that using an agent in the sale increases the duration of time on the market by about 1.6 weeks. While agents may not speed up the sale of a home relative to properties for sale by owner, they may have a role in facilitating the process after the an offer for sale has been accepted. It is also likely that the main role if an agent representing the seller is to provide a flow of appropriate buyers, for example by helping list the property on line, rather than to speed up the acceptance of such offers. Once the role of an agent to assist with the listing and sale strategy has been accounted for, the presence of an agent actually increases the time on the market.

2.6 Conclusion

Time on the market, initial asking price, the actual selling price and the decisions about the selling strategy in the residential housing market are jointly determined. Previous research in this area has not estimated this simultaneous model in its entirety, frequently resulting in endogeneity induced bias of the results. This study presents a econometric model for the joint estimation of the duration of time on the market, the selling price, the asking price and the decision to use and agent in the sale and to use on-line resources to list the home. After accounting for the endogenous determination of the five variables above, we concentrate on examining the effect of internet listing on the duration until an offer for sale is accepted.

We use a 3SLS estimation approach for the analysis of the impact of internet listing on the duration until sale in the residential housing market. We estimate a somewhat reduced form of the five equation system which allows us to discern the impact of on-line listing on the duration of time on the market for residential housing in the US. Many researchers in this area have worked on estimating the relationship between selling price and time on the market. In this study, however, we do not focus on the relationship between selling price and duration until sale. Concentrating simply on internet listing and time on the market, we can employ a more robust to exact specification form of our regression equations than a structural form model estimation would allow. We use data from the 2000 Home Buyer and Seller Survey collected by the National Association of Realtors. We find that using the internet to list a house increases its time on the market. The results presented here are consistent with previous findings pertaining to the use of the internet and the duration of search until a buyer locates a home to purchase. These results, together with the findings of the present study show for a housing market where all buyers are sellers.

We find that internet use as part of the listing strategy in the housing market in the US increases the duration of time on the market by about 1.4 weeks at the mean duration of time on the market. This result is to be expected in a situation in the housing market where all buyers are sellers. In a previous study of the effect of internet use on the buyer's side of the housing market we find a similar increase in the duration of search when the internet is used. These two results suggest a housing market as modeled by Wheaton in [12]. When all buyers are sellers, we find evidence that as buyers search for longer periods using the internet, the corresponding properties listed for sale on the internet take longer to sell as well. Internet listing as part of the sales strategy has no effect on the selling price of the home.

Table 2.1: Means and Percentages of Key Variables for Home Sellers

	Mean	Percent	N
1. Age group	5.48	—	914
2. Income Category	7.66	—	861
3. Number of Children	1.79	—	922
Has any children	—	43.4	922
4. Married	—	74.4	922
5. Black	—	2.90	897
6. Hispanic Origin	—	4.00	850
7. Location of Previous Home	—	—	920
Within a City	—	43.8	920
Suburban Neighborhood	—	41.3	920
Rural	—	12.0	920
8. In a Rush to Sell	—	47.8	655
9. Length Owned (years)	9.85	—	914
10. Previous Home	—	—	928
Sold using agent	—	68.6	928
Sold by owner	—	13.9	928
Hold as Investment	—	8.51	928
Tried to Sell Unsuccessfully	—	3.34	928
11. Times This Seller Tried to Sell	1.19	—	716
More than once	—	14.0	716
12. Time on the Market (weeks)	12.8	—	719
13. Original Asking Price (thousands)	$168.3	—	671
14. Actual Sale Price (thousands)	$161.9	—	682
15. Internet Used to List	—	27.9	928
Agent and Internet Used	—	37.5	637
For Sale by Owner and Internet Used	—	6.87	291

Note: Income categories: 1 - Under $25,000, 2 - $25,000 to 29,999, 3 - $30,000 to 34,999, 4 - $35,000 to 39,999, 5 - $40,000 to 44,999, 6 - $45,000 to 49,999, 7 - $50,000 to 59,999, 8 - $60,000 to 69,999, 9 - $70,000 to 99,999, 10 - $100,000 to 124,999, 11 - $125,000 to 149,999, 12 - $150,000 or over. Age groups: 1 - less than 25 years, 2 - 25 to 29, 3 - 30 to 34, 4 - 35 to 39, 5 - 40 to 44, 6 - 45 to 49, 7 - 50 to 54, 8 - 55 to 64, 9 - 65 years or older.

Table 2.2: Joint Estimation of Internet Use, Agent Use and Original Asking Price

	Coefficient in Equation for LN(Asking Price) *(1)*	Coefficient in Equation for Internet Use *(2)*	Coefficient in Equation for Agent Use *(3)*
1. Income	0.094	0.016	0.005
	(0.010)*	(0.008)*	(0.006)
2. Age	0.086	-0.015	0.013
	(0.015)*	(0.011)	(0.009)
3. Number of Children	-0.005	0.006	0.016
	(0.030)	(0.022)	(0.017)
4. In a Rush to Sell	0.030	-0.048	0.050
	(0.057)	(0.042)	(0.041)
5. Number of Homes Listed on	-0.038	-0.002	0
the Internet in Each State	(0.014)	(0.010)	(0.008)
6. Location:	-0.384	0.073	0.126
Within a City Neighborhood	(0.175)*	(0.128)	(0.100)
7. Location:	-0.256	0.036	0.144
Suburban	(0.174)	(0.128)	(0.100)
8. Location:	-0.252	0.142	0.178
Rural	(0.183)	(0.135)	(0.104)
9. Primary Reason for the Move:	0.033	0.029	-0.026
Space Considerations	(0.089)	(0.066)	(0.051)
10. Primary Reason for the Move:	0.094	0.029	0.028
Corporate Relocation or New Job	(0.109)	(0.081)	(0.063)
11. Primary Reason for the Move:	0.098	-0.054	-0.093
Health/Age	(0.174)	(0.128)	(0.091)
12. Primary Reason for the Move:	-0.281	-0.086	0.037
Decline of Previous Neighborhood	(0.091)*	(0.117)	(0.091)
13. Intercept	10.7	0.186	0.664
	(0.267)*	(0.197)	(0.153)*
R^2	0.31	0.06	0.04
N	514	514	514

Note: Standard errors in parenthesis, (.)* indicates significance to the 10% level. Type of home (single family, apartment, etc.) and US region indicators are included in the specification in each equation.

Table 2.3: Joint Estimation of Duration and Price Equations

	Coefficient in Equation for LN(Weeks on the Market) *(1)*	Coefficient in Equation for LN(Sale Price) *(2)*
1. Internet Used (residual)	0.347	-0.024
	(0.118)*	(0.037)
2. Agent Used (residual)	0.473	-0.019
	(0.157)*	(0.038)
3. LN(Asking Price) (residual)	0.008	0.898
	(0.075)	(0.070)*
4. Income	-0.023	0.096
	(0.021)	(0.006)*
5. In a Rush to Sell	0.115	0.014
	(0.104)	(0.032)
6. Number of Times This Seller	0.688	0.006
Previously Tried to Sell the Home	(0.086)*	(0.025)
7. Location:	0.301	-0.450
Within a City Neighborhood	(0.214)	(0.091)*
8. Location:	0.095	-0.307
Suburban	(0.213)	(0.082)*
9. Location:	0.565	-0.294
Rural	(0.226)*	(0.082)*
10. Primary Reason for the Move:	0.026	-0.090
Space Considerations	(0.148)	(0.059)
11. Primary Reason for the Move:	-0.040	-0.078
Corporate Relocation or New Job	(0.196)	(0.064)
12. Primary Reason for the Move:	-0.038	0.396
Health/Age	(0.458)	(0.215)
13. Primary Reason for the Move:	0.777	-0.570
Decline of Previous Neighborhood	(0.245)*	(0.190)*
13. Intercept	1.64	10.9
	(0.412)*	(0.149)*
R^2	0.30	0.80
N	389	389

Note: Robust standard errors in parenthesis, (.)* indicates significance to the 5% level. Parametric bootstrap with 200 replications used for calculating the robust standard errors. Type of home (single family, apartment, etc.), season and US region indicators are included in the specification (both equations).

Bibliography

[1] J. Belkin, D. J. Hempel, and D. W. McLeavey. An empirical study of time on the market using multidimentional segmentation of housing markets. *Journal of the American Real Estate and Urban Economics Association*, 4:57–75, 1976.

[2] J. S. Cubbin. Price, quality, and selling time in the housing market. *Applied Economics*, 6:171–187, 1974.

[3] C. A. Knox Lovell D. J. Aigner and P. Schmidt. Formulation and estimation of stochastic fronteer production function models. *Journal of Econometrics*, 2:21–38, 1977.

[4] 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.

[5] R. Green and K. Vandell. Optimal asking price and bid strategies for residential sales. University of Wisconsin, Madison, WI, 1994.

[6] William H. Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River, New Jersey, 07458, 1997.

[7] Joel L. Horowitz. The bootstrap. Department of Economics, University of Iowa, Iowa City, IA 52242, Prepared for the Handbook of Econometrics, Vol. 5., July 1998.

[8] Ju-Chin Huang and Raymond B. Palmquist. Environmental conditions, reservation prices and time on the market for housing. *Journal of Real Estate Finance and Economics*, 22:203–219, 2001.

[9] G. D. Jud. Real estate brokers and the market for residential housing. *Journal of the American Real Estate and Urban Economics Association*, 11:69–82, 1983.

[10] H. B. Kang and M. J. Gardner. Selling price and marketing time in the residential real estate market. *Journal of Real Estate Research*, 4:21–35, 1989.

[11] N.G. Miller. Time on the market and selling price. *Journal of the American Real Estate and Urban Economics Association*, 6:164–174, 1978.

[12] William C. Wheaton. Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, 98(6):1270–1292, December 1986.

Chapter 3

Social Capital, Economic Incentives and Voter Turnout: Who Votes in Local versus National elections? Evidence from the US General Social Survey.

3.1 Introduction

This study explores the importance of social capital as defined through the connections one has with the local community and socioeconomic status for voter turnout in national as opposed to local elections in the US. What individual characteristics determine voter turnout in the two extreme levels of political participation through voting, and is there any systematic difference of those characteristics in different size communities? The questions pertaining to political participation are crucial to the understanding of many issues in political economy. Our definition and understanding of the ways through which citizens get involved in the political process encompasses a wide variety of activities. They include voting, political activism through interest groups, lobbying and informing others. Yet, a central issue in political participation remains to be voting.

Understanding voter turnout and its relationship to individual voter characteristics is important in itself. The the socioeconomic and demographic makeup of the particular voter's community is also of interest for turnout. Knowing who votes promotes our understanding of the economic and personal trade-off decisions an individual makes when governmental and redistribution issues are at stake. In addition, understanding voter turnout is crucial for estimation procedures in political economy that involve the median voter model. Pinpointing more closely the subset of potential voters in a community that do in fact influence the political decisions permits a better identification of the median voter. Subsequent tests of the median voter model can then employ the median turning out voter, rather than the median potential voter or resident.

Given its importance for a number of questions in political economy, the notion of who votes has been studied extensively and from many angles since the 1960's. In these studies, individual level survey data or aggregate voter turnout data has been employed to establish different types of results. In terms of the aggregate level studies, lower level elections have been characterized with a lower turnout than national elections in the United States [5] [1]. Related to this issue are the findings of Campbell [3], who showed that voter turnout increases in presidential election years but declines in midterm-election years at the national level in the US, and those of Burnham [1], who discovered the tendency in the US to vote for the more prestigious offices but not for the lower level offices listed on the same ballot. Cross national studies such as [5] suggest that the above facts for national versus. lower level elections may hold

[1] This fact does not hold for some other countries such as Japan.

for some and not for other countries. The question: what are the individual characteristics of the voters that turnout to vote in one but not in the other level of election remains, at least for US national versus. lower level elections.

Similarly, the evidence for the overall amount of turnout in municipal versus. national elections for small vs. large cities or urban vs. rural areas is mixed [8]. Sidney Verba and Norman Nie have summarized two possible models that govern the rural-urban turnout differences [11]. The first one is the degradation-of-community argument that supposes a smaller number of interpersonal and social relationships in larger cities. This leads to a smaller amount of monitoring of who participates in the political process through voting. It allows for voting "free riders" in the larger communities that have an easier time avoiding their civic duties in large cities, which leads to a lower turnout in all types of elections as the city size increases. In this sense, smaller communities have a higher level of political awareness that leads to an increased level of perceived political effectiveness. The effect of this political effectiveness perceived by the members of the community is then amplified as perceived effectiveness becomes realized. The authors' investigations find support for this decline-of-community in larger cities model. The second model, widely supported by Milbrath [7], argues that increasing city size facilitates the flow of information, and with the increased number of informed potential voters comes a higher turnout in urban areas. The urban population is more frequently bombarded by political stimuli in the form of information about political issues from other voters. There may also be an additional effect of city size on voting turnout if perhaps higher income, more educated potential voters live in larger cities, and those characteristics are positively correlated with voter turnout. In contrast to the first model, this argument implies an increase in the voter turnout in larger cities for both national and sub-national elections.

There are a number of facts that have been established about individual characteristics as they relate to voter turnout starting with Milbrath [7]. He concludes that levels of participation are highly correlated with socioeconomic status, income and education levels. He finds, as mentioned above, that urbanization increases the likelihood of political participation. Wolfinger and Rosenstone [9] have determined that in national elections in the US, the likelihood that a voter turns out is very positively correlated with income. Higher levels of educational attainment also increase the probability of voting in national elections. Political participation in terms of voting peaks at mid-life, and then declines as perhaps health issues prevent older people from turning out to vote. Older individuals are never-the-less active in terms of other forms of political participation. This supports an argument that political participation in general is positively correlated with community involvement. Retired individuals possess the means in terms of time and resources to become involved in issues of local interest. They are also less mobile, and Wolfinger and Rosenstone do find that mobile individuals are less likely to vote, especially in local elections. Marriage tends to increase the probability of voter turnout for both spouses. There is an especially low turnout within young unmarried individuals, perhaps because the sense of belonging to a community is low, while mobility is high among this demographic group. These findings have not gone without question. In a study of national election voter turnout in West Virginia, Gerald Johnson [6] argues that a state with low levels of the socioeconomic variables that positively influence voter turnout (as addressed above) has had consistently high levels of voter turnout in presidential elections.

Despite the large amount of different types of evidence amassed, there still are unanswered questions. There are few studies conducted that investigate the relevant individual characteristics in large vs. small cities or more generally in urban vs. rural areas. In the US such studies are virtually non existent perhaps because of the lack of national data on municipal elections. Similar studies within any one country are rare as well. Some cross country comparisons [8] that compare urban and rural voting turnout do exist. However, there is no evidence whether the probability to turn out to vote in terms of income, for example, is the same across different size communities and if there are any significant differences of this turnout probability depending on the scope of election. In other words, we have not answered the question whether richer people in cities are more likely to vote in local elections or in national ones? This entire family of related questions can be posed a number of different ways. Maybe richer rural dwellers vote with equal probability in all levels of elections, but urban voters are more selective about the elections they participate in depending on their income because of a urban "free rider" problem. In this case, in aggregates we might observe a positive correlation of voter turnout with income, while it is possible that this relationship is mainly driven by the larger numbers of richer voters in large cities, in the absence of a similar income relationship for rural voters. The same questions can be asked about education, age, sex, race, marital status, and mobility.

If voter turnout is to some extent driven by social capital we are similarly unaware of the differential effects the amount of a voter's social capital has on voting in national vs. local elections, and if there are any substantial differences among rural and urban residents. Is social capital more important for rural or for urban voters in their decision to turnout to vote. If yes, are there any differences in its importance in the decision to vote depending on the scope of election? At this point we do not know the answers to any of these specific individual level voting turnout questions.

This study asks what type of voters, in terms of individual characteristics, social capital (personal connections to the local community), and local capital ownership (such as home-ownership), turn out to vote in national as opposed to local elections. We find that the demographic characteristics: sex, marital status, race, number of children, and income do not influence the differential decision to vote in national only as opposed to national and local level elections. However, years of education play a central role in this voting decision. More educated voters tend to vote locally as well as nationally rather than nationally only. We conclude that civic duty and lower voting costs for educated voters drive these voters to participate nationally as well as locally.

Social capital measured through length lived in the particular community and the size of the community are imperative for this differential voting decision. The more connected a voter is to his or her local municipal unit, the more this voter is likely to vote both locally and nationally rather than to vote only at the national level. However, we also find that monetary incentives such as local property taxes and values or (Federal) Social Security disbursements have little with the differential voting decision.

Finally, as the data used in this study contains self-reported voting information, we asses the need for correcting our results to account for the over-reporting of turnout recorded in the literature. In their work, Brian Silver and his co-authors [10] find evidence that more educated voters tend to misreport their actual turnout when asked whether they voted in an election. The evidence in this study is with regard to National level elections in the US. In the present work, we are interested in differential effects in national vs lower level elections and therefore we are concerned only about differential misreporting of voting between the two election modes. While our results confirm the presence of overall misreporting at the national level in the US, we do not find a large differential misreporting effect. We are confident that the results we present are not significantly affected by a potential misreporting bias.

3.2 Theoretical Discussion

The reasons behind an individual's voting turnout decision can be framed in a simple rational cost-benefit framework. A citizen turns out to vote when the benefits from doing so outweigh the cost from voting. The benefits increase in a number of variables.

As noted by [5], participation in politics (in particular through voting) is higher when the system's capacity to respond to and accommodate the voter's preferences is greater and when the voters take on their civic duty of political involvement with higher levels of responsibility and competence. A national level government has a higher system capacity in its re-distributional and legislative power. The individual voters can more effectively control a local government and steer it toward satisfying their own individual preferences, because of the relative importance of their vote. Thus, theoretically there are two opposing effects that might steer potential voters in their choice to vote or abstain in national vs. local elections.

In the decision whether to vote as a citizen in a large vs. a small community, voters similarly face two opposing effects. The civic responsibility in a small community is higher not only because of the larger share every voter carries, but also because of a potentially stricter monitoring of participation. A higher level of community involvement may result in higher level turnout in smaller communities. However, as information is more easily transmitted in urbanized areas, larger city voters may be better informed and able to carry out their civic responsibility with a higher degree of competence, leading to a higher turnout in larger cities. These arguments would hold for national as well as local elections.

Because of these opposing effects, voter turnout in in national vs. local elections, and in small vs. large communities can in theory go either way. In addition, certain demographic groups in each of the voting situations above may carry a larger share of the civic responsibility or ability to influence the direction of the vote. This would occur when the educational levels or the economic costs to voting differ between demographically distinct groups. Furthermore, when a voter is one of many voters sharing similar preferences, a "free rider" effect may be present in any type of community or election. When a voter is less likely to be the pivotal voter he or she is more likely to abstain and avoid the costs of voting.

With the above considerations in mind, the standard model of voter turnout involves a maximum likelihood estimation of the binary dependent variable in answer to the question "Did you vote?" in the appropriate election. The right hand side matrix of independent variables, X includes demographic indicators of age, sex, race, marital status, income, mobility, and social capital variables including home-ownership and membership in various organizations, as well as a variable indicating the logarithm of the city size.

The conditional probability of voting $\Pr(Y)$ is defined as

$$\Pr(Y) = F(X'\beta)$$

where $F()$ represents the normal or the logistic ($\Lambda(X'\beta) = \frac{e^{X'\beta}}{1+X'\beta}$) cumulative distribution function. The estimates of the coefficients in the above setup do not have as straight forward an interpretation as do the coefficients from OLS regressions. In particular the coefficients in the model using the normal distribution (known as probit model) do not represent an estimate of the amount of change in the dependent variable that results from a unit increase in an independent variable. Instead, the probit coefficients provide an estimate of the amount of change on the cumulative standard normal distribution that would result from a unit change in an independent variable, all other variables held constant.

To estimate the marginal effect of a single variable on the probability of voting for a particular type, or sub-population of voters, after estimating the probit equation for each respondent, this estimate can be converted to a probability by evaluating this number on the standard normal CDF. In this study we use a maximum likelihood model based on the logistic distribution in part because of the relative ease of interpretation of the exponentiated resulting coefficients. Using a normal or a logistic distribution in the model gives similar results in most cases, except for when there is a very large proportion of "yes" or a very large proportion of "no" voting outcomes.

Yet, there is no strong theoretical reason for preferring one distributional choice over the other.

3.3 Estimation Strategy

The regression analysis used in this study employs a maximum likelihood multinomial logistic regression model for the effect of voter characteristics on the propensity to vote in local and/or national level election in the United States. The multinomial (multiple modes) logistic model assumes there is a number of outcomes, four in our case, $Y \in I = \{0, 1, 2, 3\}$. The categories of outcomes Y are labeled in the above manner without any specific ordering between them.

We then estimate the multinomial logit model, where the coefficients β_1, β_2, and β_3 correspond to each category with the probability of outcome according to the logistic distribution:

$$\Pr(Y = 0) = \frac{1}{1 + e^{X'\beta_1} + e^{X'\beta_2} + e^{X'\beta_3}},$$

$$\Pr(Y = 1) = \frac{e^{X'\beta_1}}{1 + e^{X'\beta_1} + e^{X'\beta_2} + e^{X'\beta_3}},$$

$$\Pr(Y = 2) = \frac{e^{X'\beta_2}}{1 + e^{X'\beta_1} + e^{X'\beta_2} + e^{X'\beta_3}},$$

and

$$\Pr(Y = 3) = \frac{e^{X'\beta_3}}{1 + e^{X'\beta_1} + e^{X'\beta_2} + e^{X'\beta_3}}.$$

The coefficient $\beta_0 = 0$ is restricted for identification purposes of the model and the outcome $Y = 0$ is called the base category. The choice of base category and the consequent restriction of its coefficient β_0 is certainly arbitrary. Therefore, with this choice of base category, the remaining coefficients β_1, β_2, and β_3 measure the change relative to the base outcome category with

$$\frac{\Pr(Y=i)}{\Pr(Y=0)} = e^{X'\beta_i}, \quad i = \{1,2,3\}.$$

The matrix of covariates is $X = (x_1, x_{2,}, \ldots, x_k)$ and the value of $e^{\beta_{i_j}}$ can be interpreted as the (relative to the base outcome) risk ratio for one unit of change in the covariate x_j. We test the hypothesis of equality of β_j coefficients across outcome categories in order to establish whether there is more voting in both national and local elections versus voting in national elections only for certain types of voters. [2] The differential propensity to vote only in national, and not local elections, is tested using a null hypothesis for the equality of the coefficients in columns *(1)* and *(2)* of tables 3.3 and 3.7.

The differential overreporting analysis in the NES sample is examined with a probit model. The dependent variable is a dichotomous indicator of overreporting. It is set to 1 for those respondents who were validated as non-voting but said they voted and 0 for those who were confirmed as voters and in fact said they voted. Thus, the denominator used here represents all respondents who claimed they voted and reflects the aggregate amount of misreporting in each type of election. We regress this measure of vote overreporting on various demographic, geographic variables and election year indicators.

The probit model is defined as

$$\Pr(y_j \neq 0 | x_j) = \Phi(X'\gamma)$$

[2] In our study, outcome category $Y = 0$, the base category, is denoted "NN" and corresponds to an outcome of no voting in national or local elections, to which all other outcomes are compared. These include $Y = 1$, voting in national but not local elections, $Y = 2$, voting in local but not national elections and $Y = 3$ voting at both national and local levels, and are labeled by "YN", "NY" and "YY" respectively. The types of voters are defines as demographic categories according to age, sex, race, marital status, income, education and a number of social capital variables.

were Φ is the cumulative distribution function for the standard normal distribution, y_j the overreporting indicator outcome, and x_j is the vector of covariates for observation j. We then calculate the marginal effect on the probability of overreporting voting for an infinitesimal change in a covariate x_i as

$$\frac{\partial \Phi}{\partial x_i} = \phi(\bar{X}\gamma)\gamma_i$$

for continuous covariates and as the difference in the predicted probabilities at the means for dichotomous (dummy) variables. We perform hypothesis tests for the differences in the effects of the covariates on the probability of overreporting for presidential versus non presidential year elections.

3.4 Data

This study uses individual level data from the US General Social Survey (GSS), which is an almost annual "omnibus," personal interview survey of U.S. households conducted by the National Opinion Research Center (NORC). The first survey took place in 1972. Beginning in 1973, the General Social Survey was expanded to a full survey, and National Science Foundation took over the main financial support. Surveys have been conducted annually since 1972 except for the years 1979, 1981, and 1992 (a supplement was added in 1992), and biennially beginning in 1994. A repeated cross-section sample of about 1500 U.S. households is surveyed at random for each of the above years. The content of each survey changes slightly as some items are added to or deleted from the interview schedule. Main areas covered in the data include socioeconomic status, social mobility, social control, the family, race relations, sex relations, civil liberties, and morality. Our version of the data uses the additional variable of state of residence for the respondent in the year the survey was taken.

In 1987 the GSS Survey asked questions about voter participation which we use in our study. Out of the large number of variables included in the U.S. General Social Survey, we use the answer to the survey questions: "How often does the respondent vote in local elections?" and "Did the respondent vote in the presidential election in 1980 or in 1984?" This survey data provides information on self reported voting and it may not be an accurate account of actual voter turnout. However, for the lack of a complete data set including poll data together with socioeconomic and demographic variables, the GSS one of the best sources for this type of individual level information in the US.

Some local elections are held at the same time or at least in the same years as national elections. Certain states may need to be excluded from the sample in order to decouple the effects of participation in a national election from participation in a local one. We include all respondents to the 1987 GSS Survey aged between 30 and 80 years, so that the respondents would have been of voting age in 1980. We assume that there has been at least one local election in every community between the national election in 1984 and 1987 when the data was collected. Respondents to this survey would then have at least one local election in the 3 year period between 1984 and 1987 to refer to when answering the GSS questions. This assumption is adequate as most communities in the US hold municipal level elections of one sort or another most years.

While the US has 50 states only five of them hold state elections in years when no national election is held. Mississippi, New Jersey, Virginia and Louisiana hold state gubernatorial and legislative elections in years when no congressional elections are held, and Kentucky elects a governor in the off national election years. The 1500 US households surveyed at random each year since 1972, present too small of a sample if we are to look only at the above five states. As a small sample problem is unavoidable and because of the particular way in which the local elections question in the GSS is phrased (not to explicitly include state elections), we concentrate on municipal vs. national elections here, and include residents of all states and the District of Columbia.

Because of the self reported nature of the data, a check for possible differential misreporting in the two election modes is warranted. In order to investigate the differential misreporting of turnout in national versus local level elections in the US we use data from the American National Election Studies (NES) collected by the University of Michigan Center for Political Studies. The NES data used here contains self reported voting turnout, vote validation, and demographics for the November elections in the years 1978, 1980, 1984, 1986, 1988 and 1990. There are a total of 12,371 observations in our NES data from which we use the respondents aged 30 to 80 years in order to match the demographic characteristics of the respondents in the GSS data used here [3]. The validation of the self reported vote was achieved by checking voter registration records and voting records of the respondents. Thus, the vote of respondents who said they voted in the particular election was then attempted to be validated. No validation attempts were made for the respondents who indicated they did not vote. The self reported voters were then designated as validated voters, if there was a record indicating they actually voted, or as validated non-voters, if there was a record indicating they did not vote. In the event that no records were existing or no records were found, the respondent was designated as a non-validated voter. More than 90% of all vote misreports are done by people who did not actually vote, see [10], page 614. There is little point in investigating the misreporting of those who said they did not vote, but actually did.

We use the NES sample above to examine two types of elections: the presidential year elections, characterized by a relatively high turnout, and the midterm year elections, which have a consistently lower turnout than the presidential year elections. The self reported turnout in the national versus local level elections from the

[3] The results of the overreporting analysis performed on the full NES sample do not differ significantly from the results using those respondents aged 30 to 80 years.

1987 GSS sample is similar to the self reported turnout in the presidential versus the midterm election years in the NES sample. We use the differential overreporting of voting in presidential versus midterm national elections to proxy for overreporting in national versus lower level elections.

3.5 Results

3.5.1 Examining Differential Overreporting

The summary statistics for the NES data sample used here are presented in table 3.2. The overall (self reported) turnout in presidential election years is 77.6% and that in midterm election years is 58.9%. These numbers reflect a difference similar to the self reported turnout in the national versus local level elections in the GSS sample used in this study, see table 3.1. This similarity allows us to proxy for the difference in overreporting that could exist in national versus local elections with the difference we observe in vote overreporting in presidential versus midterm year elections. The differences in prestige of the office and the issues voted on, as well as the cost-benefit decision that drives the turnout must be similar in the NES and the GSS data samples. The demographic characteristics for the two samples (for the respondents aged 30 to 80 years) are quite similar. The mean number years of education and income are almost exactly the same in the two samples and the amount of time lived in one's community corresponds well in the two samples. The race and sex of the respondents in the two samples is also quite similar. There are more married respondents in the NES sample by 7%. Yet the total number of children is much lower in the NES sample. The GSS respondents we use in this study reported an average of 2.43 children. The NES respondents used here only had .805 children on average. This discrepancy is possibly due to the interpretation of the survey question as total number of children vs. total number of children currently living in the household. The number of children does not seem to play an important role in misreporting one's vote, and more importantly in differentially misreporting one's vote in the different types of election. Thus, this difference between the two data sets is not of great concern for this study. The respondents in the GSS sample tend to live in larger cities by about a hundred thousand residents than those surveyed in the NES sample. However, as the population of the local place of residence enters in logarithms in our specification, the difference amounts to one base point. The demographic characteristics of the two samples as well as the voting patterns in the two modes of election in each case allow us to use the NES data described above to investigate any differential overreporting of votes.

Table 3.5 presents the results from the probit analysis of overreporting in the pooled NES sample from all presidential and midterm election years in our sample. The results we present here largely agree with the findings in the literature. Similarly to the work of B. Silver at. al. [10] we find that education increases the probability that people report they have voted when in fact they have not. An infinitesimal increase in a respondent's amount of schooling increases the probability they incorrectly report a vote by 0.006. Other covariates that increase the probability of such vote misreporting in the November election for presidential and midterm election years include age and being married. The effect of age on vote misreporting becomes less significant in the full sample including respondents 18 years and older. Being back actually decreases the probability of respondents stating they have voted when they have not in this sample. Results from B. Silver at. al. above indicate that race does not increase the probability of misreporting, yet the evidence they cite in the literature is mixed. Here, we find that Blacks are actually less likely to overreport voting than observationally similar Whites. The result seems to be sensitive to the exact definition of overreporting of voting as fraction of those who said they voted, as a fraction of the entire population or as a fraction of those who did not vote.

The above analysis is an important confirmation of some existing results about vote overreporting. Yet, in this work we are concerned with the possibility of differential overreporting rather than the level of overall misreporting of self-reported votes. If there is a significant difference in the amount of vote overreporting in high turnout versus lower turnout elections, the results of this study would be biased. Thus, we divide the NES sample into observations from presidential (high turnout) election years and midterm election years. We then perform a probit analysis of the probability to overreport voting and do hypothesis tests for the equality of coefficients across presidential and midterm election years. Table 3.6 presents the results. The dependent variable is again the fraction of respondents who were validated as non-voters, but said they voted divided by the total number of respondents who said they voted.

The only covariate for which the null hypothesis of the equality of the coefficients in the two modes of election rejects is the length lived in the community. We are confident to the 5% level that the probability to overreport voting in the higher turnout elections is less than the probability to overreport voting in the lower turnout elections. This finding is consistent with the possibility that voters who have lived in a community for a long time feel obligated to contribute to the community through voting in the local or lower turnout elections. There may be stigma associated with long time residency and no participation in the local community including voting in the local elections. Therefore, these long time residents may be overreporting that they vote in the local (lower turnout) elections. The long time residents of a community are less likely to overreport voting in the presidential (higher turnout) elections as there is no community specific stigma for abstaining from voting in presidential elections.

This said, the differential effect for the probability to overreport voting is by no means large. Calculating the marginal effect of length lived in the community on the probability to misreport voting we find an infinitesimal increase in the length lived in a community decreases the probability to misreport a vote in the presidential election years by 0.00016. This change is not significantly different from zero. The change in the probability of overreport voting in non-presidential years associated with length lived in the community is 0.00057. It is significant to 5%. An infinitesimal increase in the length lived in the community increases the probability of misreporting voting in a lower turnout election by 0.0006. Such a small change is below the significance of our main results and there is no need to correct our results for it.

While we do find that there is overreporting of voting correlated with education, race, and marital status, the overreporting does not occur differentially in lower turnout versus higher turnout elections. If respondents state they voted when in fact they did not, this occurs with equal probability in higher and lower turnout elections. Both local and midterm elections have similar turnout levels relative to turnout in presidential elections in the US. Thus, we can proxy for local elections with midterm elections for the purposes of examining the potential problem of vote overreporting. We expect no differential overreporting problem exists in our data for the covariates examined above except length lived in a particular community. While we do find a higher probability of those who have lived in one community longer to overreport voting in lower level elections, the effect is too small to warrant correction for it in this study. Even though some of the numbers in columns *(1)*, *(2)*, and *(3)* of the main result table will be higher because of vote overreporting, the difference is unaffected. The results of the propensity to vote in national only versus national and local elections are not biased by a possible vote misreporting problem. We present those results confident that no correction is needed.

3.5.2 Results Involving "Always" Voting

The General Social Survey data used in this study includes self reported voting behavior described here as "always" voting in local or national elections. We construct voting outcome categories based on self reported always voting and we also calculate results using those respondents who reported they either always vote or that they rarely miss voting in local elections, and present them in an appendix. National level voting is similarly defined as “always” voting if a respondent voted in both 1980 and 1984 presidential elections recorded in the data.

Table 3.1 presents the descriptive statistics of key variables in the 1987 US General Social Survey used in this study. It is possible that those participants who indicated that they did not vote in the 1980 US presidential election, or that did not vote in both the 1980 and the 1984 presidential elections but did indicate that they vote in local elections may not have been of voting age at the time of the national elections recorded in the data. Therefore, we concentrate our analysis on respondents who were 30 years or older in 1987. By restricting ourselves to these respondents, we eliminate any effects in the patterns of voting behavior due to the voters who have recently started to vote on account of their age. The mean age of respondents to the survey we include in this study is 49.8 years. We divide the respondents into younger (30 years old to 44 years old), middle aged (45 to 59 years old), and senior individuals (60 to 80 years old). We also exclude the very old respondents, as those above 80 years old are more likely to be mentally or physically incapacitated to vote. There are 43.7% of younger voters, 27.4% of middle aged voters and 28.9% of senior voters in our sample. Out of those aged 30 and over voters, 79% voted in the 1980 national election, and 75% replied they voted in the 1984 national election. A slightly higher, 75.7%, of the respondents to this survey indicated that they either always vote or that they sometimes miss voting in local elections. That number falls to 39.9% among those who indicated they always vote in local elections.

The mean amount of schooling in our sample is 12.2 years, with 31.4% of respondents holding a High School degree as the highest level of education attained, 30.4% have at most some college education, and only 9.5% hold a college degree or have more than college education. The respondent's income is reported in 12 categories, and the mean income at 9.95 falls between \$10,000 and \$15,000 per year in 1987 dollars. The mean number of children is 2.43, and 56.3% of the respondents are married. Only 5.78% have lived in their particular community for less than a year at the time of the interview, 12.4% have lived there between one and three years, 17.6% of the respondents have lived in their community 4 to 10 years, the majority, 46.2% have lived in their current community for more than 10 years, and only 18.1% indicate they lived their entire life in their particular community. Homeowners make up 68.9% of the observations used. Of our sample, 57.6% are female and 11.8% are Black. We weigh our observations appropriately to account for the oversampling of Blacks in the GSS.

Table 3.3 presents the likelihood, in logarithmic terms, of a respondent to always vote in one of three different modes relative to the baseline mode of not voting in either national or local elections. The results presented in the main body of this work include state indicator variable to account for possible differences in state specific election and voter registration laws. Differences in the results with and without the state fixed effects are discussed in the appendix. In most of the demographic categories explored here, there is a statistically significant difference between those who choose not to vote in either type of election and those who always vote in both modes. Higher income, more educated and Black respondents are more likely to always vote in both national and local elections than not at all. Older individuals are also more likely to vote in both elections than not at all, except for the very old, aged 60 to 80, in our data, who are less likely to vote in both election types than not vote at all. These results are to be expected as they are consistent with general voting behavior observed in the US. [4]

[4] In a probit model analysis using the GSS data sample used here, we find probabilities of voting in national elections that are similar and consistent with the findings presented in Wolfinger and Rosenstone [9] for the variables available in both data sets.

The larger the community size, the less likely the respondents are to always vote in any elections than not vote, which is consistent with an interpretation of a free riding of voters effect in large cities. The larger the community, the smaller the share of civic duty that falls on each member of that community. Thus, many potential voters avoid the costs and responsibilities associated with voting in larger communities where that responsibility is shared among a larger number of potential voters. Homeownership increases the likelihood of voting in all types of elections, and so does the length a respondent has lived in the same community, independently of the size of the community. Homeownership and the length lived in the community must act to strengthen the civic duty, and increase not only the community specific social capital, see [4], but build an overall social and civic awareness, which at least in terms of voting, carries beyond the particular community the respondent is associated with.

Number of children, marital status and sex seem to have less of an impact on the decision to vote in all levels of election as opposed to not voting at all. Column *(3)* in table 3.3 represents a small an somewhat unusual mode of voting behavior. The respondents in this group indicated that they always vote in local elections but do not always vote in national ones. Theories of voting behavior dictate that as the stakes are higher in national elections, in aspects ranging from the amounts of money to be allocated, to the prestige of the office for which one is voting, considering the costs associated with voting, if a voter decides to vote in a lower level election, then this voter should vote in the national level election as well. However, we do see a small number of respondents who only vote in local level elections relative to not voting at all. Since we have limited our investigation to those of voting age during the national elections in 1980 and 1984 these respondents might include newly naturalized citizens since 1984 who had recently become eligible to vote. We leave considerations about the nature of voting in local elections only and not vote in national elections aside for the purposes of the current investigation. The differences between the group of respondents who indicated that they vote in all levels of elections, and those that indicated they vote only nationally are the main focus of this study.

Table 3.3 shows more educated respondents are less likely to vote in national elections only than to vote in both levels of election relative to not voting at all. This may be due to lower costs of voting at a local level for more educated voters, prompting more educated voters to be more willing to vote at both levels of election if they choose to vote at all. More educated voters may also recognize that their local vote affects them more directly than their vote at the national level [5] In the case of the effect of education on differential propensities to vote at local or national levels, the relative risk ratio between voting in both levels and voting only nationally is 1.076, so one extra year of education makes a voter almost 8% more likely to (report) vote in both types of election than in national elections only. The relative risk ratios and the corresponding marginal effects for the voter characteristics with significant differences between the two modes of election examined here are presented in table 3.4. Higher levels of education could be associated with higher levels of civic duty at a local level, where a more educated voter plays a larger role in the community. At the same time, higher education implies lower costs of voting at the local level, in terms of finding information about the candidates and issues at stake, and in terms of interpreting the available information. Since information about the issues and candidates in a national election is more widely publicized and analyzed, it is easier for voters with less education to learn and understand their voting choices in national elections.

In smaller communities, respondents are more likely to vote locally and nationally rather than only nationally by about 3% for every approximately 2700 fewer residents. The length lived in the same community increases substantially the propensity to vote

[5] We find overreporting to be positively correlated with the voter's level of education and married status. While the overall numbers pertaining to the importance of education for voter turnout we report may be overstated by as a result, see [10] and the previous section of this study, we find no evidence that there is differential overreporting of turnout in national vs. local level elections in the US. Therefore the differential propensities to vote in the two modes of election should be unaffected by the potential problem of overreporting due to the respondent's educational level.

in both types of election rather than to vote in national elections only. This effect is quite strong, and while the nature of the definition of length lived in the community in the GSS makes it harder to interpret the magnitude of the risk ratio here, it is about 1.38 as we move from less than 1 year, to 1 – 3 years, to 4 – 10 years, to more than 10 years, to entire life spent in the particular community. This on average translates to 36% more likely to vote in national and local elections as opposed to national elections only as we move from category to category. This effect points to a possibility that the costs associated with voting locally fall drastically as a voter acquires community specific information. The benefits of voting locally are also expected to increase as one accumulates time lived in a particular community. On one hand, perhaps the longer a voter lives in a community, the longer this voter expects to stay which prompts him or her to vote locally in addition to nationally in order to secure future fiscal and other benefit allocation to oneself. On the other hand, voting locally in addition to nationally will help voters promote and keep high the value of local capital such as real estate they hold.

If the latter is the incentive which causes long time local residents to vote in their local communities in addition to voting nationally, we would expect of find homeowners voting at higher rates nationally and locally rather than nationally only. However, in our data we find only a small difference between the rates at which homeowners vote nationally only versus voting nationally and locally. Thus, the homeowners in our sample are not predisposed to vote at different rates in the two modes of voting behavior explored here. Monetary incentives, such as property taxes are not driving the decision to vote locally. Perhaps homeowners are less concerned with the particular level of of taxation as long as the taxes are spent in agreement with their preferences for local amenities. The number of children[6] the respondent has

[6]The results were also calculated using a variable indicating the respondent has any children, as opposed to number of children, which did not affect the respondent's propensity to vote differently in the two levels of election in the always voting category or in the usually voting category.

does not significantly alter the respondent's propensity to vote locally and nationally versus voting only nationally. However, there is no indication in our data whether the respondent's children live with the respondent, or whether they attend local public schools, so it is difficult to assess the impact of preference for school provision on voting behavior here.

Age increases the propensity to vote both nationally and locally as opposed to nationally only. A respondent who is one year older than another is 5% more likely to vote locally as well as nationally rather than nationally only. This result could be picking up social capital in terms of one's sense of belonging to a particular community through age. This effect is different than the length lived in a community effect discussed above since we have controlled for the length lived in the particular community. That is, as life progresses, voters may be able to more accurately asses or experience the benefits of voting for local amenities, which prompts them to vote locally as well as nationally. Age groups of younger, middle aged and senior voters, however, do not show significant differential propensities of voting, and neither do the 60 – 80 year old voters. In the seniors (60 – 80 year old) voters we see a lower probability of voting in any election mode explored rather than not voting at all. The seniors in this study are less likely to vote in national and local elections than to vote nationally only, but the difference in the voting rates is not statistically significant. Theories of voting behavior predict that seniors may be largely interested in Social Security allocations which are determined at the national level in the US, and thus they would vote nationally only. In her work, A. Campbell [2] finds that senior citizens' participation in the political process in indeed influenced by Social Security allocations. Her findings are driven largely by political participation modes such as monetary contributions, contacting politicians and volunteering for political campaigns as well as voting. However, as far as voting is concerned, we do not find a differential propensity in our data for senior citizens voting at national levels only.

Thus, we conclude that while the senior citizen's decision to vote may include considerations about Social Security, it is not influenced by Social Security disbursements only. When senior citizens decide to vote, this decision is most likely based on civic duty and lower costs of voting due to the senior citizens' high levels of free time rather than simply personal monetary gain involving Social Security allocations. Of course, the poorest of the senior voters would be most affected by Social Security policy, but we do not find that an income and age interaction, or and income and senior aged indicator variable interaction has any effect on the differential voting behavior in the GSS data employed in this study. Income itself has no effect on the differential propensity to vote nationally only versus voting nationally and locally. Marital status, sex, and race have little to do with the respondent's decision to vote at the national and local versus voting at the national only level as well.

Overall, we find that homeownership does not change the propensity to vote in national and local elections, as opposed to national elections only in both always and usual voting behavior. Therefore we conclude that this differential voting decision must be based on social capital and non-monetary incentives rather than the interest a voter has in the local real estate he or she holds in a community. This non-monetary channel through which voters decide to participate in elections is reinforced with the results we find about the senior voters decision to participate in local elections at very similar rates to the rates this age group votes in national elections only. The seniors, and more importantly the poor seniors' decision to vote is not just influenced by Social Security issues. The voters' decision to participate in the electoral process at a local level though voting is driven by social capital rather than monetary capital.

3.6 Conclusion

Voting is an important mode of civic participation in the US. Voting participation studies have established that US citizen vote in local level elections at consistently lower rates that they do in national level elections. In this study, we have addressed the question concerning what type of voters are the ones who vote in nationally only and do not vote locally as opposed to the voters who turn out to vote in both levels of election. Considering the costs and benefits from political participation through voting, we expect that if voters decide to vote, they would do so in national level elections at least, as national level elections involve the allocation of larger amounts of money than are at stake in municipal level elections and also because more prestige is associated with national elections. Some voters would then decide to vote in local level elections in addition to voting at the national level. Voters are expected in theory to vote in local elections in addition to voting in national elections when their connections with the local community are strong. On the voting cost side, voters who decide to vote at all are expected to vote at the national level at least, since information about national elections is more readily available, and some voters would also vote at the local level, when it is easier for them to acquire information about the choices in the local election.

We find that education plays a central role in the voters' decision to vote in local elections as well as to vote in national ones. More education decreases the costs of voting locally. More time spent as a resident of a particular community also decreases the cost of voting in this community through the ease of acquiring information about the choices to be voted on, and it also increases the benefits from voting locally. Thus, voters who have lived in their respective communities longer, vote locally as well as nationally at much higher rates than they vote nationally only. Age also increases the this differential voting propensity toward voting locally as well as nationally. Demographic characteristics such as income, sex, race, marital status and number of children explored in this study do not play a role in the voters' decision to vote nationally only versus voting nationally and locally.

DiPasquale and Glaeser [4] find in their work that homeowners are better citizen, as they participate in social and civic life at higher rates than renters do. Here, we refine the considerations of homeownership in issues of civic participation as we find that homeowners, while they may be better citizen in general, are not better local voting participants. Homeowners do not contribute to voter turnout locally differently than other residents of the community contribute. The decision to vote locally is not based on monetary self interest and local capital incentives such as property taxes and real estate values. Instead, it is based on the amount of social capital and non-monetary connections one has in the local community, including length lived in the community, education — a proxy for local community standing and civic duty, and to some extent age — again implying a civic and local responsibility to participate in the local political process as a member of the community. This non-fiscal channel for the incentives for local voting participation is reinforced in our results through our findings that elderly, and poorer elderly voters do not tend to vote nationally only. National level decisions about Social Security distributions do not seem to play into this group of voters' decision to vote differently in the two types of elections. Therefore, they must not be considering their own fiscal benefit in their differential voting decision. Education and length lived in the community also affect this differential voting decision as higher levels of these voter characteristics act to reduce the costs of voting through the ease of gathering and processing information about the local election in question.

Given the results of this study, we conclude that social capital and civic duty and responsibility together with informational attainment cost of voting drive the decision of voters in the US to vote nationally and locally as opposed to nationally only. The reason why voters turn out to vote does not seem to be governed by personal fiscal gains and redistributions to the voters but rather by non-monetary reasons. These reasons include the citizens' responsibility to participate in the voting process when the costs of voting are sufficiently low. If the benefits from voting play into the decision to vote, our results imply that they must work through the willingness of the voter to pay higher property taxes and allow the government to have funds, as long as the funds are spent on locally and nationally allocated amenities that agree with the particular voter's preferences over these amenities. Thus, the decision to turn out to vote would not be influenced by the government's propositions to raise funds but the voter would turn out to vote in the appropriate election in order to express his or her preferences on how the funds should be spent.

Table 3.1: Means and Percentages of Key Variables From the GSS Sample

	Mean	Percent	N
1. Age (30 to 80)	49.8	—	1321
Aged 30 to 44	—	43.7	577
Aged 45 to 59	—	27.4	362
Aged 60 to 80	—	28.9	382
2. Years of Schooling	12.2	—	1316
Less than High School	—	28.8	380
High School	—	31.4	415
(Some) College	—	30.4	401
More than College	—	9.46	125
3. Mean Income Category	9.95	—	1317
4. Number of Children	2.43	—	1317
Has any children	—	81.9	1317
5. Length Lived in Community	3.58	—	1316
Less that 1 year	—	5.78	76
1 to 3 years	—	12.4	163
4 to 10 years	—	17.6	231
More that 10, less than life	—	46.2	608
Entire life	—	18.1	238
6. Population of Community in 1000s	425.9	—	1321
7. Female	—	57.6	761
8. Married	—	56.3	744
9. Black	—	11.8	158
10. Homeowner	—	68.9	910
14. Voted in 1980 National Election	—	79.0	1005
15. Voted in 1984 National Election	—	75.0	961
16. Always Votes in Local Elections	—	39.9	522
17. Usually Votes in Local Elections	—	75.7	991

Note: Income categories: 1 - Under \$1,000, 2 - \$1,000 to 2,999, 3 - \$3,000 to 3,9 99, 4 - \$4,000 to 4,999, 5 - \$5,000 to 5,999, 6 - \$6,000 to 6,999, 7 - \$7,000 to 7,999, 8 - \$8,000 to 9,999, 9 - \$10,000 to 14,999, 10 - \$15,000 to 19,999, 11 - \$20,000 to 24,999, 12 - \$25,000 or over. Length lived in community: 1 - Less than one year, 2 - One to three years, 3 - Four to ten years, 4 - More than ten years, but not entire life, 5 - Entire life. The sample excludes respondents less than 30 years old and those more than 80 years old.

Table 3.2: Means and Percentages of Key Variables From the NES Sample for Over-reporting Analysis

	Mean	Percent	N
1. Age (30 to 80)	49.9	—	8926
Aged 30 to 44	—	44.2	3946
Aged 45 to 59	—	26.7	2386
Aged 60 to 80	—	29.1	2594
2. Years of Schooling	12.3	—	8885
Less than High School	—	25.9	2304
High School	—	34.4	3058
(Some) College	—	29.1	2586
More than College	—	10.6	937
3. Mean Income Category	9.31	—	6946
4. Number of Children	.805	—	8916
Has any children	—	39.9	3559
5. Length Lived in Community (in years)	32.3	—	8886
Less that 6 months	—	2.50	222
6 mo. to 3 years	—	6.60	587
3 to 10 years	—	22.4	1989
More that 10, less than life	—	49.4	4387
Entire life	—	17.3	1534
6. Population of Community in 1000s	290	—	8926
7. Female	—	55.9	8926
8. Married	—	63.3	8898
9. Black	—	11.8	8890
10. Homeowner	—	73.5	8873
11. Turnout in Presidential Year Elections	—	77.6	3791
12. Turnout in Midterm Year Elections	—	58.9	4623

Note: Income categories: 01. none or less than $2,999, 02. $3,000 - $4,999, 03. $5,000 - $6,999, 04. $7,000 - $8,999, 05. $9,000 - $9,999, 06. $10,000 - $10,999, 07. $11,000 - $11,999, 08. $12,000 - $12,999, 09. $13,000 - $13,999, 10. $14,000 - $14,999, 11. $15,000 - $16,999, 12. $17,000 - $19,999, 13. $20,000 - $21,999, 14. $22,000 - $24,999, 15. $25,000 - $29,999, 16. $30,000 - $34,999, 17. $35,000 - $39,999, 18. $40,000 - $44,999, 19. $45,000 - $49,999, 20. $50,000 - $59,999, 21. $60,000 - $74,999, 22. $75,000 and over. Turnout is self reported.

Table 3.3: Multinomial Logit Regression for the Propensity to (Always) Vote in National and/or Local Elections

	Cell Coefficients			*Hypothesis Test of* $YN - YY = 0$
	YY	YN	NY	Chi^2
	(1)	*(2)*	*(3)*	*(4)*
1. Female	0.177	0.465	0.284	3.34
	(0.178)	(0.178)*	(0.501)	
2. Married	-0.042	-0.147	0.493	0.30
	(0.219)	(0.215)	(0.615)	
3. Black	0.509	0.345	-0.047	0.34
	(0.310)	(0.303)	(.812)	
4. Income	0.151	0.139	0.007	0.07
	(0.042)*	(0.042)*	(0.111)	
5. Years of Schooling	0.264	0.190	0.126	6.23*
	(0.034)*	(.034)*	(0.090)	
6. Log of Community Population in 1000s	-0.093	-0.062	-0.080	0.52
	(0.049)	(0.048)	(0.128)	
7. Length Lived in Community	0.362	0.053	0.022	14.6*
	(0.089)*	(0.084)	(0.224)	
8. Homeownership	0.671	0.436	-0.386	1.11
	(0.230)*	(0.223)*	(0.593)	
9. Number of Children	0.180	0.104	-0.282	0.13
	(0.239)	(0.234)	(0.641)	
10. Age	0.065	0.016	-0.029	8.78*
	(0.019)*	(0.020)	(0.056)	
11. Age groups	0.282	0.685	0.993	1.75
	(0.354)	(0.357)	(0.990)	
12. Aged 60 to 80	-0.261	-0.300	-0.058	0.01
	(0.421)	(0.428)	(1.141)	

Note: $N = 1176$. Standard errors shown in parentheses, (.)* indicates significance in columns *(1)*, *(2)*, and *(3)* and rejection of the hypothesis in column *(4)* to the 5% level. YY - the respondent votes in national and local elections, YN - the respondent votes in national but not local elections, NY - the respondent votes in local but not national elections, NN - the respondent does not vote in either election type. NN is the comparison category. Respondents ineligible to vote in either 1980 or 1984 national election are excluded. The sample excludes respondents less than 30 years old and those more than 80 years old. State fixed effects included in the analysis.

Table 3.4: Relative Risk Ratios and Marginal Effects for Voting in Both Modes of Election versus Voting Nationally Only.

	Relative Risk Ratio of YY vs. YN	*Marginal Propensity to Vote in*	
		Both Levels	*National Only*
	(1)	*(2)*	*(3)*
1. Female	0.749	—	25%
2. Years of Schooling	1.076	7.6%	—
3. Length Lived in Community (Ranges)	1.362	36.2%	—
4. Age	1.050	5.0%	—

Note: Relative risk ratios for the propensity to vote in national and local elections as opposed to voting nationally only for a unit increase, or one category to the next increase where appropriate, in the respective covariate.

Table 3.5: Probability of Vote Overreporting for the Pooled NES Data

	Coefficient *(1)*	*Marginal Effect* *(2)*
1. Years of Schooling	0.033	0.006
	(0.011)⋆	(0.002)⋆
2. Male	-0.106	-0.019
	(0.063)	(0.011)
3. Black	-0.479	-0.110
	(0.084)⋆	(0.023)⋆
4. Married	0.336	0.065
	(0.066)⋆	(0.013)⋆
5. Log of Community Size	0.0003	0.00005)
	(0.012)	(0.00216)
6. Length Lived in Community	0.0009	0.00015)
	(0.0009)	(0.00016)
7. Income Category	-0.003	-0.00056
	(0.005)	(0.00100)
8. Homeowner	0.082	0.016
	(0.069)	(0.014)
9. Number of Children	0.023	0.004
	(0.028)	(0.005)
10. Age	0.037	0.006
	(0.016)⋆	(0.003)⋆
11. Age Squared	-0.0002	-0.000045
	(0.0002)	(0.000028)
12. Aged 60 to 80	-0.040	-0.008
	(0.086)	(0.017)

Note: $N = 3984$. Standard errors shown in parentheses, (.)⋆ indicates significance to the 5% level. Marginal effects for the change in probability to overreport voting for infinitesimal change in the continuous covariates and for change from 0 to 1 in the indicator (dummy) variables at the mean (0.887) Results for the 30 to 80 years old respondents. Year and state specific effects included, state effects not significant.

Table 3.6: Differential Probability to Overreport Voting in Presidential (High Turnout) versus Midterm (Lower Turnout) Elections

	Coefficients in Election Years		*Hypothesis Test of*
	Presidential	*Midterm*	$P - NP = 0$
	P	NP	Chi^2
	(1)	*(2)*	*(3)*
1. Years of Schooling	0.017	0.054	2.87
	(0.015)	(0.016)*	
2. Male	-0.097	-0.107	0.01
	(0.086)	(0.093)	
3. Black	-0.447	-0.510	0.14
	(0.118)*	(0.122)*	
4. Married	0.369	0.273	0.51
	(0.087)*	(0.104)*	
5. Log of Community Size	-0.003	0.001	0.03
	(0.017)	(0.016)	
6. Length Lived in Community	-0.00087	0.00306	4.53*
	(0.00124)	(0.00137)*	
7. Income Category	0.003	-0.009	1.25
	(0.008)	(0.008)	
8. Homeowner	0.019	0.178	1.29
	(0.095)	(0.103)	
9. Number of Children	0.049	0.002	0.67
	(0.040)	(0.040)	
10. Age	0.031	0.048	0.29
	(0.023)	(0.023)*	
11. Age Squared	-0.0002	-0.0003	0.24
	(0.0002)	(0.0002)	
12. Aged 60 to 80	-0.092	0.227	1.63
	(0.176)	(0.177)	

Note: $N = 3984$. Standard errors shown in parentheses, (.)* indicates significance in columns *(1)* and *(2)* and rejection of the hypothesis in column *(3)* to the 5% level.

Appendix A

Results Involving Usual Voting

We present the results analogous to the main result (Table3.3) of this study calculated using the less restrictive definition of voting participation.

Table 3.7: Multinomial Logit Regression for the Propensity to Usually Vote in National and/or Local Elections

	Cell Coefficients			*Hypothesis Test of* $YN - YY = 0$
	YY	YN	NY	Chi^2
	(1)	*(2)*	*(3)*	*(4)*
1. Female	0.256	0.307	-0.078	0.05
	(0.201)	(0.275)	(0.500)	
2. Married	-0.405	-0.075	-1.15	1.42
	(0.246)	(0.331)	(0.566)*	
3. Black	0.585	0.886	0.361	0.61
	(0.349)	(0.465)	(.753)	
4. Income	0.116	-0.024	-0.51	7.87*
	(0.044)*	(0.057)	(0.094)	
5. Years of Schooling	0.251	0.149	0.085	5.88*
	(0.038)*	(.051)*	(0.089)	
6. Log of Community Population in 1000s	-0.070	0.013	0.006	1.87
	(0.053)	(0.073)	(0.132)	
7. Length Lived in Community	0.122	-0.178	-0.092	8.41*
	(0.098)	(0.124)	(0.212)	
8. Homeownership	0.812	0.706	0.058	0.13
	(0.224)*	(0.340)*	(0.564)	
9. Number of Children	0.633	0.852	0.467	0.49
	(0.263)*	(0.373)*	(0.639)	
10. Age	0.045	0.003	-0.044	3.08
	(0.021)*	(0.029)	(0.055)	
11. Age groups	0.088	0.074	0.596	0.00
	(0.406)	(0.551)	(0.998)	
12. Aged 60 to 80	-0.118	0.165	-0.137	0.27
	(0.486)	(0.674)	(1.226)	

Note: $N = 1176$. Standard errors shown in parentheses, (.)* indicates significance in columns *(1)*, *(2)*, and *(3)* and rejection of the hypothesis in column *(4)* to the 5% level. YY - the respondent votes in national and local elections, YN - the respondent votes in national but not local elections, NY - the respondent votes in local but not national elections, NN - the respondent does not vote in either election type. NN is the comparison category. Respondents ineligible to vote in either 1980 or 1984 national election are excluded. The sample excludes respondents less than 30 years old and those more than 80 years old. State fixed effects included in the analysis.

We define "usually" voting to include those who report that they always vote in local elections, and those who reported they rarely miss voting. In the national level elections, the respondent is "usually" voting if the respondent voted in at least one of the 1980 or 1984 elections recorded here.

The results using the less strict definition of usual voting, presented in table 3.7 show a similar situation to the one presented in the main body of this article. The significance of the results is reduced because of this more inclusive definition of voting, as is to be expected. The difference in the age coefficients is significant only ot 8% here. The only notable difference with the results calculated with our definition of always voting, comes when we consider the effect of income. In table 3.7 it appears that the effect of income on the respondent's propensity to vote in the two types of elections is significantly different. As income increases a respondent is less likely to vote in national elections only. However, both relevant coefficients are very close to zero, implying very small propensity to vote in either mode as opposed to not vote at all based on income, so we do not consider income to be a very important channel for the differential voting behavior explored here.

Appendix B

Results Excluding State Fixed Effects

Results including state fixed effects shown in tables 3.3 and 3.7 are similar to the ones with no state-specific indicator variables presented here.

Table 3.8: Multinomial Logit Regression for the Propensity to (Always) Vote in National and/or Local Elections (No State Specific Effects)

	Cell Coefficients			*Hypothesis Test of* $YN - YY = 0$
	YY	YN	NY	Chi^2
	(1)	*(2)*	*(3)*	*(4)*
1. Female	0.221	0.444	0.348	2.29
	(0.167)	(0.168)⋆	(0.461)	
2. Married	0.059	-0.128	0.309	0.84
	(0.205)	(0.203)	(0.561)	
3. Black	0.638	0.371	0.489	1.13
	(0.277)⋆	(0.274)	(.688)	
4. Income	0.122	0.122	0.022	0.00
	(0.039)⋆	(0.039)⋆	(0.096)	
5. Years of Schooling	0.274	0.191	0.340	8.76⋆
	(0.032)⋆	(.033)⋆	(0.087)	
6. Log of Community Population in 1000s	-0.124	-0.039	-0.176	5.60⋆
	(0.040)⋆	(0.039)	(0.113)	
7. Length Lived in Community	0.339	0.014	0.057	19.2⋆
	(0.081)⋆	(0.077)	(0.202)	
8. Homeownership	0.632	0.548	-0.307	0.17
	(0.213)⋆	(0.206)⋆	(0.548)	
9. Number of Children	0.015	-0.050	-0.088	2.14
	(0.049)	(0.051)	(0.128)	
10. Age	0.058	0.015	-0.047	7.43⋆
	(0.018)⋆	(0.018)	(0.052)	
11. Age groups	0.394	0.691	1.389	1.08
	(0.333)	(0.337)⋆	(0.927)	
12. Aged 60 to 80	-0.346	-0.273	-0.392	0.05
	(0.401)	(0.410)	(1.052)	

Note: $N = 1176$. Standard errors shown in parentheses, (.)⋆ indicates significance in columns *(1)*, *(2)*, and *(3)* and rejection of the hypothesis in column *(4)* to the 5% level. YY - the respondent votes in national and local elections, YN - the respondent votes in national but not local elections, NY - the respondent votes in local but not national elections, NN - the respondent does not vote in either election type. NN is the comparison category. Respondents ineligible to vote in either 1980 or 1984 national election are excluded. The sample excludes respondents less than 30 years old and those more than 80 years old.

The only significantly different result appears in the differential propensity to vote in the two election modes due to community size. The differences in the propensity to vote locally and nationally as opposed to the propensity to vote in national elections only due to community size are no longer statistically significant when state fixed effects for the 39 most populous states in our sample are included in the analysis.

Table 3.9: Multinomial Logit Regression for the Propensity to Usually Vote in National and/or Local Elections (No State Specific Effects)

	Cell Coefficients			*Hypothesis Test of* $YN - YY = 0$
	YY	YN	NY	Chi^2
	(1)	*(2)*	*(3)*	*(4)*
1. Female	0.287	0.351	-0.037	0.09
	(0.190)	(0.260)	(0.457)	
2. Married	-0.280	-0.003	-0.974	1.16
	(0.230)	(0.312)	(0.546)	
3. Black	0.643	0.452	0.225	0.33
	(0.310)*	(0.405)	(.649)	
4. Income	0.090	-0.025	-0.074	6.16*
	(0.040)*	(0.053)	(0.079)	
5. Years of Schooling	0.264	0.163	0.018	6.19*
	(0.036)*	(.049)*	(0.085)	
6. Log of Community Population in 1000s	-0.093	0.038	-0.093	6.91*
	(0.043)*	(0.059)	(0.102)	
7. Length Lived in Community	0.093	0.014	-0.057	12.8*
	(0.087)	(0.077)	(0.190)	
8. Homeownership	0.889	0.849	0.062	0.02
	(0.225)*	(0.313)*	(0.521)	
9. Number of Children	0.475	0.076	0.015	0.20
	(0.055)	(0.077)	(0.129)	
10. Age	0.046	0.015	-0.052	2.42
	(0.021)*	(0.018)	(0.052)	
11. Age groups	0.118	-0.052	0.937	0.16
	(0.384)	(0.527)	(0.958)	
12. Aged 60 to 80	-0.220	0.181	-0.727	0.59
	(0.468)	(0.653)	(1.121)	

Note: Standard errors shown in parentheses, (.)* indicates significance in columns *(1)*, *(2)*, and *(3)* and rejection of the hypothesis in column *(4)* to the 5% level. YY - the respondent usually votes in national and local elections, YN - the respondent usually votes in national but not local elections, NY - the respondent usually votes in local but not national elections, NN - the respondent does not usually vote in either election type. NN is the comparison category. Respondents ineligible to vote in either 1980 or 1984 national election are excluded. The sample excludes respondents less than 30 years old and those more than 80 years old.

Bibliography

[1] Walter Dean Burnham. The changing face of the american political universe. *American Political Science Review*, 59(1):7–28, 1965.

[2] Andrea Louise Campbell. *The Third Rail of American Politics: Senior Citizen Activism and the American Welfare State*. PhD thesis, University of California, Berkeley, 2000.

[3] Angus Campbell. Surge and decline: A study of electoral change. In Angus Campbell and the University of Michigan Survey Research Center, editors, *Elections and the political order*. New York: John Wiley, 1966.

[4] Denise DiPasquale and Edward L. Glaeser. Incentives and social capital: Are homeowners better citizens? *Journal of Urban Economics*, 45:354–384, 1999.

[5] Yusaku Horiuchi. Turnout twist: Higher voter turnout in lower level elections. Forthcoming: Canadian Journal of Political Science, 2001, September 2000.

[6] Gerald W. Jonson. Research note on political correlates of voter participation: A deviant case analysis. *American Political Science Review*, 65(3):768–776, 1971.

[7] Lester W. Milbrath. *Political Participation: How and Why People Get Involved in Politics*. Chicago: Rand McNally, 1965.

[8] Robert L Morlan. Municipal vs. national elections voter turnout: Europe and the united states. *Political Science Quarterly*, 99(3):457–470, 1984.

[9] Steven J. Rosenstone and Raymond E. Wolfinger. *Who votes?* New Haven: Yale University Press, 1980.

[10] Brian D. Silver, Barbara A. Anderson, and Paul R. Abramson. Who overreports voting? *The American Political Science Review*, 80(2):613–624, June 1986.

[11] Sidney Verba and Norman Nie. *Participation in America: Political Democracy and Social Equality*. Chicago: Chicago University Press, 1972.