

Housing Demand and Community Choice:

An Empirical Analysis

Carol Rapaport*

Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045
(212) 720-5657
e-mail: carol.rapaport@frbny.sprint.com

*I thank David Blau, Tom Downes, Tom Mroz, Brian Surette, and especially Dennis Coates for discussions and suggestions, seminar participants at Princeton University and the NBER Summer Institute for comments, Kirsten Madden for research assistance, and the Industrial Relations Section at Princeton.

Housing Demand and Community Choice: An Empirical Analysis

Abstract

Housing demand reflects the household's simultaneous choice of neighborhood, whether to own or rent the dwelling, and the quantity of housing services demanded. Existing literature emphasizes the final two factors, but overlooks the choice of community. This paper develops an econometric model that incorporates all three components, and then estimates this model using a sample of Tampa, Florida households. Incorporating community choice increases the price elasticity of demand and reduces the differential between white and comparable nonwhite households. The results are robust to the inclusion of permanent income and taxes.

Carol Rapaport
Federal Reserve Bank of New York
Research Department
33 Liberty Street New York, NY 10045

I. Introduction

This paper presents new estimates of the demand for owner-occupied housing. In making housing decisions, most people decide not just how many units of housing to consume, but also where to consume them. After choosing an area in which to live, households choose a specific neighborhood based on the level of public goods and housing prices in each local jurisdiction. Because households choose the community and because the price of housing varies by community, the price and quantity of housing are chosen simultaneously. By contrast, the housing demand literature has not modeled the choice of a specific community within a local housing market as part of the household's formal optimization problem.¹ This paper demonstrates that incorporating the endogeneity of price caused by community choice substantially affects estimated housing demand.

This paper models the household's simultaneous choice of the community of residence, whether to own or rent (the tenure decision), and the quantity of housing services consumed. The household maximizes a utility function whose arguments include quantities of housing, a composite commodity, and local public goods. Housing and the composite commodity are continuous variables, but the local public goods package is a discrete choice. In particular, the household is constrained in its choice of a public goods package to those offered by the relevant set of local communities; the household's choice of community determines the household's level of public goods. The model is estimated using 1980 Census of Population and Housing data from

¹While the capitalization literature incorporates the choice of community and housing prices (Yinger [20]), most capitalization papers do not estimate housing demand. An exception is Hamilton's [8] explicit modeling of capitalization leading to measurement error in demand estimation.

five communities within the Tampa, Florida area.

Following much of the literature, the estimation scheme proceeds in two stages. First, I estimate the probability that a household chooses a particular community/tenure combination as a function of community and household characteristics. Second, I estimate the demand for owner-occupied housing conditional on the community/tenure status decision. Because this approach explicitly models the choice of community and tenure, and because housing price is community and tenure specific, the estimation of housing demand incorporates the endogeneity of housing price.²

Treating housing price as endogenous substantially increases the estimated price elasticity of demand for owner-occupied housing and reduces the differences between the housing demands of white and nonwhite households. The results are robust to incorporating permanent income and federal income taxes.

The next section of this paper develops the econometric model. Section III covers data and institutional detail. Section IV presents the estimates of the demand for housing. Section V concludes.

II. An Econometric Model of Housing Demand and Community Choice

A. The Basic Model

I assume households have already chosen to reside in a particular area. Within this housing market are a finite number of separate local jurisdictions. Each household maximizes

²Housing supply is inelastic within a community. Community choice is therefore the sole source of the endogeneity of price in this model.

utility by choosing the levels of housing and the composite good, whether to own or rent the dwelling, and the mix of local public services (and other community characteristics). I assume the flow of housing services is a continuous homogeneous variable, but the level of public services is restricted to a choice among the bundles already being provided by the local jurisdictions. Each household in the local housing market faces a choice among m community/tenure combinations that are indexed $j = 1, \dots, m$. There exist $m/2$ mutually exclusive and exhaustive jurisdictions and two possible tenure statuses, owner-occupier and renter. The budget constraint variables are the price of the composite commodity, the price of housing, the property tax rate, and household income. The model draws from Dubin and McFadden [2].

The indirect utility function for household i residing in community/tenure combination j is

$$V_{ij} = V(P_{Hj}, t_j, P_x, Y_i, X_i, Z_j, \epsilon_{ij}) \quad (1)$$

where P_{Hj} is the price of a unit of housing services in community/tenure choice j , t_j is the statutory property tax rate, P_x is the price of the composite good that is the numeraire, Y_i is income, X_i is a vector of household characteristics that are independent of the choice of community/tenure, Z_j is a vector of local public goods and other community characteristics, and ϵ_{ij} contains unobservable characteristics that vary across households and communities.

Household income and the price of the composite commodity are exogenous. The price of housing reflects the supply price of housing and any capitalization of the community mix of public services and tax levels. P_{Hj} is constant across all i 's in any given j , but differs across jurisdictions and across tenure status. No single individual can affect the (gross of federal income

tax) price of housing in any community or tenure status.³ Since part of the household's maximization problem includes the choice of jurisdiction (and tenure) within a given area, the price actually paid for housing is determined endogenously. For renters, $t_j=0$.

Some examples of the elements in X_i are the gender, race, and age of the household head and the number of children. Examples of the Z_j variables are school quality, expenditures on local public goods, and population density. The V_{ij} represents household i 's evaluation of the attributes of community/tenure combination j . For example, community j may have large parks and wide avenues. Household i 's perception of these open areas is not known by the researcher yet plays a role in i 's choice of community and tenure. The probability that household i chooses community/tenure status j is

$$A_{ij} = \text{Prob}\{V(P_{Hj}, t_j, P_x, Y_i, X_i, Z_j, i_{ij}) > V(P_{Hk}, t_k, P_x, Y_i, X_i, Z_k, i_{ik}) \text{ for } j \dots k > \} \quad (2)$$

The choices of functional form and distributional assumptions are guided by existing work and by convenience. Because most of the existing research on the demand for housing estimates relatively simple functional forms, I follow along these lines to maintain comparability. In particular, I specify an index function approximation of an arbitrary indirect utility function. Equation (1) becomes

$$V_{ij} = ((1-t_j)P_{Hj})^\alpha (BP_x)^\beta (Y_i)^\gamma (X_i)^\delta (NZ_j)^\epsilon \quad (3)$$

³The deductibility of mortgage interest implies that the net price of a unit of owner-occupied housing varies with the household's tax bracket. Federal income taxes will be incorporated later in this section.

where β , α , and N are parameters. The vectors Z_j and X_i are assumed to be separable. The ϵ_{ij} are assumed to follow the Type I extreme value distribution, and to be independent of the ϵ_{ik} . The community/tenure choice, equation (2), is therefore a mixed logit estimated on the entire sample

$$A_{ij} = \frac{\exp((1/t_j)P_{Hj} + \alpha Y_i + \beta X_i + NZ_j)}{\sum_{k=1}^m \exp((1/t_k)P_{Hk} + \alpha Y_i + \beta X_i + NZ_k)} \quad (4)$$

The commonly estimated constant elasticity demand for owner-occupied housing is

$$\ln H_{ij} = \beta \ln((1/t_j)P_{Hj}) + \alpha \ln Y_i + X_i^* + O_{ij} \quad (5)$$

where H_{ij} is the quantity of housing consumed by household i in community j , β , α , and $*$ are vectors of parameters, and O_{ij} contain household-specific unobservables. Some examples of the O_{ij} are household i 's taste for open space, for bathrooms, or for gas over electric heat. In other words, O_{ij} represents household and community/tenure specific, but unobservable, tastes for housing services. Equation (5) is estimated on a sample of owners.

The appropriate estimation technique for this demand function depends on whether P_{Hj} is endogenous, which in turn depends on whether ϵ_{ij} is correlated with O_{ij} . Virtually all existing studies of the demand for housing assume implicitly that the price of housing (conditional on tenure) is uncorrelated with the unobservable household characteristics in the demand error. These studies assume $E(O_{ij} | \epsilon_{ij}) = 0$; a particularly high realization on ϵ_{ij} increases the probability that household i chooses community/tenure j , but is uncorrelated with the unobservables in household i 's demand function. If this assumption holds, then O_{ij} is uncorrelated with all

regressors conditional on the community/tenure choice made, and equation (5) can be estimated consistently with ordinary least squares. Although household i chooses housing price jointly with community, price is statistically exogenous.

A fundamental insight of the work on discrete and continuous choice, however, is that Q_{ij} and ϵ_{ij} may well be correlated. In the housing context, one plausible example of a nonzero expectation on Q_{ij} conditional on ϵ_{ij} concerns a household with a higher than average taste for open spaces. This household would probably be associated with a taste for larger rooms (more housing) than observationally equivalent households, and so would be associated with a high value of Q_{ij} . At the same time, the household would place a premium on a community/tenure, j , with large undeveloped areas, and so would be associated with a high value of ϵ_{ij} . The high ϵ_{ij} implies that the household is more likely to select community/tenure j (instead of some community/tenure k), and be observed paying housing price P_{Hj} (instead of P_{Hk}). If open land is generally considered a desirable trait, P_{Hj} will be higher than P_{Hk} as long as the value of this desirable trait is at least partially capitalized into housing prices. The household's observed price of housing will be positively correlated with the household's error in the demand equations. Under these assumptions, the OLS estimate of the price elasticity of demand for housing, ϵ , will be biased toward zero. The true relationship between price and housing is negative, but the estimated ϵ also picks up the positive correlation between price and the desirability of open space.⁴

This problem of discrete and continuous choice can be viewed as a problem of

⁴OLS estimates of ϵ could also be biased away from zero. The general point is that when $E(Q_{ij} | \epsilon_{ij}) \neq 0$, OLS cannot be used to produce consistent estimates of the behavioral parameters.

endogeneity introduced by self-selection. If the conditional expectation of the housing demand error is nonzero, price is statistically endogenous.

I use a two-step procedure to obtain consistent estimates for equation (5). This methodology was first used by Hay [10], and is discussed in Dubin and McFadden [2] and Schmertmann [19]; proofs that this procedure is consistent are given in these references. Selection terms representing the conditional mean of O_{ij} are calculated from the fitted probabilities of each of the choices. These fitted probabilities are themselves calculated from estimates of the logit choice equation (4). The selection terms are then included in the OLS specifications of housing demand equation (5). The estimating equation is

$$\ln H_{ij} = \beta_0 + \beta_1 \ln P_{Hj} + \beta_2 \ln Y_i + \beta_3 X_i + \sum_{k \neq j} \beta_k \left(\frac{\hat{A}_{ik} \ln \hat{A}_{ik}}{1 + \hat{A}_{ik}} \ln \hat{A}_{ij} \right) \quad (6)$$

where the β_k are coefficients to be estimated in the second stage and the following terms in parentheses are the selection variables representing the constructed conditional means. Each selection variable represents a pairwise comparison of the utility level associated with the choice taken, choice j , and one additional choice not taken, choice k . The coefficient on this selection variable, β_k , is a function of the covariance between O_{ij} and $(\ln \hat{A}_{ik} - \ln \hat{A}_{ij})$.

For completeness, I also estimate the standard self-selection model in which the owner's demand equation is corrected for selection into the owner-occupied state. The conventional two-step selection model involves estimating the probability of owning, constructing a selection term from the probit regressors and coefficients, and including this selection term in an OLS regression

restricted to a sample of owners. This Heckman [11] approach was first applied to the housing demand context by Lee and Trost [14] and Rosen [17], and has subsequently been used by Gillingham and Hagemann [6], Rosenthal, Duca, and Gabriel [18] and many others. Equation (3) is replaced by

$$I_i^* = \alpha Y_i + \beta X_i + \epsilon_i \quad (7)$$

where household i is an owner-occupant if $I_i^* > 0$ and a renter otherwise. This probit is estimated on the entire sample, and the estimates used to construct selection terms. Equation 6 is replaced by

$$\ln H_i = \alpha \ln((1+t_j)P_{Hij}) + \beta \ln Y_i + \gamma X_i + \delta \frac{f(I_i^*)}{F(I_i^*)} + \epsilon_i \quad (8)$$

where I_i is calculated from the deterministic part of equation (7), and $f(\cdot)$ and $F(\cdot)$ are the normal density and cumulative distribution functions, respectively. Equation (8) is estimated on the sample of owners. Finally, the paper presents summaries of results from a number of variations on all these specifications.

B. Estimating the Price and Quantity of Housing

The above discussion treats P_{Hj} and H_i as known values. In fact, the price of a standardized unit of the flow of housing services and the number of standardized units are unavailable in most data sets, and must be estimated. Since the rent paid (for rental units), the market value (for owner-occupied units), and the characteristics of the dwelling are available, I use a hedonic approach to calculate P_{Hj} . The implicit prices associated with the characteristics of

housing units in each community are estimated, and these estimates are used to form a housing price index. Separate hedonic relationships are estimated for owners and renters.

The hedonic equations from which the housing price index is formed are

$$value_{ij} = \beta_j W_{ij} + \alpha_j + \mu_{ij} \quad j = 1, \dots, \frac{m}{2} \quad \text{and} \quad (9)$$

$$rent_{ij} = \beta_j W_{ij} + \alpha_j + \mu_{ij} \quad j = \frac{m}{2} + 1, \dots, m \quad (10)$$

where $value_{ij}$ is the self-reported market value of i 's dwelling unit in community j , $rent_{ij}$ is the monthly contract rent household i pays for the dwelling unit in community j , W_{ij} is a vector of characteristics of the dwelling unit, μ are omitted characteristics and measurement error in the dependent variables, β_j are parameter vectors, and α_j is a community-specific intercept parameter. An example of W_{ij} is the presence of central air conditioning in the unit occupied by household i in community/tenure status j . The implicit price associated with this air conditioner varies by community and tenure. The community specific intercepts, the α , capture the effects of the different community-specific variables whose values may be capitalized into the price of housing services. The errors are assumed to be distributed with mean zero and constant variance.

The housing price index is calculated as

$$P_{Hj} = \frac{\hat{\beta}_j \bar{W} + \alpha_j}{\hat{\beta}_1 \bar{W} + \alpha_1} \quad j = 1, \dots, \frac{m}{2} \quad \text{and} \quad (11)$$

$$P_{Hj} = \frac{(\hat{\beta}_j \bar{W} + R_j)(12)}{\hat{\beta}_1 \bar{W} + R_1} \quad j = 1, \dots, m \quad (12)$$

where \bar{W} are the average values of each characteristic in the full sample. In the calculation for renters, equation (12), the fitted mean monthly rent is annualized (multiplied by 12) and discounted by β_1 to produce a stock value comparable to the market value for owners. Following convention, I assume the discount rate is 6%. The housing price index measures community variations in the relative price of a standard unit of housing.

The quantity of standardized units of housing, H_{ij} , is calculated for owners by dividing the market value of household i 's house by the price index for owners in community j ,

$$H_{ij} = \frac{\text{market value}_{ij}}{P_{Hj}} \quad (13)$$

The quantity of standardized units of housing for renters would be calculated analogously, but is not needed for this research.

C. Incorporating Permanent Income and Federal Income Taxes

The above model, termed the baseline model, provides consistent estimates of the demand for housing under endogenous community choice. Previous work suggests that the incorporation of permanent income and the federal tax code may influence significantly the estimated price and income elasticities. The extended model incorporates permanent income and taxes into the estimation. The extended model provides an assessment of the generality of the Florida results. If the adjustments for permanent income and income taxes influence estimated housing demand in

the Florida sample in ways similar to the existing national results, I can have more confidence in the generality of my approach.

I model permanent income as the fitted value from a regression of observed income on all individual and household characteristics in the demand for housing models plus three dummy variables indicating whether the household head's current (or most recent) job is in a white-collar occupation, a blue collar occupation, or a service occupation.^{5,6} Simply put, the maintained hypothesis is that the broad occupational category affects permanent income but not tastes for housing.⁷ Fitted permanent income is then included as the income variable in the community choice logit equation (4), the probit equation (7), and the subsequent housing demand equations (5), (6), and (8); if permanent income influences the demand for housing, then permanent income should also influence the tenure and community decisions.

The modern treatment of the federal income tax code originated with the work of Rosen [17] and King [13] and has been subsequently used by Gillingham and Hagemann [6], Henderson and Ioannides [12], Rosenthal, Duca, and Gabriel [18], and others. The net price of housing is:

⁵Harmon's [9] empirical reconciliation of the various permanent income elasticities of demand concludes that the estimates are reasonably invariant to the precise definition of permanent income.

⁶The omitted category comprises heads who have not been employed for at least five years. The Census has no occupational information for these workers. Current and permanent income are measured in logs.

⁷This assumption would be violated if households were credit constrained in the mortgage market, and lenders used occupation as a measure of credit worthiness. Testing this hypothesis is impossible with Census data. In research on the effects of credit constraints on housing demand, Rosenthal, Duca, and Gabriel [18] find that the coefficients on demographic variables are stable across households who face different credit constraints (as measured by whether they obtained a Veteran's Administration-insured loan). In any case, occupation seems a more satisfactory way to identify permanent income than more arbitrary functional form restrictions.

$$P_{NHij} = P_{Hj} \left(1 - tax_{ij} \left[\frac{m_{ij} \% t_j}{m_{ij} \% t_j \% d_{ij}} \right] \right) \quad (14)$$

where tax_{ij} is the marginal income tax rate of household i , m_{ij} is the nominal mortgage interest rate, t_j is the property tax rate, and d_{ij} is the depreciation and maintenance rate. Housing therefore receives an implicit subsidy of $\$(1 - tax_{ij} * share_{ij})$ per dollar, where $share_{ij}$ is the term in square brackets in equation (14). This term represents the share of housing expenditures that can be deducted from gross income. This specification duplicates Rosen [17], and is discussed extensively in King [13]. The required assumptions are that the opportunity cost of capital and the marginal tax rate are the same for landlords and owner-occupants. One benefit of these assumptions is that the net price of housing is independent of expected inflation; see King [13] for details.

I replace P_{Hj} in equations (4), (5), (6), and (8) with P_{NHij} from equation (12) to adjust the discrete choice and housing demand estimating equations for the provisions of the tax code. The standard tax assumptions are used. In particular, the marginal tax rate is calculated from the 1980 tax tables assuming married couples filed jointly, took an exemption for every family member, and had nonmortgage deductions equal to the average level for the income class. The nominal mortgage interest rate is set to 10%, and the depreciation and maintenance rate is set to 3.5%. The log of permanent income is calculated as fitted values from a regression of the log of net income on the basic regressor set plus the three occupational dummy variables.⁸ The results are

⁸Net income is defined as household income less the (nonhousing) tax bill. I exclude estimates of mortgage and property tax deductions from this tax bill, because these deductions are endogenous in a model of housing demand. Taxable income is therefore household income less exemptions less all other (nonhousing) deductions, and the tax bill is a function of taxable income and the marginal tax rate. This is a first dollar concept of the marginal tax rate.

insensitive to minor changes in these assumptions.

III. Data and Institutional Detail

The estimation requires data on household socioeconomic variables, dwelling characteristics, and community characteristics for all jurisdictions in the local housing market. Although cities, townships, and other incorporated areas are valid jurisdictional classifications, the paper defines community as synonymous with school district because of a prior belief in the importance of local public education to the locational choice process. Although school district is not available on any of the major micro-level data sets, the 1980 Census of Population and Housing reports the household's county of residence for those households that reside in counties with at least 100,000 residents. In 1980, school districts were approximately coterminous with counties in six states: Florida, Louisiana, Maryland, Nevada, West Virginia, and Wyoming. I defined the major local housing markets in these states, and searched for a market where all of the component counties had at least 100,000 residents and so were separately identified in the Census data. The only local housing market fully meeting these criteria is Tampa, Florida.⁹

This research uses the Public Use Microdata A (5%) sample of the 1980 Census of Population and Housing for residents of the five-county Tampa, Florida housing market. The sample used for housing demand estimation excludes elderly households. Details on the construction of the data base are provided in Appendix A.

⁹It is of course impossible to prove that these counties define one local housing market. Although Hillsborough, Pinellas, and Pasco counties comprise the Tampa SMSA, public relations literature about the area generally includes Manatee county. The National Football League franchise in Tampa considers Sarasota county part of this market. In addition, the five counties are well-linked through the interstate highway system.

Table 1 identifies the five counties in the market and provides summary statistics on the characteristics of each county and its residents. These data highlight that there are real differences in public services and property taxes between the counties. Per capita expenditures range from a low of \$448 in Pasco to a high of double that amount in Manatee. Wide differences in population density exist across the market, but the distribution of school quality, as measured by student achievement test scores, is somewhat tighter. The housing price data displayed in Table 1 were calculated from equations (11) and (12), which in turn were calculated from estimates of the hedonic equations (9) and (10). The price of a standardized unit of housing services was normalized to 1 for owners in Hillsborough, and ranges from 1.33 for owners in Sarasota to 0.91 for owners in Pasco. Further details on this calculation, including a brief discussion of the hedonic results, are provided in Appendix B.

IV. Estimates of the Demand for Housing

This section evaluates two sets of estimates of the demand for housing from each of three estimation schemes. The specifications that do not incorporate permanent income and income taxes are termed baseline specifications, while the specifications that do incorporate permanent income and income taxes are termed extended specifications. For the baseline and extended specifications, I compare the OLS estimates with estimates from the standard (Heckman) correction for the endogeneity of the tenure decision, and with estimates from the conditional expectation correction for the endogeneity of both tenure and community choice. I then evaluate the sensitivity of these estimates to various factors.

The conditional expectation correction requires first-stage input from the choice-of-

community/tenure mixed logit, equation (4). The full set of logit coefficients are available on request from the author. Table 2 presents estimates of the four conditional variables from both the baseline and extended specifications.¹⁰ The signs and significances of the coefficients differ across the two specifications. In the extended specification only, increases in the price of housing in a particular tenure/community combination significantly reduce the probability that a household will make that choice. In each specification, one of the two public goods variables has the expected sign. Increases in a community's government expenditures in the baseline specification, and increases in a community's school quality in the extended specification each increase the probability that a household will choose the community. I have no strong priors about the sign of population density, and indeed, the coefficient estimate switches from significantly positive to insignificantly negative between the baseline and extended specifications. Coefficients on the 16 multinomial variables are not reported; the variables are identical to those in the housing demand specifications presented in Table 3 below, and many of the 144 coefficient estimates are significantly different from zero. For example, observed choice patterns differ significantly between white and nonwhite households.

The first three columns of Table 3 report coefficients from the three baseline housing demand specifications: conventional OLS estimates, corrected-for-tenure estimates, and corrected-for-community-and-tenure estimates. The second three columns report the extended versions of the same specifications. The baseline OLS results presented in Column 1 are quite

¹⁰Standard errors are conditional on the estimated values of price and permanent income throughout the paper. Standard errors in the demand functions are also conditional on the estimated values of the selection terms. None of the inferences about statistical significance changes when the standard errors in the demand functions are corrected for heteroskedasticity.

conventional. The price elasticity of demand for owner-occupied housing is -0.12 , while the income elasticity of demand is 0.19 . Both elasticities are significant. The most dramatic result of the OLS estimation is the racial difference; the demand for housing is about 50% greater for white households than nonwhite households. Housing demand is positively related to the educational levels of both the head and a spouse. There exists a separate degree effect when the head (but not the spouse) graduates from college, perhaps reflecting a greater ability of college graduates to obtain mortgages. Households in which the head is an immigrant have a greater demand for housing, while households in which the head is disabled have a slightly smaller demand. The demand for housing has an inverted-U shape in the age of the head and the number of children. The maxima are at 54 years and 2.2 children, respectively. Controlling for the number of children, two-parent households demand about 11% more housing than others.¹¹

As indicated in Column 2, when the housing estimation incorporates the self-selection into the owner-occupant state, the selection term is significantly positive. This positive coefficient implies that unobservable factors that increase the probability of ownership also increase housing demand. Despite the significance of the selection term, most of the coefficients of interest are quite similar to their OLS values. In particular, the price elasticity of demand is virtually unchanged, while the income elasticity of demand is somewhat higher. The racial differential is also a little larger.

There are some differences between these baseline results and the extended OLS estimates. The extended version of Column 1 is Column 4, and the extended version of the

¹¹This comparison uses variables representing male head, spouse present, and the various spousal educational categories.

tenure-corrected Column 2 is Column 5. The most notable change is that housing demand in the extended specifications becomes more price-elastic. In addition, the precision of the price elasticity estimate increases because of the increased variation in price. The coefficient on the selection term in the tenure-corrected specification falls substantially. Apparently the differences between the measurement of permanent and transitory income, and between the measurement of net and gross housing price, comprise a large component of the unobservable variables in both the baseline tenure index function and the baseline housing demand equation. Despite these differences, there are many broad similarities between the baseline and extended specifications. For example, the income elasticity is stable, and the large racial difference remains.

The most notable changes result not from incorporating permanent income or net price, but rather from allowing for the endogeneity of community/tenure choice. Column 3 presents the baseline results, and column 6 the extended results. These specifications differ from the more conventional tenure-corrected estimates in two ways. First, treating community/tenure as endogenous increases (the absolute value of) the estimated price elasticities of demand for housing from -0.13 to -0.83 in the baseline specification, and from -0.79 to -1.63 in the extended specification. The coefficient is estimated imprecisely in the baseline specification, but is strongly significant in the extended specification. Second, the coefficient on the white dummy variable falls from 0.49 to 0.34 in the baseline specification, and from 0.41 to 0.33 in the extended specification. In short, accounting for the endogeneity of the community/tenure combination increases the estimated price elasticity of demand for housing, and reduces the racial differential between white and nonwhite households. The magnitude of these differentials suggests that the endogeneity of community choice has as great an impact on the demand for housing as the

literature's more traditional concerns of tenure, permanent income, and taxes.

Tests of the sensitivity of these results to specification changes focus on the estimates of the price elasticity of demand and the racial differential. These results are insensitive to changes in functional form, equality restrictions, and distributional assumptions. The basic results are stable across various transformations, including logs of the conditional variables in the choice logits, logs of the observed dependent variables in the hedonic equations, translog housing demands (evaluated at median prices and incomes), and variations in the functional form used to estimate permanent income. When the demand coefficients vary across communities, treating community choice as simultaneously determined with housing demand and tenure lowers the estimated racial differential in every county.¹² Results are not sensitive to changing the tenure-only error in equation (7) from normal to extreme value.

The results are somewhat more sensitive to the assumed discount rate. As indicated earlier, the above estimates assume that the flow of housing services is discounted 6% over an infinite number of periods. Increasing the discount rate to 12% actually strengthens the basic results, in that the baseline racial differential falls from 0.49 in the tenure-only estimation to 0.19 in the tenure/community estimation. The estimated price elasticity is stable.

Given that correcting for the endogeneity of community/tenure consistently reduces estimated racial differentials and increases the estimated price elasticities, it is natural to wonder about the generality of these results. Although it is never possible to draw national inferences

¹²Each housing demand equation includes nine selection terms representing the utility comparison between the observed choice and the nine alternatives. The price elasticity of demand for housing is not identified in the baseline specifications, and would only be identified by variations in the marginal tax rates in the extended specifications.

from a Tampa sample, the similarities between some specifications here and national studies suggest that the current results may indeed be quite general. Comparing the baseline OLS housing demands to existing work is surprisingly difficult because the literature often fails to present OLS results. One strong commonality, however, is that selection bias appears to be significant but quantitatively unimportant. The two original applications of the selectivity problem to housing demand, Lee and Trost [14] and Rosen [17], find exactly this result in PSID data.¹³ In a methodological review of housing demand, Olsen [16] concludes that the tenure decision is not of primary importance.

The importance of community choice in these results is broadly consistent with the few papers incorporating capitalization or other price issues into static housing demand estimation. For example, Hamilton [8] models observed house value as expenditures on housing services plus a local-government fiscal surplus. Hamilton [8] finds that the estimated income elasticity is sensitive to the incorporation of the fiscal surplus in a Philadelphia-area sample of owners. Goodman [7] uses national American Housing Survey data to demonstrate the importance of the ratio of house value to rental value as a measure of the investment potential of a property.

The magnitude of the current income elasticities are on the low end of the literature, but are not implausible. For example, my results are almost identical to Goodman [7] and Rosenthal, Duca, and Gabriel [18]. As Harmon [9] points out, the existing range of 0.14 to 1.5 makes firm conclusions difficult. My extended tenure-corrected price elasticity estimates seem to be in line with literature, and quite close to those of Rosen [17] and Rosenthal, Duca, and Gabriel [18].

¹³On the other hand, Gillingham and Hagemann's [6] analysis of CES data finds that selection bias is important for some household types.

Nevertheless, exact comparisons are almost impossible because there has been little attention given to the measurement of housing prices (see Olsen [16]). Even the research that adjusts for the federal tax code does not report estimates with and without the adjustment.

Finally, my baseline and extended results on the racial difference are somewhat larger than those typically found in the literature. For example, Goodman [7] and Rosenthal, Duca, and Gabriel [18] both find that nonwhite households demand about 20% less housing than white households. Nevertheless, 20% is itself quite a large number. Henderson and Ioannides [12] jointly estimate the demand for housing, the price of housing, and the probability of being rationed in the mortgage market. They find large (but insignificant) racial differences in the price equation, although the dissimilarity between their work and mine precludes a direct comparison. In short, the broad comparability of my results to those from national samples suggest that the endogeneity of community choice may also be important outside of Tampa.

V. Conclusion

The fundamental results of this research are quite straightforward. Incorporating the endogeneity of community choice generally increases the estimated price elasticity of demand for owner-occupied housing and decreases the differences between white and nonwhite households. Correcting for community choice therefore appears to be as important as correcting for tenure, permanent income, and the federal income tax code. While case studies may suffer from not using general data, it is reassuring that the estimated elasticities are similar to those in studies based on national data.

One plausible explanation for the price elasticity results is that an unobserved factor such

as a taste for open spaces can induce a nonzero expectation in the error in the demand function conditional on choice of community. One plausible explanation for the drop in the observed racial results is the following. Suppose nonwhites demand less housing than whites. Also suppose nonwhites prefer to and/or are constrained to live near other nonwhites. This neighborhood effect is proxied by county in the current analysis. The nonwhite dummy variable in the OLS specification captures both the demand and neighborhood effects and is therefore biased away from zero. The correction for community/tenure purges the demand estimates of the neighborhood effect. In fact, if it were possible to identify smaller geographic areas, the estimated community/tenure-corrected racial differential might be even smaller. The remaining racial differentials are likely to result from differences in tastes and constraints (see Yinger [21], Gabriel and Rosenthal ([4],[5]), Duca and Rosenthal [3], and Munnell, Browne, McEneaney, and Tootell [15]).

Appendix A: Notes on the Data Set and Variable Definitions

The basic data source for this research is the Public Use Microdata A (5%) sample of the 1980 Census of Population and Housing. The initial data set contained a 50% random sample of households. Households which meet any of the following conditions were then excluded:

- did not reside in the five-county Tampa market;
- lived in group quarters;
- lived on a rural farm;
- lived in a mobile home, trailer, boat, tent, or van;
- lived in rental unit but paid no cash rent;
- head of household employed by military;
- spouse employed by military;
- head of household under age 16;
- spouse under age 16;
- total household income reported nonpositive; and
- reported neither rent nor value of home.

The data set used to estimate the housing hedonics contained 15,643 cases. Then households whose head is over age 64 were excluded to produce the final data set of 10,484 nonelderly owners and renters. All monetary ranges were set to their midpoint values. For married couples, the head of household is redefined to be the husband.

Community characteristic variables are defined as follows:

- population density: measured in thousands per square mile;
- property tax rate (t_j): the statutory property tax rate;
- school quality: the quality of public schools, measured by the percent of students receiving 76-100% of the Florida minimum student performance standards in mathematics and writing in grades 3, 5, 8, and 11;
- expenditure: per capita expenditures on all local public goods for all governmental authorities at the county and subcounty level;
- housing prices: as defined in the text.

One could argue that an effective tax rate variable would be preferable to the statutory property tax rate. This variable would reflect the assessed value of the property (and a \$25,000

homestead exemption). Unfortunately, individual assessed value data are unavailable. Because the focus of this work is the demand for housing, however, minor errors in the tax price of the public service package should be unimportant. In addition, sensitivity analyses demonstrate that all results are insensitive to minor changes in the property tax rate.

The form of the school quality measure is likewise constrained by data availability. On the other hand, since this measure is the one Florida chooses to report, it is probably the most-watched measure, and therefore the most appropriate measure for the analysis.

Individual characteristic variables are as follows:

- property value: respondent's estimate of the current dollar worth of the owner-occupied house and the land on which it stands;
- income: household income in 1979;
- spouse: dummy=1 if spouse present;
- male head: dummy=1 if male head of household;
- white: dummy=1 if head of household is white;
- immigrant: dummy=1 if head of household is an immigrant;
- disability: dummy=1 if head of household reports any of a limited work disability, a total work disability, or a public transportation disability;
- head education: number of years of education completed by head of household;
- spouse education: if spouse present, number of years of education completed by spouse; otherwise 0;
- head high school graduate: dummy=1;
- head college graduate: dummy=1;
- spouse high school graduate: dummy=1;
- spouse college graduate: dummy=1;
- age: age of head of household in years; and
- kids: number of children living at home.

Appendix B: Estimating the Unit Price of Housing

This appendix discusses estimates from equation (9) through equation (13) used to estimate the price of a unit of the standardized flow of housing services. The underlying specifications, equations (9) and (10), are 10 OLS regressions of either monthly contract rent (for rental units) or market value (for owner-occupied units) on the characteristics of the dwelling. These coefficient estimates are combined with the full sample means of the characteristics to create the unit prices defined in equations (11) and (12) and listed in Table 1.

The independent variables are common across the 10 community/tenure specific equations. The 63 independent variables are entirely in dummy variable format, and include 7 dummy variables on the type of building (detached single family, attached single family, two units, three or four units, etc.), 6 dummies on the year constructed, 3 dummies on the number of stories, 8 dummies on the number of rooms, 5 dummies on the number of bedrooms, 3 dummies on the number of bathrooms, whether the unit is a condominium, and many additional dummy variables covering the type heating and air conditioning, the use of various types of fuel, and which utilities are included in the rent. The specification also include 4 dummy variables on the number of years the household has occupied the dwelling; these dummies capture the discounts commonly given to long-time tenants.

The coefficients from these 10 specifications are not reported, but the results contain few surprises. In general, increases in the number of rooms, bedrooms, or bathrooms are associated with higher rents and values. The relationships are typically nonlinear. Central air conditioning commands a large premium. In many of the communities, little relationship exists between rent or value and the type of fuel used for heating the unit, for heating water, or for cooking.

Although a fair number of the coefficients are of the "wrong" sign, many are not significantly different from zero. In fact, given the large number of regressors, it is not surprising that many of the coefficients are estimated with little precision. The regressors explain roughly 37 to 46 percent of the variation in value, and 47 to 69 percent of the variation in rent.

Estimation of more restrictive versions of these equations change neither the fundamental relationships revealed in the data nor the relationships between the ten fitted values of unit price. When the dummy variables indicating the number of various types of rooms and units in the structure are transformed into continuous variables, the coefficients are more often significant and of the expected sign. Pooling the data across counties for each tenure status produces similar results, although tests for the equality of slope coefficients indicate that pooling is rejected in the data. In other words, differences in the price of housing services are caused by differences in the implicit price of housing characteristics (as reflected in the slope coefficients) as well as differences in the level of community services (as reflected in the intercepts).

My confidence in the fitted implicit prices is further strengthened upon comparison of the results with those from other data sources. The data set most frequently used for determining the implicit price of housing characteristics in the American Housing Survey (AHS). The AHS contains data on many additional characteristics including the presence of a cellar, a garage, open holes or cracks in the walls, and rodents. Nevertheless, the above specification fits the data as well as many from the AHS. For example, Borsch-Supan [1] reports R-squared values between 0.50 and 0.62 for rental hedonics by city.

References

1. A. Borsch-Supan, "Econometric Analysis of Discrete Choice," Springer-Verlag, Berlin (1987).
2. J.A. Dubin and D.L. McFadden, "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica*, 52, 345-360 (1984).
3. J.V. Duca and S.S. Rosenthal, "Borrowing constraints and access to owner-occupied housing," *Regional Science and Urban Economics*, 24, 301-322 (1994).
4. S.A. Gabriel and S.S. Rosenthal, "Household location and race: estimates of a multinomial logit model," *Review of Economics and Statistics*, LXXI, 240-249 (1989).
5. S.A. Gabriel and S.S. Rosenthal, "Credit rationing, race, and the mortgage market," *Journal of Urban Economics*, 29, 71-379 (1991).
6. R. Gillingham and R. Hagemann, "Cross-sectional estimation of a simultaneous model of tenure choice and housing services demand," *Journal of Urban Economics*, 14, 16-39 (1983).
7. A.C. Goodman, "An econometric model of housing price, permanent income, tenure choice, and housing demand," *Journal of Urban Economics*, 23, 327-353 (1988).
8. B.W. Hamilton, "Capitalization and housing-demand bias," *Southern Economic Journal*, 49, 4-388 (1982).
9. O.R. Harmon, "The income elasticity of demand for single-family owner-occupied housing: an empirical reconciliation," *Journal of Urban Economics*, 24, 173-185 (1988).
10. J. Hay, "An Analysis of Occupational Choice and Income," Ph.D. dissertation, Yale University, New Haven (1979).
11. J. Heckman, "Sample selection bias as a specification error," *Econometrica*, 47, 153-161 (1979).
12. J.V. Henderson and Y.M. Ioannides, "Tenure choice and the demand for housing," *Economica*, 53, 231-246 (1986).
13. M.A. King, "An econometric model of tenure choice and demand for housing as a joint decision," *Journal of Public Economics*, 14, 137-159 (1980).

14. L.-F. Lee and R.P. Trost. 1978. "Estimation of some limited dependent variable models with application to housing demand", *Journal of Econometrics*, 8, 357-382 (1978).
15. A.H. Munnell, M.B. Tootell, L.E. Browne, and J. McEneaney, "Mortgage lending in Boston: interpreting HMDA data," *American Economic Review*, 86, 25-53 (1996).
16. E.O. Olsen, "The demand and supply of housing service: a critical survey of the empirical literature," in *Handbook of Regional and Urban Economics*, Volume II, (E.S. Mills, Ed.), Elsevier Science Publishers, The Netherlands (1987).
17. H.S. Rosen, "Housing decisions and the U.S. income tax," *Journal of Public Economics*, 11, 1-23 (1979).
18. S.S. Rosenthal, J.V. Duca, and S.A. Gabriel, "Credit rationing and the demand for owner-occupied housing," *Journal of Urban Economics*, 30, 48-63 (1991).
19. C.P. Schmertmann, "Selectivity bias correction methods in polychotomous sample selection models," *Journal of Econometrics*, 60, 101-132 (1994).
20. J. Yinger, "Capitalization and the theory of local public finance," *Journal of Political Economy*, 90, 917-942 (1982).
21. J. Yinger, "Measuring racial discrimination with fair housing audits: caught in the act," *American Economic Review*, 76, 881-893 (1986).

Table 1
The Tampa Housing Market: Descriptive Statistics

Community Characteristics

	Hillsborough	Manatee	Pasco	Pinellas	Sarasota
price to own	1.000	1.150	0.914	1.104	1.333
price to rent	1.061	1.121	1.064	1.120	1.242
property tax	0.0062	0.0064	0.0060	0.0061	0.0063
expenditures	824	886	448	574	649
pop density	614.4	198.7	262.4	2601.9	353.0
school quality	68.6	80.3	72.9	75.5	78.3

Mean Household Characteristics: Owners

	sample mean
housing	49,200
property value	53,528
income	22,921
spouse	0.76
male head	0.83
white	0.93
immigrant	0.05
disabled	0.15
head education	12.67
spouse education	9.06
head high school grad	0.36
head college grad	0.12
spouse high school grad	0.37
spouse college grad	0.06
age	45.7
number kids	1.05
sample size: owners	6,983
sample size: all	10,484

Notes: Variable definitions are in Appendix A.

Table 2
The Choice of Community and Tenure:
Coefficient Estimates of Conditional Variables
from Mixed Logit Equation (4)

	housing price	govt expenditure	population density	school quality
baseline	3.697 (2.708)	0.257 (.173)	0.581* (.205)	-0.208* (.070)
extended	-5.132* (.288)	-1.273* (.595)	-0.300 (.738)	0.071 (.183)

Notes: Standard errors are in parentheses, and are conditional on the estimated value of price. An * signifies that the coefficient is significantly different from zero in a 5%, two-tailed test. The logit coefficient and standard error on per capita government expenditures have been multiplied by 100. The logit coefficient and standard error on population density have been multiplied by 1000. The estimated logit also includes 16 multinomial variables. These variables are identical to those in the housing demand equations presented in Table 3. The log of the likelihood function is -19,016 in the baseline specification and -18,646 in the extended specification.

Table 3
Estimates of the Demand for Housing:
Baseline and Extended Specifications

	baseline			extended		
	ols	tenure	community	ols	tenure	community
constant	7.062* (.112)	4.862* (.324)	7.526* (.267)	7.296* (.372)	6.593* (.503)	7.982* (.470)
ln price	-.124* (.056)	-.130* (.056)	-.834 (1.655)	-.789* (.040)	-.786* (.040)	-1.628* (.150)
ln income	.194* (.009)	.284* (.015)	.203* (.015)	.203* (.054)	.224* (.055)	.181* (.082)
spouse	-.184* (.050)	-.001 (.056)	-.368* (.082)	-.134* (.050)	-.074 (.058)	-.175* (.077)
male head	-.015 (.025)	-.122* (.029)	.036 (.046)	-.062* (.030)	-.095* (.034)	-.183* (.051)
white	.419* (.022)	.493* (.025)	.338* (.076)	.381* (.026)	.410* (.029)	.327* (.083)
immigrant	.094* (.025)	.089* (.025)	.191* (.065)	.101* (.025)	.098* (.025)	.074 (.054)
disabled	-.034* (.016)	-.014 (.017)	.015 (.031)	.002 (.022)	.005 (.022)	.080* (.031)
head education	.047* (.002)	.049* (.002)	.057* (.006)	.042* (.003)	.043* (.003)	.045* (.006)
spouse education	.024* (.003)	.029* (.003)	.021* (.007)	.019* (.004)	.021* (.004)	.026* (.005)
head high school grad	-.004 (.012)	-.012 (.012)	-.042 (.034)	-.002 (.013)	-.006 (.013)	.029 (.029)
head college grad	.065* (.019)	.068* (.019)	.091* (.031)	.061* (.019)	.063* (.020)	.040 (.027)
spouse high school grad	.017 (.014)	.023 (.014)	.051* (.023)	.015 (.014)	.018 (.014)	-.002 (.022)
spouse college grad	-.020 (.028)	-.036 (.028)	-.068 (.072)	-.020 (.028)	-.025 (.028)	-.131* (.064)
age	.021* (.004)	.051* (.006)	.037* (.008)	.011 (.006)	.022* (.008)	.012 (.009)
age squared	-.00020* (.00005)	-.0004* (.00006)	-.0005* (.0001)	-.00009 (.00006)	-.00018* (.00008)	-.00008 (.00010)
number kids	.026* (.012)	.088* (.015)	-.059* (.030)	.025* (.012)	.048* (.017)	.015 (.023)

number kids squared	-.006* (.003)	-.017* (.003)	.005 (.005)		-.005 (.003)	-.010* (.004)	-.005 (.004)
lambda		.573* (.079)				.200* (.096)	
adj R sqrd	.316	.321	.325		.314	.314	.354

Notes: The extended specifications incorporate permanent income and taxes. Standard errors are in parentheses, and are conditional on the estimated values of price, permanent income, and the selection terms. An * signifies that the coefficient is significantly different from zero in a 5%, two-tailed test.

The row labeled lambda indicates the coefficient and standard error on the selection term in the tenure specification. The 45 selection terms in both Columns 3 and 6 are jointly significant.