Housing Tenure Choice and the Dual Income Household

Steven Carter *
Department of Economics
University of California, Irvine

November 12, 2007

Abstract

Housing tenure choice has been the subject of a very large literature. Many treatments have sought to estimate the effect of household income on the likelihood of home ownership. To date, no study has ever disaggregated the household income of married couples into the separate labor income components to see if one income has more effect than another. This paper estimates the effect of separate incomes on the tenure choice of housing. Moreover, it considers the possibility that the second income in the household may be endogenous to tenure choice. If so, then the endogeneity must be controlled to avoid bias in the estimates. Results show that failure to control for the endogeneity downward biases the coefficients. When endogeneity is controlled for, the effect of secondary income significantly increases the probability of home ownership by 4%-6%.

1 Introduction

"Home ownership is a national priority."

-Housing and Urban Development

Home ownership is seen as one of the crowning achievements in a person's life cycle. For many years, it has been a large part of the American Dream, where the model lifestyle

Address for correspondence: Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100. E-mail: sjcarter@uci.edu. I am grateful to Jan Brueckner, Marianne Bitler and David Brownstone for comments and advice. I also thank Stuart Rosenthal for providing his MSA wage index for this project.

included a family and a house in the suburbs. In addition, home ownership (and the capital gains it generates) is the primary way households generate wealth. Whether as a consumption good or investment good, owner-occupied housing is encouraged by government entities since it is seen as a means for a more stable society.

Acquiring shelter is an economic activity in which virtually all members of society participate, either in the rental or owner-occupied market. Recently, there have been sharp increases in house prices, making housing affordability an issue of concern among the public and government officials. With these increases in house prices over a short period of time, households may be constrained by a fixed income, such that affordability can only come if a non-working member of the household enters the work force. Building on this idea, the present paper seeks to model housing tenure choice when household income is disaggregated into separate incomes for both the husband and wife in a married household. The paper recognizes the potential endogeneity of the additional income and tests for biases in the estimates of the individual income coefficients when endogeneity is ignored. Results show that a second income significantly increases the probability of home ownership. Furthermore, the second income is indeed endogenous and its estimated impact is downward biased if this endogeneity is not taken into account.

Given the importance of housing as a commodity, it is no wonder that housing tenure choice has been such an intense focus of study. The first group of such studies seeks to understand the general behavior of a household and to estimate, based on the household's characteristics, the probability of home ownership (Maisel (1966), Shelton (1968), Kain and Quigley (1972), Carliner (1974)). These studies agree that the likelihood of home ownership increases with income. Shelton also argues that households are forward looking, with expected duration in a residence being a large factor in determining tenure choice. This idea of considering the household life cycle is further developed by McCarthy (1976), who finds significant differences in the likelihood of home ownership based on the life-cycle stage of the

household.

The first theoretical treatment of tenure choice was carried out by Artle and Varaiya (1978). They develop a continuous-time life cycle model in which households, under perfect foresight, choose to own or rent based on how ownership affects the lifetime consumption path. In this model, households continuously accumulate wealth, and at some point in time purchase a house using the accumulated wealth as a down payment. Brueckner (1986) proposes a simplified two period model which clarifies the trade-off between a renting and owning through the down payment mechanism.

Henderson and Ioannides (1983, 1986, 1989) extend the work of Artle and Varaiya to model tenure choice with regard to taxes and an investment portfolio that includes housing. Households who demand more "investment" housing than "consumption" housing choose to occupy part of their investment and are owner-occupiers. Households who demand more consumption than investment housing must choose between renting and owner-occupancy. This choice is based on the perceived distortion of investment levels of the household, where households who owner-occupy must equate consumption and investment demand so that 100% of the investment housing is consumed. If this distortion is too costly, the household may choose to rent instead. They also show that progressive tax systems induce more home ownership for high income households. As seen in this result, tax rate is an important concept in tenure choice studies.

Goodman (1988, 1990) introduces the notion of permanent income and ownership's relative price to renting (known as the rent-value ratio) in the tenure choice model, with the inclusion of demographic variables (also revisited by Boehm and Schlottmann (2004)). He finds that increases in permanent income and decreases in the rent-value ratio have the largest impact on tenure choice.[1]

[1]Though not directly related to the current paper, Goodman and Kawai ((1984), (1985), (1986)) examine the general demand for housing under different assumptions, including separate demand for owner-occupied and rental housing.

With the main foundations of the tenure choice model (i.e. income, life-cycle) established by the literature, tenure choice research expanded to examine more complicated models. Brownstone and Englund (1991) extend the standard binary tenure choice model to consider a third tenure option (owner-occupied apartments). As discussed above, taxes play a major roll in tenure choice, and Narwold and Sonstelie (1994) measure the effect of the combined state and federal marginal tax rate. They find that, as a household's marginal tax rate increases, owner-occupancy is more likely because home ownership shields more income from taxes.

A large body of work considers wealth and borrowing constraints and how these restrictions affect different types of households (Haurin et al. (1989, 1996a, 1996b)). Wealth and tenure choice, especially among young married couples, are jointly determined through a savings decision by the household, which makes wealth endogenous to tenure choice. After properly controlling for the endogeneity of wealth, the results show that lower levels of wealth reduce the probability of home ownership. Controlling for wealth, the effects of wage, age, and other characteristics are shown to have similar effects as in previous work.

A more recent body of work considers the tenure choice model under income uncertainty (Haurin (1991), Fu (1995), Robst et al. (1999), Ortalo-Magné and Rady (2002), Davidoff (2006)). With uncertain incomes, households can use owner-occupied housing to hedge against the risk of income volatility. This work identifies the covariance between income and house prices as a factor of home ownership, with decreases in the covariance increasing the likelihood of home ownership.

A common theme of research on tenure choice is that higher household income increases the likelihood of home ownership. What previous work does not consider is that income from different sources in the household (i.e. husband's income and wife's income) may have different effects on tenure choice. The separation of the incomes allows for the possible joint determination of tenure choice and spousal labor supply. This joint decision of the household

may especially be relevant in the presence of rising house values, which can constrain a household's ability to achieve home ownership. Households wanting to transition to home ownership may then choose to send a second laborer into the work force in the face of high local house prices. Those households who choose to send the wife into the labor force may have high unobservable preferences for home ownership, which causes her to work harder than average to make ownership possible indicating that her income has an overinflated effect on the home ownership indicator. But modelling the wife's decision to work as a binary indicator may not capture the ability to transition into home ownership as effectively as modelling the actual income, which must be treated as endogenous. This study contributes to the tenure choice literature by disaggregating the household income to measure separate effects for each component and by controlling for the potential endogeneity of the secondary income.

2 Model and Estimation

2.1 Model

To address the impact of a second income on the tenure choice of a household, the following model is used:

$$\begin{aligned} y_{i1}^{*} &= y_{i2}\gamma_2 + X_{i1}'\beta_1 + \epsilon_{i1} \\ y_{i1} &= 1 \text{ if } y_{i1}^{*} > 0 \\ y_{i1} &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

In (2), y_{i1} is the binary tenure choice indicator for household i, which takes the value $y_{i1} = 1$ (denoting home ownership) if the latent variable $y_{i1}^{*} > 0$,$y_{i1} = 0$ otherwise (denoting a renter). y_{i2} is the second income for household i while X_{i1} is a set of variables that affect

tenure choice, including the primary earner's permanent income, which allows for estimation of separate income effects. The errors follow a normal distribution, $\epsilon_{i1} \sim N(0,1)$, indicative of a probit model.

As explained above, it is possible that the second income, y_2, is endogenous to tenure choice, and ignoring this possibility could lead to bias in the estimate of γ_2. Also, the fact that not all households have a second income complicates matters in controlling for endogeneity. The censored nature of the second income suggests a tobit style regression:

$$\begin{aligned} y_{i2}^* &= X_{i2}'\beta_2 + \epsilon_{i2}, \\ y_{i2} &= y_{i2}^* \text{ if } y_{i2}^* > 0 \\ y_{i2} &= 0 \text{ otherwise} \end{aligned} \tag{2}$$

X_{i2} is a set of variables containing X_{i1} as well as a number of instruments, and $\epsilon_{i2} \sim N(0, \sigma^2)$. In contrast to standard instrumental variable procedures, which treat all or part of the model as linear, this model represents an innovation due to the attempt to handle the censored nature of the endogenous variable in the first stage.

The joint modelling of equations (2) and (3) requires a nontrivial likelihood function, part of which is analytically intractable.[2] Since maximum likelihood is thus infeasible, an available estimation method is a non-linear two-stage procedure, which accounts for the censored nature of the secondary income in the first stage as well as the binary tenure choice in the second stage.

[2]This part of the likelihood requires analyzing the double integral over the joint space where both y_1^* and y_2^* are less than zero. Technically, this could be done by a burdensome iterative algorithm like the GHK. However, the sample size is large enough such that worries about efficiency of the estimates are unwarranted.

2.2 Estimation

The two-stage procedure follows the same format as the usual two-stage least squares (2SLS) procedure used in linear models. The first stage estimates the second-income equation in (3) by maximum likelihood. Computation of the fitted values for use in the first stage requires special attention due to the non-linearity of the equation. In general linear models, the expectation $E(y|X)$ is simply $X\beta$. However, with the tobit model, the expectation[3] of y_2 is

$$E(y_2|X_2) = P(y_2 = 0|X_2) \cdot 0 + P(y_2 > 0|X_2) \cdot E(y_2|X_2, y_2 > 0). \tag{3}$$

The first term on the right hand side of (3) drops away. The term $P(y_2 > 0|X_2)$ represents the probability that y_2 is observed, given X_2. This probability is simply the normal cumulative distribution (CDF) evaluated at the linear mean scaled by the standard deviation,

$$P(y_2 > 0|X_2) = \Phi\left(\frac{X_2\beta_2}{\sigma}\right). \tag{4}$$

The expectation $E(y_2|X_2, y > 0)$ is composed of two parts; the linear mean, $X_2\beta_2$, and the Inverse Mills Ratio of $\frac{X_2\beta_2}{\sigma}$ scaled by the standard deviation of ϵ_2:

$$E(y_2|X_2, y > 0) = X_2\beta_2 + \sigma\frac{\phi(\frac{X_2\beta_2}{\sigma})}{\Phi(\frac{X_2\beta_2}{\sigma})}. \tag{5}$$

The Inverse Mills Ratio (also known as the selection hazard) adjusts the expectation to account for the excluded censored observations. Multiplying (4) and (5), the full expectation in (3) equals

$$E(y_2|X_2) = \Phi\left(\frac{X_2\beta_2}{\sigma}\right) X_2\beta_2 + \sigma\phi\left(\frac{X_2\beta_2}{\sigma}\right). \tag{6}$$

[3]The derivation of the expectation is in Wooldridge (2002), pps. 521-522.

The second stage probit regression carries out a maximum likelihood estimation of

$$y_1^* = E(y_2|X_2)\gamma_2 + X_1\beta_1 + \epsilon_1. \tag{7}$$

This equation is just equation (2) with the newly fitted values of y_2 in place of the actual values, making the resulting estimator $\hat{\gamma_2}$ consistent.

For identification, the first stage requires a set of instruments that are correlated with y_2, but uncorrelated with the errors, ϵ_1. Methods to test the satisfaction of these requirements are not available for exactly identified models (i.e. equal number of endogenous regressors and instruments). When more instruments than endogenous variables are used, however, overidentification tests are available to check the validity of the instruments (given one of the instruments is already valid).

In general, this test measures the goodness-of-fit, (i.e. R^2) from a regression of the residuals $\hat{\epsilon_1}$ from the IV procedure on the set of instruments. Intuitively, if the goodness of fit is high, then the instruments fail to satisfy the condition of low correlation with the error term. Numerically, the test statistic is the quantity NR^2 which follows a χ^2_{L-K} distribution, where L is the number of instruments and K is the number of endogenous variables. If NR^2 is greater than the critical value corresponding to the $L-K$ degrees of freedom, then the null hypothesis is rejected, indicating that the set of instruments may not be valid. A thorough discussion of IV regression procedures is found in Baum et al. (2003).[4] However, in the non-linear setting, the NR^2 statistic is inappropriate as there is no R^2 in the second stage probit regression presented here. Though there are methods to compute a "pseudo-R^2" statistic for the probit model, this paper employs the overidentification test statistic based

[4]This overidentification test statistic goes by many names depending on the setting of the estimation. For 2SLS methods assuming homoscedasticity, it is known as Sargan's statistic, which is the test used in this paper. For GMM methods without the homoscedastic assumption, it is known as Hansen's J-test which is a general case of Sargan's statistic. They both, however, measure the same thing.

on the generalized minimum distance estimator (see Lee (1992), Newey (1987)).[5]

3 Data

Multiple data sources are used to construct the variables used in the empirical model. The primary source of data is the 1992 cross section of the Panel Study of Income Dynamics (PSID). The PSID is a bi-annual survey of households that collects data on virtually all aspects of the household, with a heavy focus on household income. The extra attention paid to household income allows for easy disaggregation of household income into separate incomes for both husband and wife. Another main attraction of using the PSID is the restricted availability of detailed geographic identifiers to match observations to small regional areas of the nation, which allows for greater control of local level variation that might be present in the data. The local identification also facilitates merging other data sources.

Since the main focus is on households with two incomes (or the potential to have two incomes), the sample is restricted to married or cohabitating couples, which drops all single-head household observations. Secondary income is defined as the wife's labor income, while primary income is defined as the husband's permanent income, which is discussed later in this section. While there are cases where the wife earns more labor income than the husband, they make up only a small percentage of the observations.

Explanatory variables used in the regressions are household characteristics such as number of children (17 years or younger), age of the youngest child, age of the wife, and dummy variables measuring her education (high school drop out, some college education, and college graduate, with high school graduate being the excluded category). Other control variables include regional dummies for the Northeast, North Central, and South, with West being the

[5]This option on STATA is available in the IVPROBIT routine. However, the routine treats the endogenous variable as linear. The linear treatment slightly changes the results, but they are close enough to have confidence in the test statistic.

excluded group. A dummy variable representing whether or not the household resides in an urban area (population 250,000 or greater) controls for population size of the surrounding metropolitan area.

Because the restricted data allows the household's Metropolitan Statistical Area (MSA) to be identified, three MSA level variables are used in the regressions: a house price index, median monthly rent, and a wage index. The MSA house price index is compiled by the National Association of Realtors. This index measures quality-adjusted relative prices across MSAs, with higher values indicating more expensive areas. Median rent values are generated from the full cross section of the PSID. The median is first computed for each MSA in the cross section, and that median is then assigned to each household residing in the MSA. A quality adjusted wage index is used to control for income variation across MSAs. This index is the MSA fixed effect from a regression of individual incomes on personal characteristics, carried out by Chen and Rosenthal (2007).

To best understand how the data are put together, it is helpful to describe the merging process that combines all the data sets. The 1992 cross section of the PSID contains 9829 household level observations, each with a unique interview number. The geocoded supplement is then merged with the cross section. Any respondents not identified by an MSA are dropped, reducing the data to 9371 observations. The next step is to drop all single-head households, which reduces the sample size to 5027 households. The data are then cleaned by dropping 970 observations with missing values, leaving 4057.

The MSA house price index for 1990 is available for 113 MSAs. The MSA wage index, however, is available for 322 MSAs. This large discrepancy in the number of MSAs is due to the difference in MSA definitions between 1990 and 2000. Chen and Rosenthal's data set is based on the 2000 Census, while the MSA house price index is based on the 1990 Census.[6] When merging these data together, 209 MSAs from the wage index are lost. The indices

[6]It is possible that the National Association of Realtors' list is not exhaustive of all MSAs in the U.S.

for these MSAs are then merged with the main data set. Between the main data and the indices, there are only 70 common MSAs in the sample. After dropping observations with missing values, the data have 2431 observations with complete information.

As previously mentioned, this paper uses the husband's permanent income as a control variable. This variable is useful on two levels. First, permanent income reflects an average stream that the household would expect to earn over an extended period, a crucial factor in the home purchase decision. Also, use of permanent income overcomes the possible endogeneity of primary income raised in the female labor supply literature.

Unfortunately, the PSID does not offer a solid history of earnings for the heads of household over a long period of time, which would be desirable in modelling long term income streams. The 1979-1996 National Longitudinal Study of Youth (NLSY) panel, however, has a very complete earnings history for its participants. For males in the NLSY, income regressions are estimated using educational, regional and demographic variables (see Boehm and Schlottmann (2004)). Results from these regressions are then used to calculate the permanent income from the data in the PSID, equal to the fitted values using NLSY regression coefficients. One drawback to using the NLSY is the difference in the ages in the two data sets. Since the NLSY involves a single cohort that was sampled at a relatively young age, the ages for the two samples may not match very well with some of the older ages of the PSID, which can affect out of sample predictions.[7]

As previously mentioned, income tax rates play an important role in the home ownership decision, given that the tax system provides benefits for the owner-occupier. Therefore, to control for the effect of tax rates, this paper uses the variable Sum of Tax Rates, which is equal to the sum of the federal and state marginal tax rates. These tax rates are computed using the NBER TAXSIM program.[8] This program uses 22 variables to simulate the marginal

[7]Currently, this paper assumes that the effect of age on permanent income is linear.

[8]www.nber.org/taxsim

tax rate of each household, including the number of children, income (both husband and wife), filing status, state location, and home ownership. To avoid any possible endogeneity in the tax rate, the rate is computed using the husband's permanent income only, assuming no spousal income and no home ownership, but assuming that all households are filing a joint tax return. These restrictions imply a baseline tax rate before any changes to tenure status are made by the household, which allows for fair comparison across both renting and owner-occupying households.

Table 1 shows some selected variable means and standard deviations, separately for renters and home owners, along with difference-of-means t-statistics. It is clear from the t-statistics that there are significant mean differences across tenure choices. Both the husband's permanent income and wife's labor income are higher in the home owner category, and there are slightly more working wives in the home owner category as well, though there is no statistical difference. The MSA house price index is higher for renters than for home owners. The age variables indicate differences in tenure choice across the life cycle of the households, with older couples being owner-occupiers and younger couples renters. Also, it is important to note that, on average, home owners face a higher baseline marginal tax rate, providing an incentive to be owner-occupiers. In total, approximately 33% of households are renters and 67% are owner occupiers.

To control for endogeneity, proper instruments are needed. For the wife's income, the instruments include dummy variables representing the wife's educational level: High School Dropout, Some College Education and College Graduate (College Graduate also includes any post graduate education or degrees). Educational attainment is highly correlated with income, but the wife's education, however, is not expected to be correlated with the error term in the housing tenure choice equation, holding income constant. Other instruments used are the highest grade completed for both Father and Mother of the wife (HGC Father and HGC Mother), an interaction term equal to the product of the age of the youngest child

and the number of children in the household, and the MSA wage index described above.

4 Results

To measure the benefits from the estimation method described in Section 2, it is helpful to compare the new technique to standard IV estimation methods, as well as to a model where exogeneity is assumed. This section discusses the results of a simple (naive) probit model where endogeneity is ignored, a linear 2SLS model which accounts for endogeneity but treats the dependent variables as linear, the IVPROBIT routine, which estimates the probit second stage but still treats the first stage as linear. Finally, results for the proposed model (incorporating the endogeneity and censored nature of the wife's income) are presented.

4.1 Naive Results

Table 2 contains results of the simple probit model with no control for endogeneity. The first column uses total labor income of the head (permanent income) and wife as the income control, while the second column disaggregates the labor income. The third and fourth columns show the marginal effects of changes in the independent variables on the probability of home ownership.[9]

The coefficient of the husband's permanent income is much larger than that of the wife's labor income, and both are statistically significant. In terms of marginal effects, the wife's income has very little impact on the probability of being a home owner with a 1% increase in wife's income yielding a 1.4% increase in the probability. Comparatively, a 1% increase in the husband's permanent income increases the probability of home ownership by 13.6%.

[9]Marginal effects are the differences in the probability of choosing home ownership based on comparing two different values of X. For continuous variables in X, the marginal effect is essentially the derivative of the normal CDF evaluated at $X\beta$, measuring the rate of change in the probability for small increases in X. Marginal effects for dummy variables measure the added probability of taking a value equal to 1 compared to 0 for that dummy variable.

The other variables included in the model are the urban dummy, the regional dummies and the MSA house price index. For both naive models, the house price index has a significantly negative coefficient, but the marginal effect in both cases is quite small (-.001). The effect of Median Rent is negligable, since the marginal effect is 0 to three significant digits. Increases in the sum of tax rates lead to increases in the probability of home ownership, but the change in probability is quite small (.009). Apart from the household incomes, the largest determinant of home ownership is the regional location. All three regions reduce the probability of home ownership relative to the omitted group (West). This result is most likely due to the fact that the West region includes sparsely populated states where home ownership is easy.

4.2 IV Method Results

Table 3 presents results of a linear treatment of the tenure choice model taking into account the possible endogeneity of the wife's income. This 2SLS procedure accomplishes two things. First, it gives a general idea of what the proposed two stage results should be. Second, it produces the correct standard errors and gives a good indication of which coefficients are statistically significant. As a linear probability model, the coefficients from the linear setup are easily interpreted as marginal effects.

Table 3 shows that, when controlling for endogeneity, the effect of the wife's income is larger than in the naive probit model. A one percent increase in the wife's income leads to increases in the probability of home ownership from .05 to .07, all of which are statistically significant. The endogeneity corrected model maintains approximately the same effect of the husband's permanent income as the naive model. The coefficients of the other variables are remarkably similar to the marginal effects in Table 2. The coefficients for Residual represent the test for endogeneity. These coefficients are the result of a regression of the first stage residuals, as well as all the other variables,on the home ownership dummy. One distinction to make is that the actual wife's income (not the fitted values) are used. If these coefficients

are significantly different than zero, then the wife's income is endogenous. In columns 1-4, the coefficients are significant indicating endogeneity of the wife's income.

The overidentification p-values in columns 1 and 2, however, are less than .05, indicating that the instrument set is not valid. The instruments for column 1 include the set discussed previously as well as the wife's age squared. Column 2 drops the wife's parent's education from the instrument set, but the instrument set is still not valid. The model in column 3 drops the wife's age squared[10] and as a result, the instruments passes the overidentification test, as indicated by the p-value in column 3 (greater than .05). The interaction term equal to the product of age of the youngest child and number of children is dropped from the instrument set in the column 4 results but does not change the conclusion regarding the validity of the instruments.

The IVPROBIT routine from STATA provides an avenue for performing a probit regression that includes endogenous right-hand variables. This routine improves on the linear treatment of the 2SLS method while generating the correct standard errors. It also provides the overidentification statistic for the instrument set. It fails, however, to account for the censored nature of the endogenous variable. Table 4 provides estimates of the IVPROBIT routine.

The coefficients for the wife's income using IVPROBIT are much larger than the naive probit coefficients, further confirming the negative bias from endogeneity. The results in column 1 of Table 4 only slightly vary from those in column 2 due to the difference in instruments. As shown by the p-value, the full instrument set (wife's educational dummies, wife's parents' education, MSA wage index and the interaction term) fails to pass the validity test. Column 2 drops the wife's parents' education from the instrument set, and the overidentification test now shows the remaining instruments as valid. Columns 3 and 4, which show the

[10]During initial data exploration, the variable wife's age squared was found to have a significant coefficient on the wife's income, and was initially included as an instrument. However, this variable is not used in future instrument sets including the Heckman procedure.

marginal effects, indicate that a 1% increase in the wife's income increases the probability of home ownership by 6%, moderately higher than the 2SLS estimate of 5%. The rest of the variables have marginal effects similar to those in the 2SLS model.

4.3 Two Stage Method Results

The discussion now moves to the first stage results in the proposed model of the current paper. Table 5 presents the first stage tobit regression results with ln(1+wife's income) as the dependent variable.

The first stage regression details how the variables affect the wife's income. From column 1, a one percent increase in the permanent income of the husband reduces the wife's income by -.5%, with the coefficient representing an elasticity. An increase in the number of children decreases the wife's income, but that decrease is offset by the interaction term that takes into account the age of the youngest child. As the youngest child ages, the negative effect of the number of children decreases. As the wife ages, her income decreases by .12% per year, and living in an urban area leads to a .8% increase in the wife's income. The results indicate an income penalty for the regional variables (relative to the omitted region), all of which have negative coefficients. Only the North Central region has a significant coefficient, however. The wife's education variables all have significant coefficients and the expected signs. Relative to a high school graduate, high school dropout status leads to almost a 2% decrease in income, while some college education leads to a 1.1% increase in income. A college degree leads to a 1.7% increase in income above a high school graduate. These educational effects seem smaller than expected at first, but it should be noted that these effects are measured by incorporating the zero value incomes. The presence of the zero values may attenuate the effect of education on income. As seen in column 2, the results do not change much with the exclusion of the interaction term and the wife's father's education. The only significant change is in the coefficient of the number of children, which is smaller

without the interaction term.

An odd result that is not expected is the negative sign on the MSA house price index. Higher house prices should require higher, not lower, incomes from the working wives, holding the husband's income level constant. However, the coefficient on the MSA house price index in column 1 indicates that a one unit increase in the index reduces the wife's income by .01%. To possibly remedy this result, the ratio of the house price to the median rental rate (HPI Ratio) is used in place of the separate price and rent variables, with its coefficient reported in column 2. But the ratio's coefficient has the same negative sign and is highly significant.

One possible explanation for this result is an indirect connection between labor force participation and house prices. The first ingredient in this connection is the tendency of larger metropolitan areas to have more expensive housing. Also, larger areas also have longer commute times for workers, including working wives. Therefore, higher priced areas are likely to have longer commutes, possibly causing wives to choose not to work. Black et al. (2007) provide evidence on this effect.

To further explore this odd result, and given the conclusion in Black et al. (2007), it may be useful to compare the tobit estimates to a regression using only the observations where households have labor income from the wife. Though not reported, an OLS regression on this subsample shows that the HPI ratio coefficient becomes positive. This finding lends credibility to the idea that women who face high commuting costs would rather not work, but that if they do, they have higher incomes in high priced housing areas. Table 6 presents results from a similar regression employing the Heckman correction for sample selection bias that may be present in the OLS regression. The first column of Table 6 reports the income regression for just the non-zero wife incomes, while the second column shows the selection regression for positive wife's income. The HPI ratio variable in the first column now has the expected positive sign indicating that a one unit increase in the ratio increases the wife's income by 1.1%. The HPI ratio also has a negative sign in the selection equation, which

is consistent with the findings of Black et al. (2007). The Heckman results show that the negative coefficient of the MSA house price index in the tobit regressions (using both the zeros and positive incomes), though not initially expected, is an acceptable result.

The main purpose of the current paper is to incorporate the possible endogeneity and censored nature of the wife's income into the housing tenure choice model. With the first stage complete, the second stage probit estimates are presented in Table 4.3. The results further confirm the significant impact of the wife's income on housing tenure choice. The coefficients are similar to those from the IVPROBIT routine, though somewhat smaller, with marginal effects in the .04-.05 range. This discrepancy is most likely because of the non-linear treatment of the first stage in the two stage procedure as opposed to the linear treatment in IVPROBIT.

When controlling for endogeneity, it is clear that the secondary income produces a bias in a naive model of tenure choice. As discussed previously, the bias from endogeneity was expected to be positive due to harder working wives in the pursuit of owner-occupied housing. However, the results indicate that the bias is negative. This bias may be the result of a higher-than-average preference for home ownership coming from having a wife who concentrates on home production, which makes her care more about the quality of the home environment. This association may lead to a negative correlation between the error term and the wife's income, which constricts the naive estimates for the effect of the wife's income.

The results show that two lessons can be learned from this study. First, when endogeneity is controlled for, the estimated effect of the wife's income is three times larger than that of the effect in a naive model. Second, the result does not seem to depend on the type of endogeneity control used, whether it be a linear first stage or a more sophisticated approach that handles the censored nature of the wife's income.

5 Conclusion

This paper shows that income has a multidimensional effect on housing tenure choice. When household income is disaggregated, each component is shown to have its own significant effect on the propensity of home ownership. Moreover, the second income of the household is endogenous to tenure choice. When endogeneity is controlled for, the estimated effect of a second income is three times larger than the effect under a naive specification that ignores endogeneity. Thus, through the proposed two-stage procedure, which accounts for both the endogeneity and the censored nature of the second income, more accurate estimates of the income effect on tenure choice can be generated. With an accurate estimate of the effect of a second income, the results indicate that the policy makers can confidently pursue policies that make it less costly for a household to have two incomes, thus increasing the possibility of home ownership.

References

Black, D., N. Kolesnikova and L. Taylor (2007) "Earnings Functions When Wages and Prices Vary by Location" Federal Reserve Board Working Paper 2007-031a.

Artle, R and P. Varaiya (1978) Life Cycle Consumption and Ownership, Journal of Economic Theory 18, 35-58.

Boehm, T H. Herzog and A. Schlottmann (1991) Intra-Urban Mobility, Migration, and Tenure Choice, The Review of Economics and Statistics 73:1, 59-68

Boehm T. and A. Schlottmann (2004) "The dynamics of race, income, and homeownership" *Journal of Urban Economics* 55, 113-130.

Brownstone, D and P. Englund (1991), The demand for housing in Sweden: Equilibrium

choice of tenure and type of dwelling, Journal of Urban Economics, 29:3, 267-281

Brueckner, J.K. (1986) The Down Payment constraint and housing Tenure choice: A Simplified Exposition, Regional Science and Urban Economics, 16, 519-525

Carliner, G. (1974), Determinants of home Ownership, Land Economics 50:2, 109-119

Chen, Y. and S. Rosenthal, (2007) "Local Amenities and Life Cycle Migration: Do People Move for Jobs or Fun?", Working Paper

Davidoff, T. (2006) "Labor income, housing prices and homeownership" *Journal of Urban Economics*, 59, 209-235.

Dieleman, F., W. Clark and M. Deurloo (1994) "Tenure choice: Cross-sectional and longitudinal analyses", *Journal of Housing and the Built Environment* 9, 229-246.

Dynarski, M. and Sheffrin, S., 1985. Housing purchases and transitory income: a study with panel data. Review of Economics and Statistics 67, 195-204

Fu, Y., 1995. Uncertainty, liquidity, and housing choices. Regional Science and Urban Economics 25, pp. 223-236

Goodman, A. (1988) "An econometric model of housing price, permanent income, tenure choice, and housing demand", *Journal of Urban Economics* 23, 327

Goodman, A.C. (1990), Demographics of individual housing demand Regional Science and Urban Economics, 20:1 83-102

Goodman, A (2003) "Following a panel of stayers: Length of stay, tenure choice, and housing demand", *Journal of Housing Economics* 12, 106-133.

Goodman, A and M. Kawai (1984) Replicative Evidence on the Demand for Owner-Occupied and Rental Housing, *Southern Economic Journal* 50:4 1036-1057

Goodman, A and M. Kawai (1985) Length-of-Residence Discounts and Rental housing Demand: Theory and Evidence, *Land Economics* 61:2, 93-105

Goodman, A and M. Kawai (1986) Functional Form, Sample Selection and Housing Demand, *Journal of Urban Economics* 20:2 155-167

Haurin, D., 1991. Income variability, homeownership, and housing demand. Journal of Housing Economics 1, 60-74

Haurin, D. and Gill, L., 1987. Effects of income variability on the demand for owner-occupied housing. Journal of Urban Economics 22, pp. 136-150

Haurin, D, P Hendershott and D.C. Ling, Homeownership Rates of Married Couples: An Econometric Investigation, NBER Working Paper 2305

Haurin, D, P Hendershott and S. Wachter (1996) Expected Home Ownership and Real Wealth Accumulation of Youth, NBER Working Paper 5629

Haurin, D, P Hendershott and S. Wachter (1996) Borrowing Constraints and the Tenure Choice of Young Households, NBER Working Paper 5630

Henderson, J and Y. Ioannides (1983) A Model of Housing Tenure Choice, American Economic Review 73, 98-113

Henderson, J and Y. Ioannides (1986) Tenure Choice and the Demand for Housing, Economica 53, 231-246

Henderson, J and Y. Ioannides (1989) Dynamic Aspects of Consumer Decisions in Housing Markets, Journal of Urban Economics 26:2 212-230

Kain J. and J. Quigley (1972) Housing market discrimination, home ownership and savings behavior, American Economic Review 60, 263-277

Kan, K. (2000) "Dynamic Modelling of Tenure Choice" *Journal of Urban Economics* 48, 46-69.

Lee, L. (1992) Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables, Econometric Reviews 11:3 319-328.

Maisel, S. (1966) Rates of Ownership, Mobility and Purchase, Essays in Urban Land Economics, 88-89.

McCarthy, K. (1976), The household life-cycle and Housing, Papers in Regional Science 37:1, 55-80.

Narwold A. and J. Sonstelie (1994), 'State income taxes and homeownership: a test of the tax arbitrage theory", *Journal of Urban Economics*, 36, 249-277.

Newey, W. (1987) Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables, Journal of Econometrics 36, 231-250.

Ortalo-Magné, F. and S. Rady, "Tenure choice and the riskiness of non-housing consumption", *Journal of Housing Economics*, 11, 266279.

Painter, G. (2000) Tenure Choice with Sample Selections: Differences among alternative samples, Journal of Housing Economics 9, 197-213.

Shelton, J. (1968), The Cost of Renting versus Owning a Home, Land Economics 44:1, 59-72

Wooldridge, J. (2002) Econometric Analysis of Cross-Section and Panel Data, MIT Press

Table 1: Descriptive Statistics and Difference of Means Statistics

	Renters		Home Owners		
Variable	Mean	St. Dev	Mean	St. Dev	t-Statistic
Permanent Income	76157.5	176904.4	110179.8	162112.9	-4.58
Wife Income	9166.1	10474.5	13356.6	15032.6	-7.98
Wife Indicator*	0.68	0.48	0.71	0.45	-1.44
MSA HPI**	113.7	53.3	99.1	41.5	6.81
Age Head	39.55	14.61	46.5	13.65	-10.56
Age Wife	37.07	13.68	43.6	13.07	-11.85
Number of Children	1.3	1.3	1.1	1.2	3.69
Age of Youngest	3.5	4.4	4.2	5.2	-3.82
Marginal Tax, Federal	10.0	13.2	14.2	11.0	-7.80
Marginal Tax, State	2.1	2.8	3.0	2.9	-7.75
N	799		1633		

* This wife indicator denotes the percentage of wives with a labor income.
** This is the MSA level house price index.

Table 2: **Naive Probit Estimates with Marginal Effects**

Dependent Variable: Home Ownership Dummy

	(1)	(2)	(3)	(4)
Total Labor Income $(ln)^{\dagger}$	.508***	-	.179	-
	(.063)			
Wife's Income (ln)	-	.038***	-	.014
		(.007)		
Perm. Income (ln)	-	.514***	-	.136
		(.032)		
Number of Children	.088***	.069***	.03	.02
	(.025)	(.025)		
Age Wife	.005	.012	.001	.004
	(.004)	(.004)		
Sum Tax Rates	.027***	.027***	.009	.009
	(.002)	(.002)		
Urban	.212***	.208***	.07	.07
	(.070)	(.070)		
MSA HPI	-.005***	-.005***	-.001	-.001
	(.001)	(.001)		
MSA Median Rent	.0001	.0001	.000	.000
	(.0003)	(.0003)		
N. East	-.644***	-.653***	-.24	-.25
	(.122)	(.123)		
N. Central	-.323***	-.296***	-.118	-.11
	(.114)	(.114)		
South	-.402***	-.380***	-.14	-.14
	(.095)	(.095)		

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level
† Total Labor Income is the sum of the Permanent Income and the wife's income.
(3) is the marginal effect for the model in (1) and (4) is the marginal effect for the model in (2).

Table 3: **2SLS Results**

Dependent Variable: Home Ownership Dummy

	(1)	(2)	(3)	(4)
Wife's Inc. (ln)	.071***	.071***	.058***	.058***
	(.008)	(.008)	(.009)	(.011)
Perm. Income (ln)	.140***	.140***	.136***	.136***
	(.021)	(.021)	(.020)	(.021)
Number of Children	.054***	.054***	.048***	.048***
	(.009)	(.009)	(.009)	(.010)
Age Wife	.010***	.010***	.008***	.008***
	(.002)	(.002)	(.002)	(.002)
Sum Tax Rates	.007***	.007***	.007***	.007***
	(.0008)	(.0008)	(.0008)	(.0009)
Urban	.012	.012	.025	.025
	(.026)	(.026)	(.025)	(.026)
MSA HPI	-.001***	-.001***	-.001***	-.001***
	(.0004)	(.0004)	(.0003)	(.0003)
MSA Median Rent	-.0003*	-.0003*	-.0002	-.0002
	(.0001)	(.0001)	(.0001)	(.0001)
N. East	-.213***	-.213***	-.210***	-.209***
	(.042)	(.042)	(.041)	(.041)
N. Central	-.067*	-.067*	-.074**	-.074**
	(.039)	(.039)	(.037)	(.037)
South	-.101***	-.101***	-.104***	-.104***
	(.032)	(.032)	(.031)	(.031)
Residual	-.064***	-.064***	-.04***	-.04***
	(.007)	(.007)	(.008)	(.008)
Sargan's OverID P-value	.009	.005	.202	.113

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level
The equality of coefficients in columns 1-2 and 3-4 is due to rounding.
(1) uses the instruments: high school dropout, some college, college graduate, HGC father, HGC mother, age squared, MSA wage index and interaction term AYC*NC.
(2) uses the instruments: high school dropout, some college, college graduate, age squared, MSA wage index and interaction term AYC*NC.
(3) uses the instruments: high school dropout, some college, college graduate, MSA wage index and interaction term AYC*NC.
(4) uses the instruments: high school dropout, some college, college graduate and MSA wage index.

Table 4: **IVPROBIT Estimates**

Dependent Variable: Home Ownership Dummy

	(1)	(2)	(3)	(4)
Wife's Income (ln)	.187***	.191***	.062	.062
	(.03)	(.030)		
Perm. Inc. (ln)	.436***	.438***	.134	.134
	(.068)	(.068)		
Number of Children	.152***	.154***	.048	.048
	(.031)	(.031)		
Age Wife	.025***	.026***	.008	.008
	(.006)	(.006)		
Sum Tax Rates	.022***	.021***	.006	.006
	(.003)	(.003)		
Urban	.074	.070	.019	.017
	(.082)	(.082)		
MSA HPI	-.003***	-.003***	-.001	-.001
	(.001)	(.001)		
MSA Median Rent	-.0007	-.0007	-.0002	-.0002
	(.0005)	(.0005)		
N. East	-.695***	-.698***	-.214	-.215
	(.130)	(.130)		
N. Central	-.24*	-.238*	-.061	-.061
	(.117)	(.117)		
South	-.363***	-.362***	-.102	-.102
	(.100)	(.100)		
Walt Test p-value	.000	.000		
OverID p-value	.042	.213		

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level
(1) uses the instruments: high school dropout, some college, college graduate, HGC father, HGC mother, MSA wage index and interaction term AYC*NC.
(2) uses the instruments: high school dropout, some college, college graduate, MSA wage index and interaction term AYC*NC.
(3) and (4) are the marginal effects of (1) and (2) respectively.
The Wald Test determines the endogeneity with .05 rejection level
The OverID p-value is the result of the minimum distance estimator overidentification test with a .05 rejection level.

Table 5: **First Stage Tobit Results**

Dependent Variable: Log of 1+Wife's Income

	(1)	(2)
Perm. Inc. (ln)	-.506**	-.531**
	(.237)	(.240)
Number of Children	-1.148***	-.511***
	(.129)	(.098)
AYC*NC	.121***	-
	(.016)	
Age Wife	-.128***	-.111***
	(.020)	(.020)
Sum Tax Rates	.030***	.038***
	(.009)	(.009)
Urban	.803***	.818***
	(.285)	(.288)
MSA Wage Index	2.112	2.838**
	(1.785)	(1.260)
MSA HPI	-.014***	-
	(.005)	
MSA Median Rent	.005***	-
	(.001)	
HPI Ratio	-	-6.995***
		(1.925)
N. East	-.066	-.032
	(.540)	(.465)
N. Central	-.938*	-1.085***
	(.492)	(.392)
South	-.329	-.377
	(.374)	(.331)
H.S. Dropout	-1.942***	-1.996***
	(.319)	(.316)
Some College	1.140***	.921***
	(.305)	(.305)
College Grad	1.738***	1.574***
	(.308)	(.298)
HGC Father (w)	-.056	-
	(.063)	

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

Table 6: **Wife's Income Regression with Selection Correction**

	Income Regression	*Selection Equation*
Perm. Inc.	-.054 (.063)	-.043 (.054)
Number of Children	-.01 (.034)	-.11*** (.029)
AYC*NC	-.006 (.004)	.013* (.003)
Wife's Age	.030*** (.005)	-.025*** (.01)
Sum Tax Rates	.003 (.002)	.003 (.001)
Urban Dummy	.102 (.076)	.124* (.066)
MSA Wage Ind.	.785** (.332)	-.008 (.288)
HPI Ratio	1.17** (.505)	-1.44*** (.445)
N. East	-.210* (.123)	-.021 (.109)
N. Central	-.118 (.103)	-.135 (.09)
South	.027 (.087)	-.151** (.076)
H.S. Dropout (w)	-.128 (.085)	-.243*** (.071)
Some College (w)	.131* (.079)	.076 (.073)
College Grad (w)	.307*** (.077)	.256*** (.076)
HGC Father (w)	-	.006 (.012)
HGC Mother (w)	-	.015** (.014)
Mills ratio (λ)	-1.29 (.02)	-

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

Table 7: **Two Stage Probit Estimates with Marginal Effects**

Dependent Variable: Home Ownership Dummy

	(1)	(2)	(3)	(4)
Wife Inc. (ln)	.149***	.135***	.052	.047
	(.031)	(.038)		
Perm. Inc.	.414***	.409***	.145	.143
	(.062)	(.062)		
No. children	.124***	.116***	.007	.007
	(.028)	(.030)		
Age Wife	.021***	.020***	.043	.041
	(.006)	(.006)		
Sum Tax Rates	.024***	.024***	.008	.008
	(.002)	(.003)		
Urban	.132*	.144*	.047	.051
	(.075)	(.076)		
MSA HPI	-.004***	-.004***	-.001	-.001
	(.001)	(.001)		
MSA Med. Rent	-.0004	-.0003	-.0001	-.0001
	(.0004)	(.0004)		
N. East	-.669***	-.667***	-.254	-.254
	(.123)	(.123)		
N. Central	-.266**	-.277**	-.097	-.101
	(.115)	(.115)		
South	-.375***	-.379***	-.132	-.133
	(.096)	(.096)		

* Significant at the 10% level
** Significant at the 5% level
*** Significant at the 1% level

(1) uses the instruments: high school dropout, some college, college graduate, HGC father, MSA wage index, and interaction term AYC*NC.
(2) uses the instruments: high school dropout, some college, college graduate, and MSA wage index.
(3) and (4) are the marginal effects of (1) and (2) respectively.