IPRPD
International Journal of Business & Management Studies
ISSN 2694-1430 (Print), 2694-1449 (Online)
Volume 04; Issue no 03: March, 2023
DOI: 10.56734/ijbms.v4n3a2

FACTORS AFFECTING RESIDENTIAL REAL ESTATE INVESTMENTS OF U.S. INDIVIDUALS

Xiaoxuan Ji[1]

[1]*Commonwealth University of Pennsylvania, USA*

Abstract

This paper aims to study the causal relationships between earnings, age, number of kids, and education level of U.S. individuals in residential real estate investments. This empirical study analyzes a total of 421,911 observations based on the interview of indi- viduals from the Survey of Income and Program Participation (SIPP). By employing the econometric techniques of binomial logit and probit models, multinomial logit model, and Tobit model on censored data, the results showed that most individuals in the U.S. are willing to invest in real estate. Interestingly, as the education level increases, people tend not to invest in real estate. However, with the increase in earnings, the number of kids, and age, people are more likely to invest in real estate.

Keywords

Real estate investments; Individual earnings; Education

1 INTRODUCTION

Residential real estate investments play an important role in households and their family lives. This paper aims to study the determinant factors of residential real estate investments and explores the behavior changes of individuals' real estate investment de- cisions based on their earnings, education, and family sizes. The individual-level data is collected from the Survey of Income and Program Participation (SIPP), which is a longitu- dinal survey that provides detailed information regarding earnings, household composition, education, and employment of individuals in the U.S. The data set of this empirical study has a total of 421,911 observations, which includes the real estate status, income, age, and the number of kids of the individuals in the U.S.

In this paper, the methodology I selected to use are qualitative response regression models. Unlike quantitative regression models, qualitative response regression models are types of regression in that the dependent variable takes certain values, representing the different categories. The major types of qualitative response regressions are dichotomous, in which the dependent variable takes only two values, and polychotomous, in which the dependent variable takes more than two values. Due to the different types of regressions, I study the determinant factors of real estate investment in this paper by adopting the binomial logit and probit models, the multinomial logit and probit models, and the Tobit model.

By using three different econometric methodologies to study the behavior of individ- uals investing in residential real estate, I found that most people in the U.S. are willing to invest in a house. Individuals are less likely to invest a real estate as their education level increases. However, with the increase in earnings, the number of kids, and the age, people are more likely to invest in residential real estate.

2 RELATED LITERATURE

There is existing literature studying residential real estate investments. Brown et al. (2008) studied personal real estate investments in Australia using the logit model and data set of Australian households from 1990-2004. Özogul & Tasan-Kok (2020) studies and sum- marizes the existing literature on differentiating the investor types of residential property. By employing different types of econometric mythologies to analyze the longitudinal survey data of US individuals, this paper provides additional empirical evidence to the literature. The methodologies in this empirical study include binary choice models, multiple choice models, and models based on truncated and censored data. For example, Berman & Hericourt (2010) studies the financial factors and the margins of trade by using the probit model since the exporting decision is a discrete variable of value 0 or 1. Using the binary probit model of

Helpman et al. (2004), Oberhofer & Pfaffermayr (2012) studies how firms decide to serve foreign markets by exporting, foreign direct investment, or both. Falk (2008) applies the binary probit model and uses the maximum likelihood method to empirically study the relationship between innovation and foreign ownership. Baldwin & Yan (2011) also studies the effect of exchange rates and tariffs on the failure of plants by using the probit model. Also, Ai & Norton (2003) has studied the methodology empirically in estimating the interaction effect in nonlinear models.

The multiple choice models are applied when the dependent variable includes more than two options. Such as, Pietrovito & Pozzolo (2016) defined three indexes of foreign expansions (domestic only, exported only, and export and M&A) to study the internation- alization choices. The methodology of Pietrovito & Pozzolo (2016) is the ordered probit model. Engel et al. (2010) studies the determining factors of the firm's decision to enter and exit from the international market by adopting the multinomial probit model. Also, in studying how productivity heterogeneity affects the firms' behavior of internationalization, Wakasugi & Tanaka (2009) uses the multinomial logit model to analyze the firm-level data of Japan. By adopting the ordered probit model, Bown (2005) studies the determinants of behavior of a country participating in formal trade litigation under the World Trade Organization (WTO) between 1995 to 2000. Koru (2005) uses the multinomial logit,mul- tivariate probit, and binary logit models to study the determinant factors that variate the entry modes of firms serving foreign markets.

The Tobit model is a better option for truncated and censored data. Baldwin & Nino (2006) studies the currency effect of the Euro on trade by utilizing the Tobit model to estimate the overall usage of trade and the logit model to estimate the effect of the Euro on trade in products. To study the gravity model at the case of digital goods that are consumed online, Blum & Goldfarb (2006) estimates the number of international visits in the category by adopting the censored regression. In the study of the impact of the European Monetary Union on FDI flows, Schiavo (2007) has employed a censored regression that assumes a normal distribution to analyze the maximum amount of available information. Redding & Venables (2004) studies the impact of economic geography on cross-country variation in per capita income by using the Tobit model. In addition, McPherson et al. (2001) estimates the validity of the Linder hypothesis in East African developing countries by using the weighted maximum likelihood estimation on the fixed effect Tobit model.

3 DATA

In this paper, I use the Survey of Income and Program Participation (SIPP) data. SIPP is a household-based survey designed as a continuous series of national panels. Each panel features a nationally representative sample interviewed over a multi-year period last- ing approximately four years. SIPP is a source of data for a variety of topics and provides for the integration of information for separate topics to form a single, unified database. To study the determinant factors of real estate investment empirically, I have collected a total of 421,911 observations from the year 2008.

In the data set, *etenure* represents the status of the real estate of the individual, which is the dependent variable in this paper. In addition, 1 represent that the individual owns the real estate, 2 shows that the individual rent the real estate, and 3 represent that the individual lives in their parents' home. Also, *thearn* represents the total earning of the individual, *rfnkids* shows the number of kids the individual has, *tage* shows the individual's age, and *eeducate* represents the education level of the individual. The summaries of statistics are separately listed under three methodologies.

4 METHODOLOGIES AND EMPIRICAL RESULTS

4.1 Binary Response Models

Binary response models basically represent the models in which the dependent variable is binary. The most well-known binary response models are the logit and probit models.

Consider a binary response model of the form

$$P(y = 1|x) = G(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) \qquad (1)$$

In logit model, the function G is a logistic function, which follow a logitdistribution.

$$G(z) = \frac{e^z}{1+e^z} \qquad (2)$$

The function G is a standard normal cumulative distribution function in the probit model.

$$G(z) = \frac{1}{2\pi^{\frac{1}{2}}} e^{\frac{-z^2}{2}} \qquad (3)$$

Where

In this paper, the dependent variable is *etenure*, which is the status of the individual's real estate, where 0 represents the individual who does not own any real estate, and 1 represents that the individual owns one or more real estate. The explanatory variables are the total income of the individual, the number of kids, the individual's age, and the individual's education level. The education level includes the person who has notreceived any education, which is shown as −1. With the increase in education, the numbers will increase to 47. Therefore, the binomial logit model is

$$ln(\frac{P_i}{1-P_i}) = \beta_0 + \beta_1\, thearn_i + \beta_2\, rfnkid_i + \beta_3\, eeducate_i + \varepsilon_i. \quad (4)$$

The probability of the individual owning real estate is P_i, and the probability of the indi- vidual not owning real estate is $1-P_i$. The logistic function is shown as follows:

$$P_i = P(etenure = 1|x) = \frac{e^{\beta_0+\beta_1\, thearn_i+\beta_2\, rfnkid_i+\beta_3\, eeducate_i+\varepsilon_i}}{1+e^{\beta_0+\beta_1\, thearn_i+\beta_2\, rfnkid_i+\beta_3\, eeducate_i+\varepsilon_i}} \quad (5)$$

$$1-P_i = P(etenure = 0|x) = 1-\frac{e^{\beta_0+\beta_1\, thearn_i+\beta_2\, rfnkid_i+\beta_3\, eeducate_i+\varepsilon_i}}{1+e^{\beta_0+\beta_1\, thearn_i+\beta_2\, rfnkid_i+\beta_3\, eeducate_i+\varepsilon_i}} \quad (6)$$

To estimate how these independent variables affect an individual's likelihood of owning real estate. Table 1 shows the summary of statistics. The table shows that the 68.45% observations own real estate, the average age of the observations is 37, and the average education level of the observations is above 1st Grade.

Variable	Obs	Mean	Std.Dev.	Min	Max
etenure	421,911	0.685	0.465	0	1
thearn	421,911	5049	5946	-41176	96250
rfnkids	421,911	1.131	1.384	0	12
eeducate	421,911	31.65	17.08	-1	47

Table 1. Summary of Statistics

Table 2 shows the frequency and percentage of whether or not the total observations own real estate. It shows that 68.45% of the observations own at least one residential real estate.

HH: Ownership status of living	Quarters	Freq.	Percent	Cum.
Rent or living with parents	0	133,106	31.55	31.6
Owned or being bought by a house	1	288,805	68.45	100
Total		421,911	100	

Table 2. Frequency of Owning a House

To understand and estimate how the independent variables affect an individual's like-lihood of owning real estate, I estimate the binary response model by employing the OLS, logit, and probit estimations. The results of the three estimations are listed in the tables below.

Source	SS	df	MS	Number of obs	=	421,911
Model	5574	3	1858	Prob¿F	=	0
Residual	85540	421,907	0.203	R-squared	=	0.0612
Total	91113	421,910	0.216	Root MSE	=	0.450
etenure	Coef.	Std.Err.	t	P¿—t—	95% Conf.	Interval
thearn	1.81e-05	1.17e-07	154.3	0	1.79e-05	1.83e-05
rfnkids	-0.0176	0.000586	-30.06	0	-0.0188	-0.0165
eeducate	0.00162	4.73e-05	34.31	0	0.00153	0.00172
cons	0.562	0.00208	269.5	0	0.558	0.566

Table 3. OLS Estimation Results

Table 3 shows the ordinary least square estimation results. In the binary qualitative response model, the OLS estimation does not give the most correct results because the dependent variable of OLS regression is continuous and the binomial response regression is only coded as 1 and 0. Therefore, I interpret the logit and probit results and only report the OLS results for comparison.

Table 4 reports the logit estimation results. Compared with the OLS estimation, the signs of the coefficients are the same, but the actual values of the coefficients are different.

Iteration 0 log likelihood = -263027
Iteration 1 log likelihood = -247916
Iteration 2 log likelihood = -245507
Iteration 3 log likelihood = -245412
Iteration 4 log likelihood = -245412

Logistic regression — Number of obs = 421911; LR chi2(3) = 35230; Prob¿chi2 = 0; Log likelihood = -245412; Pseudo R2 = 0.0670

etenure	Coef.	Std.Err.	z	P¿—z—	95% Conf.	Interval
thearn	0.000147	9.79e-07	150.6	0	0.000145	0.000149
rfnkids	-0.0997	0.00287	-34.79	0	-0.105	-0.0941
eeducate	0.00760	0.000231	32.84	0	0.00715	0.00805
cons	0.0380	0.0103	3.680	0	0.0178	0.0581

Table 4. Logit Estimation Results

The result shows that people with higher earnings and higher education are likelier to own a house. However, the interesting results show that people are less likely to own a house with more kids.

The odds ratio is the ratio of the probability that the individual owns a house and the probability that the individual does not own a house. Table 5 shows the results. The odds ratio of earnings is 1, which means that the earning does not affect the individual willing to own a house as much as the individual's education level. However, with more kids, the individual is actually less likely to buy a house.

Logistic regression — Number of obs = 421911; LR chi2(3) = 35230; Prob>chi2 = 0; Log likelihood = -245412; Pseudo R2 = 0.0670

etenure	Odds Ratio	Std.Err.	z	P¿—z—	95% Conf.	Interval
thearn	1.000	9.79e-07	150.6	0	1.000	1.000
rfnkids	0.905	0.00259	-34.79	0	0.900	0.910
eeducate	1.008	0.000233	32.84	0	1.007	1.008
cons	1.038682	.010698	3.68	0.000	1.017925	1.059863

Table 5. Odds Ratio of Logistic Regression

Table 6 reports the marginal effect of the independent variables at means when all the other independent variables hold constant when the total earnings of the individual

Conditional marginal effects Model VCE: OIM — Number of obs = 421,911

Expression: Pr(etenure), predict()
dy/dx w.r.t.: thearn rfnkids eeducate
at : thearn = 5049 (mean)
rfnkids = 1.131 (mean)
eeducate = 31.65 (mean)

	dy/dx	Delta-method Std.Err.	z	P¿—z—	95% Conf.	Interval
thearn	3.02e-05	1.86e-07	161.7	0	2.98e-05	3.05e-05
rfnkids	-0.0204	0.000585	-34.84	0	-0.0215	-0.0193
eeducate	0.00156	4.73e-05	32.86	0	0.00146	0.00165

Table 6. Marginal Effect at Means

increase by one unit, the likelihood of the person willing to own a house will increase by 0.00302%. Same as the number of kids the person has, when the number of kids increases by one unit, the likelihood of the person being willing to own a house will decrease by 2.04%. When the level of education increases by one unit, the likelihood of the person being willing to invest in a house will increase by 0.156%.

Iteration 0 log likelihood = -263027	
Iteration 1 log likelihood = -254225	
Iteration 2 log likelihood = -254197	
Iteration 3 log likelihood = -254197	
Probit regression	Number of obs = 421911
	LR chi2(3) = 17661
	Prob¿chi2 = 0
Log likelihood = -254197	Pseudo R2 = 0.0336

etenure	**Coef.**	**Std.Err.**	**z**	**P¿—z—**	**95% Conf.**	**Interval**
rfnkids	0.0396	0.00177	22.36	0	0.0361	0.0431
eeducate	-0.00558	0.000166	-33.72	0	-0.00590	-0.00526
tage	0.0158	0.000134	117.8	0	0.0155	0.0161
cons	0.0496	0.00630	7.880	0	0.0373	0.0619

Table 7. Probit Estimation Results

Table 7 reports the probit estimation results, and Table 8 shows the marginal effect at means estimation results. Table 9 is the predicted probabilities of logistic regression

Conditional marginal effects						Number of obs = 421,911
Model VCE:	OIM					
Expression:	Pr(etenure), predict()					
dy/dx w.r.t.:	rfnkids eeducate tage					
at:	rfnkids = 1.131 (mean)					
	eeducate = 31.65 (mean)					
	tage = 36.86 (mean)					

	dy/dx	Delta-method Std.Err.	z	P¿—z—	95% Conf.	Interval
rfnkids	0.0139	0.000624	22.36	0	0.0127	0.0152
eeducate	-0.00196	5.82e-05	-33.74	0	-0.00208	-0.00185
tage	0.00556	4.70e-05	118.3	0	0.00547	0.00565

Table 8. Marginal Effect of Probit Model

Variable	Obs	Mean	Std.Dev.	Min	Max
plogit	421,911	0.685	0.127	0.00332	1.000
pprobit	421,911	0.684	0.0925	0.519	0.923

Table 9. Predicted Probabilities

and probit regression, where logit shows the mean value of logistic regression's predicted probabilities and probit shows the mean value of predicted probabilities of probit regression. From Table 9 results, easy to tell that the two types of regressions get very similar results regarding the predicted probabilities. The average probability of an individual owning a property is around 68%. It is also similar to the percentage of individuals who own real estate in the frequency of owning a house shown in Table 2. At the same time, Appendix 1 shows the marginal effect from the logistic regression at 25% quantile at means and at 75% quantile and Appendix 2 shows the marginal effect from the probit regression at 25% quantile at means and at 75% quantile.

4.2 Multiple Logit Model

The characteristic of multinomial response regressions is the dependent variable takes more than two values. In this section, I rearranged the data so that the dependent variable of the multinomial logit and probit models have three values. When the individual owns real estate, the value will be 1. When the individual rents a house, the value will be 2. When the individual lives at his/her parents' house the value will be 3. Table 10 summarizes the statistics of the data from the SIPP.

Other than the summary of statistics, which cannot describe the qualitative dependent variable well, Table 11 shows the frequency of owning a house. The Table shows that 68% of the interviewee own a house, 29% of the interviewee rent a house, and only less than 2% of the interviewee lives in their parents' house. This section aims to estimate how these independent variables affect the likelihood of an individual owning real estate. Therefore, following the same steps and estimation procedures from the binomial response regressions,

Variable	Obs	Mean	Std.Dev.	Min	Max
etenure	421,911	1.333	0.506	1	3
thearn	421,911	5049	5946	-41176	96250
rfnkids	421,911	1.131	1.384	0	12
tage	421,911	36.86	22.60	0	84
eeducate	421,911	31.65	17.08	-1	47

Table 10. Summary of Statistics

HH: Ownership status of living Quarters	Freq.	Percent	Cum.
Owned or being bought by a house	288,805	68.45	68.45
Rented	125,839	29.83	98.28
Occupied without payment of cash	7,267	1.72	100.00
Total	421,911	100.00	

Table 11. Frequncy of Owning a House

I empirically study the effect of the independent variables on investing a real estate by using the logit model.

The multinomial logistic regression model is as follows:

$$ln(\frac{P(y \leq j|x)}{P(y > j|x)}) = C_j - \beta_0 + \beta_1\, ethearn_i + \beta_2\, rfnkids_i + \beta_3\, tage_i + \beta_4\, eedcuate_i + \varepsilon_i \quad (7)$$

where c_j represents the different categories of the individual's living status. *ethearn* is the explanatory variable that interprets the individual's total income, *rfnkids* is the total number of kids the individual has, and *tage* is the age of the individual. The *eeducate* shows the education level of the interviewee. The response probabilities of the multinomial logit model is given by:

$$P(y = j|x) = \frac{e^{x\beta_j}}{1 + \sum_{j=1}^{J} e^{x\beta_j}}, \; j = 1, ..., J \quad (8)$$

where the total probabilities must sum up to one.

The estimation results of the multinomial logit model are listed below. First of all, for comparison, table 12 shows the estimation results of OLS. In this table, only the coefficient of education level shows positive.

Table 13 reports the estimation results of the multinomial logit model. I do not report the iterative procedure and only report the coefficients. The base outcome is the person lives at his/her parents' house, which is not the default setup. I will use it as the base outcome because I am more than willing to know how those independent variables affect the status of owning a house. As the table shows, the likelihood of renting a house will decrease with the increase in earnings, number of kids, and age. At the same time, with the increase in education level, the likelihood of owning a house will decrease.

Source	SS	df	MS		
Model	10950	4	2738	Prob > F =	0
Residual	97254	421,906	0.231	R-squared =	0.101
Total	108204	421,910	0.256	Root MSE =	0.480

Number of obs = 421,911

etenure	Coef.	Std.Err.	t	P >—t—	95% Conf.	Interval
thearn	-2.27e-05	1.27e-07	-178.4	0	-2.30e-05	-2.25e-05
rfnkids	-0.0114	0.000654	-17.47	0	-0.0127	-0.0102
tage	-0.00687	4.80e-05	-143.1	0	-0.00696	-0.00678
eeducate	0.00338	6.17e-05	54.77	0	0.00326	0.00350
cons	1.607	0.00243	660.8	0	1.602	1.611

Table 12. OLS estimation results

etenure	Coef.	Std.Err.	z	P> —z—	95% Conf.	Interval
Rented						
thearn	-0.000197	1.15e-06	-172.2	0	-0.000200	-0.000195
rfnkids	-0.0498	0.00317	-15.72	0	-0.0560	-0.0436
tage	-0.0413	0.000253	-163.2	0	-0.0418	-0.0408
eeducate	0.0222	0.000299	74.23	0	0.0216	0.0228
cons	0.805	0.0122	66.17	0	0.781	0.828
Owned						
thearn	0.000227	4.23E-06	53.51	0	0.000218	0.000235
rfnkids	0.123342	0.011721	10.52	0	0.10037	0.146314
tage	0.021616	0.000762	28.38	0	0.020123	0.023109
eeducate	-0.00844	0.00102	-8.28	0	-0.01044	-0.00644
cons	2.131069	0.041284	51.62	0	2.050155	2.211984

Table 13. Multinomial Logit Estimation Results

			Delta-method				
		dy/dx	Std. Err.	z	P¿z	[95% Conf.	Interval
thearn_predict							
	1	3.89E-05	1.99E-07	195.66	0	3.85E-05	3.93E-05
	2	-3.6E-05	1.93E-07	188.54	0	-3.7E-05	-3.6E-05
	3	-2.47E-06	4.50E-08	-54.74	0	-2.55E-06	-2.38E-06
rfnkids_predict							
	1	0.010497	0.000612	17.16	0	0.009298	0.011695
	2	-0.00895	0.000594	-15.07	0	-0.01011	-0.00778
	3	-0.00155	0.000163	-9.51	0	-0.00187	-0.00123
tage_predict							
	1	0.00786	4.69E-05	167.43	0	0.007768	0.007952
	2	-0.00771	4.59E-05	167.72	0	-0.0078	-0.00762
	3	-0.00016	1.05E-05	-14.8	0	-0.00018	-0.00013
eeducate_predict							
	1	-0.00419	5.72E-05	-73.27	0	-0.0043	-0.00408
	2	0.004153	5.56E-05	74.65	0	0.004044	0.004262
	3	3.88E-05	1.42E-05	2.73	0.006	0.000011	6.66E-05

Table 14. Marginal Effect

Unlike the binomial logit model, the multinomial logit model does not consider the case with odds ratios. Due to the coefficient only can give me the direction of the effect on real estate investing, therefore, to better interpret the results, need to estimate the marginal effects of each variable. Table 14 shows the marginal effect at means for the four independent variables. The table shows that with the earning of the individual increase by one unit the likelihood of owning a house will increase by 0.04% when others hold constant, and the likelihood of renting a house will decrease by less than 0.04%. The results show the same pattern with the number of kids and the individual's age. With the number of kids increasing by one unit, the likelihood of owning a house will increase by 1%. At the same time, with the increase of kids by one unit, the likelihood of renting a will decrease by 0.9%. The interesting result is, with the increase of the education level by one unit, the likelihood of owning a house will decrease by 0.4%, and the likelihood of other options will increase. Education level is the only one which has the total opposite effect with other independent variables. From the table, I also can see that the number of kids affects on the decision to invest a real estate more than other factors. The estimation results of the marginal effect at 25% quantile, at means, and at 75% quantile shows in Appendix 3. Here I will not interpret the detail since it is similar to the results at means.

Predicted probabilities also is a great way to learn empirical studies. The results show the predicted probabilities of each of the observations. However, I only report the predicted probabilities at means in this paper. Table 15 summarizes the predicted probabilities, where *pmlogit*1 represents the predicted probability that the person owns a house *pmlogit*2 shows the predicted probability of renting a house and *pmlogit*3 represents the probability of living

Variable	Obs	Mean	Std. Dev.	Min	Max
pmlogit1	421,911	0.6845164	0.1742311	0.0001945	1
pmlogit2	421,911	0.2982596	0.167755	5.31E-09	0.918437
pmlogit3	421,911	0.017224	0.009059	2.71E-11	0.167576

Table 15. Summary of Predicted Probabilities

at the parents' home. It is really similar to the frequency of the status of living shown in table 11, which means that the model is very good at predicting.

4.3 *Tobit Model*

In many empirical studies, situations in which the dependent variable should use censored or truncated data are very common. Censored regression models usually are adopted when the variable of interest is only observable at a certain level. Such as, when the data are equal to less than some number c, we only record the number c. There are a total of three types of censored data, which are left-censored, right-censored, and double- censored. Unlike the censored regression models, truncated regression models are usually employed when the observations' value below or above certain thresholds are automatically excluded from the sample.

In this paper, I also estimate the data by using the Tobit model with the censored data. I select to analyze censored data because, from the previous study, I notice that the binomial and multinomial response regressions give some different results. Therefore, I censored the situation that the individual lives at parents' house and rent a house as not owning real estate. Therefore, the dependent variable, *etenure*, is the right censored variable. Table 16 is the summary of statistics of the right censored data. As you can see, the situation of the third category is recorded as category 2, which is the category inthat the individual does not own real estate.

Variable	Obs	Mean	Std. Dev.	Min	Max
etenure	421,911	1.315484	0.464709	1	2
thearn	421,911	5049.081	5945.915	-41176	96250
rfnkids	421,911	1.13058	1.384335	0	12
tage	421,911	36.86436	22.59893	0	84
eeducate	421,911	31.6511	17.08447	-1	47

Table 16. Summary of Statistics of Censored Data

Table 17 shows the frequency of the status of the living house. Different from the previous study, the dependent variable here is right censored, aka upper censored. The percentage of interviewees who own real estate is 68.45%, and the percentage of interviewees not own real estate is 31.55%.

To empirically study the effect on the investment of real estate, the first thing is to build the Tobit model, which is given by:

$$y_i^* = \beta_0 + \beta_1\, ethearn_i + \beta_2\, rfnkids_i + \beta_3\, tage_i + \beta_4\, eedcuate_i + \varepsilon_i \qquad (9)$$

where yi^* is the latent variable. $Yi^* = yi$ if $yi < 2$ and $yi^* = 2$ if $yi >= 2$.

Table 18 reports the Tobit regression results. First, the number of upper censored observations is 133,106, which is a big amount. At the same time, the estimation results

HH: Ownership status of livingquarters	Freq.	Percent	Cum.
Owned or being bought by a house	288,805	68.45	68.45
Not Owned a house	133,106	31.55	100.00
Total	421,911	100.00	

Table 17. Frequency of the Status of a Living House with Censored Data

show that with the increase in earnings, number of kids, and age the individual is less likely to not own a house. However, with the increase in education levels, the individual ismore likely not to invest a real estate. Moreover, the Tobit regression results shown in Table18 have the same pattern as the multinomial logit regression models.

Due to the limitation of understanding the Tobit model by the coefficient, it is neces- sary to estimate the marginal effects of the Tobit regression model. The marginal effects can give more detailed information regarding percentage change when the independent variables changed. Table 19 reports the marginal effects of the Tobit model. It showsthat with the earning increase by one unit, the likelihood of not investing a real estate decrease by 0.002%. If the number of kids increases by one, the likelihood of people not willing to invest a real estate will decrease by 1.2%. With the age of the person increasing by one unit, the likelihood of the person not being willing to invest a real estate will decrease by 0.81%. But, with the increase of the education level, the likelihood of this person does not invest a house will increase by 0.4%. Table 20 reports the marginal effect at 25% quantile, means, 75% quantile.

Tobit regression	Number of obs =	421,911
	Uncensored =	288,805
Limits: lower = -inf	Left-censored =	0
upper = 2	Right-censored =	133,106
	LR chi2(4) =	48972.32
	Prob ¿ chi2 =	0
Log likelihood = -400541.42	Pseudo R2 =	0.0576

etenure	Coef.	Std. Err.	t	P¿t	[95% Conf.	Interval]
thearn	-.0000301	1.67E-07	-179.72	0	-3E-05	-3E-05
rfnkids	-.0146821	0.000878	-16.72	0	-0.0164	-0.01296
tage	-.0098478	6.45E-05	-152.74	0	-0.00997	-0.00972
eeducate	.0049247	8.27E-05	59.54	0	0.004763	0.005087
cons	1.804423	0.003308	545.53	0	1.79794	1.810906
var(e.etenure)	.3785074	0.001104			0.376349	0.380678

Table 18. Tobit Regression Results

Conditional marginal effects Number of obs =421,911 Model VCE
: OIM
Expression : E(etenure*etenure¡2), predict(ystar (.,2))
dy/dx w.r.t. : thearn rfnkids tage eeducate
at : thearn = 5049.081 (mean) rfnkids = 1.13058 (mean)
tage = 36.86436 (mean)
eeducate = 31.6511 (mean)

			Delta-method				
	dy/dx	Std	Err.	z	P¿z	[95% Conf.	Interval]
thearn	-2.5E-05		1.37E-07	-180.56	0	-2.5E-05	-2.5E-05
rfnkids	-0.01209		0.000723	-16.72	0	-0.01351	-0.01067
tage	-0.00811		5.29E-05	-153.37	0	-0.00821	-0.00801
eeducate	0.004055		6.81E-05	59.57	0	0.003922	0.004189

Table 19. Marginal Effects of Tobit Model

	25% quatile		Means		75% quatile	
	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
thearn	-.0000243	1.32e-07	-2.5E-05	1.37E-07	-.0000243	1.32e-07
rfnkids	-.0118831	.0007107	-0.01209	0.000723	-.0118831	.0007107
tage	-.0079703	.0000511	-0.00811	5.29E-05	-.0079703	.0000511
eeducate	.0039858	.0000668	0.004055	6.81E-05	.0039858	.0000668

Table 20. Marginal Effects of Tobit Model at 25% quatile, means, 75% quatile

5 CONCLUSION

This paper studies the effect of earnings, age, number of kids, and education level on the decision of investing a real estate. I have collected a total of 421,911 observations based on the interview of individuals from the SIPP to empirically study this topic. By employing the technique of binomial logit and probit models, multinomial logit model, and Tobit model on censored data, the results show that overall, the majority of the people are willing to invest in a house. With the education level increases people are less likely to invest a real estate, however, with the increase in earnings, the number of kids, and age people are more likely to invest a real estate.

Works Citation

Ai, C., Norton, E. C., *Interaction terms in logit and probit models*, Economics letters 80 (2003), 123–129.

Baldwin, J., Yan, B., *The death of Canadian manufacturing plants: heterogeneous responses to changes in tariffs and real exchange rates*, Review of World Economics 147 (2011), 131–167.

Baldwin, R. E., Nino, V. D., *Euros and zeros: The common currency effect on trade in new goods*, NBER Working Paper 12673 (2006).

Berman, N., H´ericourt, J., *Financial factors and the margins of trade: Evidence from cross-country firm-level data*, Journal of Development Economics 93 (2010), 206–217.

Blum, B. S., Goldfarb, A., *Does the internet defy the law of gravity?*, Journal of inter- national economics 70 (2006), 384–405.

Bown, C. P., *Participation in WTO dispute settlement: Complainants, interested par- ties, and free riders*, The World Bank Economic Review 19 (2005), 287–310.

Brown, R. M., Schwann, G., and Scott, C., *Personal residential real estate invest- ment in Australia: Investor characteristics and investment parameters* Real Estate Economics 36(1) (2008), 139–173.

Engel, D., Procher, V., *The Asymmetries of a Small World: Entry Into and Withdrawal From International Markets by French Firms*, Ruhr Economic Paper 192 (2010).

Falk, M., *Effects of foreign ownership on innovation activities: empirical evidence for twelve European countries*, National Institute Economic Review 204 (2008), 85–97.

Koru, A. T., *Is FDI indeed tariff-jumping? Firm level evidence*, Oregon State Univer- sity (2005).

McPherson, M. A., Redfearn, M. R., Tieslau, M. A., *International trade and developing countries: an empirical investigation of the Linder hypothesis*, Applied Economics 33 (2001), 649–657.

O¨ zogul, S., and Tasan-Kok, T. , *One and the same? A systematic literature review of residential property investor types* Journal of Planning Literature 35(4) (2020), 475–494.

Oberhofer, H., Pfaffermayr, M., *FDI versus exports: multiple host countries and em- pirical evidence*, The World Economy 35 (2012), 316–330.

Pietrovito, F., Pozzolo, A. F., Salvatici, L., *Internationalization choices: an ordered probit analysis at industry level*, Empirical Economics 50 (2016), 561–594.

Redding, S., Venables, A. J., *Economic geography and international inequality*, Journal of International Economics 62 (2004), 53–82.

Schiavo, S., *Common currencies and FDI flows*, Oxford Economic Papers 59 (2007), 536–560.

Wakasugi, R., Tanaka, A., *Firm Heterogeneity and the Choice of Internationalization Modes: Statistical Evidence from Japanese Firm-level Data*, RIETI Discussion Paper Series 24 (2009).

Appendices

Appendix 1

	25% Quatile		Means		75% Quatile	
	dy/dx	Std.Err.	dy/dx	Std.Err.	dy/dx	Std.Err.
thearn	.0000294	1.79e-07	0.0139	0.000624	.0000294	1.79e-07
rfnkids	-.0199225	.0005697	-0.00196	5.82e-05	-.0199225	.0005697
eeduacte	.0015189	.000046	0.00556	4.70e-05	.0015189	.000046

Appendix 2

	25% Quatile		Means		75% Quatile	
	dy/dx	Std.Err.	dy/dx	Std.Err.	dy/dx	Std.Err.
rfnkids	.0135923	.0006068	.0000294	1.79e-07	.0135923	.0006068
eeducate	-.0019144	.0000566	-.0199225	.0005697	-.0019144	.0000566
tage	.0054168	.000044	.0015189	.000046	.0054168	.000044

Table 21. Appendix 3

	25% quartile		Means		75% quatile	
	dy/dx	Std. Err.	dy/dx	Std. Err.	dy/dx	Std. Err.
thearn_predict						
1	.000037	1.83e-07	3.89E-05	1.99E-07	.000037	1.83e-07
2	-.0000344	1.83e-07	-3.6E-05	1.93E-07	-.0000344	1.83e-07
3	-2.59e-06	7.28e-08	-2.47E-06	4.50E-08	-2.59e-06	7.28e-08
rfnkids_predict						
1	.0100233	.0005807	0.010497	0.000612	.0100233	.0005807
2	-.0082545	.0005694	-0.00895	0.000594	-.0082545	.0005694
3	-.0017689	.000197	-0.00155	0.000163	-.0017689	.000197
tage_predict						
1	.0074505	.0000405	0.00786	4.69E-05	.0074505	.0000405
2	-.0073423	.0000403	-0.00771	4.59E-05	-.0073423	.0000403
3	-.0001081	.0000118	-0.00016	1.05E-05	-.0001081	.0000118
eeducate_predict						
1	-.0039706	.0000535	-0.00419	5.72E-05	-.0039706	.0000535
2	.003966	.0000525	0.004153	5.56E-05	.003966	.0000525
3	4.61e-06	.0000168	3.88E-05	1.42E-05	4.61e-06	.0000168