

HETEROSCEDASTIC PROBIT MODEL: AN APPLICATION OF HOME OWNERSHIP IN TURKEY

ÇAĞLAYAN, Ebru*
ÜN, Turgut

Abstract

This study examines the factors affecting the possibility of owning a house in Turkey using probit models. Moreover, it focuses on the heteroscedasticity problem that is ignored in most of the probit applications. The results obtained from the study indicate that besides the demographic characteristics of the head of the household, education, employment, and income are also influential factors in regard to home ownership. Besides these things, evidence indicates that the choice to live in either rural or urban areas has a significant impact on home ownership.

Keywords: probit, heteroscedastic probit, heteroscedasticity, home ownership

JEL Codes: C01, C25, C35, R21

1. Introduction

The decision to become a homeowner seems to be an important financial and social decision. Thus, the choice of individuals regarding whether to own a house or to rent one is considered as important consumer behavior (Silver, 1988). Most studies which analyze the factors affecting home ownership mainly focus on the characteristics of the Household Head. For example, Boehm and Schlottmann, 2004; Case et al., 2005; Campbell and Cocco, (2007), Aiyagari (1994), Constant et al. (2006), Berry (1980), Galor and Stark (1990), Goodman (1990), Deng et al.(2003), Haurin and Rosenthal (2004) among others. It is not unexpected that their studies determined that the characteristics of the Household Head such as gender, race, age, marital status, and educational background are influential re: the decision to buy a house. In addition to these characteristics, variables such as the size of the household and income can be important factors. The conducted studies indicated that income level has both a direct and indirect impact on home ownership. In studies examining home ownership status, qualitative preference models are commonly used. For example, Li (1977), Guris et al. (2011), Capeau et al. (2003) among others.

The aim of this study is to determine the factors affecting the probability of owning a house in Turkey. In order to achieve this aim, estimates were made with the probit model using data from the 2009 Household Budget Survey conducted by the Turkish Statistics Institute. Moreover, the study focuses on the heteroscedasticity problem, which may have a direct effect on the estimation results of the probit model.

* Ebru Çağlayan, Kyrgyzstan-Turkey Manas University, Department of Economics, Faculty of Economics and Administrative Sciences, Chyngyz Aytmatov Campus, Djal, Bishkek, Kyrgyzstan. (Marmara University, Department of Econometrics, Faculty of Economics and Administrative Sciences, Ressay Namik Ismail Sok. No.1, 34720, Bahçelievler, Istanbul, Turkey) e-mails: ecaglayan@marmara.edu.tr, ebru.caglayan@manas.edu.kg; Turgut Ün, Marmara University, Department of Econometrics, Faculty of Economics and Administrative Sciences, Istanbul, Turkey

While many studies use probit models, they ignore some of its assumptions. Heteroscedasticity, in particular, can cause problems such as incorrect standard errors, and biased and inconsistent parameters. For this reason, we examined the determinants of home ownership using the heteroscedastic probit model as well as the standard probit model. In the empirical literature, there have been some studies conducted which have applied the heteroscedastic probit models. For example, Alvarez and Brehm (1995, 1998), Busch and Reinhart (1999), Krutz (2005), Litchfield et al. (2011) among others.

This paper is organized as follows: The following section includes the introduction. Section 2 presents both the standard probit and heteroscedastic probit models. The data and variables are introduced in Section 3. Section 4 reports the estimation results. The final section presents the conclusion.

2. Heteroskedastic Probit Model

The binary probit model is based on the assumption that a latent variable y_i^* is linearly related to the observed X_i 's

$$y_i^* = X_i\beta + \varepsilon_i$$

Where X_i is the vector of values for the i th observation, β is a vector of parameters, and ε_i is the unobserved error. The relation between y_i^* and the observed binary variable y_i can be expressed as follows:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > \tau \\ 0 & \text{if } y_i^* \leq \tau \end{cases}$$

Where τ is a threshold parameter or cut point. The errors of y_i^* are assumed to follow a normal distributed, the binary probit model expressed as follows:

$$\Pr(y_i = 1) = \Phi(X_i\beta)$$

Where Φ is the cumulative normal distribution. In the probit model, the error term is assumed as a homoskedastic. Probability can be written as follows:

$$\Pr(y_i = 1) = \Phi\left(\frac{X_i\beta}{\sigma}\right)$$

When σ is a constant that equals 1, it is removed from equation to estimate β s. If the errors are heteroskedastic the parameters estimates will be biased in consistent and inefficient. In the context of heterogenous choice, σ is known or expected to vary systematically such that $\sigma = \sigma_i$ with $i=1, \dots, N$.

The heteroscedastic probit model proposed by Alvarez and Brehm (1995) can be expressed as follows:

$$\Pr(y_i = 1) = \Phi\left(\frac{X_i\beta}{\sigma_i}\right)$$

Where $\sigma_i = \exp(z_i\gamma)$ and z_i is a vector of covariates of the i th observation and γ is a vector of parameters to be estimated. If the γ 's are equal 0, the model is identified as a

probit model. The heteroskedastic probit model can be estimated by maximum likelihood like probit model. For heteroskedastic probit, loglikelihood expressed as follows:

$$\log L = \sum_i \left(y_i \log \Phi \left(\frac{X_i \beta}{\exp(z_i \gamma)} \right) - (1 - y_i) \log \left[1 - \Phi \left(\frac{X_i \beta}{\exp(z_i \gamma)} \right) \right] \right)$$

Interpretation of coefficients obtained from probit models is complex. Coefficients are interpreted with the marginal effects. Marginal effect in a standard probit model with respect to same X_k is

$$\frac{\partial \Pr(y_i = 1)}{\partial X_k} = \phi(\tilde{X}_i \tilde{\beta}) \tilde{\beta}_k$$

For heteroskedastic probit model marginal effect calculated as follows:

$$\frac{\partial \Pr(y_i = 1)}{\partial X_k} = \phi \left(\frac{X_i \beta}{\exp(z_i \gamma)} \right) \left(\frac{\tilde{\beta}_k}{\exp(z_i \gamma)} \right)$$

3. Data

Most of the explanatory variables used in the study show the characteristics of the Household Head. These variables are presented in Table 1.

Table 1: Description of Explanatory Variables

Variable	Short name	Description
Income	INCOME	Turkish Liras
Age of Household Head	AGE	year
Employment Status of Household Head	OCCU	OCCU1= if salaried employee 1, others=0 OCCU2= if casual employee 1, others=0 OCCU3= if employer 1, others=0 OCCU4= if self-employed, others=0 OCCU5= if unpaid family worker, others=0
Education	EDU	EDU1, if illiterate 1, others=0 EDU2, if elementary education 1, others=0 EDU3, if high-school 1, others=0 EDU4, if university, post-graduate, doctorate 1, others=0
Profession	PR	PR1= If lawmaker, senior manager or manager 1, others=0 PR2= If member of a profession 1, others=0 PR3=if member of a supportive profession 1, others=0 PR4=if employee working in office and customer services 1, others=0 PR5= if service and sales employee 1, others=0 PR6= if qualified agriculture, animal breeding, hunting, forestry and water products employee 1, others=0 PR7= artists and similar profession 1, others=0 PR8= if unqualified worker 1, others=0
Household Type	HT	HT1=1 if nuclear family, others=0 HT2=1 if extended family, others=0 HT3=1 if single adults, others=0
Marital Status	MAR	if married 1, otherwise (single)=0
Gender	GENDER	if male 1, otherwise (female)=0
Region	REGION	if urban 1, otherwise (rural)=0

In the study, models by which the probability of home ownership was estimated in Turkey using household survey data. The data used in the analysis was obtained from the Turkish Statistics Institute's 2009 Household Budget Survey from the period January 1, 2009 to December 31, 2009. The dependent variable is defined as follows:

OWNH=1, if the house is owner occupied, otherwise 0

The education variable considered as influencing the consumption pattern is defined with 4 dummy variables, i.e. illiterate, elementary education, high school education and higher education. The age variable is formed by calculating the average age groups of the Household Head. The marital status, gender, household type, and region are formed as dummy variables. The profession is defined by 8 dummy variables.

4. Empirical Findings

In this study, an attempt was made to determine the factors influencing the probability of home ownership using household data for the year 2009. For this purpose, both the probit and heteroscedastic probit models were evaluated. The results obtained from the probit models are presented in Table 2.

According to Table 2, all coefficients of the probit model are statistically significant and the expected sign. In the empirical studies probit models are often applied without testing the normality and homoscedastic assumptions. However, the maximum likelihood estimators of the probit model will be biased and inconsistent if the disturbances are either non-normal or heteroscedastic (Yatchew and Griliches, 1985; Davidson and MacKinnon, 2004). For these reasons, we focused on two properties of the econometric specification, which are called homoscedasticity and normality. We applied the Bera-Jarque-Lee (BJL, 1984) test to test normality¹. The result of the test showed that the normality was rejected. The LM test for homoscedasticity reported at the bottom of the regression output, rejected a model without heteroscedasticity.

As a result, the diagnostic test results revealed that the homoscedastic regression model marginally violates the normality assumption, and the null hypothesis of the homoscedasticity is decisively rejected by the data. Thus, we applied the heteroscedastic probit model to eliminate the biased and inconsistent parameters, and to correct the unequal variance with discrete outcomes. The likelihood ratio test reported, rejects a model without heteroscedasticity. The results of the heteroscedastic probit model are reported in the third and fourth columns of Table 2.

All coefficients are highly statistically significant and have the expected signs in the mean function of the heteroscedastic probit model. It is clear that some estimated coefficients of the heteroscedastic probit model are different from the probit model. All the variables are considered in the mean equation, and the suitable ones were estimated in the variance equation. When the variance function and mean function were compared, it was evident that there were some differences in the signs of coefficient. As well acknowledged, the coefficients cannot be interpreted directly in the heteroscedastic probit models. In such a case, the marginal effects can be computed as a nonlinear combination of the regression coefficient².

¹ See also Wilde (2008) for normality in probit model.

² See more details: Green (2003), Cornelißen (2005).

Table 2. Results of Probit and Heteroscedastic Probit Model

Dependent variable: OWNH			
variables	Probit model	Heteroscedastic probit model	
		Mean function	Variance function
INCOME	0.000011*** (0.0000012)	0.000022*** (0.000004)	0.000011*** (0.0000021)
GENDER	-0.1191** (0.0517)	-0.0775* (0.0470)	
EDU1	0.2373*** (0.0898)	0.4358*** (0.1041)	
EDU2	0.3155*** (0.0703)	0.4580*** (0.0940)	
EDU3	0.1908*** (0.0696)	0.3169*** (0.0822)	
PR1	-0.2165*** (0.0830)	-0.2221*** (0.0839)	
PR5	-0.2116** (0.0834)	-0.1433* (0.0762)	
PR7	-0.1527** (0.0692)	-0.1063* (0.0620)	
REGION	-0.0582*** (0.0470)	-0.6295*** (0.0780)	-0.1505** (0.0764)
OCCU1	-0.0941* (0.0549)	-0.1302** (0.0535)	
OCCU3	0.3325*** (0.1299)	0.2899** (0.1283)	
OCCU4	0.3065*** (0.0625)	0.2302*** (0.0611)	
OCCU5	0.3503*** (0.1231)	0.2993* (0.1176)	
MAR	0.2996*** (0.0811)	0.1991** (0.0937)	-0.2448*** (0.0958)
AGE	0.0410*** (0.0018)	0.0350*** (0.0038)	
HT1	-0.1879*** (0.0551)	-0.1587*** (0.0527)	
HT3	-0.2931*** (0.0907)	-0.1919** (0.0878)	
Constant	-1.6383*** (0.1430)	-1.6382*** (0.2667)	
Observations	5658	5658	
Log-likelihood	-2996.8685	-2967.119	
LR	1522.73***		
Pseudo-R2	0.2026		
Wald		139.49***	
Normality (BJL Test)	659651***		
Homoskedasticity (LM Test)	654,1197***		
Lnsigma2		59.50***	

(i) *, **, *** indicate significance at the level 10%, 5% and 1%, respectively.

(ii) Numbers in parentheses are standard errors.

Table 3. Marginal effects of Heteroscedastic Probit Model

Variables	dy/dx	variables	dy/dx
INCOME	0.000006*** (0.00000)	OCCU1	-0.05085*** (0.01966)
GENDER	-0.02969* (0.01807)	OCCU3	0.10337** (0.04088)
EDU1	0.15384*** (0.02825)	OCCU4	0.08547*** (0.02001)
EDU2	0.17467*** (0.02716)	OCCU5	0.10643*** (0.03587)
EDU3	0.11666*** (0.02479)	MAR	0.10440*** (0.0274)
PR1	-0.08901*** (0.03304)	AGE	0.01353*** (0.00072)
PR5	-0.05687** (0.03017)	HT1	-0.06002*** (0.01882)
PR7	-0.04187 (0.02423)*	HT2	-0.07670** (0.03645)
REGION	-0.19566*** (0.01449)		

- (i) *, **, *** indicate significance at the level 10%, 5% and 1%, respectively.
(ii) Numbers in parentheses are standard errors.

According to the results in Table 3, an increase of 1% in age causes an increase of 0.01% in the likelihood of home ownership when other variables were fixed. Similarly, it was found that any increase in income level also increases the likelihood of owning a house. In addition, while being married increases the likelihood of owning a house compared to being single, living in urban areas decreases the likelihood of owning a house compared to living in rural areas. Furthermore, this is also related to the fact that the cost of living is more expensive in urban areas, and thus, house prices are higher. The findings showed that it is more likely for individuals other than university graduates to own a house compared to graduates, and that it is less likely for nuclear family members and singles to own a house compared to extended family members. In addition, it was seen that being in full-time employment or being self-employed increased the probability of owning a house compared to being a casual worker, and being a salaried employee decreases the likelihood of being a homeowner.

5. Conclusion

Probit models are among the methods commonly used in empirical studies. However, probit models are often applied without testing the normality or/and the heteroscedasticity. This is rather problematic because the standard maximum likelihood estimators of the probit models are mostly biased and inconsistent if the disturbances are abnormal, or if the disturbances are heteroscedastic. For this reason, we examine the determinants of home ownership using the heteroscedastic probit model as well as the standard probit model. The heteroscedastic probit model is used to solve the problem of heteroscedasticity.

The results of the study indicated that variables such as employment status, profession, household type, and education have an impact on the likelihood of owning a house, along with demographic characteristics such as age, gender, income and marital status. In particular, the regional variable is a key variable that was reflected in this study. It was determined that living in urban areas decreases the likelihood of owning a house compared to living in rural areas.

This paper also uses some diagnostic tests for the probit model, and focuses on heteroscedasticity. Thus, it may provide researchers with a greater confidence in the empirical findings obtained when using the probit models.

References

- Aiyagari, S. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *Quarterly Journal of Economics*, 109, pp. 659-684.
- Alvarez, R. M., Brehm, J. (1995). American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit model of competing values, *American Journal of Political Science*, 39(4), 1055-1082.
- Alvarez, R. M., Brehm, J. (1998). Speaking In Two voices: American Equivocation About the Internal Revenue Service, *American Journal of Political Science*, 42(2), 418-452.
- Bera, A.K., Jarque, C.M., Lee, L.F. (1984). Testing the Normality Assumption in Limited Dependent Variable Models, *International Economic Review*, 25(3), 563-578.
- Berry, J. 1980. Acculturation as Varieties of Adaptation, in Padilla, A.M. (Ed.), *Acculturation: Theory, Models and Some New Findings*, Boulder, Westview, CO.: 9-25.
- Boehm, T.P. & Schlottman, A.M. (2004). The Dynamic of Race, Income, and Homeownership. *Journal of Urban Economics*, 55(1): 113-130.
- Busch, M.L., Reinhart, E. (1999). Industrial Location and Protection: The political and Economic Geography of US Nontariff Barriers, *American Journal of Political Science*, 43(4): 1028-1050.
- Campbell, J.Y. & Cocco, J.F. (2007). How Do House Prices Affect Consumption? Evidence From Micro Data. *Journal of Monetary Economics*, 54(3): 591-621.
- Capeau, B., Decoster, A., Vermeluen, F. (2003), Homeownership and the Life Cycle: an ordered Logit Approach, Center for Economic Studies, Discussion Paper Series, 03.08, <http://econ.kuleuven.be/ces/discussionpapers/default.htm>
- Case, K., Quigley, J. & Shiller, R. (2005). Comparing Wealth Effects: The Stock Market Versus The Housing Market. *Advances in Macroeconomics*, 5(1).<http://www.bepress.com/bejm/advances/vol5/iss1/art1>. (Accessed 14 February 2011).
- Constant, A., Gataullina, L.& Zimmermann, K.F. (2006). Gender, Ethnic Identity and Work, IZA Discussion Papers, 2420.
- Cornelißen, T. (2005). Standard Errors of marginal effects in the heteroscedastic probit model, Discussion paper, No.320, ISSN:0949-9962.

- Davidson, R., MacKinnon J.G. (2004). *Econometric Theory and Methods*, Oxford University Press, Oxford.
- Deng, Y., Ross, S.L., Wachter, S.M. (2003). Racial Differences in homeownership: the effect of residential location, 33(5): 517-556.
- Galor, O. & Stark, O. (1990). Migrants' Savings, The Probability of Return Migration and Migrants Performance. *International Economic Review*, 31(2): 463-467.
- Goodman, A.C. (1990). Demographics of Individual Housing Demand. *Regional Science and Urban Economics*, 20(1): 83-102.
- Green W., (2003) *Econometric Analysis*, Prentice Hall, New Jersey.
- Guris, S., Caglayan, E., Un, T. (2011), Estimating of Probability of homeownership in Rural and Urban Areas: Logit, Probit and Gompit Model, *European Journal of Social Sciences*, 21(3): 405-411.
- Haurin, D.R. & Stuart, S. Rosenthal. (2004). The Sustainability of Homeownership: Factors Affecting The Duration of Homeownership and Rental Spells. Washington, DC. US. Department of Housing and Urban Development. <http://www.huduser.org/publications/affhsg/homeownsustainability.html>. (Accessed 17 December 2010).
- Krutz, G.S. (2005). Issues and Institutions: Winnowing in the US Congress, *American Journal of Political Science*, 49(2): 436-477.
- Li, M. (1977). A Logit Model of Homeownership, *Econometrica*, 45(5), 1081-1097.
- Litchfield J., Reilly, B., Veneziani, M. (2011). An Analysis of Life Satisfaction in Albania: An Heteroscedastic Ordered Probit Model Approach, *Journal of Economic Behavior & Organization*, doi: 0.1016/j.jebo.2010.12.020
- Silver, S.D. (1988). Interdependencies in Social and Economic Decision Making: A conditional Logit model of the Joint Homeownership-Mobility Decision, *Journal of Consumer Research*, 15(2), 234-242.
- Wilde J. (2008). A Simple Representation of the Bera-Jarque-Lee Test for Probit Models, *Economics Letters*, 101(2): 119-121.
- Yatchew, A., Griliches, Z. (1985). Specification Error in Probit Models, *Review of Economics and Statistics*, 67(1): 134-139.