International Real Estate Society Conference 1999

Co-sponsors: Pacific Rim Real Estate Society (PRRES)
Asian Real Estate Society (AsRES)

Kuala Lumpur, 26-30 January 1999

ESTIMATING THE PROBABILITY OF A FIRST, SECOND OR THIRD HOUSE BUYER PURCHASING A PARTICULAR RESIDENTIAL PROPERTY

PETER ROSSINI

School of International Business,
Centre for Land Economics and Real Estate Research (CLEARER)
University of South Australia
Phone (61-8) 83020649, Fax (61-8) 83020512, E-mail peter.rossini@unisa.edu.au
Home Page http://www.unisanet.unisa.edu.au/staff/Homepage.asp?Name=peter.rossini

Keywords: Buyer Behavior, Purchaser Characteristics, Residential Property, Probit Modeling.

Abstract: The paper is based on research that seeks to estimate the probabilities that the purchaser of a given property will have certain characteristics. The use of ordered probit models for this use is explained. Ordered probit models are used with the number of houses previously owned as the dependent variable. This is modeled with an array of independent variables that are property attributes such as building area, condition and location. The aim is to establish a marketing tool to establish the probabilities of particular buyer groups, buying a specific property. The basis of this research is a survey of some 3000 house purchasers who purchased detached and semi detached houses in Adelaide over the period June 1994 to March 1997. Results are presented as a spreadsheet where particular property attributes are entered and the probability of specific buyer groups are calculated from the ordered probit models.

Introduction

This paper is based on research carried out over a two-year period involving a variety of residential purchaser characteristics and their links to property characteristics. This research is an attempt to quantify many of the commonly held beliefs in the residential property market in Adelaide.

Professionals in the real estate industry use informal interviews and data collection on a day to day basis to develop a set of anecdotal ideas about how markets work. Many of the ideas about how markets operate, the current state of the market and purchaser and vendor preferences are based on informal, unstructured, qualitative research. Sometimes this is supplemented with analysis of sales to infer market behavior.

Unfortunately there are few formal surveys of households (particularly purchasers) which can provide more robust information. Research that is carried out is normally commissioned through market research agency, provides only basic descriptive material and is not available in the public domain.

The most recent works involving South Australia were based on a study in 1991 (Stevens et al, 1992) considering housing and locational preferences and Rossini (1998) who provided only a basic descriptive assessment of current purchasers. Stevens work was based on a survey conducted by the Australian Bureau of Statistics and was based on a qualitative survey of approximately 3300 households in Adelaide. The research from Rossini was based on a survey of some 600 recent purchasers. However none of these studies include any robust modeling of purchaser behavior or preferences or the relationship between property attributes and purchaser characteristics.

In fact little research attempts to link the characteristics of the property with the characteristics of the final purchaser through modeling. The use of hedonic price functions to model the transaction price of properties to the property characteristics is well known and accepted. However there is little research which attempts to directly estimate the probabilities of a particular property being purchaser by a particular type of purchaser.

Henderson and Ioannides (1989) estimated probabilities through a model of joint tenure, length of stay and consumption levels based on social and economic characteristics. They found that wealthier and better-educated families were more mobile while age had a quadratic effect where the planned length of stay is minimized at about 55. In an Australian study Bourassa (1994) compares housing choice of the Australian-born population with those of 10 major immigrant groups. He estimates the probability of ownership as a function of the households permanent and transitory income, the household annual cost of ownership, the typical annual cost of renting and a vector of demographic variables. This was a cross sectional study. Gyourko and Linneman (1996) and (1997) also investigated the use of demographic and economic factors on household tenure choice. They use a similar logistic specification for their model based on cross sectional data. As in the work of Bourassa the dependent variable is expressed as a dichotomous variable for house ownership. In a similar study, Haurin et al (1994) examined the household tenure choice but considered the simultaneity of household formation and labor supply decisions. They considered the choice of living alone or with parents using a variety of social and economic indicators.

Oluwoye (1996) reported results from an empirical survey of households in Woolongong. A cluster sample was used to collect data from one hundred households through interviews. The study examined the relationship between house ownership, age and average weekly earnings. The analysis used a logit model to estimate the probabilities of home ownership based on age and income. A wide variety of other studies have used econometric techniques such as logit to model purchaser behavior as probability distributions (Iwarere, 1991, Goodman, 1988, Zorn, P.M. 1988, Li, M. 1977). These studies also use cross sectional or aggregate variables based on social, economic and demographic characteristics to examine purchaser preferences such as tenure choice.

The focus in this paper is to try to directly estimate the probabilities of a first, second or third homebuyer purchasing a property with particular physical characteristics. The reason for this research is simple. Modern marketing practice suggests that real estate agents should try to target purchasers quite accurately in order to save marketing resources while maximizing the opportunity of finding a suitable purchaser. Typically agents use anecdotal evidence to take an “educated guess” as to the most likely purchaser group. The aim of this research is to quantify some of these issues. The focus in this paper, is the question of the probability that the purchaser is a first, second or third homebuyer. The expectation is, that given a set of property characteristics including locational, site and building characteristics, an estimate of the probability that the purchaser of the property would be purchasing their first home, or a second or third home can be made.

These probabilities could then assist the property professional in their marketing strategies.

Methodology

The basis of this research was a survey of recent house purchasers conducted during 1996 and 1997. The properties were all listed as transactions of detached houses over the period from January 1995 to March 1997. Data concerning the sale was extracted from sales details of the Department of Environment and Natural Resources (DENR) using UPmarket Comparative Sales Software. A total of forty thousand nine hundred and twenty four sales were found to be probable market transactions of detached houses over the period.

The properties were selected as cluster samples of detached houses in Adelaide and South Australian regional centres. The clusters are slightly biased particularly in Adelaide where the inner and middle distance suburbs are over represented.

The survey was conducted through purchaser interviews and was designed to serve a variety of purposes but was short enough to ensure that a reasonable number of purchasers would respond. There were three sections. Sections one and two were completed by interview with the purchaser. Section one had four questions and related to internal characteristics of the properties that were not available from the DENR sales file. Section two contained twelve questions pertaining to purchaser characteristics, preferences and behavior. Section three was completed by observation of the property and the neighborhood. This section was completed even if the purchaser would not respond to an interview.

Data from the survey was merged with the corresponding sales transaction and valuation data from DENR to give a database that incorporated physical and locational details for the property as well as the results from the purchaser survey. Rossini (1998) presented some basic results, based on the six hundred and six completed surveys. The research for this paper is based on a smaller data set, as the modeling requires data for all variables used. In total 537 observations had data in all fields needed for the modeling. Also any survey where the property was purchased as an investment or speculative opportunity were not used.

The model is estimated using an ordered probit. The probit is preferred to the logit model because of the normal distribution characteristics of the data and because probabilities are easier to estimate. The dependent variable is the number of houses previously owned. Thus zero for first homebuyers, one for second homebuyers and two for third or more homebuyers. This variable fits to the requirements of the dependent variable for an ordered probit.

In the probit model we propose that the dependent variable, (number of houses owned before) can be estimated in the form

$$y^* = \alpha + \Sigma\beta_i x_i + \varepsilon$$

The x_is are an array of property attributes.

In this case the dependent variable has three categories 0,1 and 2. Rather than observe y* we observe

$$y=0 \quad \text{if } y^* \le \delta_0$$
$$y=1 \quad \text{if } \delta_0 < y^* \le \delta_1$$
$$y=2 \quad \text{if } \delta_1 < y^* \le \delta_2$$

In the process of observing α and the β_i's, through a maximum likelihood function, the unknown parameters δ_i are estimated directly from the model estimation using the LIMDEP computer package. In order to estimate the probabilities the disturbance term is assumed to be normally distributed.

Then

For y=0

$$\text{prob}[y^* = \alpha + \Sigma\beta_i x_i + \varepsilon \le \delta_0]$$

for y=1

$$\text{prob}[\delta_0 < y^* = \alpha + \Sigma\beta_i x_i + \varepsilon \le \delta_1]$$

and for y=2

$\text{prob}[\delta_1 < y^* = \alpha + \Sigma\beta_i x_i + \varepsilon \leq \delta_2]$

These probabilities can be calculated as the area under the normal distribution curve at the δ_i's. As will be shown later in the paper. An excellent summary of this procedure and the conceptual framework of how probabilities are estimated is available in Becker et al (1992). This methodology is followed in this paper.

Results

The model was estimated using a variety of independent variables that have proven in the past to be reliable in the estimation of the hedonic price function. While a wide range of satisfactory models were produced it was decided that a model incorporating a small number of the most logical variables would be used. In particular the quality and quantity of the house were included through Building area, Number of Bedrooms, Internal and External Condition. Location was included through a series of dummy variables for different regions. These locations are shown on the map in Figure 1. Details of all of the variables are shown in Table 1.

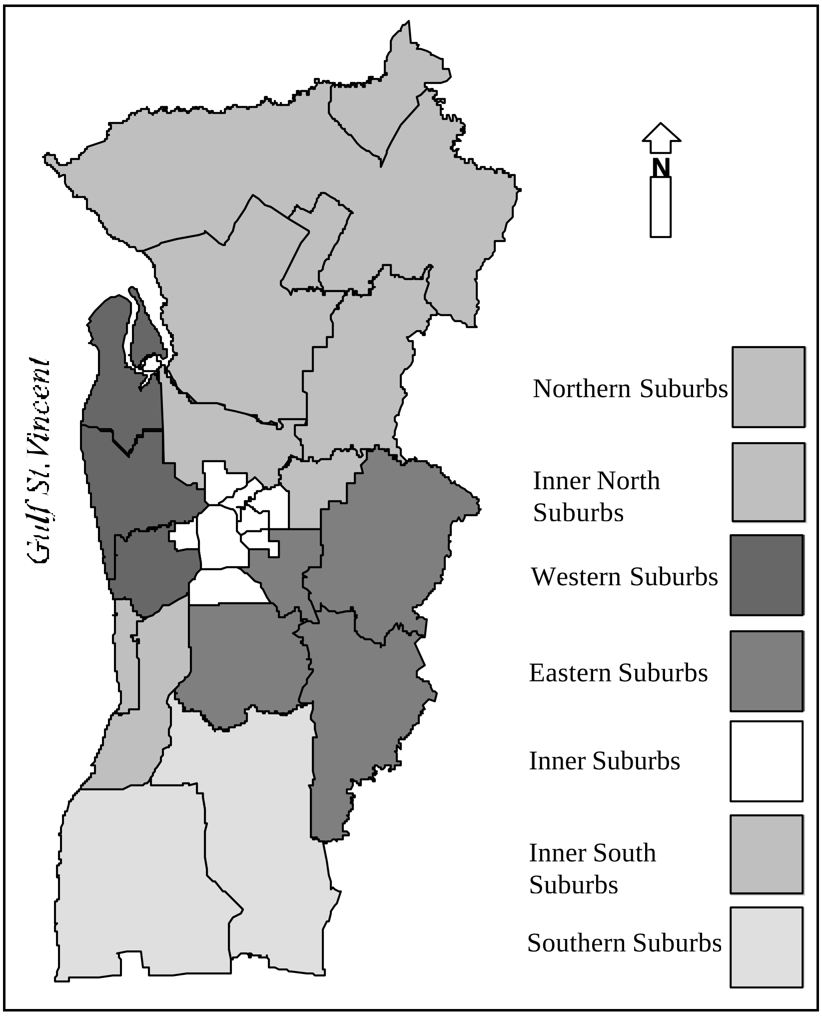

Figure 1 – Adelaide Regions

```
          +--------------------------------------------+
           | Ordered Probit Model                         |
           | Maximum Likelihood Estimates                 |
           | Dependent variable                OWNBE4     |
           | Weighting variable                   ONE     |
           | Number of observations               537     |
           | Iterations completed                  16     |
           | Log likelihood function       -530.0293     |
           | Restricted log likelihood     -581.3159     |
           | Chi-squared                    102.5732     |
           | Degrees of freedom                     9     |
           | Significance level             .0000000     |
           |    Cell frequencies for outcomes             |
           |  Y Count Freq  Y Count Freq  Y Count Freq   |
           |  0   135 .251  1   196 .364  2   206 .383   |
           +--------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Index function for probability
 Constant -2.465008656      .45772481       -5.385   .0000
 EQAREA    .5476274218E-02  .13002599E-02    4.212   .0000  154.05773
 EXCOND    .2328203578      .53521355E-01    4.350   .0000    7.08938
 EAST      .2978456478      .14756020        2.018   .0435     .21787
 INNER     .1895352331      .13605883        1.393   .1636     .22718
 INSOUTH   .4067164260      .27321208        1.489   .1366     .428540E-01
 NORTH    -.1743810106      .16226090       -1.075   .2825     .13407
 SOUTH    -.4473487971      .23878299       -1.873   .0610     .428340E-01
 INTCOND   .1478485090      .48805135E-01    3.029   .0025    2.41340
 BEDROOMS  .9688828683E-01  .91702831E-01    1.057   .2907    3.07262
          Threshold parameters for index
 Mu( 1)    1.094788341      .70236082E-01   15.587   .0000
Frequencies of actual & predicted outcomes
Predicted outcome has maximum probability.
            Predicted
------  ---------------  +  -----
Actual       0    1    2  |  Total
```

Table 1 – shows the probit output from LIMPED. The ordered probit is satisfactory but lacks a high degree of statistical significance for some variables. The Chi Squared test shows that there is good overall significance. The threshold parameter of the probability index (Mu) is calculated at 1.09478 and will be used to estimate the probabilities together with the parameter estimates (β's) for each independent variable and the constant (α). The model is reasonably accurate at predicting the original observations particularly in the case of first and third homebuyers. Of 135 first home buyers, 46 were correctly predicted as first homebuyers while only 13 were predicted as third homebuyers. For third homebuyers, some 109 of the 206 cases are correctly predicted. Only 10 of the 206 third homebuyers were predicted as probable first homebuyers based on the property purchased. Clearly there is a significant difference between the properties purchased by first and third homebuyers, however the distinction is blurred for second homebuyers.

Table 2 - Variables Using in Analysis

LIMDEP	Variable	Description
Eqarea	Equivalent Building Area (sqm)	Calculated Equivalent Area of Buildings based on weighted average formula for main buildings and other buildings
ExCond	External Condition of House	Scaled code from 1 - demolition to 9 - high quality new condition
InCond	Internal Condition of House	Scaled code from 1 – as new to 6 - needs full renovation
East	Located in Eastern Region	Dummy Variable
Inner	Located in Inner Region	Dummy Variable
InSouth	Located in Inner Southern Region	Dummy Variable
North	Located in Northern Region	Dummy Variable
South	Located in Southern Region	Dummy Variable
West	Located in Western Region	Dummy Variable
Bedrooms	Number of Bedrooms	Number of Bedrooms in the Main Building
OwnBe4	Number of Houses owned before	Discrete variable counting number of houses previously owned 0, 1, 2 or more

This output is not terribly useful unless it is converted to probability estimates since the purpose for this research is to estimate the probability of a first, second or third homebuyer purchasing a particular property.

In order to calculate probability estimates, it is necessary to input values for the independent variables into the equations and then calculate the probabilities by estimating the area under the curve for the normal distribution of outcomes for the latent variable y*.

This procedure is demonstrated in Table 3 for a typical property in the Western region of Adelaide. In this model the dummy variables for location mean that it is only reasonable to estimate the probabilities associated with a property of a particular type, for a specific location. So that while it is possible to use mean responses for the qualitative and quantitative building variables, it is clearly only possible for the property to be located in one location and therefore to "average" the locations becomes meaningless. The example in Table 3 is the estimate of the probabilities for a property in the Western (or inner northern) region, of average size and number of bedrooms and of average internal and external condition. The estimate is shown graphically in Figure 2.

Table 3 - Probability Estimtates for "Average Property" in the Western Region

Variables	Coefficients	Averages	West	Coef * West
Constant	-2.46500866		1	-2.465008656
EQAREA	0.00547627	154.06	154.06	0.843662375
EXCOND	0.23282036	7.09	7.09	1.650553269
INTCOND	0.14784851	2.41	2.41	0.356818745
EAST	0.29784565	0.22	0	0
INNER	0.18953523	0.23	0	0
INSOUTH	0.40671643	0.04	0	0
NORTH	-0.17438101	0.13	0	0
SOUTH	-0.44734880	0.04	0	0
BEDROOMS	0.09688829	3.07	3.07	0.29770144
Mu(1)	1.094788341			**0.683727173**

Estimate of latent variable Y* for a Property in the West with mean estimates of Area, Condition, and # of Bedrooms

Estimate for West

Estimates of Area under the Curve

Estimate of Z at Y*=0	-0.683727173	0.247073658		
Estimate of Z at Y*=Mu=1.09	0.411061168	0.659486118	0.412412	0.340514

Probability that house purchaser Bought **no** Homes before	**0.24707366**
Probability that house purchaser Bought **1** Homes before	**0.41241246**
Probability that house purchaser Bought **2** Homes before	**0.34051388**

Figure 2 shows the following situation. The ordered probit model is shown as the line y=-1.715 +.0055 (EQArea) with variation other than the building area being absorbed into the constant. (This is only to simplify the presentation). The normal distribution curve sits at the point estimate for the particular property and corresponds to the estimate of y* of .68 estimated in Table 3. The boundaries of the latent variable y* which correspond to the original categories are at δ_0=0.0 and at δ_1=1.09. The values associated with these points are estimates of the Z scores at these points. Z tables can then be used to calculate the area under the curve of the normal distribution, the probability value. These are shown under the normal curve.

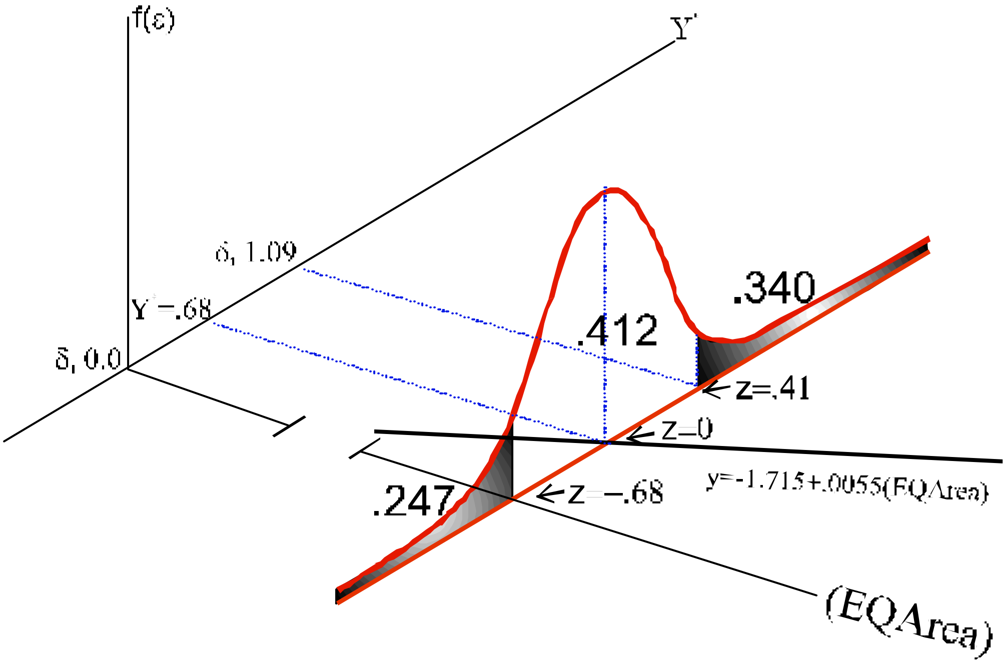

Figure 2 -Graphical Representation of the Ordered Probit Model – Based on Becker et al (1992)

In the example for an average property in the western region; there is a probability of 247 it will be purchased by a first home buyer, a probability of .412 it would be purchased by a second home buyer and the probability is it purchased by at least a third home buyer is .340. Considering different qualitative and quantitative values for the properties in the Western or Inner Northern region can extend this example. In Table 4, probabilities are estimated for situations in the western and inner northern areas at median and upper and lower quartile levels for the qualitative and quantitative variables. This highlights the effect of housing size and condition in the probable purchaser. For the small, poor quality property, there is a probability of .48 that it would be purchased by a first home buyer while the probability of it being purchased by a third home buyer is only about .15. For the larger better condition property, the probability that it is purchased by a first homebuyer is estimated at only .09 while the probability of a third homebuyer is just over .6.

Table 4 – Example of Probability Differences Due to Quantitative and Qualitative Issues – Holding Location Constant

Location	West or Inner North	West or Inner North	West or Inner North
Building Area	Small Quartile	Median	Large Quartile
External Condition	Poor Quartile	Median	Good Quartile
Internal Condition	Poor	Okay	Good
Number of Bedrooms	2 Bedrooms	3 Bedrooms	4 Bedrooms
Houses Purchased Before	**Probabilities**		
0	0.479599	0.252826	0.086563
1	0.372072	0.413273	0.308002
2	0.148328	0.333901	0.605436

Holding the quantitative and qualitative issues constant and considering different locations can also extend the earlier example. This is shown in Table 5.

Table 5 – Example of Probabilities in Different Locations – Holding Quantitative and Qualitative Issues Constant

Based Median Area (140sqm), 3 bedrooms, Median Internal and External Quality

	Homes Purchased Before		
Location	**0**	**1**	**2**
Inner	0.1962	0.3985	0.4053
East	0.1677	0.3846	0.4478
Inner South	0.1418	0.3672	0.4910
North	0.3116	0.4153	0.2731
South	0.4136	0.3960	0.1904
West-Inner North	0.2528	0.4133	0.3339

In all locations the probability that the property is purchased by a second homebuyer is around .4. However for a “typical house” there is a much greater probability of it being purchased by a purchaser who as purchased two houses previously in the Inner, Eastern and Inner Southern regions. Correspondingly there is a much greater probability of the property being purchased by a first homebuyer in the northern or southern regions.

Clearly estimates can easily be made for any particular combination of location, area, number of bedrooms, internal and external condition. To facilitate this a spreadsheet is useful. An example is shown as attachment 1.

Conclusion

The purpose of this paper is to quantify the probabilities that a first, second or third homebuyer is the purchaser of a particular property. Typically real estate agents make these assessments based on anecdotal evidence and assume certainly with their estimates. This paper has shown that it is possible to make realistic estimates of those probabilities. The results can be used to estimate the probability that the purchaser of a particular house in Adelaide would have previously purchased zero, one or two houses. This can then be used to effect marketing strategies. This research will be extended to include a wider variety of dependent variables. In particular the family and age structure of purchasers. This research is ongoing. Results from the 1997-98 survey will be incorporated with those used in this paper to extend the findings. If data can be collected for a significant period of time it would also be possible to test if the patterns have changed as trends and property markets change.

Acknowledgment

This paper would not have been possible without the kind assistance and guidance of Professor William Becker, Indiana University.

References

Becker, W.E. and Kennedy, P.E. (1992) A Graphical Exposition of the Ordered Probit, *Econometric Theory, 8, 1992, 127-131*

Bourassa, S.C. (1994) Immigration and Housing Tenure Choice in Australia, *Journal of Housing Research, Volume 5, Issue 1, 1994*

Eldred, G and Zerbst, R. (1976) Consumer Research and the Real Estate Appraiser, *The Appraisal Journal Vol. 44 510-522*

Goodman, A.C. (1988) An Econometric Model of Housing Price, Permanent Income, Tenure Choice and Housing Demand. *Journal of Urban Economics* 23, 327-353

Gyourko, J & Linneman, P (1996) Analysis of the Changing Influences on Traditional Households' Ownership Patterns*, Journal of Urban Economics, 39, 318-341, 1996*

Gyourko, J & Linneman, P (1997) The Changing Influences of Education, Income, Family Structure, and Race on Homeownership by Age over Time, *Journal of Housing Research, Volume 8, Issue 1, 1996*

Haurin, D.R., Hendershott, P.H. and Kim, D. (1994) Housing Decisions of American Youth, *Journal of Urban Economics, 35, 28-45, 1994*

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

Iwarere, L.J. Williams, J.E. (1991) A Micro-Market Analysis of Tenure Choice Using The Logit Model, *The Journal of Real Estate Research,* Fall 1991, Vol. 6 No 3 p327-339

Li, M. (1977) A Logit Model of Homeownership. *Econometrica* 45(5) 1081-1097

Oluwoye, J, (1996) An Analysis of Household Tenure Choice And Intra-State Movements: An Empirical Study of Wollongong, *Australian Land Economics Review* Volume 2 No.1. 20-22

Oluwoye, J. and Higgins, D. (1998) The Residential Property Price – On Overview, *The Valuer and Land Economist, May 1998, 174 - 178*

Rossini, P.A. (1998) Assessing Buyer Search Behavior for Residential House Purchasers in Adelaide, *4th Pacific Rim Real Estate Society Conference, Perth 1998*

Stevens, C.A., Baum, S., and Hassan, R. (1992), The Housing and Location Preferences of Adelaide Residents, *Urban Policy and Research*, Volume 10, Number 3, pp 7-22.

Wulff M (1993) An Overview of Australian Housing and Locational Preference Studies: Choices & Constraints on the Housing Market

Zorn, P.M. (1988) An Analysis of Household Mobility and Tenure Choice: An Empirical Study of Korea, *Journal of Urban Economics*, 1988 - 113-128

Attachment 1

Variables	Coefficients	Averages	WEST	Coef * WEST
Constant	-2.46500866		1	-2.465008656
EQAREA	0.00547627	154.06	180.00	0.985729359
EXCOND	0.23282036	7.09	8.00	1.862562862
INTCOND	0.14784851	2.41	4.00	0.591394036
EAST	0.29784565	0.22	0	0
INNER	0.18953523	0.23	0	0
INSOUTH	0.40671643	0.04	0	0
NORTH	-0.17438101	0.13	0	0
SOUTH	-0.44734880	0.04	0	0
BEDROOMS	0.09688829	3.07	4.00	0.387553147
Mu(1)	1.094788341			**1.362230749**

Select A Location	**WEST**
Select Eqarea Quartile	**Large Quartile**
Select Ex Cond Quartile	**Good Quartile**
Select Int Cond Code	**Good**
Select # of Bedrooms	**4 Beds**

Select variable settings from the drop downs above

Estimate of Z at Y*=0	-1.3622307	0.0865626		
Estimate of Z at Y*=Mu=1.09	-0.2674424	0.3945643	0.3080017	0.6054357

Probability that house purchaser Bought **no** Homes before	**0.086563**
Probability that house purchaser Bought **1** Homes before	**0.308002**
Probability that house purchaser Bought **2** Homes before	**0.605436**