ENDOGENEITY IN RESIDENTIAL LOCATION CHOICE MODELS

Cristian Angelo Guevara
Subsecretaria de Transportes. SECTRA. Teatinos 950 piso 16. crguevar@hotmail.com

ABSTRACT

Commonly used empirical residential location choice models have reported dwelling-unit price estimated parameters that are small, not statistically significant, or even positive. This would imply that households are non-sensitive to changes in dwelling unit prices or location taxes, which is not only against intuition, but also makes the models useless for policy analysis.

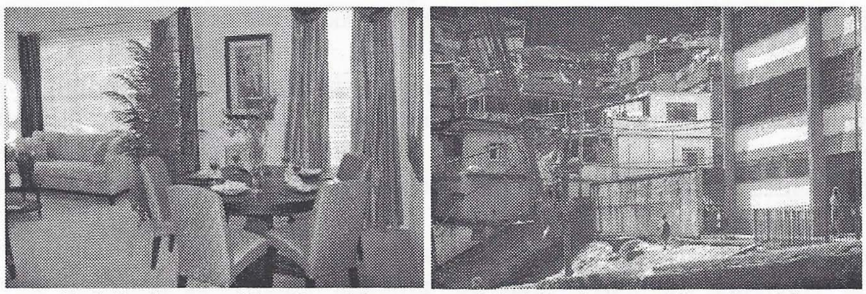
One explanation for this result is price endogeneity, which means that dwelling unit's price is correlated with the error term in the econometric model. This problem is caused either by the simultaneous determination of the supply and the demand for dwelling units in aggregated models, or by omitted attributes that are correlated with the price, in disaggregate models.

The treatment of endogeneity in discrete choice models is an area of ongoing research in econometrics. Therefore, methods to treat this problem began to be proposed only in the last decade, and have not been thoroughly analyzed for residential location models. The control-function method appears as a promising approach to address endogeneity in this framework for three reasons: 1) because it is the best method available to handle individual level endogeneity; 2) it is tractable in large scale problems with available estimation software and; 3) It can be adapted to handle more complex error structures using non-parametric methods.

This paper discusses the control-function method and tested it in terms of its robustness to different error structures using two Monte Carlo experiments. The method is also applied to an example based on real data from Santiago de Chile. The results show that the problem of price endogeneity does exist in residential location choice models, and also that the control-function method gives a satisfactory answer to the problem. However, the application of correction methods to address more general error structures gave unexpectedly poor results, which makes its study a relevant venue for further research.

1. INTRODUCTION

An econometric issue of relevance can arise in discrete choice models of residential location. Consider, for example, the problem described in Figure 1. In this case, the decision maker can perceive all the attributes of the alternatives in his or her choice set, including the non-measurable one that is explained through the photographs. However, the researcher can observe only the number of bedrooms, the size, the price and whether utilities are provided or not for each alternative. From the viewpoint of the researcher, individuals who decide to live in apartment A would be not very smart because they reject such a bargain as B or, more formally, just not sensitive to price. However, what it is really occurring is that a relevant quality attribute (in this example a neighborhood characteristic, but more generally any specific attribute), which is correlated with the price, is being omitted. Research shows that this problem implies that an upward bias for the price parameter (a negative parameter that becomes more positive) will usually occur in residential location models.



A) 1 Bedroom; 70 m2; with utilities; $2000 B) 1 Bedroom; 70 m2; with utilities; $100

Figure 1: Example of an Omitted Quality Attribute

Some empirical residential location choice models (Bhat and Guo, 2004; Sermonss and Koppelman, 2001; Levine, 1998; Waddell, 1992, 1996; Quigley 1976) have reported non-significant, small or even positive dwelling unit price parameters. The hypothesis of this paper is that this is caused by endogeneity due to omitted attributes that are correlated with the dwelling unit's price. This hypothesis has not been considered in many applications related to residential location. Furthermore, endogeneity in discrete choice models is an area of current development in econometrics, which makes the study of its effects in residential location modeling a challenging and interesting issue to address.

Section two of this paper reviews the causes, effects and the treatment of endogeneity in econometric models. Section three presents the evaluation of the control-function method, the most promising to treat endogeneity in discrete choice models of residential location, through a set of Monte Carlo experiments. Section four corresponds to the application of the control-method to an example with real data from Santiago de Chile. The final section corresponds to conclusions and recommendations for further research.

1 ENDOGENEITY IN ECONOMETRIC MODELS

Endogeneity or non-orthogonality occurs when the independent variable is correlated with the error term in an econometric model. Under this problem, the estimated parameters of a linear or a non-linear model are biased and inconsistent.

In **linear models**, endogeneity can occur by the omission of relevant attributes that are correlated with the observed ones. The way econometric theory has found to work out this problem (Greene, 2003) in linear models is called the instrumental variables (*IV*) method. It corresponds to project the dependent variable that is correlated with the error term onto a space that is orthogonal to the error's space, which is defined by another variable called instrument. Thus, to apply the method is necessary to find a variable Z which is, at the same time, correlated with the contaminated dependent variable and non-correlated with the error term.

The selection of appropriate instruments *Z* to apply the *IV* method is a relevant and controversial issue itself. Hausman (1997) and Nevo (2001) proposed to use as instruments for price observed average prices of the same product in other zones. On the other hand Berry, Levinsohn and Pakes (BLP, 1995), and Bresnaham (1997) argued that it would be better to use other attributes of the same products in other zones as instruments for price. Yet, it is not clear why attributes from other zones should be correlated with the prices in the analysis-zone's, the usage of such instruments would possibly lead to a weak instruments problem (Hahn and Hausman, 2003). Thus, this research adheres to the usage of Hausman type of instruments.

In **discrete choice models**, endogeneity is expected to occur as the result of the omission of some attributes that are correlated with the price but not observed by the researcher. It can be shown (Guevara, 2005) that an upward bias for the price parameter (a negative parameter that becomes more positive) will usually occur in the case of residential location. To address this research flaw, several methods have been proposed during the last decade, starting with the seminal work of BLP (1995), who proposed a product market fix effects procedure to solve market level endogeneity, up to the novel approach of Matzkin (2004) based on the usage of unobservable instruments. Of these methods, the most promising one to apply to discrete choice models of residential location corresponds to the control-function method (Heckman, 1978 and Hausman, 1978) applied on a discrete choice environment (Petrin and Train, 2004; Blundell and Powell, 2004).

The advantages of the control-function method under this approach are the following. First, it is able of handling individual level endogeneity, phenomenon that is common in this market because dwelling-units are almost unique and even the type of omitted attributes will differ from one to another. Second, it is tractable with available software, even for problems of the size usually implied in residential location. Furthermore, even in cases where the real error structure contradicts the assumptions of this method, Petrin and Train (2004) argue that non-parametric procedures can be used to overcome this problem.

The control-function method can be seen as a generalization of the *IV* method for linear models where, instead of using the projected dependent variable in the model, the fitted error is added as a new variable that controls-out the endogeneity problem. This can be shown as follows. Consider the utility function (1) U_{jn} that is a function of the price p_{jn}, some attributes x_{jn} and characteristics s_n. Consider also a set of appropriate instruments Z which are correlated with the price but not with the error term e_{jn}.

$$U_{jn} = V(p_{jn}, x_{jn}, s_n) + e_{jn} \tag{1}$$

The price can always be written as the sum its expectation conditional on the instruments Z and an error term μ_{jn}.

$$p_{jn} = E[p_{jn}/Z] + \mu_{jn} \tag{2}$$

Then, if expression (2) is estimated by OLS, the fitted prices $\hat{p}_{jn}^{OLS}$ will not be correlated with the unobserved part of the utility (1), e_{jn}. This follows from the fact that by doing OLS the fitted prices correspond to the projection of actual prices onto the space formed by the instruments Z, which are, by assumption, orthogonal to e_{jn}. For the same reason, the fitted errors of (2) $\hat{\mu}_{jn}^{OLS}$ will be orthogonal to the fitted prices and, because they correspond to the difference between fitted and observed prices, they will contain the part of actual prices that is correlated with the error e_{jn}. Therefore, if $\hat{\mu}$ are used to estimate the conditional expectation of e_{jn} (3), the fitted residual of this model $\hat{\varepsilon}_{jn}$ will be orthogonal to $\hat{\mu}$ and, in consequence, uncorrelated to the prices.

$$e_{jn} = E[e_{jn}/\hat{\mu}] + \varepsilon_{jn} = f_{jn}(\hat{\mu}) + \hat{\varepsilon}_{jn} \tag{3}$$

Thus, if the control-function $f_{jn}(\hat{\mu})$, which is the OLS estimation of the expectation of e_{jn} conditional on $\hat{\mu}$, is considered as an additional variable that enters the utility function linearly(1), the endogeneity problem would be solved. Petrin and Train (2004) stated that if it is assumed that the covariance matrices of e and $\hat{\mu}$ are diagonal, the control-function that enters the utility is just proportional to the price residual of the respective alternative and individual, or the "own error"

$$f_{jn}(\hat{\mu}) = E[e_{jn}/\hat{\mu}] = \lambda\hat{\mu}_{jn} \tag{4}$$

where the parameter λ is a function of the covariance of e and $\hat{\mu}$ and the variance of $\hat{\mu}$. An equivalent procedure could be pursued by estimating steps one and two simultaneously using a latent-variable approach (Walker, 2001). This would lead to an increase in efficiency. The study of this alternative method is left for further research.

2. EVALUATION OF THE CONTROL-FUNCTION METHOD USING MONTE CARLO EXPERIMENTS

The objective of this section is to analyze the robustness of the control-function method to treat endogeneity in residential location choice modeling through two Monte Carlo experiments that correspond to individual and group level endogeneity respectively.

Both experiments consider 2000 households that choose between three location alternatives based on the Random Utility Model (Ben-Akiva and Lerman, 1985). Each household (n) maximizes its utility (U_{in}), which is assumed to be a linear function (V_{in}) of the attributes (a, b, c, d and the price p) of each available dwelling unit alternative (i), with specific parameters and an error term (e_{in}).

$$V_{in} = 10a_{in} + 10b_{in} + 10c_{in} + 10d_{in} - 10p_{in} + e_{in} \tag{5}$$

The error term is distributed *iid* Extreme Value (0,1). This assumption produces a closed form for the probability that household n chooses alternative i (P_{in}) known as the Multinomial Logit model (MNL). Additionally, it was assumed that dwelling-units' prices are completely determined by the linear function ( ) of attributes c, d and z, and an error term μ_{in} that is assumed to be distributed Normal (0,σ).

$$p_{in} = 0.5c_{in} + 0.5d_{in} + 0.5z_{in} + \mu_{in} \tag{6}$$

The goal underlying this paper is to find out, for example, what occurs if the researcher does not observe an attribute, such as ***d*****, that is relevant in the decision making process and is correlated with the price**.

Two Monte Carlo experiments were developed. In the first experiment the error term of the price equation δ_{in} was constructed to be *iid* (over alternatives and households) Normal (0, 0.1). Variables a, b, c, d and the instrument z, were *iid* Uniform (0,1) generated for each household and alternative. Attribute p was generated, using the price equation (6), as a function of c, d and the exogenous instrument z. Within this setting, variables c and d are correlated with the price p but not a, nor b.

The second Monte Carlo experiment corresponds to a case where Petrin and Train (2004) state that the simplest form of the control-function method (4), is biased. The setting in this experiment is equal to the first one in terms of the true parameters considered, but differs in that variable d was defined as varying only between 40 groups of 50 household. Variables b, c and the error term δ_{in} were constructed half as varying by zone and half individually. Variable a was constructed as varying only individually.

Four MNL models were estimated with the simulated data: A) a model where all variables (a, b, c, d, p) were included; B) a model where the variable excluded (a) is not correlated with the price; C) a model where the variable excluded (d) is correlated with the price; and D) the application of the control-function method omitting d.

The parameters of the four estimated models are shown by column in Table 1, together with the direct price elasticity of each of the three alternatives, e_{11}, e_{22}, e_{33}, evaluated at the sample mean of each attribute. The first column of this table corresponds to the labels. The second one corresponds to the true values of the parameters to be estimated taken from (5).

The third column of Table 1 shows the estimated parameters of Model 1-A[1], which corresponds to the estimation of a MNL model that includes all the attributes that are relevant in the choice behavior, that is a, b, c, d and p. Not surprisingly, for this model all estimated parameters and elasticities are statistically equal to the real ones.

Table 1: Monte Carlo Experiments Models

Variable	TRUE Values	MODEL 1-A Complete Model	MODEL 1-B Omitting a Individual Error	MODEL 1-C Omitting d Individual Error	MODEL 1-D Control Function Individual Error	MODEL 2-C Omitting d Zonal Error	MODEL 2-D Control Function Zonal Error
ASC1		-0.0783	-0.0221	0.0333	0.0225	0.102	0.155
		(-0.674)	(-0.312)	(0.418)	(0.251)	(1.25)	(1.31)
ASC2		0.0883	-0.0324	0.0420	-0.0163	0.243	0.0819
		(0.777)	(-0.456)	(0.533)	(-0.183)	(2.96)	(0.712)
a	10.0	10.5		5.14	6.45	3.32	6.45
		(21.4)		(25.5)	(24.7)	(19.7)	(19.5)
b	10.0	10.6	4.04	5.25	6.52	3.21	6.57
		(21.4)	(25.2)	(25.7)	(24.7)	(24.5)	(21.3)
c	10.0	10.4	4.14	3.02	6.28	0.845	5.32
		(20.1)	(21.4)	(16.3)	(21.3)	(9.64)	(19.5)
d	10.0	10.4	4.32				
		(20.6)	(21.9)				
p	-10.0	-10.6	-4.13	-1.00	-6.61	1.90	-3.76
		(-19.0)	(-18.0)	(-5.74)	(-17.6)	(20.2)	(-14.2)
$\hat{\mu}$					8.95		9.68
					(18.2)		(20.6)
S. Size		2000	2000	2000	2000	2000	2000
LL(0)		-2197.22	-2197.22	-2197.22	-2197.22	-2197.22	-2197.22
Final LL		-517.93	-1355.03	-1098.94	-865.21	-1109.49	-555.15
Adj ρ^2		0.761	0.383	0.497	0.606	0.492	0.744
e_{11}	-11.5	-12.7	-4.82	-1.16	-7.67	2.04	-4.46
e_{22}	-11.7	-11.9	-4.82	-1.15	-7.71	1.68	-3.06
e_{33}	-11.6	-11.8	-4.78	-1.14	-7.65	2.11	-3.98

t-test in bracklets

Adj. $\rho^2 = 1 - (\text{final } LL - \#\text{Attributes})/LL(0)$

e_{ii} direct elasticity alternative i

$\hat{\mu}$ household-alternative own fitted error of the price equation

LL Log-likelihood

The fourth column of Table 1 shows the estimated parameters of Model 1-B, where attribute a is omitted. Theoretically, because by construction attribute a is not correlated with any observed attribute (in particular with the price), this model should be consistent. However, the estimated parameters are smaller than the real ones. This is because of the change in the scale parameter of the model. The error term in this model is wider because of the omission of a. This implies a bigger variance and then a smaller scale parameter (Ben-Akiva and Lerman, 1985), which explains the fact that estimated parameters are smaller than the real ones.

However, all parameters statistically have the same absolute value with a 95% of confidence, as occurred in the true model; this fact confirms the theoretical result of consistent estimates. The elasticities are also affected by the difference in the scale parameter of the model producing values that are smaller than those of the true model; not withstanding, at least the relative values of each alternative are similar, as those in the true model.

[1] Where "1" goes for the Monte Carlo Experiment considered and "A" for the econometric model estimated.

The fifth column of Table 1 shows the estimated parameters of Model 1-C where attribute *d*, which is correlated with the price, was omitted. In this case, the estimated parameters are quite different from the ones of the true model, beyond the scale parameter difference. Taking the parameters of *a* and *b* as the base, the parameter of price *p*, is 5 times smaller than it should and the parameter of c is almost two times smaller. As expected, under the omission of a relevant attribute that is correlated with the price, the price parameter is upward biased, making the models almost useless or at least not trustworthy. Elasticities are also affected by the omission of *d,* but result in values even smaller than with the omission of *a*.

In model 1-D attribute *d* is omitted. However, in this case the control-function method (Petrin and Train, 2004) was applied, in its simplest form, to correct for endogeneity. In this case, this corresponds to regressing (OLS) the dwelling-unit price *p* on c and z (but not *d*), and then to calculate the fitted errors of this model and use them as an additional variable in the choice model.

The estimation results for this model are shown in the sixth column of Table 1 (Model 1-D). In this case, the parameters of the observed attributes *a*, *b*, *c* and *p*, are, as before, smaller than the true ones, effect that is caused by the difference in the scale parameter. However it can be noted also that now these parameters have statistically (with a 95% of confidence) **the same absolute value,** as it is the case in the true model. Thus, it can be claimed that the inclusion of $\hat{u}_{in}$ as the extra variable in the choice model satisfactorily corrected the problem of the omission of *d*. Elasticities in this case are closer to the true values in comparison to those of Model 1-C without the control-function correction, and even better than the ones of Model 1-B where *a* was omitted. This last statement could be explained by the fact that the inclusion of the control-function as an additional variable accounts for a part of the variability of the error term, moving the scale parameter up again and, with it, the estimated parameters and elasticities.

For the second Monte Carlo experiment, the same four models were estimated. The conclusions extracted from models 2-A and 2-B are the same as of models 1-A and 1-B. Thus, for the sake of space only models 2-C and 2-D are reported in Table 1. For Model 2-C, in the seventh column of Table 1, it can be noted that the specific setting considered in this case increased the bias in the model far enough to turn the estimated price parameter to be positive when *d* is omitted.

Model 2-D corresponds to the application of the control-function method for the Monte Carlo experiment two. The estimated parameters of the models are shown in the last column of Table 1. In this case, the control-function correction was good enough to turn the estimated price parameter to be negative, but not as near to the absolute value of the parameters of *a*, and *b*, as in model 1-D. The same kind of bias can be appreciated for variable *c*. This problem can be attributed to the type of bias described in Petrin and Train (2004) that the control-function method applied as in (4) can produce when the endogeneity problem occurs in a zonal level. Non parametric corrections, as the ones proposed in Petrin and Train (2004), were applied to correct for this problem[2] without achieving significant improvements. This lack of improvement can be attributed to the size of the sample considered or to the non-parametric corrections themselves, making this issue a relevant research topic to address in the future.

[2] The result of these corrections is excluded form this paper for the sake of space, but can be found in Guevara (2005).

3. APPLICATION TO REAL DATA FROM SANTIAGO DE CHILE

For the best of our knowledge only two works have considered the treatment of price endogeneity in discrete choice models or residential location. The first one is Bayer et al. (2004) who used the BLP-method and the second corresponds to Ferreira (2004) who used the control-function method. Both studies conclude that price endoegeneity is a relevant issue in residential location. However the first one suffers the problem of incidental parameters becuase it is based in the inclusion of alternative specific constant for each and every household, which implies that the estimated parameters are inconsistent (Wooldridge, 2002). In the second study what is arguably is that the author used, as instruments for the price, the taxes paid by the households, which not clearly fulfils the instrument requirements.

In the this section is presented the estimation a residential location choice model based in the information collected during the 2001 Santiago de Chile Origin and Destination survey (DICTUC, 2003). The estimation is performed first in a classic form and then correcting for endogenity using the control-function method.

1.1 Base Residential Location Model

The modeling sub-sample corresponds to 630 households which are renters who have moved to their present location during the last two years. A detailed description of the original database is given in DICTUC (2003). Guevara (2005) presents all the considerations taken into account in the elaboration of the modeling sub-sample. For each observation, 10 additional alternatives to the current location were randomly selected from the sample, and a MNL model for the location choice was estimated (McFadden, 1978). Table 2 shows the explanatory variables that were found relevant in the residential location decision procedure after a process of hypothesis and statistical testing[3].

An interpretation is given for the respective estimated parameters, the values of which can be found in the second column of Table 6, at the end of this section. The most remarkable aspect of the estimated parameters of the base model is that the estimated marginal utility of income of the wealthiest household cluster is positive (-2.33 + 1.06 + 2.23), which would mean that, everything else equal, high income households prefer apartments with higher rents. This result makes the model useless for policy analysis and in a typical study it would be attributed to poor data variability or other issues and probably solved by not reporting a specific parameter for this income stratum, but a general one where this effect would be lost.

[3] All discrete choice models where estimated using the open source software BIOGEME (Bierlaire et al., 2004)

Table 2: Base Residential Location Model Variables and Parameters Interpretation

Variable	Interpretation of Results
Apartment Dummy 1 if dwelling unit is an apartment. The additional classes in the sample are Condominium and House.	The estimated parameters indicate that small families tend to favor, all other things being equal, apartments instead of houses or condos. This effect is the inverse for large families, which can be explained by the extra value that having a house has in raising children.
Apartment Dummy for Large Households (HH) The same as the last one for household (families) with more than four members, zero otherwise.	
Big Buildings Dummy for High Income HH 1 when the building where the apartment is located has more than 4 floors and household income is over 1000 US/month.	The estimated parameter indicates that some correlation exists between the newer and the highest buildings for high income households.
Condominium Dummy for High Income HH 1 if dwelling unit is a condominium and household income is over 1000 US/month, zero otherwise.	The estimated parameter indicates that condos only offer differentiated quality attributes when they are oriented to high income HH.
Cost divided by Income Ratio between the dwelling unit monthly rental cost and the household monthly income.	As expected, the parameter is negative, meaning that households prefer cheaper dwelling units, ceteris paribus.
Cost divided by Income, Non-Low Income Same as the last one if household income is **over 400 US** dollars by month, zero otherwise.	This parameter is positive, meaning that as household income increases the marginal utility of income decreases.
Cost divided by Income, High Income Same as the last one if household income is **over 1000 US** dollars by month. Zero otherwise. Note that for high income HH, this dummy **and** the last one are on.	The same conclusion as before. But now **the total cost parameter is positive (+0.96), which is against intuition and theory.** This result makes the models useless for policy analysis and claims for a correction.
Absolute Diff. HH Income and Average Zonal Income Absolute value of the difference between HH income and the zonal average income of each alternative.	As expected, the estimated parameter is negative, confirming the clustering hypothesis found also in other studies.
% of Head of the HH (hHH) with High Education Percentage of households that have a head of the HH (hHH) with more than High School education in the zone.	The estimated results of this variable combined with the parameter of the last one indicate that households prefer, other things being equal, zones where the educational level is similar to the one of the household, confirming again the clustering hypothesis.
plus Dummy if hHH with High Education Same as the last one if the head of the HH does have more than High School education. Zero otherwise.	
Distance to Work for the hHH Distance between housing alternative and the declared workplace of the hHH.	The parameter is negative, meaning that households prefer residential locations that are nearer to the workplace of the hHH.
Same as the last one plus Dummy One Worker Same as the last one if the HH has only one worker and zero otherwise.	The parameter is also negative, what implies that hHH commuting is more relevant when only his or her time has to be considered in the location desicion.
% of Housing Square Meters by Municipality As a measure of how much housing oriented, instead of industrial oriented, an area is.	The estimated parameter is positive as expected. As more housing oriented an area is, more positive attributes for housing seekers can be expected, because of scale economics in the provision and more competition between dwelling unit providers.
Dummy West Area 1 if dwelling unit belongs to the North, East or South areas of the city. Measure social or public services differentiation between the wealthiest and the poorest areas of the city.	As expected, if everything else is equal, households are more inclined not to live in the west area (poorer) of the city, because the estimated parameter is negative.

4.1 Corrected Model Using the Control-Function Method

The next step was to apply the control-function method to correct for price endogeneity as described in the preceding sections. The instruments for the price used in this case were built, following the idea of Hausman (1997), as the average of the price of other dwelling units located in the same Municipality, $\overline{p}_n$. Within this setting, the price equation model (7) was estimated doing OLS of the prices on the instrument and additional dummies indicating the type of dwelling unit under analysis.

$$p_{in} = \alpha_0 + \alpha_1 * \overline{p}_n + \alpha_2 * Apartment + \alpha_3 * Condo + \mu_{in} \quad (7)$$

The objective of this price equation is to correct the endogeneity problem and not to make a precise forecast of the dwelling unit's price. More elaborated forms of this equation, such as the one proposed by Martínez and Henríquez (2005), are left for future research.

As explained before, the estimated parameters of model(7)[4] are used to calculate a new variable $\hat{\mu}$ which is added to utility function of the residential location model. To be consistent with the formulation used in the original model, $\hat{\mu}$ was divided by the household income to enter the utility function.

The results of the estimation of the corrected residential location choice model are shown in the third column of Table 3. The estimated parameters in this case are fairly similar to the base case. The only important difference, as expected, is related to the parameter of the cost, which is corrected downwards because of endogeneity. This correction is big enough to solve the problem of having a positive cost parameter for the higher income strata (-4.86 + 1.70 + 1.97 < 0).

Table 3: Residential Location Models Using Santiago de Chile 2001 Mobility Survey

Variables	Base Residential Location Model		Control Function Linear $\hat{\mu}$	
Apartment Dummy	0.140	(1.16)	0.130	(1.08)
Apartment Dummy Large Household (HH)	-0.840	(-3.25)	-0.856	(-3.28)
Apt. Floors > 4 Dummy Income > 1000 US$	0.263	(1.18)	0.278	(1.25)
Condominium Dummy Income > 1000 US$	1.11	(2.09)	1.27	(2.39)
Cost/Income	-2.33	(-5.91)	-4.86	(-6.80)
Cost/Income Dummy I ncome > 400 US$	1.06	(1.75)	1.70	(2.59)
Cost/Income Dummy I ncome >1000 US$	2.23	(2.65)	1.97	(2.42)
Diff with Zonal Average Income	-0.630	(-6.83)	-0.459	(-4.68)
% hHH with High Educ by zone	-1.20	(-3.35)	-0.908	(-2.60)
"" plus Dummy hHH with High Education	2.89	(7.42)	2.82	(7.27)
Distance to Work hHH	-0.119	(-9.14)	-0.125	(-9.45)
Distance to Work hHH Dummy One Worker	-0.0299	(-1.55)	-0.0299	(-1.54)
% of housing Square mt. by Comuna	1.40	(4.88)	1.45	(5.10)
West Area Dummy	-0.368	(-1.90)	-0.429	(-3.55)
$\hat{\mu}$			2.50	(4.38)
Sample Size	630		630	
LL(0)	-1501.27		-1501.27	
Final *LL*	-1222.80		-1210.54	
Adj ρ^2	0.176		0.184	

t-test in bracklets
LL Log-likelihood
Income by monh
hHH: "Head of Household"
Adj. $\rho^2 = 1 - (\text{final } LL - \#\text{Attributes})/LL(0)$
$\hat{\mu}$ household-alternative own fitted error of price equation

Interestingly, the addition of the control-function variable did improve the statistical adjustment of the model, which can be noted by the significant improvement in the likelihood. This was tested using the likelihood radio test, and can be also verified by the fact that the estimated parameter of $\hat{\mu}$ is significantly different from zero at a 95% of confidence level.

[4] Which are not reported here for the sake of space but can be found in Guevara (2005).

The final step was the application of a non-parametric correction for the fact that, if the endogeneity problem does not occur at an individual level as it was implicitly assumed, the control-function correction will be biased. This objective was accomplished by considering not only own fitted errors as additional variables, but also the average fitted errors of the other dwelling units that belong to the same Municipality, variable that was defined as $\hat{\mu}2$. This procedure was used with the purpose of capturing the correlation effect between near dwelling units, following what was proposed by Petrin and Train (2004). It also considered random parameters for the control function defined. The application of this procedure did not improve the statistical adjustment of the model and thus are not reported in this paper. These result could indicate that the zonal effect does not exist, and thus the simplest version of the control-function method is appropriate; alternatively it could mean that the non-parametric corrections methods used were not appropriate. The clarification of this question is left for future research.

4. CONCLUSIONS

Three main conclusions were obtained from the development of this research. The first one is that, price endogeneity maybe is a problem in discrete choice models of residential location. This follows not only from previous studies cited where questionable results that can be attributed to price endogeneity were reported, but also from the estimation of a residential location model of Santiago de Chile, albeit with very aggregate accessibility data and restrictive error structure assumptions.

The second conclusion is that the control-function method satisfactorily corrects for endogeneity in discrete choice models of residential location when the endogeneity occurs at an individual level. In the case the existence of zonal endogeneity the control-function method presents some bias, but can be used as a first test for the presence of endogeneity, because this method at least corrected the biased parameters in the correct direction.

The third conclusion is that, by the estimation of the model with real data in section 4, it was found that an appropriate instrument for endogenous price of dwelling units in discrete choice models of residential location is, following an idea of Hausman (1997), the average price of other dwelling units in the area.

REFERENCES

Bayer, P., R. Mcmillan and K. Reuben (2004) Residential segregation in general equilibrium. Working Paper, Department Of Economics, Yale University.

Ben-Akiva M. and S. Lerman (1985) **Discrete Choice Analysis : Theory and Application To Travel Demand**. The MIT Press, Cambridge, MA.

Bhat, C. and J. Guo (2004) A mixed spatially correlated logit model: formulation and application to residential choice modeling. **Transportation Research 38B**. 147-168.

Bierlaire, M., D. Bolduc and M. Godbout (2004) An Introduction To BIOGEME (Version 1.0). URL:Roso.Epfl.Ch/Mbi/Biogeme/Doc/Tutorial.Pdf

Berry, S., J. Levinsohn and A. Pakes (1995) Automobile Prices In Market Equilibrium. **Econometrica 63**, 841-889.

Blundell, R. and J. Powell (2004) Endogeneity In Semi-Parametric Binary Response Models. **Review Of Economic Studies 71**, 655-679.

Bresnaham, T. (1997) The Apple-Cinnamon Cheerios War: Valuing New Goods, Identifying Market Power, and Economic Measurement. Working Paper, Department of Economics, Stanford University.

Ferreira, F. (2004) You Can Take It With You: Transferability Of Proposition 13 Tax Benefits, Residential Mobility, and Willingness To Pay For Housing Amenities. Working Paper No. 72. Center For Labor Economics, University Of California, Berkeley

Greene, W. (2003) **Econometric Analysis**. 5th Edition, Prentice Hall, New York.

Guevara , C. (2005) Addressing Endogeneity In Residential Location Models. MST Thesis, Massachusetts Institute Of Technology.

Hahn, J. and J. Hausman (2003), Weak instruments: diagnosis and cures in empirical econometrics. **The American Economic Review 93**, 118-128.

Hausman, J. (1997) Valuation of new goods under imperfect competition, In R. Gordon and T. Bresnaham (eds.), **The Economics Of New Goods**, University Of Chicago Press, Chicago.

Hausman, J. (1978), Specification tests in econometrics. **Econometrica 46**, 1251-1272.

Heckman, J. (1978) Dummy endogenous variables in a simultaneous equation system. **Econometrica 46**, 931-959.

Levine, J. (1998) Rethinking accessibility and jobs-housing balance. **Journal Of American Planning Association 64**, 133-149.

Martínez, F. and R. Henríquez (2005) The RB&SM: a random bidding and supply land use equilibrium model. Working Paper, Department of Civil Engineering, Universidad De Chile.

Matzkin, R. (2004) Unobservable instruments. Working Paper, Northwestern University.

Mcfadden, D.L. (1978) Modeling the choice of residential location. In Karlqvist (eds.), **Spatial Interaction Theory and Planning Models**. North-Holland, Amsterdam.

Nevo, A (2001) Measuring market power in the ready to eat breakfast cereal industry. **Econometrica 69**, 307-342

Petrin, A. and K. Train (2004) Omitted product attributes in discrete choice models. Working Paper, National Bureau Of Economic Research.

Quigley, J. (1976) Housing demand in the short run: an analysis of politomous choice, **Regional Science and Urban Economics 15**, 41-63.

DICTUC (2003) **Encuesta Origen Destino de Viajes Santiago 2001**. Informe Final a Sectra, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile.

Sermonss, M and F. Koppelman (2001) Representing the differences between female and male commute behavior in residential location choice models. **Geography 9**, 101-110.

Waddell, P. (1996) Accesibility and residential location: the interaction of workplace, residential mobility,tenure and location choices. Lincoln Land Institute TRED Conference.
Waddell, P. (1992) A multinomial logit model of race and urban structure. **Urban Geography**. **13**, 127-141.

Walker, J. (2001) Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Phd Dissertation, Massachusetts Institute of Technology.