

# Change of Scale and Forecasting with the Control-function Method in Logit Models

by

**Cristian Angelo Guevara**

Universidad de los Andes, Chile  
Facultad de Ingeniería y Ciencias Aplicadas  
San Carlos de Apoquindo, 2200  
Las Condes, Santiago, Chile  
Tel: 56-2-4129364  
Fax: 56-2-4129642  
aguevara@uandes.cl

and

**Moshe E. Ben-Akiva**

Massachusetts Institute of Technology  
Department of Civil & Environmental Engineering  
77 Massachusetts Ave., 1-181  
Cambridge, MA 02139  
Tel: 617-253-5324  
Fax: 617-253-0082  
mba@mit.edu

## ABSTRACT

The control-function method is the most suitable tool to address endogeneity for several discrete choice models that are relevant in transportation research. The estimators obtained with the control-function method are consistent only up to a scale. In this paper we first depict the determinants of this change of scale by adapting an existing result for omitted orthogonal attributes in Logit models. Then, we study the problem of forecasting under these circumstances. We show that a procedure proposed in previous literature may lead to significant biases, and we suggest novel alternatives to be used with synthetic populations. We also discuss potential extensions of these results to other non-Logit discrete choice models. We use Monte Carlo experimentation and real data on residential location choice to demonstrate these results. The paper finishes summarizing the findings of this research and suggesting future lines of research in this area.

**Key Words:** Endogeneity, Discrete Choice, Control-function, Logit, Forecasting.

## 1 Introduction

A discrete choice model suffers from endogeneity when the systematic part of the utility is correlated with the error term. This problem is common to several discrete choice models that are relevant in transportation research. It can be caused by errors in variables, simultaneous determination, or the omission of relevant attributes that are correlated with the observed ones. Endogeneity is a critical modeling failure that leads to the inconsistent estimation of model parameters. Intuitively, if a variable is endogenous, changes in the error term will be misinterpreted as resulting from changes of the endogenous variable, making impossible the consistent estimation of the model parameters.

The problem of endogeneity is relevant, for example, in modeling the choice of residential location. Consider, for example, the case of two seemingly equal dwellings that differ only in that one has been recently renovated and consequently has a higher price. If the data on dwelling's renovation is not available, observations of choices toward the dwelling with the higher price would lead to the erroneous conclusion that the sensitivity to price is smaller than it really is. Numerous empirical applications in residential location choice modeling have shown estimated coefficients of dwelling price that are non-significant or even positive when endogeneity is not taken into account (Guevara and Ben-Akiva, 2006; Bhat and Guo, 2004; Sermonss and Koppelman, 2001; Levine, 1998; Waddell, 1992; Quigley, 1976).

Two main methods have been proposed to correct for endogeneity in discrete choice models when the endogenous variable is continuous: The BLP method (Berry et al., 1995) and the control-function method (Heckman, 1978; Hausman, 1978). Other methods that have been proposed are 2SIV (Newey, 1985) and Amemiya's (1978) method. The control-function method is particularly suitable to address endogeneity in several models that are relevant in transportation research.

In this paper we study two methodological issues that arise in the application of the control-function method. We first study the problem of the change of scale that results from its application in Logit models. In this, we extend the analysis of Cramer (2007) and Daly (2008), on the change of scale produced by the omission of orthogonal attributes in Logit models. Then, we study the problem of forecasting with Logit models corrected for endogeneity using the control-function method. We show that a procedure proposed by Wooldridge (2002) may lead to significant biases and propose novel alternatives to use the control-function method with synthetic populations, making it applicable for microscopic integrated urban models such as UrbanSim (Wadell et al., 2008). We also discuss the potential extension of these results to other non-Logit discrete choice models. We use Monte Carlo experimentation and real data on residential location choice to demonstrate these results.

The paper is structured as follows. This introduction is followed by a review of the fundamentals of the control-function method in the correction of endogeneity. Then, in sections 3 and 4 we study the problems of change of scale and forecasting with the control-function method in Logit models from a theoretical perspective. Then, in Section 5 we use Monte Carlo experimentation to demonstrate the theoretical issues discussed previously. Finally, we analyze those issues in the light of an application of the control-function method to real data on residential location from the city of Lisbon, Portugal. The

paper finishes highlighting the main findings of this research, and suggesting future lines of research in this area.

## 2 Addressing Endogeneity with the Control-function Method

The control-function method to address endogeneity was originally proposed by Hausman (1978) and Heckman (1978). Rivers and Vuong (1988) used this approach in binary Probit, and Petrin and Train (2002) extended it to Logit mixture models. The control-function method consists of the construction of an auxiliary variable, which when added to the systematic part of the utility function, the remaining error of the model will no longer be correlated with observed variables.

To deploy the fundamentals of the control-function method, consider the problem described in Eq. (1), where a group of  $N$  agents ( $n$ ) face the selection of an alternative  $i$  among the  $J$  elements in the choice-set  $C_n$ .

$$\begin{aligned}
 U_{in} &= \beta_p p_{in} + \beta_x x_{in} + \overbrace{\xi_{in} + e_{in}}^{\varepsilon_{in}} \quad n=1, \dots, N; i \in C_n \\
 p_{in} &= \alpha_z z_{in} + \alpha_x x_{in} + \delta_{in} \\
 y_{in} &= 1[U_{in} = \max_{j \in C_n} \{U_{jn}\}]
 \end{aligned} \tag{1}$$

The agent  $n$  perceives a utility  $U_{in}$  from dwelling  $i$ . The utility depends linearly on variables  $p_{in}$ , an attribute  $x_{in}$ , and a zero mean error term  $\varepsilon_{in}$ , which can be decomposed into two parts  $\xi_{in}$  and  $e_{in}$  that also have zero mean. For expositional purposes, we will denominate variable  $p_{in}$  the price and will assume that this is the unique endogenous variable.  $U_{in}$  is a latent variable. The researcher observes variables  $x_{in}$ ,  $z_{in}$ ,  $p_{in}$  and the choice  $y_{in}$ , which takes value 1 if the alternative  $i$  has the largest utility among the elements in choice-set  $C_n$ , and zero otherwise. The price  $p_{in}$  is determined as a linear function of variables  $z_{in}$ ,  $x_{in}$ , and a zero mean error  $\delta_{in}$ . This expression is termed the price equation. For notational purposes, it will be considered from this point that  $U$ ,  $p$ ,  $x$ ,  $\varepsilon$ ,  $\xi$ ,  $e$ ,  $z$ ,  $\delta$  and  $y$  are vectors compounded by the respective variables stacked by alternatives  $i$  and agents  $n$ .

Variables  $x$  and  $z$  are exogenous, meaning that they are independent of all error terms  $\varepsilon$ ,  $\xi$ ,  $e$ , and  $\delta$  of the model. Variable  $x$  is said to be a control because it is exogenous and it appears in the specification of the utility function. Variable  $z$  is a suitable instrument for price  $p$  because it is relevant (correlated with  $p$ ) and valid (independent of  $\varepsilon$ ). The fact that  $z$  does not appear in the utility function allows identification. Additionally, the error term  $e$  is independent of the observed variables  $p$ ,  $x$  and  $z$ , and of the error term  $\delta$ .

Endogeneity problems arise when  $\delta$  is correlated with  $\xi$ . In that case,  $p$  will be correlated with  $\xi$  and the standard estimation methods will fail to retrieve consistent estimators of model parameters. This problem may occur, for example, if  $\xi$  contains relevant attributes that are correlated with  $p$ , but cannot be measured by the researcher.

The model described in Eq. (1) represents a triangular system. The latent and the instrumental variables are jointly independent, allowing the structural equation for the utility and the price equation to be recursive. Chesher (2010) showed that these are some of the possible forms to allow point identification, a critical requirement to be able to correct for endogeneity using instrumental variables.

Consider now the problem of constructing the control-function for the problem described in Eq. (1). The idea is to build an auxiliary variable, which when added to the systematic part of the utility function, will control for the endogenous component of the error term. It can be shown that the conditional expectation of  $\zeta$ , given  $\delta$ , can play this role (Wooldridge, 2002). Assuming then that  $\zeta$  and  $\delta$  are jointly Normal, this conditional expectation would be linear and then

$$\zeta_{in} = \beta_{\delta} \delta_{in} + v_{in}, \quad (2)$$

where  $v$  will be independent of  $\delta$  and will follow a Normal distribution with zero mean and a fixed variance  $\sigma_v^2$ . Under these conditions, the error term  $v$  will not be correlated with  $p$  or  $x$ . Therefore, assuming for the moment that  $\delta$  is observed, the endogeneity problem can be solved if this decomposed  $\zeta$  shown in Eq. (2) is replaced in the utility function as shown in Eq. (3).

$$\begin{aligned} U_{in} &= \beta_p p_{in} + \beta_x x_{in} + \varepsilon_{in} = \beta_p p_{in} + \beta_x x_{in} + \beta_{\delta} \delta_{in} + v_{in} + e_{in} \\ p_{in} &= \alpha_z z_{in} + \alpha_x x_{in} + \delta_{in} \\ y_{in} &= 1[U_{in} = \max_{j \in C_n} \{U_{jn}\}] \end{aligned} \quad (3)$$

Intuitively, the error of the price equation  $\delta$ , captures the part of price that is correlated with the error term of the model and, when it is added to the systematic part of the utility, it controls for the endogeneity problem.

The practical problem that  $\delta$  is not observed can be addressed, for example, noting that  $\delta$  can be consistently estimated by the residuals  $\hat{\delta}$  of the ordinary-least-squares regression of  $p$  on  $x$  and  $z$ . Then, if  $\hat{\delta}$  is inserted into the choice model, the consistency of the estimators of the model parameters would be guaranteed by the Slutsky theorem (see, e.g. Ben-Akiva and Lerman, 1985). This method to address endogeneity in discrete choice models is termed the two-stage control-function (2SCF) and is used, for example, by Rivers and Vuong (1988), Petrin and Train (2002) and Guevara and Ben-Akiva (2006).

An alternative way to address the fact that  $\delta$  is not observed is to consider jointly the likelihood of the price equation and the choice model in the estimation of the model parameters. This method is used by Villas-Boas and Winner (1999) and by Park and Gupta (2009), and can be seen as a full-information maximum-likelihood (FIML) version 2SCF, which can be therefore seen as a limited-information maximum-likelihood (LIML) method. This FIML approach is also equivalent to consider that Eq. (2) corresponds to a structural equation and to treat  $\zeta$  as a latent variable (Guevara, 2010). With some modifications, this approach is also equivalent to the methods used by Zimmer and Trivedi (2006) or by Bhat and Eluru (2009) to address endogeneity in discrete choice models.

There is a trade-off in using LIML or FIML to address endogeneity. On the one hand, LIML is easier to estimate and can handle a broader range of joint distributions of  $\zeta$  and  $\delta$  but, on the other hand, FIML is usually more efficient and allows the direct calculation of standard errors though the evaluation of the inverse of the Fisher-information matrix. In this paper we study methodological issues that arise both with FIML and LIML (2SCF) approaches to address endogeneity with the control-function method. For expositional purpose, we will concentrate on the 2SCF method, but will describe the implications of extending the analysis to FIML when it is relevant.

### 3 Change of Scale in the Application of the Control-function Method

The correction for endogeneity using the control-function method (both with FIML or LIML) produces consistent estimators of the model parameters but only up to a certain scale. That is, the ratios between the estimators are consistent estimators of the ratios of the parameters of the true model, but the actual estimators of the model parameters are inconsistent. This is also true, in general, for other methods to correct for endogeneity in discrete choice models, such as BLP (Berry et al., 1995), 2SIV (Newey, 1985) and Amemiya's (1978) methods.

The change of scale results from the fact that the error term with the control-function correction in Eq. (3) is  $v + e$ , whereas the error term of the original model shown in Eq. (1) was only  $e$ . Therefore, if the variance of  $v$  is not null, the control-function correction will trigger a change of scale in the estimated parameters. This effect is analogous to that of the omission of a relevant and orthogonal attribute in discrete choice models, an attribute that truly belongs to the systematic part of the utility, but is uncorrelated with other observed attributes. The problem of the change of scale due to the omission of relevant but orthogonal variables was originally studied by Yatchew and Griliches (1985) for the Probit model. Cramer (2007) extended this analysis to the binary Logit model. Here, we use their framework to study the change of scale caused by the application of the 2SCF method in correcting for endogeneity in Logit models.

Consider the true model shown in Eq. (1) where  $\xi$  is observed, and assume that the error  $e$  is distributed Extreme Value  $(0, \mu_e)$ . As with any Logit model, the scale is not identifiable and normalization is required. The usual normalization is to set  $\mu_e = 1$ . This is equivalent to normalizing the variance of the differences of  $e$  across alternatives to be equal to  $\sigma_e^2 = \pi^2/3$  (see, e.g., Ben-Akiva and Lerman, 1985). Consider now the model corrected for endogeneity using the control-function method described in Eq. (3). The first problem in the determination of the change of scale in this case is to depict the distribution of the error term  $v + e$ .

We will consider that  $v + e$  follows, or can be approximated using an Extreme Value distribution, such that the model with the control-function correction also becomes a Logit. This assumption might seem difficult to sustain at first. In Eq. (2) we said that  $v$  followed a Normal distribution and the sum of Normal and an Extreme Value distribution does not follow an Extreme Value distribution. In fact, there are no parametric distributions of  $e$  and  $v$ , which would result in that  $v + e$  is distributed Extreme Value. However if the sample is large enough, it is first possible to claim the Central Limit Theorem to say that  $v + e$  will be normally distributed. The argument is completed using the results from Lee (1982) and Ruud (1983), which state that the approximation of a Normal by an Extreme Value distribution causes only negligible discrepancies. The Monte Carlo experiments reported in Section 5 will show that this approximation is also valid in the study of the change of scale with the application of the control-function method.

Considering that  $v + e$  is distributed Extreme Value, the usual normalization for the model shown in Eq. (3) would be  $\mu_{v+e} = 1$ . However, this would imply that  $\sigma_{v+e}^2 = \pi^2/3$ , what is incompatible with the normalization assumed for the model in Eq. (1). To determine the compatible normalization, consider first the ratio between the scales of the

two models. Since  $v$  and  $e$  are uncorrelated by construction, this ratio will depend only on the variances of  $v$  and  $e$  as follows:

$$\frac{\mu_{v+e}}{\mu_e} = \frac{\sigma_e}{\sigma_{v+e}} = \frac{\sigma_e}{\sqrt{\sigma_v^2 + \sigma_e^2 + 2\text{cov}(v,e)}} = \frac{\sigma_e}{\sqrt{\sigma_v^2 + \sigma_e^2}} = \frac{1}{\sqrt{1 + \frac{\sigma_v^2}{\sigma_e^2}}}.$$

Then, if the normalization of the model in Eq. (1)  $\sigma_e^2 = \pi^2/3$  is to be maintained, the compatible scale of the model shown in Eq. (3) should be

$$\mu_{v+e} = 1/\sqrt{1 + 3\sigma_v^2/\pi^2}. \quad (4)$$

Therefore, the estimators obtained from a model that was corrected for endogeneity using the 2SCF will be smaller than those of the true model shown in Eq. (1).

This change of scale is unknown to the researcher in a practical application because the variance of  $v$  is not identifiable. This raises the natural question of what is the cost of the omission of  $v$  in the estimation of the control-function method. It turns out that the cost of this omission is negligible. First, it is usually only the ratio between the coefficients what is relevant and the ratios are indeed obtained consistently with the change of scale that results from the application of the control-function method. For example, the metric that is relevant in a mode choice model is the subjective value of travel time savings, which is calculated as the ratio of the coefficients of travel time and travel cost (see, e.g. Jara-Diaz and Guevara, 2003). Second, beyond the ratios, the other thing that is important is the effect in aggregate elasticities, the impact of this correction in forecasting. In the next section we study the problem of simulation and forecasting with the control function method. We show that some approaches that have been proposed in previous literature are valid and some may cause significant biases. Additionally, we propose novel methods to address this issue with synthetic populations.

#### 4 Simulation and Forecasting with the Control-function Method

The impact on elasticities of the change of scale resulting from the application of the control-function method is also related with the problem of omitted orthogonal attributes. The first insight into this issue comes from Wooldrige (2002, pp. 470), who proved that the omission of an attribute that is uncorrelated with other observed variables will not change the expected value of the derivative of the choice probability in a binary Probit model. There is no equivalent analytical result for Logit, but Cramer (2007), for binary Logit, and Daly (2008), for multinomial Logit, used Monte Carlo experimentation to show that the Average Sample Effect (ASE), the sample average of the derivative of the choice probability, differs insignificantly between the full model and a model that omits a variable that is uncorrelated with other observed variables.

Cramer's and Daly's results can be directly extended to the case of the change of scale caused by the application of the 2SCF method because the error term  $v$  acts as an omitted orthogonal attribute in Eq. (3). Assume that  $e$  and  $e+v$  are distributed (or can be approximated) using an Extreme Value distribution. Term:

$\hat{P}_n(i)$ : The choice probability of alternative  $i$  calculated using estimators  $\hat{\beta}$  from the model shown in Eq. (1), including variable  $\zeta$  in the utility, and

$\hat{P}_n(i)$ : The choice probability calculated using estimators  $\hat{\beta}$  of the model shown in Eq. (3), including  $\hat{\delta}$  and omitting  $v$ .

Then, the extension of Cramer's and Daly's results to the analysis of the impact of the application of the 2SCF method in the ASE of price, for alternative  $i$ , in a Logit model, can be summarized as follows:

$$ASE_p(i) = \frac{1}{N} \sum_{n=1}^N \frac{\partial \hat{P}_n(i)}{\partial p_m} = \frac{1}{N} \sum_{n=1}^N \left(1 - \hat{P}_n(i)\right) \hat{P}_n(i) \hat{\beta}_p \approx \frac{1}{N} \sum_{n=1}^N \left(1 - \hat{P}_n(i)\right) \hat{P}_n(i) \hat{\beta}_p.$$

Intuitively, although the omission of  $v$  in Eq. (3) makes the scale of that model to be smaller (Eq. 4), and therefore  $\hat{\beta}_p < \hat{\beta}_p^*$ , this change of scale is compensated by the fact that  $\hat{P}_n(i)$  becomes closer to 0.5 than  $\hat{P}_n^*(i)$ , and therefore  $\left(1 - \hat{P}_n(i)\right) \hat{P}_n(i) < \left(1 - \hat{P}_n^*(i)\right) \hat{P}_n^*(i)$ .

The fact that the combination of this two effects results in a negligible change in the sample has not been demonstrated only numerically for Logit by Crammer (2007) and Daly (2008). Formally, this result can be seen as an extension of Wooldridge's (2002) analytical result of for Probit, claiming Lee's (1982) and Ruud's (1983) approximation results for Logit.

Formally, Eq. (5) shows the expression of the simulated probabilities that would have to be used for the problem described in Eq. (3), assuming that the model is a Logit. In this case the  $\hat{\beta}$ 's are the estimators obtained from the application of the 2SCF method and the superscript 1 is used to highlight that  $p$  and  $x$  vary in the forecasting phase, but  $\hat{\delta}$  is fixed.

$$\hat{P}^1(i) = \frac{1}{N} \sum_{n=1}^N \hat{P}_n^1(i) = \frac{1}{N} \sum_{n=1}^N \frac{e^{\hat{\beta}_p p_m^1 + \hat{\beta}_x x_m^1 + \hat{\beta}_\delta \hat{\delta}_m}}{\sum_{j \in C_n} e^{\hat{\beta}_p p_j^1 + \hat{\beta}_x x_j^1 + \hat{\beta}_\delta \hat{\delta}_j}} \quad (5)$$

This expression for forecasting is suitable for the 2SCF method, the LIML version of the control-function method. In the case the FIML version of the control-function method, the equivalent expression to do forecasting corresponds to the following expression:

$$\hat{P}^1(i) = \frac{1}{N} \sum_{n=1}^N \hat{P}_n^1(i) = \frac{1}{N} \sum_{n=1}^N \frac{e^{\hat{\beta}_p p_m^1 + \hat{\beta}_x x_m^1 + \hat{\beta}_\delta (p_m^0 - \hat{\alpha}_x x_m^0 - \hat{\alpha}_z z_m^0)}}{\sum_{j \in C_n} e^{\hat{\beta}_p p_j^1 + \hat{\beta}_x x_j^1 + \hat{\beta}_\delta (p_j^0 - \hat{\alpha}_x x_j^0 - \hat{\alpha}_z z_j^0)}},$$

where the superscript 0 is to indicate the data from the sample used for estimation.

The estimator in Eq. (5) of the choice probabilities may be impractical in some cases because the data used to estimate the model might not be available for simulation, making the use of the residuals in simulating phase impossible. This occurs, for example, in microscopic integrated models of the urban system such as UrbanSim (Waddell et al., 2008), where the choice models are estimated using real data on households  $n$  and dwellings  $i$ , but are applied to synthetic populations  $\tilde{n}$  and  $\tilde{i}$ . We consider this problem for the rest of the section.

Wooldridge (2002 pp. 475) proposed a different estimator of the choice probabilities that seems to overcome the limitations that arise in simulation with synthetic populations. The idea is to avoid the need for calculating  $\hat{\delta}$  for the synthetic populations, addressing the change of scale caused by its omission. Wooldridge presents the correction required for Probit. The equivalent correction for Logit can be applied, following the same derivation used before to arrive at Eq. (4), by dividing the estimators with the factor

$$\sqrt{1 + 3\hat{\beta}_\delta^2 \hat{\sigma}_\delta^2 / \pi^2},$$

where  $\hat{\sigma}_\delta^2$  is the sample variance of the residuals of the first stage of the 2SCF. This estimator of the choice probabilities is shown in Eq. (6), where  $\sim$  denotes synthetic households and dwellings, and the superscript 1 is used to denote values from the forecasting phase.

$$\hat{P}^1(\tilde{i}) = \frac{1}{\tilde{N}} \sum_{\tilde{n}=1}^{\tilde{N}} \frac{e^{\frac{\hat{\beta}_p}{\sqrt{1+3\hat{\beta}_\delta^2 \hat{\sigma}_\delta^2 / \pi^2}} p_{\tilde{i}\tilde{n}}^1 + \frac{\hat{\beta}_x}{\sqrt{1+3\hat{\beta}_\delta^2 \hat{\sigma}_\delta^2 / \pi^2}} x_{\tilde{i}\tilde{n}}^1}}{\sum_{j \in C_{\tilde{n}}} e^{\frac{\hat{\beta}_p}{\sqrt{1+3\hat{\beta}_\delta^2 \hat{\sigma}_\delta^2 / \pi^2}} p_{j\tilde{n}}^1 + \frac{\hat{\beta}_x}{\sqrt{1+3\hat{\beta}_\delta^2 \hat{\sigma}_\delta^2 / \pi^2}} x_{j\tilde{n}}^1}} \quad (6)$$

However, this estimator of the choice probabilities is inconsistent. The problem is that Eq. (6) neglects the fact that  $\delta$  is correlated with  $p$  when the model suffers from endogeneity. Therefore, taking  $\hat{\delta}$  out of the systematic part of the utilities will impact more those alternatives with higher prices, than those with lower prices. Taking  $\hat{\delta}$  out will not be equivalent to add or to subtract an independent error and, consequently, cannot be compensated with a simple change of scale. We will explore the effect of this problem later in Section 5 using Monte Carlo experimentation.

Instead of using Eq. (6) for the case of synthetic populations, one alternative is to construct a control-function for each synthetic dwelling  $\tilde{i}$  and household  $\tilde{n}$  using the following expression:

$$\hat{\delta}_{\tilde{i}\tilde{n}}^0 = p_{\tilde{i}\tilde{n}}^0 - \alpha_z z_{\tilde{i}\tilde{n}}^0 - \alpha_x x_{\tilde{i}\tilde{n}}^0,$$

where the superscript zero indicates that the synthetic data used in the calculation of  $\hat{\delta}$  should come from the base year.

If the dwellings available for estimation in the first stage of the 2SCF are a random sample from the population, this expression can be calculated using the estimators  $\hat{\alpha}$  of the first stage of the 2SCF. Otherwise, the coefficients  $\alpha$  could be calculated by re-estimating the first stage of the 2SCF using the attributes of synthetic dwellings  $\tilde{i}$  and the characteristics of synthetic households  $\tilde{n}$ . In both cases,  $\hat{\delta}_{\tilde{i}\tilde{n}}^0$  has to be included then as an auxiliary variable in the utility, as shown in Eq. (7).

$$\hat{P}^1(\tilde{i}) = \frac{1}{\tilde{N}} \sum_{\tilde{n}=1}^{\tilde{N}} \frac{e^{\hat{\beta}_p p_{\tilde{i}\tilde{n}}^1 + \hat{\beta}_x x_{\tilde{i}\tilde{n}}^1 + \hat{\beta}_\delta \hat{\delta}_{\tilde{i}\tilde{n}}^0}}{\sum_{j \in C_{\tilde{n}}} e^{\hat{\beta}_p p_{j\tilde{n}}^1 + \hat{\beta}_x x_{j\tilde{n}}^1 + \hat{\beta}_\delta \hat{\delta}_{j\tilde{n}}^0}} \quad (7)$$



The FIML version of Eq. (7) is equivalent, but using the respective estimators of the choice model parameters and of the price equation coefficients to calculate  $\hat{\delta}_{\tilde{i}\tilde{n}}$ .

The application of this simulator may still be cumbersome because it requires the criteria used to build the instruments with the real data to be valid for the synthetic population. If the synthetic prices are reliable but the validity of the criteria used to build the instruments is uncertain or difficult to implement for the synthetic data, it would still be possible to generate a consistent estimator of the simulated probabilities by using the Logit Mixture model shown in Eq. (8), where  $f(\delta|p)$  is the conditional distribution of  $\delta$  given  $p$ .

$$\hat{P}^1(\tilde{i}) = \frac{1}{\tilde{N}} \sum_{\tilde{n}=1}^{\tilde{N}} \int \cdots \int \frac{e^{\hat{\beta}_p p_{\tilde{i}\tilde{n}}^1 + \hat{\beta}_x x_{\tilde{i}\tilde{n}}^1 + \hat{\beta}_\delta \delta_{\tilde{i}\tilde{n}}}}{\sum_{j \in C_{\tilde{n}}} e^{\hat{\beta}_p p_{j\tilde{n}}^1 + \hat{\beta}_x x_{j\tilde{n}}^1 + \hat{\beta}_\delta \delta_{j\tilde{n}}}} f(\delta | p) d\delta \quad (8)$$

In a practical application, the multifold integral shown in Eq. (8) can be calculated using Monte Carlo integration, where  $f(\delta|p)$  can be inferred from the sample (provided it is random) by estimating the auxiliary regression

$$\hat{\delta}_{in} = \gamma_0 + \gamma_p p_{in}^0 + \lambda_{in},$$

where the superscript 0 indicates that this model is estimated using data from the base year.

Then, for each synthetic dwelling  $\tilde{i}$  and household  $\tilde{n}$ , several draws  $r$  of  $\delta$  should be obtained using the expression

$$\delta_{\tilde{i}\tilde{n}r} = \hat{\gamma}_0 + \hat{\gamma}_p p_{\tilde{i}\tilde{n}}^0 + \varepsilon_{\tilde{i}\tilde{n}r},$$

where  $p_{\tilde{i}\tilde{n}}^0$  is the price of the synthetic dwelling in the estimation year,  $\hat{\gamma}$  are the estimators of the auxiliary regression for  $\hat{\delta}$ , and  $\varepsilon_{\tilde{i}\tilde{n}r}$  is a random draw distributed Normal  $(0, \hat{\sigma}_\lambda^2)$ , where  $\hat{\sigma}_\lambda^2$  is the sample variance of the residual  $\lambda$  of the auxiliary regression. Then, the choice probability for each household is obtained by averaging across draws. Finally, the probability of each synthetic dwelling shown in Eq. (8) is obtained by averaging across synthetic households. The FIML version of Eq. (8) is equivalent, but using the respective estimators of the parameters for depicting the conditional distribution of  $\delta$  given  $p$ .

Finally, regarding the extension of these results to other non-Logit models, the effect of the change of scale that occurs with the application of the 2SCF to Logit models also occurs for Probit and other non-Logit models, such as the Nested Logit. The invariance of aggregate elasticities holds also for Probit, as shown by Wooldridge (2002). However, the extension of this last principle to other non-Logit models, such as the Nested Logit is not necessarily true. The analysis of this extension is left for further research.

## 5 Monte Carlo Experiment

In this section we develop a Monte Carlo experiment to analyze the issues studied in sections 3 and 4. The true model considered in this experiment is a binary Logit ( $J=2$ ) with a latent utility that depends linearly on four attributes  $x_1$ ,  $x_2$ ,  $p$  and  $\zeta$ , and an error term  $e$  independent and identically distributed (*iid*) Extreme Value  $(0,1)$ . The coefficients of each attribute are shown in Eq. (9).

$$U_{in} = -2p_{in} + 1x_{1in} + 1x_{2in} + 1\xi_{in} + e_{in} \quad (9)$$

Variable  $p$  (price) is defined as a function of  $\zeta$ , an instrument  $z$ , and an error term  $\tilde{\delta}$  *iid* Uniform (-1,1), with the coefficients shown in Eq. (10). Variables  $x_1$ ,  $x_2$ ,  $\zeta$  and  $z$  were generated as *iid* Uniform (-3,3). The synthetic database consists of  $N=2,000$  observations and was generated 100 times.

$$p_{in} = 5 + 0.5\xi_{in} + 0.5z_{in} + \tilde{\delta}_{in} \quad (10)$$

Note that by virtue of Eq. (10) variables  $p$  and  $\zeta$  are correlated. Therefore, if  $\zeta$  is omitted in the specification of the utility function, the choice model will suffer from endogeneity. In turn, since  $x_1$  and  $x_2$  are not correlated with other variables, the model will not suffer from endogeneity if  $x_1$  or  $x_2$  are omitted. Note also that  $z$  is, by construction, a valid instrument. From Eq. (10)  $z$  is correlated with  $p$  and independent of  $e$ .

To assess the impact of endogeneity in the estimation of the model parameters and to evaluate the issues that arise in its correction using the control-function method, four models were estimated for each repetition of the Monte Carlo experiment: the true model, a model where  $x_1$  is omitted, a model where  $\zeta$  is omitted, and a model where  $\zeta$  was omitted but the problem was addressed using the 2SCF method.

For each model, the average, bias, mean squared error (MSE) and the t-test against the true values of the estimators of the model parameters are reported in Table 1. The use of repetitions avoids the risk of dealing with a singular case that may bias the analysis and avoids the need for correcting the standard errors required in the application two-stage procedures.

**Table 1 Monte Carlo Experiment: Change of Scale with Omission of Attributes and Endogeneity Correction**

	Metric	$\hat{\beta}_p$	$\hat{\beta}_{x_1}$	$\hat{\beta}_{x_2}$	$\hat{\beta}_\xi$	$\hat{\beta}_\delta$	$\hat{\beta}_p / \hat{\beta}_{x_2}$
True Model	Average	-1.990	0.9960	0.9949	0.9957		-1.980
	Bias	0.009561	-0.004032	-0.005127	-0.004288		0.02022
	MSE	0.008985	0.003247	0.002755	0.002990		0.2148
	t-test true	0.1014	-0.07094	-0.09814	-0.07867		0.04366
Omitting $x_1$	Average	-1.122		0.5627	0.5641		-1.998
	Bias	0.8778		-0.4373	-0.4359		0.002259
	MSE	0.7742		0.1923	0.1913		0.2550
	t-test true	14.53		-13.61	-12.03		0.004473
Omitting $\zeta$	Average	-0.7994	0.6675	0.6689			-1.212
	Bias	1.201	-0.3325	-0.3311			0.7881
	MSE	1.443	0.1119	0.1108			0.7276
	t-test true	26.80	-8.873	-9.359			2.415
2SCF	Average	-1.563	0.7813	0.7825		1.078	-1.992
	Bias	0.4372	-0.2187	-0.2175			0.008215
	MSE	0.1983	0.04955	0.04884			0.2581
	t-test true(*)	5.161	-5.277	-5.531		13.09(*)	0.01617
100 Repetitions. $N=2,000$ . $J=2$ . (*) t-test against zero for $\hat{\beta}_\delta$							

The first row below the labels in Table 1 shows the estimators obtained from the true model. In this case all estimators of the model parameters are statistically equal (with 95% confidence) to their true values.

The second row shows the estimators of the model that omits variable  $x_1$ . This model does not suffer from endogeneity because  $x_1$  is not correlated with other variables. The estimators in this case are consistent, but only up to a scale. It should be noted that the ratio between the coefficients of  $p$  and  $x_2$  is statistically equal (with 95% confidence) to its true value (-2). In turn, each coefficient is significantly smaller than from its respective true value. This is explained by the change of scale caused by the addition of the variance of  $x_1$  to the error of the model. The change of scale observed in Table 1 is of approximately 0.56, a value that can be approximately calculated by substituting  $\sigma_v^2$  by the variance of  $x_1$  in Eq. (4). Finally it should be noted that, although the omission of  $x_1$  did not impacted the consistency of the estimators, it did affected efficiency. There is a loss of information when  $x_1$ , which is truly relevant in the choice process, is not used in modeling it. This can be noted in the increase of the MSE of the estimator of the ratio between the coefficients of  $p$  and  $x_2$  for this model, when compared to the respective MSE of the true model.

The third row in Table 1 shows the estimators that are obtained when  $\zeta$  is omitted. This model suffers from endogeneity because  $\zeta$  is correlated with  $p$ . In this case the estimators are different from those of the true model, but not only up to a scale. The ratios between coefficients are also affected. Since  $p$  and  $\zeta$  are positively correlated, the omission of  $\zeta$  causes a positive bias in the coefficient of  $p$ . Consequently, the ratio between the coefficients of  $p$  and  $x_2$  is approximately -1.2 instead of -2, as it was in the true model. Intuitively, the problem is that positive shocks of  $\zeta$  on the utility are confounded as the results of shocks of  $p$ , causing a positive bias in the estimator of the coefficient of  $p$ .

Consider now the case of the model that omits  $\zeta$ , but is corrected using the 2SCF method. Note first that the estimator of the auxiliary variable is statistically different (with 95% confidence) from zero. This correctly confirms that endogeneity was present in the model without the correction (Rivers and Vuong, 1988). Second, although the model coefficients are not numerically equal to those of the true model, the ratios between them are the same. Particularly, the ratio between the coefficient of  $p$  and  $x_2$  is again statistically equal (with 95% confidence) to -2. The change of scale between the estimators in this case is approximately 0.78, shift that can be calculated by considering the variance of  $v$  in Eq. (4). Lastly, similarly to what occurred with the omission of  $x_1$ , although the correction for endogeneity resulted in consistent estimators up to a scale, the fact that the term  $v$  was omitted caused a reduction in efficiency. This can be noted in the increase of the MSE of the estimator of the ratio between the coefficients of  $p$  and  $x_2$  for this model, when compared to the respective MSE of the true model.

The next step in the analysis of this Monte Carlo experiment is to show how the different models behave in the forecasting or simulation phase. To do so, we use the estimators of the different models to calculate the Aggregated Direct Elasticity (ADE) of price (Eq. 11), defined as the effect of an incremental change in price on the expected share of the group choosing alternative  $i$  (Ben-Akiva and Lerman, 1985, pp. 112).

$$\text{ADE}_p(i) = \frac{\beta_p}{\sum_{n=1}^N P_n(i)} \sum_{n=1}^N (1 - P_n(i)) P_n(i) p_{in} \quad (11)$$

The experiment was repeated 100 times. Table 2 reports the average and standard errors of ADE for  $i=1$  across the repetitions. Additionally, we simulated the effect of increasing the price of alternative 1 by 50% for all  $n$ 's and calculated the average probability of choosing alternative 1 across the 2,000 observations, before ( $\hat{P}^0(i)$ ) and after ( $\hat{P}^1(i)$ ) the price shift.

The first row in Table 2 shows the metrics obtained using the estimates that result from considering  $p$ ,  $x_1$ ,  $x_2$  and  $\xi$  in the specification of the systematic utility. These statistics act as a benchmark. Table 2 shows that the 50% increase in the price of alternative 1 results in a reduction of its choice probability from approximately 50% to 19%, a 31% reduction. Additionally, the ADE is approximately -1.6 for the true model.

The results of the model where variable  $x_1$  is omitted are shown in the second row of Table 2. The results are concordant with the conclusions attained by Cramer (2007) and Daly (2008) about omitted orthogonal attributes in Logit models. Although this model resulted in an important change of scale (as it was noted in Table 1), the forecasting probabilities of the model are the same of the true model. This can be noted in that the ADE is statistically equal (with 95% confidence) to that obtained with the true model.

Instead, the results are very different when variable  $\xi$  is omitted. In this case there is an underestimation of approximately 10% of the change in the probability of choosing alternative 1 when its price is raised by 50%. The ADE is also significantly affected.

**Table 2 Monte Carlo Experiment: Forecasting with Endogeneity Correction**

Model	$\text{ADE}_p(1)$	$\hat{P}^0(1)$	$\hat{P}^1(1)$
<b>True Model</b>	-1.608 (0.05922)	0.5009 (0.00896)	0.1850 (0.008616)
<b>Omitting <math>x_1</math></b>	-1.600 (0.05868)	0.5013 (0.007275)	0.1871 (0.008294)
<b>Omitting <math>\xi</math></b>	-0.962 (0.04520)	0.5010 (0.007406)	0.2865 (0.01007)
<b>2SCF Adding <math>\hat{\delta}</math></b>	-1.608 (0.07726)	0.5013 (0.008182)	0.1852 (0.01076)
<b>2SCF Scale Adjustment</b>	-1.362 (0.05051)	0.5012 (0.008292)	0.2260 (0.009275)
<b>2SCF Logit Mixture</b>	-1.613 (0.07539)	0.5013 (0.007791)	0.1844 (0.01058)

Standard errors in parenthesis. 100 Repetitions.  $N=2,000$ .  $J=2$ .

Consider now the model corrected for endogeneity caused by the omission of  $\xi$  using the 2SCF method. Three forecasting methods were analyzed in this case. Table 2 shows that when the  $\hat{\delta}$  used for estimation is also included as an auxiliary variable during forecasting (Eq. 5), the results of the simulation of the 2SCF are indistinguishable from

those of the true model. In turn, when  $\hat{\delta}$  is not included in forecasting, but the scale is adjusted (Eq. 6) there is a significant bias in the forecast. In this case the effect of the price shift in the choice probabilities is underestimated by approximately 4% and the elasticity is underestimated by approximately 0.2. Instead, when the Logit Mixture method described in Eq. (8) is used for forecasting, the results for the simulated probabilities before and after the price shift are again statistically equal (with 95% confidence) to those obtained with the true model. The same occurs with the ADE.

In summary, this Monte Carlo experiment showed first that the omission of an orthogonal attribute causes a change of scale in the estimated coefficients but it does not impact the ratio between the coefficients nor the forecasting properties of the model. This same result holds also for the application of the 2SCF method in correcting for endogeneity. It was shown that the best alternative for forecasting with the 2SCF method is to include the residuals estimated in the first stage into the utility. In cases where the residuals are unavailable, they can be calculated from respective instruments using the estimators of the first stage of the 2SCF, or simulated using the expression shown in Eq. (8). Instead, the alternative of simply adjusting the scale when the residuals are omitted in forecasting was shown to have poor simulation properties.

## 6 Application to Real Data

The final step corresponds to the use of real data to demonstrate the issues investigated in this paper. Since the true scale is not known in an application with real data, only the impact in forecasting is studied in this section. The case study corresponds to a residential location choice model for the Portuguese municipalities of Lisbon, Odivelas and Amadora, which are located at the center of the Lisbon Metropolitan Area (LMA).

The data to estimate the model was constructed using the combination of two sources. The first source was a small convenience online survey (SOTUR) conducted in 2009 by Martinez et al. (2010) in the LMA. The second source corresponds to a snapshot of the dwellings that were advertised for sale in February 2007 within the LMA's municipalities of Lisbon, Odivelas and Amadora (Martinez and Viegas, 2009). The details on the construction of the database by matching both sources are described by Guevara (2010). The database used for estimation is compounded of 11,501 alternatives, from which only 63 correspond to chosen dwellings.

We are interested in modeling the choice of dwelling made by households in the LMA. We considered that this model can be well represented by a Logit where households chose the observed dwelling among the set of 11,501 alternatives available. The systematic utility is considered to be linear for the following variables: dwelling price in 100,000 Euros (€), the distance from the dwelling to the workplace of the head-of-the-household in kilometers (Km), the log of dwelling area in square meters ( $m^2$ ), and the log of dwelling age in years (+1). Dwelling price was interacted with household income, which was stratified in three levels defined by the thresholds of 2,000 and 5,000 Euros per month (€/M).

This residential choice model is very likely to suffer of endogeneity because of the omission of relevant dwelling attributes that are correlated with price. For example, the quality of the construction, the quality of the pipes, the type of neighbors, or the layout of a dwelling are relevant attributes considered in the choice process that are likely to

impact dwellings' prices. Therefore, the impossibility of measuring such attributes by the researcher should cause endogeneity.

To correct for endogeneity we use the 2SCF method. As instruments for price we consider the prices of similar dwellings located within certain geographic vicinity. The argument to sustain the suitability of such type of instruments is equivalent to the one used by Hausman (1996) and Nevo (2001) in other contexts. On the one hand, it is argued that dwellings within certain vicinity share some marginal costs, which make them correlated and therefore relevant. On the other hand, it is argued that demand shocks are independent beyond certain vicinity, what sustains their validity and allows identification. Further argumentation, formal statistical tests that suggest that such instruments are valid, and detailed results from the estimation of this model, can be found in and Guevara and Ben-Akiva (2011).

Table 3 reports the ADE (Eq. 11) for the four dwelling attributes considered in this residential location choice model. The dwelling used as reference for these calculations corresponds to the dwelling chosen by the first household in the sample. We report first the ADE that would result from a model that was not corrected for endogeneity and compare it with three alternatives to calculate the ADE in the model estimated using the 2SCF method.

**Table 3 Lisbon's Logit Model: Forecasting with Endogeneity Correction**

<i>Measure</i>		Without Endogeneity Correction	2SCF Adding $\hat{\delta}$	2SCF Scale Adjustment	2SCF Logit Mixture
ADE(1)	Price (in 100,000 €)	-3.813	-5.679	-5.457	-5.679
	Distance to Workplace (in Km)	-11.40	-13.73	-13.19	-13.73
	Log[Area (in m <sup>2</sup> )]	4.475	12.08	11.61	12.08
	Log[Age (in years)+1]	-0.3853	-0.5024	-0.4828	-0.5025

Consider first the ADE of the model that was not corrected for endogeneity, with that of the 2SCF in which the  $\hat{\delta}$  is included in forecasting phase using Eq. (5). It can be noted that the sensitivity of the model to dwelling's prices was significantly increased by the correction for price endogeneity. Interestingly, the ADE of other variables was also affected by the correction of price endogeneity, particularly that of dwelling area. This is because dwelling area and price are highly correlated (correlation = 0.7013) compared to other attributes, and then, the impact of price endogeneity is significantly transferred to dwelling area. In general, the correction for price endogeneity resulted in a model that is more sensitive, not only to changes in price, but also to changes in area, age, and distance to workplace. This demonstrates the importance of correcting for endogeneity on policy analysis. It shows that the misspecified model will significantly underestimate the impact, not only of a pricing policy, but also the impact of policies that may affect other attributes of dwelling-units.

Consider now the model that was corrected for endogeneity using the 2SCF method, but where the ADE was calculated with an adjustment of scale using Eq. (6). Concordant with what was found in the Monte Carlo experiments, it can be noted that the ADE of this

model is similar to that obtained when using Eq. (5), but there is a downward bias that makes unadvisable the use of the change of scale procedure in forecasting.

Finally, the last column of Table 3 shows that the Logit mixture approach (Eq. 9) to do forecasting results in values of ADE that are almost indistinguishable to those obtained with the model where  $\hat{\delta}$  is included in forecasting phase. This confirms that the use of a Logit mixture approach, where the conditional distribution of  $\delta$ , given  $p$ , is depicted from the sample, is a suitable tool to simulate the choices in synthetic populations.

## 7 Conclusion

In this paper, we analyzed two methodological issues that arise in the application of the control-function method for the correction of endogeneity in Logit models. The first issue analyzed was related with the change of scale derived from the application of the control-function method. Extending a result from Cramer (2007) and Daly (2008), we used Monte Carlo experimentation to show that the change of scale produced with the control-function method is harmless since it does neither affect the forecasting characteristics or the model nor the ratio of the estimators. Additionally, regarding simulation and forecasting, it was shown that a simple adjustment of scale (Wooldridge, 2002) is not suitable and that, instead, the inclusion the control-function correction in forecasting phase is better. In the case of simulation with synthetic populations, we proposed a novel method that will be relevant in applications of micro-simulation models such as UrbanSim. Lastly, the application to real data on residential location from the city of Lisbon gave further empirical evidence of the impact of endogeneity in this field and on the importance of using suitable procedures to forecast with the control-function method.

Finally, several potential lines of further research in this area can be identified. First, it would be interesting to assess the impact of the methodological advances under other circumstances, including diverse type of spatial choice models, and real databases. It would also be interesting to study the way to do forecasting with the control-function correction with other non Logit models, such as the Nested Logit. Finally, it would be interesting to analyze the impact in welfare analysis of using the control-function method to correct for endogeneity. Finally, it would be interesting to test the impact of this research in modeling complex systems, such as large urban areas. This might be achieved by applying these advancements in the framework of an operational microscopic integrated urban model such as UrbanSim (Waddell et al., 2008).

## Acknowledgments

Most of the research that resulted in this publication was performed as part of the Doctoral dissertation of C. Angelo Guevara at the CEE Department at MIT. Funding came in part from of the Portuguese Government through the Portuguese Foundation for International Cooperation in Science, Technology and Higher Education, undertaken by the MIT-Portugal Program. Additional funding came from the Martin Family Society of Fellows for Sustainability. We would like to thank Luis Martinez for the provision of the real data. All Monte Carlo and real data experiments were generated and estimated using the open-source software R (R Development Core Team, 2008).

## References

- Amemiya, T. (1978), "The Estimation of a Simultaneous Equation Generalized Probit Model," **Econometrica**, 46, 1193-1205.
- Bhat, C.R., and N. Eluru (2009), "A Copula-based Approach to Accommodate Residential Self-selection Effects in Travel Behavior Modeling," **Transportation Research Part B**, 43 (7), 749-765.
- Ben-Akiva, M. and S. Lerman (1985), *Discrete Choice Analysis, Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.
- Berry, S., J. Levinsohn and A. Pakes (1995), "Automobile Prices in Market Equilibrium," **Econometrica**, 63(4), 841-90.
- Bhat, C. and J. Guo (2004), "A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling," **Transportation Research Part B**, 38(2), 147-168.
- Chesher, A. (2010), "Instrumental Variables Models for Discrete Outcomes", **Econometrica**, 78, 575-601.
- Cramer, J. (2007), "Robustness of Logit Analysis: Unobserved Heterogeneity and Misspecified Disturbances," **Oxford Bulletin of Economics and Statistics**, 69(4), 545-555.
- Daly, A. (2008), "Elasticity, Model Scale and Error," presented at the European Transport Conference, Leeuwenhorst, The Netherlands.
- Guevara, C. (2010), "Endogeneity and Sampling of Alternatives in Spatial Choice Models," Ph.D. Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.
- Guevara, C. and M. Ben-Akiva (2011), "Testing for the Validity of Instrumental Variables in Discrete Choice Models," Working Paper, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Chile.
- Guevara, C. and M. Ben-Akiva (2006), "Endogeneity in Residential Location Choice Models," **Transportation Research Record**, 1977, 60-66.
- Hausman, J. (1996), "Valuation of New Goods under Perfect and Imperfect Competition," in **The Economics of New Goods, Studies in Income and Wealth** 58, Bresnahan and Gordon eds., Chicago: National Bureau of Economic Research.
- 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.
- Jara-Díaz, S. and C. Guevara (2003), "Behind the Subjective Value of Travel Time Savings: The Perception of Work, Leisure and Travel from a Joint Mode Choice - Activity Model," **Journal of Transport Economics and Policy**, 37, 29-46.
- Lee, L. (1982), "Specification Error in Multinomial Logit Models," **Journal of Econometrics**, 20, 197-209.
- Levine, J. (1998), "Rethinking Accessibility and Jobs-Housing Balance," **Journal of American Planning Association**, 64, 133-149.
- Martinez, L. and J. Viegas (2009), "Effects of Transportation Accessibility on Residential Property Values: A Hedonic Price Model in the Lisbon Metropolitan Area," **Transportation Research Record**, 2115, 127-137.



- Martinez, L., J. Abreu and J. Viegas (2010), "Assessment of Residential Location Satisfaction in Lisbon Metropolitan Area," presented at the 89<sup>th</sup> Transportation Research Board Annual Meeting, Washington, DC.
- Nevo, A. (2001) "Measuring Market Power In The Ready-To-Eat Cereal Industry," **Econometrica**, 69 (2), 307-342.
- Park, S. and S. Gupta (2009), "A Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data," **Journal of Marketing Research**, 46(4), 531-542.
- Petrin, A. and K. Train (2002), "Omitted Product Attributes in Discrete Choice Models," Working Paper, Department of Economics, University of California, Berkeley, CA.
- Quigley, J. (1976), "Housing Demand in the Short Run: An Analysis of Politomous Choice," **Explorations in Economic Research**, 3, 76–102.
- R Development Core Team (2008), **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>.
- Rivers, D. and Q. Vuong (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," **Journal of Econometrics**, 39, 347-366.
- Ruud, P. (1983), "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Models," **Econometrica**, 51, 225–228.
- Sermons, M. and F. Koppelman (2001), "Representing the Differences between Female and Male Commute Behavior in Residential Location Choice Models," **Geography**, 9, 101–110.
- Villas-Boas, A. and R. Winer (1999), "Endogeneity in Brand Choice Models," **Management Science**, 45, 1324–1338.
- Waddell, P. (1992), "A Multinomial Logit Model of Race and Urban Structure," **Urban Geography**, 13, 127–141.
- Waddell, P., L. Wang and X. Liu (2008), "UrbanSim: An Evolving Planning Support System for Evolving Communities," in Brail eds. **Planning Support Systems for Cities and Regions**, Lincoln Institute for Land Policy, Cambridge, MA, 103-138.
- Wooldridge, J. (2002), *Econometric Analysis of Cross-Section and Panel Data*, MIT Press, Cambridge, MA.
- Yatchew, A. and Z. Griliches (1985), "Specification Error in Probit Models," **The Review of Economics and Statistics**, 67, 134–139.
- Zimmer, D.M. and P.K. Trivedi (2006), "Using Trivariate Copulas to Model Sample Selection and Treatment Effects: Application to Family Health Care Demand," **Journal of Business and Economic Statistics**, 24, 63-76.