Modeling Discrete Choices in the Presence of Inertia and Serial Correlation

Víctor Cantillo

Department of Civil Engineering, Universidad del Norte, Km. 5 Antigua Vía a Puerto Colombia,
Barranquilla, Colombia, vcantill@uninorte.edu.co

Juan de Dios Ortúzar

Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile,
Cassila 306, Código 105, Santiago 22, Chile, jos@ing.puc.cl

Huw C. W. L. Williams

School of City and Regional Planning, Cardiff University, Cardiff CF1 3YN,
United Kingdom, williamsh@cardiff.ac.uk

The concept of habit or inertia in the context of (reluctance to) change in travel behavior has an important bearing on transport policy (e.g., how to break car use habits) and has remained an unresolved issue in demand modeling. Another major problem in modeling the response to policy measures is the potential correlation or dependence between the choices made by a given individual over time (i.e., serial correlation). The two phenomena are closely related. This paper discusses the effects of considering inertia and serial correlation on travel choices. We formulate a fairly general discrete choice model that incorporates randomly distributed inertia thresholds and allow for serial correlation. The inertia thresholds may also be a function of an individual's socioeconomic characteristics and choice conditions. The model can be applied with panel data as well as with mixed revealed and stated preference data. We applied it to real and simulated data, confirming that if these phenomena exist in the population but are not considered, serious errors in model estimation and prediction may arise, especially in the case of large policy impacts.

Key words: discrete choice models; thresholds; inertia; serial correlation; panel data; mixed and revealed
   preference data
History: Received: March 2005; revision received: March 2006; accepted: September 2006.

1. Introduction

Complex dynamic factors affect the choice processes of individuals and are perceived and appraised differently by them. It is also clear that in the face of a new situation, there is a process of experimentation and training, trial and error (Kitamura 1990). In the transport context it has been noted that daily travel patterns tend to repeat themselves in time (Pendyala, Parashat, and Muthyalagari 2000; Garling and Axhausen 2003), suggesting that individual travel may be habitual (i.e., that there are inertia effects).

Behavioral scientists view individuals as adaptive. Therefore, if the cost of searching for and constructing a new alternative is too high, or if it has associated with it too much uncertainty, people will tend to reuse past solutions to simplify their behavior and make it less risky (Payne, Bettman, and Johnson 1993; Verplanken et al. 1997). The challenge here is how to conceptualize and model the process of habit formation in travel choices.

Most research in travel behavior is based on cross-sectional revealed preference (RP) surveys. These have allowed the development of transport planning models to forecast travelers' choices among available alternatives on the basis of the individuals' socioeconomic characteristics and the attributes of the options (Ortúzar and Willumsen 2001). Because these models have generally ignored the individual's history and have avoided any direct indicators of habit, misspecification problems should be expected, because previous learning experiences are likely to affect subsequent decisions. Only the inclusion of questions related to car availability, number of trips per week, and ownership of long-term (e.g., monthly) tickets, say, has in some way attempted to capture inertia effects.

On the other hand, models derived from stated preference (SP) surveys, data panels, or mixed SP/RP surveys tend to include inertia indicators more often (Bradley and Daly 1997). A clear example is when a

currently chosen alternative is used to construct the hypothetical choice set in an SP survey. The dependence of SP on RP is similar to that established in a data panel, where it can be checked if behavior is influenced by previous choices. In these cases, another major problem is the potential serial correlation or dependence between responses for the same individual (Morikawa 1994).

Although the concept of inertia in travel choice modeling is not new (Goodwin 1977; Blase 1979), it has remained an important (although unresolved) topic because of its potential bearing on transport policy (e.g., how to reduce car dependency). Daganzo and Sheffi (1979, 1982) proposed a method to estimate a multinomial probit model using panel data; Johnson and Hensher (1982) applied this procedure to a two-period panel data set but did not consider the existence of serial correlation. Ben Akiva and Morikawa (1990) proposed a methodology for modeling switching behavior using mixed SP/RP data, but the inertia effect was confounded with the modal constants. Hirobata and Kawakami (1990) developed a bimodal model to predict traveler mode switching due to transport service changes, incorporating resistance to predict short-run changes in modal split. Later, Morikawa (1994) introduced a method to consider the classical problem of state dependence and serial correlation in mixed RP/SP data; he concluded that if these problems are ignored, discrete choice models would produce not only inefficient, but biased parameter estimates. The existence of inertia and other context-specific biases have frequently been considered by including alternative-specific constants or dummy variables representing the previous choice (Morikawa 1994, Bradley and Daly 1997).

Several studies, including Heckman (1981a, b), Heckman and Borjas (1980), and Kitamura (1990), provide evidence that ignoring state dependence can lead to misspecified models and biased results. On the other hand, empirical studies demonstrating state-dependence effects in the form of inertia have also been reported (Chang and Mahmassani 1988; Kitamura and Bunch 1990; Jou and Mahmassani 1998; Liu and Mahmassani 1998; Srinivasan and Mahmassani 2000).

In marketing science a popular method to consider inertia in the context of brand loyalty is the exponential smoothing model proposed by Guadagni and Little (1983). Swait, Adamowicz, and Van Bueren (2004) considered the inclusion of prior behavior and past attribute perceptions on current behavior in discrete choice models applied in a time-series context. In synthesis, much evidence exists to support the theory that individual behavior should be characterized by the formation of habit generation and as a consequence, reluctance to change. This means that the

same behavior can still prevail after a change in circumstances because altering it implies time and costs, both objective and psychological. Thus, if an individual q used alternative A_r with an associated net utility U_r, then according to random utility theory, $U_r = \{\text{Max } U_j, \forall A_j \in \underline{A}_{(q)}\}$, where $\underline{A}_{(q)}$ is the individual's choice set. However, a change can happen in such a way that even if the utility of another alternative becomes larger than the utility of A_r, the individual continues choosing the latter. What happens is that the person will only switch to alternative A_j if $U_j - U_r > \delta_{rj}$, where δ_{rj} is a threshold that reflects the reluctance to change, or inertia effect, associated with the transition from A_r to A_j.

In this context, we propose a model to estimate inertia as a function of the previous valuation of the alternatives, which also allows a consideration of serial correlation. The model is general and can be applied both to RP panel data and to mixed RP/SP data. We postulate that if inertia thresholds exist, they could be random, differ among individuals, and even be a function of individuals' socioeconomic characteristics and choice conditions.

The rest of the paper is organized as follows. First, we present the foundations of a method to estimate random utility choice models, including inertia effects and serial correlation; we also discuss the structure of the model and its estimation. Next, we present some empirical results of applying the model to simulated and real data and compare them with those of a classic no-inertia model. Finally, we provide a summary and some conclusions.

2. Formulation of an Inertia and Serial Correlation Model

We study the case of an individual q who regularly uses alternative A_r. At time t, an RP survey is carried out, and each member of the sample also reveals her or his available ($\underline{A}_{(q)}$) options. Afterward, at $t+1$, a certain policy that changes some attribute values is applied. We intend to evaluate the probability that a given individual switches mode (for example that he or she switches to A_i). The new condition could be evaluated using RP data again if it really happened; alternatively, the problem could be studied with data from an SP survey.

2.1. Hypotheses
For the rest of the analysis we will consider the following hypotheses:

• Individuals are net utility maximizers. In this sense, the model is in the framework of random utility theory.

• An inertia effect (habit) may exist; therefore, the individual may tend to place a higher value on the

alternative that he or she regularly uses. Notwithstanding this, some individuals may be endowed with a high disposition to change (negative inertia).

• The inertia is a function of the previous valuation of alternatives as well as of the set of objectives motivating and the conditions characterizing the choice process (e.g., purpose and schedule of the trip, if it is a frequent or occasional trip, or if the person owns long-term season tickets for public transport). Also, because of differences among individuals and the presence of errors, the inertia variable can be considered random within the population.

• The evaluation of the levels of service is different for individuals who use different alternatives; that is, perception differs among users of different transport modes.

• The responses of a given individual are correlated in time.

2.2. Model Specification

We will define $\mathbf{X}_{jq}^t$ as a $(K \times 1)$ vector of explanatory variables of alternative A_j at time t for individual q; $\mathbf{\Psi}_{jq}^t$ is a $(L \times 1)$ vector representing the set of objectives motivating and the conditions characterizing the choice of alternative A_j by individual q at time t. These could include some indicator relative to the previous choice and the socioeconomic characteristics of the individual.

According to random utility theory, the utility of alternative A_j as perceived by individual q at instant t is given by

$$U_{jq}^t = V_{jq}^t(\mathbf{X}_{jq}^t, \mathbf{\alpha}_q) + \varepsilon_{jq}^t, \qquad (1)$$

where $\mathbf{\alpha}_q$ is a $(1 \times K)$ vector of parameters. V_{jq}^t is the observable component of the utility function that can, for example, be expressed as linear in the parameters:

$$V_{jq}^t = \mathbf{\alpha}_q \mathbf{X}_{jq}^t. \qquad (2)$$

ε_{jq}^t is a random error term the specification of which defines the type of discrete choice model to use. If we have longitudinal data, an appropriate though complex formulation of the error term is

$$\varepsilon_{jq}^t = \nu_{jq} + \varsigma_j^t + \xi_{jq}^t, \qquad (3)$$

where ν_{jq} is a random variable representing an effect that is specific to the individual and invariant over time (it introduces serial correlation, e.g., reflected in a personal distaste of public transport); ς_j^t is a time-specific error component affecting all individuals equally; and ξ_{jq}^t is a purely random white-noise term. In the proposed model below, we do not consider the time-specific error, so $\varsigma_j^t = 0$, and we will assume that the white-noise term distributes independent and identically (IID) Gumbel; in addition, the serial correlation term is specified as $\nu_{jq} = \sigma_j \upsilon_{jq}$,

where υ_{jq} is a random variable standard Normal distributed and σ_j is the standard deviation of ν_j in the population.[1] Similar approaches to consider the problem of repeated observations have been proposed by Heckman (1981a, b), Ouwersloot and Rietveld (1996), and Abdel-aty, Kitamura, and Jovanis (1997), among others.

If the individual uses alternative A_r at the initial time t, then

$$P_q(A_r^t) = P(U_{rq}^t - U_{jq}^t \geq 0, \forall A_j \in \underline{A}_{(q)}). \qquad (4)$$

If the error vector[2] $\mathbf{\nu}_q$ were known, the model would correspond to a multinomial logit (MNL) formulation (Domencich and McFadden 1975):

$$\Lambda_q^t(r \mid \mathbf{\nu}_q) = \frac{\exp(V_{rq}^t + \sigma_r \upsilon_{rq})}{\sum_{A_j \in \underline{A}_{(q)}^t} \exp(V_{jq}^t + \sigma_j \upsilon_{jq})}, \qquad (5)$$

where $\Lambda^t(r \mid \mathbf{\nu}_q)$ is the probability that the choice is A_r given $\mathbf{\nu}_q$.

On the other hand, if at time $t+1$ a change happens in some of the attributes of the alternatives and we want to evaluate the probability that the individual switches from its usual choice A_r to alternative A_i, we have

$$P_q(A_i^{t+1}) = P(U_{iq}^{t+1} - U_{rq}^{t+1} \geq I_{irq}^{t+1} \text{ and } U_{iq}^{t+1} - U_{jq}^{t+1} \geq I_{irq}^{t+1}$$
$$- I_{jrq}^{t+1}, \forall A_j \in \underline{A}_{(q)}, \text{ such that } j \neq r), \qquad (6)$$

where I_{irq}^{t+1} is the resistance or inertia (threshold) to switch from A_r to A_i. The inertia value is calculated with reference to the alternative previously used (A_r).[3]

Moreover, the probability that individual q keeps using A_r is

$$P_q(A_r^{t+1}) = P(U_{rq}^{t+1} - U_{jq}^{t+1} + I_{jrq}^{t+1} \geq 0, \forall A_j \in \underline{A}_{(q)}). \qquad (7)$$

According to our hypotheses, it is possible to postulate the following general form for the inertia variable:

$$I_{irq}^{t+1} = I_{irq}^{t+1}(\mathbf{\Psi}_{rq}^t, \underbrace{(V_{rq}^t - V_{iq}^t)}_{\Delta V_{irq}^t}, \mathbf{\gamma}, \lambda_q), \qquad (8)$$

where $\mathbf{\gamma}$ is a $(1 \times L)$ vector of parameters characterizing the set of objectives motivating the choice,[4] while

---

[1] The conditional errors introducing serial correlation are not distributed Normal in a general case. We proposed independent normally distributed errors, but they can follow other kind of distributions.

[2] The dimensions of this vector depend on the number of alternatives in the set $J = /\underline{A}_{(q)}/$. The discussion presented later about the theme of identification is relevant here.

[3] Note that (6) can be written in the standard form (4) if we define $\hat{U}_{irq}^{t+1} = U_{iq}^{t+1} - I_{irq}^{t+1}$. By definition we have that $I_{rrq}^{t+1} = 0$.

[4] $\mathbf{\Psi}$ includes variables as purpose and schedule of the trip, if it is a frequent or occasional trip, if the person owns long-term season tickets for public transport, or even socioeconomic characteristics. These variables are generally specified as binary dummies.

λ_q is a parameter reflecting individual preferences. For example, we can propose the following expression for (8):

$$I_{irq}^{t+1} = \lambda_q (\boldsymbol{\gamma} \boldsymbol{\Psi}_{iq}^t + \Delta V_{irq}^t). \qquad (9)$$

Here, λ_q is an unknown coefficient that varies randomly among individuals. If it is greater than zero, inertia exists; if it is equal to zero, there is no inertia; and if it is negative, it implies that the individual has a high disposition to change. The latter may occur because the person is not satisfied with the previously chosen alternative and wants a change; it also denotes the influence or weight of the previous valuation of alternatives on the current choice. In particular, λ_q may be expressed as

$$\lambda_q = \bar{\lambda} + \zeta_q \sigma_\lambda, \qquad (10)$$

where $\bar{\lambda}$ is the population mean and the individual deviations distribute according to a density function (e.g., Normal), with mean zero and standard deviation σ_λ; ζ_q is a standard factor (e.g., standard Normal distributed).[5]

If we know $\boldsymbol{v}_q$ and ζ_q, the conditional probability (6) can be written as

$$\Lambda_q^{t+1}(i \mid \boldsymbol{v}_q, \zeta_q)$$
$$= (\exp(V_{iq}^{t+1} - (\bar{\lambda} + \zeta_q \sigma_\lambda)(\boldsymbol{\gamma} \boldsymbol{\Psi}_{rq}^t + \Delta V_{irq}^t) + \sigma_i v_{iq}))$$
$$\cdot \left(\sum_{A_j \in \underline{A}_{(q)}, j \neq r} \exp(V_{jq}^{t+1} - (\bar{\lambda} + \zeta_q \sigma_\lambda)(\boldsymbol{\gamma} \boldsymbol{\Psi}_{rq}^t + \Delta V_{jrq}^t) \right.$$
$$\left. + \sigma_j v_{jq}) + \exp(V_{rq}^{t+1} + \sigma_r v_{rq}) \right)^{-1} \qquad (11)$$

and (7) as

$$\Lambda_q^{t+1}(r \mid \boldsymbol{v}_q, \zeta_q) = (\exp(V_{rq}^{t+1} + \sigma_r v_{rq}))$$
$$\cdot \left(\sum_{A_j \in \underline{A}_{(q)}, j \neq r} \exp(V_{jq}^{t+1} - (\bar{\lambda} + \zeta_q \sigma_\lambda) \right.$$
$$\cdot (\boldsymbol{\gamma} \boldsymbol{\Psi}_{rq}^t + \Delta V_{jrq}^t) + \sigma_j v_{jq})$$
$$\left. + \exp(V_{rq}^{t+1} + \sigma_r v_{rq}) \right)^{-1}. \qquad (12)$$

2.3. Model Estimation

To estimate the above model panel data would ideally be required. In such a case, we would have information corresponding to at least two instants in time (i.e., two cross-sections) for the same group of individuals; alternatively, we could resort to mixed RP/SP survey

data. To estimate the model we first need to build a joint likelihood function for times t and $t+1$. Let us remember that because A_r is the alternative chosen by individual q, it is not the same for all individuals. The conditional choice probability for the initial time t, $\Lambda^t(r \mid \boldsymbol{v}_q)$, is calculated as defined in (5). On the other hand, for time $t+1$ the conditional choice probability is calculated as (11) or (12), depending on the chosen alternative. Then the conditional probability of a person's sequence of choices is given by a product of logit terms:[6]

$$P_q(A_r^t \wedge A_j^{t+1} \mid \boldsymbol{v}_q, \zeta_q) = \Lambda_q^t(r \mid \boldsymbol{v}_q) \Lambda_q^{t+1}(i \mid \boldsymbol{v}_q, \zeta_q). \quad (13)$$

Because $\boldsymbol{v}_q$ and ζ_q are actually unknown, the unconditional probability of the person's choice sequence of interest is of mixed logit form:

$$P_q = \int_{\boldsymbol{v}, \zeta} \Lambda_q^t(r \mid \boldsymbol{v}_q) \Lambda_q^{t+1}(i \mid \boldsymbol{v}, \zeta) \eta(\boldsymbol{v}, \zeta) \, d\zeta \, d\boldsymbol{v}, \quad (14)$$

where $\eta(\boldsymbol{v}, \zeta)$ is the joint density function of $\boldsymbol{v}$ and ζ, which we could, for example, define as a product of standard univariate Normal functions (assuming independence).[7] We can estimate (14) with an unbiased, smooth, tractable simulator, computed as

$$\hat{P}_q = \frac{1}{N} \sum_{n=1}^{N} \Lambda_q^t(r \mid \boldsymbol{v}_q^n) \Lambda_q^{t+1}(i \mid \boldsymbol{v}_q^n, \zeta_q^n), \qquad (15)$$

where $\boldsymbol{v}_q^n$ and ζ_q^n denote the nth draw from the distribution of $\boldsymbol{v}$ and ζ (Train 2003).

Proceeding as usual, we will attempt to maximize the log-likelihood function, which is given by

$$l(\boldsymbol{\theta}) = \sum_{q=1}^{Q} \text{Ln}(\hat{P}_q). \qquad (16)$$

If we had more than two waves in the panel, it would be necessary to extend the series using a similar procedure.

Note that if both waves of the panel are RP surveys, the estimation is direct. However, if the data for instant $t+1$ come from an SP survey, it would be necessary to incorporate an appropriate scale factor, such that the variances of both sets of random white-noise errors are equalled (Ben Akiva and Morikawa 1990); that is, we need a scale factor μ such that $\text{var}(\xi_{jq}^t) = \text{var}(\mu \xi_{jq}^{t+1})$. As an illustration, note that in this case

---

[5] Here we propose a generic inertia parameter. However, it is possible to specify alternative specific inertia parameters, in which case the notation would be λ_{iq}.

[6] For an individual choosing A_r at time t and A_j at time $t+1$, we have that $P_q(A_r^t \wedge A_j^{t+1}) = P_q(A_r^t) \cdot P_q(A_j^{t+1}/A_r^t)$. $P_q(A_r^t)$ is given by (5), and the conditional probability $P_q(A_j^{t+1}/A_r^t)$ by (11). Note that a model without inertia assumes independence, then $P_q(A_j^{t+1}/A_r^t) = P_q(A_j^{t+1})$.

[7] Here we assume that parameters α in the utility function are constant, but they can be also specified as random variables; in that case, however, the dimension of the integral increases. On the other hand, it is also possible to consider other types of distributions.

(12) would be expressed as

$$\Lambda_q^{t+1}(r \mid \boldsymbol{v}_q, \zeta_q) = (\exp \mu(V_{riq}^{t+1} + \sigma_r v_{rq}))$$

$$\cdot \left(\sum_{A_j \in \underline{A}_{(q)}, \, j \neq r} \exp \mu(V_{jq}^{t+1} - (\bar{\lambda} + \zeta_q \sigma_\lambda) \right.$$

$$\cdot (\boldsymbol{\gamma} \boldsymbol{\Psi}_{rq}^t + \Delta V_{jrq}^t) + \sigma_j v_{jq})$$

$$\left. + \exp \mu(V_{rq}^{t+1} + \sigma_r v_{rq}) \right)^{-1}. \quad (17)$$

It is important to examine whether it is actually possible to theoretically identify all parameters in this model. The following conclusions are obtained from an analysis of the utility functions and model structure following Walker (2002):

• The inclusion of vector $\boldsymbol{v}$ that considers agent effect parameters introduces heteroscedasticity. According to the rank and order conditions in the case of data belonging to a cross section, if we have only two alternatives, none of the variances in $\boldsymbol{v}$ can be identified, but if we have $J > 2$ alternatives, $J - 1$ variances in $\boldsymbol{v}$ can be identified and we need to normalize by constraining the term of the minimum variance option in $\boldsymbol{v}$ to zero. For a panel with two waves or responses per person, if we have two alternatives, we can estimate only one of the variances in $\boldsymbol{v}$, the normalization being arbitrary; however, if we have three or more alternatives, we can estimate all the variances.

• Because the attribute inertia term $\boldsymbol{\gamma} \boldsymbol{\Psi}_{iq}^t + \Delta V_{irq}^t$ is continuous, there is no identification issue per se concerning the inertia parameter λ, and its variance can be estimated. This is true for the specification defined here but not true in general, because the number of serial correlation parameters affects the number of state-dependence parameters that can be identified (Heckman 1981a, b). Also, it is necessary to have enough variability in the attribute across alternatives to get appropriate estimations.

3. Empirical Results

The model proposed above was applied to two data sets. The first is a simulated panel data bank and the second a real mixed RP/SP data set. To estimate the model we wrote a special code using the Maxlik library in GAUSS[8] (Aptech Systems 1994).

3.1. Application to Simulated Data

To examine the performance of the proposed model for a population where inertia and serial correlation exist, we followed the classical procedure of Williams and Ortúzar (1982) and generated a simulated data bank with three hypothetical alternatives—*taxi*, *bus*,

and *metro*—and three attributes—*cost*, *travel time*, and *access time*. A total of 10,000 individuals were simulated, and for each one we generated two waves: the first one corresponded to $t = 0$ and the other to $t = 1$; therefore, we generated 20,000 observations. As mentioned, the purely random errors were assumed to be IID Gumbel.

To generate the data, we considered two factors: the existence of inertia at $t = 1$, as a function of the valuation of alternatives at $t = 0$, and the existence of serial correlation for the alternatives *taxi* and *bus*.

The parameters used to generate the data are shown in Table 1, including the "true" (target) parameters of the utility function; for the serial correlation and inertia variables we assumed Normal distributions. In the table, we also present the results of estimating two types of models. The first is a classical MNL model without considering serial correlation or inertia, the second (MNLID) is an MNL model including at $t = 1$ the choice at $t = 0$ as an inertia dummy variable (Morikawa 1994, Bradley and Daly 1997), and the third is the proposed model that incorporates inertia and serial correlation. In parentheses two t-tests appear: The first corresponds to the traditional null hypothesis $\theta_k = 0$, and the second refers to the null hypothesis $\theta_k = \theta_v$, where θ_v are the target values.

The t-tests in Table 1 allow us to conclude that the MNL parameters for all attributes and also the scale parameter are significantly different from their targets at the 95% level; the same happens for the MNLID model, although its likelihood is better (an increase of 32 for three additional parameters). However, in the case of the proposed model, we note a reasonable recovery of the parameters in the utility function and of those relative to serial correlation. The null hypothesis $\theta_k = \theta_v$ is only rejected at the 95% level for the mean and standard deviation of the inertia variable, which are overestimated. The improvement in fit of the proposed model is noteworthy. In fact, the likelihood ratio test[9] LR = 419.2 is much larger than the critical value $\chi^2_{95\%, \, 4df} = 9.49$. Therefore, we conclude that the restricted model (MNL) is not acceptable, although if it were judged on the basis of all the usual goodness-of-fit criteria, it would be labeled a reasonable model. The ρ^2 values are modest but fairly typical for discrete choice models.

If we analyze the implied subjective values of time in Table 2, we note that in the proposed model the targets are recovered well within their confidence intervals (Armstrong, Garrido, Ortúzar 2001). Meanwhile,

---

[8] The optimization algorithms used were BFGS, BHHH, and Newton Raphson, or a combination of these.

[9] The LR statistic $-2\{l^*(\theta_r) - l^*(\theta)\}$, where $l^*(\theta)$ and $l^*(\theta_r)$ are the log-likelihood values at convergence for the nonrestricted and restricted models, respectively, is asymptotically distributed χ^2 with r degrees of freedom, where r is the number of linear restrictions (Ortúzar and Willumsen 2001, p. 262).

Table 1   **Model Parameters for the Simulated Database**

Parameter	Parameter target	MNL	MNLID	Proposed model
Cost	−0.06	−0.0491 (−58.54; 13.63)	−0.0504 (−53.84; 10.67)	−0.0617 (−13.14; 0.36)
Travel time	−0.12	−0.0878 (−60.45; 21.47)	−0.0873 (−57.07; 21.80)	−0.1193 (−13.38; 0.08)
Access time	−0.18	−0.1555 (−55.81; 8.75)	−0.1550 (−54.39; 8.62)	−0.1803 (−15.37; 0.03)
Serial correlation for taxi (standard deviation)	1.0			1.2506 (4.97; 1.00)
Serial correlation for bus (standard deviation)	2.0			1.9699 (8.80; 0.13)
Inertia mean $\bar{\lambda}$	0.40			0.5467 (12.86; 3.45)
Inertia standard deviation (σ_λ)	0.30			0.6075 (8.20; 4.15)
Inertia dummy for taxi			0.3002 (7.23)	
Inertia dummy for bus			0.0325 (0.769)	
Inertia dummy for train			0.0995 (2.54)	
Sample size		20,000	20,000	20,000
Log likelihood		−18,502.5	−18,470.8	−18,292.9
ρ^2		0.158	0.159	0.168

the estimates from the MNL are clearly outside them and in the case of the MNLID, only the target relative to *access time* is within its confidence interval.

To examine the practical consequences of adopting models with and without inertia and serial correlation when there is evidence of their presence, we also tested the performance of the MNLID and the proposed model in terms of their response capabilities. We used scenarios representing "policy changes," which ranged from slightly to quite different from the base data used for estimation, simulating a panel with two waves. This entailed changing attribute values for the options and reexecuting the choice simulation procedure. We generated a series of simulated future scenarios that could be compared with the models' predictions.

Doing an analysis similar to Munizaga, Heydecker, and Ortúzar (2000), we tested six policies, P1 to P6. Table 3 shows the specific percentage changes to the *cost*, *travel time*, and *access time* attributes associated with each policy. Policies P1–P3 correspond to small

changes; in contrast, P4–P6 represent large policy changes. Initial market shares were 0.378, 0.276, and 0.346 for *taxi*, *bus*, and *metro*, respectively.

The error measure considered was the percentage difference between observed or "true" behavior (i.e., that simulated for the modified attribute values) and that estimated by the models. The minimum response error that can be considered a prediction error is given by the standard deviation of the simulated observations generated with different seeds, as it reflects the inherent variability of the simulation process. We performed 15 repetitions of the data-generation process using different seeds for the pseudorandom numbers and found, reassuringly, that the percentage difference tends to increase with decreasing market share. The largest value of the coefficient of variation was 7.7%; given these results, and considering that in some policies we expected smaller market shares, we decided to take 10% as a reasonable threshold. Therefore, any discrepancy exceeding this value was considered an estimation error.

Table 2   **Subjective Values of Time and Confidence Intervals for Simulated Data**

		MNL		MNLID		Proposed model	
SVT	Target	Point estimate	Confidence interval	Point estimate	Confidence interval	Point estimate	Confidence interval
Travel time	2.00	1.79	1.73–1.85	1.73	1.67–1.80	1.94	1.84–2.03
Access time	3.00	3.17	3.08–3.25	3.08	2.99–3.16	2.93	2.83–3.04

Table 3   Policy Changes: Percentage Change in Attribute Values

Policy	Cost bus	Cost metro	Travel time bus	Travel time metro	Access time taxi	Access time metro
P1			+20			
P2		−15				
P3					+25	
P4			+60			
P5		−50				
P6	+50	+50	−40	−50		−40

To test goodness of fit we used a chi-squared test given by

$$\chi^2 = \sum_i \frac{(\widehat{N}_i - N_i)^2}{N_i},\qquad(18)$$

where $\widehat{N}_i$ is the model estimate of the number of individuals choosing alternative A_i, and N_i is the actual (simulated) number. The result should be contrasted with the critical χ^2 value at the 5% level with $J-1=2$ degrees of freedom (i.e., 5.99).

Table 4 shows that the proposed model yields superior results. The MNLID model produces response errors outside the thresholds in all cases except P3; meanwhile, the percentage errors in the proposed model are always within the threshold (although in P6, errors are close to the limit). A similar analysis using the χ^2 index shows that the errors for the proposed model are not statistically significant at the 5% level for all policies except P6, where the changes are dramatic. Conversely, the errors of the MNLID model are significant in all policies except P2 and P3. These results allow us to conclude that if there are inertia and serial correlation effects, a misspecified model might lead to significant response errors, especially in the case of large policy impacts.

To evaluate the effect of sample size on model estimation, we generated simulated data with only 2,000 draws; the results are presented in Table 5. In this case, the recovery of the true parameters is similar to the previous experiment with a larger sample size, but the improvements in log-likelihood are less pronounced. The χ^2 test shows that the proposed model can be distinguished from the MNL at the 5% level. It can be noted also that the standard deviation of the serial correlation for taxi is not significantly different from zero; furthermore, the null hypothesis $\theta_k = \theta_v$ for the inertia standard deviation is rejected at the 95% level. This confirms the requirement for adequate estimation sample sizes, especially when dealing with complex models (Munizaga, Heydecker, and Ortúzar 2000).

To verify the incidence of the levels of inertia and serial correlation, we generated simulated data combining high and low levels in both cases. The results of these simulations allow us to conclude that in all cases, the parameters of the utility function are

well recovered, but the inertia and serial correlation parameters are not always significant. In fact, when one effect is strong and the other weak, there is dominance, and the latter parameters are not significant.

3.2. Application to a Mixed SP/RP Database
These data came from a mixed RP and SP survey carried out in Cagliari, Italy. The RP data involved the choice among car, bus, and train; the SP data considered the binary choice between a new train service (quicker, more frequent, and with more stations than the current one) and the alternative currently chosen by users (i.e., car or bus). It was assumed that current train users would also prefer the much-improved (in all attributes, including the fare) new service, so they were not considered in the SP survey (Cherchi and Ortúzar 2002).

The database includes general and socioeconomic information, such as family income, gender, trip purpose,[10] motorization, and family structure. In addition, the variables included in the survey were in-vehicle and walking time (for bus, train, and car); waiting time (for the public transport (PT) modes only); frequency (for the PT modes, it is equal to the number of buses/trains that pass in one hour); number of transfers (for PT modes); in-vehicle cost (for bus, train, and car), and comfort.[11]

After many tests a final sample of 310 individuals and 1,998 observations was selected for model estimation, including 110 individuals who only responded to the RP survey. To examine the practical significance of the proposed model in terms of improved predictive ability, prior to model estimation we selected a hold-out validation sample. Its size (160 observations[12]) was computed using the method described by Ortúzar and Willumsen (2001, p. 267). The remaining observations (1,838) were used to estimate the models.

The estimation results are shown in Table 6, including the traditional mixed RP/SP nested logit (NL) with a scale parameter and two versions of the proposed model considering serial correlation for train

[10] Following the recommendation of one referee, we included trip purpose in the inertia term as a motivating objective. The dummy was included in the car option and took the value 1 if the trip was compulsory (work or study) and 0 otherwise.

[11] The user was asked to provide a simple judgment of the comfort experienced during the journey described. The variable was precoded into three levels—poor, sufficient, and good—and two dummy variables were used; *Comf 1* equals 1 if the level of comfort was poor and 0 otherwise, and *Comf 2* equals 1 if comfort was sufficient and 0 otherwise. The "good" level was left as reference because the comfort variable was only introduced in the PT alternatives; therefore, it was implicitly assumed that the car had high comfort.

[12] These observations were mixed RP/SP data.

Table 4   **Comparison of Simulated and Modeled Forecasts**

	Targets			MNLID				Proposed model			
Policy	Taxi	Bus	Metro	Taxi	Bus	Metro	χ^2	Taxi	Bus	Metro	χ^2
P1	396	222	382	431 8.8%	183 −17.8%	387 1.3%	10.1	404 2.0%	204 −8.4%	392 2.8%	2.0
P2	364	263	373	379 4.0%	232 −11.7%	389 4.4%	4.9	362 −0.8%	247 −6.2%	392 5.1%	2.0
P3	343	272	385	342 −0.2%	267 −1.9%	391 1.5%	0.2	335 −2.3%	271 −0.3%	394 2.3%	0.4
P4	435	141	424	471 8.2%	103 −26.7%	426 0.5%	13.3	439 0.8%	128 −9.1%	434 2.2%	1.4
P5	330	241	429	330 −0.1%	202 −16.1%	469 9.1%	9.9	319 −3.5%	222 −7.6%	459 7.0%	3.8
P6	243	258	499	229 −5.9%	203 −21.1%	568 13.8%	21.4	227 −6.8%	232 −10.0%	541 8.5%	6.9

and inertia.[13] They differ in that the second one includes a trip purpose dummy variable in the inertia term as a choice motivating objective. It can be noted that in all three models the signs are consistent.

Although our proposed models fit the data better (the LR test for both models, 9.2 and 14.0, respectively, are significantly larger than the critical values $\chi^2_{95\%,\,3df} = 7.81$ and $\chi^2_{95\%,\,4df} = 9.49$, respectively), we note that the serial correlation parameter and the mean of the inertia variable are not significant. Nevertheless, the standard deviation of inertia is significant, indicating that although there is inertia for some individuals, others seem to have a high disposition to change. Thus, interestingly, although the drag effect of inertia was expected to dominate, in fact those individuals for whom inertia exerts the expected drag effect on change are approximately balanced by those who appear to be hyper-ready for change. However, this fact deserves greater discussion; given the well-known biases inherent in SP surveys, it seems quite likely that in general an RP/SP model will underestimate the mean of the inertia effect and overestimate its variance because it will show a lot more apparent change (many more people overcoming inertia) than would actually be the case.[14]

On the other hand, it is interesting to note that the scale parameters in the proposed models are less significant but closer to one (in proposed model 2 it is greater than one, suggesting less variance for the SP survey), meaning that the variances of the purely random terms are more similar in this case than in the mixed RP/SP NL model. The trip purpose variable included in inertia as motivating objective was significant, and its sign suggests that the inertia for

car decreases for compulsory trips; however, the log-likelihood improved only slightly in this case.

Unfortunately, we did not have a variety of additional variables to incorporate in the inertia specification; we tried several model specifications, but results did not improve. Although we did not find evidence of multicolinearity between factors affecting the inertial terms and other utility terms (the SP survey came from an orthogonal design) in the correlation matrix, it is possible that other factors—such as endogeneity due to measurement errors in the inertial variables and other error sources—may have an incidence in the significance of the parameters.

Observed differences in the coefficients of variables in the utility function are not very informative, because they may be due to different scale factors; it is more interesting to examine coefficient ratios. As an illustration, note that the ratios between the travel time and cost/income coefficients (i.e., the subjective values of time) are 2.59, 2.96, and 2.46, respectively, for the models in the order they were presented. On the other hand, it may be noted that some variables, such as transfer, lose significance in the proposed models.

Finally, we used the hold-out validation sample to examine the practical consequences of adopting models with and without inertia and serial correlation in forecasting; for this we contrasted the mixed RP/SP NL models versus our proposed model 2. Note first that if we use a direct likelihood ratio test (Ortúzar and Willumsen 2001, p. 271) to discriminate between both models (as they are nonnested), we get

$$\frac{L_{Proposed\ model\ 2}}{L_{Mixed\ RP/SP\ NL}} = 2.8;$$

that is, the data are approximately three times more probable under the proposed model 2 than under the mixed RP/SP NL model. The forecast results for the validation sample (Table 7) imply that the

---

[13] Cherchi and Ortúzar (2002) also tried an inertia variable following the approach of Bradley and Daly (1997), but it was clearly not significantly different from zero.

[14] We are grateful to a referee for having made this clearer.

Table 5   Model Parameters for a Smaller Simulated Database

Parameter	Parameter target	MNL	MNLID	Proposed model
Cost	−0.06	−0.465 (−17.74; 5.19)	−0.0484 (−16.57; 4.00)	−0.0535 (−4.91; 0.60)
Travel time	−0.12	−0.0853 (−19.02; 7.71)	−0.0860 (−18.11; 7.23)	−0.1050 (−4.88; 0.70)
Access time	−0.18	−0.1466 (−16.96; 3.88)	−0.1462 (−16.60; 3.84)	−0.1574 (−5.74; 0.82)
Serial correlation for taxi (standard deviation)	1.0			0.9501 (1.34; 0.07)
Serial correlation for bus (standard deviation)	2.0			1.6918 (2.82; 0.51)
Inertia mean $\bar{\lambda}$	0.40			0.6704 (4.43; 1.79)
Inertia standard deviation (σ_λ)	0.30			0.8138 (3.44; 2.17)
Inertia dummy for taxi			0.3493 (2.68)	
Inertia dummy for bus			0.0650 (0.47)	
Inertia dummy for train			0.0068 (0.096)	
Sample size		2,000	2,000	2,000
Log likelihood		−1,870.4	−1,866.6	−1,846.9
ρ^2		0.149	0.151	0.160

performance of both models is both excellent and indistinguishable in this case.

As with simulated data, we tested four policies to compare models with and without inertia and serial correlation with real data. Table 8 shows the specific percentage changes to the variables cost/income, walking time, and travel time; these were applied to all 310 RP surveys. It can be noted that changes tend to stimulate the use of public transport.

In this case it is not possible to compare our modeled results with the true responses, only to compare both estimations. Results are presented in Table 9; it is important to recall that initially we had 172 car users, 125 bus users, and only 13 (old) train users.

It can be noted that in the proposed model, individuals show more disposition to change toward the train, reflecting the negative sign of inertia. The mixed RP/SP NL model only reflects a more significant change in modal split for the very strong P4 policy; but even in this case the proposed model is more sensitive to changes.

4.   Conclusions

When working with a panel data set or with mixed RP/SP data, it is necessary to consider that inertia and serial correlation effects may exist. We propose a model allowing for inertia thresholds that can be random, that can differ among individuals, and that can even be a function of their socioeconomic characteristics, choice conditions, and the previous valuation of alternatives. The estimation of the model is fairly simple using maximum simulated likelihood.

The model was applied to real and simulated data, and our results add more evidence to that provided in previous work (Williams and Ortúzar 1982; Heckman 1981a, b; Kitamura 1990; Chang and Mahmassani 1988, among others) in the sense that if there are inertia and serial correlation effects, a misspecified model might lead to bias in coefficient estimates and produce significant response errors, especially in the case of large policy impacts. However, we also found that for modest adjustments in the policy variables, the potential bias in the parameter estimates of standard models are relatively minor. On another line, when the proposed model was applied to a mixed RP/SP data set, the scale parameter became closer to one than in a traditional NL model, suggesting that the variances of the purely random terms are more similar when these phenomena are included. This type of finding has also been reported by Smith (2005).

There are several aspects of interest for future research. One is to evaluate in more depth the impact of inertia and serial correlation on the estimation of subjective values of time and other marginal rates of substitution, including their microeconomic implications; their influence on model elasticities also remains to be examined. In addition, issues of multicolinearity between factors affecting the inertial terms and

Table 6   Models with Mixed RP/SP Database

Parameter	Mixed RP/SP NL	Proposed model 1	Proposed model 2
Cost/income	−0.0152	−0.0141	−0.0135
	(−2.58)	(−2.50)	(−2.43)
Walking time	−0.0597	−0.0555	−0.0537
	(−3.40)	(−3.21)	(−3.49)
Travel time	−0.0394	−0.0418	−0.0332
	(−2.78)	(−2.83)	(−2.40)
Transfer (bus and train)	−0.7661	−0.4308	−0.6530
	(−2.63)	(−1.16)	(−1.49)
Frequency (bus and train)	0.2141	0.2056	0.1964
	(2.91)	(2.85)	(3.10)
Comfort 1	−2.4221	−2.7521	−2.1597
	(−4.41)	(−4.26)	(−4.07)
Comfort 2	−1.2506	−1.8526	−1.2272
	(−3.58)	(−3.64)	(−3.38)
Car/licenses (car in RP)	3.9455	3.4946	4.0429
	(3.52)	(2.67)	(4.03)
K_train RP	−2.8280	−3.1352	−2.8266
	(−6.22)	(−5.63)	(−5.53)
K_train SP	0.6970	1.2236	0.4556
	(1.62)	(1.49)	(1.04)
K_car RP	−1.7892	−1.8578	−1.8001
	(−1.98)	(−2.00)	(−2.21)
K_car SP	0.2558	−0.5827	0.2665
	(0.45)	(−0.60)	(0.52)
Serial correlation car (standard deviation)	—	0.1704	0.0336
		(0.12)	(0.06)
Scale factor μ	0.4049	0.6623	1.3100
	(3.94)	(2.28)	(1.67)
Purpose (inertia car)			−2.0004
			(−4.04)
Mean of inertia $\bar{\lambda}$	—	0.1983	−0.0955
		(0.80)	(−0.79)
Standard deviation of inertia (σ_λ)	—	1.0695	1.4073
		(3.54)	(3.96)
Sample size	1,838	1,838	1,838
Log likelihood	−1,078.0	−1,073.4	1,071.0

Table 7   Comparison of Observed Choices and Modeled Forecasts for Validation Sample

Targets			Mixed RP/SP NL				Proposed model 2			
Car	Bus	Train	Car	Bus	Train	χ^2	Car	Bus	Train	χ^2
70	35	57	68	41	51	1, 72	69	40	51	1, 36
			−3%	17%	−11%		−1%	14%	−11%	

Table 8   Percentage Changes in Attribute Values for Real Database

Policy	Cost/income Train	Waiting time Train	Cost/income Car
P1	−10		
P2	−30		
P3		−30	
P4		−30	+100

Table 9   Comparison of Forecasts Using Real Data

	Mixed RP/SP NL			Proposed model 2		
Policy	Car	Bus	Train	Car	Bus	Train
P1	172	124	14	170	123	17
P2	172	123	15	167	121	22
P3	171	121	18	165	119	26
P4	153	136	21	141	139	30

endogeneity due to measurement errors in inertial variables and other error sources should also be considered, because they can affect the significance of the estimated parameters. Finally, we recognize that the proposed model needs further empirical analysis, in different choice contexts, to establish its widespread value; unfortunately, appropriate panel data are very difficult to find.

Acknowledgments

References

Abdel-Aty, M. A., R. Kitamura, P. P. Jovanis. 1997. Using stated preference data for studying the effect of advanced traffic information on driver's route choice. *Transportation Res. Part C* **5** 35–50.

Aptech Systems. 1994. *GAUSS User's Manuals.* Aptech Systems Inc., Black Diamond, WA.

Armstrong, P. M., R. A. Garrido, J. de D. Ortúzar. 2001. Confidence intervals to bound the value of time. *Transportation Res. Part E* **37** 143–161.

Ben Akiva, M., T. Morikawa. 1990. Estimation of switching models from revealed preferences and stated intentions. *Transportation Res. Part A* **24** 485–495.

Blase, J. H. 1979. Hysteresis and catastrophe theory: A demonstration of habit and thresholds in travel behaviour. *Proc. 7th PTRC Summer Annual Meeting* (July), University of Warwick, England.

Bradley, M. A., A. J. Daly. 1997. Estimation of logit models using mixed stated preference and revealed preference information. P. R. Stopher, M. Lee-Gosselin, eds. *Understanding Travel Behaviour in an Era of Change.* Pergamon Press, Oxford.

Chang, G. L., H. S. Mahmassani. 1988. Travel time prediction and departure time adjustment behaviour dynamics in a traffic system. *Transportation Res. Part B* **22** 217–232.

Cherchi, E., J. de D. Ortúzar. 2002. Mixed RP/SP models incorporating interaction effects: Modelling new suburban train services in Cagliari. *Transportation* **29** 371–395.

Daganzo, C. F., Y. Sheffi. 1979. Estimation of choice models from panel data. *26th Annual Meeting Regional Sci. Assoc.* (November), Los Angeles, CA.

Daganzo, C. F., Y. Sheffi. 1982. Multinomial probit with time-series data: Unifying state dependence and serial correlation models. *Environment Planning* **14A** 1377–1388.

Domencich, T., D. Mcfadden. 1975. *Urban Travel Demand: A Behavioural Analysis*. North Holland, Amsterdam, The Netherlands.

Garling, T., K. Axhausen. 2003. Introduction: Habitual travel choice. *Transportation* **30** 1–11.

Goodwin, P. 1977. Habit and hysteresis in mode choice. *Urban Stud.* **14** 95–98.

Guadagni, P., J. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2** 203–238.

Heckman, J. J. 1981a. Statistical models for discrete panel data. C. F. Manski, D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA.

Heckman, J. J. 1981b. Heterogeneity and state dependence. S. Rosen, ed. *Studies in Labour Markets*. University of Chicago Press, Chicago, IL.

Heckman, J. J., G. Borjas. 1980. Does employment cause future employment? Definitions, questions and answers from a continuous time model for heterogeneity and state-dependence. *Econometrica* **47** 247–283.

Hirobata, Y., S. Kawakami. 1990. Modelling disaggregate behavioural modal switching models based on intention data. *Transportation Res. Part B* **24** 15–25.

Johnson, L., D. Hensher. 1982. Application of multinomial Probit to a two-period panel data set. *Transportation Res. Part A* **16** 457–464.

Jou, R.-C., H. S. Mahmassani. 1998. Day-to-day dynamics of urban commuter departure time and route switching decisions: Joint model estimation. J. de D. Ortúzar, D. Hensher, S. R. Jara-Diaz, eds. *Travel Behaviour Research: Updating the State of Play*. Elsevier Science, Amsterdam.

Kitamura, R. 1990. Panel analysis in transportation planning: An overview. *Transportation Res. Part A* **24** 401–415.

Kitamura, R., D. Bunch. 1990. Heterogeneity and state dependence in household car ownership: Panel analysis with ordered response Probit models with error components. M. Koshi, ed. *Transportation and Traffic Theory*. Elsevier, New York.

Liu, Y. H., H. S. Mahmassani. 1998. Dynamic aspects of commuter decisions under advance traveller information systems: Modelling framework and experimental results. *Transportation Res. Record* **1645** 111–119.

Morikawa, T. 1994. Correcting state dependence and serial correlation in the RP/SP combined estimation method. *Transportation* **21** 153–165.

Munizaga, M. A., B. G. Heydecker, J. de D. Ortúzar. 2000. Representation of heteroscedasticity in discrete choice models. *Transportation Res. Part B* **34** 219–240.

Ortúzar, J. de D., L. G. Willumsen. 2001. *Modelling Transport*, 3rd ed. John Wiley and Sons, Chichester.

Ouwersloot, H., P. Rietveld. 1996. Stated choice experiments with repeated observations. *J. Transport Econom. Policy* **30** 203–212.

Payne, J., J. Bettman, E. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge University Press, New York.

Pendyala, R. M., A. Parashat, G. R. Muthyalagari. 2000. Measuring day-to day variability in travel characteristics using GPS data. *79th Annual Meeting Transportation Res. Board* (January), Washington, D.C.

Smith, M. D. 2005. State dependence and heterogeneity in fishing location choice. *J. Environ. Econom. Management* **50** 319–340.

Srinivasan, K. K., H. S. Mahmassani. 2000. Modelling inertia and compliance mechanisms in route choice behaviour under real-time information. *Transportation Res. Record* **1725** 45–53.

Swait, J. D., W. Adamowicz, M. Van Bueren. 2004. Choice and temporal welfare impacts: Incorporating history into discrete choice models. *J. Environ. Econom. Management* **47** 94–116.

Train, K. E. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.

Verplanken, B., H. Aarts, A. Van Knippenberg, C. Van Knippenberg. 1997. Habit, information acquisition and the process of making travel mode choices. *Eur. J. Soc. Psych.* **27** 539–560.

Walker, J. L. 2002. Mixed logit (or logit kernel) model: Dispelling misconceptions of identification. *Transportation Res. Record* **1805** 86–98.

Williams, H. C. W. L., J. de D. Ortúzar. 1982. Behavioural theories of dispersion and the misspecification of travel demand models. *Transportation Res. Part B* **16** 167–219.