6 Random versus Fixed Coefficient Quantal Choice Models

Gregory W. Fischer and Daniel Nagin

6.1 Introduction

A large class of decision problems is most appropriately characterized as a choice dilemma in which an individual selects one element from a discrete set of decision alternatives. Examples of such dilemmas are almost limitless, ranging from the relatively profound (selection of a job or a spouse) to the relatively mundane (selection of a brand of toothpaste or shaving cream). Quantal choice theory attempts to explain and predict the behavior of individuals confronted by such decision dilemmas.[1] To date, quantal choice models have been most widely employed in transportation demand studies. Here the models have been used to examine such issues as the selection of transportation modes for home-to-work trips (e.g., Ben-Akiva and Haus 1973, Charles Rivers Associates 1972) and automobile ownership decisions (Lerman and Ben-Akiva 1975). A key motivation for all of these studies has been the development of models that predict the impact of changes in certain factors which can be manipulated by policy makers (e.g., gasoline prices, travel time). In view of the potential significance of these applications of quantal choice models, it is essential that the statistical procedures used in developing these models lead to accurate inferences and predictions.

6.2 Quantal Choice Theory and Variation in Tastes

Notationally let $\mathbf{R} = (\mathbf{X}^1, \ldots, \mathbf{X}^r)$ be a mutually exclusive and exhaustive set of r choice alternatives, where each alternative is characterized by a vector of m value relevant attributes. That is, $\mathbf{X}^i = (x_1^i, \ldots, x_m^i)$. For example, if the alternatives are transportation modes, the attribute vector might include such factors as time per trip, cost per trip, and so forth. Further let $\mathbf{Y}^j = (y_1^j, \ldots, y_n^j)$ be a vector of attributes characterizing the jth individual choosing from choice set $\mathbf{R}$. These might include the

This research was supported by a grant from the Duke University Research Council. The authors are indebted to Charles Manski for numerous insights which have greatly enhanced our analysis. The observations of several anonymous reviewers were also of substantial value in preparing the final draft. Any errors remain the responsibility of the authors alone.

1. For general historical reviews of quantal choice theory see Bock and Jones (1968) and Luce and Suppes (1965). For more recent developments in quantal choice theory see Hausman and Wise (1978), McFadden (1973), Manski (1973), and Tversky (1972).

individual's age, sex, income, and so forth. Then for any such choice set **R**, and for any individual described by the set of attributes $\mathbf{Y}^j$, quantal choice models generate a vector of choice probabilities $(P^{1j}, \ldots, P^{rj})$. Here P^{ij} is the probability that a person characterized by $\mathbf{Y}^j$ will choose alternative $\mathbf{X}^i$ from choice set **R**. Thus $\Sigma_{i=1}^r P^{ij} = 1$.

The models considered here assume that each individual is a utility maximizer. Thus $P^{ij} = Pr(U^{ij} \geq U^{kj}$, for all $k \neq i)$, where U^{ij} is the subjective utility of alternative $\mathbf{X}^i$ to the jth individual. Statistically the models considered here assume that U^{ij} is a linear function of the attributes of alternative $\mathbf{X}^i$ and the individual's attributes, $\mathbf{Y}^j$. More precisely let $\mathbf{Z}^{ij} = (z_1^{ij}, \ldots, z_s^{ij})$ be a vector of arithmetic combinations of $\mathbf{X}^i$ and $\mathbf{Y}^j$. This **Z**-vector might include simple attributes (say, income or price per trip), transformations of attributes (say, the log of income or price), or explicit interactions of the attributes of the alternatives and the individual (e.g., price/income).

The first statistical specification we consider here is the linear in parameters, independent and identically distributed (LPIID) disturbances model

$$U^{ij} = \mathbf{Z}^{ij}\boldsymbol{\beta} + \varepsilon^{ij}. \tag{6.1}$$

Here U^{ij} is the unobservable utility of the ith alternative to the jth individual. This utility is assumed to be linear in the elements of $\mathbf{Z}^{ij}$. Thus the coefficient vector $\boldsymbol{\beta}$ reflects the tastes of the individuals in the population. Random variation in the U^{ij} is introduced through the additive disturbance term ε^{ij} which is assumed to be independently and identically distributed across individuals and alternatives. Manski (1973) provides an insightful analysis of the possible sources of apparent random variation in behavior. For our purposes two relatively straightforward interpretations seem adequate. First the ε^{ij} may arise due to choice relevant but unobserved attributes of alternatives or individuals. Such factors are necessarily excluded from the attribute vector $\mathbf{Z}^{ij}$ and are, with the LPIID specification, assumed to be independent of the elements of $\mathbf{Z}^{ij}$. A second (and not incompatible) possibility is that the ε^{ij} reflect true random variation in choice behavior, an interpretation that is commonly invoked by psychologists (see Coombs, Dawes, and Tversky 1970).

Although the U^{ij} are unobservable, the coefficients of equation (6.1) are estimable. Consider a binary choice situation in which an individual described by the vector $\mathbf{Y}^j$ is choosing between two alternatives described by the vectors $\mathbf{X}^1$ and $\mathbf{X}^2$. Then $\mathbf{Z}^{1j}$ and $\mathbf{Z}^{2j}$ are the vectors of the appropriate arithmetic combinations of the attributes of the alternatives and of the individual choosing between them. According to equation (6.1) the probability that the individual will choose alternative $\mathbf{X}^1$ is given by

$$\begin{aligned} Pr(U^{1j} \geq U^{2j}) &= Pr[\mathbf{Z}^{1j}\boldsymbol{\beta} + \varepsilon^{1j} \geq \mathbf{Z}^{2j}\boldsymbol{\beta} + \varepsilon^{2j}] \\ &= Pr[(\mathbf{Z}^{1j} - \mathbf{Z}^{2j})\boldsymbol{\beta} \geq \varepsilon^{2j} - \varepsilon^{1j}]. \end{aligned}$$

Thus if the ε^{ij} are assumed to be normally distributed, and if a sample of binary choices has been observed, $\boldsymbol{\beta}$ can be estimated using IID probit estimation procedures (Albright, Lerman, and Manski 1977, Hausman and Wise 1978); if the ε are assumed to be extreme value distributed, then β can be estimated using IID logit estimation methods (McFadden 1972). Here we consider only the probit specification of the LPIID model.

Note that the LPIID model assumes that all individuals' tastes are identical with respect to the observed attributes embodied in the $\mathbf{Z}^{ij}$ vectors. Consequently the LPIID formulation implies that all individuals of identical observed characteristics have identical tastes with respect to the observed attributes of alternatives (except for random additive disturbances). Empirically, this need not be the case. For example, two individuals of identical observed characteristics may attach different disutilities to price per trip and time per trip when making transportation mode choices. If this is the case, the specification in equation (6.1) is incorrect, for apparently random variation in behavior is due not only to an additive disturbance effect but also to variation in tastes, that is, to variation in the elements of $\boldsymbol{\beta}$.

Recognizing this difficulty, Hausman and Wise (1978), and Albright, Lerman and Manski (1977) have developed the random coefficients, covarying disturbances (RCCD) model, a quantal choice model that explicitly incorporates variation in the tastes of individuals with identical observed characteristics. The RCCD model is given by

$$U^{*ij} = \mathbf{Z}^{ij}\boldsymbol{\beta}^* + \mathbf{Z}^{ij}\boldsymbol{\delta}^j + \gamma^{ij}, \tag{6.2}$$

where $\boldsymbol{\beta}^*$ is the mean coefficient vector for the population of interest, $\boldsymbol{\delta}^j$ is a coefficient vector describing the deviations of the jth individual's tastes from the tastes embodied in the mean coefficient vector, and γ^{ij} is an additive disturbance term assumed to be independently and identically distributed across individuals but possibly correlated across alternatives. In general RCCD models assume that the $\boldsymbol{\delta}^j$ and $\gamma^j = (\gamma^{1j}, \ldots, \gamma^{rj})$ are multivariate normally distributed with $\boldsymbol{\delta}^j \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\gamma^j \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$. The parameters $\boldsymbol{\beta}^*$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Sigma}_\gamma$ are not jointly identified. The necessary identification restrictions are generally imposed on the components of $\boldsymbol{\Sigma}_\gamma$, and the remaining parameters are estimated up to these restrictions. Referring again to the binary choice example used earlier, the probability that an individual described by the attribute vector $\mathbf{Y}^j$ will prefer alternative $\mathbf{X}^1$ to $\mathbf{X}^2$ is

$$Pr(U^{*1j} \geq U^{*2j}) = Pr[\mathbf{Z}^{1j}(\boldsymbol{\beta}^* + \boldsymbol{\delta}^j) + \gamma^{1j} \geq \mathbf{Z}^{2j}(\boldsymbol{\beta}^* + \boldsymbol{\delta}^j) + \gamma^{2j}]$$
$$= Pr[(\mathbf{Z}^{1j} - \mathbf{Z}^{2j})(\boldsymbol{\beta}^* + \boldsymbol{\delta}^j) \geq \gamma^{2j} - \gamma^{1j}].$$

If the distribution assumptions are satisfied, then $\boldsymbol{\beta}^*$ and $\boldsymbol{\Sigma}$, the covariance matrix for the $\boldsymbol{\delta}^j$, can be estimated using maximum likelihood random coefficients probit estimation procedures (Hausman and Wise 1978).

With the RCCD formulation the total disturbance term for the ith alternative and jth individual is given by

$$\varepsilon^{*ij} = \mathbf{Z}^{ij}\boldsymbol{\delta}^j + \gamma^{ij}.$$

Note that, because $\boldsymbol{\delta}^j$ is constant for all alternatives evaluated by the jth individual, the total disturbances for all alternatives evaluated by the jth individual will covary.

This study uses the results of a choice experiment to address the following questions:

1. To what extent do taste parameters vary across individuals?
2. Is RCCD probit more immune to specification errors than LPIID probit?
3. Does the RCCD estimator provide a better fit to choice data?
4. Does the RCCD estimator lead to more accurate predictions concerning the impact of marginal changes in attributes on choice probabilities?
5. How precise are the RCCD estimates of the mean tastes coefficient vector? Of the variation in tastes?
6. Do the RCCD and LPIID estimators yield similar estimates of the elements of the mean tastes vector?

Question 1 addresses the fundamental premise motivating the development of the RCCD estimator. Taste variations undoubtedly exist, but to our knowledge no empirical studies have assessed the magnitude of such taste variations. Nor, to our knowledge, has any empirical study examined the shape of taste distributions (e.g., normal, exponential). It is apparent that the extent to which the RCCD estimator yields improved predictive accuracy will be directly related to the magnitude of the variation in taste parameters, and also to their distribution.

This speculation suggests a corollary speculation addressed in question 2. The degree to which RCCD estimation procedures are superior to LPIID procedures should be inversely related to the extent to which we are able to explicitly model the sources of variation in tastes. To illustrate this argument, suppose that attribute x_1 is a quality measure, attribute x_2 a price measure, and y is a measure of income.

Consider the following RCCD specifications:

$$U_1^{*ij} = (\beta_1^* x_1^i + \beta_2^* x_2^i) + (\delta_1^j x_1^i + \delta_2^j x_2^i) + \gamma_1^{ij},$$

$$U_2^{*ij} = \left[\alpha_1^* x_1^i + \alpha_2^* \left(\frac{x_2^i}{y^j}\right)\right] + \left[\theta_1^j x_1^i + \theta_2^j \left(\frac{x_2^i}{y^j}\right)\right] + \gamma_2^{ij}.$$

On a priori grounds we would expect much of the variation in the disutility of price to be explained by income. Consequently we expect Var $(\theta_2) <$ Var (δ_2). Also the covariance of the total disturbance term across alternatives should be smaller for the second specification, since one important source of this covariation (namely, income) is now explicitly incorporated in the utility model. The parallel LPIID models are

$$U_1^{ij} = \beta_1 x_1^i + \beta_2 x_2^i + \varepsilon_1^{ij},$$

$$U_2^{ij} = \alpha_1 x_1^i + \alpha_2 \left(\frac{x_2^i}{y^j}\right) + \varepsilon_2^{ij}.$$

Both are misspecified if these RCCD models hold. Nevertheless it would appear that the problems associated with incorrectly using an LPIID estimator should be greater in the first case. For here there is greater variation in tastes and greater covariation of the ε^{ij}. Thus the relative superiority of an RCCD estimator should be greater in the first case (where we do not model the source of taste variation) than in the second (where in part we do).

This observation has important practical implications for the potential usefulness of RCCD estimators. Whether through ignorance or lack of data, an analyst will frequently be unable to model explicitly the major sources of taste variation. Under such circumstances RCCD estimators may prove a useful tool for reducing the resulting losses in explanatory power and predictive accuracy. We hasten to add that we are not advocating the use of RCCD estimators as a remedy to excuse sloppy or ill-considered analyses. It must be recognized, however, that our understanding of the sources of taste variation is extremely limited. Consequently our ability to model explicitly the sources of taste variation is also extremely limited. Even in the most careful analyses numerous specification errors are almost certain to arise. Even though each of these errors may be small in its impact, the total effect may still be large. Because RCCD estimators may reduce the costs of such unavoidable specification errors, they are deserving of careful scrutiny.

Questions 3 and 4 address closely related, but not identical, issues. The LPIID model is a special case of the RCCD model that arises when $\Sigma_y = \sigma^2\mathbf{I}$ and Σ is a zero matrix. Thus it is apparent that the RCCD model will provide a better fit to a set of data than will the corresponding LPIID model. In principle one should not go wrong by obtaining maximum likelihood (ML) estimates of the RCCD model parameters. For when there is in fact random variation in tastes, ML estimators of the RCCD parameters are consistent.[2] If there is no variation in tastes, the ML estimators will still be consistent and will asymptotically reveal the true LPIID structure. By contrast ML estimators of the LPIID model parameters are consistent only in the absence of variation in tastes. The degree to which RCCD estimators are superior is clearly an empirical matter, depending both on the criterion one uses and on the amount of taste variation present in the empirical context studied. And if the speculation embodied in question 2 is correct, the superiority of the RCCD model should be inversely related to the extent to which we are able to explicitly model the source of variation in tastes.[3] With regard to predicting the effect of changes in choice relevant attributes (question 4), it is not obvious to what extent RCCD estimators will be superior, or even that they will be superior at all. Because RCCD models have more coefficients than the corresponding LPIID models, the standard errors of the RCCD model coefficients are likely to be larger. Predictions of the effects of changes in

2. Provided of course that variations in the taste parameters are multivariate normally distributed as assumed in the RCCD specification.

3. In general we explicitly model variation in tastes by including terms in the utility function constructed from explicit interactions of the attributes of alternatives and the attributes of individuals, for example, price/income.

choice relevant attributes that are based on imprecisely measured coefficients may be correspondingly imprecise. Also actual estimation of the parameters of either LPIID or RCCD models depends on ML search procedures that provide no assurance of attaining the global maximum. The fallibility of these search procedures, in conjunction with the greater complexity of RCCD models, may in practice more than offset the theoretical advantages of RCCD estimators.[4]

Questions 5 and 6 address related issues. Note that the first part of question 5 does not involve a comparison of the RCCD and LPIID estimators. We state it in this form because we are confident (even absent a formal proof) that the LPIID estimator of $\boldsymbol{\beta}$ is not a consistent estimator of $\boldsymbol{\beta}^*$, the population tastes vector. If the true total disturbance terms are of the form

$$\varepsilon^{*ij} = \mathbf{Z}^{ij}\boldsymbol{\delta}^{j} + \gamma^{ij},$$

then these ε^{*ij} are not independently distributed as assumed by the LPIID model. Under these conditions we conjecture that application of LPIID estimation procedures to choice data generated by an RCCD process will result in inconsistent estimates of $\boldsymbol{\beta}^*$, the true mean tastes coefficient vector. In particular we speculate that the probability limit of elements of the LPPID estimator $\boldsymbol{\beta}$ will be too small in absolute value terms (relative to the true mean tastes vector $\boldsymbol{\beta}^*$).[5] The reasoning behind this speculation is briefly outlined here and developed more fully in section 6.9. The unit of the utility measure (and consequently $\boldsymbol{\beta}$) in quantal choice models is arbitrary. In statistical applications the standard procedure has been to normalize coefficients with respect to the variance of the independent additive disturbance term. In the LPIID model this disturbance is ε^{ij}. In the RCCD model it is γ^{ij}. Note that, when tastes do in fact vary, Var (ε^{ij}) = Var (ε^{*ij}) > Var (γ^{ij}), for under these circumstances, Var (ε^{ij}) = Var $(\mathbf{Z}^{ij}\boldsymbol{\delta}^{j} + \gamma^{ij})$. Consequently the elements of the LPIID estimator $\boldsymbol{\beta}$ should be too small in absolute value (relative to $\boldsymbol{\beta}^*$). The reasoning behind this speculation leads to a related speculation addressed in question 6. The unit of measure cancels out when we look at the ratios of coefficients. Thus we

4. Also, RCCD estimation procedures are much more costly. These costs must also be weighed in any decision to use RCCD methods.

5. This argument, and the one which immediately follows, were suggested to us by Charles Manski.

suspect that on the average the ratios of the LPIID estimated coefficients should be close to the ratios of the RCCD coefficients, that is,

$$\frac{\hat{\beta}_s}{\hat{\beta}_t} \cong \frac{\hat{\beta}^*_s}{\hat{\beta}^*_t}, \quad \text{for all } s, t.$$

6.3 An Empirical Comparison of the LPIID and RCCD Models

The experiment described in this section was designed to provide a data base appropriate for addressing the six questions. The general design follows. To answer these questions we asked a group of respondents to choose between pairs of alternatives, with each alternative defined by two attributes. Each respondent made a relatively large number of hypothetical choices between pairs of alternatives, thus permitting us to estimate (for each respondent) the coefficients of the LPIID model:

$$U(x_1^i, x_2^i) = \beta_1 x_1^i + \beta_2 x_2^i + \varepsilon^i.$$

That is, we assumed that apart from an additive disturbance term each respondent's tastes were fixed (for the duration of the experiment) and linear in the attributes x_1 and x_2. The coefficients obtained from these individual choice models were used first to assess the degree of variation in tastes and second as a benchmark for evaluating the LPIID and RCCD models when they were applied to the group data.

We are of course aware that the data used in this study involve hypothetical choices. But given our objective of precisely modeling each individual's choice process, we need to observe many choices by each individual. It is extremely difficult to do this in a real choice setting. The hypothetical choices made by the individuals studied here are systematic and sensible. Thus we believe that they provide an appropriate basis for evaluating the statistical properties of the two estimation procedures.

6.4 Details of the Experiment

In implementing the experimental strategy outlined in section 6.3, we sought a choice task that was realistic, interesting, and yet involved alternatives described by only two major attributes. At the time we conducted the study, Duke University was embroiled in a debate over procedures for allocating parking permits. Some participants in that debate suggested that a pricing mechanism be used, with higher prices being levied for parking spots closer to the center of the campus. We exploited this situation by recruiting twenty respondents from the faculty, administration, and secretarial staff of Duke University's Institute of Policy Sciences and Public Affairs. Each respondent was asked to consider 60 pairs of parking lot alternatives, with each alternative being characterized by the attributes price per year and walking distance (in minutes) from the parking lot to the building where the respondent worked. The respondents were asked to choose one option from each pair. In constructing alternatives we used five price levels ($20, $40, $60, $90, and $120) and four walking distances (1.5, 2.5, 4 to 5, and 9 to 10 minutes).[6] From the 20 basic alternatives we created 60 pairs of alternatives in which, for every pair, the less expensive lot always involved the greater walking distance. The order of alternatives within pairs was randomly determined, with the low price-long distance option being presented first in half the pairs, and second in the other half. After the respondent had worked through all 60 pairs of alternatives, he or she was asked to complete a short questionnaire which elicited background information on age, position, income, and some details concerning where the respondent typically parked.

6.5 Results

The analysis proceeds in two stages. In the first stage a linear stochastic utility function is estimated for each respondent. In the second the observations for all respondents are pooled and stochastic utility functions for the entire sample are estimated. The alternative model specifications estimated in the second stage include both LPIID and RCCD models.[7]

The results of the first stage analyses will serve several functions. First, they will provide some empirical evidence on the magnitude of taste variations in the respondent population and on shape of the distributions

6. Since three of the four lots in question were real, we gave ranges for the more distant lots to avoid an aura of phony precision.

7. The first-stage models were estimated using a well-tested LPIID probit program developed by Richard McKelvey of Carnegie-Mellon University. The second-stage models (both LPIID and RCCD) were estimated using a program developed by Cambridge Systematics, Inc., Cambridge, Mass. In fact we could also have used the latter program to obtain the first-stage estimates. The decision to estimate the first-stage models with the McKelvey program was purely pragmatic. We had that program first and used it to complete the first stage analyses before receiving the Cambridge Systematics program.

of taste coefficients (question 1). Additionally, summary statistics characterizing the taste distribution estimated in the first stage analysis provide a useful basis for evaluating the parameter estimates from the second stage analysis. For example, are the second stage RCCD parameter estimates of β^* close to some measure of the central tendency of the estimated distribution of tastes (question 5)? How well does the estimation method allowing for taste variation capture the estimated distribution in tastes from the first stage analysis (question 5)? Finally, the models estimated in the first stage can be used as the basis for an analysis of the predictive accuracy of the models estimated in the second stage (question 4).

6.6 Analysis of Individual Respondents

A linear stochastic utility function (eq. 6.3) for each respondent was estimated under the assumption that the disturbance term in the utility function was normally distributed:[8]

$$U(D,F) = \beta_1 D + \beta_2 F + \varepsilon, \tag{6.3}$$

where

D = distance,
F = fee,
$\varepsilon \sim N(0, \sigma^2)$.

The results are summarized in table 6.1. No coefficient estimates are given for three respondents. The reasons for β_1 and β_2 being unestimable for these respondents are discussed in detail in section 6.9. In brief model (6.3) is unestimable for these respondents because their choice behavior is perfectly explained by a model of the form:

$$U(D,F) = \beta_1 D + \beta_2 F. \tag{6.4}$$

Model (6.4) is a special case of model (6.3) where the disturbance has zero variance (there is no stochastic component, ε, in the utility function). When

8. A log-linear utility model was also estimated for each respondent. The explanatory power (the log likelihood) of model (6.3) was almost always greater than the log-linear model.

Table 6.1
The model estimates for the individual respondents

Respondent	β_1	β_2	Percentage of choices correctly predicted
1	−4.57 (1.07)	−0.635 (1.05)	98
2	−0.623 (3.89)	−0.0509 (3.89)	85
3	−0.858 (3.66)	−0.0675 (3.66)	90
4	−0.230 (1.92)	−0.0751 (3.01)	92
5	−0.543 (4.37)	−0.0340 (3.85)	87
6	−0.266 (1.71)	−0.101 (2.59)	93
7	—	—	100
8	−0.0614 (0.42)	−0.0730 (2.22)	97
9	−1.75 (2.80)	−0.0924 (2.89)	95
10	−1.70 (2.83)	−0.0730 (2.74)	87
11	−0.569 (4.04)	−0.0366 (3.75)	83
12	−0.218 (1.55)	−0.0892 (2.65)	93
13	−0.628 (2.71)	−0.146 (2.92)	93
14	−0.476 (4.24)	−0.0381 (4.17)	83
15	−0.905 (3.61)	−0.0620 (3.45)	90
16	−0.382 (3.82)	−0.0133 (2.30)	82
17	−0.452 (4.24)	−0.0283 (3.75)	80
18	—	—	100
19	−1.73 (2.83)	−0.0887 (2.83)	93
20	—	—	100

Note: Ratio of parameter estimate to its standard error in parenthesis.

$\sigma^2 = 0$, only the ratio β_1/β_2 is potentially identifiable; β_1 and β_2 cannot be individually estimated.[9]

For the respondents with estimable models, the ratios of the parameter estimates to their standard errors are generally greater than two which suggests that the estimates are reasonably precise. The explanatory power of the models is also excellent. The percentage of choices predicted correctly ranges from 80 to 100 percent. For 70 percent of the individuals the percentage of correct predictions is 90 percent or better.

The results are examined from a different perspective in table 6.2. In this table the estimates of β_1, β_2, and β_1/β_2 (the value of time) are rank ordered from smallest to largest. (The basis for including the respondents with unestimable models in the ranking is discussed in section 6.9.)

These rankings can be interpreted as estimates of the distribution of tastes in the respondent population. The results suggest that the taste variations in the respondent population are large. The interquartile range (the difference between 25th and 75th percentile values) for each distribution is greater than the distribution's median value. (Because some respondent's models are not estimable, we cannot calculate sample means for the taste coefficients. But for reasons discussed in section 6.9 we are able to confidently estimate the sample medians.) While the results in table 6.2 may exaggerate the magnitude of the actual taste variations due to statistical variations in the parameter estimates for each respondent, the estimated distributions are sufficiently dispersed to leave little question that the actual taste variations are large. The results thus imply that in principle the idea of developing estimation methods that account for taste variations is well founded and that, if these methods can successfully capture the actual variation, they would be of considerable value.

Under an assumption of the number of trips made per year, the estimated values of β_1/β_2 can be transformed into estimates of the value of time (in a restricted sense) on a scale of dollars per hour. Assuming 500 trips to and from a lot per year, the median value of time is \$1.53; the 25th to 75th percentile range is \$.50 to \$2.27.

Information on the demographic and economic characteristics of each respondent was collected to assess which individual characteristics, if any, are systematically related to taste variation. The results of this analysis were used as a basis for properly specifying the population models

9. The parameters β_1, β_2, and σ^2 are not jointly identifiable. β_1 and β_2 are only estimable up to some assumed value of $1/\sigma$.

Table 6.2
β_1, β_2, and β_1/β_2 rank ordered from smallest to largest

	β_1		β_2		β_1/β_2	
1	—	(7, 18)	—	(8, 20)	—	(20)
2	—		—		0.842	(8)
3	−4.57	(1)	−0.635	(1)	2.44	(12)
4	−1.75	(9)	−0.146	(13)	2.64	(6)
5	−1.73*	(19)	−0.101*	(6)	3.06	(4)
					3.68*	
6	−1.70	(10)	−0.0924	(9)	4.29	(13)
7	−0.905	(15)	−0.0892	(12)	4.50	(18)
8	−0.858	(3)	−0.0887	(19)	7.20	(1)
9	−0.628	(13)	−0.0751	(4)	12.3	(2)
10	−0.623*	(2)	−0.0730*	(10)	12.5	(14)
					12.6*	
11	−0.569	(11)	−0.0730	(8)	12.7	(3)
12	−0.544	(5)	−0.0675	(3)	14.6	(15)
13	−0.476	(14)	−0.0620	(15)	15.6	(11)
14	−0.452	(17)	−0.0509	(2)	16.0	(17)
15	−0.382*	(16)	−0.0381*	(14)	16.0	(5)
					17.5*	
16	−0.266	(6)	−0.0366	(17)	18.9	(9)
17	−0.230	(4)	−0.0340	(5)	19.5	(19)
18	−0.218	(12)	−0.0283	(17)	23.2	(10)
19	−0.0614	(8)	−0.0133	(16)	28.6	(16)
20					95.0	(7)

Note: Respondent number in parentheses. Asterisks indicate 25th, 50th, and 75th percentile values.

estimated in the second stage analysis. Regressions of β_1/β_2 on respondent characteristics suggested that the only significant determinant of β_1/β_2 is income.[10] The income effect is positive as expected. The distributions of β_1 and β_2 (income adjusted) are shown in figures 6.1 and 6.2. (The method

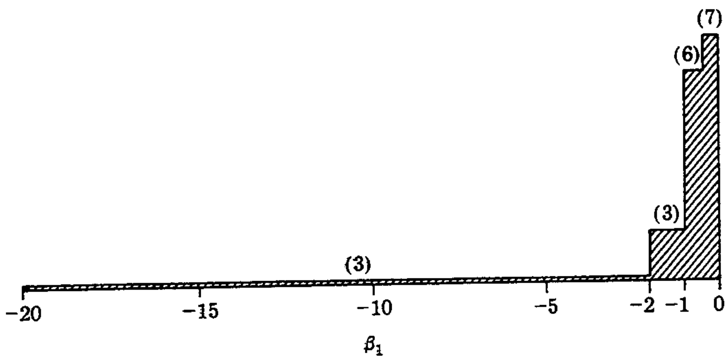

Figure 6.1
Relative frequency distribution for β_1 (the distance coefficient). Ths histogram is drawn so that areas are proportional to relative frequencies. At point (3) one respondent has an estimable coefficient value of 4.57. Two respondents have unestimable coefficients. Arbitrarily −20 is used as the lower bound for the interval.

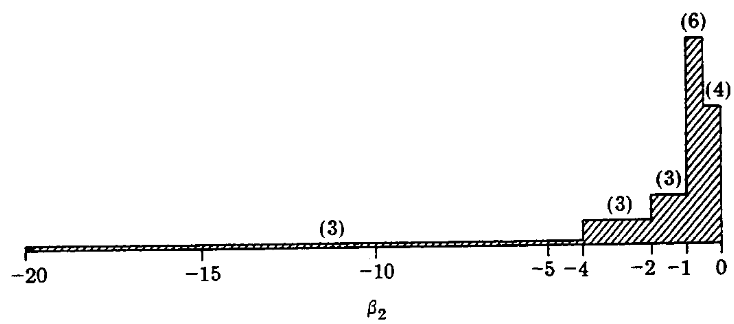

Figure 6.2
Relative frequency distribution for β_2 (the income adjusted fee coefficient). The histogram is drawn so that areas are proportional to relative frequency. At (3) one respondent has an estimable coefficient value of 15.88. Two others have unestimable coefficients. Arbitrarily −20 is used as the lower bound for the interval.

10. Regressions of β_1 and β_2 individually on respondent characteristics were not estimated because estimates of β_1 and β_2 are not available for the three respondents with unestimable models. Regressions involving only the β's of respondents with estimable models would yield inconsistent parameter estimates because, in effect, we would be sampling on the basis of the value of the dependent variable. Regressions involving the ratio of β_1/β_2 were feasible because an estimate was available for 19 of 20 respondents (see section 6.9). For the remaining respondent (number 20) β_1/β_2 could safely be approximated as zero.

used to make the income adjustment is described in section 6.7.) Note that even with the income adjustment there is still appreciable variation in both β_1 and β_2. This result suggests that attempts to model explicitly the sources of variation in tastes will meet with limited success at best. The potential usefulness of RCCD probit is enhanced accordingly.

6.7 A Comparison of LPIID Probit and RCCD Probit Estimation

To examine the merits of RCCD probit estimation relative to LPIID probit extimation, the responses for the entire respondent population were pooled, and several model specifications were estimated on the pooled data set using both estimation methods.[11]

Using the simulated cross section, we estimate model (6.3) using LPIID probit and a generalized version of model (6.3) that allows for taste variation using RCCD probit. The generalized model is of the form

$$U(D, F) = (\beta_1 + \delta_1)D + (\beta_2 + \delta_2)F + \gamma, \tag{6.5}$$

where

$$\delta_1, \delta_2 \sim N(\mathbf{0}, \mathbf{\Sigma}),$$
$$\gamma \sim N(0, \sigma^2).$$

In view of our finding that β_1/β_2 is significantly associated with the respondent's income, neither models (6.3) nor (6.5) are actually appropriate specifications for analyzing the pooled data set. Nevertheless we estimate the coefficients of models (6.3) and (6.5) in order to empirically evaluate the speculation that RCCD probit may offer valuable protection against the ill effects of specification error.

11. In estimating the various models with the pooled data, all responses were treated as independent. The responses are of course not independent since the 1,200 observations in the data set were not the responses of 1,200 different respondents but of 20 respondents making 60 choices. The LPIID and RCCD probit estimation algorithms are not designed to allow for the possibility that successive observations might be nonindependent, and we therefore made no attempt to account for the probable absence of independence among observations. While we have not analytically examined the impact of the nonindependence of the observations, standard results for least squares regression provide some insights into the probable impact. The estimated standard errors of the parameters are almost certainly incorrect because they are computed under the assumption that the observations are independent. It is unclear whether the parameter estimates themselves are inconsistent due to failure of the independence assumption. Standard regression results concerning the effect of correlation among the disturbances are probably not transferable because knowledge of the distribution of the disturbances is generally not necessary to make consistent estimates of the structural parameters.

To model the income effect we interact β_2 with income (I) as follows:

$$\text{Without taste variations: } U(D, F; I) = \beta_1 D + \frac{\beta_2}{I} F + \gamma \tag{6.6}$$

$$= \beta_1 D + \beta_2 \frac{F}{I} + \gamma,$$

$$\text{With taste variations: } U(D, F; I) = (\beta_1 + \delta_1) D + \left(\frac{\beta_2 + \delta_2}{I}\right) F + \gamma$$

$$= (\beta_1 + \delta_1) D + (\beta_2 + \delta_2)\frac{F}{I} + \gamma. \tag{6.7}$$

The parameter estimates for the four models are shown in table 6.3. Also included in the table are the median values of β_1, β_2, β_2/I, β_1/β_2, and $\beta_1/(\beta_2/I)$ from the first stage analysis of the individual respondents.

For the distance-fee specification the explanatory power of the RCCD probit model estimate is substantially greater than that for the LPIID probit model estimate. The log likelihood of the former is nearly twice as large as for the latter. While the RCCD probit model estimate has appreciably greater explanatory power than the LPIID probit model estimate, the estimated values of β_1 and β_2 from the LPIID probit model are much closer to the sample median estimates of β_1 and β_2. For the reasons discussed in the introduction, and elaborated upon in section 6.10, we believe the seemingly greater precision of the LPIID estimates is merely coincidental. In section 6.10 we argue that the LPIID estimates of β_1 and β_2 are probably not comparable to the estimated population medians because the LPIID estimates are made under an incorrect assumption about the variance of the disturbance.

While we suspect that the LPIID estimates of β_1 and β_2 are not comparable to the population medians, we do believe that their ratio, β_1/β_2, is comparable to the estimated population median ratio. The results in table 6.3 reveal that the LPIID ratio is moderately closer to the sample median ratio than the RCCD ratio. But the RCCD and LPIID estimates of β_1/β_2 are very similar, and both provide reasonably precise estimates of the median ratio in the sample.

Table 6.3
Parameter estimates for models estimated from the pooled data[a]

	Median value among respondents	Distance, fee model		Distance, fee/income model	
		LPIID probit	RCCD probit	LPIID probit	RCCD probit
β_1 (distance)	−0.623	−0.590 (−85.61)	−1.38 (−1.58)	−0.421 (−18.40)	−0.764 (−2.94)
β_2 (fee)	−0.07512	−0.0502 (−65.88)	−0.135 (−1.59)	—	—
β_1/β_2	12.71	11.77	10.21	—	—
β_2 (fee/income)	−1.20	—	—	−0.597 (−21.74)	−1.09 (−2.99)
β_1/β_2 (with income interaction)	0.639	—	—	0.705	0.702
Log likelihood	−224[b]	−1168	−641	−596	−559

[a] Ratio of parameter estimate with its standard error in parentheses.
[b] Sum of the log likelihoods for the models estimated for each respondent.

Turning to the more appropriate income interaction specification the difference in the explanatory power of the LPIID and RCCD probit estimation is greatly diminished; the log likelihood of the RCCD probit model estimate is only 10% greater than that for the LPIID probit model estimate. The improvement in explanatory power resulting from the introduction of the income interaction displays an interesting pattern. For the models estimated with LPIID probit the log likelihood increases by a factor of two with the introduction of the income interaction. In contrast for the models estimated with RCCD probit the introduction of the income interaction improves the log likelihood by less than 20%. This result is consistent with our speculation that RCCD probit estimation may be more immune to specification error than estimation methods that do not allow for taste variations.

While there is only a moderate difference in the explanatory power of LPIID and RCCD probit estimation for the income interaction model, the resulting parameter estimates differ appreciably. The RCCD probit estimates of β_1 and β_2 (income adjusted) are quite close to the corresponding median values of the individual taste coefficients from the first stage analysis. By contrast, the LPIID estimates of β_1 and β_2 are only slightly more than half as large (in absolute magnitude) as the corresponding RCCD estimates. This latter finding may be due to the fact that even when the income effect is explicitly modeled, substantial taste variation remains to be accounted for (see figure 6.1).

We next examine the ability of RCCD probit estimation to capture the taste variation in the population. Table 6.4 gives the RCCD probit estimates of Σ, the variance-covariance matrix of tastes, for both RCCD model specifications. As can be seen in table 6.4, the estimate of Σ for the fee-distance model is very imprecise; the ratios of the estimated parameters to their standard errors are never greater than one. The estimate of the population variance of β_2 is particularly poor; the ratio of the estimate to its standard error is nearly zero. Albright, Lerman, and Manski (1977) and Hausman and Wise (1978) experienced similar difficulties in estimating the variances of taste parameters.

The estimates of Σ for the income-adjusted model appear to be moderately precise. Nevertheless the more important question is how well do the point estimates of the population variance of β_1 and β_2 for the income-adjusted model capture the variation measured in the first-stage analysis. Since RCCD probit estimation assumes that tastes are normally distributed, the 25th to 75th percentile range, the interquartile range, of the estimated taste distributions from the first-stage analyses should be approximately equal to

$$\beta_i^{25\text{th}} = \beta_1 - 0.667\sigma_i,$$
$$\beta_i^{75\text{th}} = \beta_i + 0.667\sigma_i,$$

where

$\beta_i^{25\text{th}}, \beta_i^{75\text{th}}$ = respectively, the 25th and 75th percentile estimates of β_i,
β_i = RCCD probit estimate of the mean (or median) β_i,
σ_i = RCCD probit estimate of the standard deviation of the taste distribution.

In table 6.5 the interquartile range of β_1 and β_2 (income adjusted) estimated from the first-stage analysis is compared with the estimate of that

Table 6.4
Estimates of Σ for RCCD models

	Fee distance	Income adjusted
Population variance β_1	0.518 (0.73)	0.277 (1.20)
Population variance β_2	0.00358 (0.01)	0.247 (1.18)
Population covariance β_1 and β_2	−0.0574 (−0.75)	0.0689 (0.47)

Note: Ratio of estimate to its standard error in parentheses.

Table 6.5
The 25th to 75th percentile range of taste parameters

	β_1		β_2 (income adjusted)	
	25th	75th	25th	75th
From first-stage analysis	−1.72	−0.38	−3.11	−0.58
From RCCD probit parameter estimates	−1.11	−0.41	−1.42	−0.76

range generated from the RCCD probit parameter estimates of the income-adjusted model.

As can be seen in table 6.5, the RCCD probit estimate of the 75th percentile value of β_1 corresponds quite closely with the estimate from the first stage analysis. There is also a reasonably close correspondence between the two estimates of the 75th percentile value of β_2 (income adjusted). In contrast the RCCD probit estimates of the 25th percentile values of β_1 and β_2 differ markedly from the first-stage estimates. The reason for the lack of correspondence between the RCCD probit and first-stage estimates of the 25th percentile values of β_1 and β_2 is due to the marked leftward skew in the distribution of each in the respondent population. Histograms of the distribution of β_1 and β_2 (income adjusted) from the first-stage analysis are shown in figures 6.1 and 6.2. Since RCCD probit estimation assumes tastes to be symmetrically distributed, it is not surprising that the RCCD probit and first-stage estimates of the 25th percentile values of β_1 and β_2 (income adjusted) differ markedly.

The leftward skew of the distributions in figure 6.1 suggests that the assumption that tastes are normally distributed may not be a good approximation for taste parameters which for theoretical reasons are thought to be bounded. A more appropriate distributional approximation for β_1 and β_2 might be the negative of either a log-normal or exponential distribution. Both of these distributions are bounded from above by zero. Theoretically these two distributions are also more appealing. While it is conceivable that a small proportion of the population might value walking time positively ($\beta_1 > 0$), it is not plausible that even a minority of the population would positively value paying more for parking privileges ($\beta_2 > 0$).

We close this section with the results of an experiment comparing the predictive accuracy of the two estimation methods. Using log likelihood as a measure of explanatory power, the results suggest that, if the specification of explanatory variables is approximately correct, then LPIID and RCCD probit estimation are about equivalently powerful. However, if the explanatory variables are misspecified, then RCCD probit estimation is appreciably more powerful than LPIID probit estimation. Nevertheless, even if the explanatory variables are correctly specified, it is not clear that the predictive accuracy of the RCCD and LPIID models will be equal. The predictions of the two models may potentially be quite different. In view of the current and potential applications of RCCD and LPIID estimation, their relative predictive accuracy is a crucial test of their relative overall merits. The following experiment is intended to provide such a test.

For any given value of the fee and distance differential between two lots, A and B, the expected change in the proportion of the population choosing A because of a one unit increase in ΔF can be approximated by

$$E_{\Delta F}(\Delta P_A; \Delta F, \Delta D) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial P_A^i(\Delta F, \Delta D)}{\partial(\Delta F)}, \tag{6.8}$$

where

$E_{\Delta F}(\Delta P_A;\ \Delta F,\ \Delta D)$ = expected change in the proportion of the population choosing lot A, resulting from a one unit increase in ΔF for given values of ΔF and ΔD,

$P_A^i(\Delta F,\ \Delta D)$ = probability of individual i choosing lot A for given values of ΔF and ΔD,

n = number of individuals in the population.

Similarly the expected change in the proportion choosing A resulting from a one unit increase in ΔD can be approximated by

$$E_{\Delta D}(\Delta P_A; \Delta F, \Delta D) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial P_A^i(\Delta F, \Delta D)}{\partial(\Delta D)}. \tag{6.9}$$

For selected values of ΔF and ΔD we computed equations (6.8) and (6.9), using as our population the 20 respondents in the sample. The partial derivatives of P_A^i w.r.t. ΔD and ΔF for each respondent with an estimable model were computed on the basis of the estimated values of β_1 and β_2 for that respondent. For the respondents with unestimable models the partial derivatives of P_A^i could be assumed to be zero for the values of ΔF and ΔD used in the experiment.

Estimation of $E_{\Delta D}$ for the models estimated in the second-stage analysis can be computed as follows (we show only equations for $E_{\Delta D}$; the corresponding equations for $E_{\Delta F}$ are the same except that the partial derivatives are computed with respect to ΔF):

Fee-distance model: LPIID probit

$$E_{\Delta D} = \frac{\partial P_A^I(\Delta F, \Delta D)}{\partial(\Delta D)},$$

RCCD probit

$$E_{\Delta D} = \frac{\partial P_A^R(\Delta F, \Delta D)}{\partial(\Delta D)};$$

Income-adjusted model: LPIID probit

$$E_{\Delta D} = \sum_{i=1}^{k} \pi_k \cdot \frac{\partial P_A^I(\Delta F, \Delta D; Y_k)}{\partial(\Delta D)},$$

RCCD probit

$$E_{\Delta D} = \sum_{i=1}^{k} \pi_k \cdot \frac{\partial P_A^R(\Delta F, \Delta D; Y_k)}{\partial(\Delta D)};$$

where $P_A^I(\Delta F, \Delta D), P_A^R(\Delta F, \Delta D)$ = predicted proportion of the population choosing lot A for given ΔF and ΔD for, respectively, the LPIID and RCCD probit estimates of the fee-distance model,

$P_A^I(\Delta F, \Delta D; Y_k), P_A^R(\Delta F, \Delta D; Y_k)$ = predicted proportion of the population choosing lot A for given ΔF, ΔD and Y_k for, respectively, the LPIID and RCCD probit estimates of the income-adjusted model,

Y_k = income of respondents in income class k,

π_k = proportion of respondents in income class k.

In the experiment we treat the estimates of $E_{\Delta D}$ and $E_{\Delta F}$ from the first-stage model estimates (eqs. (6.8) and (6.9)) as the actual changes in P_A that would occur from one unit increases in ΔD and ΔF, respectively. We then compare these estimates of $E_{\Delta D}$ and $E_{\Delta F}$ with those generated from the second-stage model estimates. Estimates of $E_{\Delta D}$ and $E_{\Delta F}$ are computed for nine different values of $(\Delta F, \Delta D)$ and three different values of $(\Delta F/\Delta D)$.[12] The results are shown in tables 6.6 and 6.7.

12. One value of $-(\Delta F/\Delta D)$ used in the experiment is 12 which is about equal to the median estimate of β_1/β_2 in the population. The remaining two values of $-(\Delta F/\Delta D)$ are 8 and 16. The population estimate of the 75th percentile value of β_1/β_2 is 16, but due to the leftward skew of the distributions of β_1 and β_2 the 25th percentile estimate of β_1/β_2 is nearly 3.5. For the purpose of the experiment we chose not to use a value of $-(\Delta F/\Delta D)$ that was so far from the population median.

Table 6.6
The results of the fee-distance model prediction experiment

			$E_{\Delta F}$			$E_{\Delta D}$		
$-\frac{\Delta F}{\Delta D}$	ΔF	ΔD	"Actual" $E_{\Delta F}$	Deviation of LPIID probit estimates from "actual"	Deviation of RCCD probit estimates from "actual"	"Actual" $E_{\Delta D}$	Deviation of LPIID probit estimates from "actual"	Deviation of RCCD probit estimates from "actual"
8	−48	6.0	0.00212	0.00842	0.00358	0.02731	0.09662	0.01835
12	−72	6.0	0.00542	0.01456	−0.00024	0.07610	0.15888	0.03988
16	−96	6.0	0.00563	0.00324	−0.00385	0.08707	0.01727	−0.05864
8	−24	3.0	0.01168	0.00538	−0.00046	0.10432	0.09631	−0.01393
12	−36	3.0	0.00723	0.01278	−0.00123	0.10274	0.13269	−0.03097
16	−48	3.0	0.00892	0.00742	−0.00536	0.14496	0.04722	−0.08833
8	−12	1.5	0.02419	0.00495	−0.00300	0.20626	0.02005	−0.03221
12	−18	1.5	0.01323	0.00679	−0.00144	0.16321	0.07233	−0.02320
16	−24	1.5	0.01237	0.00666	−0.00526	0.18528	0.03861	−0.07376
Sum of absolute deviations				0.07020	0.02655		0.6710	0.3794

Note: Positive deviations are overpredictions, and negative deviations are underpredictions.

Table 6.7
The results of the income-adjusted model prediction experiment

			$E_{\Delta F}$			$E_{\Delta D}$		
$-\frac{\Delta F}{\Delta D}$	ΔF	ΔD	"Actual" $E_{\Delta F}$	Deviation of LPIID probit estimates from "actual"	Deviation of RCCD probit estimates from "actual"	"Actual" $E_{\Delta D}$	Deviation of LPIID probit estimates from "actual"	Deviation of RCCD probit estimates from "actual"
8	−48	6.0	0.00212	0.00459	−0.00306	0.02731	0.04105	−0.03749
12	−72	6.0	0.00542	−0.00045	−0.00169	0.07610	−0.01169	0.00105
16	−96	6.0	0.00563	−0.00172	−0.00289	0.08707	−0.02822	0.00769
8	−24	3.0	0.01168	−0.00281	0.00217	0.10432	−0.02007	−0.02237
12	−36	3.0	0.00723	−0.00011	0.00026	0.10274	−0.02422	−0.03763
16	−48	3.0	0.00892	−0.00324	0.00373	0.14496	−0.07341	−0.00049
8	−12	1.5	0.02419	−0.01422	0.00894	0.20626	−0.11602	−0.00694
12	−18	1.5	0.01323	−0.00413	0.00144	0.16321	−0.07576	−0.06039
16	−24	1.5	0.01237	−0.00435	0.00322	0.18528	−0.10217	−0.04057
Sum of absolute deviations				0.03564	0.02736		0.4926	0.2147

Note: Positive deviations are overpredictions, and negative deviations are underpredictions.

For the fee-distance model the LPIID model consistently overpredicts the magnitude of the change, whereas the RCCD model underpredicts the change in all but one trial. In terms of the absolute magnitude of the prediction errors, the predictions from the RCCD model are appreciably more precise than those of the LPIID model. The RCCD estimate of $E_{\Delta F}$ is closer to the actual value of $E_{\Delta F}$ in 8 of 9 trials. Moreover the sum of the absolute deviations of the RCCD estimate of $E_{\Delta F}$ from the actual value of $E_{\Delta F}$ across the nine trials is about 40 percent of that sum for the LPIID estimates of $E_{\Delta F}$. For $E_{\Delta D}$ the RCCD estimates are more accurate in only 5 of 9 trials. But using the sum of absolute deviations as a measure of predictive accuracy, the RCCD sum is roughly half that of the LPIID sum.

For the income-adjusted model the patterns in the direction of prediction errors is markedly different from that pattern for the fee-distance model. The LPIID estimates of $E_{\Delta F}$ and $E_{\Delta D}$ are smaller than the actual in 8 of 9 trials. In contrast the patterns in the sign of the prediction errors for the RCCD model are less distinct. A pattern of overprediction of $E_{\Delta F}$ and underprediction of $E_{\Delta D}$ does appear to be present, however. In terms of the absolute magnitudes of the prediction errors, RCCD probit again appears to be distinctly more accurate than LPIID probit estimation. The RCCD probit estimates of $E_{\Delta D}$ are more accurate than the LPIID probit estimates in 7 of 9 trials, and the sum of absolute deviations for the RCCD probit estimates is less than half that sum for the LPIID probit estimates. For $E_{\Delta F}$ the improvement in predictive accuracy offered by RCCD probit estimation is more moderate. The RCCD estimates are more accurate than the LPIID estimates in only 5 of 9 trials, but the sum of absolute deviations for the RCCD estimates is moderately smaller than the sum for the LPIID estimates.

Overall, the results of this experiment suggest that models allowing for taste variation have appreciably better predictive power than models that do not account for such variations.

6.8 Conclusions

The results discussed in section 6.7 prompt several observations. The first-stage analysis clearly indicates that the magnitude of taste variation is appreciable. Regressions of β_1/β_2 on various respondent characteristics revealed only one significant correlate—income. Yet figures 6.1 and 6.2 reveal that, even when the effects of income are explicitly modeled, substantial taste variation remains unexplained. This observation suggests that attempts to model explicitly the sources of taste variation using an LPIID formulation are unlikely to succeed. Thus the motivation underlying the development of RCCD estimation methods appears well founded.

The results of the second-stage analysis suggest that RCCD estimators are more robust than LPIID estimators against errors involving inappropriate specification of explanatory variables. Here our results are at odds with those of Albright, Lerman, and Manski (1977) and Hausman and Wise (1978). They found that the explanatory power of RCCD probit was not appreciably greater than that of LPIID logit. Thus it would be premature to conclude that RCCD probit is more robust to specification errors than the LPIID formulation (either probit or logit).

To account for the conflict between our results and those obtained in the two studies cited, we speculate that the difference may be attributable to the relative contributions of $\mathbf{Z}\boldsymbol{\delta}$ (error due to taste variation) and ε (additive error) to the total disturbance term. On the basis of our first-stage analyses we are confident that the bulk of the disturbance in our second-stage model specifications is attributable to taste variation.[13] This may well be due to the fact that in the problem studied we knew what the relevant attributes of the alternatives were. We knew because we specified them as part of our experimental design. By contrast the Albright, Lerman, and Manski (1977) and Hausman and Wise (1978) analyses used nonexperimental data concerning transportation mode selection for work commuting trips. In such nonexperimental contexts there may well be many excluded but choice relevant attributes of alternatives. These omitted variables are reflected in ε, the additive disturbance term. When important attributes of alternatives are omitted (as they are likely to be), ε will make a large contribution to the total disturbance term; and $\mathbf{Z}\boldsymbol{\delta}$ (error due to variation in tastes for observed attributes) may be a relatively minor contributor to the total disturbance term. This speculation suggests that RCCD probit will become an increasingly valuable tool as our ability to recognize and measure choice relevant attributes increases.

13. Recall the excellent explanatory power of the first-stage models of individual respondents.

6.9 Appendix: The Unestimable Models

For three respondents a model of the form

$$U(D,F) = \beta_1 D + \beta_2 F + \varepsilon \tag{6.10}$$

was not estimable. To motivate a discussion of the reasons model (6.10) is not estimable for these respondents, it is useful to consider the implications for estimating taste parameters when choice behavior is perfectly explained by a model of the form

$$U(D,F) = \beta_1 D + \beta_2 F. \tag{6.11}$$

Model (6.11) is a special case of model (6.10) where the disturbance has zero variance (there is no stochastic component, ε, in the utility function). When $\sigma_\varepsilon^2 = 0$, only the ratio β_1/β_2 is identifiable. It can be shown that, when $\sigma_\varepsilon^2 = 0$, the prefered alternative in a binary choice set (A, B) can always be predicted by the following rule: A will be chosen if

$$\begin{aligned} \frac{\beta_1}{\beta_2} &> -\frac{F_A - F_B}{D_A - D_B} \quad \text{if} \quad D_A - D_B < 0, \\ \frac{\beta_1}{\beta_2} &< -\frac{F_A - F_B}{D_A - D_B} \quad \text{if} \quad D_A - D_B > 0. \end{aligned} \tag{6.12}$$

Otherwise B will be chosen. Thus in instances where $\sigma_\varepsilon^2 = 0$, only the ratio β_1/β_2 is identifiable from actual choice behavior.

Table 6.8 shows how choice behavior predicted perfectly by (6.11) would closely identify β_1/β_2.

The choice behavior of the three respondents with unestimable models is consistent with the predictions of model (6.10). The choice behavior of respondent 18 is fully consistent with that predicted by model (6.11), and accordingly β_1/β_2 could confidently be estimated at 4.15.

Respondent 7 picked the most convenient lot in all except one trial. In that trial the choice of the most convenient lot required the individual to pay an additional \$105 annually to reduce walking time by about 1 minute per trip; this trial involved the largest ($\Delta F/\Delta D$) in the experiment. Although respondent 7 demonstrated a willingness to sacrifice convenience for a savings in the parking fee in only one instance, his behavior is fully consistent with model (6.10), and we estimate β_1/β_2 for this respondent to be about 95.

Table 6.8
Estimating β_1/β_2 when model (6.11) is the applicable utility function

(ranked from largest to smallest) $-\frac{F_A - F_B}{D_A - D_B}$	$D_A - D_B < 0$ (chosen alternative)	$D_A - D_B > 0$ (chosen alternative)
K_1	B	A
K_2	B	A
K_3	B	A
K_4	B	A
$\frac{\beta_1}{\beta_2} = \frac{(K_4 + K_5)}{2}$		
K_5	A	B
K_6	A	B
K_7	A	B

Respondent 20 displayed lexicographic type behavior; the individual always picked the most inexpensive lot. If this respondent actually used a lexicographic choice rule, then neither models (6.10) nor (6.11) would be an appropriate specification. However, his behavior is not inconsistent with model (6.11). It is possible that the reason we observed no trade-offs between D and F is that the individual's value of time was sufficiently small that it was not detected by the experiment. Indeed this interpretation seems more plausible than the lexicographic interpretation, since the lexicographic interpretation implies that the individual would be unwilling to pay a trivial sum (e.g., 1 cent/annum) in return for any savings in walking time (e.g., 30 minutes/trip). We thus assume that a value of β_1/β_2 exists for this individual and ranked this respondent's β_1/β_2 as the smallest in the sample without specifying a value (see table 6.2).

While estimates of β_1 and β_2 for the three respondents could not be identified with the data at hand, this does not necessarily imply that a unique value for each does not exist. Estimates do not exist only if $\sigma_\varepsilon^2 = 0$. The fact that model (6.11) is sufficient for explaining the behavior of these respondents for this data set does not imply that it would be sufficient for explaining their behavior in all possible data sets. In fact it seems safe to assume that $\sigma_\varepsilon^2 \neq 0$ for these individuals. Suppose these individuals were confronted with successive choice sets where $-(F_A - F_B)/(D_A - D_B) \cong \beta_1/\beta_2$. It is highly unlikely that their choices would be predicted exactly by (6.12); human cognitive capabilities are simply not sufficiently acute to distinguish very small differences. Additionally any excluded determinants of choice that only marginally effect the utility of each alternative could become decisive for individuals close to the margin.

The observation that σ_ε^2 probably does not equal zero for these three respondents has implications for the rankings of the β_1's and β_2's for these respondents. In model (6.10) β_1 and β_2 are estimable only up to some assumed value of $1/\sigma_\varepsilon$; β_1, β_2, and σ_ε are not jointly identified. The estimated values of β_1 and β_2 should actually be interpreted as $\beta_1/\sigma_\varepsilon$ and $\beta_2/\sigma_\varepsilon$, respectively, where the value of σ_ε is arbitrarily assigned.

Suppose two respondents, X and Y, both have identical β_1 and β_2 but $\sigma_\varepsilon^X < \sigma_\varepsilon^Y$. We would thus predict that the estimates of β_1 and β_2 for X would be larger in absolute magnitude than those for Y, since

$$|E(\hat{\beta}_i^X)| = |[\beta_i^X/\sigma_\varepsilon][\sigma_\varepsilon/\sigma_\varepsilon^X]| > |E(\hat{\beta}_i^Y)| = |[\beta_i^Y/\sigma_\varepsilon][\sigma_\varepsilon/\sigma_\varepsilon^Y]|, \qquad (6.13)$$

$i = 1, 2$, where

$\hat{\beta}_i^X$, $\hat{\beta}_i^Y$ = estimated value of β_i for X and Y, respectively,

σ_ε = assumed value of σ_ε^X and σ_ε^Y.

This result provides a basis for making some reasonable assumptions about the appropriate rankings of β_1 and β_2 for respondents with unestimable models. The results suggest that the value of σ_ε^2 for these respondents is probably small relative to its value for the remainder of the respondents.

In the case of respondent 18, whose β_1/β_2 is very precisely bracketed by the experiment, we are quite confident that this individual has a relatively small σ_ε^2. The result in equation (6.13) thus suggests that this individual's β_1 and β_2 are best assumed to be among the largest in the sample (see table 6.2). Also included among the largest values of β_1 and β_2 in the population are respondent 7's β_1 and 20's β_2, respectively.

We have made no assumptions about the ranking of respondent 7's β_2 and 20's β_1. Recall that respondent 7 had a very high value of time and respondent 20 a very low value of time. Loosely speaking, this implies 7's β_2 and 20's β_1 are small, since the value of time is estimated by β_1/β_2. Thus for respondent 7's β_2 and 20's β_1 to be ranked among the largest in their respective distributions, σ_ε^2 for 7 and 20 would have to be extremely small relative to the remainder of the respondents. This seemed to us to be an overly strong assumption, and we thus have chosen to exclude them from the rankings. This is an admittedly tenuous assumption but in any event the alternative of ranking each among the largest in their respective distributions does not alter any of our conclusions.

6.10 Appendix: Mean Taste Estimates in the LPIID and RCCD Models

Among the issues we had initially intended to explore in this analysis was the correspondence between the LPIID probit estimates of β and some measure of the central tendency of the taste distribution estimated in the first-stage analysis. In the course of the analysis the question arose of whether the LPIID estimates of β are theoretically comparable to the first-stage estimates of the central tendency of the taste distribution. Stated differently, is there any theoretical reason for suspecting that the LPIID estimates of β will be biased estimates of the mean (or median) of the population's taste distribution? Although we have not addressed this question in a fully rigorous fashion, we suspect that LPIID estimate of β is not comparable to the mean (or median) of the population taste distribution.

The generalized forms of the stochastic utility functions considered in this chapter are

$$\text{Without taste variations (LPIID): } U(Z) = \mathbf{Z}\boldsymbol{\beta} + \varepsilon, \tag{6.14}$$

$$\text{With taste variations (RCCD): } U(Z) = \mathbf{Z}(\boldsymbol{\beta} + \boldsymbol{\delta}) + \gamma = \mathbf{Z}\boldsymbol{\beta} + (\mathbf{Z}\boldsymbol{\delta} + \gamma), \tag{6.15}$$

where

$\mathbf{Z} = (1 \times K)$ vector of individual and alternative characteristics,
$\boldsymbol{\beta} = (K \times 1)$ vector of parameters,
$\boldsymbol{\delta} = (K \times 1)$ disturbance vector assumed to be distributed $N(0, \boldsymbol{\Sigma})$,
γ, ε = random variables assumed to be distributed $N(0, \sigma^2)$.

In both models (6.14) and (6.15) all the parameters are not jointly identified: $\boldsymbol{\beta}$ and σ^2 in (6.14) and $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, and σ^2 in (6.15). For both models a necessary condition for identification is that the value of one parameter be assumed; the remaining parameters are identified up to that assumed value. For both models identification is typically accomplished by assuming the value of σ^2. It can be shown that for model (6.14) the expected value of $\hat{\boldsymbol{\beta}}$ equals $\boldsymbol{\beta}/\sigma$ and that for model (6.15) the expected values of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$ are $\boldsymbol{\beta}/\sigma$ and $\boldsymbol{\Sigma}/\sigma^2$, respectively.

Now suppose model (6.14) is estimated when the true model is (6.15). The actual disturbance is thus not ε but $\alpha = \varepsilon + \mathbf{Z}\boldsymbol{\delta}$, and therefore the actual variance of the disturbance is not σ^2 but $\sigma^2 + \mathbf{Z\Sigma Z}'$, where $\sigma^2 + \mathbf{Z\Sigma Z}' > \sigma^2$. We thus speculate that, when model (6.14) is estimated where (6.15) is the true model, the following result will hold:

$$|E(\hat{\boldsymbol{\beta}})| = \frac{|\boldsymbol{\beta}|}{(\sigma^2 + \mathbf{Z\Sigma Z}')^{\frac{1}{2}}} < \frac{|\boldsymbol{\beta}|}{\sigma}. \tag{6.16}$$

Furthermore we would predict that the absolute magnitude of each element of the LPIID estimate of $\boldsymbol{\beta}$ would be less than the corresponding element of the RCCD estimate of $\boldsymbol{\beta}$. The basis for this prediction is that in RCCD estimation, the presence of $\mathbf{Z}\boldsymbol{\delta}$ in the disturbance is explicitly taken into account in estimating $\boldsymbol{\beta}$.

While we have not formally proven these results, the empirical results in table 6.3 are consistent with our speculation. For both the fee-distance and income-adjusted models the absolute magnitude of RCCD estimates of β_1 and β_2 are greater than the corresponding LPIID estimates. An inspection of the alternative estimates of the ratio β_1/β_2 also supports our speculation. If we are correct in supposing the source of the bias in the LPIID estimate of $\boldsymbol{\beta}$ is the presence of the term $\mathbf{Z\Sigma Z}'$ in the denominator of eq. (6.16), then the ratio of the LPIID estimates of β_1 and β_2 should be comparable to both the ratio of the RCCD estimates of β_1 and β_2 and to the median estimate of β_1/β_2 from the first-stage analysis. An examination of table 6.3 reveals all three estimates of the true mean value of β_1/β_2 in the population are about equal.

For these reasons we do not believe the LPIID estimates of β_1 and β_2 can be meaningfully compared with the estimates of the median values of β_1 and β_2 in the respondent population. Thus we suspect that it is only coincidental that the LPIID estimates of β_1 and β_2 in the fee-distance model are closer to the first-stage medians than the RCCD estimates. We attribute the large divergence of the RCCD estimates of β_1 and β_2 in the fee-distance model from first-stage estimates of the population medians to the absence of an income adjustment in the fee-distance model.

References

Albright, R. L., S. Lerman, and C. F. Manski. 1977. Report on the Development of an Estimation Program for Multinomial Probit Model. Prepared for the Federal Highway Administration.

Ben-Akiva, M., and P. Haus. 1973. Estimation of a Work Mode Split Model Which Includes the Carpool Mode. Working paper WP-1. U.S. Department of Transportation, Washington, D. C.

Bock, R. D., and L. V. Jones. *The Measurement and Prediction of Judgment and Choice.* San Francisco: Holden-Day.

Charles Rivers Associates. 1972. *A Disaggregate Behavioral Model of Urban Travel Demand*. Cambridge, Mass.

Coombs, C., R. Dawes, and A. Tversky. 1970. *Mathematical Psychology*. Englewood Cliffs, N. J.: Prentice-Hall.

Hausman, J. A., and D. A. Wise. 1978. A Conditional Probit Model for Qualitative Choice: Descrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica*. 46: 403–426.

Kohn, M., C. Manski, and D. Mundel. 1971. A Study of College Choice: Report on Research in Progress. Unpublished.

Lerman, S., and M. Ben-Akiva. 1975. A Disaggregate Model of Auto Ownership. Presented at the Annual meetings of the Transportation Research Board, Washington, D. C.

Luce, R. D., and P. Suppes. 1965. Preference, Utility, and Subjective Probability. In *Handbook of Mathematical Psychology*, vol. 3, ed. R. D. Luce. New York: Wiley, pp. 249–410.

Manski, C. F. 1973. The Stochastic Utility Model of Choice. School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh.

McFadden, D. 1973. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka. New York: Academic Press.

Tversky, A. 1972. Choice by Elimination. *Journal of Mathematical Psychology*. 79: 281–299.