Mode Choice — Problems and Extensions

Philip A. Viton

November 21, 2011

Introduction

We now consider some problems with the models we have developed (principally logit) and what can be done to avoid them.

Note: the more technical material on mixed logit will not be part of the final examination.

IIA

- Suppose that the logit model holds; and consider the odds that individual i will selected mode j over mode h.
- With logit choice probabilities, this is a routine calculation: the denominators cancel and we have

$$\frac{P_{ij}}{P_{ih}} = \frac{e^{v_{ij}}}{e^{v_{ih}}}$$

- We see that the odds depends only on the systematic (observable) utility of the two modes in question.
- Put another way, the odds do not depend on (are independent of) the characteristics of any other (irrelevant) alternatives — only the two alternatives (j and h) in question.
- This is the *Independence of Irrelevant Alternatives* (IIA) property of the logit model.

So What?

- Although the IIA property is easiest to see for the logit model, it will in fact hold for any model in which the idiosyncratic terms (the η's) are independent and identically distributed random variables.
- But again, so what?
- It turns out that this property can lead to *demonstrably wrong* predictions about mode choices.

Red Bus / Blue Bus

- To see this, consider a situation in which there are two available modes, auto and a bus transit mode. And suppose that the transit system paints all its buses red.
- Suppose that we are looking at the aggregate choice probabilities, rather than those of a single individual. In this case, the choice probabilities will be the modal shares.
- Suppose, finally, that the observed shares are

$$\begin{aligned} P_A &= 0.7 \\ P_{RB} &= 0.3 \end{aligned}$$

where the subscripts A and RB refer to auto and the (red) bus modes, respectively.

Red Bus / Blue Bus

- Now suppose that a new transit mode becomes available.
- It is *exactly* like the existing mode in all respects, except that its buses are painted blue rather than red.
- We are now asked to predict the mode shares in the 3-mode setting.
- Surely this is completely clear: the two transit modes will split the previous transit-using population, and we will have

$$\begin{aligned} P'_A &= 0.7 \\ P'_{RB} &= 0.15 \\ P'_{BB} &= 0.15 \end{aligned}$$

where the primes refer to the 3-mode setting.

Red Bus / Blue Bus

- However *the logit model does not agree*. For the 3-mode setting, it reasons as follows:
- Because of IIA, the odds of selecting auto over red bus do not depend on whether there is a blue bus in the picture or not (the blue bus is an irrelevant alternative here) so:

$$\frac{P'_A}{P'_{RB}} = \frac{P_A}{P_{RB}} = \frac{0.7}{0.3} = 2.3333$$

- And because the two transit modes are identical, their choice probabilities must be the same:

$$P'_{RB} = P'_{BB}$$

- And finally the choice probabilities must add up to one:

$$P'_A + P'_{RB} + P'_{BB} = 1$$

Red Bus / Blue Bus

- Therefore we have:

$$\begin{aligned} P'_A &= 2.3333\ P'_{RB} \\ 1 &= 2.3333\ P'_{RB} + P'_{RB} + P'_{RB} \\ &= 4.3333\ P'_{RB} \end{aligned}$$

- So:

$$P'_{RB} = \frac{1}{4.3333} = 0.23077$$

- And thus:

$$\begin{aligned} P'_A &= 2.3333 \times 0.23077 = 0.53846 \\ P'_{RB} &= 0.23077 \\ P'_{BB} &= 0.23077 \end{aligned}$$

:

Red Bus / Blue Bus

But this is just wrong : it conflicts with what we know to be the case.

We have, in the 3-mode setting:

Mode	We know	Logit says
Auto	0.70	0.538
Red Bus	0.15	0.231
Blue Bus	0.15	0.231

So the logit model has significantly mis-predicted the outcome.

Diagnosing the Problem

- What has gone wrong?
- Think back to the formulation of the logit model. In it we assumed that the random terms (the η's) were i.i.d T1EV random variates.
- But we constructed our red bus / blue bus story by implicitly assuming that (for the two bus modes) not only were the random terms not independent, they were *perfectly correlated*.
- In other words, our story was inconsistent with a fundamental assumption about the logit model.
- This is why a similar counter-example will hold for any choice model assuming independent and identically distributed random terms: it will therefore hold for independent probit, but not for the general case of multinomial probit (which allows for correlated random terms; though the choice probabilities in this case may be very hard to compute).

Another Version of the Problem

- In the problem set, you were asked to formulate an alternative version of the problem.
- This can be stated as: the impact (percent change) on the choice probabilities of two alternatives of a percent change in a characteristic of a *third* (irrelevant) alternative, will be the same.
- In our example, we would expect that the impact of Hyundai changing the price of its cars would be different (greater) on the demand (choice probability) for a KIA than for a Lexus.
- But the logit model says that the impact (in percentage terms) will be the same. But this is surely just wrong.
- This illustrates the fact that perfect correlation (as in the red bus / blue bus case) isn't necessary in order to get anomalous results.

More Questions

This raises at least three questions for our use of the logit model.

1. When can this model be used?
   If the choices in our sample are consistent with the IIA property, then it is safe to use logit. If choices are inconsistent with IIA, then logit may give seriously wrong predictions.
2. How can we tell whether it is safe to use the logit model? That is, how can we tell if are the observed choices in our sample are consistent with the IIA property characterizing the logit model?
3. If logit is not a safe model for our particular context, are there models available that do not assume the IIA property (and hence would be safe to use in this situation)?

Diagnosing IIA

- The literature contains several statistical tests that can be used to decide whether observed choices are (or are not) consistent with IIA.
- They are all based on the following simple idea.
- Suppose choices are consistent with IIA, and we estimate a logit model.
- Then if we (randomly) remove some un-selected (irrelevant) alternatives from the choice sets and re-estimate a logit model with the smaller choice sets, then, up to the randomness induced by our incomplete information, the two sets of estimates should be (approximately) the same.

Diagnosing IIA

- The analytical difficulty consists in deriving a statistical test for the "sameness" of the two logit models.
- A number of authors have done this: among the most often used tests are those by Hausmann and McFadden, and by Small and Hsiao.
- We will not pursue the details of these tests here; but some of the more sophisticated discrete-choice computer packages (including LIMDEP) have ways to compute these tests at least semi-automatically.

Models Without IIA

- Over the years, researchers have developed several choice models that attempt to retain the easy computability of the logit model but do not imply the IIA property.
- Some of these are:
  - nested logit, which groups the alternatives into classes such that IIA holds within a class but not between classes
  - heteroskedastic extreme value, which retains independence but relaxes the assumption of identically distributed random variables over the alternatives.
- In the rest of this note, we describe a third, more recent development in this area.

Random Coefficients

- Think back to our formulation of the statistical version of the logit model.
- There we were forced to assume that all the decision-makers in our sample utilized the *same* decision rule (the same weighting vector β).
- But if you think about it, this is not a very plausible assumption. We would surely expect different people to place different emphases (weights) on the various observable characteristics of the modes in the choice set.
  For example, a poor person may think that mode cost is very important, but a millionaire probably wouldn't care much about cost.

Random Coefficients

- So it would seem to be desirable to allow different individuals to have different weights (β's)
- But how could we estimate a model in which different individuals used different β's? After all, the inability to estimate a single individual's weighting vector was the entire motivation for our taking the statistical (random sampling) approach to estimation in the first place.

Random Coefficients

- The answer is to take an intermediate view between the one-β-for-the-entire-sample, and one-β-per-individual possibilities.
- Instead we will say that the elements of the weighting vector β have a *joint distribution* in the population. Some individuals' β's may be in one tail of the distribution, some in the other tail, and some may be near the center of the distribution.
- So we will assume that β has some joint probability distribution (density) function $g(\beta; \alpha)$ where α is a set of characteristics describing the distribution g. For example, if g was the normal distribution, then α might be the mean vector and covariance matrix.

Random Coefficients

- Note what is going on here: we are introducing yet another level of randomness into the problem.
- There is the randomness associated with our incomplete information (the η's) and now, the additional randomness associated with the variation of the β's in the population.
- This new source of randomness induces another change of perspective: as we will see later, we will not be able to say anything about the β's themselves: instead, we will be able to discuss only what the data reveals about their distribution in the population: that is, our inference will be about the elements of α.

Mixed Logit

- A model in which the η's retain their i.i.d T1EV formulation but where the coefficients β are themselves also random is called the Mixed-Logit (MXL) model.
- So our first question is, does this model help us to resolve the IIA issue? That is, does the mixed-logit model avoid IIA? If not, then nothing will have been achieved (though a model with randomly varying coefficients may be of independent interest).

Mixed Logit — Choice Probabilities

- To answer this question, we now derive an expression for the choice probabilities of the mixed logit model.
- Consider individual i, and let her coefficient vector be β^i.
- Suppose we knew β^i. Then we'd be back in the original model formulation and the *conditional* choice probability — the choice probability *conditional* on our knowledge of β^i — would be the usual ("standard") logit model:

$$P_{ij|\beta^i} = \frac{e^{x_{ij}\beta^i}}{\sum_m e^{x_{im}\beta^i}}$$

  (assuming a linear-in-parameters formulation of v_{ij}).
- But of course we do *not* know β^i. However, we *do* know its distribution in the population: this is $g(\beta^i; \alpha)$

Mixed Logit — Choice Probabilities

- Then we can use Bayes' Theorem (sometimes called in this context the Theorem of Total Probability) to un-conditionalize the conditional choice probabilities.
- We just weight each possible value of β^i by its probability, and then average (integrate) over all possible values.
- Note that this is a version of the strategy we used to derive the mode choice probabilities for the standard (i.i.d) logit model.

Mixed Logit — Choice Probabilities

- The result is:

$$\begin{aligned} P_{ij}(\alpha) &= \int P_{ij|\beta^i}\ g(\beta^i; \alpha)\ d\beta^i \\ &= \int \frac{e^{x_{ij}\beta^i}}{\sum_m e^{x_{im}\beta^i}}\ g(\beta^i; \alpha)\ d\beta^i \end{aligned}$$

  where the region of integration is where g is positive.
- Note that this is deceptively simple-looking: the integral is over all K elements of β^i, that is, this is a K-fold multidimensional integral.
- Also note that when we do the integration, β^i is integrated out, and the result will depend just on α. This justifies our earlier remark that we will be able to discuss only the distribution of the β's, not the β values themselves. (But see the discussion of panel data for a hint as to how this could change).

Mixed Logit and IIA

- If you examine this unconditional choice probability, it should be apparent that if we now compute the odds of choosing mode j over mode h then the denominators of the logit expression do not simply cancel as they did for standard logit.
- The result is that the mixed logit choice probabilities do not imply the IIA property.
- This will be true even if we assume that the elements of β^i are statistically independent.

Mixed Logit : Estimation

- But now we seem to have jumped out of the frying pan into the fire.
- It is true that we have formulated a model that does not imply IIA (and that was the aim of the exercise).
- But on the other hand, we have a model that, in order to compute the choice probabilities, requires us to do a multivariate integration.
- And note that, in our maximum-likelihood setting, we will need to do this potentially many, many times (each time we try out a new guess for a possible value for an element of β).
- So have we really gained anything?

Mixed Logit : Estimation

- Until relatively recently, the received answer was No.
- We have a perfectly reasonable theoretical model of mode choice, and one that does not involve the restrictive IIA property.
- But in practice, because of the integrations required, it was computationally intractable.

Mixed Logit : Estimation

Two things occurred that make the mixed-logit model usable in practice.

- The first was the development of Monte Carlo or simulation methods for evaluating difficult integrals. This has been around for some time — since the 1950's, when it was associated with the mathematician Stanislaw Ulam — and has often been used to compute a single difficult integral.
- The second was the development of very fast computers, with large addressable memory space. This enabled researchers to do the *many* evaluations of the integrals that would be required for estimation of the mixed logit model.

Integration by Simulation

Without going into all the details, the idea behind Monte Carlo evaluation of an integral is as follows:

1. The integral we need to evaluate may be considered as the expectation of a function of a random variable. That is, for a function $f(\beta)$ of the random variable β, its expectation is by definition:

$$E[f(\beta)] = \int f(\beta)\ g(\beta)\ d\beta$$

   where g is the density (frequency) function of β.
2. Expectations are averages. So with a sufficiently large sample, you can compute an expectation by averaging the relevant data.

Integration by Simulation

Putting these two ideas together suggests how we might approximate (simulate) the value of the integral defining the mixed-logit choice probability:

1. Draw a random value (number) — call if $\beta^{i(r)}$ — from the distribution g of the weighting vector β^i. At this point the value of β^i is known (it is $\beta^{i(r)}$) so we are back with the standard logit model.
2. Compute the logit choice probability:

$$P_{ij}^{(r)} = \frac{e^{x_{ij}\beta^{i(r)}}}{\sum_m e^{x_{ij}\beta^{i(r)}}}$$

which is just the logit model evaluated using $\beta^{i(r)}$.

Integration by Simulation

3. Repeat this many times (R times). The result is $P_{ij}^{(1)}, P_{ij}^{(2)}, \ldots P_{ij}^{(R)}$.
4. Average the results:

$$\hat{P}_{ij} = \sum_{r=1}^{R} P_{ij}^{(r)}$$

5. It can be shown that $\hat{P}_{ij}$ is an unbiased estimate of the multidimensional integral defining the mode-choice probability, and an estimate whose variance decreases with R.

.

Estimation by Simulation

- We can use this result to do the equivalent of maximum-likelihood estimation for our mode-choice problem.
- The new procedure, which uses the simulated estimates $\hat{P}_{ij}$ in place of the computed P_{ij} in the ordinary logit model, is called estimation by "maximum simulated likelihood".
- It has been shown that the large sample properties of (ordinary) maximum-likelihood estimation apply here too, when R (the number of random draws used to do the simulation) is large.

Estimation by Simulation

- In practice, R may range from a few hundred to several thousands.
- It should be clear why we need fast computers in order to make this usable in practice: we will need to simulate these choice probabilities repeatedly for each individual in our sample, and for many iterations of our search for the maximum of the (simulated) likelihood function.
- Whereas estimation of the standard logit model usually takes a few seconds, estimation of the mixed logit model can take many minutes, sometimes even hours.

Estimation by Simulation

- Remember that what you get from the mixed logit model is not an estimate of the individual specific weighting parameters, but an estimate of the *distributional* parameters (the α's)
- This means that you will need to make a specific assumption about the joint distribution of the elements of β.
  The computer packages that support estimation of this model allow you to choose between a few selected distributions, usually normal, uniform over some range (and this range is what you will be estimating), and lognormal.

Estimation by Simulation

- The lognormal case is for when you know in advance the sign of a particular element of β : for example, if you are looking at the weight for trip cost, you would expect this to be negative, for all decision-makers.
  A model that assumed that the distribution of this element was normal — which would allow for both positive and negative values — would not be appropriate. (Since the lognormal distribution is positive, in this case you'd say that *minus* the particular element of β is lognormally distributed.

Extension — Panel Data

- Suppose we have a sample in which each individual is represented more than once: that is, we observe the choices that an individual makes on more than one occasion.
- This is known in the literature as observations on a panel of individuals.
- Under the standard logit model, there is no way to distinguish the case where the second, third ... observations are for a single individual and the case where they represent different individuals.
- But under mixed logit the situation changes.

Extension — Panel Data

- That is because when we estimate this model by simulation we draw $\beta^{i(r)}$ for individual i. With repeated observations on this individual, we would use this same draw for all the observations.
- The result is that the likelihood for individual i will be, with repeated observations, the *product* of logit functions (one element for each time that i is observed). This product will then be averaged over, in order to estimate P_{ij}.
- For the next individual, say h, the same thing will apply: we will draw $\beta^{h(r)}$ and use it for all the observations on this individual.

Individual-Level Parameters

- Suppose we observe individual i making multiple mode choices.
- Intuitively, these multiple observations may allow us to say something about the individual parameter vector applicable to this individual.
- This turns our to be true: in the mixed-logit setting we can estimate the expected value of β^i for each individual i. This turns out to be another long simulation of the choice probabilities.

Multinomial Probit

- Recall that we previously noted that the general multinomial probit model was thought to be inestimable, for the case of more than about 4 alternatives.
- As you might suspect, the simulation techniques we have discussed here can also be adopted to this choice model.
- It turns out that the crucial underlying task — drawing a random vector $\beta^{i(r)}$ from a general multinormal (correlated) distribution — is fairly easy, provided we can draw from the standard normal distribution.
- The result is that multinomial probit has now become estimable via simulation.
- But note that in the general case — a covariance matrix with non-zero cross-covariances — there are restrictions on the number that can be separately identified. This is an issue that does not arise for mixed logit. See Train's book (below) for a clear discussion of this.

Software

- Virtually all statistics packages can estimate the standard multinomial logit model, or the binary probit model.
- NLOGIT (a superset of LIMDEP), a commercial program only for Windows, can estimate all the discrete choice models mentioned here (plus some others not mentioned). It is also a full featured general econometrics package, and very convenient to use.
- In the free R system (available for most operating systems) you use add-on packages (also free) to estimate specific models. For the models discussed here:
  - the `mlogit` package can estimate the mixed-logit model in both cross-section and panel data form. It can also estimate other T1EV-based models, including nested logit and heteroskedastic logit.
  - the `mnp` package can estimate the multinomial probit model.
  - packages exist for most other moderately well-known statistical procedures (including spatial statistics, which is not supported in LIMDEP/NLOGIT).

References

The following two books are available free from Ken Train's website at UC Berkeley: they are both excellent texts. The first does not cover mixed logit; the second concentrates on it.

Kenneth E. Train.
Qualitative Choice Analysis.
MIT Press, Cambridge, MA., 1986.

Kenneth E. Train.
Discrete Choice Methods with Simulation.
Cambridge University Press, Cambridge, UK, 2002.