Notes on Data Stacking in NLOGIT

It is important to get a grasp on the stacking stacking of data that is used to estimate multinomial models in NLOGIT (the conversion of 151 observations into 453 rows).

To help understand what is going on, let's look at the data for Assignments #3, #4 and #5. For observation #1 there are 3 rows of data (you can look at the actual data file to see all 453 rows):

x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
1	1	0	0	460	14	48	0	0	2	0	1	0	1	86	0	28
0	0	1	0	440	7	44	0	0	2	0	1	0	1	86	0	28
0	0	0	1	130	7	61	0	0	2	0	1	0	1	86	0	28

The first row has the arterial data, the second row the rural data, the third row has the freeway data for observation #1.

x1 for observation #1 (in column 1, row 1) shows the arterial was chosen (it is one and the other two x1's are zero meaning the rural (in column 1 row 2) and freeway (column 1 row 3) were not chosen). Note that x2, x3, and x4 (columns 2, 3 and 4) are really not needed for anything and should never be used. I have them in these data because I used to run these models with a different software package which required the data in this form.

The variables that vary across alternatives (x5, x6 and x7) change in each row (see columns 5, 6 and 7). The variables that do not (X8 through X17) are the same value in each of the three rows (because all three rows are for the same person).

If you look at x7 (column 7) the values in these stacked data they are 48 for the arterial (meaning that the arterial is 4.8 miles origin to destination), 44 for the rural and 61 for the freeway).

When you do the NLOGIT commands with

u(a)=dist*x7
u(r)=dist*x7
u(f)=dist*x7

Your are estimating one beta (dist) for all three utility equations and NLOGIT is using 48, 44 and 61 for the x7's in the three equations. When you do the NLOGIT commands with

u(a)=dista*x7
u(r)=distr*x7
u(f)=distf*x7

NLOGIT is again using 48, 44 and 61 for the x7's but now a different beta is being estimated for each utility function (dista, distr, distf).

Hope this helps.