HW 4.2

Hayden Atchley

October 13, 2021

In order to estimate the models, we first need to load the data:

```
worktrips <- read_rds("data/worktrips_dfidx.rds")
```

4.5

The first model I estimated with cost, travel time, OVTT, and employment density. I nested this model based on car vs. non-car modes, and constrained the nesting parameter to a single value:

```
nl1 <- mlogit(CHOSEN ~ COST + TVTT  + OVTT | WKEMPDEN,
              data = worktrips,
              nests = list(auto = c('Drive Alone', 'Share 2', 'Share 3+'),
                           nonauto = c('Bike', 'Walk', 'Transit')),
              un.nest.el = T)
my_msummary(list(nested_model = nl1))
```

	nested_model
(Intercept) × Share 2	-2.639 (0.164)***
(Intercept) × Share 3++	-4.289 (0.280)***
(Intercept) × Transit	-1.539 (0.144)***
(Intercept) × Bike	-3.386 (0.225)***
(Intercept) × Walk	-1.149 (0.188)***
COST	-0.003 (0.000)***
TVTT	-0.042 (0.005)***
OVTT	-0.003 (0.008)
WKEMPDEN × Share 2	0.001 (0.000)***
WKEMPDEN × Share 3++	0.003 (0.001)***
WKEMPDEN × Transit	0.003 (0.000)***
WKEMPDEN × Bike	0.001 (0.001)
WKEMPDEN × Walk	0.002 (0.001)**
iv	1.174 (0.071)***

	nested_model
Num.Obs.	5029
Log.Lik.	-3590.769
rho2	0.261
rho20	0.602

Looking at the nesting value gives:

```
nl1$coefficients["iv"]
```

```
##       iv
## 1.173543
```

This is a problematic value, as it is greater than 1. If we used this value, then a decrease in utility would cause people to choose that mode more. Because of this, we should either set the nesting parameter to equal 1 (effectively un-nesting the mode choice), or possibly set the `un.nest.el` argument to be `FALSE`, which would give us different nesting parameters for each nest (some of which may be less than 1 and usable).

4.6

The next models are segmented based on income, with $50,000 as the cutoff point. I also added vehicles per worker (`VEHBYWRK`) as a covariate:

```
nl2A <- mlogit(CHOSEN ~ COST + TVTT  + OVTT | WKEMPDEN + VEHBYWRK,
               data = worktrips %>% filter(HHINC < 50),
               nests = list(auto = c('Drive Alone', 'Share 2', 'Share 3+'),
                            nonauto = c('Bike', 'Walk', 'Transit')),
               un.nest.el = T)

nl2B <- mlogit(CHOSEN ~ COST + TVTT  + OVTT | WKEMPDEN + VEHBYWRK,
               data = worktrips %>% filter(HHINC > 50),
               nests = list(auto = c('Drive Alone', 'Share 2', 'Share 3+'),
                            nonauto = c('Bike', 'Walk', 'Transit')),
               un.nest.el = T)

my_msummary(list(`< $50,000` = nl2A, `> $50,000` = nl2B))
```

	< $50,000	> $50,000
(Intercept) × Share 2	-1.376 (0.193)***	-1.556 (0.240)***
(Intercept) × Share 3++	-2.794 (0.385)***	-2.945 (0.458)***
(Intercept) × Transit	-0.226 (0.234)	-0.633 (0.241)**
(Intercept) × Bike	-1.487 (0.363)***	-1.981 (0.510)***
(Intercept) × Walk	0.609 (0.292)*	-0.276 (0.448)

	< $50,000	> $50,000
COST	-0.003 (0.000)***	-0.003 (0.000)***
TVTT	-0.049 (0.006)***	-0.035 (0.007)***
OVTT	0.015 (0.011)	-0.009 (0.012)
WKEMPDEN × Share 2	0.001 (0.001)	0.001 (0.000)*
WKEMPDEN × Share 3++	0.002 (0.001)***	0.001 (0.001)
WKEMPDEN × Transit	0.004 (0.001)***	0.003 (0.000)***
WKEMPDEN × Bike	0.001 (0.002)	0.001 (0.001)
WKEMPDEN × Walk	0.003 (0.001)***	0.001 (0.001)
VEHBYWRK × Share 2	-0.268 (0.088)**	-0.479 (0.112)***
VEHBYWRK × Share 3++	-0.128 (0.122)	-0.345 (0.167)*
VEHBYWRK × Transit	-1.199 (0.135)***	-0.795 (0.146)***
VEHBYWRK × Bike	-0.835 (0.237)***	-1.028 (0.351)**
VEHBYWRK × Walk	-0.882 (0.159)***	-1.001 (0.264)***
iv	0.862 (0.096)***	0.915 (0.120)***
Num.Obs.	2438	2591
Log.Lik.	-1816.535	-1696.004
rho2	0.297	0.245
rho20	0.584	0.635

In this case, the nesting parameters are both reasonable (between 0 and 1), so we can take them as given. A few interesting items of note:

- The employment density affects all modes other than `Drive Alone` positively, but especially so for transit and walking. Employment density has a greater effect on the mode choice for lower-income households than for high-income ones. This makes sense especially for walking, since walking costs nothing but is usually only feasible in more urban areas.
- The value of time (VOT) can be calculated for the two groups by taking the `TVTT` coefficient divided by the `COST` coefficient and adjusting units:

```
tibble(VOT = "$ / hr",
       `< $50,000` = nl2A$coefficients["TVTT"] / nl2A$coefficients["COST"] * 60/100,
       `> $50,000` = nl2B$coefficients["TVTT"] / nl2B$coefficients["COST"] * 60/100) %>%
  mutate_if(is.numeric, round, 2) %>%
  my_flextable() %>%
  vline(j = 1)
```

VOT	< $50,000	> $50,000
$ / hr	9.81	6.84

This is a bit counter-intuitive, as we would expect the higher-income people to have a higher value of time. However, it may make sense, as higher-income jobs can often be more forgiving of tardiness than lower-income jobs (which includes many service and similar jobs).

4.7

Of all the models calculated so far (including the previous homework segment), the final two seem to be the best in terms of likelihood, with the best ρ^2 and log likelihood values. They also seem to be relatively reasonable, with most coefficients having expected signs and believable values. However, the lower-income model does have a couple oddities, in the form of the `Walk` intercept and the `OVTT` coefficient: both are positive, and probably should be negative. It's possible that at least the positive `Walk` intercept may in part be explained by the fact that lower-income people are less likely to own as many vehicles. Since we don't have household vehicle ownership as a factor, this may be picked up in the intercept. However, the `OVTT` coefficient is less explainable. However, it is not statistically significant, so it may be better not to include it in the model at all.