ECONOMETRICS II, Fall 2023.

Bent E. Sørensen

1 Multivariate, Multinomial, Ordered, Sequential Discrete Probability Models

1.1 Multivariate Logit

A Multivariate Logit is a model where an agent can choose several alternatives. Buy a new car, buy a new house, etc. These would just the probability of say, buying a house AND buying a car would just be

$$\frac{\exp(X_i \beta_k)}{1+\exp(X_{ik}\beta_k)} \frac{\exp(W_i \gamma_k)}{1+\exp(W_i \gamma_k)} ,$$

which is just the product of the product of two logit probabilities. (Of course, if you have theory that some of the γ coefficients are identical to some of the β coefficients, you need to estimate it simultaneously.)

1.2 Multivariate Probit

The multivariate probit often makes more sense as it allows for correlated outcomes. Maybe you buy a new car and a new house at the same time because you got a big raise. You can think of many examples. Assume you have, for each agent i, a vector of latent Normal random variables which for $y^0_{i1}, ..., y^0_{iK}$ (independent across agents) with mean

$$E y^0_{ik} = X_i \beta_k \,,$$

and variance-covariance matrix Σ. You set

$$P(y_{ik} = 1) = P(y^0_{ik} > 0) \,.$$

So, for the case of $K = 2$, you get, for example,

$$P(y_{i1} = 0, y_{i2} = 0) = \int_{-\infty}^{0} \int_{-\infty}^{0} f(y^0_{i1}, y^0_{i2}) \, dy^0_{i1} \, dy^0_{i2} \,.$$

Here, f is the bivariate normal density and you plug in the means and the variance. If there are many choices, this may be time consuming and there are numerical "short-cuts" that I will not cover here.

1.3 Multivariate Linear Probability Model

As for the bivariate choice model, you might simply estimate a multivariate regression model, ignoring that some predicted probabilities would not be between 0 and 1. We will cover the multivariate regression model soon.

1.4 Multinomial Logit

Consider the case of a discrete probability model where the outcome is *one* of the values $k = 1, .., K$. For the multinomial logit, the model it is simple:

$$P(y_i = k) = \frac{\exp(X_i \beta_k)}{\text{Denom}} \, ,$$

where you normalize (say) $\beta_1 = 0$ and

$$\text{Denom} = \Sigma_1^K \exp(X_i \beta_k) \, .$$

Because only one event can happen at a time, the probabilities have to sum to unity. This is a simple generalization of the bivariate logit.

1.5 Multinomial Probit

The multinomial probit is a lot harder computationally. Assume you have, for each agent i, a vector of latent Normal random variables which for $y_{i1}^0, ..., y_{iK}^0$ (independent across agents) with mean

$$y_{ik}^0 = X_i \beta_k + u_{ik} \, ,$$

where the error terms are normally distributed with mean 0 and and variance-covariance matrix Σ with variances on the diagonal normalized to unity. You set

$$P(y_i = k) = P(y_{ik}^0 > y_{ij}^0; \;\; \forall \, j \neq k) \, .$$

Normalize y^0_{i1} to 0. This becomes the standard probit then for $K = 2$. For $K = 3$ you then get the probability that $y_i = 2$ as the probability that $X_i\beta_2 + u_{i2} > 0$ and $X_i\beta_3 + u_{i3} < X_i\beta_2 + u_{i2}$:

$$P(y_i = 2) = \int_{-X_i\beta_2}^{\infty} \int_{-\infty}^{X_i\beta_2+u_{i2}-X_i\beta_3} f(u_{i2}, u_{i3})\, du_{i3}\, du_{i2}\,,$$

and the probability that $y_i = 3$:

$$P(y_i = 3) = \int_{-X_i\beta_2}^{\infty} \int_{X_i\beta_2+u_{i2}-X_i\beta_3}^{\infty} f(u_{i2}, u_{i3})\, du_{i3}\, du_{i2}\,.$$

This may look a bit evil, but it simple integrates the joint density of the latent outcomes 2 and 3 over the area where $y^0_{i3} > 0$ (because otherwise, the first outcome would "be chosen" over the third) and where $y^0_{i3} > y^0_{i2}$. But unless you have optimized software, the integration (over $K - 1$ latent variables) quickly becomes daunting. And if K is large, there is a lot of parameters in the variance-covariance matrix. (Even if you have to normalize one of them to unity.) I will not ask you to write down the integral, but you should know the rest.

If you use the multinomial logit, you can get criticized for assuming that the outcomes are independent. This is known as the "red bus-blue bus problem." Assume you need to commute to work and you can drive your car or take a bus. However, the bus company has some blue busses and some red busses. Assume, as is likely, that no-one cares much about the color of the bus, you will get imprecise estimates if you model the three outcomes (drive, red bus, blue bus). Or worse, the blue busses get painted red, and now you assign the probability of taking a blue bus partly to the red bus and partly to driving. Which of course is nuts, but the multivariate logit does not allow you estimate that the bus-outcomes are highly correlated. (For this case, I would use a nested logit or probit, see below.) Sometimes it is a less obvious, so you need to consider whether a multivariate logit is suitable. In Stata, there is an "mprobit" command that does multivariate probit assuming independent outcomes, but there isn't much point in using this a logit and probits give similar results, but there is also an "asmprobit," that allows for correlation between outcomes and uses indirect inference (think GMM) rather than taking high dimensional integrals.

1.6 Ordered Logit/Probit

Sometimes we have clearly ordered outcomes. For example, the events getting promoted, not getting promoted, getting laid off can be modeled as ordered outcomes (although there may be a few people who would prefer to be laid off, in applications we typically ignore that). Scoring A, B, or C at an exam is a clear example. (Winning Gold, Silver, or Bronze is not a good example, because we would here need to model competing agents and not just one agent.)

We label the outcomes $k = 1, ..., K$. To limit notation, I will assume $K = 3$ here and that 1 is the best outcome. For the probit, we use the latent variable model again (for $\kappa_1 > \kappa_2$):

$$P(y_i = 1) = P(X_i\beta + u_i > \kappa_1)\,,$$

$$P(y_i = 2) = P(\kappa_2 < X_i\beta + u_i < \kappa_1)\,,$$

$$P(y_i = 3) = P(X_i\beta + u_i < \kappa_2)\,.$$

(This formulation assume there is no constant in X_i otherwise on κ has to be set to zero. In the two dimensional case, that is what we implicitly assumed, but I include to constants here, so you can see how it goes if you have more than two choices.) Or, in the likelihood function

$$P(y_i = 1) = \Phi(X_i\beta - \kappa_1)\,,$$

$$P(y_i = 2) = \Phi(X_i\beta - \kappa_2) - \Phi(X_i\beta - \kappa_1)\,,$$

$$P(y_i = 3) = 1 - \Phi(X_i\beta - \kappa_2)\,.$$

You estimate both β and the κ's and if you model makes sense the estimated κ's will have the right relative magnitudes.

For the ordered logit you have

$$P(y_i = 1) = \frac{\exp(X_i\beta - \kappa_1)}{1 + \exp(X_i\beta - \kappa_1)}\,,$$

$$P(y_i = 2) = \frac{\exp(X_i\beta - \kappa_2)}{1 + \exp(X_i\beta - \kappa_2)} - \frac{\exp(X_i\beta - \kappa_1)}{1 + \exp(X_i\beta - \kappa_1)}\,,$$

$$P(y_i = 3) = 1 - \frac{\exp(X_i\beta - \kappa_2)}{1 + \exp(X_i\beta - \kappa_2)}\,,$$

etc. where you replace the Normal CDF in the previous with logistic probabilites.

1.7 Sequential Logit/Probit

For a problem like the choosing to drive or take the red or blue bus (or more realistically train/bus), it often make more sense to use sequential model. First you decide whether to drive or take public transportation, and if you decide to take public transportation you choose between train or bus. If P_{car} is the probability of taking the car, and P_{train}

and P_{bus} are the other probabilities. The probability of taking the care is P_{car}, the probability of taking the bus is $(1 - P_{car})P_{bus}$, etc. For the probabilities you put your favorite probit or logit which would be a function of covariates. It is simple to generalize this to the slightly more complicated problem where the car drive further chooses between his/her Toyota or BMW or even whether you stop for gas. The first of second step can be a multinomial or a multivariate model, it is all modular.