Chapter 16 - Random Parameter Models

Traditional models

- Treat parameters as constant across observations

- Problem:
  - Consider the price of fuel effect on miles driven
  - Will effect be the same across all observations?
  - Or, will high-income households be less sensitive to fuel prices?

Unobserved Heterogeneity

- Factors other than income may affect individuals' sensitivity to gas prices, for example
- Unobserved factors may result in parameters that vary across observations.

Solution: Random Parameter models

- Allows parameter values to vary across the population according to some pre-specified distribution
- If a parameter is found to vary significantly across observations, it implies that each observation has its own parameter
- This makes model estimation much more complex

Random Parameters Multinomial Logit Model (Mixed Logit Model)

- As shown in Chapter 13, the assumption that the disturbances are extreme value Type I distributed gives the standard multinomial logit form as,

$$P_n\left(i\right) = \frac{EXP\left[\boldsymbol{\beta}_i \mathbf{X}_{in}\right]}{\sum_{\forall I} EXP\left(\boldsymbol{\beta}_I \mathbf{X}_{In}\right)}$$

Where:

$P_n(i)$ = probability of observation n having discrete outcome i ($i \in I$ with I denoting all possible outcomes for observation n).

- Now define a mixed model (a model with a mixing distribution) whose outcome probabilities are defined as $P_n^m(i)$ with

$$P_n^m(i) = \int_{\mathbf{X}} P_n(i) \;\; f(\boldsymbol{\beta} \mid \boldsymbol{\varphi})\, d\boldsymbol{\beta}$$

Where:

$f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ = density function of $\boldsymbol{\beta}$ with $\boldsymbol{\varphi}$ referring to a vector of parameters of that density function (mean and variance), and all other terms are as previously defined.

Substituting this equation into the standard logit equation gives the mixed logit model,

$$P_n^m(i) = \int_{\mathbf{X}} \frac{EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]}{\sum_I EXP[\boldsymbol{\beta}_i \mathbf{X}_{In}]} f(\boldsymbol{\beta} \mid \boldsymbol{\varphi}) \, d\boldsymbol{\beta}$$

- Here the mixed logit probabilities $P_n^m(i)$ are the weighted average of the standard multinomial logit probabilities $P_n(i)$ with the weights determined by the density function $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$.
- Note that in the simplified case where $f(\boldsymbol{\beta}|\boldsymbol{\varphi}) = 1$, the model reduces to the standard multinomial logit.

- With the mixed logit, $\boldsymbol{\beta}$ can now account for observation-specific variations of the effect of $\mathbf{X}$ on outcome probabilities, with the density function $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}$.

- Mixed logit probabilities are thus a weighted average for different values of $\boldsymbol{\beta}$ across observations where some elements of the parameter vector $\boldsymbol{\beta}$ are fixed parameters and some are random.

- Any form of the density function $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ in model estimation (such as a normal distribution) can be used.

- Mixed logit models do not suffer from the independence of irrelevant alternatives problem because the ratio of any two outcome probabilities is no longer independent of any other outcomes' probabilities

Random Parameter Model Estimation

For the mixed logit, the log-likelihood is:

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} \delta_{in} LN\left[P_n^m(i)\right]$$

Where:

- N is the total number of observations,
- I is the total number of outcomes,
- δ_{in} is defined as being equal to 1 if the observed discrete outcome for observation n is i and zero otherwise.

- The mixed logit probabilities $P_n^m(i)$ are approximated by drawing values of $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ given values of $\boldsymbol{\varphi}$ and using these drawn values to estimate the simple logit probability $P_n(i) = EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]/\Sigma_i EXP[\boldsymbol{\beta}_i \mathbf{X}_{in}]$

- This procedure is repeated across many samples and the computed logit probabilities are summed and averaged to obtain the simulated probability $\hat{P}_n^m(i)$ to compute the likelihood function.

- How best to draw values of $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ so that accurate approximations of the probabilities are obtained with as few draws as possible?
- Random draws?
- Use Halton sequences (or Halton draws), which are based on a technique developed by Halton (1960) to generate a systematic non-random sequence of numbers.
- Halton draws(samples) are significantly more efficient than purely random draws, arriving at accurate probability approximations with far fewer draws

Random Parameter Count Models

From Chapter 11:

Poisson:

$$P\left(y_n\right)=\frac{EXP\left(-\lambda_n\right)\lambda_n^{y_n}}{y_n!} \qquad (16.6)$$

With:

$$\lambda_n = EXP\left(\boldsymbol{\beta}\mathbf{X}_n\right) \qquad (16.7)$$

Negative Binomial has:

$$\lambda_n = EXP(\boldsymbol{\beta}\mathbf{X}_n + \varepsilon_n)$$

Where:

- $EXP(\varepsilon_n)$ is a Gamma-distributed error term with mean 1 and variance α.

To introduce random parameters:

$$\beta_n = \beta + \omega_n$$

Where:

- ω_n is a randomly distributed term (for example a normally distributed term with mean zero and variance σ^2).
- With this equation, the Poisson parameter becomes $\lambda_n|\boldsymbol{\omega}_n = EXP(\boldsymbol{\beta}_n\mathbf{X}_n)$ in the Poisson model
- And $\lambda_n|\boldsymbol{\omega}_n = EXP(\boldsymbol{\beta}_n\mathbf{X}_n + \varepsilon_n)$ in the negative binomial
- With the corresponding probabilities for Poisson or negative binomial now $P(y_i|\boldsymbol{\omega}_i)$. With this random parameters

- With this random parameters version of the model, the log-likelihood is written as,

$$LL = \sum_{\forall n} ln \int_{\boldsymbol{\omega}_n} g(\boldsymbol{\omega}_n) P(y_n | \boldsymbol{\omega}_n) \, d\boldsymbol{\omega}_n$$

- Where $g(.)$ is the probability density function of the $\boldsymbol{\omega}_i$.

- Simulation-based maximum likelihood method is again used (with Halton draws again being an efficient alternative to random draws).

Random Parameter Duration Models

- Instead of having the explanatory variables act as $EXP(\boldsymbol{\beta}\mathbf{X}_n)$ as in

$$h(t|\mathbf{X}) = h_0(t)EXP(\boldsymbol{\beta}\mathbf{X})$$

- A randomly distributed term ($\boldsymbol{\omega}_n$) is introduced and explanatory variables now act on the hazard as $EXP(\boldsymbol{\beta}_n\mathbf{X}_n)$, where $\boldsymbol{\beta}$ now varies across n observations.

- As with the two random parameter models presented previously, a simulation-based maximum likelihood method is again used (with Halton draws again being an efficient alternative to random draws).