Random Parameter Models

- Previous modeling approaches treat parameters as constant across observations (effect of any individual explanatory variable is the same for each observation).

- But unobserved factors may suggest that the estimated parameters may vary from one observation to the next.

- Random parameters models account for the influences of this unobserved heterogeneity.

Random Parameters Multinomial Logit Model (Mixed Logit Model)

- Consider a function determining discrete outcome probabilities as shown in Chapter 13,

$$T_{in} = \boldsymbol{\beta}_i \mathbf{X}_{in} + \varepsilon_{in} \quad . \tag{16.1}$$

Where:

$\boldsymbol{\beta}_i$ is a vector of estimable parameters for discrete outcome i,

$\mathbf{X}_{in}$ is a vector of the observable characteristics (covariates) that determine discrete outcomes for observation n, and

ε_{in} is a disturbance term.

As before, if the disturbances are extreme value Type I the standard multinomial logit results,

$$P_n(i) = \frac{EXP\left[\boldsymbol{\beta}_i \mathbf{X}_{in}\right]}{\sum_{\forall I} EXP\left(\boldsymbol{\beta}_I \mathbf{X}_{In}\right)} \tag{16.2}$$

Where: $P_n(i)$ is the 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

$$P_n^m(i) = \int_{\mathbf{X}} P_n(i)\ f(\boldsymbol{\beta} \mid \boldsymbol{\varphi})\, d\boldsymbol{\beta} \tag{16.3}$$

Where: $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$ is the 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 Equation 16.2 into Equation 16.3 gives the mixed logit model,

$$P_n^m\left(i\right) = \int_{\mathbf{X}} \frac{EXP\left[\boldsymbol{\beta}_i \mathbf{X}_{in}\right]}{\sum_I EXP\left[\boldsymbol{\beta}_i \mathbf{X}_{In}\right]} f\left(\boldsymbol{\beta} \mid \boldsymbol{\varphi}\right) d\boldsymbol{\beta} \tag{16.4}$$

- Mixed logit probabilities are the weighted average of the standard multinomial logit probabilities $P_n(i)$ with the weights determined by the density function $f(\boldsymbol{\beta}|\boldsymbol{\varphi})$.

- In the simplified case where $f(\boldsymbol{\beta}|\boldsymbol{\varphi}) = 1$, the model reduces to the standard multinomial logit.

- In the mixed logit, $\boldsymbol{\beta}$ can 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 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.

- Estimation of the mixed logit model is undertaken using simulation approaches due to the difficulty in computing the probabilities.

- 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 a "simulated" probability

- The most popular alternative to random draws are 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) have been shown to be significantly more efficient than purely random draws, arriving at accurate probability approximations with far fewer draws.

- 200 Halton draws is the usual standard for acceptable accuracy.

Random Parameter Count Models

Consider the basic Poisson model presented previously in Equation 11.1,

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

Where: y_n is non-negative integer count,

$P(y_n)$ is the probability of observation n having y_n counts per some time period (for example one year) and

λ_n is the Poisson parameter for observation n and is equal to observation n's expected number of counts per year, $E[y_n]$.

Recall from Chapter 11 that a Poisson regression is estimated by setting,

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

where $\mathbf{X}_n$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of estimable parameters. Also recall that negative binomial model is derived by assuming,

$$\lambda_n = EXP(\boldsymbol{\beta}\mathbf{X}_n + \varepsilon_n) \qquad (16.8)$$

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

To allow for random parameters in count-data models, estimable parameters are written as,

$$\beta_n = \beta + \omega_n \tag{16.9}$$

where ω_n is a randomly distributed term (for example a normally distributed).

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 $P(y_i|\boldsymbol{\omega}_i)$.

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

$$LL = \sum_{\forall n} ln \int_{\boldsymbol{\omega}_n} g(\boldsymbol{\omega}_n) P(y_n \mid \boldsymbol{\omega}_n) \, d\boldsymbol{\omega}_n \qquad (16.10)$$

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

- Because probability estimations are computationally cumbersome much like the case for the mixed logit, a simulation-based maximum likelihood method is again used (with Halton draws again being an efficient alternative to random draws).

Random Parameter Duration Models

- Hazard-based models are applied to study the conditional probability of a time duration ending at some time t, given that the duration has continued until time t.

- Developing hazard-based duration models begins with the **cumulative distribution function**,

$$F(t) = P(T < t)$$

  - Where:

    P denotes probability,

    T is a random time variable, and

    t is some specified time.

- The **density function** corresponding to this distribution function (the first derivative of the cumulative distribution with respect to time) is

$$f(t) = dF(t)/dt$$

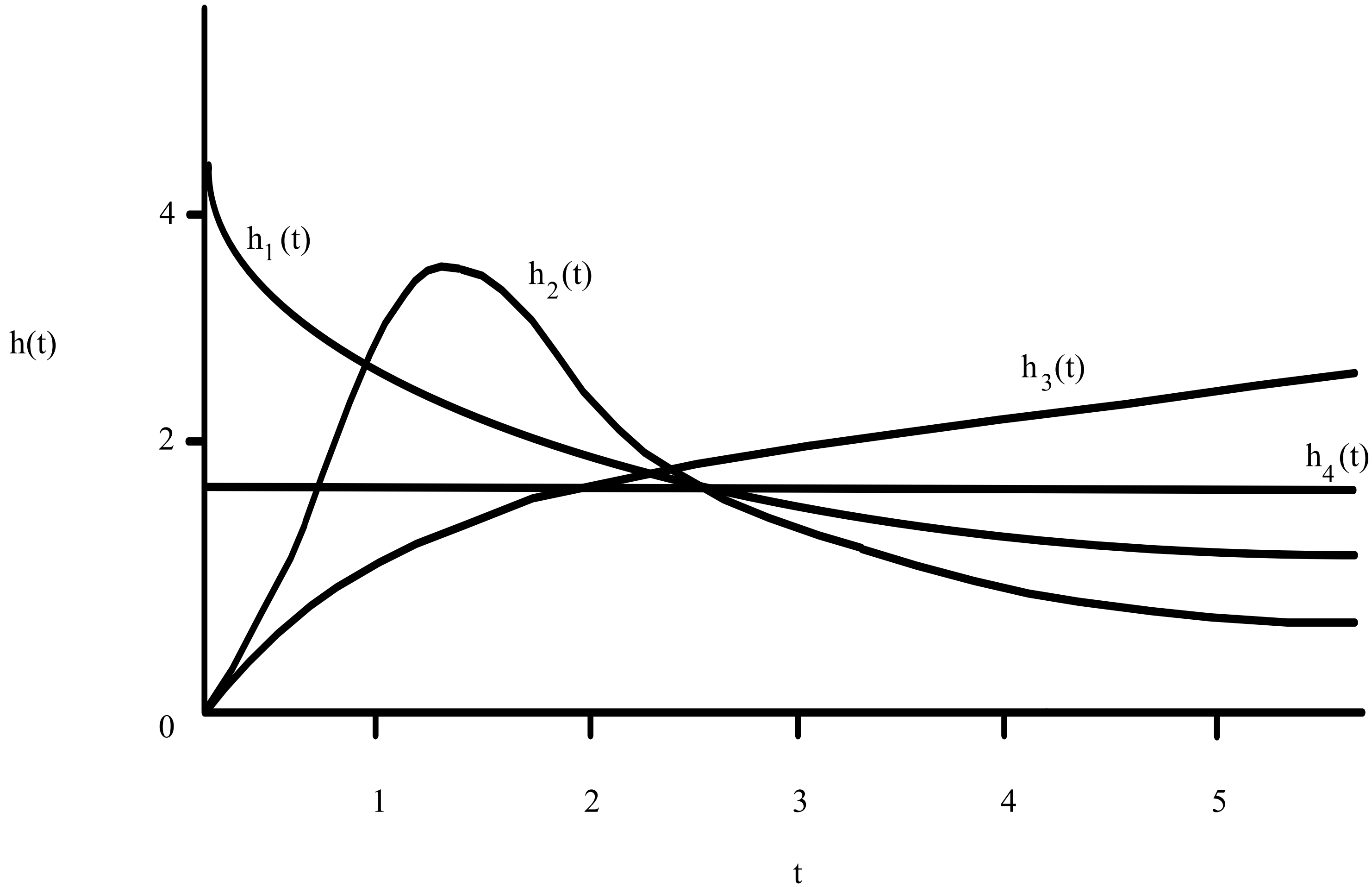
- and the **hazard function** is

$$h(t) = f(t)/[1 - F(t)]$$

  - where: $h(t)$ is the conditional probability that an event will occur between time t and $t + dt$, given that the event has not occurred up to time t.

  - $h(t)$ gives the rate at which event durations are ending at time t, given that the event duration has not ended up to time t.

The slope of the hazard function (the first derivative with respect to time) captures **Duration Dependence**:



- **Fully-Parametric Models**
  - With fully-parametric models, typical distributions of the hazard function include gamma, exponential, Weibull, log-logistic, log-normal and Gompertz distributions, among others.
  - The choice of a specific distribution has important implications relating not only to the shape the underlying hazard, but also to the efficiency and potential biasedness of the estimated parameters.
- The hazard rate with covariates is

$$h(t|X) = h_O(t)EXP(\beta X)$$

Exponential

$$f(t) = \lambda EXP(-\lambda t)$$

- with hazard,

$$h(t) = \lambda$$

- This distribution's hazard is constant, as illustrated by $h_4(t)$ in the previous figure.
- This means that the probability of a duration ending is independent of time and there is no duration dependence.

Weibull

- A more generalized form of the exponential.
- It allows for positive duration dependence (hazard is monotonic increasing in duration and the probability of the duration ending increases over time), negative duration dependence (hazard is monotonic decreasing in duration and the probability of the duration ending decreases over time) or no duration dependence (hazard is constant in duration and the probability of the duration ending is unchanged over time).
- With parameters $\lambda > 0$ and $P > 0$, the Weibull distribution has the density function,

$$f(t) = \lambda P(\lambda t)^{P\text{-}1} EXP[-(\lambda t)^{P}]$$

- with hazard

$$h(t) = (\lambda P)(\lambda t)^{P-1}$$

- If the Weibull parameter *P* is greater than one, the hazard is monotone increasing in duration (see $h_3(t)$ in Figure);
- If *P* is less than one, it is monotone decreasing in duration (see $h_1(t)$ in Figure);
- If *P* equals one, the hazard is constant in duration and reduces to the exponential distribution's hazard with $h(t) = \lambda$ (see $h_4(t)$ in Figure).
- Because the Weibull distribution is a more generalized form of the exponential distribution, it provides a more flexible means of capturing duration dependence. However, it is still limited because it requires the

hazard to be monotonic over time. In many applications, a nonmonotonic hazard is theoretically justified.

Log-logistic

- The log-logistic distribution allows for nonmonotonic hazard functions and is often used as an approximation of the more computationally cumbersome lognormal distribution.

- The log-logistic with parameters $\lambda > 0$ and $P > 0$ has the density function,

$$f(t) = \lambda P(\lambda t)^{P\text{-}1}[1+(\lambda t)^{P}]^{-2}$$

  - and hazard function

$$h(t) = \frac{(\lambda P)(\lambda t)^{P-1}}{\left[1+(\lambda t)^{P}\right]}$$

- The log-logistic's hazard is identical to the Weibull's except for the denominator.
- If $P < 1$, then the hazard is monotone decreasing in duration (see $h_1(t)$ in Figure 9-2);
- If $P = 1$, then the hazard is monotone decreasing in duration from parameter λ; and if $P > 1$, then the hazard increases in duration from zero to an inflection point, $t_i = (P\text{-}1)^{1/P}/\lambda$, and decreases toward zero thereafter (see $h_2(t)$ in Figure 9-2).

Duration Models with Random Parameters:

- Following the same procedure used for count models, random parameters are introduced into duration models.

- That is, instead of having the explanatory variables act as $EXP(\boldsymbol{\beta}\mathbf{X}_n)$ as shown in $h(t|X) = h_0(t)EXP(\beta X)$, a randomly distributed term ($\boldsymbol{\omega}_n$) is introduced as in Equation 16.9 and explanatory variables now act on the hazard as $EXP(\boldsymbol{\beta}_n\mathbf{X}_n)$, where $\boldsymbol{\beta}$ now varies across n observations.

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