Notes on Non-linear Panel Model

Max Leung

January 16, 2022

1 General Results

1.1 Individual-specific Effects Model

1.1.1 Parametric Models

Specify

$$f(y_{it}|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta}, \gamma)$$

1.1.2 Conditional Mean Model

$$\mathbb{E}[y_{it}|\alpha_i, \boldsymbol{x}_{it}] = g(\alpha_i, \boldsymbol{x}_{it}, \boldsymbol{\beta}) \qquad i = 1, \cdots, N, t = 1, \cdots, T$$

Additive individual-specific effects model:

$$g(\alpha_i, \boldsymbol{x}_{it}, \boldsymbol{\beta}) = \alpha_i + \tilde{g}(\boldsymbol{x}_{it}, \boldsymbol{\beta})$$

Multiplicative individual-specific effects model:

$$g(\alpha_i, \boldsymbol{x}_{it}, \boldsymbol{\beta}) = \alpha_i \tilde{g}(\boldsymbol{x}_{it}, \boldsymbol{\beta})$$

Single-index individual-specific effects model:

$$g(\alpha_i, \boldsymbol{x}_{it}, \boldsymbol{\beta}) = \tilde{g}(\alpha_i + \boldsymbol{x}_{it}\boldsymbol{\beta})$$

Only the first two models can be manipulated by quasi-difference to remove the nuisance variables α_i.

1.1.3 Assumption for Conditional Mean Model

Weakly Exogenous

$$\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{it}) = g(\alpha_i, \boldsymbol{x}_{it}, \boldsymbol{\beta})$$

Strongly/Strictly Exogeneoys

$$\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) = g(\alpha_i, \boldsymbol{x}_{it}, \boldsymbol{\beta})$$

1.2 Fixed Effect Model

1.2.1 Incidental Parameters Problem

When $N \to \infty$ but T is fixed, the number of α_i (incidental parameters) estimated also approach ∞. The estimates of α_i are inconsistent, if other parameters' estimates such as $\boldsymbol{\beta}$ (common parameters) are also"contaminated" and thus inconsistent as a result, we call this Incidental Parameters Problem.

1.2.2 Parametric model - Conditional MLE

If there exists a Sufficient statistics $\boldsymbol{s}_i$ for α_i, we have

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i) &= f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i) \\ &= \Pi_{t=1}^T f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i) && \text{if independence of t given i} \end{aligned}$$

$$\begin{aligned} ln[L_{COND}(\boldsymbol{\beta},\boldsymbol{\gamma})] &= ln[\Pi_{i=1}^N f(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i)] \\ &= \sum_{i=1}^N ln[f(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i)] \\ &= \sum_{i=1}^N ln[\Pi_{t=1}^T f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i)] \\ &= \sum_{i=1}^N \sum_{t=1}^T ln[f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{s}_i] && \text{assume independence of t given i} \end{aligned}$$

Sufficient statistics exist for static linear panel model with normality, static and dynamic FE logit model (but not probit), static FE Poisson model, etc.

1.2.3 Conditional Mean Model - Quasi-difference and then GMM estimation

For Additive individual-specific effects model (assuming single index) with strong/strict exogeneity, mean-differenced transformation eliminates α_i,

$$\begin{aligned} \mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) &= \alpha_i + g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \\ \mathbb{E}(\bar{y}_i|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) &= \alpha_i + \bar{g}_i(\boldsymbol{\beta}) && \text{where } \bar{g}_i(\boldsymbol{\beta}) = T^{-1}\textstyle\sum_t g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \end{aligned}$$

$$\begin{aligned} \mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) - \mathbb{E}(\bar{y}_i|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) &= \alpha_i - \alpha_i + g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) - \bar{g}_i(\boldsymbol{\beta}) \\ \mathbb{E}(y_{it} - \bar{y}_i|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) &= g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) - \bar{g}_i(\boldsymbol{\beta}) \\ \mathbb{E}((y_{it} - \bar{y}_i) - (g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) - \bar{g}_i(\boldsymbol{\beta}))|\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{iT}) &= 0 \end{aligned}$$

For the same model with weak exogeneity, first-differences transformation also eliminates α_i,
First note that

$$\begin{aligned} \mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{it}) &= \alpha_i + g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \\ \mathbb{E}(y_{it} - \alpha_i - g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{it}) &= 0 \\ \implies \mathbb{E}(y_{it} - \alpha_i - g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) &= 0 \end{aligned}$$

because

$$\mathbb{E}(y_{it} - \alpha_i - g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) = \mathbb{E}(\underbrace{\mathbb{E}(y_{it} - \alpha_i - g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{it})}_{0}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) = 0$$

Thus,

$$\mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) = \alpha_i + g(\boldsymbol{x}'_{it}\boldsymbol{\beta})$$

$$\begin{aligned} \mathbb{E}(y_{i,t-1}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) &= \alpha_i + g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}) \\ \mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) - \mathbb{E}(y_{i,t-1}|\alpha_i,\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) &= \alpha_i + g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) - \alpha_i - g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}) \\ \mathbb{E}((y_{it} - y_{i,t-1}) - (g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) - g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}))|\boldsymbol{x}_{i1},\cdots,\boldsymbol{x}_{i,t-1}) &= 0 \end{aligned}$$

With these population moments conditions, GMM estimation can be performed.

For Multiplicative individual-specific effects model (assuming single index) with strong/strict exogeneity, mean-differenced transformation eliminates α_i,

$$\begin{aligned}
\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) &= \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \\
\mathbb{E}(\bar{y}_i|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) &= \alpha_i \bar{g}_i(\boldsymbol{\beta}) \qquad \text{where } \bar{g}_i(\boldsymbol{\beta}) = T^{-1}\textstyle\sum_t g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \\
\frac{\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT})}{\mathbb{E}(\bar{y}_i|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT})} &= \frac{\alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\alpha_i \bar{g}_i(\boldsymbol{\beta})} \\
\mathbb{E}(y_{it}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) &= \frac{g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\bar{g}_i(\boldsymbol{\beta})}\mathbb{E}(\bar{y}_i|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) \\
&= \mathbb{E}(\frac{g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\bar{g}_i(\boldsymbol{\beta})}\bar{y}_i|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) \\
\mathbb{E}(y_{it} - \frac{\bar{y}_i}{\bar{g}_i(\boldsymbol{\beta})}g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) &= 0
\end{aligned}$$

For the same model with weak exogeneity, first differences transformation also eliminates α_i,

$$\begin{aligned}
\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{it}) &= \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta}) \\
\mathbb{E}(y_{it} - \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{it}) &= 0 \\
\Longrightarrow \mathbb{E}(y_{it} - \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) &= 0
\end{aligned}$$

because

$$\mathbb{E}(y_{it} - \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) = \mathbb{E}(\underbrace{\mathbb{E}(y_{it} - \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{it})}_{0}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) = 0$$

Thus,

$$\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) = \alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})$$

$$\begin{aligned}
\mathbb{E}(y_{i,t-1}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) &= \alpha_i g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}) \\
\frac{\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1})}{\mathbb{E}(y_{i,t-1}|\alpha_i, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1})} &= \frac{\alpha_i g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\alpha_i g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})} \\
\mathbb{E}(y_{it}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) &= \frac{g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})}\mathbb{E}(y_{i,t-1}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) \\
&= \mathbb{E}(\frac{g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})}y_{i,t-1}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) \\
\mathbb{E}(y_{it} - \frac{g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})}y_{i,t-1}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) &= 0
\end{aligned}$$

Similarly,

$$\mathbb{E}(\frac{g(\boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})}{g(\boldsymbol{x}'_{it}\boldsymbol{\beta})}y_{it} - y_{i,t-1}|\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{i,t-1}) = 0$$

1.2.4 Dummy Variable Parametric Model

In general, Incidental Parameters Problem exists for Dummy Variable Parametric Model when $N \to \infty$ with fixed T, except for two special cases: First, $y_{it} \sim N(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta}, \sigma^2)$ (MLE of $\boldsymbol{\beta}$ and α_i is the same as Dummy Variable Least Squares (DVLS) estimates in static linear panel model. As DVLS estimate of $\boldsymbol{\beta}$ is the same as Fixed Effect/Within estimator, which is thus consistent. So, no Incidental Parameter Problem for $\boldsymbol{\beta}$).

Second, $y_{it} \sim P(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))$, there are no Incidental Parameters Problem when using Concentrated Maximum Likelihood Estimation.
The general form is

$$\begin{aligned}
ln[L_{DV}(\overbrace{\boldsymbol{\beta}, \boldsymbol{\gamma}}^{\boldsymbol{\delta}}, \boldsymbol{\alpha})] &= \sum_{i=1}^{N}\sum_{t=1}^{T} ln[f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})] \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} ln[f(y_{it}|\boldsymbol{d}'_{it}\boldsymbol{\alpha} + \boldsymbol{x}'_{it}\boldsymbol{\beta}, \boldsymbol{\gamma})]
\end{aligned}$$

FOC:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\partial ln[f(y_{it}|\boldsymbol{d}_{it}'\hat{\boldsymbol{\alpha}}+\boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}},\hat{\gamma})]/\partial\boldsymbol{\delta}=\mathbf{0}$$
$$\sum_{i=1}^{N}\sum_{t=1}^{T}\partial ln[f(y_{it}|\boldsymbol{d}_{it}'\hat{\boldsymbol{\alpha}}+\boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}},\hat{\gamma})]/\partial\boldsymbol{\alpha}=\mathbf{0}$$

Even $N=dim(\boldsymbol{\alpha})$ is large, it is still computationally feasible to obtain the maximizer of the log-likelihood function (see details in Greene (2004)).

Greene (2004)'s simulation shows that the inconsistency in Incidental Parameters Problem is moderate when $T=20$. Also, the extent of inconsistency are different for different non-linear panel models.

1.2.5 Dynamic FE model

Given weak exogeneity assumption, Additive individual-specific effects model and Multiplicative individual-specific effects model with lagged dependent variable as regressor can be estimated by GMM after first difference transformation. For the former case (additive model), It is a non-linear generalization of Arellano-Bond (AB) estimator in the sense that if $g(.)$ is the identity function, it reduces to AB estimator.

1.3 Random Effect Model

1.3.1 Parametric model - Integrate out α_i and then MLE

Given the joint density of $\boldsymbol{y}_i,\alpha_i$

$$\begin{aligned}
f(\boldsymbol{y}_i,\alpha_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta}) &= f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},\alpha_i)g(\alpha_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})\\
&= f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\alpha_i)g(\alpha_i|\boldsymbol{\eta})\\
&= [\Pi_{t=1}^{T}f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\alpha_i)]g(\alpha_i|\boldsymbol{\eta}) && \text{assume independence of t given i}
\end{aligned}$$

$$\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta}) &= \int f(\boldsymbol{y}_i,\alpha_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})d\alpha_i\\
&= \int [\Pi_{t=1}^{T}f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\alpha_i)]g(\alpha_i|\boldsymbol{\eta})d\alpha_i\\
&= \mathbb{E}[\Pi_{t=1}^{T}f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\alpha_i)]
\end{aligned}$$

In general, there is no closed form solution for the integration (except for special cases like static RE Poisson model with Poisson f and gamma g, or other conjugate pairs). The one-dimension integration can be calculated by Gauss-Hermite Quadrature (a kind of deterministic numerical integration). Moreover, Direct Monte Carlo Integral Estimate (a kind of simulation) can also be used e.g., draw S number of α_i by Rejection Sampling or Importance Sampling and then compute $S^{-1}\sum_s[\Pi_{t=1}^{T}f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\alpha_i^s)]$. If the latter method is used, MLE with the simulated density is called Maximum Simulated Likelihood Estimation (MSLE).

$$\begin{aligned}
ln[L_{RE}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})] &= ln[\Pi_{i=1}^{N}f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})]\\
&= \sum_{i=1}^{N}ln[f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta})]\\
&= \sum_{i=1}^{N}ln[\int[\Pi_{t=1}^{T}f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\boldsymbol{\gamma},\alpha_i)]g(\alpha_i|\boldsymbol{\eta})d\alpha_i]
\end{aligned}$$

1.3.2 Correlated Random Effect Model / Quasi Fixed Effect Model

The Random Effect Model assume α_i is uncorrelated with $\boldsymbol{X}_i$, so there is no OVB. This assumption is too strong, so relax it a bit. Chamberlain (1980 and 1982) suggested

$$\alpha_i=\sum_{t=1}^{T}\boldsymbol{x}_{it}'\boldsymbol{\pi}_t+\xi_i$$

Mundlak (1978) suggested a special case of it ($\boldsymbol{\pi}_t = \boldsymbol{\pi}/T \ \forall t$),

$$\alpha_i = \sum_{t=1}^{T} \boldsymbol{x}_{it}' \boldsymbol{\pi}/T + \xi_i = \bar{\boldsymbol{x}}_i' \boldsymbol{\pi} + \xi_i$$

1.3.3 Dynamic RE model

$y_{i,t-1}$ is one of the regressors. As y_{i0} does not exist/unobservable, we need to care the initial condition.

$$f(\boldsymbol{y}_i | \boldsymbol{X}_i, \overbrace{\boldsymbol{\beta}, \boldsymbol{\gamma}}^{\boldsymbol{\delta}}, \overbrace{\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1}^{\boldsymbol{\delta}_1}, \boldsymbol{\eta}) = \int [\Pi_{t=2}^{T} f(y_{it} | y_{i,t-1}, \boldsymbol{x}_{it}, \boldsymbol{\delta}, \alpha_i)] f_1(y_{i1} | \boldsymbol{x}_{i1}, \boldsymbol{\delta}_1, \alpha_i) g(\alpha_i | \boldsymbol{\eta}) d\alpha_i$$

$f_1(y_{i1} | \boldsymbol{x}_{i1}$ is specified by econometricians. If $T \to \infty$, initial condition does not matter. However, it is important if T is small. Two approaches: Heckman's approach and Wooldridge's approach.

2 Binary Data

2.1 Parametric Model

$$
\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta}) &= \Pi_{t=1}^{T} f(y_{it}|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) && \text{assume independence of t given i} \\
&= \Pi_{t=1}^{T} Pr(y_{it}=1|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})^{y_{it}} Pr(y_{it}=0|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})^{1-y_{it}} \\
&\text{As } y_{it} \text{ is binary, it must follow Bernoulli distribution, its pdf must be correctly specified.} \\
&= \Pi_{t=1}^{T} Pr(y_{it}=1|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})^{y_{it}} (1-Pr(y_{it}=1|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}))^{1-y_{it}} \\
&= \Pi_{t=1}^{T} F_{\varepsilon}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})^{y_{it}} (1-F_{\varepsilon}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}))^{1-y_{it}}
\end{aligned}
$$

$$
\begin{aligned}
Pr(y_{it}=1|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) &= Pr(y_{it}^*>0|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) \\
&= Pr(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i+\varepsilon_{it}>0|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) \\
&= Pr(\varepsilon_{it}>-\boldsymbol{x}_{it}'\boldsymbol{\beta}-\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) \\
&= Pr(\varepsilon_{it}\le \boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) && \text{If pdf of } \varepsilon_{it} \text{ is symmetric} \\
&= F_{\varepsilon}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})
\end{aligned}
$$

If ε_{it} follows standard normal distribution, $F_{\varepsilon}(.)$ is the cdf of standard normal r.v. $\Phi(.)$, the model is called Probit model. If ε_{it} follows logistic distribution, $F_{\varepsilon}(.)$ is the cdf of logistic r.v. $\Lambda(z)=\frac{e^z}{1+e^z}=\frac{1}{1+e^{-z}}$ (the logistic function), the model is called Logit model.

2.2 Conditional Mean Model

$$
\begin{aligned}
\mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{it}) &= Pr(y_{it}=1|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})\cdot 1 + Pr(y_{it}=0|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})\cdot 0 \\
&= Pr(y_{it}=1|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}) \\
&= F_{\varepsilon}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})
\end{aligned}
$$

which is Single-index individual-specific effects model.

$$
F_{\varepsilon}^{-1}(\mathbb{E}(y_{it}|\alpha_i,\boldsymbol{x}_{it})) = \boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i
$$

$F_{\varepsilon}(.)$ is called mean function. $F_{\varepsilon}^{-1}(.)$ is called link function.
If $F_{\varepsilon}(.)=\Phi(.)$, $F_{\varepsilon}^{-1}(.)=\Phi^{-1}(.)$ which is called probit function (no closed form).
If $F_{\varepsilon}(.)=\Lambda(.)$ (the logistic function), $F_{\varepsilon}^{-1}(z)=\Lambda^{-1}(z)=ln(\frac{z}{1-z})$ which is called logit function.

Note that Quasi Difference cannot eliminate the α_i.

2.3 Fixed Effect Model

2.3.1 Parametric model - Conditional MLE: Static FE Logit Model

$$
\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta}) &= \Pi_{t=1}^{T} F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta})^{y_{it}}(1-F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i|\boldsymbol{x}_{it},\alpha_i,\boldsymbol{\beta}))^{1-y_{it}} \\
&= \Pi_{t=1}^{T}(\frac{exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)}{1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)})^{y_{it}}(1-\frac{exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)}{1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)})^{1-y_{it}} \\
&= \Pi_{t=1}^{T}(\frac{exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)}{1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)})^{y_{it}}(\frac{1}{1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i)})^{1-y_{it}} \\
&= \Pi_{t=1}^{T}\frac{exp(y_{it}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))}{(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))^{y_{it}}}\frac{1}{(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))^{1-y_{it}}} \\
&= \frac{\Pi_{t=1}^{T}exp(y_{it}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))}{\Pi_{t=1}^{T}(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))} \\
&= \frac{exp(\sum_{t=1}^{T}y_{it}(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))}{\Pi_{t=1}^{T}(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))} \\
&= \frac{exp(\sum_{t=1}^{T}y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta}+\sum_{t=1}^{T}y_{it}\alpha_i)}{\Pi_{t=1}^{T}(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))} \\
&= \frac{exp(\alpha_i\sum_{t=1}^{T}y_{it})exp((\sum_{t=1}^{T}y_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})}{\Pi_{t=1}^{T}(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))}
\end{aligned}
$$

The sufficient statistic for α_i is $\sum_{t=1}^{T}y_{it}$. Define $\boldsymbol{B}_c := \{\boldsymbol{d}_i\in\boldsymbol{R}^T|\sum_{t=1}^{T}d_{it}=c\}$ and suppose $\boldsymbol{y}_i\in\boldsymbol{B}_c$

$$
\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta},\sum_{t=1}^{T}y_{it}=c) &= \frac{Pr(\boldsymbol{y}_i\cap\sum_{t=1}^{T}y_{it}=c|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}{Pr(\sum_{t=1}^{T}y_{it}=c|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})} && \text{conditional probability} \\
&= \frac{Pr(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}{Pr(\sum_{t=1}^{T}y_{it}=c|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})} && \text{As } \boldsymbol{y}_i\in\boldsymbol{B}_c \\
&= \frac{Pr(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}{Pr(\cup_{\boldsymbol{d}_i\in\boldsymbol{B}_c}(\boldsymbol{d}_i\cap\sum_{t=1}^{T}y_{it}=c)|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})} && \text{total probability} \\
&= \frac{Pr(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i\in\boldsymbol{B}_c}Pr(\boldsymbol{d}_i\cap\sum_{t=1}^{T}y_{it}=c|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})} && \text{As mutually exclusive} \\
&= \frac{Pr(\boldsymbol{y}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i\in\boldsymbol{B}_c}Pr(\boldsymbol{d}_i|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})} && \text{As } \boldsymbol{d}_i\in\boldsymbol{B}_c
\end{aligned}
$$

Above result can be obtained by applying Bayes Theorem directly: $= \frac{Pr(\boldsymbol{y}_i\cap\sum_{t=1}^{T}y_{it}=c|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i\in\boldsymbol{B}_c}Pr(\boldsymbol{d}_i\cap\sum_{t=1}^{T}y_{it}=c|\boldsymbol{X}_i,\alpha_i,\boldsymbol{\beta})}$

$$
\begin{aligned}
&= \frac{\frac{exp(\alpha_i\sum_{t=1}^{T}y_{it})exp((\sum_{t=1}^{T}y_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})}{\Pi_{t=1}^{T}(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))}}{\sum_{\boldsymbol{d}_i\in\boldsymbol{B}_c}\frac{exp(\alpha_i\sum_{t=1}^{T}d_{it})exp((\sum_{t=1}^{T}d_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})}{\Pi_{t=1}^{T}(1+exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}+\alpha_i))}} \\
&= \frac{exp(\alpha_i\overbrace{\sum_{t=1}^{T}y_{it}}^{c})exp((\sum_{t=1}^{T}y_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i\in\boldsymbol{B}_c}exp(\alpha_i\underbrace{\sum_{t=1}^{T}d_{it}}_{c})exp((\sum_{t=1}^{T}d_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})} \\
&= \frac{exp((\sum_{t=1}^{T}y_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i\in\boldsymbol{B}_c}exp((\sum_{t=1}^{T}d_{it}\boldsymbol{x}_{it}')\boldsymbol{\beta})} \\
&= f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\sum_{t=1}^{T}y_{it}=c) && \alpha_i \text{ gone}
\end{aligned}
$$

$$
\begin{aligned}
L_{COND}(\boldsymbol{\beta}) &= \Pi_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = c) \\
&= \Pi_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = c) \\
&= \Pi_{i=1}^{N} \frac{exp((\sum_{t=1}^{T} y_{it}\boldsymbol{x}'_{it})\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i \in \boldsymbol{B}_c} exp((\sum_{t=1}^{T} d_{it}\boldsymbol{x}'_{it})\boldsymbol{\beta})}
\end{aligned}
$$

c cannot be 0 or T as $c = 0$ mean all $y_{it} = 0$ and $c = T$ mean all $y_{it} = 1$. The very high or very low y^*_{it} may only be the result of very high or very low α_i, not related to $\boldsymbol{\beta}$.

Example: $T = 2 \implies \sum_t y_{it} = 1$. c cannot be 0 or 2, so it is 1. $\boldsymbol{B}_1 = \{\boldsymbol{d}_i \in \boldsymbol{R}^2 | \sum_{t=1}^{2} d_{it} = 1\} = \{(d_{i1} = 1, d_{i2} = 0), (d_{i1} = 0, d_{i2} = 1)\}$.

$$
\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{2} y_{it} = 1) &= f(y_{i1}, y_{i2}|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{2} y_{it} = 1) \\
&= \frac{exp((\sum_{t=1}^{2} y_{it}\boldsymbol{x}'_{it})\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i \in \boldsymbol{B}_1} exp((\sum_{t=1}^{2} d_{it}\boldsymbol{x}'_{it})\boldsymbol{\beta})} \\
f(y_{i1} = 1, y_{i2} = 0|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = 1) &= \frac{exp((\sum_{t=1}^{2} y_{it}\boldsymbol{x}'_{it})\boldsymbol{\beta})}{\sum_{\boldsymbol{d}_i \in \boldsymbol{B}_1} exp((\sum_{t=1}^{2} d_{it}\boldsymbol{x}'_{it})\boldsymbol{\beta})} \\
&= \frac{exp((1 \cdot \boldsymbol{x}'_{i1} + 0 \cdot \boldsymbol{x}'_{i2})\boldsymbol{\beta})}{exp((1 \cdot \boldsymbol{x}'_{i1} + 0 \cdot \boldsymbol{x}'_{i2})\boldsymbol{\beta}) + exp((0 \cdot \boldsymbol{x}'_{i1} + 1 \cdot \boldsymbol{x}'_{i2})\boldsymbol{\beta})} \\
&= \frac{exp(\boldsymbol{x}'_{i1}\boldsymbol{\beta})}{exp(\boldsymbol{x}'_{i1}\boldsymbol{\beta}) + exp(\boldsymbol{x}'_{i2}\boldsymbol{\beta})} \\
&= \frac{\frac{exp(\boldsymbol{x}'_{i1}\boldsymbol{\beta})}{exp(\boldsymbol{x}'_{i2}\boldsymbol{\beta})}}{\frac{exp(\boldsymbol{x}'_{i1}\boldsymbol{\beta}) + exp(\boldsymbol{x}'_{i2}\boldsymbol{\beta})}{exp(\boldsymbol{x}'_{i2}\boldsymbol{\beta})}} \\
&= \frac{exp((\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})'\boldsymbol{\beta})}{1 + exp((\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})'\boldsymbol{\beta})} \\
&= \Lambda((\boldsymbol{x}_{i1} - \boldsymbol{x}_{i2})'\boldsymbol{\beta}) \\
&= \Lambda(-(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta}) \\
&= 1 - \Lambda((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta}) && \text{As } \Lambda(.) \text{ is symmetric}
\end{aligned}
$$

Similarly,

$$
f(y_{i1} = 0, y_{i2} = 1|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = 1) = \Lambda((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})
$$

$$
\begin{aligned}
L_{COND}(\boldsymbol{\beta}) &= \Pi_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{2} y_{it} = 1) \\
&= \Pi_{i=1}^{N} f(y_{i1}, y_{i2}|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{2} y_{it} = 1) \\
&= \Pi_{i=1}^{N} [f(y_{i1} = 0, y_{i2} = 1|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = 1)]^{y_{i2}} [f(y_{i1} = 1, y_{i2} = 0|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = 1)]^{1-y_{i2}} \\
&= \Pi_{i=1}^{N} [\Lambda((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})]^{y_{i2}} [1 - \Lambda((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})]^{1-y_{i2}}
\end{aligned}
$$

So, it becomes cross-sectional Logit model with $\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1}$ as regressor and y_{i2} as dependent variable (individual i with $(y_{i1} = 0, y_{i2} = 0)$ and $(y_{i1} = 1, y_{i2} = 1)$ are excluded.

Example: $T = 3 \implies \sum_t y_{it} = 1$ or $\sum_t y_{it} = 2$. c cannot be 0 or 3, so it is 1 or 2. $\boldsymbol{B}_1 = \{\boldsymbol{d}_i \in \boldsymbol{R}^3 | \sum_{t=1}^{3} d_{it} = 1\}$

and $\boldsymbol{B}_2 = \{\boldsymbol{d}_i \in \boldsymbol{R}^3 | \sum_{t=1}^{3} d_{it} = 2\}$.

$$\begin{aligned}
L_{COND}(\boldsymbol{\beta}) &= \Pi_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{3} y_{it} = 1) f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{3} y_{it} = 2) \\
&= \Pi_{i=1}^{N} f(y_{i1}, y_{i2}, y_{i3}|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{3} y_{it} = 1)^{1(\sum_{t=1}^{3} y_{it}=1)} f(y_{i1}, y_{i2}, y_{i3}|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{3} y_{it} = 2)^{1(\sum_{t=1}^{3} y_{it}=2)}
\end{aligned}$$

In general T,

$$L_{COND}(\boldsymbol{\beta}) = \Pi_{i=1}^{N} \Pi_{c\neq 0, c\neq T} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it} = c)^{1(\sum_{t=1}^{T} y_{it}=c)}$$

2.3.2 Dynamic FE Logit Model

Assume no covariate $\boldsymbol{x}_{it}$. Similar to static case,

$$\begin{aligned}
f(\boldsymbol{y}_i|\alpha_i, \gamma) &= \frac{exp(\alpha_i \sum_{t=2}^{T} y_{it}) exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{\Pi_{t=2}^{T}(1 + exp(\gamma y_{i,t-1} + \alpha_i))} \\
&= \frac{exp(\alpha_i \sum_{t=2}^{T} y_{it}) exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{(1 + exp(\alpha_i))^{\sum_{t=2}^{T}(1-y_{i,t-1})}(1 + exp(\alpha_i + \gamma))^{\sum_{t=2}^{T} y_{i,t-1}}} \\
&= \frac{exp(\alpha_i \sum_{t=2}^{T} y_{it}) exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{(1 + exp(\alpha_i))^{(T-1)-\sum_{t=2}^{T} y_{i,t-1}}(1 + exp(\alpha_i + \gamma))^{\sum_{t=2}^{T} y_{i,t-1}}} \\
&= \frac{exp(\alpha_i \sum_{t=2}^{T} y_{it}) exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{(1 + exp(\alpha_i))^{(T-1)-(y_1 - y_T + \sum_{t=2}^{T} y_{it})}(1 + exp(\alpha_i + \gamma))^{y_1 - y_T + \sum_{t=2}^{T} y_{it}}}
\end{aligned}$$

Define $\boldsymbol{C}_i := \{\boldsymbol{d}_i \in R^T | d_{i1} = y_{i1}, d_{iT} = y_{iT}, \sum_{t=1}^{T} d_{it} = \sum_{t=1}^{T} y_{it}\}$. Similarly to static case,

$$\begin{aligned}
f(\boldsymbol{y}_i|\alpha_i, \gamma, y_{i1}, y_{iT}, \sum_{t=1}^{T} y_{it}) &= \frac{Pr(\boldsymbol{y}_i|\alpha_i, \gamma)}{\sum_{\boldsymbol{d}_i \in \boldsymbol{C}_i} Pr(\boldsymbol{d}_i|\alpha_i, \gamma)} \\
&= \frac{\frac{exp(\alpha_i \sum_{t=2}^{T} y_{it}) exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{(1+exp(\alpha_i))^{(T-1)-(y_1 - y_T + \sum_{t=2}^{T} y_{it})}(1+exp(\alpha_i+\gamma))^{y_1 - y_T + \sum_{t=2}^{T} y_{it}}}}{\sum_{\boldsymbol{d}_i \in \boldsymbol{C}_i} \frac{exp(\alpha_i \sum_{t=2}^{T} d_{it}) exp((\sum_{t=2}^{T} d_{it} d_{i,t-1})\gamma)}{(1+exp(\alpha_i))^{(T-1)-(d_1 - d_T + \sum_{t=2}^{T} d_{it})}(1+exp(\alpha_i+\gamma))^{d_1 - d_T + \sum_{t=2}^{T} d_{it}}}} \\
&= \frac{exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{\sum_{\boldsymbol{d}_i \in \boldsymbol{C}_i} exp((\sum_{t=2}^{T} d_{it} d_{i,t-1})\gamma)} \\
&= f(\boldsymbol{y}_i|\gamma, y_{i1}, y_{iT}, \sum_{t=1}^{T} y_{it}) \qquad \alpha_i \text{ is gone}
\end{aligned}$$

$$\begin{aligned}
L_{COND}(\boldsymbol{\beta}) &= \Pi_{i=1}^{N} f(\boldsymbol{y}_i|\alpha_i, \gamma, y_{i1}, y_{iT}, \sum_{t=1}^{T} y_{it}) \\
&= \Pi_{i=1}^{N} f(\boldsymbol{y}_i|\gamma, y_{i1}, y_{iT}, \sum_{t=1}^{T} y_{it}) \\
&= \Pi_{i=1}^{N} \frac{exp((\sum_{t=2}^{T} y_{it} y_{i,t-1})\gamma)}{\sum_{\boldsymbol{d}_i \in \boldsymbol{C}_i} exp((\sum_{t=2}^{T} d_{it} d_{i,t-1})\gamma)}
\end{aligned}$$

T must be at least 4. If $T = 2$ or $T = 3$, $\boldsymbol{C}_i$ is singleton. If $\boldsymbol{y}_i$ is $(0, 1, 0, 1)$, $\boldsymbol{C}_i = \{(0, 1, 0, 1), (0, 0, 1, 1)\}$.

2.4 Random Effect Model

2.4.1 Parametric model - Integrate out α_i and then MLE: Static RE Binary Model

Assume $\alpha_i \sim N(0, \sigma_\alpha^2)$

$$\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma_\alpha^2) &= \int f(\boldsymbol{y}_i, \alpha_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma_\alpha^2) d\alpha_i \\
&= \int [\Pi_{t=1}^T f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{\beta}, \alpha_i)] g(\alpha_i|\sigma_\alpha^2) d\alpha_i \\
&= \int [\Pi_{t=1}^T F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta})^{y_{it}} (1 - F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta}))^{1-y_{it}}] \phi(\frac{\alpha_i}{\sigma_\varepsilon^2}) d\alpha_i \\
&= \mathbb{E}[\Pi_{t=1}^T F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta})^{y_{it}} (1 - F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta}))^{1-y_{it}}]
\end{aligned}$$

The integration can be solved by numerical integration or simulation method.
If ε_{it} follow standard normal distribution, it is called Static RE Probit Model. If ε_{it} follow logistic distribution, it is called Static RE Logit Model.

2.4.2 Dynamic RE Binary Model

Assume $\alpha_i \sim N(0, \sigma_\alpha^2)$

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\beta}_1, \sigma_\alpha^2) = \int [\Pi_{t=2}^T F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta})^{y_{it}} (1 - F_\varepsilon(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta}))^{1-y_{it}}] f_1(y_{i1}|\boldsymbol{x}_{i1}, \boldsymbol{\beta}_1, \alpha_i) \phi(\frac{\alpha_i}{\sigma_\varepsilon^2}) d\alpha_i$$

3 Count Data

3.1 Parametric Model

Assume $y_{it} \sim P(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))$,

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}) &= \Pi_{t=1}^T f(y_{it}|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta}) \\ &= \Pi_{t=1}^T \frac{exp(-\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{y_{it}!} \end{aligned}$$

3.2 Conditional Mean Model

$$\mathbb{E}(y_{it}|\alpha_i, \boldsymbol{x}_{it}) = \alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})$$

This is a kind of multiplicative individual-specific effects model.

3.3 Fixed Effect Model

3.3.1 Parametric Model - Conditional MLE: Static FE Poisson Model

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}, \sum_{t=1}^T y_{it}) &= \frac{f(\boldsymbol{y}_i \cap \sum_{t=1}^T y_{it}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta})}{f(\sum_{t=1}^T y_{it}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta})} && \text{Conditional density} \\ &= \frac{f(\boldsymbol{y}_i|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta})}{f(\sum_{t=1}^T y_{it}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta})} \\ &= \frac{\Pi_{t=1}^T f(y_{it}|\boldsymbol{x}_{it}, \alpha_i, \boldsymbol{\beta})}{f(\sum_{t=1}^T y_{it}|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta})} \\ &= \frac{\Pi_{t=1}^T \frac{exp(-\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{y_{it}!}}{\frac{exp(-\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))(\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{\sum_t y_{it}}}{(\sum_t y_{it})!}} \\ &= \frac{\frac{exp(-\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))\Pi_{t=1}^T(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{\Pi_{t=1}^T y_{it}!}}{\frac{exp(-\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))\Pi_t(\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{(\sum_t y_{it})!}} \\ &= \frac{\frac{\Pi_{t=1}^T(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{\Pi_{t=1}^T y_{it}!}}{\frac{\Pi_t(\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{(\sum_t y_{it})!}} \\ &= \frac{(\sum_t y_{it})!}{\Pi_{t=1}^T y_{it}!} \frac{\Pi_{t=1}^T(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{\Pi_t(\alpha_i \sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}} \\ &= \frac{(\sum_t y_{it})!}{\Pi_{t=1}^T y_{it}!} \frac{\Pi_{t=1}^T(exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{\Pi_t(\sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}} && \alpha_i \text{ is gone} \\ &= \frac{(\sum_t y_{it})!}{\Pi_{t=1}^T y_{it}!} \Pi_{t=1}^T (\frac{exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})})^{y_{it}} && \text{Multinomial distribution} \\ &= f(\boldsymbol{y}_i|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}) \\ &\propto \Pi_{t=1}^T (\frac{exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\sum_t exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})})^{y_{it}} \end{aligned}$$

Log likelihood function is

$$
\begin{aligned}
ln[L_{COND}(\boldsymbol{\beta})] &= ln[\Pi_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \alpha_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it})] \\
&= ln[\Pi_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it})] \\
&= \sum_{i=1}^{N} ln[f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sum_{t=1}^{T} y_{it})] \\
&= \sum_{i=1}^{N} ln[\frac{(\sum_t y_{it})!}{\Pi_{t=1}^{T} y_{it}!} \Pi_{t=1}^{T} (\frac{exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})}{\sum_t exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})})^{y_{it}}] \\
&= \sum_{i=1}^{N} \{ln[\frac{(\sum_t y_{it})!}{\Pi_{t=1}^{T} y_{it}!}] + \sum_{t=1}^{T} ln(\frac{exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})}{\sum_t exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})})^{y_{it}}\} \\
&\propto \sum_{i=1}^{N} \sum_{t=1}^{T} ln(\frac{exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})}{\sum_t exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})})^{y_{it}} && \text{Ommit terms without beta} \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}[\boldsymbol{x}_{it}'\boldsymbol{\beta} - ln(\sum_t exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))]
\end{aligned}
$$

3.3.2 Dummy Variable Parametric Model

Poisson y_{it} is one of the special case that does not lead to Incidental Parameter Problem to $\boldsymbol{\beta}$.

$$
\begin{aligned}
ln[L(\boldsymbol{\beta}, \boldsymbol{\alpha})] &= ln[\Pi_{i=1}^{N} \Pi_{t=1}^{T} \frac{exp(-\alpha_i exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))(\alpha_i exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))^{y_{it}}}{y_{it}!}] \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} ln[\frac{exp(-\alpha_i exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))(\alpha_i exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))^{y_{it}}}{y_{it}!}] \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} ln[\frac{exp(-\alpha_i exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))(\alpha_i^{y_{it}} exp(y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta})}{y_{it}!}] \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} [-\alpha_i exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + y_{it} ln(\alpha_i) + y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} - ln(y_{it}!)] \\
&= \sum_{i=1}^{N} [-\alpha_i \sum_{t=1}^{T} exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + ln(\alpha_i) \sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} - \sum_{t=1}^{T} ln(y_{it}!)]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial ln[L(\boldsymbol{\beta}, \boldsymbol{\alpha})]}{\partial \alpha_i'} &= \frac{\partial \sum_{i=1}^{N} [-\alpha_i \sum_{t=1}^{T} exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + ln(\alpha_i) \sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T} y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} - \sum_{t=1}^{T} ln(y_{it}!)]}{\partial \alpha_i'} \\
&= \frac{\partial [-\alpha_{i'} \sum_{t=1}^{T} exp(\boldsymbol{x}_{i't}'\boldsymbol{\beta}) + ln(\alpha_{i'}) \sum_{t=1}^{T} y_{i't}]}{\partial \alpha_{i'}} \\
&= -\sum_{t=1}^{T} exp(\boldsymbol{x}_{i't}'\boldsymbol{\beta}) + \frac{\sum_{t=1}^{T} y_{i't}}{\alpha_{i'}}
\end{aligned}
$$

FOC:

$$
\begin{aligned}
-\sum_{t=1}^{T} exp(\boldsymbol{x}_{i't}'\boldsymbol{\beta}) + \frac{\sum_{t=1}^{T} y_{i't}}{\widehat{\alpha}_{i'}} &= 0 \\
\frac{\sum_{t=1}^{T} y_{i't}}{\widehat{\alpha}_{i'}} &= \sum_{t=1}^{T} exp(\boldsymbol{x}_{i't}'\boldsymbol{\beta}) \\
\widehat{\alpha}_{i'} &= \frac{\sum_{t=1}^{T} y_{i't}}{\sum_{t=1}^{T} exp(\boldsymbol{x}_{i't}'\boldsymbol{\beta})}
\end{aligned}
$$

Substitute back to $ln[L(\boldsymbol{\beta},\boldsymbol{\alpha})]$ function

$$\begin{aligned}
ln[L(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}})] &= \sum_{i=1}^{N}\sum_{t=1}^{T}[-\widehat{\alpha}_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}) + y_{it}ln(\widehat{\alpha}_i) + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} - ln(y_{it}!)] \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}[-\frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}) + y_{it}ln(\frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})}) + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} - ln(y_{it}!)] \\
&= \sum_{i=1}^{N}\{-\frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})}\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}) + \sum_{t=1}^{T}[y_{it}ln(\frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})}) + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} - ln(y_{it}!)]\} \\
&= \sum_{i=1}^{N}\{-\sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T}[y_{it}ln(\sum_{t=1}^{T} y_{it}) - y_{it}ln(\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})) + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} - ln(y_{it}!)]\} \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\{-y_{it} + y_{it}ln(\sum_{t=1}^{T} y_{it}) - y_{it}ln(\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})) + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta} - ln(y_{it}!)\} \\
&\propto \sum_{i=1}^{N}\sum_{t=1}^{T}\{-y_{it}ln(\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})) + y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta}\} \qquad \text{Ommit terms without beta}
\end{aligned}$$

Which is the same as the Conditional Likelihood Function. Thus, $\widehat{\boldsymbol{\beta}}_{DV} = \widehat{\boldsymbol{\beta}}_{COND}$. As $\widehat{\boldsymbol{\beta}}_{COND}$ is consistent, $\widehat{\boldsymbol{\beta}}_{DV}$ is also consistent. So, there is no Incidental Parameter Problem for beta.

3.4 Random Effect Model

3.4.1 Parametric model - Integrate out α_i and then MLE: Static RE Poisson Model

It is one of the examples that has closed form solution of the integration under some specifications.

$$\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{\beta},\eta) &= \int_0^\infty f(\boldsymbol{y}_i,\alpha_i|\boldsymbol{X}_i,\boldsymbol{\beta},\eta)d\alpha_i \\
&= \int_0^\infty [\Pi_{t=1}^{T} f(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\beta},\alpha_i)]g(\alpha_i|\eta)d\alpha_i \\
&= \int_0^\infty [\Pi_{t=1}^{T}\frac{exp(-\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))(\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))^{y_{it}}}{y_{it}!}]g(\alpha_i|\eta)d\alpha_i \\
&= \int_0^\infty [\frac{\Pi_{t=1}^{T} exp(-\alpha_i exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))\Pi_{t=1}^{T}\alpha_i^{y_{it}}\Pi_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})^{y_{it}}}{\Pi_{t=1}^{T} y_{it}!}]g(\alpha_i|\eta)d\alpha_i \\
&= \int_0^\infty [\frac{\Pi_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})^{y_{it}}}{\Pi_{t=1}^{T} y_{it}!} exp(-\alpha_i\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))\alpha_i^{\sum_{t=1}^{T} y_{it}}]g(\alpha_i|\eta)d\alpha_i \\
&= \Pi_{t=1}^{T}\frac{\exp(\boldsymbol{x}'_{it}\boldsymbol{\beta})^{y_{it}}}{y_{it}!}\int_0^\infty exp(-\alpha_i\sum_{t=1}^{T} exp(\boldsymbol{x}'_{it}\boldsymbol{\beta}))\alpha_i^{\sum_{t=1}^{T} y_{it}} g(\alpha_i|\eta)d\alpha_i
\end{aligned}$$

If α_i follows gamma distribution i.e., $g(\alpha_i|\eta) = \frac{\eta^\eta \alpha_i^{\eta-1} exp(-\alpha_i \eta)}{\Gamma(\eta)}$

$$= \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \int_0^\infty exp(-\alpha_i \sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))\alpha_i^{\sum_{t=1}^T y_{it}} \frac{\eta^\eta \alpha_i^{\eta-1} exp(-\alpha_i\eta)}{\Gamma(\eta)} d\alpha_i$$
$$= \frac{\eta^\eta}{\Gamma(\eta)} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \int_0^\infty exp(-\alpha_i \sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}))\alpha_i^{\sum_{t=1}^T y_{it}} \alpha_i^{\eta-1} exp(-\alpha_i\eta) d\alpha_i$$
$$= \frac{\eta^\eta}{\Gamma(\eta)} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \int_0^\infty exp(-\alpha_i \sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) - \alpha_i\eta)\alpha_i^{\sum_{t=1}^T y_{it}+\eta-1} d\alpha_i$$
$$= \frac{\eta^\eta}{\Gamma(\eta)} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \int_0^\infty exp(-\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta])\alpha_i^{\sum_{t=1}^T y_{it}+\eta-1} \frac{d\alpha_i}{d\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]} d\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]$$
$$= \frac{\eta^\eta}{\Gamma(\eta)} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \int_0^\infty exp(-\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta])\alpha_i^{\sum_{t=1}^T y_{it}+\eta-1} [\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]^{-1} d\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]$$
$$= \frac{\eta^\eta}{\Gamma(\eta)[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]^{\sum_{t=1}^T y_{it}+\eta}} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!}$$
$$\int_0^\infty exp(-\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta])\alpha_i^{\sum_{t=1}^T y_{it}+\eta-1} [\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]^{\sum_{t=1}^T y_{it}+\eta-1} d\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]$$
$$= \frac{\eta^\eta}{\Gamma(\eta)[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]^{\sum_{t=1}^T y_{it}+\eta}} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!}$$
$$\int_0^\infty exp(-\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta])(\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta])^{\sum_{t=1}^T y_{it}+\eta-1} d\alpha_i[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]$$
$$= \frac{\eta^\eta}{\Gamma(\eta)[\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]^{\sum_{t=1}^T y_{it}+\eta}} \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \Gamma(\sum_{t=1}^T y_{it} + \eta)$$
$$= \Pi_{t=1}^T \frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})^{y_{it}}}{y_{it}!} \cdot (\frac{\eta}{\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta})^\eta \cdot [\sum_{t=1}^T exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) + \eta]^{-\sum_{t=1}^T y_{it}} \cdot \frac{\Gamma(\sum_{t=1}^T y_{it} + \eta)}{\Gamma(\eta)}$$

4 References

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications