CHAPTER 8

Some Nonlinear Models

There is no general concensus about incorporating unobserved individual- and time-specific effects in nonlinear models. If the effects are considered independent of the conditional covariates, the random-effects specification will lead to complicated multiple integrations. If the effects are treated as fixed, the estimation of structural parameters is often functions of the specific effects which require the time series dimension T to be large to obtain consistent estimators of individual-specific effects. To obtain a consistent estimator of time-specific effects, the cross-section dimension N needs to go to infinity. In this chapter we consider some popular nonlinear models for empirical analysis. Section 8.1 considers duration models. Section 8.2 considers count data models. Section 8.3 considers nonparametric models.

8.1 DURATION MODEL

8.1.1 Cumulative Distribution of Duration, Hazard Function, Survival Function

The length of time of experiencing an event is a widely considered issue in economics, finance, or health (e.g., the duration of unemployment, the life expectancy of a patient with or without being subject to a given treatment, etc.). The length of the time interval in a given state before transition to another state is called a *duration*. The duration in a given state is a nonnegative random variable, denoted by D. The *cumulative distribution function* of D is defined as

$$\begin{aligned} F(t) &= \text{Prob}\ (D < t) \\ &= \int_0^t f(s)ds, \end{aligned} \tag{8.1.1}$$

where the probability density function of *exit* at t is

$$f(t) = \frac{dF(t)}{dt}. \tag{8.1.2}$$

The function form of $f(t)$ and $F(t)$ are derived from the *hazard function* – the instantaneous probability of leaving the current state (the *death* of a process) conditional on that it is in the state from 0 to t (survives from 0 to t).

$$\mu_t = \lim_{\Delta t \to 0} \frac{\text{Prob}\,[t \le D < t + \Delta t \mid D \ge t]}{\Delta t}.$$
$$= \frac{f(t)}{1 - F(t)}. \tag{8.1.3}$$

To see this, let $A_{it,t+\Delta}$ denote the event that the ith individual stay in a state between t and $t+\Delta t$, then exit at $t+\Delta$. Suppose the probability of event $A_{it,t+\Delta}$ occurs where the time distance between two adjacent time periods approaches zero is given by

$$P(A_{it,t+\Delta}) = \mu_{it}\Delta t. \tag{8.1.4}$$

Under the assumption that μ_{it} stays constant between 0 and $t+s$, $\mu_{it} = \mu_i$, the probability that an individual stayed in a state (say unemployment) from 0 to s and moved out at $t+s = s+\Delta t$ is

$$P(A_{its}) = (1 - \mu_i \Delta t)^{\frac{s}{\Delta t}} \mu_i \Delta t. \tag{8.1.5}$$

Let $n = \frac{s}{\Delta t}$, then $\Delta t \longrightarrow 0, n \longrightarrow \infty$. Using the identity $\lim_{n \longrightarrow \infty} (1 - n^{-1})^n = e^{-1}$, we obtain that for small Δt,

$$P(A_{its}) = \exp\,(-\mu_i s)\mu_i \Delta t = f_i(s)dt. \tag{8.1.6}$$

Then the probability for the ith individual survives at a state over duration t is given by the *survival function*,

$$S_i(t) = \text{Prob}\,(D_i \ge t) = \exp\,(-\mu_i t). \tag{8.1.7}$$

The cumulative distribution function of D_i is

$$\begin{aligned} F_i(t) &= \text{Prob}(D_i < t) \\ &= 1 - \text{Prob}(D_i \ge t) \\ &= 1 - \exp(-\mu_i t). \end{aligned} \tag{8.1.8}$$

The expected duration for the ith individual in a state under the assumption that $S(\infty) = 0$ is

$$E(D_i) = \int_0^\infty t\mu_i \exp\,(-\mu_i t)dt = \frac{1}{\mu_i}. \tag{8.1.9}$$

The relationships between the hazard rate, survival function and the probability density function for exit at time s when μ_{it} is not necessarily a constant,

$$\mu_{it} = \lim_{\Delta t \to 0} \frac{\text{Prob}\,[t \le D_i < t + \Delta t \mid D_i \ge t]}{\Delta t}.$$
$$= \frac{f_i(t)}{1 - F_i(t)}, \tag{8.1.10}$$

can be seen through breaking up the interval $(0,t)$ to n sub-intervals, $n = \frac{t}{\Delta s}$, and noting that the ith individual survives from s to $s+\Delta s$ is equal to $e^{-\mu_{is}\Delta s}$. Using $S(0) = 1$, the probability that the ith individual survives from zero to time t is equal to

$$\prod_{s=1}^{n} \exp\,(-\mu_{is}\Delta s) = \exp\left\{-\sum_{s=1}^{n} \mu_{is}\Delta s\right\}.$$

As $\Delta s \to 0$, it yields the survival function as

$$S_i(t) = \exp\left(-\int_0^t \mu_{is} ds\right). \tag{8.1.11}$$

It follows that

$$\mu_{it} = \text{Prob}\ (D_i = t \mid D_i \geq t) = -\frac{d \ln S_i(t)}{dt} = \frac{f_i(t)}{S_i(t)}, \tag{8.1.12}$$

where $f_i(t) = S_i(t)\mu_{it}$ is the probability density of exit at t.

8.1.2 The Likelihood Function

The likelihood function for an individual with complete spell (or duration) t_i is given by $f(t_i)$. However, many duration data are *right- or left-censored*. Suppose the data on N individuals take the form that each individual experiences either one complete spell at time t_i, i.e., $D_i = t_i$, or right-censored at time t_i^* , i.e., $D_i \geq t_i^*$. Suppose $\mu_{it} = \mu_i$ in a sample of N independent individuals with $i = 1, \ldots, n$ complete their spells of duration t_i; then their joint likelihood function is

$$\prod_{i=1}^{n} f_i(t_i) = \prod_{i=1}^{n} \mu_i \exp\ (-\mu_i t_i). \tag{8.1.13}$$

Suppose $i = n+1, \ldots, N$ are right-censored at t_i^*; then,

$$\prod_{i=n+1}^{N} S_i(t_i^*) = \prod_{i=n+1}^{N} (1 - F_i(t_i^*)) = \prod_{i=n+1}^{N} \exp\ (-\mu_i t_i^*). \tag{8.1.14}$$

The data are *left-censored* if the starting time of a state is unknown; then for those individuals with complete spell at t_i, their likelihood function is given by $S(t_i)$. For those individuals in the sample where neither the starting time nor the exiting time is observed, (i.e. both left- and right-censored), their likelihood function is $E(D_i)$. So, considering a data set that for $i = 1, \ldots, n_1$, neither their starting time nor their exit time is observed; for $i = n_1 + 1, \ldots, n_2$, the exit time $\tilde{t}_i$ is observed, but not the starting time, for $i = n_2 + 1, \ldots, n_3$, the complete spell (duration) is observed; and for $i = n_3 + 1, \ldots, N$, the duration exceeding t_i^* is observed (right-censored), the likelihood function for N independent individual samples takes the form

$$\prod_{i=1}^{n_1} E(D_i) \prod_{i=n_1+1}^{n_2} [1 - F_i(\tilde{t}_i)] \prod_{i=n_2+1}^{n_3} f_i(t_i) \prod_{i=n_3+1}^{N} [1 - F_i(t_i^*)]. \tag{8.1.15}$$

The hazard rate μ_{it} or μ_i is often assumed to be a function of socio-demographic variables $\boldsymbol{x}_i$. Because duration is a nonnegative random variable, μ_i or μ_{it} should be nonnegative. A simple way to ensure this is to let

$$\mu_i = \exp\ (\boldsymbol{x}_i' \boldsymbol{\beta}). \tag{8.1.16}$$

Substituting (8.1.16) into the (8.1.15), one can obtain the maximum likelihood estimator of $\boldsymbol{\beta}$ by maximizing the logarithm of (8.1.15).

Alternatively, noting that

$$\begin{aligned} E \log t_i &= \int_0^\infty (\log t)\ \mu_i \exp\ (-\mu_i t) dt \\ &= -\log \mu_i \int \mu_i \exp(-\mu_i t) dt + \int (\log \mu_i t) \mu_i \exp(-\mu_i t) dt \\ &= -c - \log \mu_i, \end{aligned} \tag{8.1.17}$$

where $c = \int \log z(\exp(-z)dz) \simeq 0.577$ is Euler's constant, and

$$\begin{aligned} \text{Var}\,(\log t_i) &= E(\log t_i)^2 - [E \log t_i]^2 \\ &= \frac{\pi^2}{6}, \end{aligned} \tag{8.1.18}$$

one can put the duration model in a regression framework,

$$\log t_i + 0.557 = -\boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \tag{8.1.19}$$

where $E(u_i) = 0$ and var $(u_i) = \frac{\pi^2}{6}$. Consistent estimate of $\boldsymbol{\beta}$ can be obtained by either the least squares method using the n subsample of individuals that experiences one complete spell or applying the censored least squares or least deviation method to the complete data. However, the least squares estimator has covariance matrix $\frac{\pi^2}{6}\left(\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}$, which is greater than the covariance matrix of the MLE of (8.1.15).

8.1.3 Proportional Hazard Model

Cox (1972) proposed a proportional hazard model to control for the heterogeneity across individuals (and over time) by letting

$$\mu_{it} = \mu(t)g(\boldsymbol{x}_i), \tag{8.1.20}$$

where $\mu(t)$ is the so-called *baseline hazard* function that gives the shape of the hazard function, and $g(\cdot)$ is a known function of observable exogenous variables $\boldsymbol{x}_i$ that shifts the level of the hazard function. A common formulation for $g(\boldsymbol{x}_i)$ is to let

$$g(\boldsymbol{x}_i) = \exp\,(\boldsymbol{x}_i'\boldsymbol{\beta}). \tag{8.1.21}$$

Then

$$\frac{\partial \mu_{it}}{\partial x_{ki}} = \beta_k \cdot \exp\,(\boldsymbol{x}_i'\boldsymbol{\beta})\mu(t) = \beta_k \cdot \mu_{it}, \tag{8.1.22}$$

has a constant proportional effect on the instantaneous conditional probability of leaving the state. However, $\mu_{it} = \mu(t)cc^{-1}g(\boldsymbol{x}_i)$ for any $c > 0$. We need to define a reference individual. A common approach is to choose a particular value of $x = x^*$ such that $g(\boldsymbol{x}^*) = 1$.

One can simultaneously estimate the *baseline hazard* function $\mu(t)$ and $\boldsymbol{\beta}$ by maximizing the logarithm of the likelihood function (8.1.15). However, Cox's proportional hazard model allows the separation of the estimation of $g(\boldsymbol{x}_i)$ from the estimation of the baseline hazard $\mu(t)$. Suppose the data are only right-censored. Let $t_1 < t_2 < \ldots < t_j < \ldots < t_n$ denote the observed ordered discrete exit times of the complete spell (it is referred as *failure* time when the date of change is interpreted as a breakdown or a failure) for $i = 1, \ldots, n$, in a sample consisting of N individuals, $N \geq n$, and let $t_i^*, i = n+1, \ldots, N$ be the censored time for those with censored durations. Substituting (8.1.20) and (8.1.21) into the right-censored likelihood function yields Cox's proportional hazard model likelihood function,

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \exp\,(\boldsymbol{x}_i'\boldsymbol{\beta})\mu(t_i) \cdot \exp\left[-\exp\,(\boldsymbol{x}_i'\boldsymbol{\beta}) \int_0^{t_i} \mu(s)ds\right] \\
&\quad \cdot \prod_{i=n+1}^{N} \exp\left[-\exp\,(\boldsymbol{x}_i'\boldsymbol{\beta}) \int_0^{t_i^*} \mu(s)ds\right] \\
&= \prod_{i=1}^{n} \exp\,(\boldsymbol{x}_i'\boldsymbol{\beta})\mu(t_i) \cdot \exp\left\{-\int_0^{\infty}\left[\sum_{h \in R(t_\ell)} \exp\,(\boldsymbol{x}_h'\boldsymbol{\beta})\right]\mu(t)dt\right\} \\
&= L_1 \cdot L_2,
\end{aligned}
\tag{8.1.23}
$$

where

$$R(t_\ell) = \{i \mid S_i(t) \geq t_\ell\}$$

denotes the set of individuals who are at risk of exiting just before the ℓth ordered exiting, L_1 is the likelihood of an individual exits among all those who could exit at t_i,

$$L_1 = \prod_{i=1}^{n} \frac{\exp\,(\boldsymbol{x}_i'\boldsymbol{\beta})}{\sum_{h \in R(t_i)} \exp\,(\boldsymbol{x}_h'\boldsymbol{\beta})}\quad, \tag{8.1.24}$$

and

$$
\begin{aligned}
L_2 &= \sum_{i=1}^{n} \left[\sum_{h \in R(t_i)} \exp\,(\boldsymbol{x}_h'\boldsymbol{\beta})\mu(t_i)\right] \\
&\quad \cdot \exp\left\{-\int_0^{\infty}\left[\sum_{h \in R(t_j)} \exp\,(\boldsymbol{x}_h'\boldsymbol{\beta})\right]\mu(s)ds\right\}.
\end{aligned}
\tag{8.1.25}
$$

Since L_1 does not depend on $\mu(t)$, Cox (1975) suggested maximizing the partial likelihood function L_1 to obtain the PMLE estimator $\hat{\boldsymbol{\beta}}_p$. Tsiatis (1981) showed that the partial MLE of $\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}_p$ is consistent and asymptotically normally distributed with the asymptotic covariance

$$\text{Cov}(\hat{\boldsymbol{\beta}}_p) = -\left[E\frac{\partial^2 \log L_1}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]^{-1}. \tag{8.1.26}$$

Once $\hat{\boldsymbol{\beta}}_p$ is obtained, one can estimate $\mu(t)$ parametrically by substituting $\hat{\boldsymbol{\beta}}_p$ for $\boldsymbol{\beta}$ in the likelihood function (8.1.15) or semiparametrically through the relation

$$-\log S_i(t_i) = \boldsymbol{x}_i'\boldsymbol{\beta} + \int_0^{t_i} \mu(s)ds + \epsilon_i \tag{8.1.27}$$

For detail, see Han and Hausman (1990) or Florens, Fougére, and Mouchart (2008).

The generalization of duration models for cross-sectional data analysis to panel data analysis raises several complicated issues:

1. Where the unobserved individual- and time-specific effects fit into a model.
2. The hazard function may depend on time-varying covariates.

When the conditional covariates are time-invariant, $\boldsymbol{x}_{it} = \boldsymbol{x}_i$, and analogous to the linear model specification, the heterogeneous hazard function can be specified as

$$
\begin{aligned}
\mu_{it} &= \mu(t) \exp\,(\boldsymbol{x}_i'\boldsymbol{\beta} + \alpha_i) \\
&= \mu(t)\, \alpha_i^* \cdot \exp\,(\boldsymbol{x}_i'\boldsymbol{\beta}),
\end{aligned}
\tag{8.1.28}
$$

where $\alpha_i^* = \exp(\alpha_i)$ is the unobserved heterogeneity term for the ith individual, normalized with $E(\alpha_i^*) = 1$. When α_i^* are independent of $\boldsymbol{x}_i$, common assumptions for the heterogeneity in a mixture model are Gamma, inverse Gamma, or log-normal (e.g., Lancaster 1990). Once the heterogeneity distribution is specified, one can integrate out α_i^* to derive the marginal survival function or expected duration conditional on $\boldsymbol{x}_i$.

When $\alpha_i^* = \exp(\alpha_i)$ are treated as fixed constants, the probability density of exiting at time t_i is

$$f(t_i) = \alpha_i^* \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\mu(t_i) \cdot \exp\left[-\alpha_i^* \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \int_0^{t_i} \mu(s)ds\right]. \tag{8.1.29}$$

The survival function at t_i is

$$S(t_i) = \exp(-\alpha_i^*) \exp\left[-\exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \int_0^{t_i} \mu(s)ds\right]. \tag{8.1.30}$$

Depending on the configuration of panel data, one can write down the likelihood function using the specification of (8.1.29) and (8.1.30).

When the conditional covariates are time-varying, the problem becomes much more complicated (e.g. Van den Berg 2001). A simple transformation of an observed duration t_i into a linear regression format does not exist. Consider a hazard function for the ith individual at time t is given by

$$\mu_{it} = \mu(t) \cdot \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}), \tag{8.1.31}$$

then, from (8.1.10) and (8.1.11), the probability density of y_{it} exits at time t takes the form

$$f_i(t) = \mu(t) \cdot \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) \cdot \exp\left[-\int_0^t \mu(s)\exp(-\boldsymbol{x}_{is}'\boldsymbol{\beta})ds\right] \tag{8.1.32}$$

The maximization of the likelihood function of the form (8.1.23) is not computationally feasible when $f_i(t)$ takes the form (8.1.32).

8.1.4 Accelerated Failure Time Model

As an alternative, an *accelerated failure time model* (AFT) is suggested. The AFT model suggests that instead of using the actual duration, t, the model rescales the duration as $\tilde{t}$, where $\tilde{t} = t\phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})$. Then, from $\exp(-\int_0^t \mu_{is}ds) = \exp(-\int_0^{\tilde{t}} \tilde{\mu}_{is}ds)$ and $d\tilde{t} = \phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})dt$, an alternative hazard function in terms of the rescaled duration $\tilde{t}$ is suggested:

$$\tilde{\mu}_{it} = \mu[t \cdot \phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})]\,\phi(\boldsymbol{x}_{it}'\boldsymbol{\beta}). \tag{8.1.33}$$

Then the survival function takes the form

$$S_i(t_i) = \exp\left[-\int_0^{-t\phi(\boldsymbol{x}_{it}'\boldsymbol{\beta})} \tilde{\mu}_{is}ds\right] \tag{8.1.34}$$

Using the similar manipulation as (8.1.17), the ith individual exits at the time t_i for the AFT model can be approximated as a log linear function of the form

$$\log t_i = c + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \xi_{it}, \tag{8.1.35}$$

when $\phi(\boldsymbol{x}_{it}'\boldsymbol{\beta}) = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})$, where the density of ξ_{it} is extreme value distributed. The parameter $\boldsymbol{\beta}$ can then be obtained by the least squares method using those individuals with complete spells.

8.1.5 Discrete Time Markov Chain Modeling

Alternatively, one may model duration of staying in a state with time-varying covariates $\boldsymbol{x}_{it}$ by considering a discrete-time Markov Chain model. At each time t, there are two possible states, state 1 and state 0. Let the transition probability at state l in period t when one is in the state l at time $t-1$, $P_{it,ll}$ be denoted as

$$P_{it,ll} = F(\boldsymbol{x}_{it}'\boldsymbol{\beta}_l),\ l = 0 \text{ or } 1. \tag{8.1.36}$$

Then the transition probability from state 1 to state 0 or from state 0 to state 1 at t is given by

$$\begin{aligned} P_{it,10} &= 1 - F(\boldsymbol{x}_{it}'\boldsymbol{\beta}_1), \quad \text{or} \\ P_{it,01} &= 1 - F(\boldsymbol{x}_{it}'\boldsymbol{\beta}_0). \end{aligned} \tag{8.1.37}$$

For an individual starting at state 1, the probability of this individual exiting state 1 at time T is

$$\left[\prod_{t=1}^{T-1} F(\boldsymbol{x}_{it}'\boldsymbol{\beta}_1)\right](1 - F(\boldsymbol{x}_{it}'\boldsymbol{\beta}_1)). \tag{8.1.38}$$

The expected duration in state 1 is

$$\begin{aligned} D_i =\ & 1 \cdot F(\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1) \cdot (1 - F(\boldsymbol{x}_{i2}'\boldsymbol{\beta}_1)) \\ & + 2 \cdot F(\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1) \cdot F(\boldsymbol{x}_{i2}'\boldsymbol{\beta}_1) \cdot (1 - F(\boldsymbol{x}_{i3}'\boldsymbol{\beta}_1)) \\ & + \cdots, \end{aligned} \tag{8.1.39}$$

When $\boldsymbol{x}_{it} = \boldsymbol{x}_i$, the expected duration at state 1 is

$$D_i = (1 - F(\boldsymbol{x}_i'\boldsymbol{\beta}_0)) \cdot \frac{F(\boldsymbol{x}_i'\boldsymbol{\beta}_1)}{(1 - F(\boldsymbol{x}_i'\boldsymbol{\beta}_1))^2}. \tag{8.1.40}$$

The parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ can be obtained by splitting the observed y_{it} series into two sequences, $y_{it}(l)$, $l = 0$ or 1. Let $y_{it}(l)$ be the two sequences of random variables that take the value 1 if an individual at period $t-1$ is in state l and stays at state l in t, and 0 otherwise, $l = 0$ or 1. Suppose

$$y_{it}(1) = \begin{cases} 1, & \text{if} \quad d_{it}^*(1) > 0, \\ 0, & \text{if} \quad d_{it}^*(1) < 0. \end{cases} \tag{8.1.41}$$

and

$$y_{it}(0) = \begin{cases} 1, & \text{if} \quad d_{it}^*(0) > 0, \\ 0, & \text{if} \quad d_{it}^*(0) < 0. \end{cases} \tag{8.1.42}$$

where

$$d_{it}^*(1) = \boldsymbol{x}_{it}'\boldsymbol{\beta}_1 + \epsilon_{it}^1, \tag{8.1.43}$$

$$d_{it}^*(0) = \boldsymbol{x}_{it}'\boldsymbol{\beta}_0 + \epsilon_{it}^0. \tag{8.1.44}$$

Then the transition probability of starting at state 1 and staying at state 1 is

$$P_{it,11} = \text{Prob}(y_{it}(1) = 1) = \int_{-\boldsymbol{x}_{it}'\beta}^{\infty} f(\epsilon_{it}^1) d\epsilon_{it}^1, \tag{8.1.45}$$

and similarly for the derivation of the probability of starting at state 0 and ending at state 0 in t. Thus, methods discussed in Chapter 6 can be used to obtain consistent estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$.

8.1.6 An Example – Rating a Firm in Terms of Default Probability

Duration models can also be applied to predict exit of an event in future based on current state variables. For instance, Duan, Sun, and Wang (2012) define the average exit intensity for the period $[t, t+\tau]$ as:

$$\mu_{it}(\tau) = -\frac{\ln[1 - F_{it}(\tau)]}{\tau}, \tag{8.1.46}$$

where $F_{it}(\tau)$ is the conditional distribution function of the exit at time $t+\tau$ evaluated at time t for the ith individual. (When $\tau = 0, \mu_{it}(0)$ is the hazard function defined at (8.1.8) or (8.1.12), which they call the forward intensity rate.) Then, as in (8.1.12), the average exit intensity for the period $[t, t+\tau]$ becomes

$$\begin{aligned} \mu_{it}(\tau) &= -\frac{\ln[1 - F_{it}(\tau)]}{\tau} \\ &= -\frac{\ln[\exp\,(-\int_t^{t+\tau} \mu_{is} ds)]}{\tau}. \end{aligned} \tag{8.1.47}$$

Hence, the survival probability over $[t, t+\tau]$ becomes $\exp\,[-\mu_{it}(\tau)\tau]$. Assume $\mu_{it}(\tau)$ is differentiable with τ, it follows from (8.1.11) that the instantaneous forward exit intensity at time $t+\tau$ is:

$$\psi_{it}(\tau) = \frac{F'_{it}(\tau)}{1 - F_{it}(\tau)} = \mu_{it}(\tau) + \mu'_{it}(\tau)\tau. \tag{8.1.48}$$

Then $\mu_{it}(\tau) \cdot \tau = \int_0^\tau \psi_{it}(s) ds$. The forward exit probability at time t for the period $[t+\tau, t+\tau+1]$ is then equal to

$$\int_0^1 e^{-\mu_{it}(\tau+s)s} \psi_{it}(\tau + s) ds. \tag{8.1.49}$$

A firm can exit either due to bankruptcy or other reasons such as mergers or acquisitions. In other words, $\psi_{it}(s)$ is a combined exit intensity of default and other exit. Let $\phi_{it}(\tau)$ denote the forward default intensity. Then the default probability over $[t+\tau, t+\tau+1]$ at time t is

$$\int_0^1 e^{-\mu_{it}(\tau+s)s} \phi_{it}(\tau + s) ds. \tag{8.1.50}$$

The actual exit is recorded at a discrete time, say in a month or a year. Discretizing the continuous version by Δt for empirical implementation yields the forward (combined) exit probability and forward default probability at time t for the period $[t+\tau, t+\tau+1]$ as

$$e^{-\mu_{it}(\tau)\tau\Delta t}\left[1 - e^{-\psi_{it}(\tau)\Delta t}\right], \tag{8.1.51}$$

and

$$e^{-\mu_{it}(\tau)\tau\Delta t}\left[1 - e^{-\phi_{it}(\tau)\Delta t}\right], \tag{8.1.52}$$

respectively with spot (instantaneous) exit intensity at time t for the period $[t, t+\tau]$ being

$$\mu_{it}(\tau) = \frac{1}{\tau}[\mu_{it}(\tau-1)(\tau-1) + \psi_{it}(\tau-1)]. \tag{8.1.53}$$

Default is only one of the possibilities for a firm to exit; $\phi_{it}(\tau)$ must be no greater than $\psi_{it}(\tau)$. Suppose $\psi_{it}(\tau)$ and $\phi_{it}(\tau)$ depend on a set of macroeconomic factors and firm-specific attributes, $\boldsymbol{x}_{it}$, a convenient specification to ensure $\phi_{it}(\tau) \leq \psi_{it}(\tau)$ is to let

$$\phi_{it}(\tau) = \exp\,(\boldsymbol{x}'_{it}\boldsymbol{\gamma}(\tau)), \tag{8.1.54}$$

and

$$\psi_{it}(\tau) = \phi_{it}(\tau) + \exp\,(\boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau)). \tag{8.1.55}$$

Duan, Sun, and Wang (2012) use the monthly data $\left(\Delta t = \frac{1}{12}\right)$ of 12,268 publicly traded firms for the period 1991 to 2011 to predict the multiperiod ahead default probabilities for the horizon τ from 0 to 35 months. Table 8.1 provides the summaries of the number of active companies, defaults/bankruptcies, and other exits each year. The overall default rate ranges between 0.37% and 3.26% of the firms in each sample year. Other forms of exit are significantly higher, ranging from 4.91% to 13.61%. The macro and firm-specific attributes for $\phi_{it}(\tau)$ and $\psi_{it}(\tau)$ include: (trailing) one-year S&P 500 return (SP500); 3-month US Treasury bill rate; firm's distance to default (DTD), which is a volatility-adjusted leverage measure based on Merton (1974) (for detail, see Duan and Wang 2012); ratio of cash and short-term investments to the total assets (CASH/TA); ratio of net income to total assets (NI/TA); logarithm of the ratio of a firm's market equity value to the average market equity value of the S&P 500 firms (SIZE); market-to-book asset ratio (M/B); and

Table 8.1. *Total number of active firms, defaults/bankruptcies, and other exits for each year over the sample period 1991–2011*

Year	Active firms	Defaults/bankruptcies	(%)	Other exit	(%)
1991	4,012	32	0.80	257	6.41
1992	4,009	28	0.70	325	8.11
1993	4,195	25	0.6	206	4.91
1994	4,433	24	0.54	273	6.16
1995	5,069	19	0.37	393	7.75
1996	5,462	20	0.37	463	8.48
1997	5,649	44	0.78	560	9.91
1998	5,703	64	1.12	753	13.20
1999	5,422	77	1.42	738	13.61
2000	5,082	104	2.05	616	12.12
2001	4,902	160	3.26	577	11.77
2002	4,666	81	1.74	397	8.51
2003	4,330	61	1.41	368	8.50
2004	4,070	25	0.61	302	7.42
2005	3,915	24	0.61	291	7.43
2006	3,848	15	0.39	279	7.25
2007	3,767	19	0.50	352	9.34
2008	3,676	59	1.61	285	7.75
2009	3,586	67	1.87	244	6.80
2010	3,396	25	0.74	242	7.13
2011	3,224	21	0.65	226	7.01

The number of active firms is computed by averaging over the number of active firms across all months of the year.
Source: Duan, Sun, and Wang (2012, Table 1).

one-year idiosyncratic volatility, calculated by regressing individual monthly stock return on the value-weighted CRSP monthly return over the preceding 12 months (SIGMA). Both level and trend measures for DTD, CASH/TA, NI/TA, and SIZE are employed in the empirical analysis. To account for the impact of the massive US governmental interventions during the 2008–2009 financial crisis, Duan, Sun, and Wang (2012) also included a common bailout term, $\lambda(\tau)\exp\{-\delta(\tau)(t-t_B)\}\cdot 1[(t-t_B)>0]$ for $\tau = 0, 1, \ldots, 11$, to the forward default intensity function where t_B denotes the end of August 2008 and 1(A) is the indicator function that equals 1 if event A occurs and zero otherwise. Specifically

$$\phi_{it}(\tau) = \exp\{\lambda(\tau)\exp[-\delta(\tau)(t-t_B)]1((t-t_B)>0) + \boldsymbol{x}_{it}'\boldsymbol{\gamma}(\tau)\}, \qquad (8.1.56)$$

for $\tau = 0, 1, \ldots, 11$.

Assuming the firms are cross-sectional independent conditional on $\boldsymbol{x}_{it}$, and ignoring the time dependence, Duan, Sun, and Wang (2012) obtain the estimated $\phi_{it}(\tau)$ and $\psi_{it}(\tau)$ by maximizing the pseudo-likelihood function,

$$\prod_{i=1}^{N}\prod_{t=0}^{T-1}\mathcal{L}_{it}(\tau). \qquad (8.1.57)$$

Let t_{oi}, τ_{oi}, and τ_{ci} denote the first month appeared in the sample, default time, and combined exit time for firm i, respectively; $\mathcal{L}_{it}(\tau)$ is defined as,

$$\begin{aligned}\mathcal{L}_{it}(\tau) = {} & 1\{t_{oi} \le t, \tau_{ci} > t+\tau\}P_t(\tau_{ci} > t+\tau) \\ & + 1\{t_{oi} < t, \tau_{ci} = \tau_{oi} \le t+\tau\}P_t(\tau_{ci}; \tau_{ci} = \tau_{oi} \le t+\tau) \\ & + 1\{t_{oi} < t, \tau_{ci} \ne \tau_{oi}, \tau_{ci} \le t+\tau\}P_t(\tau_{ci}; \tau_{ci} \ne \tau_{oi} \\ & \text{and } \tau_{ci} \le t+\tau) + 1\{t_{oi} > t\} + 1\{\tau_{ci} \le t\},\end{aligned} \qquad (8.1.58)$$

$$P_t(\tau_{ci} > t+\tau) = \exp\left[-\sum_{s=0}^{\tau-1}\psi_{it}(s)\Delta t\right], \qquad (8.1.59)$$

$$\begin{aligned}& P_t(\tau_{ci} \mid \tau_{ci} = \tau_{oi} \le t+\tau) \\ & = \begin{cases} 1-\exp[-\phi_{it}(0)\Delta t], \text{ if } \tau_{oi} = t+1, \\ \exp\left[-\sum_{s=0}^{\tau_{ci}-t-2}\psi_{it}(s)\Delta t\right] \\ \quad \cdot[1-\exp[-\phi_{it}(\tau_{c_i}-t-1)\Delta t]\}, \text{ if } t+1 < \tau_{ci} \le t+\tau, \end{cases}\end{aligned} \qquad (8.1.60)$$

$$\begin{aligned}& P_t(\tau_{ci} \mid \tau_{ci} \ne \tau_{oi} \le t+\tau) \\ & = \begin{cases} \exp[-\phi_{it}(0)\Delta t] - \exp[-\psi_{it}(0)\Delta t], \text{ if } \tau_{oi} = t+1, \\ \exp\left[-\sum_{s=0}^{\tau_{ci}-t-2}\psi_{it}(s)\Delta t\right]\cdot\{\exp[-\phi_{it}(\tau_{ci}-t-1)\Delta t] \\ \qquad - \exp[-\psi_{it}(\tau_{ci}-t-1)]\}, \text{ if } t+1 < \tau_{ci} \le t+\tau, \end{cases}\end{aligned} \qquad (8.1.61)$$

with $\Delta t = \frac{1}{12}$, and $\psi_{it}(s)$ and $\phi_{it}(s)$ take the form of (8.1.54) and (8.1.55), respectively. The first term on the right-hand side of $\mathcal{L}_{it}(\tau)$ is the probability of surviving both forms of exit. The second term is the probability that the firm defaults at a particular time point. The third term is the probability that the firm exits due to other reasons at a particular time point. If the firm does not appear in the sample in month t, it is set equal to 1, which is transformed to 0 in the log-pseudo-likelihood function. The forward intensity approach allows an investigator to predict the forward exiting time of interest τ, $\phi_{it}(\tau)$, and $\psi_{it}(\tau)$

as functions of conditional variables available at time t without the need to predict future conditional variables.

Figure 8.1 plots each of the estimated $\boldsymbol{\gamma}(\tau)$ and $\boldsymbol{\beta}(\tau)$ and its 90% confidence interval, with τ ranging from 0 to 35 months. They show that some firm-specific attributes influence

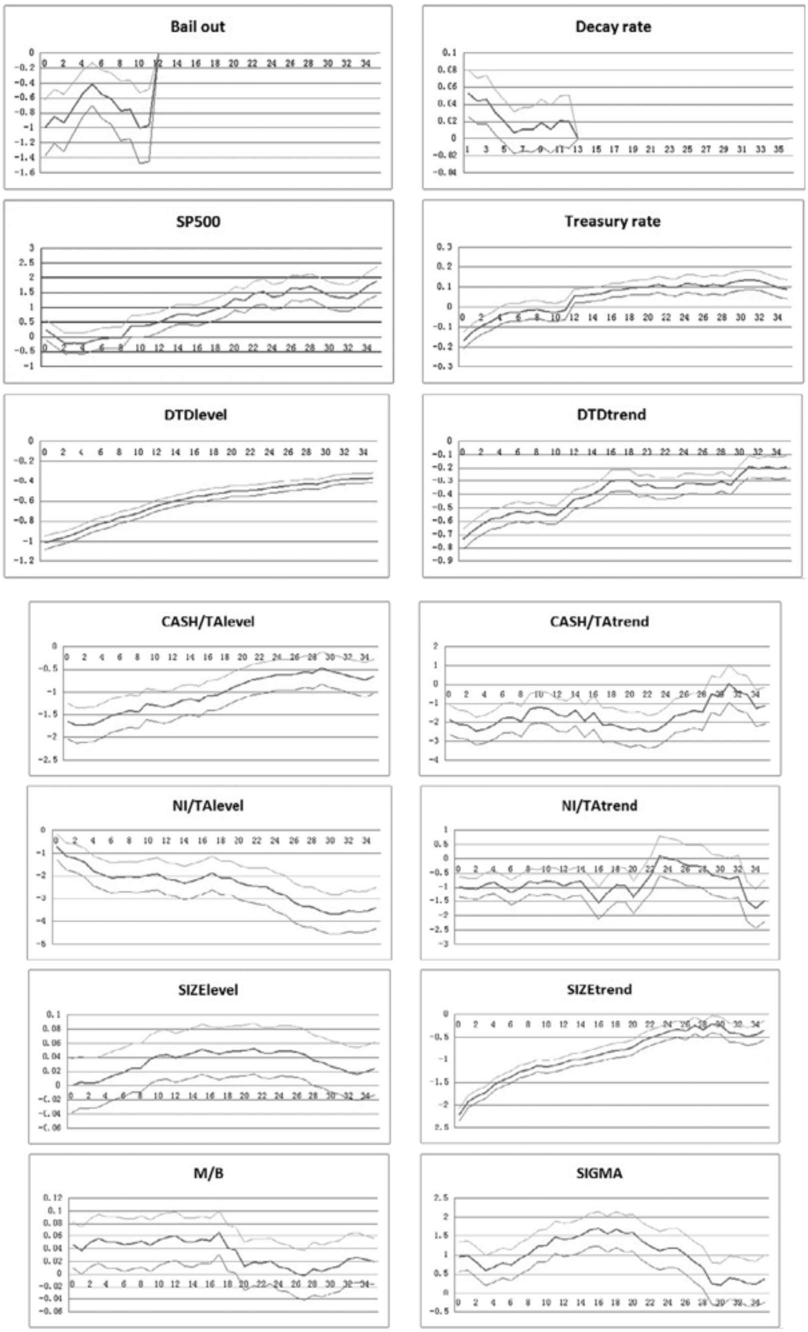

Figure 8.1. Parameter estimates for the forward default intensity function. The solid line is for the parameter estimates and the dotted lines depict the 90% confidence interval. *Source:* Duan, Sun, and Wang (2012, Fig. 1).

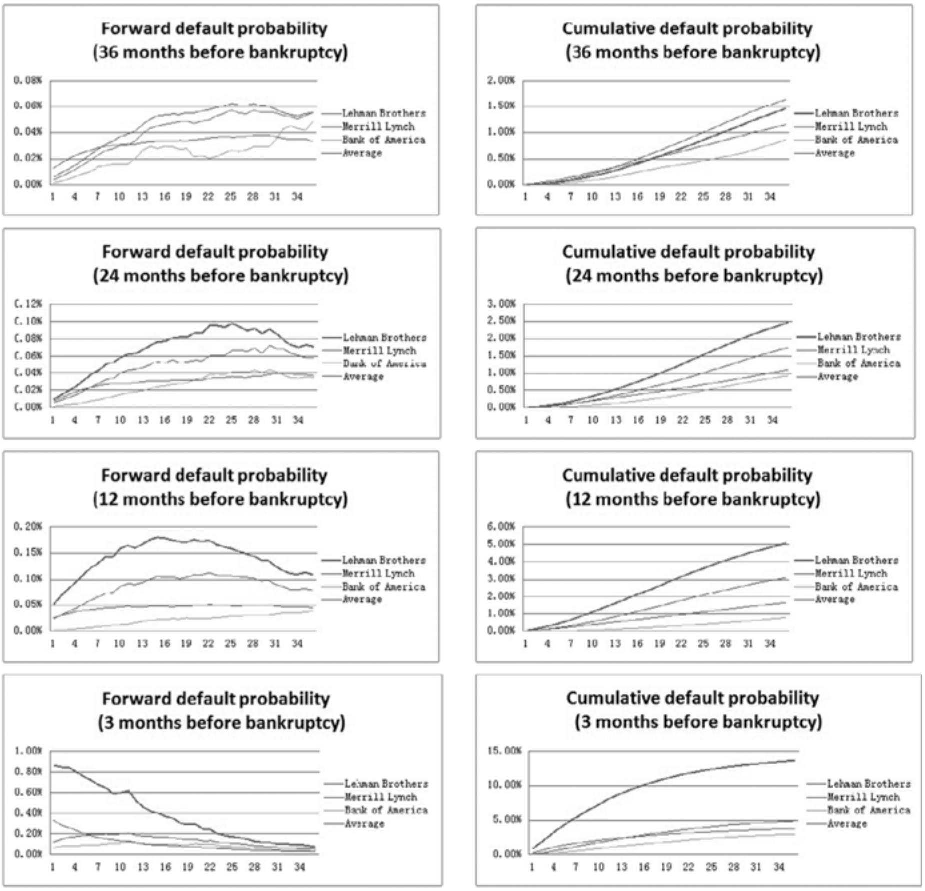

Figure 8.2. Lehman Brothers' term structure of forward and cumulative default probabilities relative to Merrill Lynch, Bank of America as well as the average values of financial sector at 36 months, 24 months, 12 months, and 3 months before Lehman Brothers' bankruptcy filing date (September 15, 2008).
Source: Duan, Sun, and Wang (2012, Fig. 4).

the forward intensity both in terms of level and trend. Figure 8.2 plots the estimated term structure of predicted default probabilities of Lehman Brothers, Merrill Lynch, Bank of America, and the averages of the US financial sector at several time points prior to Lehman Brothers' bankruptcy filing on September 15, 2008. The term structures for Lehman Brothers in June 2008, three months before its bankruptcy filing, show that the company's short-term credit risk reached its historical high. The peak of the forward default probability is one month. The one-year cumulative default probability increased sharply to 8.5%, which is about 35 times of the value three years earlier. This case study appears to suggest that the forward intensity model is quite informative for short prediction horizons.

8.2 COUNT DATA MODEL

8.2.1 Poisson Model

The count data model may be considered as the dual of the duration model. The duration model considers the probability that an event stays for a certain time period before another event occurs. The count data models consider the probability that a certain number of an event would occur during a fixed period of time. The probability of exit at t is derived from the hazard function. The probability of the nonnegative integer count number y can be derived by supposing an instant arrival rate at a given interval is μ. Breaking up a unit interval into n subintervals Δt, then probability of no arrival is $(1-\mu\Delta t)^{\frac{1}{\Delta t}}$, and the probability of one arrival is $\mu(1-\mu\Delta t)^{\frac{1}{\Delta t}}$. Let $\Delta t \to 0$, $(1-\mu\Delta t)^{\frac{1}{\Delta t}} = (1-\frac{\mu}{n})^n \to e^{-\mu}$ as $n \to \infty$. Then in a given interval, no arrival is synonymous to no death in this interval in duration models,

$$P(y=0) = e^{-\mu}. \tag{8.2.1}$$

One arrival is synonymous to death at the end of the interval in duration models,

$$P(y=1) = \mu e^{-\mu}. \tag{8.2.2}$$

The probability of two arrivals can be derived by letting z_1 and z_2 denote the arrival time interval for the first and second events, respectively, then $0 < z_1 < 1$, $0 < z_2 < 1 - z_1$, and the interval for no arrival is $1 - z_1 - z_2$. Then

$$\begin{aligned} P(y=2) &= \int_0^1 \mu e^{-\mu z_1} \left[\int_0^{1-z_1} \mu e^{-\mu z_2} \cdot e^{-\mu(1-z_1-z_2)} dz_2 \right] dz_1 \\ &= \mu^2 e^{-\mu} \int_0^1 (1-z_1) dz_1 \\ &= \frac{\mu^2 e^{-\mu}}{2} \end{aligned} \tag{8.2.3}$$

Using the similar manipulation as (8.2.3), the probability of the number of occurrences in a given interval is Poisson distributed,

$$P(y=j) = \frac{\mu^y e^{-\mu}}{y!}, \quad j = 0, 1, 2, \dots. \tag{8.2.4}$$

The Poisson distribution has mean

$$E(y) = \mu, \tag{8.2.5}$$

and variance

$$E(y-\mu)^2 = \mu. \tag{8.2.6}$$

8.2.2 Negative Binomial Model

The Poisson distribution has the mean equal to the variance. However, most observed count data have the dispersion greater than the mean and the frequency of the observed zero greater than that predicted by a Poisson model (excess zeros). To deal with the first issue, a random variable ν is introduced into the modeling of the instant arrival rate, $\lambda = \mu\nu$ for the Poisson model,

$$f(y \mid \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad j = 0, 1, 2, \dots, \tag{8.2.5}$$

where the random variable ν is nonnegative with density $g(\nu|\delta)$ and $E(\nu) = 1$. For instance, suppose $g(\nu|\delta)$ takes the one parameter Gamma distribution,

$$g(\nu \mid \delta) = \nu^{\delta-1} e^{-\nu\delta} \frac{\delta^\delta}{\Gamma(\delta)}. \tag{8.2.6}$$

Then $E(\nu) = 1$, $\text{Var}(\nu) = \frac{1}{\delta}$, and hence $E(\lambda|\mu) = \mu$. Under (8.2.5) and (8.2.6), the probability distribution conditioning on μ and δ becomes a negative binomial distribution,

$$
\begin{aligned}
h(y \mid \mu, \delta) &= \int_0^\infty f(y \mid \mu, \nu) g(\nu \mid \delta) d\nu \\
&= \frac{\Gamma(\delta + y)}{\Gamma(\delta)\Gamma(y+1)} \left(\frac{\delta}{\delta+\mu}\right)^{\delta} \left(\frac{\mu}{\delta+\mu}\right)^{y}, \qquad (8.2.7)
\end{aligned}
$$

where $\Gamma(\delta) = \int_0^\infty e^{-t} t^{(\frac{1}{\delta})-1} dt$. The negative binomial distribution has mean μ and variance $\mu(1 + \frac{\mu}{\delta})$. When $\frac{1}{\delta} = 0$, (8.2.7) becomes a Poisson model.

8.2.3 Excess Zeros

To deal with the excess zeros in the observed sample, the easiest remedy is to consider a two-model specification. A dichotomous model estimates the probability of $y = 0$ and $y \neq 0$. A second model estimates $P(y = j \mid y \neq 0)$. For instance, one may consider the probability of modeling $y = 0$ or not equal to zero in the form,

$$P(y = 0) = F_1(\cdot), \qquad (8.2.8)$$

$$P(y \neq 0) = 1 - F_1(\cdot), \qquad (8.2.9)$$

and

$$P(y = j \mid y \neq 0) = F_2(\cdot), \qquad (8.2.10)$$

then the unconditional probability of

$$P(y = j) = (1 - F_1(\cdot))(F_2(\cdot)), j = 1, 2, 3, \ldots, \qquad (8.2.11)$$

where $F_1(\cdot)$ could be a Probit or logit form, and $F_2(\cdot)$ could be a conditional Possion form $\frac{P(y=j)}{1-P(y=0)}$.

8.2.4 Estimation

Generalizing the count data model developed for the analysis of cross-sectional data to the panel data, we may assume the instant arrival rate for the ith individual at time t is μ_{it}. Then a Poisson model for $y_{it} = j$ takes the form

$$P(y_{it}) = \frac{e^{-\mu_{it}} (\mu_{it})^{y_{it}}}{y_{it}!}, \quad y_{it} = 0, 1, 2, \ldots. \qquad (8.2.12)$$

The Poisson model implies the distribution of y_{it} is independent over t,

$$\text{Prob } (y_{it} = r \mid y_{i,t-s} = \ell) = P(y_{it} = r), \qquad (8.2.13)$$

with

$$E(y_{it}) = \mu_{it}, \qquad (8.2.14)$$

and

$$\text{Var } (y_{it}) = \mu_{it}. \qquad (8.2.15)$$

Therefore, under the assumption that y_{it} is independently distributed across i, the log-likelihood function is given by

$$\log L = \sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it}\log(\mu_{it}) - \mu_{it} - \log(y_{it}!)]. \tag{8.2.16}$$

The intensity μ_{it} is often assumed to be a function of K strictly exogenous variables, $\boldsymbol{x}_{it}$, and individual-specific effects, α_i. Because μ_{it} has to be nonnegative, popular specifications to ensure nonnegative μ_{it} without the need to impose restrictions on the parameters are to let

$$\mu_{it} = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i) = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}) \cdot \exp(\alpha_i). \tag{8.2.17}$$

or to let

$$\mu_{it} = \eta_i \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta}), \quad \eta_i > 0, \tag{8.2.18}$$

here, $\eta_i = \exp(\alpha_i)$. The possible distribution for η_i should be subject to $E(\eta_i) = 1$.

When η_i is treated as random and independent of $\boldsymbol{x}_{it}$ with known density function $g(\eta)$, the marginal distribution of $(y_{i1}, \ldots, y_{iT})$ takes the form

$$f(y_{i1}, \ldots, y_{iT}) = \int \prod_{t=1}^{T}\left[\frac{(\mu_{it})^{y_{it}}\exp(-\mu_{it})}{y_{it}!}\right] g(\eta)d\eta. \tag{8.2.19}$$

The MLE of $\boldsymbol{\beta}$ is consistent and asymptotically normally distributed when either N or T or both tend to infinity. However, the computation can be tedious because of the need to take multiple integrations. For instance, suppose $g(\eta)$ has Gamma density $g(\eta) = \eta^{\nu-1}\exp(-\eta)/\Gamma(\nu)$ with $E(\eta) = 1$ and variance $\nu(\nu > 0)$. When $T = 1$ (cross-sectional data),

$$f(y_i) = \frac{[\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]^{y_i}\Gamma(y_i + \nu)}{y_i!\,\Gamma(\nu)}\left(\frac{1}{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}) + \nu}\right)^{y_i+\nu}, \; i = 1, \ldots, N, \tag{8.2.20}$$

has a *negative binomial* distribution. But if $T > 1$, (8.2.19) no longer has the closed form. One computationally simpler method to obtain a consistent estimator of $\boldsymbol{\beta}$ is to ignore the serial dependence of y_{it} due to the presence of time-persistent η_i by considering the marginal (or unconditional) distribution of y_{it}. For instance, if μ_{it} takes the form of (8.2.18) and η is Gamma distributed, then the unconditional distribution of y_{it} takes the form of (8.2.20). Conditional on η_i, $f(y_{it} \mid y_{is}) = f(y_{it})$ for $t \neq s$. Maximizing the pseudo joint log-likelihood function $\prod_{i=1}^{N}\prod_{t=1}^{T} f(y_{it})$ of the form (8.2.16) can yield a consistent estimator of unknown structural parameters when either N or T or both tend to infinity. The pseudo MLE can also be used as initial values for the iterative schemes to obtain the MLE.

There is an advantage of treating α_i (or η_i) fixed when μ_{it} takes the form (8.2.17) because, conditional on α_i (or η_i), the log-likelihood function is of the simple form (8.2.16). It can also allow correlations between α_i and $\boldsymbol{x}_{it}$ just like in the linear model. When η_i is treated as fixed and μ_{it} takes the form (8.2.18), the maximum likelihood estimator of $\boldsymbol{\beta}$ and η_i, where $\eta_i = \exp(\alpha_i)$, can be obtained by simultaneously solving the first-order conditions

$$\frac{\partial \log L}{\partial \eta_i} = \sum_{t=1}^{T}\left[\frac{y_{it}}{\eta_i} - \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta})\right] = 0, \quad i = 1, \ldots, N, \tag{8.2.21}$$

and

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it} - \eta_i \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta})] \boldsymbol{x}_{it} = \mathbf{0}. \tag{8.2.22}$$

Solving (8.2.21) yields the MLE of η_i conditional on $\boldsymbol{\beta}$ as

$$\hat{\eta}_i = \frac{\bar{y}_i}{\bar{\mu}_i}, \quad i = 1, \ldots, N. \tag{8.2.23}$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}, \bar{\mu}_i = T^{-1} \sum_{t=1}^{T} \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta})$. The MLE of $\boldsymbol{\beta}$ and η_i can be obtained by iterating between (8.2.22) and (8.2.23) until the solutions converge.

Substituting $\hat{\eta}_i$ for η_i in (8.2.21), the MLE of $\boldsymbol{\beta}$ is the solution to

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} - \frac{\bar{y}_i}{\hat{\bar{\mu}}_i} \exp(\boldsymbol{x}_{it}' \hat{\boldsymbol{\beta}}) \right] \boldsymbol{x}_{it} = \mathbf{0}, \tag{8.2.24}$$

where $\hat{\bar{\mu}}_i = \frac{1}{T} \sum_{t=1}^{T} \exp(\boldsymbol{x}_{it}' \hat{\boldsymbol{\beta}})$. Substituting (8.2.23) into (8.2.24), (8.2.24) is equivalent to finding $\hat{\boldsymbol{\beta}}$ that satisfies the moment condition,

$$\operatorname*{plim}_{N \longrightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} [u_{it} - \bar{u}_i] \boldsymbol{x}_{it} = \mathbf{0}, \tag{8.2.25}$$

where

$$u_{it} = y_{it} - \eta_i \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta}), \tag{8.2.26}$$

and $\bar{u}_i = \frac{1}{T} \sum_{t=1}^{T} u_{it}$.

The transformation of $y_{it} = \eta_i \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta}) + u_{it}$ into a model

$$u_{it} - \bar{u}_i = y_{it} - \frac{\bar{y}_i}{\bar{\mu}_i} \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta}), \tag{8.2.27}$$

is to remove the individual-specific effects η_i by taking the ratio of individual means $\bar{y}_i$ and $\bar{\mu}_i = \left(\frac{1}{T} \sum_{t=1}^{T} \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta}) \right)$. Thus, solving the first-order condition of the conditional MLE of the Poisson model, $\boldsymbol{\beta}$, may be viewed as a generalized covariance estimator. Just like the linear static model discussed in Chapter 2, solving

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} - \frac{\bar{y}_i}{\bar{\mu}_i} \exp(\boldsymbol{x}_{it}' \boldsymbol{\beta}) \right] \boldsymbol{x}_{it} = 0 \tag{8.2.28}$$

yields consistent and asymptotically normally distributed estimator $\boldsymbol{\beta}$ if $\boldsymbol{x}_{it}$ is strictly exogenous with respect to u_{it}. Blundell, Griffith, and Windmeijer (2002) call this estimator the *within group mean scaling estimator*.

8.2.5 Dynamic Models

If $\boldsymbol{x}_{it}$ contains the lag dependent variable $y_{i,t-1}$, then $E(\boldsymbol{x}_{it} u_{i,t-s}) \neq 0$ for $s \geq 1$, (8.2.25) will not converge to zero. Just like the dynamic linear regression model discussed in Chapter 3, the solution of (8.2.25) (generalized covariance estimator) is inconsistent. Analogous to Anderson and Hsiao's (1981, 1982) simple intrumental variable estimator for the dynamic linear panel data model, one may consider a quasi-difference equations to remove η_i from the model (8.2.17) or (8.2.18),

$$y_{it} - \theta_t y_{i,t-1} = u_{it} - \theta_t u_{i,t-1}, \quad t = 2, \ldots, T, \tag{8.2.29}$$

where

$$\theta_t = \frac{\eta_i \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})}{\eta_i \exp(\boldsymbol{x}_{i,t-1}\boldsymbol{\beta})} = \exp\big((\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})'\boldsymbol{\beta}\big). \tag{8.2.30}$$

If u_{it} is independently distributed over t, then solving

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it} - \hat{\theta}_t y_{i,t-1})\boldsymbol{x}_{i,t-2} = 0, \tag{8.2.31}$$

or

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{it} - \hat{\theta}_t y_{i,t-1})(\boldsymbol{x}_{i,t-2} - \boldsymbol{x}_{i,t-3}) = 0 \tag{8.2.32}$$

yields a consistent and asymptotically normally distributed estimator of $\boldsymbol{\beta}$ when either N or T or both tend to infinity.

When both time-invariant variables, z_i, and time-varying variables $\boldsymbol{x}_{it}$, appear in $\mu_{it} = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + z_i'\boldsymbol{\gamma} + \alpha_i)$, one can obtain consistent estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by solving the first-order condition of the pseudo likelihood function $\prod_{i=1}^{N}\prod_{t=1}^{T} f(y_{it}|\mu_{it})$. (Gourieroux, Monfort, and Trognon 1984). However, the solution of the form (8.2.28) is highly nonlinear and can be computationally complex. Numerical methods have to be used to obtain the consistent estimator. On the other hand, estimating $\boldsymbol{\beta}$ using the form of either (8.2.31) or (8.2.32) is much simpler. The solution of (8.2.31) or (8.2.32) where $\boldsymbol{x}_{i,t-j}$ is replaced by $\boldsymbol{x}_{it}$ can also be used as an initial estimator of $\boldsymbol{\beta}$ to obtain the more efficient conditional MLE (8.2.24) when $\boldsymbol{x}_{it}$ are strictly exogenous. θ_t (8.2.30) is in terms of the difference of the explanatory variables. The time-invariant variable is removed in (8.2.31) or (8.2.32). To obtain a consistent estimator of $\boldsymbol{\gamma}$, Honoré and Kesina (2017) suggest a two-step procedure. In the first step, estimate $\boldsymbol{\beta}$ using either (8.2.31) or (8.2.32). In the second step, substitute consistently estimated $\boldsymbol{\beta}$ for the moment condition that the mean of a Poisson model that $E(y_{it}|\boldsymbol{x}_{it}, z_i, \alpha_i) = \mu_{it} = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + z_i'\gamma + \alpha_i)$,

$$E\left[\exp(-\boldsymbol{x}_{it}'\hat{\beta})y_{it}|(z_i, \alpha_i)\right] = \exp(z_i'\boldsymbol{\gamma} + \alpha_i). \tag{8.2.33}$$

Under the assumption that z_i and α_i are independent, $\boldsymbol{\gamma}$ can be estimated by the nonlinear least square method,

$$\text{Min} \quad \sum_{i=1}^{N}\sum_{t=1}^{T}\big(\exp(-\boldsymbol{x}_{it}'\boldsymbol{\beta})y_{it} - \exp(z_i'\boldsymbol{\gamma})\big)^2. \tag{8.2.34}$$

It should be noted that inclusion of the lagged dependent variable in an exponential mean function may lead to rapidly exploding series. Crépon and Duguet (1997) suggest specifying

$$\mu_{it} = h(y_{i,t-1})\exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i). \tag{8.2.35}$$

Possible choice for $h(\cdot)$ could be

$$h(y_{i,t-1}) = \exp(\gamma(1 - d_{i,t-1})), \tag{8.2.36}$$

or

$$h(y_{i,t-1}) = \exp(\gamma_1 ln(y_{i,t-1} + cd_{i,t-1}) + \gamma_2 d_{i,t-1}), \tag{8.2.37}$$

where c is a pre-specified constant, $d_{it} = 1$ if $y_{it} = 0$ and 0 otherwise. In this case, $lny_{i,t-1}$ is included as a regressor for positive $y_{i,t-1}$ and 0 values of $y_{i,t-1}$ have a separate effect on current values of y_{it}. Alternatively, Blundell, Griffith, Windmeijer (2002) propose a linear feedback model of the form

$$\mu_{it} = \gamma y_{i,t-1} + \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i). \tag{8.2.38}$$

Unfortunately, neither specification leads to easily derivable MLE (because of the complications in formulating the distribution for the initial values) or moment conditions (because of the nonlinear nature of the moment functions).

8.3 NONPARAMETRIC MODELS

We have been assuming that the conditional expection of y_{it} conditional on some covariates and individual- and/or time-specific effects is parametrically specified. The panel nonparametric model assumes that conditional on the unobserved specific effects, the functional form for the conditional mean, conditional on some observable driving covariates, can take any form.

Suppose

$$y_{it} = m(\boldsymbol{x}_{it}) + v_{it}, \quad \begin{aligned} i &= 1, \ldots, N, \\ t &= 1, \ldots, T, \end{aligned} \tag{8.3.1}$$

$$v_{it} = \alpha_i + u_{it}, \tag{8.3.2}$$

where $m(\boldsymbol{x}_{it})$ can take any form of $\boldsymbol{x}_{it}$, $\boldsymbol{x}_{it}$ are the $K \times 1$ causal variables that drive y_{it}. We assume $\boldsymbol{x}_{it}$ are strictly exogenous variables with respect to u_{it}, $E(u_{it} \mid \boldsymbol{x}_{is}) = 0$ for all t and s.

If α_i is treated as random and uncorrelated with $\boldsymbol{x}_{it}$, then $m(\boldsymbol{x}^*)$ can be estimated by the kernel method or the local linear least squares method in which

$$m(\boldsymbol{x}_{it}) = m(\boldsymbol{x}^*) + (\boldsymbol{x}_{it} - \boldsymbol{x}^*)'\boldsymbol{\beta}(\boldsymbol{x}^*) \tag{8.3.3}$$

for $\boldsymbol{x}_{it}$ close to $\boldsymbol{x}^*$, where the "closeness" is defined in terms of some kernel function, $\sigma_N^{-K} K\left(\frac{x_{it} - x^*}{\sigma_N}\right)$, with $K(\boldsymbol{v}) \geq 0, K(\boldsymbol{v}) \longrightarrow 0$ as $\boldsymbol{v} \longrightarrow \pm\infty$, and σ_N is a bandwidth parameter. Substituting (8.3.3) into (8.3.1), one can estimate $m(\boldsymbol{x}^*)$ and $\boldsymbol{\beta}(\boldsymbol{x}^*)$ by the local linear least squares method (Li and Racine 2007). However, the least squares method ignores the error-components structure of v_{it} is inefficient. Under the standard assumption of the random-effects error-components model (e.g., (2.3.3)–(2.3.4)).

$$\boldsymbol{y}_i = \boldsymbol{m}(\boldsymbol{x}_i) + \boldsymbol{v}_i, \tag{8.3.4}$$

$$E\boldsymbol{v}_i\boldsymbol{v}_i = \sigma_u^2 I_T + \sigma_\alpha^2 \boldsymbol{e}_T\boldsymbol{e}_T' = V, \tag{8.3.5}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{m}_i(\boldsymbol{x}_i) = (m(\boldsymbol{x}_{i1}), \ldots, m(\boldsymbol{x}_{iT}))'$, $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iT})'$, $\boldsymbol{e}_T = (1, \ldots, 1)'$, $\sigma_u^2 = E(u_{it}^2)$, $\sigma_\alpha^2 = E(\alpha_i^2)$. Estimating $\boldsymbol{m}(\boldsymbol{x}_i)$ while ignoring the serial correlation of v_{it} is inefficient. An efficient estimator is the one that minimizes

$$S = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{m}(\boldsymbol{x}_i))'V^{-1}(\boldsymbol{y}_i - \boldsymbol{m}(\boldsymbol{x}_i)). \tag{8.3.6}$$

The first-order condition of minimizing (8.3.6) with respect to $\boldsymbol{m}(\boldsymbol{x}_i)$ is

$$\boldsymbol{e}_T' V^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{m}}(\boldsymbol{x}_i)) = \boldsymbol{0}. \tag{8.3.7}$$

However, V is unknown. Ruckstuhl et al. (2000) and Martins-Filho and Yao (2009) propose a two-step feasible estimator:

Step 1: Estimate $m(\boldsymbol{x}_{it})$ using local-polynomial least squares and obtain the residuals $\hat{v}_{it} = (y_{it} - \hat{m}(\boldsymbol{x}_{it}))$. Construct $\hat{V}$ similar to the method described for applying the feasible generalized least squares method (Chapter 2, Equations 2.3.17 and 2.3.18).

Step 2: Run the local-polynomial least squares regression of $[y_{it} - (1 - \psi^{1/2})\bar{y}_i]$ on $(1 - (1 - \psi^{1/2})\, \bar{m}(\boldsymbol{x}_i))$, where $\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$, $\bar{y}_i = \frac{1}{T}\sum_{t=1}^T y_{it}$, $\bar{m}(\boldsymbol{x}_i) = \frac{1}{T}\sum_{t=1}^T \hat{m}(x_{it})$.

Alternatively, Ullah and Roy (1998), Lin and Carroll (2000), and Henderson and Ullah (2005) propose using the local-polynomial weighted least squares method by taking a Taylor series expansion of $m(\boldsymbol{x}_{it})$. Ma et al. (2015) consider a B-spline regression approach. For a survey of nonparametric random effects error component model, see Parmeter and Racine (2018).

When α_i is treated as a fixed constant, the kernel approach is not a convenient method to estimate $m(\boldsymbol{x}_{it})$ because linear difference of y_{it} has to be used to eliminate α_i. (e.g., Li and Stengos 1996). A convenient approach is to put $m(\boldsymbol{x}_{it})$ into the following general index format,

$$m(\boldsymbol{x}_{it}) = \nu_0(\boldsymbol{x}_{it}, \boldsymbol{\theta}) + \sum_{j=1}^{m} h_{j0}(\nu_j(\boldsymbol{x}_{it}, \boldsymbol{\theta})), \tag{8.3.8}$$

where $\nu_j(\boldsymbol{x}_{it}, \boldsymbol{\theta})$ for $j = 0, 1, \ldots, m$ are known functions of $\boldsymbol{x}_{it}$ and $h_{j0}(\cdot)$ for $j = 1, \ldots, m$ are unknown functions.

To uniquely identify the parameters of interest of the index model (8.3.8), the following normalization conditions are imposed:

(i) $h_{j0}(0) = 0$ for $j = 1, \ldots, m$;
(ii) the scaling restriction, say $\boldsymbol{\theta}'\boldsymbol{\theta} = 1$ or the first element of $\boldsymbol{\theta}^0$ being normalized to 1 if it is known different from zero;
(iii) the exclusion restriction when $\nu_j(\boldsymbol{x}, \boldsymbol{\theta})$ and $\nu_s(\boldsymbol{x}, \boldsymbol{\theta})$ are homogeneous of degree 1 in the regressors for some $s \neq j$.

If (i) does not hold, it is not possible to distinguish $(h_{j0}(\cdot), \alpha_i)$ from $(h_{j0}(\cdot) - \mu, \alpha_i + \mu)$ for any constant μ and for any j in (8.3.8). If (ii) does not hold, it is not possible to distinguish $(\boldsymbol{\theta}, h_0(\cdot))$ from $(c\boldsymbol{\theta}^0, \tilde{h}_0(\cdot) = h_0(\cdot/c))$ for any nonzero constant c. If (iii) does not hold, say $(h_{10}(\cdot), h_{20}(\cdot))$ contains a common element $\boldsymbol{x}_{3it}$, then (h_{10}, h_{20}) is not distinguishable from $(h_{10} + g(\boldsymbol{x}_{3it}), h_{20} - g(\boldsymbol{x}_{3it}))$ for any function of $g(\cdot)$ (for further detail, see Ai and Li 2008).

A finite sample approximation for $h_j(\cdot)$ is to use series approximations

$$h_{j0}(\cdot) \simeq \boldsymbol{p}_j(\cdot)'\boldsymbol{\pi}_j \tag{8.3.9}$$

The simplest series base function is to use the power series. However, the power series can be sensitive to outliers. Ai and Li (2008) suggest using the piecewise local polynomial spline as a base function in nonparametric series estimation. An lth-order univariate B-spline base function is given by Chui (1992, chapter 4):

$$B_r(x \mid l_0, \ldots, l_r) = \frac{1}{(r-1)!} \sum_{j=1}^{r} (-1)^j \binom{r}{j} \, [\max\,(0, x - l_j)]^{r-1}, \tag{8.3.10}$$

where $l_0, l_1, \ldots, l_r$ are the evenly spaced design knots on the support of X. When $r = 2$, (8.3.10) gives a piecewise linear spline, and when $r = 4$, it gives piecewise cubic splines (i.e., the third-order polynomials). Substituting the parametric specification (8.3.9) in lieu of $h_{j0}(\cdot)$ into (8.3.1), one obtains a parametric analog of (8.3.8). Just like the parametric case, one can remove α_i by taking the deviation of y_{it} from the ith individual time series mean $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ for the model (8.3.8) under (8.3.9) and (8.3.10). The consistent estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\pi}_j$, $j = 1, \ldots, m$ can then be obtained by minimizing

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(y_{it} - \nu_0(\boldsymbol{x}_{it}, \boldsymbol{\theta}) - \sum_{j=1}^{m} \boldsymbol{p}_j(\nu_j(\boldsymbol{x}_{it}, \boldsymbol{\theta}))'\boldsymbol{\pi}_j\right)\right. \\ \left. -\frac{1}{T}\sum_{s=1}^{T}\left(y_{is} - \nu_0(\boldsymbol{x}_{is}, \boldsymbol{\theta}) - \sum_{j=1}^{m} \boldsymbol{p}_j(\nu_j(\boldsymbol{x}_{is}, \boldsymbol{\theta})'\boldsymbol{\pi}_j)\right)\right\}^2 \tag{8.3.11}$$

Shen (1997), Newey (1997), and Che and Shen (1998) show that both $\hat{\boldsymbol{\theta}}$ and $\hat{h}_j(\cdot)$, $j = 1, \ldots, m$ are consistent and asymptotically normally distributed if $k_j \longrightarrow \infty$ while $\frac{k_j}{N} \longrightarrow 0$ (at a certain rate) where k_j denotes the dimension of $\boldsymbol{\pi}_j$.

The series approach can also be extended to the sample selection model (or partial linear model) discussed in Chapter 7, where

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + m(z_{it}) + \alpha_i + u_{it}, \tag{8.3.12}$$

if the dummy variable $d_{it} = 1$. The sample selection effect $m(z_{it})$ given $d_{it} = 1$ can be approximated as

$$m(z_{it}) \sim \sum_{j=1}^{m} h_j(z_{jit}), \tag{8.3.13}$$

where $h_j(\cdot)$ is an unknown function. For identification purpose, $h_j(\cdot)$ is commonly assumed to satisfy the local restriction $h_j(0) = 0$ for all j and the exclusive restriction that $z_{1it}, \ldots, z_{mit}$ are mutually exclusive. Then each $h_j(\cdot)$ can be approximated by the linear sieve $\boldsymbol{p}_j^{k_j}(\cdot)'\boldsymbol{\pi}_j$, where $\boldsymbol{p}_j^{k_j}(\cdot)$ is a vector of approximating functions satisfying $\boldsymbol{P}_j^{k_j}(0) = \boldsymbol{0}$. The unknown parameters $\boldsymbol{\beta}$ and the coefficients $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \ldots, \boldsymbol{\pi}_m')'$ can be estimated by the generalized least squares estimator if α_i are treated as random and uncorrelated with $(\boldsymbol{x}_{it}, z_{it})$, or by minimizing

$$\sum_{i=1}^{N}\sum_{s<t}\left[(y_{it} - y_{is}) - (\boldsymbol{x}_{it} - \boldsymbol{x}_{is})'\boldsymbol{\beta} - \sum_{j=1}^{m}(\boldsymbol{P}_j^{k_j}(z_{jit}) - \boldsymbol{P}_j^{k_j}(z_{is}))'\boldsymbol{\pi}_j\right]^2. \tag{8.3.14}$$

Ai and Li (2005) show that the resulting estimator is consistent and derive its asymptotic distribution.[1]

The methods discussed above are based on the assumption that α_i and $m(\boldsymbol{x}_{it})$ are separable whether α_i are treated as random or fixed. When the individual-specific effects α_i are not separable in the nonlinear conditional mean function $E(y_{it} \mid \boldsymbol{x}_{it}, \alpha_i) = m(\boldsymbol{x}_{it}, \alpha_i)$, the problem becomes very complicated.

Consider a model of the form

$$y_{it} = m(\boldsymbol{x}_{it}, \alpha_i) + u_{it}, \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T. \end{array} \tag{8.3.15}$$

[1] Ai and Li (2005) show that the nonlinear least squares estimator of $\boldsymbol{\beta}$ is asymptotically normally distributed, but not $\boldsymbol{\pi}_j$.

Suppose $(\boldsymbol{y}_i, \boldsymbol{x}_i)$ are randomly drawn, where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT}), \boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \ldots, x_{iT})'$. Under the assumption that $E(u_{it}|\boldsymbol{x}_{it}, \alpha_i) = 0$ and u_{it} is independently identically distributed over i and t with finite variance σ_u^2 and the normalization condition that there exists an $\boldsymbol{x}^*$ such that $m(\boldsymbol{x}^*, \alpha_i) = \alpha_i$, Ju, Gan, and Li (2019) suggest the following steps to identify the functional $m(\boldsymbol{x}, \alpha)$ and the cumulative distribution function of $f(u_{it}|\boldsymbol{x}_{it})$ conditional on $\boldsymbol{x}_{it} = \boldsymbol{x}_{il}$ using the method of deconvolution of characteristic functions[2]:

Step 1: Identification of the characteristics function of $u, \phi_u(s)$, where $\phi_w(\cdot) = \int e^{sw\sqrt{-1}} f(w)dw$ and $f(w)$ is the density function of the random variable w.

Conditional on $\boldsymbol{x}_{it} = \boldsymbol{x}_{il}$,

$$\begin{aligned} y_{it} - y_{il} &= [m(\boldsymbol{x}_{it}, \alpha_i) - m(\boldsymbol{x}_{il}, \alpha_i)] + u_{it} - u_{il} \\ &= u_{it} - u_{il}, \end{aligned} \tag{8.3.16}$$

Then independence of $(u_{it}|\boldsymbol{x}_{il}, \alpha_i)$ gives

$$\phi_u(s) = \sqrt{\phi_{y_{it}-y_{il}}(s|\boldsymbol{x}_{it} = \boldsymbol{x}_{il})}, \tag{8.3.17}$$

where $\phi_{y_{it}-y_{il}}(s|\boldsymbol{x}_{it} = \boldsymbol{x}_{il})$ can be identified from the data.

Step 2: Identification of distributions $\left\{m(\boldsymbol{x}_{it}, \alpha_i)|\boldsymbol{x}_{il} = \boldsymbol{x}^*\right\}$ and $\left\{\alpha_i|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*\right\}$.

$$\phi_{y_{it}}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*) = \phi_{m(\boldsymbol{x}_{it}, \alpha_i)}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*) \cdot \phi_u(s) \tag{8.3.18}$$

$$\phi_{y_{il}}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*) = \phi_{\alpha_i}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*) \cdot \phi_u(s). \tag{8.3.19}$$

Then,

$$\phi_{m(\boldsymbol{x}_{it}, \alpha_i)} = \frac{\phi_{y_{it}}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*)}{\phi_u(s)}, \tag{8.3.20}$$

$$\phi_{\alpha_i}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*) = \frac{\phi_{y_{il}}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*)}{\phi_u(s)}, \tag{8.3.21}$$

where $\phi_{y_{it}}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*)$ and $\phi_{y_{il}}(s|\boldsymbol{x}_{it}, \boldsymbol{x}_{il} = \boldsymbol{x}^*)$ are readily identifiable from the data and $\phi_u(s)$ is identified by step 1. The conditional cumulative distribution function (CDF) is obtained as

$$F_{\alpha_i|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(\alpha|x, \bar{x}) = \frac{1}{2} - \lim_{x\to\infty} \int_{-x}^{x} \left(\frac{\exp(-s\alpha\sqrt{-1})}{2\pi s\sqrt{-1}}\right) \phi_{\alpha_i|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(s|\boldsymbol{x}, \boldsymbol{x}^*)ds, \tag{8.3.22}$$

$$F_{m(x,\alpha_i)|X_{it}X_{i\tau}}(w|x, \boldsymbol{x}^*) = \frac{1}{2} - \lim_{x\to\infty} \int_{-x}^{x} \left(\frac{\exp(-sw\sqrt{-1})}{2\pi s\sqrt{-1}}\right) \phi_{m(x,\alpha_i)|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(s|x, \bar{x})ds, \tag{8.3.23}$$

Step 3: The conditional quantile function,

$$Q_{m(x,\alpha_i)|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(q|\boldsymbol{x}, \boldsymbol{x}^*) = \inf\left\{w : F_{m(x,\alpha_i)|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(w|\boldsymbol{x}, \boldsymbol{x}^*) \geq q\right\}, q \in (0, 1). \tag{8.3.24}$$

According to the property of quantiles, we have identification of the functional $m(\boldsymbol{x}, \cdot)$,

$$m(\boldsymbol{x}, a) = Q_{m(x,\alpha_i)|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(F_{\alpha_i|\boldsymbol{x}_{it}\boldsymbol{x}_{i\tau}}(a|\boldsymbol{x}, \boldsymbol{x}^*)|\boldsymbol{x}, \boldsymbol{x}^*), \tag{8.3.25}$$

for all $\boldsymbol{x}$ and a.

[2] Ju, Gan, and Li's (2019) conditions are actually less restrictive than the ones stated here.

Step 4: Identification of the quantile function $Q_{\alpha_i|\boldsymbol{x}_{it}}(q|x)$.

From the conditional independence, de-convoluting y conditional on contemporeneous covariates $\boldsymbol{x}_{it} = \boldsymbol{x}$ yields

$$Q_{m(\boldsymbol{x}_{it},\alpha_i)}(s|\boldsymbol{x}_{it} = \boldsymbol{x}) = \phi_{y_{it}}(s|\boldsymbol{x}_{it} = \boldsymbol{x})/\phi_{u_{it}}(s|\boldsymbol{x}_{it} = \boldsymbol{x}) \tag{8.3.26}$$

and $Q_{m(\boldsymbol{x}_{it},\alpha_i)}(s|\boldsymbol{x}_{it} = \boldsymbol{x})$ is identifiable. Hence, the CDF $F_{m(\boldsymbol{x}_{it},\alpha_i)}(w|\boldsymbol{x})$ can be identified.

The conditional characteristics functions $\phi_{y_{it}}(s|\boldsymbol{x}_{it},\boldsymbol{x}_{il})$, $\phi_{y_{il}}(s|\boldsymbol{x}_{it},\boldsymbol{x}_{il})$, $\phi_{y_{it}}(s|\boldsymbol{x})$, $\phi_{y_{il}}(s|\boldsymbol{x})$, and $\phi_{y_{it}-y_{il}}(s|\boldsymbol{x}_{it} = \boldsymbol{x}_{il} = \boldsymbol{x})$ can be estimated using any of the nonparametric methods such as kernel, series, or k-nearest-neighbor methods (e.g., Li and Racine 2007). Ju, Gan, and Li (2019) further revise these methods to take account the issues of selecting proper bandwidths and how to identify $m(\boldsymbol{x},\alpha)$ when the joint density of $f(\boldsymbol{x},\boldsymbol{x}^*)$ is small.

When α_i and $m(\boldsymbol{x}_{it})$ are separable, as in (8.3.8), the nonparametric estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\pi}_j$ can be used to test the parametric specification of the model following the idea of Hong and White (1995). However, the strict exogeneity assumption of $\boldsymbol{x}_{it}$ excludes the inclusion of lagged dependent variables. Neither is this approach of replacing unknown $h_j(\cdot)$ by series expansion easily generalizable to some or nonlinear panel data models (for further discussion, see Ai and Li 2008; Su and Ullah 2011). When α_i is nonseparable from $E(y_{it}|\boldsymbol{x}_{it}) = m(\boldsymbol{x}_{it},\alpha_i)$, as the above discussions show, statistical inference becomes extremely complicated even though $m(\boldsymbol{x},\alpha)$ in principle is consistently estimable.