CHAPTER 12

Miscellaneous Topics

In this chapter we give brief introduction to some miscellaneous topics. We shall first consider panel duration and count data models in Section 12.1 and 12.2, respectively. Section 12.3 introduces the quantile regression model. Section 12.4 considers statistical inference using simulation methods. Section 12.5 discusses the conventional error components formulation for panels with more than two dimensions. Section 12.6 considers issues of measurement errors and indicates how one can take advantage of the panel structure to identify and estimate an otherwise unidentified model. Section 12.7 discusses nonparametric approaches for panel data modeling.

12.1 DURATION MODEL

Duration models study the length of time spent in a given state before transition to another state, such as the length of time unemployed. The length of the interval between two successive events is called a *duration*. A *duration* is a nonnegative random variable, denoted by D, representing the length of a time period spent by an individual or a firm in a given state. The cumulative distribution function of D is defined as

$$\begin{aligned} F(t) &= \text{Prob}(D < t) \\ &= \textstyle\int_0^t f(s)\,ds, \end{aligned} \tag{12.1.1}$$

where

$$f(t) = \frac{dF(t)}{dt}. \tag{12.1.2}$$

Let A_{its} denote the event that nothing happens between time period t and $t + s$ for individual i; then an event of interest (say doctor visit) occurs at $t + s$. Suppose the probability of event $A_{it,t+\Delta}$ occurs where the time distance between two adjacent time periods approaches 0 is given by

$$P(A_{it,t+\Delta}) = \mu_{it}\Delta t. \tag{12.1.3}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

Under the assumption that μ_{it} stays constant between period t and $t+s$,

$$P(A_{its}) = (1-\mu_{it}\Delta t)^{\frac{s}{\Delta t}}\mu_{it}\Delta t. \tag{12.1.4}$$

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,

$$P(A_{its}) = \exp(-\mu_{it}s)\mu_{it}\Delta t. \tag{12.1.5}$$

Suppose μ_{it} stays constant between time period 0 and $t+\Delta t$, the probability that an individual stayed in a state (say unemployment) from 0 to t and moved out at $t+\Delta t$, $f_i(t)\Delta t$, is given by (12.1.5). Then

$$\text{Prob}(D_i \geq t) = \exp(-\mu_{it}t). \tag{12.1.6}$$

The cumulative distribution function of D,

$$\begin{aligned} F_i(t) &= \text{Prob}(D_i < t) \\ &= 1 - \text{Prob}(D_i \geq t) \\ &= 1 - \exp(-\mu_{it}t). \end{aligned} \tag{12.1.7}$$

Hence

$$\begin{aligned} \mu_{it} &= \lim_{\Delta t\to 0}\frac{\text{Prob}[t \leq D_i < t+\Delta t \mid D_i \geq t]}{\Delta t} \\ &= \frac{f_i(t)}{1-F_i(t)}, \end{aligned} \tag{12.1.8}$$

where μ_{it} is called the *hazard function* of the duration variable D_i and $f_i(t) = \exp(-\mu_{it}t)\mu_{it}$. The hazard function, μ_{it}, gives the instantaneous conditional probability of leaving the current state (*the death of a process*) and

$$S_i(t) = 1 - F_i(t) = \text{Prob}[D_i \geq t], \tag{12.1.9}$$

is called the *survival function*. Let μ_{is} be the instant hazard rate at time s, the probability that the ith individual survives from time s to $s+\Delta s$ is equal to $e^{-\mu_{is}\Delta s}$ ((12.1.6)). Breaking up the interval $(0,t)$ to n subintervals, $n = \frac{t}{\Delta s}$, and using $S(0)=1$ yields the probability that the ith individual survives at time t equals to

$$\prod_{s=1}^{n}\exp(-\mu_{is}\Delta s) = \exp\left\{-\sum_{s=1}^{n}\mu_{is}\Delta s\right\}. \tag{12.1.10}$$

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

$$S_i(t) = \exp\left(-\int_0^t \mu_{is}ds\right). \tag{12.1.11}$$

It follows that

$$\mu_{it} = \text{Prob}(D_i = t \mid D_{it} \geq t) = -\frac{d\, lnS_i(t)}{dt}. \tag{12.1.12}$$

When $\mu_{it} = \mu_i$, the expected duration for the ith individual is

$$E(D_i) = \int_0^\infty t\mu_i \exp(-\mu_i t)dt = \frac{1}{\mu_i}. \tag{12.1.13}$$

Suppose the data on N individuals take the form that each individual either experiences one complete spell at time t_i, that is, $D_i = t_i$, or right-censored at time t_i^*, that is, $D_i \geq t_i^*$. Suppose $\mu_{it} = \mu_i$ and $i = 1, \ldots, n$ complete their spells of duration t_i; then

$$f_i(t_i) = \mu_i \exp(-\mu_i t_i), \quad i = 1, \ldots, n, \tag{12.1.14}$$

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

$$S_i(t_i^*) = 1 - F_i(t_i^*) = \exp(-\mu_i t_i^*), \quad i = n+1, \ldots, N. \tag{12.1.15}$$

Under the assumption that cross-sectional units are independently distributed, the likelihood function for the N units takes the form,

$$L = \prod_{i=1}^{n} f_i(t_i) \cdot \prod_{i=n+1}^{N} [1 - F_i(t_i^*)]. \tag{12.1.16}$$

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

$$\mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}). \tag{12.1.17}$$

Substituting (12.1.17) into (12.1.16), one can obtain the maximum-likelihood estimator of $\boldsymbol{\beta}$ by maximizing the logarithm of (12.1.16).

Alternatively, noting that

$$\begin{aligned} E \log t_i &= \mu_i \int_0^\infty (\log t) \exp(-\mu_i t)\, dt \\ &= -c - \log \mu_i, \end{aligned} \tag{12.1.18}$$

where $c \simeq 0.577$ is Euler's constant, and

$$\begin{aligned} \text{Var}(\log t_i) &= E(\log t_i)^2 - [E \log t_i]^2 \\ &= \frac{\pi^2}{6}, \end{aligned} \tag{12.1.19}$$

one can put the duration model in a regression framework,

$$\begin{aligned} \log t_i + 0.557 &= -\mathbf{x}_i' \boldsymbol{\beta} + u_i, \\ i &= 1, \ldots, n, \end{aligned} \tag{12.1.20}$$

where $E(u_i) = 0$ and $\text{var}(u_i) = \frac{\pi^2}{6}$. Consistent estimate of $\boldsymbol{\beta}$ can be obtained by the least-squares method using the n subsample of individuals who experience one complete spell. However, the least-squares estimator has covariance matrix

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

$\frac{\pi^2}{6}\left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right)^{-1}$, which is greater than the covariance matrix of the MLE of (12.1.16).

Cox (1972) proposes a proportional hazard model to take account the heterogeneity across individuals (and over time) by letting

$$\mu_{it} = \mu(t)g(\mathbf{x}_i), \tag{12.1.21}$$

where $\mu(t)$ is the so-called *baseline hazard* function and $g(\cdot)$ is a known function of observable exogenous variables $\mathbf{x}_i$. To ensure nonnegativity of μ_{it}, a common formulation for $g(\mathbf{x}_i)$ is to let

$$g(\mathbf{x}_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}). \tag{12.1.22}$$

Then

$$\frac{\partial \mu_{it}}{\partial x_{ki}} = \beta_k \cdot \exp(\mathbf{x}_i'\boldsymbol{\beta})\mu(t) = \beta_k \cdot \mu_{it} \tag{12.1.23}$$

has a constant proportional effect on the instantaneous conditional probability of leaving the state. However, $\mu_{it} = \mu(t)cc^{-1}g(\mathbf{x}_i)$ 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 $g(\mathbf{x}^*) = 1$.

One can simultaneously estimate the *baseline hazard* function $\mu(t)$ and $\boldsymbol{\beta}$ by maximizing the logarithm of the likelihood function (12.1.16). However, Cox's proportional hazard model allows the separation of the estimation of $g(\mathbf{x}_i)$ from the estimation of the baseline hazard $\mu(t)$. Let $t_1 < t_2 < \cdots < t_j < \cdots < t_n$ denote the observed ordered discrete exit times of the 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 (12.1.21) and (12.1.22) into the likelihood function (12.1.16) yields Cox's proportional hazard model likelihood function,

$$\begin{aligned} L &= \prod_{i=1}^{n} \exp(\mathbf{x}_i'\boldsymbol{\beta})\mu(t_i) \cdot \exp[-\exp(\mathbf{x}_i'\boldsymbol{\beta}) \int_0^{t_i} \mu(s)ds] \\ &\quad \cdot \prod_{i=n+1}^{N} \exp[-\exp(\mathbf{x}_i'\boldsymbol{\beta}) \int_0^{t_i^*} \mu(s)ds] \\ &= \prod_{i=1}^{n} \exp(\mathbf{x}_i'\boldsymbol{\beta})\mu(t_i) \\ &\quad \cdot \exp\left\{ -\int_0^{\infty} \left[\sum_{h \epsilon R(t_\ell)} \exp\,(\mathbf{x}_h'\boldsymbol{\beta}) \right] \mu(t)dt \right\} \\ &= L_1 \cdot L_2, \end{aligned} \tag{12.1.24}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

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,

$$L_1 = \prod_{i=1}^{n} \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\sum_{h \epsilon R(t_i)} \exp(\mathbf{x}_h'\boldsymbol{\beta})}, \tag{12.1.25}$$

$$\begin{aligned} L_2 = \sum_{i=1}^{n} &\left[\sum_{h \epsilon R(t_i)} \exp(\mathbf{x}_h'\boldsymbol{\beta})\mu(t_i) \right] \\ &\cdot \exp\left\{ -\int_0^\infty \left[\sum_{h \epsilon R(t_j)} \exp(\mathbf{x}_h'\boldsymbol{\beta}) \right] \mu(s)ds \right\}. \end{aligned} \tag{12.1.26}$$

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

$$\text{Cov}(\hat{\boldsymbol{\beta}}_p) = -\left[E \frac{\partial^2 \log L_1}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right]^{-1}. \tag{12.1.27}$$

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 (12.1.16) or semiparametrically through the relation

$$-\log S_i(t_i) = \mathbf{x}_i'\boldsymbol{\beta} + \int_0^{t_i} \mu(s)ds + \epsilon_i \tag{12.1.28}$$

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

To allow for the presence of unobserved heterogeneity, mixture models have been proposed for the hazard rate,

$$\mu_{it} = \mu(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})\alpha_i, \alpha_i > 0 \tag{12.1.29}$$

where α_i denotes an unobserved heterogeneity term for the ith individual, independent of $\mathbf{x}_i$ and normalized with $E(\alpha_i) = 1$. Common assumptions for the heterogeneity are gamma, inverse gamma, or log-normal. Once the heterogeneity distribution is specified, one can integrate out α_i to derive the marginal survivor function or expected duration conditional on $\mathbf{x}_i$.

The idea of studying the duration of an event based on the hazard rate (instant rate of exit) has wide applications in economics and finance, for example, duration of a strike, unemployment, or a mortgage. It can also be applied to predict exit of an event in the future based on current state variables. For instance, Duan et al. (2012) define the average exit intensity for the period $[t, t+\tau]$ as:

$$\mu_{it}(\tau) = -\frac{ln[1 - F_{it}(\tau)]}{\tau}, \tag{12.1.30}$$

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 (12.1.8) or (12.1.12), which they call the forward intensity rate.) Then as in (12.1.9) and (12.1.12), the average exit intensity for the period $[t, t+\tau]$ becomes

$$\begin{aligned} \mu_{it}(\tau) &= -\frac{ln[1 - F_{it}(\tau)]}{\tau} \\ &= -\frac{ln[\exp(-\int_t^{t+\tau} \mu_{is} ds)]}{\tau}. \end{aligned} \tag{12.1.31}$$

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 (12.1.8) that the instantaneous forward exit intensity at time $t+\tau$ is:

$$\psi_{it}(\tau) = \frac{F'_{it}(\tau)}{1 - F_{it}(\tau)} = \mu_{it}(\tau) + \mu'_{it}(\tau)\tau. \tag{12.1.32}$$

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

$$\int_0^1 e^{-\mu_{it}(\tau+s)s} \psi_{it}(\tau+s)\, ds. \tag{12.1.33}$$

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

$$\int_0^1 e^{-\mu_{it}(\tau+s)s} \phi_{it}(\tau+s)\, ds. \tag{12.1.34}$$

The actual exit is recorded at discrete time, say in a month or 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

$$e^{-\mu_{it}(\tau)\tau\Delta t}\left[1 - e^{-\psi_{it}(\tau)\Delta t}\right], \tag{12.1.35}$$

and

$$e^{-\mu_{it}(\tau)\tau\Delta t}\left[1 - e^{-\phi_{it}(\tau)\Delta t}\right], \tag{12.1.36}$$

Table 12.1. *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	4012	32	0.80	257	6.41
1992	4009	28	0.70	325	8.11
1993	4195	25	0.6	206	4.91
1994	4433	24	0.54	273	6.16
1995	5069	19	0.37	393	7.75
1996	5462	20	0.37	463	8.48
1997	5649	44	0.78	560	9.91
1998	5703	64	1.12	753	13.20
1999	5422	77	1.42	738	13.61
2000	5082	104	2.05	616	12.12
2001	4902	160	3.26	577	11.77
2002	4666	81	1.74	397	8.51
2003	4330	61	1.41	368	8.50
2004	4070	25	0.61	302	7.42
2005	3915	24	0.61	291	7.43
2006	3848	15	0.39	279	7.25
2007	3767	19	0.50	352	9.34
2008	3676	59	1.61	285	7.75
2009	3586	67	1.87	244	6.80
2010	3396	25	0.74	242	7.13
2011	3224	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).

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

$$\mu_{it}(\tau) = \frac{1}{\tau}[\mu_{it}(\tau-1)(\tau-1) + \psi_{it}(\tau-1)]. \tag{12.1.37}$$

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 attribute, $\mathbf{x}_{it}$, a convenient specification to ensure $\phi_{it}(\tau) \leq \psi_{it}(\tau)$ is to let

$$\phi_{it}(\tau) = \exp(\mathbf{x}_{it}'\boldsymbol{\gamma}(\tau)), \tag{12.1.38}$$

and

$$\psi_{it}(\tau) = \phi_{it}(\tau) + \exp(\mathbf{x}_{it}'\boldsymbol{\beta}(\tau)). \tag{12.1.39}$$

Duan, Sun, and Wang (2012) use the monthly data ($\Delta t = \frac{1}{12}$) 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 12.1 provides the summaries of the number of active companies, defaults/bankruptcies, and other exits each year. The overall default rate ranges between 0.37 percent and 3.26 percent of the firms in each sample year. Other forms of exit are significantly higher, ranging from 4.91 percent to 13.61 percent. The macro and firm-specific attributes for $\phi_{it}(\tau)$ and $\psi_{it}(\tau)$ include (trailing) one-year S&P 500 return (SP500); three-month US Treasury bill rate; the firm's distance to default (DTD), which is a volatility-adjusted leverage measure based on Merton (1974) (for details, see Duan and Wang 2012); ratio of cash and short-term investments to the total assets (CASH/TA); ratio of net income to the 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); one-year idiosyncratic volatility, calculated by regressing individual monthly stock return on the value-weighted the Center for Research in Security Prices (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 take account the impact of the massive US governmental interventions during the 2008–09 financial crisis, Duan et al. (2012) also include a common bail out 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 0 otherwise. Specifically

$$\begin{aligned}\phi_{it}(\tau) = \exp\{&\lambda(\tau)\exp\left[-\delta(\tau)(t-t_B)\right]1((t-t_B)>0) \\ &+ \mathbf{x}_{it}'\boldsymbol{\gamma}(\tau)\},\end{aligned} \tag{12.1.40}$$

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

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

$$\prod_{i=1}^{N}\prod_{t=0}^{T-1}\mathcal{L}_{it}(\tau). \tag{12.1.41}$$

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

$$\begin{aligned}\mathcal{L}_{it}(\tau) = {}& 1\{t_{oi} \le t, \tau_{ci} > t+\tau\}P_t(\tau_{ci} > t+\tau) \\ &+ 1\{t_{oi} < t, \tau_{ci} = \tau_{oi} \le t+\tau\}P_t(\tau_{ci}; \tau_{ci} = \tau_{oi} \le t+\tau) \\ &+ 1\{t_{oi} < t, \tau_{ci} \ne \tau_{oi}, \tau_{ci} \le t+\tau\}P_t(\tau_{ci}; \tau_{ci} \ne \tau_{oi} \\ \text{and } &\tau_{ci} \le t+\tau) + 1\{t_{oi} > t\} + 1\{\tau_{ci} \le t\},\end{aligned} \tag{12.1.42}$$

$$P_t(\tau_{ci} > t+\tau) = \exp\left[-\sum_{s=0}^{\tau-1}\psi_{it}(s)\Delta t\right], \tag{12.1.43}$$

$$P_t(\tau_{ci} \mid \tau_{ci} = \tau_{oi} \leq t + \tau)$$
$$= \begin{cases} 1 - \exp[-\phi_{it}(0)\Delta t], \text{ if } \tau_{oi} = t+1, \\ \exp\left[-\sum_{s=0}^{\tau_{ci}-t-2} \psi_{it}(s)\Delta t\right] \\ \cdot [1 - \exp[-\phi_{it}(\tau_{c_i} - t - 1)\Delta t]\}, \text{ if } t+1 < \tau_{ci} \leq t + \tau, \end{cases} \tag{12.1.44}$$

$$P_t(\tau_{ci} \mid \tau_{ci} \neq \tau_{oi} \leq t + \tau)$$
$$= \begin{cases} \exp[-\phi_{it}(0)\Delta t] - \exp\,[-\psi_{it}(0)\Delta t], \text{ if } \tau_{oi} = t+1, \\ \exp[-\sum_{s=0}^{\tau_{ci}-t-2} \psi_{it}(s)\Delta t] \\ \cdot \{\exp[-\phi_{it}(\tau_{ci} - t - 1)\Delta t] - \exp[-\psi_{it}(\tau_{ci} - t - 1)]\}, \\ \text{if } t+1 < \tau_{ci} \leq t + \tau, \end{cases} \tag{12.1.45}$$

with $\Delta t = \frac{1}{12}$, and $\psi_{it}(s)$ and $\phi_{it}(s)$ take the form of (12.1.38) and (12.1.39), 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 12.1 plots each of the estimated $\boldsymbol{\gamma}(\tau)$ and $\boldsymbol{\beta}(\tau)$ and its 90% confidence interval with τ ranging from 0 month to 35 months. They show that some firm-specific attributes influence the forward intensity both in terms of level and trend. Figure 12.2 plot 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 percent, 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.

12.2 COUNT DATA MODEL

The count data model is the dual of the duration model. The duration model considers the probability that an event stays for certain time period before another event occurs. The count data models consider the probability that a

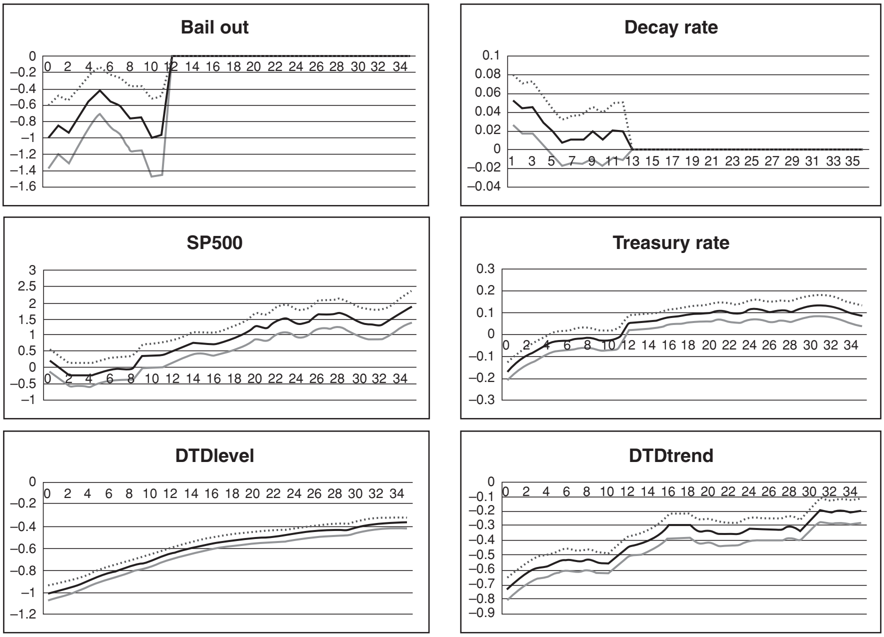

Figure 12.1. Parameter estimates for the forward default intensity function. The solid line is for the parameter estimates and the dotted lines depict the 90% confidence interval. *Source:* Duan, Sun, and Wang (2012, Fig. 1).

certain number of an event would occur during a fixed period of time. Under the assumption that the instant arrival rate is μ_{it}. The probability of the nonnegative integer count number y_{it} in a unit interval is given by a Poisson process.

$$P(y_{it}) = \frac{e^{-\mu_{it}}(\mu_{it})^{y_{it}}}{y_{it}!}, \quad y_{it} = 0, 1, 2, \ldots \tag{12.2.1}$$

To see this, suppose $y_{it} = 2$. Let $t + s_1$ and $t + s_1 + s_2$ be the time that the first and the second occurrence of the event of interest. Then $0 \leq s_1 < 1, 0 < s_2 < 1 - s_1$, and the probability that $y_{it} = 2$ is equal to the probability that one event occurs at $t + s_1$, another at $t + s_1 + s_2$ (or s_2 between 0 and $1 - s_1$), and no event occurs between $t + s_1 + s_2$ and $t + 1$),

$$\begin{aligned} P(y_{it} = 2) &= \int_0^1 \mu_{it} \exp\left(-\mu_{it} s_1\right) \left\{ \int_0^{1-s_1} \mu_{it} \exp\left[-\mu_{it} s_2\right] \right. \\ &\quad \left. \cdot \exp[-\mu_{it}(1 - s_1 - s_2)] ds_2 \right\} ds_1 \qquad (12.2.2) \\ &= \frac{(\mu_{it})^2 e^{-\mu_{it}}}{2}. \end{aligned}$$

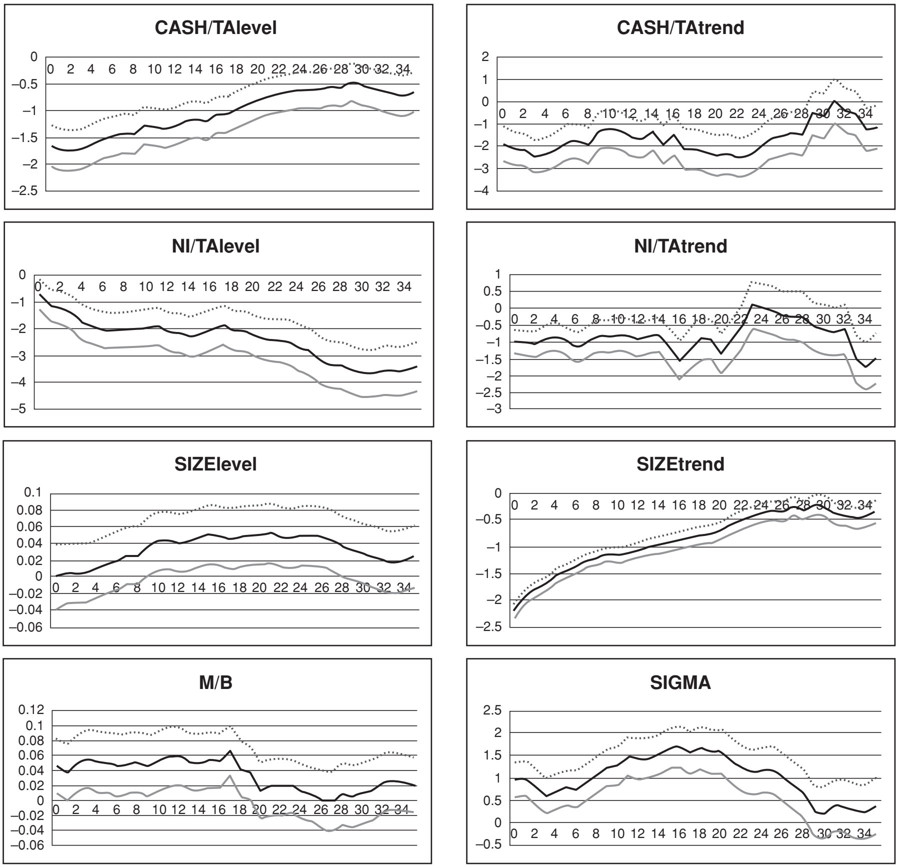

Figure 12.1 (*continued*).

Similarly, one can show that

$$P(y_{it} = r) = \frac{(\mu_{it})^r \exp(-\mu_{it})}{r!}. \tag{12.2.3}$$

The Poisson model implies y_{it} is independent over time,

$$\text{Prob}(y_{it} = r \mid y_{i,t-s} = \ell) = P(y_{it} = r), \tag{12.2.4}$$

$$E(y_{it}) = \mu_{it}, \tag{12.2.5}$$

and

$$\text{Var}(y_{it}) = \mu_{it}. \tag{12.2.6}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

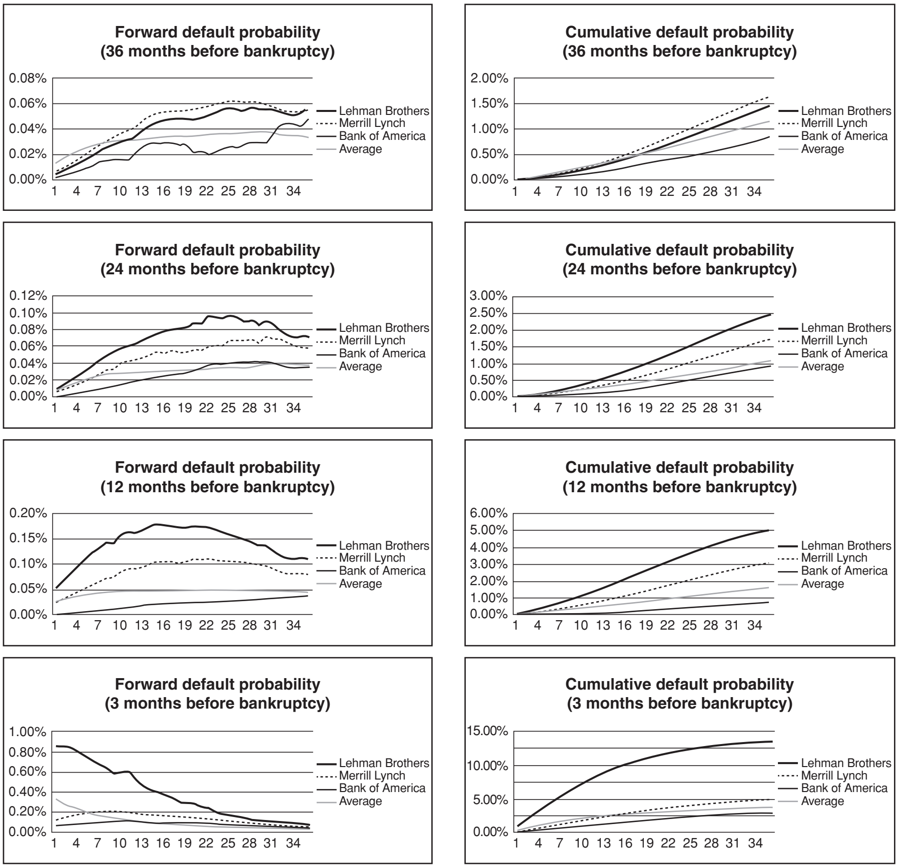

Figure 12.2. Lehman Brothers' term structure of forward and cumulative default probabilities. This figure shows the estimated term structure of forward default probabilities and that of cumulative default probabilities for Lehman Brothers, Merrill Lynch, Bank of America as well as the average values of the 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).

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

$$\log L = \sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it}\log(\mu_{it}) - \mu_{it} - \log(y_{it}!)] \tag{12.2.7}$$

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

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

restrictions on the parameters are to let

$$\mu_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i), \tag{12.2.8}$$

or to let

$$\mu_{it} = \alpha_i \exp(\mathbf{x}'_{it}\boldsymbol{\beta}), \quad \alpha_i > 0 \tag{12.2.9}$$

and $E(\alpha_i) = 1$.

When α_i is treated random and independent of $\mathbf{x}_{it}$ with known density function $g(\alpha)$, the marginal distribution of $(y_{i1}, \dots, y_{iT})$ takes the form

$$f(y_{i1}, \dots, y_{iT}) = \int \prod_{t=1}^{T} \left[\frac{(\mu_{it})^{y_{it}} \exp(-\mu_{it})}{y_{it}!} \right] g(\alpha) d\alpha. \tag{12.2.10}$$

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

$$f(y_i) = \frac{[\exp(\mathbf{x}'_i\boldsymbol{\beta})]^{y_i} \Gamma(y_i + \nu)}{y_i! \Gamma(\nu)} \left(\frac{1}{\exp(\mathbf{x}'_i\boldsymbol{\beta}) + \nu} \right)^{y_i + \nu}, \quad i = 1, \dots, N, \tag{12.2.11}$$

has a *negative binomial* distribution. But if $T > 1$, (12.2.10) no longer has the closed form. One computationally simpler method to obtain consistent estimator of $\boldsymbol{\beta}$ is to ignore the serial dependence of y_{it} because of the presence of α_i by considering the marginal (or unconditional) distribution of y_{it}. For instance, if μ_{it} takes the form of (12.2.9) and α is gamma distributed, then the unconditional distribution of y_{it} takes the form of (12.2.11). Maximizing the pseudo-joint likelihood function $\prod_{i=1}^{N} \prod_{t=1}^{T} f(y_{it})$ yields consistent estimator of $\boldsymbol{\beta}$ either N or T or both tend to infinity. The pseudo-MLE can also be used as initial values of the iterative schemes to obtain the MLE.

Conditional on α_i, the log-likelihood function remains of the simple form (12.2.7). When α_i is treated as fixed and μ_{it} takes either the form (12.2.8) or (12.2.9), the maximum-likelihood estimator of $\boldsymbol{\beta}$ and η_i, where $\eta_i = \exp(\alpha_i)$ or $\eta_i = \alpha_i$ if μ_{it} takes the form (12.2.8) or (12.2.9), respectively, are obtained by simultaneously solving the first-order conditions

$$\frac{\partial \log L}{\partial \eta_i} = \sum_{t=1}^{T} \left[\frac{y_{it}}{\eta_i} - \exp(\mathbf{x}'_{it}\boldsymbol{\beta}) \right] = 0, \quad i = 1, \dots, N, \tag{12.2.12}$$

$$\frac{\partial \log L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it} - \eta_i \exp(\mathbf{x}'_{it}\boldsymbol{\beta})]\mathbf{x}_{it} = \mathbf{0}. \tag{12.2.13}$$

Solving (12.2.12) yields the MLE of η_i conditional on $\boldsymbol{\beta}$ as

$$\hat{\eta}_i = \frac{\bar{y}_i}{\bar{\mu}_i}, \quad i = 1, \dots, N. \tag{12.2.14}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\bar{\mu}_i = T^{-1} \sum_{t=1}^{T} \exp(\mathbf{x}_{it}' \boldsymbol{\beta})$. Substituting $\hat{\eta}_i$ for η_i in (12.2.13), the MLE of $\boldsymbol{\beta}$ is the solution to

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} - \frac{\bar{y}_i}{\hat{\bar{\mu}}_i} \exp(\mathbf{x}_{it}' \hat{\boldsymbol{\beta}}) \right] \mathbf{x}_{it} = \mathbf{0}, \tag{12.2.15}$$

where $\hat{\bar{\mu}}_i = \frac{1}{T} \sum_{t=1}^{T} \exp(\mathbf{x}_{it}' \hat{\boldsymbol{\beta}})$. When $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, (12.2.15) is equivalent to

$$\underset{N \longrightarrow \infty}{\text{plim}} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[u_{it} - \frac{\bar{u}_i}{\bar{\mu}_i} \exp(\mathbf{x}_{it}' \boldsymbol{\beta}) \right] \mathbf{x}_{it} = \mathbf{0}, \tag{12.2.16}$$

where

$$\begin{aligned} u_{it} &= y_{it} - E(y_{it} \mid \mathbf{x}_{it}, \alpha_i), \\ &= y_{it} - \eta_i \exp(\mathbf{x}_{it}' \boldsymbol{\beta}), \end{aligned} \tag{12.2.17}$$

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

The strict exogeneity of $\mathbf{x}_{it}$ implies that

$$E(y_{it} \mid \mathbf{x}_{it}, \alpha_i) = E(y_{it} \mid \mathbf{x}_i, \alpha_i), \tag{12.2.18}$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')'$. Therefore, $E(u_{it} \mid \mathbf{x}_i) = 0$, and hence (12.2.17) follows. However, the MLE of α_i (or η_i) is consistent only when $T \longrightarrow \infty$.

The sufficient statistic for η_i is $\sum_{t=1}^{T} y_{it}$. Conditional on $\sum_{t=1}^{T} y_{it}$, the Poisson conditional log-likelihood function is given by (Hausman, Hall, and Griliches 1984)

$$\log L^* = \sum_{i=1}^{N} \sum_{t=1}^{T} \Gamma(y_{it} + 1) - \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \log \left\{ \sum_{s=1}^{T} \exp[-(\mathbf{x}_{it} - \mathbf{x}_{is})' \boldsymbol{\beta}] \right\}, \tag{12.2.19}$$

where $\Gamma(\cdot)$ is the gamma function. Equation (12.2.19) no longer involves the incidental parameters α_i. Maximizing (12.2.19) yields consistent and asymptotic normally distributed estimator under the usual regularity conditions. As a matter of fact, $\frac{\partial \log L^*}{\partial \boldsymbol{\beta}}$ is identical to (12.2.16) (for details, see Windmeijer (2008)).

The limitations of Poisson models are the mean-variance equality restriction ((12.2.5) and (12.2.6)) and conditional on α_i, y_{it} independent of $y_{i,t-1}$. These features often contradict to the observed phenomena that the (conditional) variance usually exceeds the (conditional) mean and y_{it} are not independent of $y_{i,t-1}$. The introduction of individual-specific effects, α_i, partially get around the overdispersion problem. For instance, under the assumption that μ_{it} takes the form of (12.2.9) and α_i follows a gamma distribution, (12.2.11) leads to

$$E(y \mid \mathbf{x}) = \exp(\mathbf{x}' \boldsymbol{\beta}) \tag{12.2.20}$$

and

$$\text{Var}\,(y \mid \mathbf{x}) = \exp(\mathbf{x}' \boldsymbol{\beta})[1 + \nu \exp\,(\mathbf{x}' \boldsymbol{\beta})] > E(y \mid \mathbf{x}). \tag{12.2.21}$$

One way to explicitly take account of serial dependence of y_{it} is to include lagged $y_{i,t-1}$ into the specification of the mean arrival function μ_{it}. However, inclusion of the lagged dependent variable in an exponential mean function may lead to rapidly exploding series. Crépon and Duguet (1997) suggest specifying

$$\mu_{it} = h(y_{i,t-1}) \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i). \tag{12.2.22}$$

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

$$h(y_{i,t-1}) = \exp(\gamma(1 - d_{i,t-1})), \tag{12.2.23}$$

or

$$h(y_{i,t-1}) = \exp(\gamma_1 \ell n(y_{i,t-1} + c d_{i,t-1}) + \gamma_2 d_{i,t-1}), \tag{12.2.24}$$

where c is a pre-specified constant, $d_{it} = 1$ if $y_{it} = 0$ and 0 otherwise. In this case, $\ell n y_{i,t-1}$ 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, and Windmeijer (2002) propose a linear feedback model of the form

$$\mu_{it} = \gamma y_{i,t-1} + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i). \tag{12.2.25}$$

Unfortunately, neither specification leads to easy to device 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).

Another often observed phenomena in data is that there is a much larger probability mass at the 0 (count) value than predicted by the Poisson model. One way to deal with this "excess zeros" in the data is to assume a two-part model or zero-inflated model in which the 0's and the positives come from two different data-generating process (e.g., Gurmu and Trivedi 1996; Harris and Zhao 2007), in which the probability of $y = 0$ or not is given by a binary process, say $F_1^*(0)$ and $(1 - F_1^*(0))$, and the probability y takes the count values of 0,1,2, ... from the count probability $F_2(y = r)$. Then the two part model assumes

$$\text{Prob}(y = 0) = F_1^*(0) \tag{12.2.26}$$

$$\text{Prob}(y = r) = [1 - F_1^*(0)]F_2(y = r), \quad r \geq 1. \tag{12.2.27}$$

The zero-inflated model assumes the zero-inflated model has probability

$$P(y = 0) = F_1^*(0) + [1 - F_1^*(0)]F_2(y = 0) \tag{12.2.28}$$

and

$$P(y = r) = [1 - F_1^*(0)]F_2^*(y = r) \text{ if } r \geq 1. \tag{12.2.29}$$

For additional discussions on modifying Poisson model to take account endogeneity, etc., see Trivedi and Munkin (2011) and Windmeijer (2008).

12.3 PANEL QUANTILE REGRESSION

The τth quantile of a random variable y, y_τ, for $0 < \tau < 1$ is defined as

$$\operatorname{Prob}(y \leq y_\tau) = \int_{-\infty}^{y_\tau} f(y)dy = F(y_\tau) = \tau, \tag{12.3.1}$$

where $f(y)$ denotes the probability density function of y. The sample location quantiles estimator for the τth sample quantile, $0 < \tau < 1$, for N random sample y_i is the solution to the minimization problem

$$\underset{c}{\text{Min}} \left\{ \sum_{i \in \psi_c} \tau \mid y_i - c \mid + \sum_{i \in \bar{\psi}_c} (1-\tau) \mid y_i - c \mid \right\} \tag{12.3.2}$$

where $\psi_c = \{i \mid y_i \geq c\}$ and $\bar{\psi}_c = \{i \mid y_i < c\}$.

As $N \to \infty$, eq. (12.3.2) divided by N converges to

$$\begin{aligned} S(c) &= (1-\tau) \int_{-\infty}^{c} \mid y - c \mid f(y)\, dy \\ &\quad + (\tau) \int_{c}^{\infty} \mid y - c \mid f(y)\, dy. \end{aligned} \tag{12.3.3}$$

Suppose $0 < c < y_\tau$. For $y < c, \mid y - c \mid = \mid y - y_\tau \mid - \mid y_\tau - c \mid$. For $c < y < y_\tau, \mid y - c \mid = \mid y_\tau - c \mid - \mid y - y_\tau \mid$. For $y > y_\tau$, $\mid y - c \mid = \mid y - y_\tau \mid + \mid y_\tau - c \mid$. Equation (12.3.3) can be written as

$$\begin{aligned} S(c) &= (1-\tau) \int_{-\infty}^{c} \mid y - c \mid f(y)\, dy \\ &\quad + \tau \int_{c}^{y_\tau} \mid y - c \mid f(y)\, dy \\ &\quad + \tau \int_{y_\tau}^{\infty} \mid y - c \mid f(y)\, dy \\ &= S(y_\tau) + \mid y_\tau - c \mid (\tau - F(c)) - \int_{c}^{y_\tau} \mid y - y_\tau \mid f(y) dy \\ &\geq S(y_\tau), \end{aligned} \tag{12.3.4}$$

where $S(y_\tau) = (1-\tau)\int_{-\infty}^{y_\tau} \mid y - y_\tau \mid f(y)dy + \tau \int_{y_\tau}^{\infty} \mid y - y_\tau \mid f(y)dy$. Similarly, one can show that for other values of c where $c \neq y_\tau$, $S(c) \geq S(y_\tau)$. Therefore, as $N \to \infty$, the solution (12.3.2) yields a consistent estimator of y_τ.

Koenker and Bassett (1978) generalize the ordinary notion of sample quantiles based on an ordering sample observations to the regression framework

$$\underset{\mathbf{b}}{\min} \left\{ \sum_{i \in \psi_b} \tau \mid y_i - \mathbf{x}_i'\mathbf{b} \mid + \sum_{i \in \bar{\psi}_b} (1-\tau) \mid y_i - \mathbf{x}_i'\mathbf{b} \mid \right\}, \tag{12.3.5}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

where $\psi_b = \{i \mid y_i \geq \mathbf{x}_i'\mathbf{b}\}$ and $\bar{\psi}_b = \{i \mid y_i < \mathbf{x}_i'\mathbf{b}\}$. When $\tau = \frac{1}{2}$, the quantile estimator (12.3.5) is the least absolute deviation estimator. Minimizing (12.3.5) can also be written in the form

$$\underset{\mathbf{b}}{\text{Min}} \sum_{i=1}^{N} \rho_\tau(y_i - \mathbf{x}_i\mathbf{b}), \tag{12.3.6}$$

where $\rho_\tau(u) := [\tau - 1(u \leq 0)]u$, and $1(A) = 1$ if A occurs and 0 otherwise. Equation (12.3.6) is equivalent to the linear programming form,

$$\text{Min } [\tau\mathbf{e}'\mathbf{u}^+ + (1-\tau)\mathbf{e}'\mathbf{u}^-] \tag{12.3.7}$$

subject to

$$\mathbf{y} = X\mathbf{b} + \mathbf{u}^+ - \mathbf{u}^-, \tag{12.3.8}$$

$$(\mathbf{u}^+, \mathbf{u}^-) \in R_+^{2N}, \tag{12.3.9}$$

where $\mathbf{e}$ is an $N \times 1$ vector of $(1, \ldots, 1)$, R_+^{2N} denotes the positive quadrant of the $2N$ dimensional real space such that if $u_i^+ > 0, u_i^- = 0$ and if $u_i^- > 0, u_i^+ = 0$. Sparse linear algebra and interior point methods for solving large linear programs are essential computational tools.

The quantile estimator for the panel data model,

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it}, \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{array} \tag{12.3.10}$$

is the solution of

$$\underset{\mathbf{b}(\tau), \alpha_i(\tau)}{\text{Min}} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(y_{it} - \mathbf{x}_{it}'\mathbf{b}(\tau) - \alpha_i(\tau)), \tag{12.3.11}$$

where

$$Q_\tau(y_{it} \mid \mathbf{x}_{it}, \alpha_i) = \mathbf{x}_{it}'\mathbf{b}(\tau) + \alpha_i(\tau) \tag{12.3.12}$$

is the τth conditional quantile.

The main idea of regression quantile is to break up the common assumption that u_{it} are independently, identically distributed. The conditional quantile (12.3.12) provides information on how $\mathbf{x}$ influence the location, scale and shape of the conditional distribution of the response. For instance,

$$u_{it} = (1 + \mathbf{x}_{it}'\boldsymbol{\gamma})\epsilon_{it}, \tag{12.3.13}$$

where $\mathbf{x}_{it}'\boldsymbol{\gamma} > 0$ and $\boldsymbol{\epsilon}_{it}$ has distribution function $F_\epsilon(\cdot)$. Then

$$\begin{aligned} Q_\tau(y_{it} \mid \mathbf{x}_{it}, \alpha_i) &= \mathbf{x}_{it}'\big(\boldsymbol{\beta} + \boldsymbol{\gamma}F_\epsilon^{-1}(\tau)\big) + \big(\alpha_i + F_\epsilon^{-1}(\tau)\big) \\ &= \mathbf{x}_{it}'\boldsymbol{\beta}(\tau) + \alpha_i(\tau). \end{aligned} \tag{12.3.14}$$

In other words, (12.3.14) is just a straightline describing the τth quantile of y_{it} given $\mathbf{x}_{it}$. One should not confuse (12.3.14) with the traditional meaning of $E(y_{it} \mid \mathbf{x}_{it}, \alpha_i)$.

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

Kato, Galvao, and Montes-Rojas (2012) show that the quantile estimator of $(\mathbf{b}(\tau), \alpha_i(\tau))$ of (12.3.11) is consistent and asymptotically normally distributed provided $\frac{N^2(\log N)^3}{T} \longrightarrow 0$ as $N \longrightarrow \infty$. The requirement that the time-dimension of a panel, T, to grow much faster than the cross-sectional dimension, N, as N increases is because directly estimating the individual-specific effects significantly increases the variability of the estimates of $\mathbf{b}(\tau)$. Standard linear transformation procedures such as first-differencing or mean differencing are not applicable in quantile regression. Koenker (2004) noted that shrinking the individual-specific effects toward a common mean can reduce the variability due to directly estimating the large number of individual-specific effects. He suggested a penalized version of (12.3.11),

$$\underset{\mathbf{b}(\tau),\alpha_i(\tau)}{\text{Min}} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(y_{it} - \mathbf{x}_{it}'\mathbf{b}(\tau) - \alpha_i(\tau)) + d\sum_{i=1}^{N} \mid \alpha_i(\tau) \mid \tag{12.3.15}$$

The penalty $d \mid \sum_{i=1}^{N} \alpha_i(\tau) \mid$ serves to shrink the individual effects estimates toward zero. When $d \longrightarrow 0$, (12.3.15) yields the quantile fixed effects estimator (12.3.11). When $d \longrightarrow \infty$, $\hat{\alpha}_i(\tau) \longrightarrow 0$ for all $i = 1, \ldots, N$. Minimizing (12.3.15) leads to improved performance for the estimates of the slope parameter $\boldsymbol{\beta}(\tau)$.

One trouble with (12.3.11) or (12.3.15) is that the individual-specific effects could change because the realized value of y_{it} at different time periods could fall into different quantiles. One way to get around this problem is to view the individual-specific effect summarizing the impact of some time-invariant latent variables while the error, u_{it}, bounces the responses y_{it} around from quantile to quantile. In other words, we condition not only on the observed covariates, $\mathbf{x}_{it}$, but also on the individual fixed effects, α_i, and replace the objective function (12.3.15) by pooling the estimates of individual quantile through

$$\text{Min} \sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{t=1}^{T} \omega_j \rho_{\tau_j}(y_{it} - \mathbf{x}_{it}'\mathbf{b}(\tau) - \alpha_i) + d\sum_{i=1}^{N} \mid \alpha_i \mid, \tag{12.3.16}$$

where ω_j is a relative weight given to the τ_j-th quantile. Monte Carlo studies show that shrinking the unconstrained individual-specific effects toward a common value helps to achieve improved performance for the estimates of the individual-specific effects and $\mathbf{b}(\tau_j)$.

Although introducing the penalty factor $d\sum_{i=1}^{N} \mid \alpha_i(\tau) \mid$ achieves the improved performance of panel quantile estimates, deciding d is a challenging question. Lamarche (2010) shows that when the individual-specific effects α_i are independent of $\mathbf{x}_{it}$ the penalized quantile estimator is asymptotically unbiased and normally distributed if the individual-specific effects, α_i, are drawn from a class of zero-median distribution functions. The regularization parameter, d, can thus be selected accordingly to minimize the estimated asymptotic variance.

12.4 SIMULATION METHODS

Panel data contains two dimensions – a cross-sectional dimension and a time dimension. Models using panel data also often contain unobserved heterogeneity factors. To transform a latent variable model involving missing data, random coefficients, heterogeneity, etc., into an observable model often requires the integration of latent variables over multiple dimensions (e.g., Hsiao 1989, 1991a,b, 1992a). The resulting panel data model estimators can be quite difficult to compute. Simulation methods have been suggested to get around the complex computational issues involving multiple integrations (e.g., Geweke 1991; Gourieroux and Monfort 1996; Hajivassiliou 1990; Hsiao and Wang 2000; Keane 1994; McFadden 1989; Pakes and Pollard 1989; and Richard and Zhang 2007).

The basic idea of simulation approach is to rely on the law of large numbers to obtain the approximation of the integrals through taking the averages of random drawings from a known probability distribution function. For instance, consider the problem of computing the conditional density function of $\mathbf{y}_i$ given $\mathbf{x}_i$, $f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ or some conditional moments $\mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta})$ say $E(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ or $E(\mathbf{y}_i \mathbf{y}_i' \mid \mathbf{x}_i; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of parameters characterizing these functions. In many cases, it is difficult to compute these functions because they do not have closed forms. However, if the conditional density or moments conditional on $\mathbf{x}$ and another vector $\boldsymbol{\eta}$, $f^*(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\eta}; \boldsymbol{\theta})$ or $\mathbf{m}^*(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\eta}; \boldsymbol{\theta})$, have closed forms and the probability distribution of $\boldsymbol{\eta}$, $P(\boldsymbol{\eta})$, is known, then from

$$f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \int f^*(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\eta}; \boldsymbol{\theta}) dP(\boldsymbol{\eta}), \tag{12.4.1}$$

and

$$\mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}) = \int \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}; \boldsymbol{\theta}) dP(\boldsymbol{\eta}), \tag{12.4.2}$$

we may approximate (12.4.1) and (12.4.2) by

$$\tilde{f}_H(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \frac{1}{H} \sum_{h=1}^{H} f^*(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\eta}_{ih}; \boldsymbol{\theta}), \tag{12.4.3}$$

and

$$\tilde{\mathbf{m}}_H(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}) = \frac{1}{H} \sum_{h=1}^{H} \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{ih}; \boldsymbol{\theta}), \tag{12.4.4}$$

where $(\boldsymbol{\eta}_{i1}, \ldots, \boldsymbol{\eta}_{iH})$ are H random draws from $P(\boldsymbol{\eta})$.

For example, consider the random effects panel Probit and Tobit models defined by the latent response function

$$y_{it}^* = \boldsymbol{\beta}' \mathbf{x}_{it} + \alpha_i + u_{it}, \tag{12.4.5}$$

where α_i and u_{it} are assumed to be independently normally distributed with mean 0 and variance σ_α^2 and 1, respectively, and are mutually independent. The

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

Probit model assumes that the observed y_{it} takes the form

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0, \\ 0 & \text{if } y_{it}^* \leq 0. \end{cases} \tag{12.4.6}$$

The Tobit model assumes that

$$y_{it} = \begin{cases} y_{it}^* & \text{if } y_{it}^* > 0, \\ 0 & \text{if } y_{it}^* \leq 0. \end{cases} \tag{12.4.7}$$

We note that the density function of α_i and u_{it} can be expressed as transformations of some standard distributions, here, standard normal, so that the density function of $\mathbf{y}_i' = (y_{i1}, \ldots, y_{iT})$ becomes an integral of a conditional function over the range of these standard distributions A:

$$f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \int_A f^*(\mathbf{y}_i \mid \mathbf{x}_i, \eta; \boldsymbol{\theta})\, dP(\eta) \tag{12.4.8}$$

with $p(\eta) \sim N(0, 1)$. For instance, in the case of Probit model,

$$f^*(\mathbf{y}_i \mid \mathbf{x}_i, \eta; \boldsymbol{\theta}) = \sum_{t=1}^{T} \Phi(\mathbf{x}_{it}'\boldsymbol{\beta} + \sigma_\alpha \eta_i)^{y_{it}} [1 - \Phi(\mathbf{x}_{it}'\boldsymbol{\beta} + \sigma_\alpha \eta_i)]^{1-y_{it}}, \tag{12.4.9}$$

and in the case of Tobit model,

$$\begin{aligned} f^*(\mathbf{y}_i \mid \mathbf{x}_i, \eta; \boldsymbol{\theta}) = & \prod_{t \epsilon \Psi_1} \phi(y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta} - \sigma_\alpha \eta_i) \\ & \cdot \prod_{t \epsilon \Psi_0} \Phi(-\mathbf{x}_{it}'\boldsymbol{\beta} - \sigma_\alpha \eta_i), \end{aligned} \tag{12.4.10}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and integrated normal, respectively, and $\Psi_1 = \{t \mid y_{it} > 0\}$ and $\Psi_0 = \{t \mid y_{it} = 0\}$. Since conditional on $\mathbf{x}_{it}$ and each of the H random draws of η from a standard normal distribution, $\eta_{ih}, h = 1, \ldots, H$, the conditional density function (12.4.9) on (12.4.10) is well defined in terms of $\boldsymbol{\beta}, \sigma_\alpha^2$, the approximation of $f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\beta}, \sigma_\alpha^2)$ can be obtained by taking their averages as in (12.4.3).

Random draws of η_h from $P(\eta)$ can be obtained through the *inversion* technique from a sequence of independent uniform [0, 1] pseudo-random draws

$$\eta_h = P^{-1}(\epsilon_h),$$

where $P^{-1}(\cdot)$ denote the inverse of P. For instance, if ϵ is normally distributed with mean μ and variance σ_ϵ^2, then $\eta_h = \Phi^{-1}(\frac{\epsilon_h - \mu}{\sigma_\epsilon})$. If η is a Weibull random variable with parameters a and b, $P(\eta_h) = 1 - \exp(-b\eta_h^a)$, then $\eta_h = \left[-\frac{1}{b} ln \epsilon_h\right]^{\frac{1}{a}}$.

The generation of a multivariate $\boldsymbol{\eta}_h$ can be obtained through recursive factorization of its density into lower dimensional density (e.g., Liesenfeld and Richard 2008). The basic idea of factorization of a k-dimensional

$\boldsymbol{\eta}_h = (\eta_{1h}, \ldots, \eta_{kh})$ is to write

$$P(\boldsymbol{\eta}_h) = P(\eta_{kh} \mid \boldsymbol{\eta}^*_{k-1,h})P(\eta_{k-1,h} \mid \boldsymbol{\eta}^*_{k-2,h}) \ldots P(\eta_{2h} \mid \eta_{1h})P(\eta_{1h}), \tag{12.4.11}$$

where $\eta^*_{j,h} = (\eta_{1h}, \ldots, \eta_{jh})$. For example, random draws from a multivariate normal density are typically obtained based on Cholesdy decomposition of its covariance matrix $\sum = \bigwedge \bigwedge'$, $\boldsymbol{\eta}_h = \bigwedge \boldsymbol{\eta}^*_h$, where $\bigwedge$ is a lower triangular matrix and $\boldsymbol{\eta}^*$ is standard multivariate normal with identity covariance matrix.

A particularly useful technique for evaluating high-dimensional integrals is known as *Importance Sampling*. The idea of importance sampling is to replace $P(\boldsymbol{\eta}_i)$ by an alternative simulator with density $\mu(\cdot)$. Substituting $\mu(\cdot)$ into (12.4.11)

$$f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \int f^*(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\eta}; \boldsymbol{\theta})\omega(\boldsymbol{\eta}_i)\mu(\boldsymbol{\eta}_i)\, d\boldsymbol{\eta}_i, \tag{12.4.12}$$

where $dP(\boldsymbol{\eta}_i) = p(\boldsymbol{\eta}_i)\, d\boldsymbol{\eta}_i$,

$$\omega(\boldsymbol{\eta}_i) = \frac{p(\boldsymbol{\eta}_i)}{\mu(\boldsymbol{\eta}_i)}. \tag{12.4.13}$$

Then the corresponding Monte Carlo simulator of (12.4.3), known as the "importance sampling" estimator, is given by

$$\tilde{f}_H(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \frac{1}{H} \sum_{h=1}^{H} \omega(\boldsymbol{\eta}^*_{ih})\mu(\boldsymbol{\eta}^*_{ih}) f^*(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\eta}^*_{ih}; \boldsymbol{\theta}), \tag{12.4.14}$$

where $\boldsymbol{\eta}^*_{ih}$ are random draws from $\mu(\boldsymbol{\eta}^*_i)$.

If u_{it} in the above example follows a first-order autoregressive process

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad \mid \rho \mid < 1, \tag{12.4.15}$$

then we can rewrite (12.4.5) as

$$y^*_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\alpha \eta_i + \sum_{\tau=1}^{t} a_{t\tau}\eta^*_{i\tau}, \tag{12.4.16}$$

where $\eta^*_{i\tau}$, $\tau = 1, \ldots, T$ are random draws from independent $N(0, 1)$, and $a_{t\tau}$ are the entries of the lower triangular matrix Λ. It turns out that here $a_{t\tau} = (1 - \rho^2)^{-\frac{1}{2}}\rho^{t-\tau}$ if $t \geq \tau$ and $a_{t\tau} = 0$ if $t < \tau$.

Using the approach described above, we can obtain an unbiased, differentiable and positive simulator of $f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_\alpha, \rho)'$, in the Probit case by considering the following drawings:

η_{ih} is drawn from N(0,1).
η^*_{i1h} is drawn from N(0,1) restricted to

$[-(\boldsymbol{\beta}'\mathbf{x}_{i1} + \sigma_\alpha \eta_{ih})/a_{11}, \infty]$ if $y_{i1} = 1$ or $[-\infty, -(\boldsymbol{\beta}'\mathbf{x}_{i1} + \sigma_\alpha \eta_{ih})/a_{11}]$ if $y_{i1} = 0$, η^*_{i2h} is drawn from N(0,1) restricted to

$$[-(\boldsymbol{\beta}'\mathbf{x}_{i2} + \sigma_\alpha \eta_{ih} + a_{21}\eta^*_{i1h})/a_{22}, \infty] \text{ if } y_{i2} = 1,$$

and

$$[-\infty, -(\boldsymbol{\beta}'\mathbf{x}_{i2} + \sigma_\alpha \eta_{ih} + a_{21}\eta^*_{i1h})/a_{22}] \text{ if } y_{i2} = 0,$$

and so on. The simulator of $f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ is

$$\tilde{f}_H(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \frac{1}{H}\sum_{h=1}^{H}\prod_{t=1}^{T}\Phi\left[(-1)^{1-y_{it}}\left(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\alpha \eta_{ih} + \sum_{\tau=1}^{t-1} a_{t\tau}\eta^*_{i\tau h}\right)/a_{tt}\right], \tag{12.4.17}$$

where for $t = 1$, the sum over τ disappears.

In the Tobit case, the same kind of method can be used. The only difference is that the simulator of $f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ becomes

$$\tilde{f}_H(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}) = \frac{1}{H}\sum_{i=1}^{H}\left[\prod_{t\epsilon\Psi_1}\frac{1}{a_{tt}}\phi\left(\left[y_{it} - \left(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\alpha \eta_{ih} + \sum_{\tau=1}^{t-1} a_{t\tau}\eta^*_{i\tau h}\right)\right]/a_{tt}\right)\right]$$
$$\cdot \prod_{t\epsilon\Psi_0}\Phi\left[-\left(\boldsymbol{\beta}'\mathbf{x}_{it} + \sigma_\alpha \eta_{ih} + \sum_{\tau=1}^{t-1} a_{t\tau}\eta^*_{i\tau h}\right)/a_{tt}\right]. \tag{12.4.18}$$

The simulated maximum likelihood estimator (SMLE) is obtained from maximizing the simulated log-likelihood function. The simulated method of moments estimator (SMM) is obtained from the simulated moments. The simulated least squares estimator (SLS) is obtained if we let $\mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}) = E(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ and minimize $\sum_{i=1}^{N}[\mathbf{y}_i - E(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})]^2$.

Although we need $H \to \infty$ to obtain consistent simulator of $f(\mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ and $\mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta})$, it is shown by McFadden (1989) that when finite H vectors $(\boldsymbol{\eta}_{i1}, \ldots, \boldsymbol{\eta}_{iH})$ are drawn by simple random sampling and independently for different i from the marginal density $P(\boldsymbol{\eta})$, the simulation errors are independent across observations; hence the variance introduced by simulation will be controlled by the law of large numbers operating across observations, making it unnecessary to consistently estimate each theoretical $\mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta})$ for the consistency of SMM, $\hat{\boldsymbol{\theta}}_{\text{SGMM}}$, as $N \to \infty$.

The asymptotic covariance matrix of $\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{SMM}} - \boldsymbol{\theta})$ obtained from minimizing $[\hat{\mathbf{m}}(\boldsymbol{\theta}) - \mathbf{m}(\boldsymbol{\theta})]'A[\hat{\mathbf{m}}(\boldsymbol{\theta}) - \mathbf{m}(\boldsymbol{\theta})]$ where A is a positive definite matrix such as moments of the form (4.3.38) can be approximated by

$$(R'AR)^{-1}R'AG_{NH}AR(R'AR)^{-1}, \tag{12.4.19}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

where

$$R = \frac{1}{N}\sum_{i=1}^{N} W_i' \frac{\partial \tilde{\mathbf{m}}_H(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'},$$

$$G_{NH} = \frac{1}{N}\sum_{i=1}^{N} W_i \left(\Omega + \frac{1}{H}\Delta_H\right) W_i', \tag{12.4.20}$$

$$\Omega = \text{Cov}\,(\mathbf{m}_i(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}))$$

$$\Delta_H = \text{Cov}\,[\tilde{\mathbf{m}}_H(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}) - \mathbf{m}(y_i, \mathbf{x}_i; \boldsymbol{\theta})],$$

and W_i is given by (4.3.41). When $A = [\text{plim Cov}\,(\hat{\mathbf{m}}_i(y_i, \mathbf{x}_i; \boldsymbol{\theta})]^{-1}$, the SMM is the simulated generalized method of moments estimator (SGMM). It is clear that as $H \to \infty$, the SGMM has the same asymptotic efficiency as the GMM. However, even with finite H, the relative efficiency of SGMM is quite high. For instance, for the simple frequency simulator, $\Delta_H = \Omega$, one draw per observation gives fifty percent of the asymptotic efficiency of the corresponding GMM estimator, and nine draws per observation gives 90 percent relative efficiency.

The consistency of SMLE or SLS needs consistently estimated conditional density or moments. With a finite H, the approximation error of the conditional density or moments is of order H^{-1}. This will lead to the asymptotic bias of $O(1/H)$ (e.g., Gourieroux and Monfort 1996; Hsiao, Wang, and Wang 1997). Nevertheless, with a finite H it is still possible to propose SLS estimator that is consistent and asymptotically normally distributed as $N \to \infty$ by noting that for the sequence of 2H random draws $(\boldsymbol{\eta}_{i1}, \ldots, \boldsymbol{\eta}_{iH}, \boldsymbol{\eta}_{i,H+1}, \ldots, \boldsymbol{\eta}_{i,2H})$ for each i,

$$E\left[\frac{1}{H}\sum_{h=1}^{H} \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{ih}; \boldsymbol{\theta})\right] = E\left[\frac{1}{H}\sum_{h=1}^{H} \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{i,H+h}; \boldsymbol{\theta})\right] = \mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}), \tag{12.4.21}$$

and

$$E\left[\mathbf{y}_i - \frac{1}{H}\sum_{h=1}^{H} \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{ih}; \boldsymbol{\theta})\right]'\left[\mathbf{y}_i - \frac{1}{H}\sum_{h=1}^{H} \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{i,H+h}; \boldsymbol{\theta})\right] = E\left[\mathbf{y}_i - \mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta})\right]'\left[\mathbf{y}_i - \mathbf{m}(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta})\right], \tag{12.4.22}$$

because of the independence between $(\boldsymbol{\eta}_{i1}, \ldots, \boldsymbol{\eta}_{iH})$ and $(\boldsymbol{\eta}_{i,H+1}, \ldots, \boldsymbol{\eta}_{i,2H})$. Then the SLS estimator that minimizes

$$\sum_{i=1}^{N}\left[\mathbf{y}_i - \frac{1}{H}\sum_{h=1}^{H} \mathbf{m}^*(\mathbf{y}_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{ih}; \boldsymbol{\theta})\right]'\left[\mathbf{y}_i - \frac{1}{H}\sum_{h=1}^{H} \mathbf{m}^*(y_i, \mathbf{x}_i \mid \boldsymbol{\eta}_{i,H+h}; \boldsymbol{\theta})\right] \tag{12.4.23}$$

is consistent as $N \to \infty$ even H is fixed (e.g., Gourieroux and Monfort 1996; Hsiao and Wang 2000).

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

12.5 DATA WITH MULTILEVEL STRUCTURES

We have illustrated panel data methodology by assuming the presence of individual and/or time effects only. However, panel data need not be restricted to two dimensions. We can have a more complicated "clustering" or "hierarchical" structure. For example, Antweiler (2001), Baltagi, Song, and Jung (2001), and Davis (2002), following the methodology developed by Wansbeek (1982) Wansbeek and Kapteyn (1978), consider the multiway error components model of the form

$$y_{ij\ell t} = \mathbf{x}'_{ij\ell t}\boldsymbol{\beta} + v_{ij\ell t}, \tag{12.5.1}$$

for $i = 1, \ldots, N$, $j = 1, \ldots, M_i$, $\ell = 1, \ldots, L_{ij}$, and $t = 1, \ldots, T_{ij\ell}$. For example, the dependent variable $y_{ij\ell t}$ could denote the air pollution measured at station ℓ in city j of country i in time period t. This means that there are N countries, and each country i has M_i cities in which L_{ij} observation stations are located. At each station, air pollution is observed for $T_{ij\ell}$ periods. The $\mathbf{x}_{ij\ell t}$ denotes a vector of K explanatory variables, and the disturbance is assumed to have a multiway error components structure,

$$v_{ij\ell t} = \alpha_i + \lambda_{ij} + \nu_{ij\ell} + \epsilon_{ij\ell t} \tag{12.5.2}$$

where α_i, λ_{ij}, $\nu_{ij\ell}$ and $\epsilon_{ij\ell t}$ are assumed to be independently, identically distributed and are mutually independent with mean 0 and variances σ_α^2, σ_λ^2, σ_ν^2, and σ_ϵ^2, respectively.

In the case that the data are balanced, the variance–covariance matrix of $\mathbf{v}$, has the form

$$\Omega = \sigma_\alpha^2(I_N \otimes J_{MLT}) + \sigma_\lambda^2(I_{NM} \otimes J_{LT}) + \sigma_\nu^2(I_{NML} \otimes J_T) + \sigma_\epsilon^2 I_{LMNT}, \tag{12.5.3}$$

where J_s be a square matrix of dimension s with all elements equal to 1. Rewrite (12.5.3) in the form representing the spectral decomposition Ω (e.g., as in Appendix 3B), we have

$$\begin{aligned}\Omega &= MLT\sigma_\alpha^2(I_N \otimes P_{MLT}) + LT\sigma_\lambda^2(I_{NM} \otimes P_{LT}) \\ &\quad + T\sigma_\nu^2(I_{NML} \otimes P_T) + \sigma_\epsilon^2 I_{LMNT} \\ &= \sigma_\epsilon^2(I_{NML} \otimes Q_T) + \sigma_1^2(I_{NM} \otimes Q_L \otimes P_T) \\ &\quad + \sigma_2^2(I_N \otimes Q_M \otimes P_{LT}) + \sigma_3^2(I_N \otimes P_{MLT})\end{aligned} \tag{12.5.4}$$

where $P_s \equiv \frac{1}{s}J_s$, $Q_s = I_s - P_s$, and

$$\sigma_1^2 = T\sigma_\nu^2 + \sigma_\epsilon^2, \tag{12.5.5}$$

$$\sigma_2^2 = LT\sigma_\lambda^2 + T\sigma_\nu^2 + \sigma_\epsilon^2, \tag{12.5.6}$$

$$\sigma_3^2 = MLT\sigma_\alpha^2 + LT\sigma_\lambda^2 + T\sigma_\nu^2 + \sigma_\epsilon^2, \tag{12.5.7}$$

σ_ϵ^2 are the characteristic roots of Ω. As each of the terms of (12.5.4) is orthogonal to each other and sum to I_{NMLT}, it follows that

$$\begin{aligned}\Omega^{-1/2} &= \sigma_\epsilon^{-1}(I_{NML} \otimes Q_T) + \sigma_1^{-1}(I_{NM} \otimes Q_L \otimes P_T) \\ &\quad + \sigma_2^{-1}(I_N \otimes Q_M \otimes P_{LT}) + \sigma_3^{-1}(I_N \otimes P_{MLT}) \end{aligned} \tag{12.5.8}$$

Expanding all the Q matrices as the difference of I and P, multiplying both sides of the equation by σ_ϵ, and collecting terms yield

$$\begin{aligned}\sigma_\epsilon \Omega^{-1/2} &= I_{NMLT} - \left(1 - \frac{\sigma_\epsilon}{\sigma_1}\right)(I_{\text{NML}} \otimes P_T) \\ &\quad - \left(\frac{\sigma_\epsilon}{\sigma_1} - \frac{\sigma_\epsilon}{\sigma_2}\right)(I_{NM} \otimes P_{\text{LT}}) \\ &\quad - \left(\frac{\sigma_\epsilon}{\sigma_2} - \frac{\sigma_\epsilon}{\sigma_3}\right)(I_N \otimes P_{\text{MLT}}). \end{aligned} \tag{12.5.9}$$

The generalized least-squares estimator (GLS) of (12.5.1) is equivalent to the least squares estimator of

$$y_{ij\ell t}^* = y_{ij\ell t} - \left(1 - \frac{\sigma_\epsilon}{\sigma_1}\right)\bar{y}_{ij\ell .} - \left(\frac{\sigma_\epsilon}{\sigma_1} - \frac{\sigma_\epsilon}{\sigma_2}\right)\bar{y}_{ij\cdot\cdot} - \left(\frac{\sigma_\epsilon}{\sigma_2} - \frac{\sigma_\epsilon}{\sigma_3}\right)\bar{y}_{i\ldots}, \tag{12.5.10}$$

on

$$\mathbf{x}_{ij\ell t}^* = x_{ij\ell t} - \left(1 - \frac{\sigma_\epsilon}{\sigma_1}\right)\bar{\mathbf{x}}_{ij\ell .} - \left(\frac{\sigma_\epsilon}{\sigma_1} - \frac{\sigma_\epsilon}{\sigma_2}\right)\bar{\mathbf{x}}_{ij\cdot\cdot} - \left(\frac{\sigma_\epsilon}{\sigma_2} - \frac{\sigma_\epsilon}{\sigma_3}\right)\bar{\mathbf{x}}_{i\ldots}, \tag{12.5.11}$$

where $\bar{y}_{ij\ell .}(\bar{\mathbf{x}}_{ij\ell .})$, $\bar{y}_{ij..}(\bar{\mathbf{x}}_{ij..})$ and $\bar{y}_{i\ldots}(\bar{\mathbf{x}}_{i\ldots})$ indicate group averages. The application of feasible GLS can be carried out by replacing the variances in (12.5.10) and (12.5.11) by their estimates obtained from the three groupwise between estimates and the within estimate of the innermost group.

The pattern exhibited in (12.5.10) and (12.5.11) is suggestive of solutions for higher order hierarchy with a balanced structure. If the hierarchical structure is unbalanced, Kronecker product operation can no longer be applied. It introduces quite a bit of notational inconvenience into the algebra (e.g., Baltagi (1995, Chapter 9) and Wansbeek and Kapteyn (1978)). Neither can the GLS estimator be molded into a simple transformation for least-squares estimator. However, an unbalanced panel is made up of N top level groups, each containing M_i second-level groups, the second-level groups containing the innermost L_{ij} subgroups, which in turn containing $T_{ij\ell}$ observations, the number of observations in the higher-level groups are thus $T_{ij} = \sum_{\ell=1}^{L_{ij}} T_{ij\ell}$ and $T_i = \sum_{j=1}^{M_i} T_{ij}$, and the total number of observations is $H = \sum_{i=1}^{N} T_i$. The number of top-level groups is N, the number of second level groups is $F = \sum_{i=1}^{N} M_i$, and the bottom-level groups is $G = \sum_{i=1}^{N}\sum_{j=1}^{M_i} L_{ij}$. We can redefine J matrices to be block diagonal of size $H \times H$, corresponding in structure to the groups or subgroups they represent. They can be constructed explicitly by using "group membership" matrices consisting of 1's and 0's that uniquely assign each of the H observations to one of the G (or F or N) groups. Antweiler (2001) has derived the maximum-likelihood estimator for the panels with unbalanced hierarchy.

When data contains a multilevel hierarchical structure, the application of a simple error component estimation, although inefficient, remains consistent under the assumption that the error component is independent of the regressors. However, the estimated standard errors of the slope coefficients are usually biased downward.

12.6 ERRORS OF MEASUREMENT

Thus far we have assumed that variables are observed without errors. Economic quantities, however, are frequently measured with errors, particularly if longitudinal information is collected through one-time retrospective surveys, which are notoriously susceptible to recall errors. If variables are indeed subject to measurement errors, exploiting panel data to control for the effects of unobserved individual characteristics using standard differenced estimators (deviations from means, etc.) may result in even more biased estimates than simple least-squares estimators using cross-sectional data alone.

Consider, for example, the following single-equation model (Solon 1985):

$$y_{it} = \alpha_i^* + \beta x_{it} + u_{it}, \quad \begin{matrix} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{matrix} \tag{12.6.1}$$

where u_{it} is independently identically distributed, with mean 0 and variance σ_u^2, and $\text{Cov}(x_{it}, u_{is}) = \text{Cov}(\alpha_i^*, u_{it}) = 0$ for any t and s, but $\text{Cov}(x_{it}, \alpha_i^*) \neq 0$. Suppose further that we observe not x_{it} itself, but rather the error-ridden measure

$$x_{it}^* = x_{it} + \tau_{it}, \tag{12.6.2}$$

where $\text{Cov}(x_{is}, \tau_{it}) = \text{Cov}(\alpha_i^*, \tau_{it}) = \text{Cov}(u_{it}, \tau_{is}) = 0$, and $\text{Var}(\tau_{it}) = \sigma_\tau^2$, $\text{Cov}(\tau_{it}, \tau_{i,t-1}) = \gamma_\tau \sigma_\tau^2$.

If we estimate (12.6.1) by ordinary least-squares (OLS) with cross-sectional data for period t, the estimator converges to (as $N \to \infty$)

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}_{LS} = \beta + \frac{\text{Cov}(x_{it}, \alpha_i^*)}{\sigma_x^2 + \sigma_\tau^2} - \frac{\beta \sigma_\tau^2}{\sigma_x^2 + \sigma_\tau^2}, \tag{12.6.3}$$

where $\sigma_x^2 = \text{Var}(x_{it})$. The inconsistency of the least-squares estimator involves two terms, the first due to the failure to control for the individual effects α_i^* and the second due to measurement error.

If we have panel data, say $T = 2$, we can alternatively first difference the data to eliminate the individual effects, α_i^*,

$$y_{it} - y_{i,t-1} = \beta(x_{it}^* - x_{i,t-1}^*) + [(u_{it} - \beta\tau_{it}) - (u_{i,t-1} - \beta\tau_{i,t-1})], \tag{12.6.4}$$

and then apply least squares. The probability limit of the differenced estimator as $N \to \infty$ becomes

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} \hat{\beta}_d &= \beta\left[1 - \frac{2(1-\gamma_\tau)\sigma_\tau^2}{\operatorname{Var}(x_{it}^* - x_{i,t-1}^*)}\right] \\ &= \beta - \frac{\beta\sigma_\tau^2}{[(1-\gamma_x)/(1-\gamma_\tau)]\sigma_x^2 + \sigma_\tau^2}, \end{aligned} \tag{12.6.5}$$

where γ_x is the first-order serial-correlation coefficient of x_{it}. The estimator $\hat{\beta}_d$ eliminates the first source of inconsistency, but may aggravate the second. If $\gamma_x > \gamma_\tau$, the inconsistency due to measurement error is larger for $\hat{\beta}_d$ than for $\hat{\beta}_{LS}$. This occurs because if the serial correlation of the measurement error is less than that of the true x (as seems often likely to be the case), first differencing increases the noise-to-signal ratio for the measured explanatory variable.

The standard treatment for the errors-in-variables models requires extraneous information in the form of either additional data (replication and/or instrumental variables) or additional assumptions to identify the parameters of interest (e.g., Aigner et al. (1984)). The repeated measurement property of panel data allows a researcher to use different transformations of the data to induce different and deducible changes in the biases in the estimated parameters that can then be used to identify the importance of measurement errors and recover the "true" parameters (Ashenfelter, Deaton, and Solon (1984); Griliches and Hausman (1986). For instance, if the measurement error, τ_{it}, is independently identically distributed across i and t and x is serially correlated, then in the foregoing example we can use $x_{i,t-2}^*$ or $(x_{i,t-2}^* - x_{i,t-3}^*)$ as instruments for $(x_{it}^* - x_{i,t-1}^*)$ as long as $T > 3$. Thus, even though T may be finite, the resulting IV estimator is consistent when N tends to infinity.

Alternatively, we can obtain consistent estimates through a comparison of magnitudes of the bias arrived at by subjecting a model to different transformations (Griliches and Hausman 1986). For instance, if we use a covariance transformation to eliminate the contributions of unobserved individual components, we have

$$(y_{it} - \bar{y}_i) = \beta(x_{it}^* - \bar{x}_i^*) + [(u_{it} - \bar{u}_i) - \beta(\tau_{it} - \bar{\tau}_i)], \tag{12.6.6}$$

where $\bar{y}_i, \bar{x}_i^*, \bar{u}_i$, and $\bar{\tau}_i$ are individual time means of respective variables. Under the assumption that the measurement errors are independently identically distributed, the LS regression of (12.6.6) converges to

$$\operatorname*{plim}_{N\to\infty} \beta_w = \beta\left[1 - \frac{T-1}{T}\frac{\sigma_\tau^2}{\operatorname{Var}(x_{it}^* - \bar{x}_i^*)}\right]. \tag{12.6.7}$$

Then consistent estimators of β and σ_τ^2 can be solved from (12.6.5) and (12.6.7),

$$\hat{\beta} = \left[\frac{2\hat{\beta}_w}{\text{Var}\,(x_{it}^* - x_{i,t-1}^*)} - \frac{(T-1)\hat{\beta}_d}{T\ \text{Var}(x_{it}^* - \bar{x}_i^*)}\right] \tag{12.6.8}$$

$$\cdot \left[\frac{2}{\text{Var}(x_{it}^* - x_{i,t-1}^*)} - \frac{T-1}{T\ \text{Var}(x_{it}^* - \bar{x}_i^*)}\right]^{-1},$$

$$\hat{\sigma}_\tau^2 = \frac{\hat{\beta} - \hat{\beta}_d}{\hat{\beta}} \cdot \frac{\text{Var}(x_{it}^* - x_{i,t-1}^*)}{2}. \tag{12.6.9}$$

In general, if the measurement errors are known to possess certain structures, consistent estimators may be available from a method of moments and/or from an IV approach by utilizing the panel structure of the data. Moreover, the first difference and the within estimators are not the only ones that will give us an implicit estimate of the bias. In fact, there are $T/2$ such independent estimates. For a six-period cross section with τ_{it} independently identically distributed, we can compute estimates of β and σ_τ^2 from $y_6 - y_1$, $y_5 - y_2$, and $y_4 - y_3$ using the relationships

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \hat{\beta}_{61} &= \beta - 2\beta\sigma_\tau^2/\text{Var}(x_{i6}^* - x_{i1}^*),\\
\operatorname*{plim}_{N\to\infty} \hat{\beta}_{52} &= \beta - 2\beta\sigma_\tau^2/\text{Var}(x_{i5}^* - x_{i2}^*),\\
\operatorname*{plim}_{N\to\infty} \hat{\beta}_{43} &= \beta - 2\beta\sigma_\tau^2/\text{Var}(x_{i4}^* - x_{i3}^*).
\end{aligned} \tag{12.6.10}$$

Thus, there are alternative consistent estimators. This fact can be exploited to test the assumption with regard to measurement errors,which provide the rationale for the validity of the instruments, by comparing whether or not the alternative estimates of β are mutually coherent (e.g., Griliches and Hausman 1986). The moment conditions (12.6.5), (12.6.7), and (12.6.10) can also be combined together to obtain efficient estimates of β and σ_τ^2 by the use of Chamberlain π method (Chapter 3, Section 3.8) or generalized method of moments estimator.

For instance, transforming $\mathbf{y}$ and $\mathbf{x}$ by the transformation matrix P_s such that $P_s\mathbf{e}_T = \mathbf{0}$ eliminates the individual effects from the model (12.6.1). Regressing the transformed $\mathbf{y}$ on transformed $\mathbf{x}$ yields estimator that is a function of $\beta, \sigma_x^2, \sigma_\tau$ and the serial correlations of x and τ. Wansbeek and Koning (1989) have provided a general formula for the estimates based on various transformation of the data by letting

$$Y^* = \mathbf{e}_{NT}\mu + X^*\boldsymbol{\beta} + \mathbf{v}^* \tag{12.6.11}$$

where $Y^* = (\mathbf{y}_1^{*'}, \ldots, \mathbf{y}_T^{*'})'$, $\mathbf{y}_t^* = (y_{1t}, \ldots, y_{Nt})'$, $X^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*)'$, $\mathbf{x}_t^* = (\mathbf{x}_{1t}', \ldots, \mathbf{x}_{Nt}')$, $\mathbf{v}^* = (\mathbf{v}_1^{*'}, \ldots, \mathbf{v}_T^{*'})'$, and $\mathbf{v}_t^* = (v_{1t}, \ldots, v_{NT})'$. Then

$$\begin{aligned}\hat{\mathbf{b}}_s &= [X^{*'}(Q_s \otimes I_N)X^*]^{-1}[X^{*'}(Q_s \otimes I_N)Y^*] \\ &= \boldsymbol{\beta} + [X^{*'}(Q_s \otimes I_N)X^*]^{-1}[X^{*'}(Q_s \otimes I_N)(\mathbf{u}^* - \boldsymbol{\tau}^*\boldsymbol{\beta})],\end{aligned} \tag{12.6.12}$$

where $Q_s = P_s' P_s$, $\mathbf{u}^* = (\mathbf{u}_1^{*'}, \ldots, \mathbf{u}_T^{*'})'$, $\mathbf{u}_t^* = (u_{1t}, \ldots, u_{Nt})'$, $\boldsymbol{\tau}^* = (\boldsymbol{\tau}_1^*, \ldots, \boldsymbol{\tau}_T^*)'$, and $\boldsymbol{\tau}_t^* = (\boldsymbol{\tau}_{1t}, \ldots, \boldsymbol{\tau}_{Nt})$. In the case of $K = 1$ and measurement errors are serially uncorrelated, Wansbeek and Koning (1989) show that the m different transformed estimators $\mathbf{b} = (b_1, \ldots, b_m)'$

$$\sqrt{N}(\mathbf{b} - \beta(\mathbf{e}_m - \sigma_\tau^2 \boldsymbol{\phi})] \sim N(\mathbf{0}, V), \tag{12.6.13}$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)'$, $\phi_s = (tr\, Q_s / tr\, Q_s \Sigma_{x^*})$,

$$\Sigma_{x^*} = \text{Cov}\,(\mathbf{x}_i^*),\ x_i^* = (\mathbf{x}_{i1}^* \ldots, \mathbf{x}_{iT}^*)',$$

$$V = F'\{\sigma_u^2 \Sigma_{x^*} \otimes I_T + \beta^2 \sigma_\tau^2 (\Sigma_{x^*} + \sigma_\tau^2 I_T) \otimes I_T\} F,$$

and F is the $T^2 \times m$ matrix with the sth column $\mathbf{f}_s = \text{vec}\,(Q_s)/(tr\, Q_s \Sigma_{x^*})$, where vec (A) denotes the operation of transforming an $m \times n$ matrix A into the $mn \times 1$ vector by stacking the columns of A one underneath the other (Magnus and Neudecker (1999, p. 30). Then one can obtain an efficient estimator by minimizing

$$[\mathbf{b} - \beta(\mathbf{e}_m - \sigma_\tau^2 \boldsymbol{\phi})]' V^{-1} [\mathbf{b} - \beta(\mathbf{e}_m - \sigma_\tau^2 \boldsymbol{\phi})], \tag{12.6.14}$$

with respect to β and σ_τ^2, which yields

$$\hat{\beta} = \left\{ \frac{\boldsymbol{\phi}' V^{-1} \mathbf{b}}{\boldsymbol{\phi}' V^{-1} \boldsymbol{\phi}} - \frac{\mathbf{e}_m' V^{-1} \mathbf{b}}{\mathbf{e}_m' V^{-1} \boldsymbol{\phi}} \right\} \bigg/ \left\{ \frac{\boldsymbol{\phi}' V^{-1} \mathbf{e}}{\boldsymbol{\phi}' V^{-1} \boldsymbol{\phi}} - \frac{\mathbf{e}_m' V^{-1} \mathbf{e}_m}{\mathbf{e}_m' V^{-1} \boldsymbol{\phi}} \right\} \tag{12.6.15}$$

and

$$\sigma_\tau^2 = \left\{ \frac{\boldsymbol{\phi}' V^{-1} \mathbf{e}_m}{\boldsymbol{\phi}' V^{-1} \mathbf{b}} - \frac{\mathbf{e}_m' V^{-1} \mathbf{e}_m}{\mathbf{e}_m' V^{-1} \mathbf{b}} \right\} \bigg/ \left\{ \frac{\boldsymbol{\phi}' V^{-1} \boldsymbol{\phi}}{\boldsymbol{\phi}' V^{-1} \mathbf{b}} - \frac{\mathbf{e}_m' V^{-1} \boldsymbol{\phi}}{\mathbf{e}_m' V^{-1} \mathbf{b}} \right\}. \tag{12.6.16}$$

Extensions of this simple model to the serially correlated measurement errors are given by Biørn (2000), Hsiao and Taylor (1991). Wansbeek and Kapteyn (1978) consider simple estimators for dynamic panel data models with measurement errors. In the case of only one regressor for a linear panel data model, Wansbeek (2001) has provided a neat framework to derive the moment conditions under a variety of measurement errors assumption by stacking the matrix of covariances between the vector of dependent variables over time and the regressors, then projecting out nuisance parameters. To illustrate the basic idea, consider a linear model,

$$y_{it} = \alpha_i^* + \beta x_{it} + \boldsymbol{\gamma}' \mathbf{w}_{it} + \mathbf{u}_{it} \quad \begin{array}{l} i = 1, \ldots, N, \\ t = 1, \ldots, T, \end{array} \tag{12.6.17}$$

https://doi.org/10.1017/CBO9781139839327.015 Published online by Cambridge University Press

where x_{it} is not observed. Instead one observes x_{it}^*, which is related to x_{it} by (12.6.2). Suppose that the $T \times 1$ measurement error vector $\boldsymbol{\tau}_i = (\tau_{i1}, \ldots, \tau_{iT})'$ is i.i.d. with mean zero and covariance matrix $\Omega = E(\boldsymbol{\tau}_i \boldsymbol{\tau}_i')$.

Suppose Ω has a structure of the form

$$\text{vec}\ \Omega = R_0 \boldsymbol{\lambda}, \tag{12.6.18}$$

where vec denotes the operation that stacks the rows of a matrix one after another in a column vector form, R is a matrix of order $T^2 \times m$ with known elements, and $\boldsymbol{\lambda}$ is an $m \times 1$ vector of unknown constants. Using the covariance transformation matrix $Q = I_T - \frac{1}{T}\mathbf{e}_T \mathbf{e}_T'$ to eliminate the individual effects, α_i^*, yields

$$Q\mathbf{y}_i = Q\mathbf{x}_i + QW_i\boldsymbol{\gamma} + Q\mathbf{u}_i, \tag{12.6.19}$$

$$Q\mathbf{x}_i^* = Q\mathbf{x}_i + Q\boldsymbol{\tau}_i, \tag{12.6.20}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{iT})'$, $W_i = (\mathbf{w}_{it}')$. Let

$$R \equiv (I_T \otimes Q)R_0. \tag{12.6.21}$$

From (12.6.2), we have

$$\begin{aligned} E(\boldsymbol{\tau}_i \otimes Q\boldsymbol{\tau}_i) &= (I_T \otimes Q)E(\boldsymbol{\tau}_i \otimes \boldsymbol{\tau}_i) \\ &= (I_T \otimes Q)R_0\boldsymbol{\lambda} \\ &= R\boldsymbol{\lambda}. \end{aligned} \tag{12.6.22}$$

It follows that

$$\begin{aligned} E(\mathbf{x}_i^* \otimes Q\mathbf{x}_i) &= E(\mathbf{x}_i^* \otimes Q\mathbf{x}_i^*) - E[(\mathbf{x}_i + \boldsymbol{\tau}_i) \otimes Q\boldsymbol{\tau}_i] \\ &= E(\mathbf{x}_i^* \otimes Q\mathbf{x}_i^*) - R\boldsymbol{\lambda}. \end{aligned} \tag{12.6.23}$$

Therefore

$$E(\mathbf{x}_i^* \otimes Q\mathbf{y}_i) = E(\mathbf{x}_i^* \otimes Q\mathbf{x}_i^*)\beta + E(\mathbf{x}_i^* \otimes QW_i)\boldsymbol{\gamma} - \mathbf{R}\boldsymbol{\lambda}\boldsymbol{\beta}. \tag{12.6.24}$$

Equation (12.6.24) contains the nuisance parameter $\boldsymbol{\lambda}$. To eliminate $\boldsymbol{\lambda}$ from (12.6.24), multiplying $M_R = I_{T^2} - R(R'R)^{-1}R'$ to both sides of (12.6.24), we have the orthogonality conditions:

$$M_R E\{\mathbf{x}_i^* \otimes Q(\mathbf{y}_i - \mathbf{x}_i^*\beta - W_i\gamma)\} = \mathbf{0} \tag{12.6.25}$$

Combining (12.6.25) with the moment conditions $E(W_i'Q\mathbf{u}_i) = \mathbf{0}$, we have the moment conditions for the measurement error model (12.6.17)

$$E[M(\mathbf{d}_i - C_i\boldsymbol{\theta})] = \mathbf{0}, \tag{12.6.26}$$

where

$$M = \begin{bmatrix} M_R & \mathbf{0} \\ \mathbf{0} & I_K \end{bmatrix}, \mathbf{d}_i = \begin{bmatrix} \mathbf{x}_i^* \otimes I_T \\ W_i' \end{bmatrix} Q\mathbf{y}_i,$$

$$C_i = \begin{bmatrix} \mathbf{x}_i^* \otimes I_T \\ W_i' \end{bmatrix} Q[\mathbf{x}_i^*, W_i], \quad \boldsymbol{\theta}' = (\beta, \boldsymbol{\gamma}').$$

A GMM estimator is obtained by minimizing

$$\frac{1}{N}\left[\sum_{i=1}^{N} M(\mathbf{d}_i - C_i\boldsymbol{\theta})\right]' A_N \left[\sum_{i=1}^{N} M(\mathbf{d}_i - C_i\boldsymbol{\theta})\right]. \tag{12.6.27}$$

An optimal GMM estimator is to let

$$A_N^{-1} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{d}_i - C_i\hat{\boldsymbol{\theta}})(\mathbf{d}_i - \mathbf{C}_i\hat{\boldsymbol{\theta}})', \tag{12.6.28}$$

where $\hat{\boldsymbol{\theta}}$ is some consistent estimator of $\boldsymbol{\theta}$ such as

$$\hat{\boldsymbol{\theta}} = \left[\left(\sum_{i=1}^{N} C_i'\right) M \left(\sum_{i=1}^{N} C_i\right)\right]^{-1}\left[\left(\sum_{i=1}^{N} C_i\right)' M \left(\sum_{i=1}^{N} \mathbf{d}_i\right)\right]. \tag{12.6.29}$$

In the case when τ_{it} is i.i.d. across i and over t, Ω is diagonal with equal diagonal element. Then $m = 1$ and $R_0 = \text{vec } I_T$, $R = (I_T \otimes Q)$ vec $I_T = \text{vec} Q$, $R'R = tr\, Q = T - 1$, and $M_R = I_{T^2} - \frac{1}{T-1}(\text{vec } Q)(\text{vec } Q)'$. When Ω is diagonal with distinct diagonal elements, $m = T$ and $R_0 = \mathbf{i}_t\mathbf{i}_t' \otimes \mathbf{i}_t$, where $\mathbf{i}_t$ is the tth unit vector of order T. When τ_{it} is a first-order moving average process and $T = 4$,

$$\Omega = \begin{bmatrix} a & c & 0 & 0 \\ c & b & c & 0 \\ 0 & c & b & c \\ 0 & 0 & c & a \end{bmatrix},$$

then

$$R_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix},$$

and $\boldsymbol{\lambda} = (a, b, c)'$.

In general, the identifiability of the slope parameters $\boldsymbol{\beta}$ for a linear regression model depends on whether the moment equations involving observables in levels and differences for different order of lags are sufficient to obtain a unique solution for $\boldsymbol{\beta}$ given the assumption about the autocorrelation patterns of measurement errors. For additional references, see Biørn (2000), Biørn and Klette (1998), Biørn and Krishnakumar (2008), Wansbeek (2001), and Wansbeek and Meijer (2000, Chapter 6, Section 6.6).

The measurement errors for nonlinear models are much more difficult to handle (e.g., Hsiao 1992c). For binary choice models with measurement errors, see Kao and Schnell (1987a,b) and Hsiao (1991b).

12.7 NONPARAMETRIC PANEL DATA MODELS

Our discussion of panel data models have been confined to parametrically or semiparametrically specified models. The generalization to panel nonparametric models can be very complicated. However, in a static framework, the generalization to the nonparametric setup is fairly straightforward although the computation can be tedious. To see this, let

$$y_{it} = m(\mathbf{x}_{it}) + \nu_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{12.7.1}$$

$$\nu_{it} = \alpha_i + u_{it}, \tag{12.7.2}$$

where $\mathbf{x}_{it}$ denotes the $K \times 1$ strictly exogenous variables with respect to u_{it}, $E(u_{it} \mid \mathbf{x}_{is}) = 0$ for all t and s.

If α_i is treated as random and uncorrelated with $\mathbf{x}_{it}$, then $m(\mathbf{x}_{it})$ can be estimated by kernel method or the local linear least-squares method $(\min \sum_{i=1}^{N} \sum_{t=1}^{I} K(\frac{\mathbf{X}_{it}-\mathbf{X}^*}{\sigma_N})(y_{it} - m(\mathbf{x}_{it}))^2)$ in which

$$m(\mathbf{x}_{it}) = m(\mathbf{x}^*) + (\mathbf{x}_{it} - \mathbf{x}^*)'\boldsymbol{\beta}(\mathbf{x}^*) \tag{12.7.3}$$

for $\mathbf{x}_{it}$ close to $\mathbf{x}^*$, where the "closeness" is defined in terms of some kernel function, $\sigma_N^{-K} K\left(\frac{\mathbf{x}_{it}-\mathbf{x}^*}{\sigma_N}\right)$, with $K(\mathbf{v}) \geq 0$, $K(\mathbf{v}) \longrightarrow 0$ as $\mathbf{v} \longrightarrow \pm\infty$ and σ_N is a bandwidth parameter. Substituting (12.7.3) into (12.7.1), one can estimate $m(\mathbf{x}^*)$ and $\boldsymbol{\beta}(\mathbf{x}^*)$ by the least squares method (Li and Racine (2007, ch. 2)). However, the least squares method ignores the error components structure of v_{it}. Martins-Filho and Yao (2009), Su, Ullah, and Wang (2010), etc. have considered more efficient two-step estimators.

When α_i is treated as fixed constant, Kernel approach is not a convenient method to estimate $m(\mathbf{x}_{it})$ 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 $m(\mathbf{x}_{it})$ into the following general index format,

$$m(\mathbf{x}_{it}) = \nu_0(\mathbf{x}_{it}, \boldsymbol{\theta}^0) + \sum_{j=1}^{m} h_{j0}(\nu_j(\mathbf{x}_{it}, \boldsymbol{\theta}^0)), \tag{12.7.4}$$

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

However, to uniquely identify the parameters of interest of the index model (12.7.4) one needs to impose the following normalization conditions:

(1) $h_{j0}(0) = 0$ for $j = 1, \dots, m$.
(2) The scaling restriction, say $\boldsymbol{\theta}^{0'}\boldsymbol{\theta}^0 = 1$ or the first element of $\boldsymbol{\theta}^0$ be normalized to 1 if it is known different from 0.
(3) The exclusion restriction when $\nu_j(\mathbf{x}, \boldsymbol{\theta})$ and $\nu_s(\mathbf{x}, \boldsymbol{\theta})$ are homogeneous of degree 1 in the regressors for some $s \neq j$.

When (1) 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 (1). When (2) does not hold, it is not possible to distinguish $(\boldsymbol{\theta}^0, h_0(\cdot))$ from $(c\boldsymbol{\theta}^0, \tilde{h}_0(\cdot) = h_0(\cdot/c))$ for any nonzero constant c. When (3) does not hold, say $(h_{10}(\cdot), h_{20}(\cdot)$ containing a common element $\mathbf{x}_{3it}$, then (h_{10}, h_{20}) is not distinguishable from $(h_{10} + g(\mathbf{x}_{3it}), h_{20} - g(\mathbf{x}_{3it}))$ for any function g (For further details, see Ai and Li (2008)).

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

$$h_{j0}(\cdot) \simeq \mathbf{p}_j(\cdot)'\boldsymbol{\pi}_j \tag{12.7.5}$$

The simplest series base function is to use the power series. However, 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 tth order univariate B-spline base function is given by (see Chui (1992, Chapter 4).

$$B_r(x \mid t_0, \ldots, t_r) = \frac{1}{(r-1)!} \sum_{j=1}^{r} (-1)^j \binom{r}{j} [\max(0, x - t_j)]^{r-1}, \tag{12.7.6}$$

where $t_0, t_1, \ldots, t_r$ are the evenly spaced design knots on the support of X. When $r = 2$, (12.7.6) gives a piecewise linear spline, and when $r = 4$, it gives piecewise cubic splines (i.e., the third-order polynomials). Substituting the parametric specification (12.7.5) in lieu of $h_{j0}(\cdot)$ into (12.7.1), we obtain the parametric analog of (12.7.4). Then, just like in the parametric case, one can remove α_i by taking the deviation of y_{it} from the ith individual time series mean $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$. Therefore one can obtain consistent estimators of $\boldsymbol{\theta}^0$ and $\boldsymbol{\pi}_j$, $j = 1, \ldots, m$ by minimizing

$$\begin{aligned} &\sum_{i=1}^{N}\sum_{t=1}^{T}\{(y_{it} - \nu_0(\mathbf{x}_{it}, \boldsymbol{\theta}) - \sum_{j=1}^{m} \mathbf{p}_j(\nu_j(\mathbf{x}_{it}, \boldsymbol{\theta}))'\boldsymbol{\pi}_j) \\ &\quad - \frac{1}{T}\sum_{t=1}^{T}(y_{is} - \nu_0(\mathbf{x}_{is}, \boldsymbol{\theta}) - \sum_{j=1}^{m} \mathbf{p}_j(\nu_j(\mathbf{x}_{is}, \boldsymbol{\theta})'\boldsymbol{\pi}_j))\}^2 \end{aligned} \tag{12.7.7}$$

Shen (1997), Newey (1997), and Chen 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 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 8,

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + m(\mathbf{z}_{it}) + \alpha_i + u_{it}, \tag{12.7.8}$$

where y_{it} is observed if the dummy variable $d_{it} = 1$. The sample selection effect $m(\mathbf{z}_{it})$ given $d_{it} = 1$ can be approximated

$$m(\mathbf{z}_{it}) \sim \sum_{j=1}^{m} h_j(\mathbf{z}_{jit}), \tag{12.7.9}$$

where $h_j(\cdot)$ are 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 $\mathbf{z}_{1it}, \dots, \mathbf{z}_{mit}$ are mutually exclusive. Then each $h_j(\cdot)$ can be approximated by the linear sieve $\mathbf{p}_j^{k_j}(\cdot)'\boldsymbol{\pi}_j$, where $\mathbf{p}_j^{k_j}(\cdot)$ is a vector of approximating functions satisfying $\mathbf{p}_j^{k_j}(0) = \mathbf{0}$. The unknown parameters $\boldsymbol{\beta}$ and the coefficients $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \dots, \boldsymbol{\pi}_m')'$ can be estimated by the generalized least-squares estimator if α_i are treated as random and uncorrelated with $(\mathbf{x}_{it}, \mathbf{z}_{it})$, or by minimizing

$$\sum_{i=1}^{N} \sum_{s<t} \left[(y_{it} - y_{is}) - (\mathbf{x}_{it} - \mathbf{x}_{is})'\boldsymbol{\beta} - \sum_{j=1}^{m} (\mathbf{p}_j^{k_j}(\mathbf{z}_{jit}) - \mathbf{p}_j^{k_j}(\mathbf{z}_{is}))'\boldsymbol{\pi}_j \right]^2 . \tag{12.7.10}$$

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

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 $\mathbf{x}_{it}$ excludes the inclusion of lagged dependent variables. Neither is this approach of replacing unknown $h_j(\cdot)$ by series expansion easily generalizable to censored or nonlinear panel data models [for further discussion, see Ai and Li (2008), and Su and Ullah (2011)].

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