vation is the family. Therefore, inference is done as the number of families in the sample tends to infinity.

The assumptions that we make about how the unobservables u_g are related to the explanatory variables $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_G)$ are crucial for determining which estimators of the $\boldsymbol{\beta}_g$ have acceptable properties. Often, when system (7.1) represents a structural model (without omitted variables, errors-in-variables, or simultaneity), we can assume that

$$\mathrm{E}(u_g \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_G) = 0, \qquad g = 1, \ldots, G \tag{7.2}$$

One important implication of assumption (7.2) is that u_g is uncorrelated with the explanatory variables in *all* equations, as well as all functions of these explanatory variables. When system (7.1) is a system of equations derived from economic theory, assumption (7.2) is often very natural. For example, in the set of demand functions that we have presented, $\mathbf{x}_g \equiv \mathbf{x}$ is the same for all g, and so assumption (7.2) is the same as $\mathrm{E}(u_g \mid \mathbf{x}_g) = \mathrm{E}(u_g \mid \mathbf{x}) = 0$.

If assumption (7.2) is maintained, and if the $\mathbf{x}_g$ are not the same across g, then any explanatory variables excluded from equation g are assumed to have no effect on expected y_g once $\mathbf{x}_g$ has been controlled for. That is,

$$\mathrm{E}(y_g \mid \mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_G) = \mathrm{E}(y_g \mid \mathbf{x}_g) = \mathbf{x}_g \boldsymbol{\beta}_g, \qquad g = 1, 2, \ldots, G \tag{7.3}$$

There are examples of SUR systems where assumption (7.3) is too strong, but standard SUR analysis either explicitly or implicitly makes this assumption.

Our next example involves panel data.

Example 7.2 (Panel Data Model): Suppose that for each cross section unit we observe data on the same set of variables for T time periods. Let $\mathbf{x}_t$ be a $1 \times K$ vector for $t = 1, 2, \ldots, T$, and let $\boldsymbol{\beta}$ be a $K \times 1$ vector. The model in the population is

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + u_t, \qquad t = 1, 2, \ldots, T \tag{7.4}$$

where y_t is a scalar. For example, a simple equation to explain annual family saving over a five-year span is

$$sav_t = \beta_0 + \beta_1 inc_t + \beta_2 age_t + \beta_3 educ_t + u_t, \qquad t = 1, 2, \ldots, 5$$

where inc_t is annual income, $educ_t$ is years of education of the household head, and age_t is age of the household head. This is an example of a **linear panel data model**. It is a **static model** because all explanatory variables are dated contemporaneously with sav_t.

The panel data setup is conceptually very different from the SUR example. In Example 7.1, each equation explains a different dependent variable for the same cross section unit. Here we only have one dependent variable we are trying to explain—*sav*—but we observe *sav*, and the explanatory variables, over a five-year period. (Therefore, the label "system of equations" is really a misnomer for panel data applications. At this point, we are using the phrase to denote more than one equation in any context.) As we will see in the next section, the statistical properties of estimators in SUR and panel data models can be analyzed within the same structure.

When we need to indicate that an equation is for a particular cross section unit i during a particular time period t, we write $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}$. We will omit the i subscript whenever its omission does not cause confusion.

What kinds of exogeneity assumptions do we use for panel data analysis? One possibility is to assume that u_t and $\mathbf{x}_t$ are orthogonal in the conditional mean sense:

$$\mathrm{E}(u_t \mid \mathbf{x}_t) = 0, \qquad t = 1, \ldots, T \tag{7.5}$$

We call this **contemporaneous exogeneity** of $\mathbf{x}_t$ because it only restricts the relationship between the disturbance and explanatory variables in the same time period. It is very important to distinguish assumption (7.5) from the stronger assumption

$$\mathrm{E}(u_t \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = 0, \qquad t = 1, \ldots, T \tag{7.6}$$

which, combined with model (7.4), is identical to $\mathrm{E}(y_t \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = \mathrm{E}(y_t \mid \mathbf{x}_t)$. Assumption (7.5) places no restrictions on the relationship between $\mathbf{x}_s$ and u_t for $s \neq t$, while assumption (7.6) implies that each u_t is uncorrelated with the explanatory variables in *all* time periods. When assumption (7.6) holds, we say that the explanatory variables $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T\}$ are **strictly exogenous**.

To illustrate the difference between assumptions (7.5) and (7.6), let $\mathbf{x}_t \equiv (1, y_{t-1})$. Then assumption (7.5) holds if $\mathrm{E}(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_0) = \beta_0 + \beta_1 y_{t-1}$, which imposes first-order dynamics in the conditional mean. However, assumption (7.6) *must* fail since $\mathbf{x}_{t+1} = (1, y_t)$, and therefore $\mathrm{E}(u_t \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = \mathrm{E}(u_t \mid y_0, y_1, \ldots, y_{T-1}) = u_t$ for $t = 1, 2, \ldots, T-1$ (because $u_t = y_t - \beta_0 - \beta_1 y_{t-1}$).

Assumption (7.6) can fail even if $\mathbf{x}_t$ does *not* contain a lagged dependent variable. Consider a model relating poverty rates to welfare spending per capita, at the city level. A **finite distributed lag (FDL) model** is

$$poverty_t = \theta_t + \delta_0 welfare_t + \delta_1 welfare_{t-1} + \delta_2 welfare_{t-2} + u_t \tag{7.7}$$

where we assume a two-year effect. The parameter θ_t simply denotes a different aggregate time effect in each year. It is reasonable to think that welfare spending reacts to lagged poverty rates. An equation that captures this feedback is

$$welfare_t = \eta_t + \rho_1 poverty_{t-1} + r_t \tag{7.8}$$

Even if equation (7.7) contains enough lags of welfare spending, assumption (7.6) would be violated if $\rho_1 \neq 0$ in equation (7.8) because $welfare_{t+1}$ depends on u_t and $\mathbf{x}_{t+1}$ includes $welfare_{t+1}$.

How we go about consistently estimating $\boldsymbol{\beta}$ depends crucially on whether we maintain assumption (7.5) or the stronger assumption (7.6). Assuming that the $\mathbf{x}_{it}$ are fixed in repeated samples is effectively the same as making assumption (7.6).

7.3 System OLS Estimation of a Multivariate Linear System

7.3.1 Preliminaries

We now analyze a general multivariate model that contains the examples in Section 7.2, and many others, as special cases. Assume that we have independent, identically distributed cross section observations $\{(\mathbf{X}_i, \mathbf{y}_i): i = 1, 2, \ldots, N\}$, where $\mathbf{X}_i$ is a $G \times K$ matrix and $\mathbf{y}_i$ is a $G \times 1$ vector. Thus, $\mathbf{y}_i$ contains the dependent variables for all G equations (or time periods, in the panel data case). The matrix $\mathbf{X}_i$ contains the explanatory variables appearing anywhere in the system. For notational clarity we include the i subscript for stating the general model and the assumptions.

The multivariate linear model for a random draw from the population can be expressed as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i \tag{7.9}$$

where $\boldsymbol{\beta}$ is the $K \times 1$ parameter vector of interest and $\mathbf{u}_i$ is a $G \times 1$ vector of unobservables. Equation (7.9) explains the G variables $y_{i1}, \ldots, y_{iG}$ in terms of $\mathbf{X}_i$ and the unobservables $\mathbf{u}_i$. Because of the random sampling assumption, we can state all assumptions in terms of a generic observation; in examples, we will often omit the i subscript.

Before stating any assumptions, we show how the two examples introduced in Section 7.2 fit into this framework.

Example 7.1 (SUR, continued): The SUR model (7.1) can be expressed as in equation (7.9) by defining $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iG})'$, $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iG})'$, and

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2} & & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & & \vdots \\ \vdots & & & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{iG} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_G \end{pmatrix} \tag{7.10}$$