CHAPTER 4

Simultaneous-Equations Models

4.1 INTRODUCTION

In Chapters 2 and 3, we discussed the approach of decomposing the effect of a large number of factors that affect the dependent variables, but are not explicitly included as explanatory variables, into effects specific to individual units, to time periods, and to both individual units and time periods as a means to take account of the heterogeneity in panel data in estimating single-equation models. However, the consistency or asymptotic efficiency of various estimators discussed in previous chapters depends on the validity of the single-equation model assumptions. If they are not true, this approach may solve one problem but aggravate other problems.

For instance, consider the income-schooling model,

$$y = \beta_0 + \beta_1 S + \beta_2 A + u, \tag{4.1.1}$$

where y is a measure of income, earnings, or wage rate, S is a measure of schooling, and A is an unmeasured ability variable that is assumed to be positively related to S. The coefficients β_1 and β_2 are assumed positive. Under the assumption that S and A are uncorrelated with u, the least squares estimate of β_1 that ignores A is biased upward. The standard left-out-variable formula gives the size of this bias as

$$E(\hat{\beta}_{1,\text{LS}}) = \beta_1 + \beta_2 \frac{\sigma_{AS}}{\sigma_S^2}, \tag{4.1.2}$$

where σ_S^2 is the variance of S, and σ_{AS} is the covariance between A and S.

If the omitted variable A is a purely "family" one,[1] that is, if siblings have exactly the same level of A, then estimating β_1 from within-family data (i.e., from differences between the brothers' earnings and differences between the brothers' education) will eliminate this bias. But if ability, apart from having a family component, also has an individual component, and this individual component is not independent of the schooling variable, the within-family estimates are not necessarily less biased.

Suppose

$$A_{it} = \alpha_i + \omega_{it}, \tag{4.1.3}$$

where i denotes the family, and t denotes members of the family. If ω_{it} is uncorrelated with S_{it}, the combination of (4.1.1) and (4.1.3) is basically of the same form as (2.3.3).

[1] Namely, the family effect A_i has the same meaning as α_i in Chapters 3 and 4.

The expected value of the within (or LSDV) estimator is unbiased. On the other hand, if the within-family covariance between A and $S, \sigma_{S\omega}$, is not equal to zero, the expected value of the within estimator is

$$E(\hat{\beta}_{1,w}) = \beta_1 + \beta_2 \frac{\sigma_{S\omega}}{\sigma^2_{S|w}}, \tag{4.1.4}$$

where $\sigma^2_{S|w}$ is the within-family variance of S. The estimator remains biased. Furthermore, if the reasons for the correlation between A and S are largely individual rather than familial, then going to within data will drastically reduce $\sigma^2_{S|w}$, with little change to σ_{AS} (or $\sigma_{S\omega}$), which would make this source of bias even more serious.

Moreover, if S is also a function of A and other social–economic variables, (4.1.1) is only one behavioral equation in a simultaneous-equations model. Then the probability limit of the least-squares estimate, $\hat{\beta}_{1,\mathrm{LS}}$, is no longer (4.1.2) but is of the form

$$\operatorname{plim} \hat{\beta}_{1,\mathrm{LS}} = \beta_1 + \beta_2 \frac{\sigma_{AS}}{\sigma^2_S} + \frac{\sigma_{uS}}{\sigma^2_S}, \tag{4.1.5}$$

where σ_{uS} is the covariance between u and S. If, as argued by Griliches (1977, 1979), schooling is the result, at least in part, of optimizing behavior by individuals and their family, σ_{uS} could be negative. This opens the possibility that the least squares estimates of the schooling coefficient may be biased downward rather than upward. Furthermore, if the reasons for σ_{uS} being negative are again largely individual rather than familial, and the within-family covariance between A and S reduces σ_{AS} by roughly the same proportion as $\sigma^2_{S|w}$ is to σ^2_S, there will be a significant decline in the $\hat{\beta}_{1,w}$ relative to $\hat{\beta}_{1,\mathrm{LS}}$. The size of this decline will be attributed to the importance of ability and "family background," but in fact it reflects nothing more than the simultaneity problems associated with the schooling variable itself. In short, the simultaneity problem could reverse the single-equation conclusions.

In this chapter we focus on the issues of correlations arising from the joint dependence of G endogenous variables $\mathbf{y}_{it} = \left(y_{1,it}, y_{2,it}, \ldots, y_{G,it}\right)'$ given K exogenous variables $\mathbf{x}_{it} = \left(x_{1,it}, x_{2,it}, \ldots, x_{K,it}\right)'$. Suppose the model is[2]

$$\Gamma \mathbf{y}_{it} + \mathbf{B}\mathbf{x}_{it} + \boldsymbol{\mu} = \mathbf{v}_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{4.1.6}$$

where Γ and $\mathbf{B}$ are $G \times G$ and $G \times K$ matrices of coefficients; $\boldsymbol{\mu}$ is the $G \times 1$ vector of intercepts. We assume that the $G \times 1$ errors $\mathbf{v}_{it}$ has a component structure,

$$\mathbf{v}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\lambda}_t + \mathbf{u}_{it}, \tag{4.1.7}$$

where $\boldsymbol{\alpha}_i$ and $\boldsymbol{\lambda}_t$ denote the $G \times 1$ individual varying but time-invariant and individual invariant but time-varying specific effects, respectively, and $\mathbf{u}_{it}$ denote the $G \times 1$ random vector that varies across i and over t and are uncorrelated with $\mathbf{x}_{it}$,

$$E\left(\mathbf{x}_{it}\mathbf{u}'_{js}\right) = \mathbf{0} \tag{4.1.8}$$

The issue of dynamic dependence will be discussed in Section 4.5.

[2] The speed of convergence of a fixed-effects linear simultaneous-equations model is the same as that of the single-equation fixed-effects linear static model (see Chapter 2). The MLE of $\boldsymbol{\alpha}_i$ is consistent only when T tends to infinity. The MLE of $\boldsymbol{\lambda}_t$ is consistent only when N tends to infinity. However, just as in the linear static model, the MLE of Γ and B do not depend on the MLE of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\lambda}_t$. They are consistent when either N or T or both tend to infinity (Schmidt 1984).

Model (4.1.6) could give rise to two sources of correlations between the regressors and the errors of the equations: (i) the potential correlations between the individual- and time-specific effects, α_i and λ_t, with $\mathbf{x}_{it}$ and (ii) the correlations between the joint dependent variables and the errors. The first source of correlations could be eliminated through some linear transformation of the original variables. For instance, the covariance transformation of $\mathbf{y}_{it}$ and $\mathbf{x}_{it}$,

$$\dot{\mathbf{y}}_{it} = \mathbf{y}_{it} - \bar{\mathbf{y}}_i - \bar{\mathbf{y}}_t + \bar{\mathbf{y}} \tag{4.1.9}$$

$$\dot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\mathbf{x}} \tag{4.1.10}$$

yields

$$\Gamma\dot{\mathbf{y}}_{it} + \mathbf{B}\dot{\mathbf{x}}_{it} = \dot{\mathbf{v}}_{it} \tag{4.1.11}$$

where $\dot{\mathbf{v}}_{it} = \mathbf{v}_{it} - \bar{\mathbf{v}}_i - \bar{\mathbf{v}}_t + \bar{\mathbf{v}}$ and $(\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i, \bar{\mathbf{v}}_i)$, $(\bar{\mathbf{y}}_t, \bar{\mathbf{x}}_t, \bar{\mathbf{v}}_t)$, $(\bar{\mathbf{y}}, \bar{\mathbf{x}}, \bar{\mathbf{v}})$ denote the individual time series mean, cross-sectional mean at t, and overall mean of respective variable, e.g. $\bar{\mathbf{y}}_i = \frac{1}{T}\sum_{t=1}^T \mathbf{y}_{it}$, $\bar{\mathbf{y}}_t = \frac{1}{N}\sum_{i=1}^N \mathbf{y}_{it}$, and $\bar{\mathbf{y}} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbf{y}_{it}$.

Under the assumption (4.1.8),

$$E\left(\dot{\mathbf{x}}_{it}\dot{\mathbf{v}}'_{js}\right) = \mathbf{0} \tag{4.1.12}$$

standard identification and estimation methods for a Cowles Commission structural equation model can be applied to model (4.1.11) to obtain consistent and asymptotically normally distributed estimators (e.g., Hood and Koopmans 1953; Hsiao 1983; Intriligator, Bodkin, and Hsiao 1996). However, exploitation of the component structure could lead to more efficient inference of (4.1.6) than those based on the two- or three-stage least squares methods.

We assume α_i, λ_t, and $\mathbf{u}_{it}$ are each $G \times 1$ random vectors that have zero means and are independent of one another, and

$$\begin{aligned}
&E\mathbf{x}_{it}\mathbf{v}'_{js} = \mathbf{0},\\
&E\alpha_i\alpha'_j = \begin{cases}\Omega_\alpha = \left(\sigma^2_{\alpha g\ell}\right) & \text{if } i = j,\\ \mathbf{0} & \text{if } i \neq j,\end{cases}\\
&E\lambda_t\lambda'_s = \begin{cases}\Omega_\lambda = \left(\sigma^2_{\lambda g\ell}\right) & \text{if } t = s,\\ \mathbf{0} & \text{if } t \neq s,\end{cases}\\
&E\mathbf{u}_{it}\mathbf{u}'_{js} = \begin{cases}\Omega_u = \left(\sigma^2_{ug\ell}\right) & \text{if } i = j, \quad \text{and} \quad t = s,\\ \mathbf{0} & \text{otherwise.}\end{cases}
\end{aligned} \tag{4.1.13}$$

Multiplying (4.1.6) by Γ^{-1}, we have the reduced form

$$\mathbf{y}_{it} = \mu^* + \Pi\mathbf{x}_{it} + \boldsymbol{\epsilon}_{it}, \tag{4.1.14}$$

where $\mu^* = -\Gamma^{-1}\boldsymbol{\mu}, \Pi = -\Gamma^{-1}\mathbf{B}$, and $\boldsymbol{\epsilon}_{it} = \Gamma^{-1}\mathbf{v}_{it}$. The reduced-form error term $\boldsymbol{\epsilon}_{it}$ again has an error-component structure[3]

$$\boldsymbol{\epsilon}_{it} = \boldsymbol{\alpha}^*_i + \boldsymbol{\lambda}^*_t + \mathbf{u}^*_{it}, \tag{4.1.15}$$

[3] Note that the meaning of these asterisks has been changed from what they were in previous chapters.

with

$$
\begin{aligned}
&E\boldsymbol{\alpha}_i^* = E\boldsymbol{\lambda}_t^* = E\mathbf{u}_{it}^* = \mathbf{0}, \qquad E\boldsymbol{\alpha}_i^*\boldsymbol{\lambda}_t^{*\prime} = E\boldsymbol{\alpha}_i^*\mathbf{u}_{it}^{*\prime} = E\boldsymbol{\lambda}_t^*\mathbf{u}_{it}^{*\prime} = \mathbf{0},\\
&E\boldsymbol{\alpha}_i^*\boldsymbol{\alpha}_j^{*\prime} = \begin{cases} \Omega_\alpha^* = \left(\sigma_{\alpha g\ell}^{*2}\right) & \text{if } i = j,\\ \mathbf{0} & \text{if } i \neq j,\end{cases}\\
&E\boldsymbol{\lambda}_t^*\boldsymbol{\lambda}_s^{*\prime} = \begin{cases} \Omega_\lambda^* = \left(\sigma_{\lambda g\ell}^{*2}\right) & \text{if } t = s,\\ \mathbf{0} & \text{if } t \neq s,\end{cases}\\
&E\mathbf{u}_{it}^*\mathbf{u}_{js}^{*\prime} = \begin{cases} \Omega_u^* = \left(\sigma_{ug\ell}^{*2}\right) & \text{if } i = j \quad \text{and} \quad t = s,\\ \mathbf{0} & \text{otherwise.}\end{cases}
\end{aligned}
\tag{4.1.16}
$$

If the $G \times G$ covariance matrices $\Omega_\alpha, \Omega_\lambda$, and Ω_u are unrestricted, there are no restrictions on the variance–covariance matrix. The usual order and rank conditions are the necessary and sufficient conditions for identifying a particular equation in the system (e.g., Hsiao 1983). If there are restrictions on $\Omega_\alpha, \Omega_\lambda$, or Ω_u, we can combine these covariance restrictions with the restrictions on the coefficient matrices to identify a model and obtain efficient estimates of the parameters. We shall first discuss estimation of the simultaneous-equations model under the assumption that there are no restrictions on the variance–covariance matrix, but the rank condition for identification holds. Estimation of reduced-form or stacked equations will be discussed in Section 4.2, and estimation of the structural form will be dealt with in Section 4.3. We then discuss the case in which there are restrictions on the variance–covariance matrix in Section 4.4. Because a widely used structure for longitudinal microdata is the triangular structure (e.g., Chamberlain 1976, 1977a, 1977b; Chamberlain and Griliches 1975), we shall use this special case to illustrate how the covariance restrictions can be used to identify an otherwise unidentified model and to improve the efficiency of the estimates.

4.2 JOINT GENERALIZED LEAST SQUARES ESTIMATION TECHNIQUE

We can write an equation of a reduced form (4.1.14) in the more general form in which the explanatory variables in each equation can be different:[4]

$$
\mathbf{y}_g = \mathbf{e}_{NT}\mu_g^* + \mathbf{X}_g\boldsymbol{\pi}_g + \boldsymbol{\epsilon}_g, \quad g = 1, \ldots, G, \tag{4.2.1}
$$

where $\mathbf{y}_g$ and $\mathbf{e}_{NT}$ are $NT \times 1$ with $\mathbf{e}_{NT}$ denoting a vector of 1's, $\mathbf{X}_g$ is $NT \times K_g, \mu_g^*$ is the 1×1 intercept term for the gth equation, $\boldsymbol{\pi}_g$ *is* $K_g \times 1$, and $\boldsymbol{\epsilon}_g = (I_N \otimes \mathbf{e}_T)\boldsymbol{\alpha}_g^* + (\mathbf{e}_N \otimes I_T)\boldsymbol{\lambda}_g^* + \mathbf{u}_g^*$, where $\boldsymbol{\alpha}_g^* = (\alpha_{1g}^*, \alpha_{2g}^*, \ldots, \alpha_{Ng}^*)'$, $\boldsymbol{\lambda}_g^* = (\lambda_{1g}^*, \lambda_{2g}^*, \ldots, \lambda_{Tg}^*)'$, and $\mathbf{u}_g^* = (u_{11g}^*, u_{12g}^*, \ldots, u_{1Tg}^*, u_{21g}^*, \ldots, u_{NTg}^*)'$ are $N \times 1, T \times 1$, and $NT \times 1$ random vectors, respectively. Stacking the set of G equations, we get

$$
\underset{GNT\times 1}{\mathbf{y}} = (I_G \otimes \mathbf{e}_{NT})\boldsymbol{\mu}^* + \mathbf{X}\boldsymbol{\pi} + \boldsymbol{\epsilon}, \tag{4.2.2}
$$

[4] By allowing X to be different, the discussion of estimation of reduced-form equations can proceed along the more general format of seemingly unrelated regression models (Avery 1977; Baltagi 1980).

where

$$\underset{GNT\times 1}{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_G \end{bmatrix}, \qquad \underset{GNT\times\left(\sum_{g=1}^{G} K_g\right)}{\mathbf{X}} = \begin{bmatrix} X_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & X_2 & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & X_G \end{bmatrix},$$

$$\underset{G\times 1}{\boldsymbol{\mu}^*} = \begin{bmatrix} \mu_1^* \\ \mu_2^* \\ \vdots \\ \mu_G^* \end{bmatrix}, \qquad \underset{\left(\sum_{g=1}^{G} K_g\right)\times 1}{\boldsymbol{\pi}} = \begin{bmatrix} \boldsymbol{\pi}_1 \\ \vdots \\ \boldsymbol{\pi}_G \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_G \end{bmatrix},$$

with

$$V = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = [V_{g\ell}], \tag{4.2.3}$$

where $V_{g\ell}$ denotes the $g\ell$th block submatrix of V, which is given by

$$\underset{NT\times NT}{V_{g\ell}} = E\left(\boldsymbol{\epsilon}_g\boldsymbol{\epsilon}_\ell'\right) = \sigma_{\alpha_{g\ell}}^{*2} A + \sigma_{\lambda_{g\ell}}^{*2} D + \sigma_{u_{g\ell}}^{*2} I_{NT}, \tag{4.2.4}$$

where $A = I_N \otimes \mathbf{e}_T\mathbf{e}_T'$ and $D = \mathbf{e}_N\mathbf{e}_N' \otimes I_T$. Equation (4.2.4) can also be written as

$$\begin{aligned} V_{g\ell} &= \sigma_{1_{g\ell}}^{*2}\left(\frac{1}{T}A - \frac{1}{NT}J\right) + \sigma_{2_{g\ell}}^{*2}\left(\frac{1}{N}D - \frac{1}{NT}J\right) \\ &\quad + \sigma_{u_{g\ell}}^{*2}\tilde{Q} + \sigma_{4_{g\ell}}^{*2}\left(\frac{1}{NT}J\right), \end{aligned} \tag{4.2.5}$$

where $J = \mathbf{e}_{NT}\mathbf{e}_{NT}'$, $\tilde{Q} = I_{NT} - (1/T)A - (1/N)D + (1/NT)J$, $\sigma_{1_{g\ell}}^{*2} = \sigma_{u_{g\ell}}^{*2} + T\sigma_{\alpha_{g\ell}}^{*2}$, $\sigma_{2_{g\ell}}^{*2} = \sigma_{u_{g\ell}}^{*2} + N\sigma_{\lambda_{g\ell}}^{*2}$, and $\sigma_{4_{g\ell}}^{*2} = \sigma_{u_{g\ell}}^{*2} + T\sigma_{\alpha_{g\ell}}^{*2} + N\sigma_{\lambda_{g\ell}}^{*2}$. It was shown in Appendix 2B that $\sigma_{1_{g\ell}}^{*2}, \sigma_{2_{g\ell}}^{*2}, \sigma_{u_{g\ell}}^{*2}$, and $\sigma_{4_{g\ell}}^{*2}$ are the distinct characteristic roots of $V_{g\ell}$ of multiplicity $N-1, T-1, (N-1)(T-1)$, and 1, with C_1, C_2, C_3, and C_4 as the matrices of their corresponding characteristic vectors.

We can rewrite V as

$$\begin{aligned} V &= V_1 \otimes \left(\frac{1}{T}A - \frac{1}{NT}J\right) + V_2 \otimes \left(\frac{1}{N}D - \frac{1}{NT}J\right) \\ &\quad + \Omega_u^* \otimes \tilde{Q} + V_4 \otimes \left(\frac{1}{NT}J\right), \end{aligned} \tag{4.2.6}$$

where $V_1 = (\sigma_{1_{g\ell}}^{*2})$, $V_2 = (\sigma_{2_{g\ell}}^{*2})$, and $V_4 = (\sigma_{4_{g\ell}}^{*2})$ all of dimension $G \times G$. Using the fact that $[(1/T)A - (1/NT)J]$, $[(1/N)D - (1/NT)J]$, $\tilde{Q}$, and $[(1/NT)J]$ are symmetric idempotent matrices, mutually orthogonal, and sum to the identity matrix I_{NT}, we can write down the inverse of V explicitly as (Avery 1977; Baltagi 1980)[5]

[5] One can check that (4.2.7) is indeed the inverse of (4.2.6) by repeatedly using the formulas of the Kronecker products: $(B + C) \otimes A = B \otimes A + C \otimes A$, $(A \otimes B)(C \otimes D) = AC \otimes BD$, provided the product of these matrices exists (Theil 1971, Chapter 7, Section 7.2).

$$V^{-1} = V_1^{-1} \otimes \left(\frac{1}{T}A - \frac{1}{NT}J\right) + V_2^{-1} \otimes \left(\frac{1}{N}D - \frac{1}{NT}J\right)$$
$$+ \Omega_u^{*-1} \otimes \tilde{Q} + V_4^{-1} \otimes \left(\frac{1}{NT}J\right). \tag{4.2.7}$$

The GLS estimators of $\boldsymbol{\mu}^*$ and $\boldsymbol{\pi}$ are obtained by minimizing the distance function

$$[\mathbf{y} - (I_G \otimes \mathbf{e}_{NT})\boldsymbol{\mu}^* - \mathbf{X}\boldsymbol{\pi}]' V^{-1} [\mathbf{y} - (I_G \otimes \mathbf{e}_{NT})\boldsymbol{\mu}^* - \mathbf{X}\boldsymbol{\pi}]. \tag{4.2.8}$$

Taking partial derivatives of (4.2.8) with respect to $\boldsymbol{\mu}^*$ and $\boldsymbol{\pi}$, we obtain the first-order conditions

$$(I_G \otimes \mathbf{e}_{NT})' V^{-1} [\mathbf{y} - (I_G \otimes \mathbf{e}_{NT})\boldsymbol{\mu}^* - \mathbf{X}\boldsymbol{\pi}] = \mathbf{0}, \tag{4.2.9}$$

$$-X' V^{-1} [\mathbf{y} - (I_G \otimes \mathbf{e}_{NT})\boldsymbol{\mu}^* - \mathbf{X}\boldsymbol{\pi}] = \mathbf{0}. \tag{4.2.10}$$

Solving (4.2.9) and making use of the relations $[(1/T)A - (1/NT)J]\mathbf{e}_{NT} = \mathbf{0}$, $[(1/N)D - (1/NT)J]\mathbf{e}_{NT} = \mathbf{0}$, $\tilde{Q}\mathbf{e}_{NT} = \mathbf{0}$, and $(1/NT)J\mathbf{e}_{NT} = \mathbf{e}_{\mathrm{NT}}$, we have

$$\hat{\boldsymbol{\mu}}^* = \left(I_G \otimes \frac{1}{NT}\mathbf{e}'_{NT}\right)(\mathbf{y} - \mathbf{X}\boldsymbol{\pi}). \tag{4.2.11}$$

Substituting (4.2.11) into (4.2.10), we have the GLS estimator of $\boldsymbol{\pi}$ as[6]

$$\hat{\boldsymbol{\pi}}_{\mathrm{GLS}} = [\mathbf{X}' \tilde{V}^{-1} \mathbf{X}]^{-1} (\mathbf{X}' \tilde{V}^{-1} \mathbf{y}), \tag{4.2.12}$$

where

$$\tilde{V}^{-1} = V_1^{-1} \otimes \left(\frac{1}{T}A - \frac{1}{NT}J\right) + V_2^{-1} \otimes \left(\frac{1}{N}D - \frac{1}{NT}J\right)$$
$$+ \Omega_u^{*-1} \otimes \tilde{Q}. \tag{4.2.13}$$

If $E(\boldsymbol{\epsilon}_g \boldsymbol{\epsilon}'_\ell) = \mathbf{0}$ for $g \neq \ell$, then V is block-diagonal, and Equation (4.2.12) is reduced to applying the GLS estimation method to each equation separately. If both N and T tend to infinity and N/T tends to a nonzero constant, then lim $V_1^{-1} = \mathbf{0}$, lim $V_2^{-1} = \mathbf{0}$, and lim $V_4^{-1} = \mathbf{0}$. Equation (4.2.12) becomes the least squares dummy-variable (or fixed-effects) estimator for the seemingly unrelated regression case,

[6] If only the first M out of G equations have nonzero intercepts, we estimate the first M intercepts by $\{[I_M, (V_4^{MM})^{-1} V_4^{M(G-M)}] \otimes (1/NT)\boldsymbol{e}'_{NT}\}(\boldsymbol{y} - X\boldsymbol{\pi})$ and estimate $\boldsymbol{\pi}$ by $[X' V^{*-1} X]^{-1} [X' V^{*-1} \boldsymbol{y}]$, where I_M is the M-rowed identity matrix, V_4^{MM} and $V_4^{M(G-M)}$ are the corresponding $M \times M$ and $M \times (G-M)$ partitioned matrices of

$$V_4^{-1} = \begin{bmatrix} V_4^{MM} & V_4^{M(G-M)} \\ V_4^{(G-M)M} & V_4^{(G-M)(G-M)} \end{bmatrix},$$

and

$$V^{*-1} = \tilde{V}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & V_4^{(G-M)(G-M)} - V_4^{(G-M)M} (V_4^{MM})^{-1} V_4^{M(G-M)} \end{bmatrix} \otimes \frac{1}{NT} J.$$

For details, see Prucha (1983).

$$\operatorname{plim} \hat{\boldsymbol{\pi}}_{\text{GLS}} = \operatorname*{plim}_{\substack{N\to\infty \\ T\to\infty}} \left[\frac{1}{N}\mathbf{X}'\left(\Omega_u^{*-1} \otimes \tilde{Q}\right)\mathbf{X}\right]^{-1} \times \left[\frac{1}{NT}\mathbf{X}'\left(\Omega_u^{*-1} \otimes \tilde{Q}\right)\mathbf{y}\right]. \tag{4.2.14}$$

In the case of the standard reduced form, $X_1 = X_2 = \cdots = X_G = \bar{X}$,

$$\begin{aligned} \hat{\boldsymbol{\pi}}_{\text{GLS}} = \bigg[& V_1^{-1} \otimes \bar{X}'\left(\frac{1}{T}A - \frac{1}{NT}J\right)\bar{X} \\ & + V_2^{-1} \otimes \bar{X}'\left(\frac{1}{N}D - \frac{1}{NT}J\right)\bar{X} + \Omega_u^{*-1} \otimes \bar{X}'\tilde{Q}\bar{X}\bigg]^{-1} \\ & \times \left\{\left[V_1^{-1} \otimes \bar{X}'\left(\frac{1}{T}A - \frac{1}{NT}J\right)\right]\mathbf{y}\right. \\ & \left. + \left[V_2^{-1} \otimes \bar{X}'\left(\frac{1}{N}D - \frac{1}{NT}J\right)\right]\mathbf{y} + \left[\Omega_u^{*-1} \otimes \bar{X}'\tilde{Q}\right]\mathbf{y}\right\}. \end{aligned} \tag{4.2.15}$$

We know that in the conventional case when no restriction is imposed on the reduced-form coefficients vector $\boldsymbol{\pi}$, estimating each equation by the least squares method yields the best linear unbiased estimate. Equation (4.2.15) shows that in a seemingly unrelated regression model with error components, the fact that each equation has an identical set of explanatory variables is not a sufficient condition for the GLS performed on the whole system to be equivalent to estimating each equation separately.

Intuitively, by stacking different equations together, we shall gain efficiency in the estimates, because knowing the residual of the ℓth equation helps in predicting the gth equation when the covariance terms between different equations are nonzero. For instance, if the residuals are normally distributed, $E(\boldsymbol{\epsilon}_g \mid \boldsymbol{\epsilon}_\ell) = \text{Cov}(\boldsymbol{\epsilon}_g, \boldsymbol{\epsilon}_\ell)\text{Var}(\boldsymbol{\epsilon}_\ell)^{-1}\boldsymbol{\epsilon}_\ell \neq \mathbf{0}$. To adjust for this nonzero mean, it would be appropriate to regress $\mathbf{y}_g - \text{Cov}(\boldsymbol{\epsilon}_g, \boldsymbol{\epsilon}_\ell)\text{Var}(\boldsymbol{\epsilon}_\ell)^{-1}\boldsymbol{\epsilon}_\ell$ on $(\mathbf{e}_{NT}, X_g)$. Although in general ϵ_ℓ is unknown, asymptotically there is no difference if we replace it with the least squares residual, $\hat{\boldsymbol{\epsilon}}_\ell$. However, if the explanatory variables in different equations are identical, namely, $X_g = X_\ell = \bar{X}$, there is no gain in efficiency by bringing different equations together when the cross-equation covariances are unrestricted; because $\text{Cov}(\boldsymbol{\epsilon}_g, \boldsymbol{\epsilon}_\ell) = \sigma_{\epsilon_{g\ell}} I_{NT}$, $\text{Var}(\boldsymbol{\epsilon}_\ell) = \sigma_{\epsilon_{\ell\ell}} I_{NT}$, and $\hat{\boldsymbol{\epsilon}}_\ell$ is orthogonal to $(\mathbf{e}_{NT}, X_g)$ by construction, the variable $\sigma_{\epsilon_\ell}\sigma_{\epsilon_{\ell\ell}}^{-1}\hat{\boldsymbol{\epsilon}}_\ell$ can have no effect on the estimate of $(\boldsymbol{\mu}_g, \boldsymbol{\pi}'_g)$ when it is subtracted from $\mathbf{y}_g$. But the same cannot be said for the error-components case, because $\text{Cov}(\boldsymbol{\epsilon}_g, \boldsymbol{\epsilon}_\ell)\text{Var}(\boldsymbol{\epsilon}_\ell)^{-1}$ is not proportional to an identity matrix. The weighted variable $\text{Cov}(\boldsymbol{\epsilon}_g, \boldsymbol{\epsilon}_\ell)\text{Var}(\boldsymbol{\epsilon}_\ell)^{-1}\hat{\boldsymbol{\epsilon}}_\ell$ is no longer orthogonal to $(\mathbf{e}_{NT}, \bar{X})$. Therefore, in the error-components case, it remains fruitful to exploit the covariances between different equations to improve the accuracy of the estimates.

When V_1, V_2, and Ω_u^* are unknown, we can replace them by their consistent estimates. In Chapter 2, we discussed methods of estimating variance components. These techniques can be straightforwardly applied to the multiple-equations model as well (Avery 1977; Baltagi 1980).

The model discussed earlier assumes the existence of both individual and time effects. Suppose we believe that the covariances of some of the components are zero. The same procedure can be applied to the simpler model with some slight modifications. For example, if the covariance of the residuals between equations g and ℓ is composed of only two components (an individual effect and overall effect), then $\sigma^2_{\lambda_{g\ell}} = 0$. Hence, $\sigma^{*2}_{1_{g\ell}} = \sigma^{*2}_{4_{g\ell}}$, and $\sigma^{*2}_{2_{g\ell}} = \sigma^{*2}_{u_{g\ell}}$. These adjusted roots can be substituted into the appropriate positions in (4.2.6) and (4.2.7), with coefficient estimates following directly from (4.2.12).

4.3 ESTIMATION OF STRUCTURAL EQUATIONS

4.3.1 Estimation of a Single Equation in the Structural Model

As (4.2.12) shows, the generalized least squares estimator of the slope coefficients is invariant against centering the data around overall sample means; so for ease of exposition, we shall assume that there is an intercept term and that all sample observations are measured as deviations from their respective overall means, and we shall consider the gth structural equation as

$$\begin{aligned} \underset{NT\times 1}{\mathbf{y}_g} &= \mathbf{Y}_g\boldsymbol{\gamma}_g + \mathbf{X}_g\boldsymbol{\beta}_g + \mathbf{v}_g \\ &= \mathbf{W}_g\boldsymbol{\theta}_g + \mathbf{v}_g, \quad g = 1, \ldots, G, \end{aligned} \tag{4.3.1}$$

where $\mathbf{Y}_g$ is an $NT \times (G_g - 1)$ matrix of NT observations of $G_g - 1$ included joint dependent variables, X_g is an $NT \times K_g$ matrix of NT observations of K_g included exogenous variables, $W_g = (Y_g, X_g)$, and $\boldsymbol{\theta}_g = (\gamma'_g, \beta'_g)'$, The $\mathbf{v}_g$ is an $NT \times 1$ vector of error terms,

$$\mathbf{v}_g = (I_N \otimes \mathbf{e}_T)\boldsymbol{\alpha}_g + (\mathbf{e}_N \otimes I_T)\boldsymbol{\lambda}_g + \mathbf{u}_g, \tag{4.3.2}$$

with $\alpha_g = (\alpha_{1g}, \ldots, \alpha_{Ng})', \lambda_g = (\lambda_{1g}, \ldots, \lambda_{Tg})'$, and $\mathbf{u}_g = (u_{11g}, \ldots, u_{1Tg}, u_{21g}, \ldots, u_{NTg})'$ satisfying assumption (4.2.3). So the covariance matrix between the gth and the ℓth structural equations is

$$\begin{aligned} \Sigma_{g\ell} &= E(\mathbf{v}_g\mathbf{v}'_\ell) = \sigma^2_{\alpha_{g\ell}} A + \sigma^2_{\lambda_{g\ell}} D + \sigma^2_{u_{g\ell}} I_{NT} \\ &= \sigma^2_{1_{g\ell}}\left(\frac{1}{T}A - \frac{1}{NT}J\right) + \sigma^2_{2_{g\ell}}\left(\frac{1}{N}D - \frac{1}{NT}J\right) \\ &\quad + \sigma^2_{3_{g\ell}}\tilde{Q} + \sigma^2_{4_{g\ell}}\left(\frac{1}{NT}J\right), \end{aligned} \tag{4.3.3}$$

where $\sigma^2_{1_{g\ell}} = \sigma^2_{u_{g\ell}} + T\sigma^2_{\alpha_{g\ell}}, \sigma^2_{2_{g\ell}} = \sigma^2_{u_{g\ell}} + N\sigma^2_{\lambda_{g\ell}}, \sigma^2_{3_{g\ell}} = \sigma^2_{u_{g\ell}}$, and $\sigma^2_{4_{g\ell}} = \sigma^2_{u_{g\ell}} + T\sigma^2_{\alpha_{g\ell}} + N\sigma^2_{\lambda_{g\ell}}$. We also assume that each equation in (4.3.1) satisfies the rank condition for identification with $K \geq G_g + K_g - 1, g = 1, \ldots, G$, where K denotes the number of distinct exogenous variables in the system.

We first consider estimation of a single equation in the structural model. To estimate the gth structural equation, we take into account only the a priori restrictions affecting that equation and ignore the restrictions affecting all other equations. Therefore, suppose we are interested in estimating the first equation. The *limited-information* principle of estimating this equation is equivalent to the full-information estimation of the system

$$\begin{aligned} y_{1_{it}} &= \mathbf{w}'_{1_{it}}\boldsymbol{\theta}_1 + v_{1_{it}}, \\ y_{2_{it}} &= \mathbf{x}'_{it}\boldsymbol{\pi}_2 + \epsilon_{2_{it}}, \\ &\;\;\vdots \\ y_{G_{it}} &= \mathbf{x}'_{it}\boldsymbol{\pi}_G + \epsilon_{G_{it}}, \qquad i = 1, \ldots, N, \\ &\qquad\qquad\qquad\qquad\quad t = 1, \ldots, T, \end{aligned} \tag{4.3.4}$$

where there are no restrictions on $\boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_G$.

We can apply the usual two-stage least-squares (2SLS) method to estimate the first equation in (4.3.4). The 2SLS estimator is consistent. However, if the $v_{1_{it}}$ are not independently identically distributed over i and t, the 2SLS estimator is not efficient even within the limited-information context. To allow for arbitrary heteroscedasticity and serial correlation in the residuals, we can generalize Chamberlain's (1982, 1984) minimum-distance or generalized 2SLS estimator.

We first consider the minimum-distance estimator. Suppose T is fixed and N tends to infinity. Stacking the T period equations for a single individual's behavioral equation into one system, we create a model of GT equations,

$$\begin{aligned}
\underset{T\times 1}{\mathbf{y}_{1_i}} &= \mathbf{W}_{1_i}\boldsymbol{\theta}_1 + \mathbf{v}_{1_i},\\
\underset{T\times 1}{\mathbf{y}_{2_i}} &= \mathbf{X}_i\boldsymbol{\pi}_2 + \boldsymbol{\epsilon}_{2_i},\\
&\ \ \vdots \\
\mathbf{y}_{G_i} &= \mathbf{X}_i\boldsymbol{\pi}_G + \boldsymbol{\epsilon}_{G_i}, \qquad i = 1,\ldots,N.
\end{aligned} \tag{4.3.5}$$

Let $\mathbf{y}_i' = (\mathbf{y}_{1_i}', \ldots, \mathbf{y}_{G_i}')$. The reduced form of $\mathbf{y}_i$ is

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{1_i} \\ \mathbf{y}_{2_i} \\ \vdots \\ \mathbf{y}_{G_i} \end{bmatrix} = (I_G \otimes \tilde{\mathbf{X}}_i)\tilde{\boldsymbol{\pi}} + \boldsymbol{\epsilon}_i, \qquad i = 1,\ldots,N, \tag{4.3.6}$$

where

$$\begin{aligned}
\underset{T\times TK}{\tilde{\mathbf{X}}_i} &= \begin{bmatrix} \mathbf{x}_{i1}' & & & \mathbf{0} \\ & \mathbf{x}_{i2}' & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{x}_{iT}' \end{bmatrix},\\
\tilde{\boldsymbol{\pi}} &= \operatorname{vec}(\tilde{\Pi}'),
\end{aligned} \tag{4.3.7}$$

$$\underset{GT\times K}{\tilde{\Pi}} = \Pi \otimes \mathbf{e}_T, \quad \text{and} \quad \Pi = E(\mathbf{y}_{it} \mid \mathbf{x}_{it}). \tag{4.3.8}$$

The unconstrained least squares regression of $\mathbf{y}_i$ on $(I_G \otimes \tilde{\mathbf{X}}_i)$ yields a consistent estimate of $\tilde{\boldsymbol{\pi}}, \hat{\tilde{\boldsymbol{\pi}}}$. If $\boldsymbol{\epsilon}_i$ are independently distributed over i, then $\sqrt{N}(\hat{\tilde{\boldsymbol{\pi}}} - \tilde{\boldsymbol{\pi}})$ is asymptotically normally distributed, with mean zero and variance–covariance matrix

$$\underset{GTK\times GTK}{\tilde{\Omega}} = \left(I_G \otimes \Phi_{xx}^{-1}\right)\tilde{V}\left(I_G \otimes \Phi_{xx}^{-1}\right), \tag{4.3.9}$$

where $\Phi_{xx} = E\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i = \operatorname{diag}\{E(\mathbf{x}_{i1}\mathbf{x}_{i1}'), \ldots, E(\mathbf{x}_{iT}\mathbf{x}_{iT}')\}$, and $\tilde{V}$ is a $GTK \times GTK$ matrix, with the $g\ell$th block a $TK \times TK$ matrix of the form

$$\tilde{V}_{g\ell} = E\begin{bmatrix} \epsilon_{gi1}\epsilon_{\ell i1}\mathbf{x}_{i1}\mathbf{x}_{i1}' & \epsilon_{gi1}\epsilon_{\ell i2}\mathbf{x}_{i1}\mathbf{x}_{i2}' & \cdots & \epsilon_{gi1}\epsilon_{\ell iT}\mathbf{x}_{i1}\mathbf{x}_{iT}' \\ \epsilon_{gi2}\epsilon_{\ell i1}\mathbf{x}_{i2}\mathbf{x}_{i1}' & \epsilon_{gi2}\epsilon_{\ell i2}\mathbf{x}_{i2}\mathbf{x}_{i2}' & \cdots & \epsilon_{gi2}\epsilon_{\ell iT}\mathbf{x}_{i2}\mathbf{x}_{iT}' \\ \vdots & \vdots & & \vdots \\ \epsilon_{giT}\epsilon_{\ell i1}\mathbf{x}_{iT}\mathbf{x}_{i1}' & \epsilon_{giT}\epsilon_{\ell i2}\mathbf{x}_{iT}\mathbf{x}_{i2}' & \cdots & \epsilon_{giT}\epsilon_{\ell iT}\mathbf{x}_{iT}\mathbf{x}_{iT}' \end{bmatrix}. \tag{4.3.10}$$

One can obtain a consistent estimator of $\tilde{\Omega}$ by replacing the population moments in $\tilde{\Omega}$ by the corresponding sample moments (e.g., $E\mathbf{x}_{i1}\mathbf{x}_{i1}'$ is replaced by $\sum_{i=1}^{N}\mathbf{x}_{i1}\mathbf{x}_{i1}'/N$).

Let $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \pi_2', \ldots, \pi_G')$, and specify the restrictions on $\tilde{\pi}$ by the condition that $\tilde{\pi} = \mathbf{f}(\boldsymbol{\theta})$. Choose $\boldsymbol{\theta}$ to minimize the following distance function:

$$[\hat{\tilde{\pi}} - \tilde{\mathbf{f}}(\boldsymbol{\theta})]'\hat{\tilde{\Omega}}^{-1}[\hat{\tilde{\pi}} - \tilde{\mathbf{f}}(\boldsymbol{\theta})]. \tag{4.3.11}$$

Then $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed with mean zero and variance–covariance matrix $(\tilde{F}'\tilde{\Omega}^{-1}\tilde{F})^{-1}$, where $\tilde{F} = \partial\tilde{\mathbf{f}}/\partial\boldsymbol{\theta}'$. Noting that $\tilde{\Pi} = \Pi \otimes \mathbf{e}_T$, and evaluating the partitioned inverse, we obtain the asymptotic variance–covariance matrix of $\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)$ as

$$\left(\tilde{\Phi}_{w_1x}\Psi_{11}^{-1}\tilde{\Phi}_{w_1x}'\right)^{-1}, \tag{4.3.12}$$

where $\tilde{\Phi}_{w_1x} = [E(\mathbf{w}_{1_{i1}}\mathbf{x}_{i1}'), E(\mathbf{w}_{1_{i2}}\mathbf{x}_{i2}'), \ldots, E(\mathbf{w}_{1_{iT}}\mathbf{x}_{iT}')]$, and

$$\Psi_{11} = E\begin{bmatrix} v_{1_{i1}}^2\mathbf{x}_{i1}\mathbf{x}_{i1}' & v_{1_{i1}}v_{1_{i2}}\mathbf{x}_{i1}\mathbf{x}_{i2}' & \cdots & v_{1_{i1}}v_{1_{i1}}\mathbf{x}_{i1}\mathbf{x}_{iT}' \\ v_{1_{i2}}v_{1_{i1}}\mathbf{x}_{i2}\mathbf{x}_{i1}' & v_{1_{i2}}^2\mathbf{x}_{i2}\mathbf{x}_{i2}' & \cdots & v_{1_{i2}}v_{1_{iT}}\mathbf{x}_{i2}\mathbf{x}_{iT}' \\ \vdots & \vdots & & \vdots \\ v_{1_{iT}}v_{1_{i1}}\mathbf{x}_{iT}\mathbf{x}_{i1}' & v_{1_{iT}}v_{1_{i2}}\mathbf{x}_{iT}\mathbf{x}_{i2}' & \cdots & v_{1_{iT}}v_{1_{iT}}\mathbf{x}_{iT}\mathbf{x}_{iT}' \end{bmatrix}. \tag{4.3.13}$$

The limited-information minimum-distance estimator of (4.3.11) is asymptotically equivalent to the following generalization of the 2SLS estimator:

$$\hat{\boldsymbol{\theta}}_{1,\mathrm{G2SLS}} = \left(\tilde{\mathbf{S}}_{w_1x}\hat{\Psi}_{11}^{-1}\tilde{\mathbf{S}}_{w_1x}'\right)^{-1}\left(\tilde{\mathbf{S}}_{w_1x}\hat{\Psi}_{11}^{-1}\mathbf{s}_{xy_1}\right), \tag{4.3.14}$$

where

$$\tilde{\mathbf{S}}_{w_1x} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{1_{i1}}\mathbf{x}_{i1}', \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{1_{i2}}\mathbf{x}_{i2}', \ldots, \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{1_{iT}}\mathbf{x}_{iT}'\right),$$

$$\mathbf{s}_{xy_1} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{i1}y_{1_{i1}} \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{i2}y_{1_{i2}} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{iT}y_{1_{iT}} \end{bmatrix},$$

$$\hat{\Psi}_{11} = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N}\hat{v}_{1_{i1}}^2\mathbf{x}_{i1}\mathbf{x}_{i1}' & \sum_{i=1}^{N}\hat{v}_{1_{i1}}\hat{v}_{1_{i2}}\mathbf{x}_{i1}\mathbf{x}_{i2}' & \cdots & \sum_{i=1}^{N}\hat{v}_{1_{i1}}\hat{v}_{1_{iT}}\mathbf{x}_{i1}\mathbf{x}_{iT}' \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N}\hat{v}_{1_{iT}}\hat{v}_{1_{i1}}\mathbf{x}_{iT}\mathbf{x}_{i1}' & \sum_{i=1}^{N}\hat{v}_{1iT}\hat{v}_{1i2}\mathbf{x}_{iT}\mathbf{x}_{i2}' & \cdots & \sum_{i=1}^{N}\hat{v}_{1_{iT}}\hat{v}_{1_{iT}}\mathbf{x}_{iT}\mathbf{x}_{iT}' \end{bmatrix},$$

and $\hat{v}_{1_{it}} = y_{1_{it}} - \mathbf{w}_{1_{it}}'\hat{\boldsymbol{\theta}}_1$, with $\hat{\boldsymbol{\theta}}_1$ any consistent estimator of $\boldsymbol{\theta}_1$. The generalized 2SLS coverges to the 2SLS if $v_{1_{it}}$ is independently identically distributed over i and t and $E\mathbf{x}_{it}\mathbf{x}_{it}' = E\mathbf{x}_{is}\mathbf{x}_{is}'$. But the generalized 2SLS, like the minimum-distance estimator of (4.3.11), makes allowance for the heteroscedasticity and arbitrary serial correlation in $\nu_{1_{it}}$, whereas the 2SLS does not.

When the variance–covariance matrix $\sum_{gg}$ possesses an error-component structure as specified in (4.3.3), although both the 2SLS estimator and the minimum-distance estimator of (4.3.11) (or the generalized 2SLS estimator) remain consistent, they are no longer efficient even within a limited-information framework; because, as shown in the last section, when there are restrictions on the variance–covariance matrix, the least squares estimator of the unconstrained Π is not as efficient as the generalized least squares estimator[7]. An efficient estimation method has to exploit the known restrictions on the error structure. Baltagi (1981a) suggested using the following error-component two-stage least squares (EC2SLS) method to obtain a more efficient estimator of the unknown parameters in the gth equation.

Transforming (4.3.1) by the eigenvectors of $\sum_{gg}, C_1', C_2'$, and C_3', we have[8]

$$\mathbf{y}_g^{(h)} = Y_g^{(h)}\boldsymbol{\gamma}_g + \mathbf{X}_g^{(h)}\boldsymbol{\beta}_g + \mathbf{v}_g^{(h)} = \mathbf{W}_g^{(h)}\boldsymbol{\theta}_g + \mathbf{v}_g^{(h)}, \tag{4.3.15}$$

where $\mathbf{y}_g^{(h)} = C_h'\mathbf{y}_g, W_g^{(h)} = C_h'W_g, \mathbf{v}_g^{(h)} = C_h'\mathbf{v}_g$ for $h = 1, 2, 3$, and C_1', C_2', and C_3' are as defined in Appendix 2B. The transformed disturbance term $\mathbf{v}_g^{(h)}$ is mutually orthogonal and has a covariance matrix proportional to an identity matrix. We can therefore use $X^{(h)} = C_h'X$ as the instruments and apply the Aitken estimation procedure to the system of equations

$$\begin{bmatrix} X^{(1)\prime}\mathbf{y}_g^{(1)} \\ X^{(2)\prime}\mathbf{y}_g^{(2)} \\ X^{(3)\prime}\mathbf{y}_g^{(3)} \end{bmatrix} = \begin{bmatrix} X^{(1)\prime}W_g^{(1)} \\ X^{(2)\prime}W_g^{(2)} \\ X^{(3)\prime}W_g^{(3)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_g \\ \boldsymbol{\beta}_g \end{bmatrix} + \begin{bmatrix} X^{(1)\prime}\mathbf{v}_g^{(1)} \\ X^{(2)\prime}\mathbf{v}_g^{(2)} \\ X^{(3)\prime}\mathbf{v}_g^{(3)} \end{bmatrix}. \tag{4.3.16}$$

The resulting Aitken estimator of $(\boldsymbol{\gamma}_g', \boldsymbol{\beta}_g')$ is

$$\hat{\boldsymbol{\theta}}_{g,\mathrm{EC2SLS}} = \left\{\sum_{h=1}^{3}\left[\frac{1}{\sigma_{hgg}^2}W_g^{(h)\prime}P_X(h)W_g^{(h)}\right]\right\}^{-1} \left\{\sum_{h=1}^{3}\left[\frac{1}{\sigma_{hgg}^2}W_g^{(h)\prime}P_X(H)\mathbf{y}_g^{(h)}\right]\right\}, \tag{4.3.17}$$

where $P_X(h) = X^{(h)}(X^{(h)\prime}X^{(h)})^{-1}X^{(h)\prime}$. It is a weighted combination of the between-groups, between-time-periods, and within-groups 2SLS estimators of $(\boldsymbol{\gamma}_g', \boldsymbol{\beta}_g')$. The weights σ_{hgg}^2 can be estimated by substituting the transformed 2SLS residuals in the usual variance formula,

$$\hat{\sigma}_{hgg}^2 = \left(\mathbf{y}_g^{(h)} - W_g^{(h)}\hat{\boldsymbol{\theta}}_{g,2\mathrm{SLS}}^{(h)}\right)'\left(\mathbf{y}_g^{(h)} - W_g^{(h)}\hat{\boldsymbol{\theta}}_{g,2\mathrm{SLS}}^{(h)}\right)/n(h), \tag{4.3.18}$$

where $\hat{\boldsymbol{\theta}}_{g,2\mathrm{SLS}}^{(h)} = [W_g^{(h)\prime}P_X(h)W_g^{(h)}]^{-1}[W_g^{(h)\prime}P_X(h)\mathbf{y}_g^{(h)}]$, and $n(1) = N-1, n(2) = T-1$, $n(3) = (N-1)(T-1)$. If $N \to \infty, T \to \infty$, and N/T tends to a nonzero constant, then the probability limit of the EC2SLS tends to the 2SLS estimator based on the within-groups variation alone.

In the special case in which the source of correlation between some of the regressors and residuals comes from the unobserved time-invariant individual effects alone, the correlations between them can be removed by removing the time-invariant component from the corresponding variables. Thus, instruments for the correlated regressors can be

[7] See Chapter 2, footnote 20.

[8] As indicated earlier, we have assumed here that all variables are measured as deviations from their respective overall means. There is no loss of generality in this formulation, because the intercept μ_g is estimated by $\hat{\mu}_g = (1/NT)\boldsymbol{e}_{NT}'(\boldsymbol{y}_g - W_g\hat{\boldsymbol{\theta}}_g)$. Because $C_h'\boldsymbol{e}_{NT} = \mathbf{0}$ for $h = 1, 2, 3$, the only terms pertinent to our discussion are C_h for $h = 1, 2, 3$.

chosen from "inside" the equation, as opposed to the conventional method of being chosen from "outside" the equation. Hausman and Taylor (1981) noted that for variables that are time-varying and are correlated with α_{ig}, transforming them into deviations from their corresponding time means provides legitimate instruments, because they will no longer be correlated with α_{ig}. For variables that are time-invariant, the time means of those variables that are uncorrelated with α_{ig} can be used as instruments. Hence, a necessary condition for identification of all the parameters within a single-equation framework is that the number of time-varying variables that are uncorrelated with α_{ig} be at least as great as the number of time-invariant variables that are correlated with α_{ig}. They further showed that when the variance-component structure of the disturbance term is taken account of, the instrumental-variable estimator with instruments chosen this way is efficient among the single-equation estimators.

4.3.2 Estimation of the Complete Structural System

The single-equation estimation method considered earlier ignores restrictions in all equations in the structural system except the one being estimated. In general, we expect to get more efficient estimates if we consider the additional information contained in the other equations. In this subsection we consider the full-information estimation methods.

Let $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_G')'$, $\mathbf{v} = (\mathbf{v}_1', \ldots, \mathbf{v}_G')'$,

$$W = \begin{bmatrix} W_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & W_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & W_G \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_G \end{bmatrix}.$$

We write the set of G structural equations as

$$\mathbf{y} = W\boldsymbol{\theta} + \mathbf{v}. \tag{4.3.19}$$

We can estimate the system (4.3.19) by the three-stage least-squares (3SLS) method. But just as in the limited-information case, the 3SLS estimator is efficient only if $(v_{1_{it}}, v_{2_{it}}, \ldots, v_{G_{it}})$ are independently identically distributed over i and t. To allow for arbitrary heteroscedasticity or serial correlation, we can use the full-information minimum-distance estimator or the generalized 3SLS estimator.

We first consider the minimum-distance estimator. When T is fixed and N tends to infinity, we can stack the T period equations for an individual's behavioral equation into a system to create a model of GT equations,

$$\begin{aligned} \underset{T\times 1}{\mathbf{y}_{1_i}} &= W_{1_i}\boldsymbol{\theta}_1 + \mathbf{v}_{1_i}, \\ \mathbf{y}_{2_i} &= W_{2i}\boldsymbol{\theta}_2 + \mathbf{v}_{2_i}, \\ &\vdots \\ \mathbf{y}_{G_i} &= W_{G_i}\boldsymbol{\theta}_G + \mathbf{v}_{G_i}, \qquad i = 1, \ldots, N. \end{aligned} \tag{4.3.20}$$

We obtain a minimum-distance estimator of $\boldsymbol{\theta}$ by choosing $\hat{\boldsymbol{\theta}}$ to minimize $[\hat{\tilde{\boldsymbol{\pi}}} - \tilde{\mathbf{f}}(\boldsymbol{\theta})]' \hat{\tilde{\Omega}}^{-1}[\hat{\tilde{\boldsymbol{\pi}}} - \tilde{\mathbf{f}}(\boldsymbol{\theta})]$, where $\hat{\tilde{\boldsymbol{\pi}}}$ is the unconstrained least-squares estimator of regressing $\mathbf{y}_i$ on $(I_G \otimes \tilde{\mathbf{X}}_i)$, and $\hat{\tilde{\Omega}}$ is a consistent estimate of $\tilde{\Omega}$ (Equation 4.3.9). Noting that $\tilde{\Pi} = \boldsymbol{\Pi} \otimes \mathbf{e}_T$ and $\text{vec}(\Pi') = \boldsymbol{\pi} = \text{vec}[-\Gamma^{-1}B]'$ for all elements of Γ and B not known a priori, and making use of the formula $\partial \boldsymbol{\pi}/\partial \boldsymbol{\theta}'$ (Equation 2.9.25), we can show that if $\mathbf{v}_i$ are independently distributed over i, then $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed, with mean zero and variance–covariance matrix

$$\{\Phi_{wx}\Psi^{-1}\Phi'_{wx}\}^{-1}, \tag{4.3.21}$$

where

$$\Phi_{wx} = \begin{bmatrix} \tilde{\Phi}_{w_1x} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\Phi}_{w_2x} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tilde{\Phi}_{w_Gx} \end{bmatrix},$$

$$\tilde{\Phi}_{w_gx} = [E(\mathbf{w}_{g_{i1}}\mathbf{x}'_{i1}), E(\mathbf{w}_{g_{i2}}\mathbf{x}'_{i2}), \ldots, E(\mathbf{w}_{g_{iT}}\mathbf{x}'_{iT})], \tag{4.3.22}$$

$$\underset{GTK\times GTK}{\Psi} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \cdots & \Psi_{1G} \\ \Psi_{21} & \Psi_{22} & \cdots & \Psi_{2G} \\ \vdots & \vdots & & \vdots \\ \Psi_{G1} & \Psi_{G2} & \cdots & \Psi_{GG} \end{bmatrix},$$

$$\underset{TK\times TK}{\Psi_{g\ell}} = E\begin{bmatrix} v_{gi1}v_{\ell_{i1}}\mathbf{x}_{i1}\mathbf{x}'_{i1} & v_{gi1}v_{\ell_{i2}}\mathbf{x}_{i1}\mathbf{x}'_{i2} & \cdots & v_{gi1}v_{\ell_{iT}}\mathbf{x}_{i1}\mathbf{x}'_{iT} \\ \vdots & \vdots & & \vdots \\ v_{giT}v_{\ell i1}\mathbf{x}_{iT}\mathbf{x}'_{i1} & v_{giT}v_{\ell_{i2}}\mathbf{x}_{iT}\mathbf{x}'{}_{i2} & \cdots & v_{giT}v_{\ell_{iT}}\mathbf{x}_{iT}\mathbf{x}'_{iT} \end{bmatrix}.$$

We can also estimate (4.3.20) by using a generalized 3SLS estimator,

$$\hat{\boldsymbol{\theta}}_{\text{G3SLS}} = (S_{wx}\hat{\Psi}^{-1}S'_{wx})^{-1}(S_{wx}\hat{\Psi}^{-1}S_{xy}), \tag{4.3.23}$$

where

$$S_{wx} = \begin{bmatrix} \tilde{S}_{w_1x} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{S}_{w_2x} & & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tilde{S}_{w_Gx} \end{bmatrix},$$

$$\tilde{S}_{w_gx} = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{g_{i1}}\mathbf{x}'_{i1}, \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{g_{i2}}\mathbf{x}'_{i2}, \ldots, \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_{g_{iT}}\mathbf{x}'_{iT}\right],$$

$$S_{xy} = \begin{bmatrix} \mathbf{s}_{xy_1} \\ \mathbf{s}_{xy_2} \\ \vdots \\ \mathbf{s}_{xy_G} \end{bmatrix},$$

$$\underset{TK\times 1}{\mathbf{s}_{xy_g}} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{i1}y_{g_{i1}} \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{iT}y_{g_{iT}}, \end{bmatrix},$$

and $\hat{\Psi}$ is Ψ (Equation 4.3.22) with $\mathbf{v}_{it}$ replaced by $\hat{\mathbf{v}}_{it} = \hat{\Gamma}\mathbf{y}_{it} + \hat{B}\mathbf{x}_{it}$, where $\hat{\Gamma}$ and $\hat{B}$ are any consistent estimates of Γ and B. The generalized 3SLS is asymptotically equivalent to the minimum-distance estimator.

Both the 3SLS and the generalized 3SLS are consistent. But just as in the limited-information case, if the variance–covariance matrix possesses an error-component structure, they are not fully efficient. To take advantage of the known structure of the covariance matrix, Baltagi (1981a) suggested the following error-component three-stage least-squares estimator (EC3SLS).

The $g\ell$th block of the covariance matrix Σ is of the form (4.3.3). A key point that is evident from Appendix 2B is that the set of eigenvectors C_1, C_2, C_3, and C_4 of (4.3.3) is invariant with respect to changes in the parameters $\sigma^2_{\lambda_{g\ell}}, \sigma^2_{\alpha_{g\ell}}$, and $\sigma^2_{u_{g\ell}}$. Therefore, premultiplying (4.3.19) by $I_G \otimes C_h'$, we have[9]

$$\mathbf{y}^{(h)} = W^{(h)}\theta + \mathbf{v}^{(h)}, \qquad h = 1, 2, 3, \tag{4.3.24}$$

where $\mathbf{y}^{(h)} = (I_G \otimes C_h')\mathbf{y}$, $W^{(h)} = (I_G \otimes C_h')W$, $\mathbf{v}^{(h)} = (I_G \otimes C_h')\mathbf{v}$, with $E(\mathbf{v}^{(h)}\mathbf{v}^{(h)\prime}) = \Sigma^{(h)} \otimes I_{n(h)}$, where $\Sigma^{(h)} = (\sigma^2_{h_{g\ell}})$ for $h = 1$, 2, and 3. Because $W^{(h)}$ contains endogenous variables that are correlated with $\mathbf{v}^{(h)}$, we first premultiply (4.3.24) by $(I_G \otimes X^{(h)})'$ to purge the correlation between $W^{(h)}$ and $\mathbf{v}^{(h)}$. Then we apply the GLS estimation procedure to the resulting systems of equations to obtain

$$\hat{\boldsymbol{\theta}}_{\text{GLS}} = \left[\sum_{h=1}^{3}\{W^{(h)\prime}[(\Sigma^{(h)})^{-1} \otimes P_X(h)]W^{(h)}\}\right]^{-1} \times \left[\sum_{h=1}^{3}\{W^{(h)\prime}[(\Sigma^{(h)})^{-1} \otimes P_X(h)]\mathbf{y}^{(h)}\right]. \tag{4.3.25}$$

Usually we do not know $\Sigma^{(h)}$. Therefore, the following three-stage procedure is suggested:

1. Estimate the $\hat{\boldsymbol{\theta}}_g^{(h)}$ by 2SLS.
2. Use the residuals from the hth 2SLS estimate to estimate $\hat{\sigma}^2_{h_{gl}}$ (Equation 4.3.18).
3. Replace $\Sigma^{(h)}$ by the estimated covariance matrix. Estimate $\boldsymbol{\theta}$ by (4.3.25).

The resulting estimator is called the EC3SLS estimator. It is a weighted combination of three 3SLS (within, between-groups, and between-time-periods) estimators of the structural parameters (Baltagi 1981a).

The EC3SLS estimator is asymptotically equivalent to the full-information maximum-likelihood estimator. In the case in which Σ is block-diagonal, the EC3SLS reduces to the EC2SLS. But, contrary to the usual simultaneous-equations models, when the error terms have an error-component structure, the EC3SLS does not necessarily reduce to the EC2SLS, even if all the structural equations are just identified. For details, see Baltagi (1981a).

In the study of the relationship between health and labor supply, a commonly cited issue is the endogeneity of self-reported health status. Some authors have resorted to instrument self-reported health status (e.g., Bound 1991; Campolieti 2002). However, the instrumental variable approach does not necessarily resolve the endogeneity of self-reported health (e.g., Kreider 1991). A simultaneous equation model is probably better able to deal with the complexity of the source of endogeneity. A panel simultaneous equation model not only

[9] Again, we ignore $C_4 = \boldsymbol{e}_{NT}/\sqrt{NT}$ because we have assumed that there is an intercept for each equation, and because $C_h'\boldsymbol{e}_{NT} = \mathbf{0}$ for $h = 1, 2, 3$.

can study the sources of endogeneity of health but also can allow an investigator to control unobserved heterogeneity better. For instance, Cai and Kalb (2006) employed a panel random-effects simultaneous-equation model with the first four waves of the Household, Income, and Labour Dynamics in Australia (HILDA) survey to explore the relationship between health and labor force status. They found that health had a positive and significant effect on labor force participation for both males and females. On the other hand, they also found that labor force participation had a negative effect on male health but a positive effect on female health.

4.4 TRIANGULAR SYSTEM

The model discussed earlier assumes that residuals of different equations in a multiequation model have an unrestricted variance-component structure. Under this assumption, the panel data only improve the precision of the estimates by providing a large number of sample observations. It does not offer additional opportunities that are not standard. However, quite often the residual correlations may simply be due to one or two common omitted or unobservable variables (Chamberlain 1976, 1977a, 1977b; Chamberlain and Griliches 1975; Goldberger 1972; Zellner 1970). For instance, in the estimation of income and schooling relations or individual-firm production and factor-demand relations, it is sometimes postulated that the biases in different equations are caused by a common left-out “ability” or “managerial-differences” variable. When panel data are used, this common omitted variable is again assumed to have a within- and between-group structure. The combination of this factor-analytic structure with error-components formulations puts restrictions on the residual covariance matrix that can be used to identify an otherwise unidentified model and improve the efficiency of the estimates. Because a widely used structure for longitudinal microdata is the triangular structure, and because its connection with the general simultaneous-equations model in which the residuals have a factor-analytic structure holds in general, in this section we focus on the triangular structure to illustrate how such information can be used to identify and estimate a model.

4.4.1 Identification

A convenient way to model correlations across equations, as well as the correlation of a given individual at different times (or different members of a group), is to use latent variables to connect the residuals. Let $y_{g_{it}}$ denote the value of the variable y_g for the ith individual (or group) at time t (or tth member). We can assume that

$$v_{g_{it}} = d_g h_{it} + u_{g_{it}}, \tag{4.4.1}$$

where the u_g are uncorrelated across equations and across i and t. The correlations across equations are all generated by the common omitted variable h, which is assumed to have a variance-component structure:

$$h_{it} = \alpha_i + \omega_{it}, \tag{4.4.2}$$

where α_i is invariant over t but is independently identically distributed across i (groups), with mean zero and variance σ_α^2, and ω_{it} is independently identically distributed across i and t, with mean zero and variance σ_ω^2, and is uncorrelated with α_i.

An example of the model with Γ lower-triangular and $\mathbf{v}$ of the form (4.4.1) is (Chamberlain 1977a, 1977b; Chamberlain and Griliches 1975; Griliches 1979)

$$
\begin{aligned}
y_{1_{it}} &= \boldsymbol{\beta}_1'\mathbf{x}_{it} + d_1 h_{it} + u_{1_{it}},\\
y_{2_{it}} &= -\gamma_{21} y_{1_{it}} + \boldsymbol{\beta}_2'\mathbf{x}_{it} + d_2 h_{it} + u_{2_{it}},\\
y_{3_{it}} &= -\gamma_{31} y_{1_{it}} - \gamma_{32} y_{2_{it}} + \boldsymbol{\beta}_3'\mathbf{x}_{it} + d_3 h_{it} + u_{3_{it}},
\end{aligned}
\tag{4.4.3}
$$

where y_1, y_2, and y_3 denote years of schooling, a late (postschool) test score, and earnings, respectively, and $\mathbf{x}_{it}$ are exogenous variables (which may differ from equation to equation via restrictions on $\boldsymbol{\beta}_g$). The unobservable h can be interpreted as early "ability," and u_2 as measurement error in the test. The index i indicates groups (or families), and t indicates members in each group (or family).

Without the h variables, or if $d_g = 0$, Equation (4.4.3) would be only a simple recursive system that could be estimated by applying least squares separately to each equation. The simultaneity problem arises when we admit the possibility that $d_g \neq 0$. In general, if there were enough exogenous variables in the first (schooling) equation that did not appear again in the other equations, the system could be estimated using 2SLS or EC2SLS procedures. Unfortunately, in the income–schooling–ability model using sibling data (e.g., see the survey by Griliches 1979), there usually are not enough distinct $\mathbf{x}$'s to identify all the parameters. Thus, restrictions imposed on the variance–covariance matrix of the residuals will have to be used.

Given that h is unobservable, we have an indeterminate scale

$$
d_g^2\left(\sigma_\alpha^2 + \sigma_\omega^2\right) = c d_g^2\left(\frac{1}{c}\sigma_\alpha^2 + \frac{1}{c}\sigma_\omega^2\right). \tag{4.4.4}
$$

So we normalize h by letting $\sigma_\alpha^2 = 1$. Then

$$
E\mathbf{v}_{it}\mathbf{v}_{it}' = \left(1 + \sigma_\omega^2\right)\mathbf{d}\mathbf{d}' + \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_G^2\right) = \Omega, \tag{4.4.5}
$$

$$
E\mathbf{v}_{it}\mathbf{v}_{is}' = \mathbf{d}\mathbf{d}' = \Omega_w \qquad \text{if } t \neq s, \tag{4.4.6}
$$

$$
E\mathbf{v}_{it}\mathbf{v}_{js}' = \mathbf{0} \qquad \text{if } i \neq j, \tag{4.4.7}
$$

where $\mathbf{d} = (d_1, \ldots, d_G)$, and $\operatorname{diag}(\sigma_1^2, \ldots, \sigma_G^2)$ denotes a $G \times G$ diagonal matrix with $\sigma_1^2, \sigma_2^2, \ldots, \sigma_G^2$ on the diagonal.

Under the assumption that α_i, ω_{it}, and $u_{g_{it}}$ are normally distributed, or if we limit our attention to second-order moments, all the information with regard to the distribution of $\mathbf{y}$ is contained in

$$
C_{y_{tt}} = \Gamma^{-1}\mathbf{B}C_{x_{tt}}\mathbf{B}'\Gamma'^{-1} + \Gamma^{-1}\Omega\Gamma'^{-1}, \tag{4.4.8}
$$

$$
C_{y_{ts}} = \Gamma^{-1}\mathbf{B}C_{x_{ts}}\mathbf{B}'\Gamma'^{-1} + \Gamma^{-1}\Omega_w\Gamma'^{-1}, \qquad t \neq s, \tag{4.4.9}
$$

$$
C_{yx_{ts}} = -\Gamma^{-1}\mathbf{B}C_{x_{ts}}, \tag{4.4.10}
$$

where $C_{y_{ts}} = E\mathbf{y}_{it}\mathbf{y}_{is}'$, $C_{yx_{ts}} = E\mathbf{y}_{it}\mathbf{x}_{is}'$, and $C_{x_{ts}} = E\mathbf{x}_{it}\mathbf{x}_{is}'$.

Stack the coefficient matrices Γ and $\mathbf{B}$ into a $1 \times G(G+K)$ vector $\boldsymbol{\theta}' = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_G', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_G')$. Suppose $\boldsymbol{\theta}$ is subject to M a priori constraints:

$$
\Phi(\boldsymbol{\theta}) = \boldsymbol{\phi}, \tag{4.4.11}
$$

where $\boldsymbol{\phi}$ is an $M \times 1$ vector of constants. Then a necessary and sufficient condition for local identification of $\Gamma, \mathbf{B}, \mathbf{d}, \sigma_\omega^2$, and $\sigma_1^2, \ldots, \sigma_G^2$ is that the rank of the Jacobian formed by taking partial derivatives of (4.4.8)–(4.4.11) with respect to the unknowns is equal to $G(G+K)+2G+1$ (e.g., Hsiao 1983).

Suppose there is no restriction on the matrix **B**. The GK equations (4.4.10) can be used to identify **B** provided that Γ is identifiable. Hence, we can concentrate on

$$\Gamma\left(C_{y_{tt}} - C_{yx_{tt}}C_{x_{tt}}^{-1}C_{yx_{tt}}'\right)\Gamma' = \Omega, \tag{4.4.12}$$

$$\Gamma\left(C_{y_{ts}} - C_{yx_{ts}}C_{x_{ts}}^{-1}C_{yx_{ts}}'\right)\Gamma' = \Omega_w, \qquad t \neq s, \tag{4.4.13}$$

We note that Ω is symmetric, and we have $G(G+1)/2$ independent equations from (4.4.12). But Ω_w is of rank 1; therefore, we can derive only G independent equations from (4.4.13). Suppose Γ is lower-triangular and the diagonal elements of Γ are normalized to be unity; there are $G(G-1)/2$ unknowns in Γ, and $2G+1$ unknowns of $(\mathrm{d}_1,\ldots,d_G),(\sigma_1^2,\ldots,\sigma_G^2)$, and σ_ω^2. We have one less equation than the number of unknowns. In order for the Jacobian matrix formed by (4.4.12), (4.4.13), and a priori restrictions to be nonsingular, we need at least one additional a priori restriction. Thus, for the system

$$\Gamma\mathbf{y}_{it} + \mathbf{B}\mathbf{x}_{it} = \mathbf{v}_{it}, \tag{4.4.14}$$

where Γ is lower-triangular, **B** is unrestricted, and $\mathbf{v}_{it}$ satisfies (4.4.1) and (4.4.2), a necessary condition for the identification under exclusion restrictions is that at least one $\gamma_{g\ell} = 0$ for $g > \ell$ (for details, see Chamberlain 1976; Hsiao 1983).

4.4.2 Estimation

We have discussed how the restrictions in the variance–covariance matrix can help identify the model. We now turn to the issues of estimation. Two methods are discussed: the purged-instrumental-variable method (Chamberlain 1977a) and the maximum-likelihood method (Chamberlain and Griliches 1975). The latter method is efficient but computationally complicated. The former method is inefficient, but it is simple and consistent. It also helps to clarify the previous results on the sources of identification.

For simplicity, we assume that there is no restriction on the coefficients of exogenous variables. Under this assumption we can further ignore the existence of exogenous variables without loss of generality, because there are no excluded exogenous variables that can legitimately be used as instruments for the endogenous variables appearing in the equation. The instruments have to come from the group structure of the model. We illustrate this point by considering the following triangular system:

$$\begin{aligned} y_{1_{it}} &= && + v_{1_{it}}, \\ y_{2_{it}} &= \gamma_{21}y_{1_{it}} && + v_{2_{it}}, \\ &\;\vdots && \\ y_{G_{it}} &= \gamma_{G1}y_{1_{it}} + \cdots + \gamma_{G,G-1}y_{G-1_{it}} && + v_{G_{it}}, \end{aligned} \tag{4.4.15}$$

where $v_{g_{it}}$ satisfy (4.4.1) and (4.4.2). We assume one additional $\gamma_{\ell k} = 0$ for some ℓ and $k, \ell > k$, for identification.

The reduced form of (4.4.15) is

$$y_{g_{it}} = a_g h_{it} + \epsilon_{g_{it}}, \qquad g = 1,\ldots,G, \tag{4.4.16}$$

where

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_G \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 + \gamma_{21} d_1 \\ d_3 + \gamma_{31} d_1 + \gamma_{32}(d_2 + \gamma_{21} d_1) \\ \vdots \end{bmatrix}, \tag{4.4.17}$$

$$\boldsymbol{\epsilon}_{it} = \begin{bmatrix} \epsilon_{1_{it}} \\ \epsilon_{2_{it}} \\ \epsilon_{3_{it}} \\ \vdots \\ \epsilon_{g_{it}} \\ \vdots \end{bmatrix} = \begin{bmatrix} u_{1_{it}} \\ u_{2_{it}} + \gamma_{21} u_{1_{it}} \\ u_{3_{it}} + \gamma_{31} u_{1_{it}} + \gamma_{32}(u_{2_{it}} + \gamma_{21} u_{1_{it}}) \\ \vdots \\ u_{g_{it}} + \sum_{k=1}^{g-1} \gamma_{gk}^* u_{k_{it}} \\ \vdots \end{bmatrix}, \tag{4.4.18}$$

where $\gamma_{gk}^* = \gamma_{gk} + \sum_{i=k+1}^{g-1} \gamma_{gi} \gamma_{ik}^*$ if $g > 1$ and $k + 1 < g$, and $\gamma_{gk}^* = \gamma_{gk}$ if $k + 1 = g$.

4.4.2.1 Instrumental-Variable Method

The trick of the purged instrumental-variable (IV) method is to leave h in the residual and construct instruments that are uncorrelated with h. Before going to the general formula, we use several simple examples to show where the instruments come from.

Consider the case that $G = 3$. Suppose $\gamma_{21} = \gamma_{31} = 0$. Using y_1 as a proxy for h in the y_3 equation, we have

$$y_{3_{it}} = \gamma_{32} y_{2_{it}} + \frac{d_3}{d_1} y_{1_{it}} + u_{3_{it}} - \frac{d_3}{d_1} u_{1_{it}}. \tag{4.4.19}$$

If $T \geq 2$, then $y_{1_{is}}, s \neq t$, is a legitimate instrument for $y_{1_{it}}$, because it is uncorrelated with $u_{3_{it}} - (d_3/d_1) u_{1_{it}}$ but it is correlated with $y_{1_{it}}$ provided that $d_1 \sigma_\alpha^2 \neq 0$. Therefore, we can use $(y_{2_{it}}, y_{1_{is}})$ as instruments to estimate (4.4.19).

Next, suppose that only $\gamma_{32} = 0$. The reduced form of the model becomes

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} &= \begin{bmatrix} d_1 \\ d_2 + \gamma_{21} d_1 \\ d_3 + \gamma_{31} d_1 \end{bmatrix} h_{it} + \begin{bmatrix} u_{1_{it}} \\ u_{2_{it}} + \gamma_{21} u_{1_{it}} \\ u_{3_{it}} + \gamma_{31} u_{1_{it}} \end{bmatrix} \\ &= \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} h_{it} + \begin{bmatrix} \epsilon_{1_{it}} \\ \epsilon_{2_{it}} \\ \epsilon_{3_{it}} \end{bmatrix}. \end{aligned} \tag{4.4.20}$$

In this case, the construction of valid instruments is more complicated. It requires two stages. The first stage is to use y_1 as a proxy for h in the reduced-form equation for y_2:

$$y_{2_{it}} = \frac{a_2}{a_1} y_{1_{it}} + \epsilon_{2_{it}} - \frac{a_2}{a_1} \epsilon_{1_{it}}. \tag{4.4.21}$$

Equation (4.4.21) can be estimated by using $y_{1_{is}}, s \neq t$, as an instrument for $y_{1_{it}}$, provided that $d_1 \sigma_\alpha^2 \neq 0$. Then form the residual, thereby purging y_2 of its dependence on h:

$$z_{2_{it}} = y_{2_{it}} - \frac{a_2}{a_1} y_{1_{it}} = \epsilon_{2_{it}} - \frac{a_2}{a_1} \epsilon_{1_{it}}. \tag{4.4.22}$$

The second stage is to use z_2 as an instrument for y_1 in the structural equation y_3:

$$y_{3_{it}} = \gamma_{31} y_{1_{it}} + d_3 h_{it} + u_{3_{it}}. \tag{4.4.23}$$

The variable z_2 is an appropriate IV because it is uncorrelated with h and u_3, but it is correlated with y_1, provided $d_2\sigma_1^2 \neq 0$. (If $d_2 = 0$, then $z_2 = y_2 - \gamma_{21} y_1 = u_2$. It is no longer correlated with y_1.) Therefore, we require that h appear directly in the y_2 equation and that y_1 not be proportional to h – otherwise, we could never separate the effects of y_1 and h.

In order to identify the y_2 equation

$$y_{2_{it}} = \gamma_{21} y_{1_{it}} + d_2 h_{it} + u_{2_{it}}, \tag{4.4.24}$$

we can interchange the reduced-form y_2 and y_3 equations and repeat the two stages. With γ_{21} and γ_{31} identified, in the third stage we form the residuals

$$\begin{aligned} v_{2_{it}} &= y_{2_{it}} - \gamma_{21} y_{1_{it}} = d_2 h_{it} + u_{2_{it}}, \\ v_{3_{it}} &= y_{3_{it}} - \gamma_{31} y_{1_{it}} = d_3 h_{it} + u_{3_{it}}. \end{aligned} \tag{4.4.25}$$

Then we use y_1 as a proxy for h:

$$\begin{aligned} v_{2_{it}} &= \frac{d_2}{d_1} y_{1_{it}} + u_{2_{it}} - \frac{d_2}{d_1} u_{1_{it}}, \\ v_{3_{it}} &= \frac{d_3}{d_1} y_{1_{it}} + u_{3_{it}} - \frac{d_3}{d_1} u_{1_{it}}. \end{aligned} \tag{4.4.26}$$

Now d_2/d_1 and d_3/d_1 can be identified by a third application of instrumental variables, using $y_{1_{is}}, s \neq t$, as an instrument for $y_{1_{it}}$. (Note that only the ratio of the d's is identified, because of the indeterminate scale of the latent variable.)

Now let's come back to the construction of IVs for the general system (4.4.15)–(4.4.18). We assume that $T \geq 2$. The instruments are constructed over several stages. At the first stage, let y_1 be a proxy for h. Then the reduced-form equation for y_g becomes

$$y_{g_{it}} = \frac{a_g}{a_1} y_{1_{it}} + \epsilon_{g_{it}} - \frac{a_g}{a_1} \epsilon_{1_{it}}, \qquad g = 2, \ldots, \ell - 1. \tag{4.4.27}$$

If $T \geq 2$, then a_g/a_1 can be consistently estimated by using different members in the same group (e.g., $y_{1_{is}}$ and $y_{1_{it}}, t \neq s$) as instruments for the y_g equation (4.4.27) when $d_1\sigma_\alpha^2 \neq 0$. Once a_g/a_1 is consistently estimated, we form the residual

$$z_{g_{it}} = y_{g_{it}} - \frac{a_g}{a_1} y_{1_{it}} = \epsilon_{g_{it}} - \frac{a_g}{a_1} \epsilon_{1_{it}}, \qquad g = 2, \ldots, \ell - 1. \tag{4.4.28}$$

The z_g are uncorrelated with h. They are valid instruments for y_g provided $d_g\sigma_1^2 \neq 0$. There are $\ell - 2$ IVs for the $\ell - 2$ variables that remain on the right-hand side of the ℓth structural equation after y_k has been excluded.

To estimate the equations that follow y_ℓ, we form the transformed variables

$$\begin{aligned} y_{2_{it}}^* &= y_{2_{it}} - \gamma_{21} y_{1_{it}}, \\ y_{3_{it}}^* &= y_{3_{it}} - \gamma_{31} y_{1_{it}} - \gamma_{32} y_{2_{it}}, \\ &\vdots \\ y_{\ell_{it}}^* &= y_{\ell_{it}} - \gamma_{\ell 1} y_{1_{it}} - \cdots - \gamma_{\ell,\ell-1} y_{\ell-1_{it}}, \end{aligned} \tag{4.4.29}$$

and rewrite the $y_{\ell+1}$ equation as

$$y_{\ell+1_{it}} = \gamma^*_{\ell+1,1} y_{1_{it}} + \gamma^*_{\ell+1,2} y^*_{2_{it}} + \cdots + \gamma^*_{\ell+1,\ell-1} y^*_{\ell-1_{it}} + \gamma_{\ell+1,\ell} y^*_{\ell_{it}} \\ + d_{\ell+1} h_{it} + u_{\ell+1_{it}}, \tag{4.4.30}$$

where $\gamma^*_{\ell+1,j} = \gamma_{\ell+1,j} + \sum_{m=j+1}^{\ell} \gamma_{\ell+1,m} \gamma^*_{mj}$ for $j < \ell$. Using y_1 as a proxy for h, we have

$$y_{\ell+1_{it}} = \gamma^*_{\ell+1,2} y^*_{2_{it}} + \cdots + \gamma_{\ell+1,\ell} y^*_{\ell_{it}} \\ + \left(\gamma^*_{\ell+1,1} + \frac{d_{\ell+1}}{d_1} \right) y_{1_{it}} + u_{\ell+1_{it}} - \frac{d_{\ell+1}}{d_1} u_{1_{it}}, \tag{4.4.31}$$

Because u_1 is uncorrelated with y^*_g for $2 \leq g \leq \ell$, we can use $y^*_{g_{it}}$ together with $y_{1_{is}}, s \neq t$ as instruments to identify $\gamma_{\ell+1,j}$. Once $\gamma_{\ell+1,j}$ are identified, we can form $y^*_{\ell+1} = y_{\ell+1} - \gamma_{\ell+1,1} y_1 - \cdots - \gamma_{\ell+1,\ell} y_\ell$, proceed in a similar fashion to identify the $y_{\ell+2}$ equation, and so on.

Once all the γ are identified, we can form the estimated residuals, $\hat{\mathbf{v}}_{it}$. From $\hat{\mathbf{v}}_{it}$ we can estimate d_g/d_1 by the same procedure as (4.4.26). Or we can form the matrix $\hat{\Omega}$ of variance–covariances of the residuals, and the matrix $\hat{\bar{\Omega}}$ of variance–covariances of averaged residuals $(1/T)\sum_{t=1}^{T} \hat{\mathbf{v}}_{it}$, and then solve for $\mathbf{d}$, $(\sigma_1^2, \ldots, \sigma_G^2)$, and σ_ω^2 from the relations

$$\hat{\Omega} = \left(1 + \sigma_\omega^2\right) \mathbf{dd}' + \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_G^2\right), \tag{4.4.32}$$

$$\hat{\bar{\Omega}} = \left(1 + \sigma_\omega^2\right) \mathbf{dd}' + \frac{1}{T} \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_G^2\right). \tag{4.4.33}$$

The purged IV estimator is consistent. It also will often indicate quickly if a new model is identified. For instance, to see the necessity of having at least one more $\gamma_{g\ell} = 0$ for $g > \ell$ to identify the foregoing system, we can check if the instruments formed by the foregoing procedure satisfy the required rank condition. Consider the example where $G = 3$ and all $\gamma_{g\ell} \neq 0$ for $g > \ell$. In order to follow the strategy of allowing h to remain in the residual, in the third equation we need IVs for y_1 and y_2 that are uncorrelated with h. As indicated earlier, we can purge y_2 of its dependence on h by forming $z_2 = y_2 - (a_2/a_1) y_1$. A similar procedure can be applied to y_1. We use y_2 as a proxy for h, with $y_{2_{is}}$ as an IV for $y_{2_{it}}$. Then we form the residual $z_1 = y_1 - (a_1/a_2) y_2$. Again, z_1 is uncorrelated with h and u_3. But $z_1 = -(a_1/a_2) z_2$, and so an attempt to use both z_2 and z_1 as IVs fails to meet the rank condition.

4.4.2.2 Maximum-Likelihood Method

Although the purged IV method is simple to use, it is likely to be inefficient, because the correlations between the endogenous variables and the purged IVs will probably be small. Also, the restriction that (4.4.6) is of rank 1 is not being utilized. To obtain efficient estimates of the unknown parameters, it is necessary to estimate the covariance matrices simultaneously with the equation coefficients. Under the normality assumptions for α_i, ω_{it}, and u_{it}, we can obtain efficient estimates of (4.4.15) by maximizing the log likelihood function

$$\log L = -\frac{N}{2} \log |V| - \frac{1}{2} \sum_{i=1}^{N} (\mathbf{y}'_{1i}, \mathbf{y}'_{2i}, \ldots, \mathbf{y}'_{Gi}) V^{-1} (\mathbf{y}'_{1i}, \ldots, \mathbf{y}'_{Gi})', \tag{4.4.34}$$

where

$$\begin{aligned}
\underset{T\times 1}{\mathbf{y}_{gi}} &= (y_{gi1}, \ldots, y_{giT})', \qquad g = 1, \ldots, G, \\
\underset{GT\times GT}{V} &= \Lambda \otimes I_T + \mathbf{aa}' \otimes \mathbf{e}_T \mathbf{e}'_T, \qquad (4.4.35) \\
\underset{G\times G}{\Lambda} &= E(\boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}'_{it}) + \sigma^2_\omega \mathbf{aa}'.
\end{aligned}$$

Using the relations[10]

$$V^{-1} = \Lambda^{-1} \otimes I_T - \mathbf{cc}' \otimes \mathbf{e}_T \mathbf{e}'_T, \tag{4.4.36}$$

$$|V| = |\Lambda|^T |1 - T\mathbf{c}'\Lambda\mathbf{c}|^{-1}, \tag{4.4.37}$$

we can simplify the log likelihood function as[11]

$$\log L = -\frac{NT}{2} \log |\Lambda| + \frac{N}{2} \log(1 - T\mathbf{c}'\Lambda\mathbf{c}) - \frac{NT}{2} \operatorname{tr}(\Lambda^{-1} R) + \frac{NT^2}{2} \mathbf{c}' \bar{R} \mathbf{c}, \tag{4.4.38}$$

where $\mathbf{c}$ is a $G \times 1$ vector proportional to $\Lambda^{-1}\mathbf{a}$, R is the matrix of the sums of the squares and cross-products of the residuals divided by NT, and $\bar{R}$ is the matrix of sums of squares and cross-products of the averaged residuals (over t for i) divided by N. In other words, we simplify the log likelihood function (4.4.34) by reparameterizing it in terms of $\mathbf{c}$ and Λ.

Taking partial derivatives of (4.4.38), we obtain the first-order conditions[12]

$$\frac{\partial \log L}{\partial \Lambda^{-1}} = \frac{NT}{2}\Lambda + \frac{NT}{2} \frac{1}{(1 - T\mathbf{c}'\Lambda\mathbf{c})} \Lambda\mathbf{cc}'\Lambda - \frac{NT}{2} R = \mathbf{0}, \tag{4.4.39}$$

$$\frac{\partial \log L}{\partial \mathbf{c}} = -\frac{NT}{1 - T\mathbf{c}'\Lambda\mathbf{c}} \Lambda\mathbf{c} + NT^2 \bar{R}\mathbf{c} = \mathbf{0}. \tag{4.4.40}$$

Postmultiplying (4.4.39) by $\mathbf{c}$ and regrouping the terms, we have

$$\Lambda\mathbf{c} = \frac{1 - T\mathbf{c}'\Lambda\mathbf{c}}{1 - (T-1)\mathbf{c}'\Lambda\mathbf{c}} R\mathbf{c}. \tag{4.4.41}$$

[10] For the derivations of (4.4.36) and (4.4.37), see Appendix 4A.

[11] From $V \cdot V^{-1} = I_{GT}$, we have $-\Lambda \boldsymbol{cc}' - T\boldsymbol{aa}'\boldsymbol{cc}' + \boldsymbol{aa}'\Lambda^{-1} = \mathbf{0}$. Premultiplying this equation by $\boldsymbol{c}'$, we obtain $(b_1 + Tb_2^2)\boldsymbol{c}' = b_2\boldsymbol{a}'\Lambda^{-1}$, where $b_1 = \boldsymbol{c}'\Lambda\boldsymbol{c}$ and $b_2 = \boldsymbol{c}'\boldsymbol{a}$. In Appendix 4A we give the values of b_1 and b_2 explicitly in terms of the eigenvalue of $| \boldsymbol{aa}' - \lambda\Lambda | = 0$.

[12] We make use of the formula $\partial \log | \Lambda | / \partial\Lambda^{-1} = -\Lambda'$ and $\partial(\boldsymbol{c}'\Lambda\boldsymbol{c})/\partial\Lambda^{-1} = -\Lambda\boldsymbol{cc}'\Lambda$ (Theil 1971, pp. 32, 33).

Combining (4.4.40) and (4.4.41), we obtain

$$\left(\bar{R} - \frac{1}{T[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]}R\right)\mathbf{c} = \mathbf{0}. \tag{4.4.42}$$

Hence, the MLE of $\mathbf{c}$ is a characteristic vector corresponding to a root of

$$|\bar{R} - \lambda R| = 0. \tag{4.4.43}$$

The determinate equation (4.4.43) has G roots. To find which root to use, substitute (4.4.39) and (4.4.40) into (4.4.38):

$$\begin{aligned}\log L &= -\frac{NT}{2}\log|\Lambda| + \frac{N}{2}\log(1 - T\mathbf{c}'\Lambda\mathbf{c}) \\ &\quad - \frac{NT}{2}(G + T \operatorname{tr} \mathbf{c}'\bar{R}\mathbf{c}) + \frac{NT^2}{2}\operatorname{tr}(\mathbf{c}'\bar{R}\mathbf{c}) \\ &= -\frac{NT}{2}\log|\Lambda| + \frac{N}{2}\log(1 - T\mathbf{c}'\Lambda\mathbf{c}) - \frac{NTG}{2}. \end{aligned}\tag{4.4.44}$$

Let the G characteristic vectors corresponding to the G roots of (4.4.43) be denoted as $\mathbf{c}_1(=\mathbf{c}), \mathbf{c}_2, \ldots, \mathbf{c}_G$. These characteristic vectors are determined only up to a scalar. Choose the normalization $\mathbf{c}_g^{*\prime}R\mathbf{c}_g^* = 1, g = 1, \ldots, G$, where $\mathbf{c}_g^* = (\mathbf{c}_g' R\mathbf{c}_g)^{-1/2}\mathbf{c}_g$. Let $C^* = [\mathbf{c}_1^*, \ldots, \mathbf{c}_G^*]$; then $C^{*\prime}RC^* = I_G$. From (4.4.39) and (4.4.41), we have

$$\begin{aligned} C^{*\prime}\Lambda C^* &= C^{*\prime}RC^* - \frac{1 - T\mathbf{c}'\Lambda\mathbf{c}}{[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]^2}C^{*\prime}R\mathbf{c}\mathbf{c}'RC^* \\ &= I_G - \frac{1 - T\mathbf{c}'\Lambda\mathbf{c}}{[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]^2} \\ &\quad \times \begin{bmatrix} (\mathbf{c}'R\mathbf{c})^{1/2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} [(\mathbf{c}'R\mathbf{c})^{1/2}, 0 \cdots 0]. \end{aligned}\tag{4.4.45}$$

Equation (4.4.41) implies that $(\mathbf{c}'R\mathbf{c}) = \{[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]/[1-T\mathbf{c}'\Lambda\mathbf{c}]\}\mathbf{c}'\Lambda\mathbf{c}$. Therefore, the determinant of (4.4.45) is $\{[1-T\mathbf{c}'\Lambda\mathbf{c}]/[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]\}$. Using $C^{*\prime-1}C^{*-1} = R$, we have $|\Lambda| = \{[1-T\mathbf{c}'\Lambda\mathbf{c}]/[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]\}|R|$. Substituting this into (4.4.44), the log likelihood function becomes

$$\begin{aligned}\log L &= -\frac{NT}{2}\{\log|R| + \log(1 - T\mathbf{c}'\Lambda\mathbf{c}) \\ &\quad - \log[1-(T-1)\mathbf{c}'\Lambda\mathbf{c}]\} \\ &\quad + \frac{N}{2}\log[1 - T\mathbf{c}'\Lambda\mathbf{c}] - \frac{NTG}{2}, \end{aligned}\tag{4.4.46}$$

which is positively related to $\mathbf{c}'\Lambda\mathbf{c}$ within the admissible range $(0, 1/T)$.[13] So the MLE of $\mathbf{c}$ is the characteristic vector corresponding to the largest root of (4.4.43). Once $\mathbf{c}$ is obtained, from Appendix 4A and (4.4.39) and (4.4.40), we can estimate $\mathbf{a}$ and Λ by

$$\mathbf{a}' = T(1 + T^2\mathbf{c}'\bar{R}\mathbf{c})^{-1/2}\mathbf{c}'\bar{R}, \tag{4.4.47}$$

[13] See Appendix 4A, equation (4A.7), in which ψ_1 is positive.

and

$$\Lambda = R - \mathbf{a}\mathbf{a}'. \tag{4.4.48}$$

Knowing $\mathbf{a}$ and Λ, we can solve for the coefficients of the joint dependent variables Γ.

When exogenous variables also appear in the equation, and with no restrictions on the coefficients of exogenous variables, we need only replace the exponential term of the likelihood function (4.4.34),

$$-\frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}'_{1i},\ldots,\mathbf{y}'_{Gi})V^{-1}(\mathbf{y}'_{1i},\ldots,\mathbf{y}'_{Gi})',$$

with

$$-\frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}'_{1i}-\pi'_1\mathbf{X}'_i,\ldots,\mathbf{y}'_{Gi}-\pi'_G X'_i)$$
$$\times V^{-1}(\mathbf{y}'_{1i}-\pi'_1 X'_i,\ldots,\mathbf{y}'_{Gi}-\pi'_G X'_i)'.$$

The MLEs of $\mathbf{c}$, $\mathbf{a}$, and Λ remain the solutions of (4.4.43), (4.4.47), and (4.4.48). From knowledge of Λ and $\mathbf{a}$, we can solve for Γ and σ^2_ω. The MLE of Π conditional on V is the GLS of Π. Knowing Π and Γ, we can solve for $B = -\Gamma\Pi$.

Thus, Chamberlain and Griliches (1975) suggested the following iterative algorithm to solve for the MLE. Starting from the least squares reduced-form estimates, we can form consistent estimates of R and $\bar{R}$. Then we estimate $\mathbf{c}$ by maximizing[14]

$$\frac{\mathbf{c}'\bar{R}\mathbf{c}}{\mathbf{c}'R\mathbf{c}}. \tag{4.4.49}$$

Once $\mathbf{c}$ is obtained, we solve for $\mathbf{a}$ and Λ by (4.4.47) and (4.4.48). After obtaining Λ and $\mathbf{a}$, the MLE of the reduced-form parameters is just the generalized least-squares estimate. With these estimated reduced-form coefficients, one can form new estimates of R and $\bar{R}$ and continue the iteration until the solution converges. The structural-form parameters are then solved from the convergent reduced-form parameters.

4.4.3 An Example

Chamberlain and Griliches (1975) used the Gorseline (1932) data of the highest grade of schooling attained (y_1), the logarithm of the occupational (Duncan's (1980) SES) standing (y_2), and the logarithm of 1927 income (y_3) for 156 pairs of brothers from Indiana (U.S.) to fit a model of the type (4.4.1)–(4.4.3). Specifically, they let

$$\begin{aligned} y_{1_{it}} &= \boldsymbol{\beta}'_1\mathbf{x}_{it} + d_1 h_{it} + u_{1_{it}},\\ y_{2_{it}} &= \gamma_{21}y_{1_{it}} + \boldsymbol{\beta}'_2\mathbf{x}_{it} + d_2 h_{it} + u_{2_{it}},\\ y_{3_{it}} &= \gamma_{31}y_{1_{it}} + \boldsymbol{\beta}'_3\mathbf{x}_{it} + d_3 h_{it} + u_{3_{it}}. \end{aligned} \tag{4.4.50}$$

The set X contains a constant, age, and age squared, with age squared appearing only in the income equation.

[14] Finding the largest root of (4.4.43) is equivalent to maximizing (4.4.49). If we normalize $\boldsymbol{c}'R\boldsymbol{c} = 1$, then to find the maximum of (4.4.49) we can use Lagrangian multipliers and maximize $\boldsymbol{c}'\bar{R}\boldsymbol{c} + \lambda(1 - \boldsymbol{c}'R\boldsymbol{c})$. Taking partial derivatives with respect to $\boldsymbol{c}$ gives $(\bar{R} - \lambda R)\boldsymbol{c} = \mathbf{0}$. Premultiplying by $\boldsymbol{c}'$, we have $\boldsymbol{c}'\bar{R}\boldsymbol{c} = \lambda$. Thus, the maximum of (4.4.49) is the largest root of $|\bar{R} - \lambda R| = 0$, and $\boldsymbol{c}$ is the characteristic vector corresponding to the largest root.

The reduced form of (4.4.50) is

$$\mathbf{y}_{it} = \Pi \mathbf{x}_{it} + \mathbf{a} h_{it} + \epsilon_{it}, \tag{4.4.51}$$

where

$$\Pi = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \gamma_{21}\boldsymbol{\beta}_1' + \boldsymbol{\beta}_2' \\ \gamma_{31}\boldsymbol{\beta}_1' + \boldsymbol{\beta}_3' \end{bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} d_1 \\ d_2 + \gamma_{21} d_1 \\ d_3 + \gamma_{31} d_1 \end{bmatrix}, \tag{4.4.52}$$

$$\boldsymbol{\epsilon}_{it} = \begin{bmatrix} u_{1_{it}} \\ u_{2_{it}} + \gamma_{21} u_{1_{it}} \\ u_{3_{it}} + \gamma_{31} u_{1_{it}} \end{bmatrix}.$$

Therefore,

$$E\epsilon_{it}\epsilon_{it}' = \begin{bmatrix} \sigma_{u1}^2 & \gamma_{21}\sigma_{u1}^2 & \gamma_{31}\sigma_{u1}^2 \\ & \sigma_{u2}^2 + \gamma_{21}^2\sigma_{u1}^2 & \gamma_{21}\gamma_{31}\sigma_{u1}^2 \\ & & \sigma_{u3}^2 + \gamma_{31}^2\sigma_{u1}^2 \end{bmatrix}, \tag{4.4.53}$$

and

$$\Lambda = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{bmatrix} = E(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{it}') + \sigma_\omega^2 \mathbf{a}\mathbf{a}'. \tag{4.4.54}$$

We show that knowing $\mathbf{a}$ and Λ identifies the structural coefficients of the joint dependent variables as follows: For a given value of σ_ω^2, we can solve for

$$\sigma_{u1}^2 = \sigma_{11} - \sigma_\omega^2 a_1^2, \tag{4.4.55}$$

$$\gamma_{21} = \frac{\sigma_{12} - \sigma_\omega^2 a_1 a_2}{\sigma_{u1}^2}, \tag{4.4.56}$$

$$\gamma_{31} = \frac{\sigma_{13} - \sigma_\omega^2 a_1 a_3}{\sigma_{u1}^2}. \tag{4.4.57}$$

Equating

$$\gamma_{21}\gamma_{31} = \frac{\sigma_{23} - \sigma_\omega^2 a_2 a_3}{\sigma_{u1}^2} \tag{4.4.58}$$

with the product of (4.4.56) and (4.4.57), and making use of (4.4.55), we have

$$\sigma_\omega^2 = \frac{\sigma_{12}\sigma_{13} - \sigma_{11}\sigma_{23}}{\sigma_{12}a_1a_3 + \sigma_{13}a_1a_2 - \sigma_{11}a_2a_3 - \sigma_{23}a_1^2}. \tag{4.4.59}$$

The problem then becomes one of estimating $\mathbf{a}$ and Λ. Chamberlain and Griliches (1975) provided the MLE estimate of the coefficient of schooling and (unobservable) ability variables with σ_α^2 normalized to equal 1. They also provided the least squares estimates without including familial information as explanatory variables; and the covariance estimates in which each brother's characteristics (his income, occupation, schooling, and age) are measured around his own family's mean. The covariance estimate of the coefficient-of-schooling variable in the income equation is smaller than the least squares estimate (0.080 vs. 0.082). However, the simultaneous-equations model estimate of the coefficient for the ability variable is negative in the schooling equation (−0.092). As discussed in Section 4.1, if schooling and ability are negatively correlated, the single-equation within-family estimate of the schooling coefficient could be less than the least squares estimate. To attribute this decline to "ability" or "family background" is erroneous. In fact, when schooling and ability were treated symmetrically, the coefficient-of-schooling variable (0.088 with standard error 0.009) became greater than the least squares estimate 0.082.

APPENDIX 4A THE DETERMINANT AND INVERSE OF THE TRIANGULAR SYSTEM COVARIANCE MATRIX

Let

$$V = \Lambda \otimes I_T + \mathbf{a}\mathbf{a}' \otimes \mathbf{e}_T \mathbf{e}_T'. \tag{4A.1}$$

Because Λ is positive definite and $\mathbf{a}\mathbf{a}'$ is positive semidefinate, there exists a $G \times G$ nonsingular matrix F such that (Anderson 1985, p. 341)

$$F'\Lambda F = I_G \text{ and } F'\mathbf{a}\mathbf{a}'F = \begin{bmatrix} \psi_1 & & & \mathbf{0} \\ & 0 & & \\ & & \ddots & \\ \mathbf{0} & & & 0 \end{bmatrix},$$

where ψ_1 is the root of

$$\left|\mathbf{a}\mathbf{a}' - \lambda\Lambda\right| = 0. \tag{4A.2}$$

Next, choose a $T \times T$ orthogonal matrix E, with the first column of E being the vector $(1/\sqrt{T})\mathbf{e}_T$. Then

$$E'E = I_T \quad \text{and} \quad E'\mathbf{e}_T\mathbf{e}_T'E = \begin{bmatrix} T & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{4A.3}$$

Now $F \otimes E$ can be used to diagonalize V,

$$(F \otimes E)'V(F \otimes E) = I_{GT} + \underbrace{\begin{bmatrix} \psi_1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{G \times G} \otimes \underbrace{\begin{bmatrix} T & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{T \times T}, \tag{4A.4}$$

and factor V^{-1},

$$\begin{aligned} V^{-1} &= \Lambda^{-1} \otimes I_T - F' \begin{bmatrix} \frac{\psi_1}{1+T\psi_1} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} F \otimes \mathbf{e}_T\mathbf{e}_T' \\ &= \Lambda^{-1} \otimes I_T - \mathbf{c}\mathbf{c}' \otimes \mathbf{e}_T\mathbf{e}_T', \end{aligned} \tag{4A.5}$$

where $\mathbf{c}' = \left[\psi_1/(1 + T\psi_1)\right]^{\frac{1}{2}} \mathbf{f}_1'$, and $\mathbf{f}_1$ is the first column of F.

The determinant of V can be obtained from (4A.4):

$$|V| = |\Lambda|^T \cdot (1 + T\psi_1). \tag{4A.6}$$

This can be expressed in terms of $\mathbf{c}$ and Λ by noting that

$$\mathbf{c}'\Lambda\mathbf{c} = \frac{\psi_1}{1 + T\psi_1}. \tag{4A.7}$$

Thus, we have

$$1 - T\mathbf{c}'\Lambda\mathbf{c} = \frac{1}{1 + T\psi_1}, \tag{4A.8}$$

and

$$|V| = |\Lambda|^T \cdot (1 - T\mathbf{c}'\Lambda\mathbf{c})^{-1}. \tag{4A.9}$$

From $V \cdot V^{-1} = I_{GT}$, it is implied that

$$-\Lambda\mathbf{c}\mathbf{c}' + \mathbf{a}\mathbf{a}'\Lambda^{-1} - T\mathbf{a}\mathbf{a}'\mathbf{c}\mathbf{c}' = 0. \tag{4A.10}$$

Premultiplying (4A.10) by $\mathbf{c}'$, we obtain

$$\mathbf{a} = \frac{\mathbf{c}'\mathbf{a}}{\mathbf{c}'\Lambda\mathbf{c} + (\mathbf{c}'\mathbf{a})^2}\Lambda\mathbf{c}. \tag{4A.11}$$

Also, from $\mathbf{f}_1'\mathbf{a} = \psi^{\frac{1}{2}}$ and $\mathbf{a}$ proportional to $\mathbf{c}_1$ (Equation 4A.11) and hence $\mathbf{f}_1$, we have

$$\mathbf{a} = \frac{\psi^{\frac{1}{2}}}{\mathbf{f}_1'\mathbf{f}_1}\mathbf{f}_1 = \frac{1}{(1 + T\psi_1)^{\frac{1}{2}}(\mathbf{c}\mathbf{c}')}\mathbf{c} \tag{4A.12}$$

Premultiplying (4.4.40) by $\mathbf{c}'$, we obtain

$$\mathbf{c}'\bar{R}\mathbf{c} = \frac{\mathbf{c}'\Lambda\mathbf{c}}{T(1 - T\mathbf{c}'\Lambda\mathbf{c})} = \frac{1}{T}\psi_1. \tag{4A.13}$$

Combining (4.4.40) with (4A.8), (4A.12), and (4A.13), and using $\Lambda\mathbf{f}_1 = \left(1 \big/ \mathbf{f}_1'\mathbf{f}_1\right)\mathbf{f}_1$, we obtain

$$\begin{aligned}
\bar{R}\mathbf{c} &= \frac{1}{T}(1 + T\psi_1)\Lambda\mathbf{c} \\
&= \frac{1}{T}(1 + T\psi_1)^{\frac{1}{2}}\mathbf{a} \\
&= \frac{1}{T}(1 + T^2\mathbf{c}'\bar{R}\mathbf{c})^{\frac{1}{2}}\mathbf{a}.
\end{aligned} \tag{4A.14}$$

From (4.4.39) and (4A.12), we have

$$\begin{aligned}
\Lambda &= R - \frac{1}{1 - T\mathbf{c}'\Lambda\mathbf{c}}\Lambda\mathbf{c}\mathbf{c}'\Lambda \\
&= R - \mathbf{a}\mathbf{a}'.
\end{aligned} \tag{4A.15}$$