System of Linear Equations

Li Gan

October 2009

<u>Linear Equation System without Endogeneity (Seemingly Unrelated Regressions):</u>

Example: A basic utility maximization problem:

Suppose we are interested in estimating household demand functions for a set of goods. The basic utility maximization problem is given by:

$$\max_{housing, food, closth} \quad U(housing, food, clothing; c)$$
$$s.t. \quad housing * houseprc + food * foodprc + clothing * clothprc + c = income$$

To characterize the solution of the problem, the demand function is a function of prices and income *y*:

housing = H(*houseprc, foodprc, clothprc; income*)
food = F(*houseprc, foodprc, clothprc; income*)
clothing = C(*houseprc, foodprc, clothprc; income*)

If we know the functional form the utility function U(·), then we can precisely obtain the functional forms of H(·), F(·), and C(·). In most cases, all three functions are likely to be nonlinear in prices and in y. So we may use the nonlinear least square to estimate the model. The empirical econometric model is:

housing = H(*houseprc, foodprc, clothprc; income;θ*) + u_1
food = F(*houseprc, foodprc, clothprc; income;θ*) + u_2
clothing = C(*houseprc, foodprc, clothprc; income; θ*) + u_3

However, if we do not know or we are not willing to make assumptions on the utility functional forms, then we consider a linear model as an approximation of the first-order Taylor expansion of any nonlinear model:

housing = β_{10} + β_{11} houseprc + β_{12} foodprc + β_{13} clothprc + β_{14} income + β_{15} size + .. u_1
food = β_{20} + β_{21} houseprc + β_{22} foodprc + β_{23} clothprc + β_{24} income + β_{25} size + .. u_2
clothing = β_{30} + β_{31} houseprc + β_{32} foodprc + β_{33} clothprc + β_{34} income + β_{35} size + .. u_3

After the model is estimated, it is possible to "recover" the implied direct and indirect functional form by Roy's identify. For example, the demand for housing can be derived based on the Roy's identify:

$$food = -\frac{\partial V / \partial foodprc}{\partial V / \partial y}$$

One solve for the differential equation to recover the functional form of the indirect utility function $V(\cdot)$. Hausman (1981, *AER*) argues that the importance of this method.

Notes:

(i) In the previous equation, we assume that all regressors are uncorrelated with errors. A system of equations with such an assumption is called **seemingly unrelated regression** (SUR) model.

(ii) If the model is estimated equation by equation, consistency only requires that u_g is uncorrelated with x_g for the gth equation, i.e., $\text{cov}(u_g, x_g) = 0$. This assumption is called the weak exogeneity. However, if we are interested in obtaining more efficient estimates by working with all three equations simultaneously, it is necessary to have the so-called the strong exogeneity, i.e., all regressors in all equations are uncorrelated with all errors in all equations, $\text{cov}(u_g, x_k) = 0$, for all g and k. Furthermore, there are correlations among error terms.

(iii) We include the family size in the model. We may also include, for example, more detailed family composition variables, including the number of children in certain age ranges, if the household lives in urban areas, and other variables. These variables are often called control variables (as a modification of the constant term.)

(iv) In the previous model, regressors in all three equations are the same. This is NOT necessary. In fact, in many cases regrssors are different across different models. Further, in the previous model, coefficients are different across different equations. Again this is NOT necessary. Some coefficients can be the same across different equations.

Example (Panel Data Model): Consider one of the most famous panel data sets: Panel Study of Income Dynamics (PSID): since 1969, we observe the set of about 10,000 same households year after year.

Suppose we are interested in knowing how savings vary:

$$\begin{aligned} \log(saving_{it}) &= a_0 + a_1 \log(income_{it}) + a_2 return_{it} + z_{it}\gamma + u_{it} \\ &\equiv x_{it}\beta + u_{it} \end{aligned}$$

where $return_{it}$ is the investment return, for example. It may depend on individuals because of different compositions in their investment portfolio. In this example, the strong exogeneity assumption is that:

$$\text{Cov}(x_{it}, u_{is}) = 0 \text{ for all } t \text{ and } s.$$

In such a model, however, it may be difficult to maintain the strong exogeneity assumption. For example, when the lagged dependent variable is present:

$$\log(saving_{it}) = a_0 + \rho \log(saving_{it-1}) + a_1 \log(income_{it}) + a_2 return_{it} + z_{it}\gamma + u_{it}$$

In this equation, the strong exogeneity assumption is violated, cov(x_{it}, u_{it-1}) ≠0.

In general, we write SUR models as following:

$$\begin{aligned} y_1 &= x_1\beta_1 + u_1 \\ y_2 &= x_2\beta_2 + u_2 \\ &\vdots \\ y_G &= x_G\beta_G + u_G \end{aligned},$$

where $E(x_g^{'} u_g) = 0$ for $g = 1, \ldots, G$. Note x_g is N×K_g, and β_g is K_g×1.

Write the system of equations in the matrix form: $y_i = (y_{i1}, y_{i2}, \cdots y_{iG})'$, and $u_i = (u_{i1}, u_{i2}, \cdots u_{iG})'$.

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iG} \end{pmatrix} = \begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & x_{iG} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iG} \end{pmatrix} \equiv X_i\beta + u_i$$

where $i = 1, \ldots, N$

- It is possible to write: $X_i = I_G \otimes x_i$
- Since y_i is G×1, and therefore y is NG×1.
- X_i is G×(K_1+K_1+ … + K_G). Define K=K_1+K_1+ … + K_G, and X is NG×K.
- β is (K_1+K_1+ … + K_G)×1 = K×1.

With the weak exogeneity assumption: E(x_g'u_g) = 0, the simple OLS estimator is consistent and asymptotically normal, similar to what we have before.

However, it is possible to improve efficiency of the estimator if correlations Cov(u_g, u_k) ≠ 0. The method is the Generalized Least Square, or GLS:

A transformation (by multiplying a matrix) of X_i would typically lead to a linear combination of X_i and u_g, and the weak exogeneity assumption would not be sufficient.

Generalized Least Squares (GLS)

Let $\Omega \equiv E(u_i u_i^{'})$, which is G×G.

The GLS estimator is to obtained by pre-multiplying $\Omega^{-1/2}$:

$$\Omega^{-1/2} y_i = \Omega^{-1/2} X_i \beta + \Omega^{-1/2} u_i, \quad \text{or:} \qquad y_i^* = X_i^* \beta + u_i^*$$

It is easy to show that $E\left(u_i^* u_i^{*}{}'\right) = I$

Now we need to show that consistency:

$$\beta^* = \beta + \left(N^{-1} \sum_{i=1}^{N} X_i' \Omega^{-1} X_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} X_i' \Omega^{-1} u_i \right) \tag{1}$$

It is therefore sufficient to show that: $E\left(X_i' \Omega^{-1} u_i\right) = 0$.

This is where the strong exogeneity assumption becomes necessary. In this equation, $X_i' \Omega^{-1}$ is a linear combination of x_{ig}, $g = 1, \ldots, G$. The weak exogeneity assumption is no longer sufficient for consistency.

For example, consider a two equation system, and let:

$$\Omega^{-1} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$$

Then:

$$E\left(X_i' \Omega^{-1} u_i\right) = E\left[\begin{pmatrix} x_{i1}' \omega_{11} & x_{i1}' \omega_{12} \\ x_{i2}' \omega_{12} & x_{i2}' \omega_{22} \end{pmatrix} \begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix}\right] = E\begin{pmatrix} x_{i1}' \omega_{11} u_{i1} + x_{i1}' \omega_{12} u_{i2} \\ x_{i2}' \omega_{12} u_{i1} + x_{i2}' \omega_{22} u_{i2} \end{pmatrix} = 0$$

Therefore, the weak exogeneity ensures that $E\left(x_{i1}' \omega_{11} u_{i1}\right) = 0$, and $E\left(x_{i2}' \omega_{22} u_{i2}\right) = 0$. But the strong exogeneity assumption is necessary to ensure $E\left(x_{i1}' \omega_{12} u_{i2}\right) = 0$ and $E\left(x_{i2}' \omega_{12} u_{i1}\right) = 0$.

However, if $\Omega^{-1} = \begin{pmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} \end{pmatrix}$, i.e., there are no correlations between the errors in two equations, the strong exogeneity is not necessary, and will not help.

From the (1), the asymptotic distribution of the estimator is given by:

$$\sqrt{N}\left(\beta^* - \beta\right) \rightarrow$$
$$N\left(0, \left(N^{-1} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} \hat{u}_i \hat{u}_i' \hat{\Omega}^{-1} X_i \right) \left(N^{-1} \sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \right)$$

where $\hat{\Omega} \equiv \frac{1}{N}\sum_{i=1}^{N}\hat{u}_i\hat{u}_i'$, where $\hat{u}_i$ is the residual from the OLS estimation.

Notes: (1) Note in general:

$$E\left(X_i'\Omega^{-1}u_iu_i'\Omega^{-1}X_i\right) \neq E\left(X_i'\Omega^{-1}X_i\right)$$

(2) To see how the equality would hold:

$$\begin{aligned}
&E\left(X_i'\Omega^{-1}u_iu_i'\Omega^{-1}X_i\right) = E\left[E\left(X_i'\Omega^{-1}u_iu_i'\Omega^{-1}X_i \mid X_i\right)\right] \\
&= E\left[X_i'\Omega^{-1}E\left(u_iu_i' \mid X_i\right)u_iu_i'\Omega^{-1}X_i\right] \\
&= E\left(X_i'\Omega^{-1}E\left(u_iu_i'\right)^{-1}\Omega^{-1}X_i\right) = E\left(X_i'\Omega^{-1}\Omega\Omega^{-1}X_i\right) \\
&= E\left(x_i'\Omega^{-1}x_i\right)
\end{aligned}.$$

$$\text{if } E\left(u_iu_i' \mid X_i\right) = E\left(u_iu_i'\right)$$

Therefore, to write this model into the usual GLS or FGLS format, additional assumption $E\left(u_iu_i' \mid X_i\right) = E\left(u_iu_i'\right)$ is necessary.

SUR Revisited:

(1) If Ω is a diagonal matrix, i.e., there are no correlations among error terms, GLS and OLS are the same.

(2) If $x_{i1} = x_{i2} = \ldots = x_{iG}$, i.e., if the same regressors show up in each equations (for all observations), then OLS equation by equation and FGLS are identical.

Example: A model of the demand for inputs in the production process.

Consider a firm who minimizes its cost:

$$\begin{aligned}
\max_{K_i, L_i, M_i} \quad & C = p_{iK}K + p_{iL}L + p_{iM}M \\
s.t. \quad & Y_i = F\left(K_i, L_i, M_i\right)
\end{aligned}$$

Solving this equation to get a demand for *K*, *L* and *M* for firm *i*.

$$\begin{aligned}
K_i^* &= k\left(Y_i; p_{ik}, p_{iL}, p_{iM}\right) \\
L_i^* &= l\left(Y_i; p_{ik}, p_{iL}, p_{iM}\right) \\
M_i^* &= m\left(Y_i; p_{ik}, p_{iL}, p_{iM}\right)
\end{aligned}$$

The total cost of production is given by the cost function:

$$C_i^* = p_{ik} K_i^* + p_{iL} L_i^* + p_{iM} M_i^*$$

If there are constant return to scale, then

$$C_i = Y_i c(p_{ik}, p_{iL}, p_{iM}), \text{ or } \ln C_i = \ln Y_i + \ln c(p_{ik}, p_{iL}, p_{iM}),$$

where $c(p_{ik}, p_{iL}, p_{iM})$ is the unit or average cost function.

The cost minimization factor demands are obtained by applying Shephard's lemma,

$$K_i^* = \frac{\partial C(Y_i; p_{ik}, p_{iL}, p_{iM})}{\partial p_{iK}}$$

With constant returns to scale,

$$s_{ik} \equiv \frac{p_{ik} K_i^*}{C(Y_i; p_{ik}, p_{iL}, p_{iM})} = \frac{p_{ik}}{C(Y_i; p_{ik}, p_{iL}, p_{iM})} \cdot \frac{\partial C(Y_i; p_{ik}, p_{iL}, p_{iM})}{\partial p_{iK}}$$
$$= \frac{p_{ik} Y_i}{Y_i c(p_{ik}, p_{iL}, p_{iM})} \frac{\partial c(p_{ik}, p_{iL}, p_{iM})}{\partial p_{iK}} = \frac{\partial \ln c(p_{ik}, p_{iL}, p_{iM})}{\partial \ln p_{iK}}$$

By expanding $\ln c(p_{ik}, p_{iL}, p_{iM})$ in a second-order Taylor series at the point $lnp_{ij} = 0$, $j = K, L, M$, we obtain:

$$\ln c \approx \beta_0 + \sum_{j=K,L,M} \left(\frac{\partial \ln c}{\partial \ln p_j} \right) \ln p_j + \frac{1}{2} \sum_{k=K,L,M} \sum_{j=K,L,M} \left(\frac{\partial^2 \ln c}{\partial \ln p_j \partial \ln p_k} \right) \ln p_j \ln p_k$$

where all derivatives are evaluated at the expansion point, i.e., $lnp_{ij} = 0$.

If we let the derivatives be coefficients (parameters), then we have:

$$\ln c \approx \beta_0 + \sum_{j=K,L,M} \beta_j \ln p_j + \frac{1}{2} \sum_{k=K,L,M} \sum_{j=K,L,M} \gamma_{jk} \ln p_j \ln p_k$$

This is the transcendental logarithmic, or translog, cost function. Note in the previous equation, it is necessary the case that $\gamma_{jk} = \gamma_{kj}$ because $\ln p_j \ln p_k = \ln p_k \ln p_j$.

The cost share equations are given by:

$$s_{iK} = \gamma_K + \gamma_{KK} \ln p_{iK} + \gamma_{KL} \ln p_{iL} + \gamma_{KM} \ln p_{iM}$$
$$s_{iL} = \gamma_L + \gamma_{KL} \ln p_{iK} + \gamma_{LL} \ln p_{iL} + \gamma_{LM} \ln p_{iM}$$
$$s_{iM} = \gamma_M + \gamma_{KM} \ln p_{iK} + \gamma_{LM} \ln p_{iL} + \gamma_{MM} \ln p_{iM}$$

Note because shares are used, we must have:

$$s_{iK} + s_{iL} + s_{iM} = 1 \Rightarrow$$

$$(\gamma_K + \gamma_L + \gamma_M) + \ln p_{ik}(\gamma_{KK} + \gamma_{KL} + \gamma_{KM}) +$$
$$\ln p_{iL}(\gamma_{KL} + \gamma_{LL} + \gamma_{LM}) + \ln p_{iM}(\gamma_{KM} + \gamma_{LM} + \gamma_{MM}) = 0$$

Because this is true for all i, it is therefore necessarily true that:

$$\gamma_K + \gamma_L + \gamma_M = 1$$
$$\gamma_{KK} + \gamma_{KL} + \gamma_{KM} = 0$$
$$\gamma_{KL} + \gamma_{LL} + \gamma_{LM} = 0$$
$$\gamma_{KM} + \gamma_{LM} + \gamma_{MM} = 0$$

With these restrictions, we can rewrite previous models as:

$$s_{iK} = \gamma_K + \gamma_{KK} \ln\left(\frac{p_{iK}}{p_{iM}}\right) + \gamma_{KL} \ln\left(\frac{p_{iL}}{p_{iM}}\right)$$
$$s_{iL} = \gamma_L + \gamma_{KL} \ln\left(\frac{p_{iK}}{p_{iM}}\right) + \gamma_{LL} \ln\left(\frac{p_{iL}}{p_{iM}}\right)$$

Further, you may want to add some control variables and an error term:

$$s_{iK} = z_i\delta_K + \gamma_{KK} \ln\left(\frac{p_{iK}}{p_{iM}}\right) + \gamma_{KL} \ln\left(\frac{p_{iL}}{p_{iM}}\right) + u_K$$
$$s_{iL} = z_i\delta_L + \gamma_{KL} \ln\left(\frac{p_{iK}}{p_{iM}}\right) + \gamma_{LL} \ln\left(\frac{p_{iL}}{p_{iM}}\right) + u_L$$

This two share equations consists a SUR where u_K and u_L are expected to be correlated.

Linear Equation System with Endogeneity: General Method of Moments (GMM)

Example: labor supply and wage offers.

$$h_i = \gamma_1 w_i + z_{i1}\delta_1 + u_{i1}$$
$$w_i = \gamma_2 h_i + z_{i2}\delta_2 + u_{i2}$$

In this model, it is obvious that E(w_i, u_{i1}) ≠ 0, and E(h_i, u_{i2}) ≠ 0.

In general, consider a system of equations:

$$\begin{aligned} y_1 &= x_1\beta_1 + u_1 \\ y_2 &= x_2\beta_2 + u_2 \\ &\vdots \\ y_G &= x_G\beta_G + u_G \end{aligned} , \qquad \text{where } E(x_g' u_g) \neq 0.$$

Again, write the system of equations in the matrix form. Define $y_i = (y_{i1}, y_{i2}, \cdots y_{iG})'$, and $u_i = (u_{i1}, u_{i2}, \cdots u_{iG})'$.

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iG} \end{pmatrix} = \begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & x_{iG} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iG} \end{pmatrix} \equiv X_i\beta + u_i$$

where i = 1, …, N

- Since y_i is Gx1, and therefore y is NGx1.
- X_i is Gx(K_1+K_1+ … + K_G). Define K=K_1+K_1+ … + K_G, and X is NGxK.
- β is (K_1+K_1+ … + K_G)x1 = Kx1.

A general approach to the endogeneity problem is to find a set of instruments z_g for equation g:

Assumption 1: E($z_g' u_g$) = 0, z_g is NxL_g.
Assumption 2: E($z_g' x_g$) ≠ 0, x_g is NxK_g, therefore E($z_g' x_g$) is L_gxK_g, and L_g≥K_g.

Alternatively, Assumption 2 can be rephrased as the rank condition: Rank E($z_g' x_g$) = K_g.

We already know how to estimate the model equation by equation consistently using 2SLS. Again the issue here is to improve efficiency.

Define the matrix of all instruments:

$$\begin{pmatrix} z_{i1} & 0 & \cdots & 0 \\ 0 & z_{i2} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & z_{iG} \end{pmatrix} = Z_i$$

where Z_i is $G\times(L_1+L_2+\ldots+L_g)\equiv G\times L$. It is possible to write $Z_i = I_G \otimes z_i$.

S-Assumption 1: $E\left(Z_i^{'}u_i\right)=0$.
S-Assumption 2: rank $E\left(Z_i^{'}X_i\right)=K$

Note: S-Assumption 1 requires that IVs at equation g, z_g, to be uncorrelated with all errors in all equations, analogous to the strong exogeneity condition in the SUR model.

S-Assumption 2 requires that each of the equation to be identifiable, $L_g \geq K_g$, for all g, and therefore, $L \geq K$.

S-Assumption 1 suggests comment conditions:

$$\mathrm{E}(Z_i^{'}u_i) = \mathrm{E}(Z_i^{'}(y_i - X_i\beta)) = 0$$

Suppose z is $L\times N$, and β is $K\times 1$, and $L\geq K$. The previous equation has L linear equations. Let the solution to the previous equation be given by:

$$\frac{1}{N}\sum_{i=1}^{N} Z_i^{'}\left(y_i - X_i\hat{\beta}\right)=0 \qquad (2)$$

When $L = K$, we can write previous equation to be:

$$\hat{\beta}=\left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{'}X_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{'}y_i\right)=\left(Z'X\right)^{-1}Z'Y$$

However, when $L > K$, we have more IVs than necessary. In other words, we have more equations than unknowns. The intuition is that we choose $\hat{\beta}$ such that the Lx1 vector in (2) as small as possible. One possibility is the squared term of (2) is as small as possible.

Example: In the simplest case, suppose we have two unknowns, a, and b, and two equations:

$$y_{1i} = a + bx_{1i}$$
$$y_{2i} = a + bx_{2i}$$

It is trivial to solve to a, and b, given x_{gi}, and y_{gi} where i = 1, 2, …, N, and g = 1, 2.

Now instead, we have y_{gi} and x_{gi}, g =1, 2, 3, such that:

$$y_{1i} = a + bx_{1i}$$
$$y_{2i} = a + bx_{2i}$$
$$y_{3i} = a + bx_{3i}$$

For this system of three equations and two unknowns, it is unlikely that we can find *a* and *b* such that all three equations are perfectly satisfied. A solution to this problem is to find *a* and *b* to minimize the following equation:

$$\sum_{i=1}^{N}\left(y_{1i}-a-bx_{1i}\right)^2+\left(y_{2i}-a-bx_{2i}\right)^2+\left(y_3-a-bx_{3i}\right)^2$$

In general, the objective is to find β, such that:

$$\min_{\beta}\quad\left(\sum_{i=1}^{N}Z_i'\left(y_i-X_i\beta\right)\right)^{/}W\left(\sum_{i=1}^{N}Z_i'\left(y_i-X_i\beta\right)\right)$$

The solution to this problem:

$$\begin{aligned}\hat{\beta}_{MM}&=\left(X'ZWZ'X\right)^{-1}X'ZWZ'Y\\&=\beta+\left(X'ZWZ'X\right)^{-1}X'ZWZ'u\end{aligned},$$

and the covariance of the estimator:

$$Var\left(\hat{\beta}_{MM}\right)=\left(X'ZWZ'X\right)^{-1}X'ZW\Lambda WZ'X\left(X'ZWZ'X\right)^{-1}$$

where $\Lambda = E(Z_i' u_i u_i' Z_i)$. This is a very long but intuitive covariance matrix. It is also easy to establish the asymptotic normality by using the central limit theorem.

One of the choices for W is given by:

$$\hat{W}=\left(\frac{1}{N}\sum_{i=1}^{N}Z_i'Z_i\right)^{-1}=\left(\frac{1}{N}Z'Z\right)^{-1}$$

We have:

$$\hat{\beta}_{MM}=\left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1}X'Z\left(Z'Z\right)^{-1}Z'Y$$

This is precisely the 2SLS, called pooled 2SLS estimator.

The Optimal Weighting Matrix and GMM:

The next step is to find the optimal weighting matrix W, which is the *GMM*.

It turns out, when $W = \Lambda^{-1}$, then the optimal covariance matrix is reached. In this case,

$$\hat{\beta}_{GMM}=\left(X'Z\Lambda^{-1}Z'X\right)^{-1}X'Z\Lambda^{-1}Z'Y\text{, and}$$

$$
\begin{aligned}
Var\left(\hat{\beta}_{GMM}\right) &= \left(X'Z\Lambda^{-1}Z'X\right)^{-1} X'Z\Lambda^{-1}\Lambda\Lambda^{-1}Z'X\left(X'Z\Lambda^{-1}Z'X\right)^{-1} \\
&= \left(X'Z\Lambda^{-1}Z'X\right)^{-1} X'Z\Lambda^{-1}Z'X\left(X'Z\Lambda^{-1}Z'X\right)^{-1} \\
&= \left(X'Z\Lambda^{-1}Z'X\right)^{-1}
\end{aligned}
$$

How to obtain $\hat{\beta}_{GMM}$ in practice?

a. Let $\hat{\hat{\beta}}$ be an initial consistent estimator of β. In most cases this is the 2SLS estimator.
b. Obtain the Gx1 residual vectors:

$$\hat{\hat{u}}_i = y_i - X_i\hat{\hat{\beta}}.$$

c. A generally consistent estimator of Λ is:

$$\hat{\Lambda} = \frac{1}{N}\sum_{i=1}^{N} Z_i'\hat{\hat{u}}_i\hat{\hat{u}}_i'Z_i$$

and choose: $\hat{W} = \hat{\Lambda}^{-1}$

Discussions:

(1) We develop the GMM with linear models. The moment conditions used here are $E\left(Z_i'u_i\right) = 0$. However, GMM is an estimator that is much more general than linear models or linear regressions. In fact, any expectations can serve as moment conditions. One may write a general form:

$$E\left[q\left(w_i;\theta\right)\right] = 0$$

where w_i represents the set of data, θ is the set of Kx1 parameters to be estimated, $q(w_i;\theta)$ is Lx1, with $L \geq K$.

Examples of $q(w_i;\theta)$:

(a) If $q(w_i;\theta)$ is the first-order conditions of the log-likelihood function, MLE can be considered as a special case of GMM.

(b) If $q(w_i;\theta)$ is the first order condition of $(y_i - X_i\beta)^2$, which is, $X_i'\left(y_i - X_i\beta\right)$, the OLS is a special case of GMM.

(c) One of the popular method to obtain moment conditions is by the first order conditions.

(2) Hansen and Singleton (1982, *Econometrica*).

Consider a maximization problem over time:

$$\underset{C_t}{Max} \quad E_0\left[\sum_{t=0}^{\infty}\beta^t U(C_t)\right]$$

where C_t is consumption in period t, β is a discount factor. Suppose the consumer has the choice of investing in a collection of N assets, with maturities M_j, j = 1, …, N. Let Q_{jt} denote the quantities of asset j held at t he end of period t, P_{jt} be the price of asset j at t, and R_{jt} be the date t payoff from holding a unit of M_j-period asset purchased at date t- M_j. The labor income is W_t. The budget set is given by:

$$C_t + \sum_{j=1}^{N} P_{jt}Q_{jt} \leq \sum_{j=1}^{N} R_{jt}Q_{jt-M_j} + W_t$$

The first order condition is given by:

$$P_{jt}U'(C_t) = \beta^{M_j} E_t\left[R_{jt+M_j}U'\left(C_{t+M_j}\right)\right], \qquad j = 1, \ldots, N,$$

Or:

$$E_t\left[\beta^{M_j}\frac{U'\left(C_{t+M_j}\right)}{U'(C_t)}\frac{R_{jt+M_j}}{P_{jt}} - 1\right] = 0$$

The previous equation suggests N moment conditions, with each moment condition for each asset j. Note the moment conditions work even if one does not have information on all assets but rather a subset of assets.

Note if asset j represents a stock or a one-year deposit, $M_j = 1$, then $R_{jt+1}/P_{jt} = 1 + r_{jt,}$ where r_{jt} is the one-year return of this stock. Further, let $U(C_t) = C_t^{1-\gamma}/(1-\gamma)$, we obtain a familiar model:

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_t) - 1\right] = 0$$

The parameters to be estimated are γ and β.

<u>The Three Stage Least Squares Estimator (3SLS)</u>

3SLS is a GMM estimator that uses a particular weighting matrix.

Assume that: E($Z_i' u_i u'_i Z_i$) = E($Z_i' \Omega Z_i$).

This equality will hold if E($u_i u_i'|Z_i$) = E($u_i u_i'$)

Define: $\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\hat{u}_i\hat{u}_i'$. The 3SLS is the GMM with the weighting matrix W being:

$$\hat{W} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'\hat{\Omega}Z_i\right)$$

Simultaneous Equation Models

Example: consider a model of demand and supply:

$$q^S = q = \gamma_1 p + z_1\delta_1 + u_1$$
$$q^D = q = \gamma_2 p + z_2\delta_2 + u_2$$

Suppose $\gamma_1 \neq \gamma_2$, i.e., the demand and supply curves have different slopes. The difference between the two equations, the second equation – the first equation, and we have:

$$p = \frac{\delta_1}{\gamma_2 - \gamma_1} z_1 - \frac{\delta_2}{\gamma_2 - \gamma_1} z_2 + \frac{u_1 - u_2}{\gamma_2 - \gamma_1}$$
$$= z_1\pi_{21} + z_2\pi_{22} + v_2$$

The identification of this model is rather simple and intuitive. The reduced form model of *p* serves as the instrumental regression for the two structural equations.

- To identify the supply equation, there is at least one element of z_2 that does not appear in z_1. For example, prices of other substitutable goods should be part of z_2 but not necessary in z_1.

- To identify the demand equation, there is at least one element of z_1 that does not appear in z_2. For example, prices of input (to produce this good) should be part of z_1 but not necessary in z_2.

Example: labor supply and demand for married women.

$$hours = \gamma_1 \log(wage) + \delta_{10} + \delta_{11}educ + \delta_{12}age + \delta_{13}kids + \delta_{14}othinc + u_1$$
$$hours = \gamma_2 \log(wage) + \delta_{20} + \delta_{21}educ + \delta_{22}expr + u_2$$

The first equation is the labor supply model while the second equation is the labor demand equation. The variables (*age*, *kids*, and *othinc*) appear in the labor supply equation but not in labor demand equation. These variables can serve as IVs for the wage in the labor demand equation. The variable (*expr*) appears in the labor demand but not in labor supply and can serve as the IV for the labor supply equation. However, if *expr* = *age* – *educ* – 6, as typically calculated, then it cannot be used as a valid IV.

More formally, consider a system of equations,

$$\begin{aligned} &y\gamma_1 + z\delta_1 + u_1 = 0 \\ &\vdots \\ &y\gamma_G + z\delta_G + u_G = 0 \end{aligned}$$

where $y = (y_1, ..., y_G)$, 1 x G, $\gamma_k = (\gamma_{k1}, ..., \gamma_{kG})'$, G x 1.
$z = (z_1, ..., z_G)$, 1 x M, $\delta_k = (\delta_{k1}, ..., \delta_{kM})'$, M x 1

Write in the matrix form,

$$\begin{aligned} &y\Gamma + z\Delta + u = 0 \\ &= (y \quad z)\begin{pmatrix}\Gamma \\ \Delta\end{pmatrix} = (y \quad z)B = 0 \end{aligned} \tag{3}$$

where $u = (y_1, ..., y_G)$, 1 x G,
Γ is G x G with gth column γ_g
Δ is M x G with gth column δ_g
B is $(M+G)$ x G

When the endogenous variables are represented in terms of exogenous variables, we call this type of models as the reduced form model. Here, the reduced form is given by:

$$y = z(-\Delta\Gamma^{-1}) + u(-\Gamma^{-1}) \equiv z\Pi + v$$

An unconstrained model such as (3) is not identified. To see this point, one may multiply (3) by a matrix of ***F***, the equality still holds.

Without loss of generality, we consider identification of the first equation:

$$y\gamma_1 + z\delta_1 + u_1 = 0$$

Or, $\gamma_{11}y_1 + \gamma_{12}y_2 + \cdots \gamma_{1G}y_G + \delta_{11}z_1 + \delta_{12}z_2 + \cdots + \delta_{1M}z_M + u_1 = 0$

Since the unconstrained model is not identified, it is necessary to add some constraints or restrictions to the model. Therefore, our problem becomes if a given restriction, denoted as R_1, is necessary and sufficient to identify the first equation.

The first restriction is normalization – one element of γ_1 is -1. Let:

$$\beta_1 = \begin{pmatrix} \gamma_1 \\ \delta_1 \end{pmatrix}$$

which is $(G+M)\times 1$. However, because of the normalization, there are only $G+M-1$ unknown elements in β_1.

Assume that the restrictions about β_1 can be expressed as:

$$R_1\beta_1 = 0$$

where R_1 is $J_1\times(G+M)$ matrix of known constants, and J_1 is the number of constraints on β_1 (in addition to the normalization restriction.

To understand if R_1 is sufficient to identify the model:

Define $\boldsymbol{F} = (f_1, f_2, .., f_G)$, and let $\beta_1^* = Bf_1$. We are interested in finding out: $\beta_1^* = \beta_1$ if and only if $f_1 = (1, 0, \ldots, 0)'$.

$$R_1\beta_1^* = (R_1B)f_1 = (R_1\beta_1, R_1\beta_2, \cdots, R_1\beta_G)f_1$$
$$= (0, R_1\beta_2, \cdots, R_1\beta_G)f_1$$

(i) When $\boldsymbol{R_1B}$ has a rank of G-1, then the only possibility that $R_1\beta_1^* = 0$ is $f_1 = (1, 0, \ldots, 0)'$ (subject to a constant).

(ii) When $f_1 = (1, 0, \ldots, 0)'$, it is obvious that $R_1\beta_1^* = 0$ if $\boldsymbol{R_1B}$ has a rank of G-1. Therefore,

Theorem: Rank $\boldsymbol{R_1B} = G-1$ is the sufficient and necessary condition for identification for the identification of $\boldsymbol{\beta_1}$.

Note since $\boldsymbol{R_1B}$ is $J_1\text{x}G$, and $\boldsymbol{R_1\beta_1} = \mathbf{0}$. Therefore, $\boldsymbol{R_1B}$ cannot have rank that is larger than G-1 because $\boldsymbol{R_1\beta_1} = \mathbf{0}$. Since Rank $R_1 = J_1$, then a necessary condition for the identification of $\boldsymbol{\beta_1}$ is $J_1 \geq G-1$.

To intuitively understand this model, J_1 suggests how many excluded exogenous variables, and G-1 is the number of endogenous variables (one endogenous variable, y_1, is now the dependent variable after the normalization). Therefore, when $J_1 \geq G_1$, the model is identified in most cases.

Check Identification:

(1) Set one element of J_1 to -1 as a normalization.
(2) Define the $J_1\text{x}(G+M)$ matrix R_1.
(3) If $J_1 < G-1$, then the first equation is not identified.
(4) If $J_1 \geq G-1$, the equation might be identified. Check the rank condition of $\boldsymbol{R_1B}$.

Example: ($G = 3$, $M = 4$):

$$
\begin{aligned}
y_1 &= \gamma_{12} y_2 + \gamma_{13} y_3 + \delta_{11} z_1 + \delta_{13} z_3 + u_1 \\
y_2 &= \gamma_{21} y_1 + \delta_{21} z_1 + u_2 \\
y_3 &= \delta_{31} z_1 + \delta_{32} z_z + \delta_{33} z_3 + \delta_{34} z_4 + u_3
\end{aligned}
$$

where $z_1 = 1$, $E(u_g) = 0$, $g = 1, 2, 3$, and each z_j is uncorrelated with u_g. Note the third equation is already a reduced form.

Consider the identification of equation (1). In this equation, $\gamma_{11} = -1$ (normalization), and two constraints, $\delta_{12} = 0$, and $\delta_{14} = 0$, i.e., $J_1 = 2$, but $G = 3$ ➔ $J_1 = G$-1. Therefore, the model is potentially identifiable.

Intuitively, there are two excluded exogenous variables, z_2 and z_4, and there are two endogenous variables, y_2 and y_3. The model satisfies the necessary condition for identifications.

To check the rank condition, let $\boldsymbol{\beta_1} = (-1, \gamma_{12}, \gamma_{13}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14})'$. The constraints $\delta_{12} = 0$, and $\delta_{14} = 0$ can be written as:

$$
R_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}
$$

Let $\boldsymbol{B}$ be the full 7x3 matrix of parameters with only the three normalizations imposed, so: $\boldsymbol{\beta_2} = (\gamma_{21}, -1, \gamma_{23}, \delta_{21}, \delta_{22}, \delta_{23}, \delta_{24})'$ and $\boldsymbol{\beta_3} = (\gamma_{31}, \gamma_{32}, -1, \delta_{31}, \delta_{32}, \delta_{33}, \delta_{34})'$.

$$
R_1 B = \begin{pmatrix} \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \delta_{32} \\ 0 & 0 & \delta_{34} \end{pmatrix}
$$

since from y_2 equation, we also have additional restrictions: $\delta_{22} = 0$, and $\delta_{24} = 0$.

Therefore, with all restrictions, the rank of $\boldsymbol{R_1 B}$ is 1. The model is not identified.

Intuition: Note in this example, although from y_1 equation, it looks like we have two excluded exogenous variables, z_2 and z_4 that may serve as IVs for y_2 and y_3. However, neither z_2 nor z_4 is correlated with y_2. So in fact no valid IVs exist for y_2. This is the reason that why locally that R_1 may be sufficient to identify the model but globally it may not be.

<u>Estimation after Identification</u>

To estimate the model, we rewrite the model in a form,

$$y_1 = y_{(1)}\gamma_{(1)} + z_1\delta_1 + u_1 = x_1\beta_1 + u_1$$
$$\vdots$$
$$y_G = y_{(G)}\gamma_{(G)} + z_G\delta_G + u_G = x_G\beta_G + u_G$$

where $y_{(1)}$ includes all y except y_1. This is also a normalized model.

Estimating this model is same as estimating previous models with endogenous variables – we use GMM.

Basic Panel Data Models

Many panel data sets are publicly available. The most well-known ones for economists include Panel Study of Income Dynamics (PSID), Health and Retirement Study (HRS), NLSY (National Longitudinal Study of Youth), and COMPUSTAT. The first three are at household or individual level, and the last one is at firm level.

Collecting panel data sets is typically a lot more expensive per observation than collecting cross section data sets. Compare with the cross section data sets, panel data sets have two distinguishing advantages.

(1) Dynamics. When studying dynamic behavior, using panel data sets is more appropriate than cross section data. Studying dynamics typically requires understanding the evolvement of some stock variables. For example, at household level, how wealth, human capital, and health capital evolve is typically very important in understanding many household behaviors. Studying these stock variables often requires following how these variables evolve over time. In this regard, observing the same households across time is very critical. At the firm level, it is even more important since almost all important factors such as capital stock, inventory stock, human capital, etc are stock variables.

(2) Endogeneity. Having observed the same household or firm multiple times can take care of one important source of endogeneity – the time-invarying unobserved heterogeneity. This can be seen more directly by the following example:

Consider a simple linear model:

$$E(BMI_i\,|x, c) = \alpha_0 + \alpha_0 * calorie\text{-}intake_i + X\beta + c_i. \qquad (4)$$

In (4), c_i represents unobserved factors such as genetics, for example.

If Cov(x,c) = 0, then OLS is consistent. If Cov(x,c) ≠ 0, then OLS is NOT consistent. We have two ways to solve this problem:

(i) using a proxy -- *BMI* of siblings, etc.

(ii) using the *IV* approach: $Cov(z, c) = 0$, and $Cov(z, x) \neq 0$, where z may include income, number of people in the family, # of kids in the family, etc.

However, if we observe the same person repeatedly, then more options arise. Suppose y and x are observed at two time periods.

$$E(y_t|X_t, c) = \beta_0 + X_t \beta + c, \qquad t = 1, 2 \tag{5}$$

Again, assume c is consistent (time invariant). Add an error term:

$$E(y_t \mid X_t, c) = \beta_0 + X_t \beta + c + u_t \tag{6}$$

By definition $E(X_t' u_t) = 0$. If we assume $E(X_t' c) = 0$, then OLS produces consistent estimates, otherwise no consistency.

Alternatively, if $E(X_t' c) \neq 0$, take the difference over time of (6):

$$\Delta y_t = \Delta X_t \beta + \Delta u_t \, .$$

We now can use OLS if: (a) $E(\Delta X_t' \Delta u_t) = 0$
(b) rank of $E(\Delta X_t' \Delta X_t)$ has full rank.

Condition (a) implies:

$$\begin{aligned} E\left(\Delta X_t' \Delta u_t\right) &= E\left(X_2' u_2\right) + E\left(X_1' u_1\right) - E\left(X_2' u_1\right) - E\left(X_1' u_2\right) \\ &= -E\left(X_2' u_1\right) - E\left(X_1' u_2\right) \end{aligned}$$

So $E(X_t' u_t) = 0$ cannot guarantee $E(\Delta X_t' \Delta u_t) = 0$. More conditions are necessary.

<u>Discussions:</u> consider the general panel model:

$$y_{it} = X_{it} \beta + c_i + u_{it}, \quad t = 1, 2, \ldots T, \text{ and } i = 1, 2, \ldots N. \tag{7}$$

(1) Random effect or fixed effect? (c_i)

Random effect : c_i is random. More precisely, $E(x_{it}' c_i) = 0$.
Fixed effect: c_i is an arbitrary constant. More precisely, $E(X_{it}' c_i) \neq 0$.

If random effect, c_i is drawn independent of x_{it}. Example: $N(0, \sigma_c^2)$.
If fixed effect, c_i can be drawn from a distribution of x_{it}. Example: $N(g(x_{it}), \sigma^2(x_{it}))$.

Note here in the fixed effect model, c_i could be random. Therefore, the difference between the so-called random effect model and the fixed effect model is NOT about randomness of the c_i. The difference is all about if c_i is correlated with x_{it}. More recently, the panel data model is often referred as *unobserved effects model*.

(2) Strict exogeneity of x_{it}:

It is possible for y_{it} to affect x_{it} for $s > t$. For example, in the example that studies the effect of debt level on inflation, inflation at t will affect debt level at later years.

Strict exogeneity:

$$E(y_{it}|x_{i1}, x_{i2}, \ldots x_{iT}) = E(y_{it}|x_{it}) = x_{it}\beta + c_i$$

This assumption implies:

$$E(u_{it}|x_{i1}, x_{i2}, \ldots x_{iT}, c_i) = 0, \text{ or } E(x_{it}' u_{is}) = 0, \text{ for any } i \text{ and } s.$$

This is a necessary condition for estimation.

(3) Failures of the strict exogeneity condition:

Example 1: Cross country data. Study the effect of debt on inflation.

$$CPI_{it} = X_{it}\beta + \gamma Debt_{it} + c_i + u_{it}$$

Problem: past debt level may affect future inflation level.

Example 2: Program evaluation, consider a work-training program, denoted as pwg_{it}:

$$\log(wage_{it}) = \theta_t + z_{it}\gamma + \delta\, pwg_{it} + c_i + u_{it}$$

Problem: participation in the program may not be random. It is possible that:

$$\text{Cov}(u_{it}, pwg_{it}) > 0.$$

Those who are more active in the labor are more likely to seek work training program.

Example 3: Distribution lag model:

.

$$patents_{it} = \theta_t + z_{it}\gamma + \sum_{\tau=0}^{k} \delta_\tau RD_{it-\tau} + c_i + u_{it}$$

Is RD_{it-j} correlated with today's u_{it}? A shock in patents may affect the earning ability of the firm, and hence affect future spending in R&D. Cov(u_{it},x_{it+1}) $\neq 0$, So strict exogeneity fails.

Example 4: Lagged dependent variable

$$\log(wage_{it}) = \theta \log(wage_{it-1}) + z_{it}\gamma + c_i + u_{it}.$$

In this case, since Cov(log($wage_{it}$),c_{it}) > 0, it must be the case that: Cov(x_{it+1},c_{it}) > 0, since x_{it+1} includes log($wage_{it}$). So the Strict Exogeneity condition fails.

Example 1 again: it may be useful to include lagged dependent variable in the model.

$$CPI_{it} = \sum_{\tau}^{K} \alpha_\tau CPI_{it-\tau} + X_{it}\beta + \gamma Debt_{it} + c_i + u_{it}$$

<u>Estimation: Random Effect Model</u>

The basic model is:

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

*Assumptions RE*1:

(a) E(u_{it}|x_{it}, c_i) = 0, where $t=1,\ldots\ldots,T$
(b) E(c_i|x_{it}) = 0, where $t=1,\ldots\ldots,T$

Rewrite (1) into: $y_t = X_t\beta + c_i j_T + u_t$, where j_T is the $Tx1$ vector of ones:

$$j_T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{Tx1}. \qquad \text{Let } v_i = \begin{pmatrix} c_i \\ \vdots \\ c_i \end{pmatrix} + \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix} = \begin{pmatrix} c_i + u_{i1} \\ \vdots \\ c_i + u_{iT} \end{pmatrix}$$

The model becomes: $y_i = x_i\beta + v_i$. Let $\Omega = E(v_i v_i')$.

Assumption RE2: Rank $E(x_i'\Omega^{-1}x_i)=K$.
Assumption RE3: Homoskedasticity, which can be written as:

$$E(u_i u_i' | x_i, c_i) = \sigma_c^2 I, \text{ and } E(c_i | x_i) = \sigma_c^2.$$

With these two assumptions, we have:

$$E\left(v_i^{'}v_i\right)=E\left[\begin{pmatrix} c_i+u_{i1} \\ \vdots \\ c_i+u_{iT}\end{pmatrix}\begin{pmatrix} c_i+u_{i1} & \cdots & c_i+u_{i1}\end{pmatrix}\right]$$
$$=\begin{pmatrix} \sigma_c^2+\sigma_u^2 & \cdots & \sigma_c^2 \\ \vdots & \ddots & \vdots \\ \sigma_c^2 & \cdots & \sigma_c^2+\sigma_u^2 \end{pmatrix}\equiv\Omega$$

Therefore, in the model $y_i = x_i\beta + v_i$, the error term v_i is no longer iid. An estimator that can do better than *OLS* is *GLS* (generalized least square).

Consider $y_i = x_i\beta + v_i$, (8) or:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N\end{pmatrix}=\begin{pmatrix} x_1 \\ \vdots \\ x_N\end{pmatrix}\beta+\begin{pmatrix} v_1 \\ \vdots \\ v_N\end{pmatrix} \qquad (8')$$

The covariance matrix of (8) is given by

$$\Lambda=\begin{pmatrix} \Omega & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Omega\end{pmatrix}$$

$$\hat{\beta}_{GLS}=\left(x'\Lambda^{-1}x\right)^{-1}x'\Lambda^{-1}y$$
$$=\left(\left(x_1^{'}\cdots x_N^{'}\right)\begin{pmatrix} \Omega^{-1} & & 0 \\ & \ddots & \\ 0 & & \Omega^{-1}\end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_N\end{pmatrix}\right)^{-1}\left(x_1^{'}\cdots x_N^{'}\right)\begin{pmatrix} \Omega^{-1} & & 0 \\ & \ddots & \\ 0 & & \Omega^{-1}\end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_N\end{pmatrix}$$
$$=\left(\begin{pmatrix} x_1^{'}\Omega^{-1} & \cdots & x_1^{'}\Omega^{-1}\end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_N\end{pmatrix}\right)^{-1}\begin{pmatrix} x_1^{'}\Omega^{-1} & \cdots & x_1^{'}\Omega^{-1}\end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_N\end{pmatrix}$$
$$=\left(\sum_{i=1}^{N}x_i^{'}\Omega^{-1}x_i\right)^{-1}\sum_{i=1}^{N}x_i^{'}\Omega^{-1}y_i$$
$$\equiv\hat{\beta}_{RE}$$

In this model, Ω is known. However, when Ω is unknown, we could use $\hat{\Omega}\rightarrow\Omega$, which requires us to obtain estimates for both σ_c^2 and σ_u^2. Note we have:

$$\sigma_v^{\ 2} = \sigma_c^{\ 2} + \sigma_u^{\ 2} = \frac{1}{T}\sum_{t=1}^{T} E(v_{it}^{\ 2}) \quad \text{for all } i.$$

Therefore, averaging v_{it}^2 across all i and t would give a consistent estimate of σ_v^2:

$$\hat{\sigma}_v^{\ 2} = \frac{1}{NT-K}\sum_{i=1}^{N}\sum_{i=1}^{T}\hat{\hat{v}}_{it}^{\ 2},$$

where $\hat{\hat{v}}_{it}$ is the residual from the OLS regression:

$$\hat{\hat{v}}_{it} = y_{it} - x_{it}\hat{\beta}_{OLS}.$$

Next we need to find a consistent estimator for σ_c^2. Recall that σ_c^2 =E(v_{it} v_{is}) all $t \neq s$. Therefore, for all i, there are $T/(T\text{-}1)/2$ non-redundant error products that can be used to estimate σ_c^2:

$$\hat{\sigma}_c^{\ 2} \ = \frac{1}{\left[NT(T-1)/2-K\right]}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{\hat{v}}_{it}\hat{\hat{v}}_{is}$$

is a consistent estimator of σ_c^2. Given $\hat{\sigma}_c^2$ and $\hat{\sigma}_v^2$, then we can obtain $\hat{\sigma}_u^2$:

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2 \qquad (9)$$

Note in practice (9) may not be positive. A negative value for $\hat{\sigma}_u^2$ is indicative of a substantial amount of negative serial correlation of u_{it}. If this occurs, we must use heteroscedastic-robust covariance matrix to construct $\hat{\beta}_{RE}$.

Robust variance matrix estimator:

If *Assumption 3* (homoscedastic) doesn't hold, then we must calculate the heteroscedastic-robust covariance matrix. Define the *Tx1* residual vector from the pooled OLS regressions: $\hat{\hat{v}}_i = y_i - x_i\hat{\beta}_{OLS}$, where i = 1, …, N. Define:

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{n}\hat{\hat{v}}_i\hat{\hat{v}}_i^{'}.$$

The reason that $\hat{\Omega} \xrightarrow{p} \Omega$ because it is the average over N. The random effect estimator is given by:

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} x_i^{'}\hat{\Omega}^{-1}x_i\right)^{-1}\left(\sum_{i=1}^{N} x_i^{'}\hat{\Omega}^{-1}y_i\right)$$

The necessary condition for the estimator to work is iid across individuals.

Estimation: Fixed-Effect model

Consider the panel model again:

$y_{it} = x_{it}\beta + c_i + u_{it}$, $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots T$.

Assumption FE1: $E(u_{it}|c_i, x_i) = 0$, but $E(c_i|x_{it}) \neq 0$, where $t = 1, \ldots, T$.

One way to estimate the model is by running a regression with a dummy variable for each individual:

$$y_{it} = x_{it}\beta + \sum_{i=1}^{N} c_i D_{it} + v_{it},$$

where $D_{it} = 1$ if ith person; and $D_{it} = 0$ otherwise.

In this case, it is important to note that $\hat{c}_i \xrightarrow{p} c_i$ iff $T \to \infty$. Note this is not the typical assumption of $N \to \infty$. In fact, we often assume T is fixed but $N \to \infty$ for the purpose of robust covariance estimation.

Therefore, typically $\hat{c}_i$ is not a consistent estimator of c_i. However, $\hat{c}_i$ is an unbiased estimator of c_i. In many applications, c_i is a nuisance parameter, so we don't really need to know them.

There are many ways to get rid of c_i before we estimate our model.

Method 1: Time-Demeaning:

Take away c_i by taking way the average across i:

$$\bar{y}_i = \bar{x}_i\beta + c_i + \bar{u}_i$$

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i), \text{ or } \ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it} \qquad (10)$$

Given the strict exogeneity condition, u_{it} is uncorrelated with x_{it} for all t=1,…..,T:

$$E(\ddot{x}_{it}'\ddot{u}_{it}) = E((x_{it} - \bar{x}_i)'(u_{it} - \bar{u}_i)) = 0$$

The OLS estimator of (5) is consistent:

$$\hat{\beta}_{FE} = \left(\sum_{i}^{N} \ddot{x}_i{}'\ddot{x}_i\right)^{-1}\left(\sum_{i}^{N} \ddot{x}_i{}'\ddot{y}_i\right)^{-1}$$

$$= \left(\sum_{i}^{N}\sum_{t}^{T} \ddot{x}_{it}\ddot{x}_{it}\right)^{-1}\left(\sum_{i}^{N}\sum_{t}^{T} \ddot{x}_{it}\ddot{y}_{it}\right)^{-1}$$

Is $\hat{\beta}_{FE}$ efficient asymptotically? Or the question: is $\ddot{u}_{it}$ iid?

For the same individual,

$$E\left(\ddot{u}_{it}{}^2\right) = E\left(u_{it} - \overline{u}_i\right)^2$$
$$= \sigma_u^2\left(1 - \frac{1}{T}\right)$$

$\Rightarrow$ homoscedasticity across individual t. For $t \neq s$, the covariance is:

$$E\left(\ddot{u}_{it}\ddot{u}_{is}\right) = -\frac{\sigma_u^2}{T}$$

Which shows the time demeaned error $\ddot{u}_{it}$ serially correlated. As T gets large, this correlation becomes smaller.

As it can be seen later, however, this serial correlation does not cause any problems. Because asymptotically, it is as if the covariance structure were iid with some minor correlation.

This set of de-meaned equation can be obtained by premultiplying a time demeaning matrix, Q_T, defined as:

$$Q_T \equiv I_T - j_T\left(j_T{}'j_T\right)^{-1}j_T{}'$$
$$= I_T - \frac{1}{T}j_Tj_T{}'$$
$$= \begin{pmatrix} 1-\frac{1}{T} & & -\frac{1}{T} \\ & \ddots & \\ -\frac{1}{T} & & 1-\frac{1}{T} \end{pmatrix}$$

Q_T is symmetric, idempotent with rank *T-1*. We have:

$$Q_Tj_T \equiv I_Tj_T - j_T\left(j_T{}'j_T\right)^{-1}j_T{}'j_T = 0$$

$$
\begin{aligned}
Q_T y_i &\equiv I_T y_i - j_T\left(j_T{}' j_T\right)^{-1} j_T{}' y_i \\
&= y_i - \frac{1}{T}\begin{pmatrix} 1 & & 1 \\ & \ddots & \\ 1 & & 1 \end{pmatrix} y_i \\
&= y_i - \bar{y}_i
\end{aligned}
$$

Therefore, we have:

$Q_T y_i = y_i - \bar{y}_i = \ddot{y}_i$. Similarly, $Q_T x_i = x_i - \bar{x}_i = \ddot{x}_i$.
$Q_T u_i = u_i - \bar{u}_i = \ddot{u}_i$. Therefore, (10) can be written as:

$$
Q_T y_i = Q_T x_i \beta + Q_T u_i \tag{11}
$$

The correlation between the transformed x_i and transformed u_i is given by:

$$
\begin{aligned}
\ddot{x}_{it}'\ddot{u}_{it} &= \left(Q_T x_i\right)' Q_T u_i \\
&= x_i' Q_T{}' Q_T u_i \\
&= x_i' Q_T{}' u_i = \ddot{x}_{it}' u_i
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sqrt{N}\left(\hat{\beta}_{FE} - \beta\right) &= \left(\frac{1}{N}\sum_i^N \ddot{x}_i'\ddot{x}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i^N \ddot{x}_i'\ddot{u}_i\right) \\
&= \left(\frac{1}{N}\sum_i^N \ddot{x}_i'\ddot{x}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_i^N \ddot{x}_i' u_i\right)
\end{aligned}
$$

Note:

$E\left(\ddot{x}_i' u_i\right) = 0$, and

$$
Var\left(\ddot{x}_i' u_i\right) = E\left(\ddot{x}_i' u_i u_i' \ddot{x}_i\right) = \sigma_u^2 E\left(\ddot{x}_i'\ddot{x}_i\right)
$$

By *CLT*,

$$
\frac{1}{\sqrt{N}}\sum_i^N \ddot{x}_i' u_i \xrightarrow{d} N\left(0, \sigma_u^2 E\left(\ddot{x}_i'\ddot{x}_i\right)\right).
$$

Then we must have:

$$
\sqrt{N}\left(\hat{\beta}_{FE} - \beta\right) \xrightarrow{d} N\left(0, \sigma_u^2\left(\frac{1}{N}\sum_i^N \ddot{x}_i'\ddot{x}_i\right)^{-1}\right)
$$

Therefore, the asymptotic covariance of $\hat{\beta}_{FE}$ is the same as if $\ddot{u}_{it}$ were iid, with one difference:

If *OLS*, $Var(\hat{\beta}_{FE}) = Var(\ddot{u}_{it})\left(\sum_{i}^{N} \ddot{x}_i' \ddot{x}_i\right)^{-1}$.

Note $Var(\ddot{u}_{it}) \neq \sigma_u^2 = Var(u_{it})$. We need to make an adjustment of the variance estimate directly from *OLS*.

Now how to estimate σ_u^2?

Note the error term in *FE* model is $\ddot{u}_{it}$, so we cannot directly use the *SSR*. $\sum_{t=1}^{T} E(\ddot{u}_{it}^2) = (T-1)\sigma_u^2$ which means $E(\ddot{u}_{it}^2) \leq \sigma_u^2$. Therefore, directly using OLS variance estimate would yield a smaller variance estimate. In another words, OLS still problematic in calculating variance, but since the difference is only a constant, we can correct it – in this sense that OLS of the demeaned equation does not cause problems.

$$\frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} E(u_{it}^2) = \sigma_u^2$$

Now define fixed effect residual: $\hat{u}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\hat{\beta}_{FE}$.

A consistent and unbiased estimator of $E(\ddot{u}_{it}^2)$ is:

$$E(\ddot{u}_{it}^2) = \frac{1}{NT-K} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2$$

Then a consistent estimator of σ_u^2 is:

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\hat{u}_{it}^2)$$

.

In the denominator is *N(T-1)-K* instead of *NT-K*. This difference could be substantial if *T* is small.

More complications:

1. Serial correlation: $u_{it} = \rho u_{it-1} + \varepsilon_{it}$

Typically, we simply use the residual to test if there is a serial correlation. However, it is not simple in this case because the error term after the transformation is $\ddot{u}_{it}$.

Since $E(\ddot{u}_{it}\ddot{u}_{is}) = -\frac{\sigma_u^2}{T}$, for the equation $\ddot{u}_{it} = \delta\ddot{u}_{it-1} + \varepsilon_{it}$, the H_0 of no-serial correlation is: H_0: $\delta = -1/(T-1)$.

To find if there a serial correlation, let $\hat{u}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\hat{\beta}_{FE}$, run a regression of $\hat{u}_{it}$ on $\hat{u}_{it-1}$, $t = 2,,\ldots\ldots T$, and i=1,….N. Test if the coefficient is -1/(T-1). Note the standard error for the coefficient used in the test should be robust standard error.

If we find serial correlation, then we have to use robust variance estimator in the original FE estimator.

$$A\,\mathrm{var}(\hat{\beta}_{FE}) = \left(\sum_{i=1}^{N}\ddot{x}_i'\ddot{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{x}_i'\hat{u}_i\hat{u}_i'\ddot{x}_i\right)\left(\sum_{i=1}^{N}\ddot{x}_i'\ddot{x}_i\right)^{-1}$$

This robust variance matrix estimator is valid in the presence of heteroskedasticity or serial correlation provided that T is small relation to N.

Method 2: First Differencing

Again, consider the linear panel data model,

$$y_{it} = x_{it}\beta + c_i + u_{it},\ \ i = 1, 2, \ldots, N;\ t = 1, 2, \ldots T. \qquad (12)$$

To get rid of the unobserved c_i, take the first differencing, we have:

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it},. \qquad (13)$$

A pooled regression of Δy_{it} on Δx_{it}, denoted as $\hat{\beta}_{FD}$, can yield consistent estimator of β, given the assumptions:

Assumption *FD1*: $E(\Delta x_{it}'\Delta u_{it}) = 0$, $t = 2, 3,\ldots,T$.
Assumption *FD2*: rank $\sum_{t=2}^{T} E(\Delta x_{it}'\Delta x_{it})$ is K. $t = 2, 3,\ldots,T$.

With the two assumptions, the OLS estimate of (13) is consistent.

If u_{it} in (12) follows random walk, then the OLS estimates of (13), $\hat{\beta}_{FD}$, is efficient. If u_{it} in (12) is iid, then Δu_{it} in (13) will be serially correlated:

$Var(\Delta u_{it}) = 2\sigma_u^2$, and $E(\Delta u_{it}\Delta u_{it-1}) = -\sigma_u^2$.

Therefore, we would adopt robust variance matrix:

$$A\,\mathrm{var}\left(\hat{\beta}_{FD}\right)=\left(\sum_{i=1}^{N}\Delta x_i^{'}\Delta x_i\right)^{-1}\left(\sum_{i=1}^{N}\Delta x_i^{'}\Delta\hat{u}_i\Delta\hat{u}_i{}'\Delta x_i\right)\left(\sum_{i=1}^{N}\Delta x_i^{'}\Delta x_i\right)^{-1}$$

where $\Delta\hat{u}_{it}=\Delta y_{it}-\Delta x_{it}\hat{\beta}_{FD}$.

Discussions: serial correlation of u_{it}.

Again, when u_{it} are iid, then a regression of the following model should yield:

$$\Delta u_{iT} = \rho\Delta u_{iT-1} + \varepsilon_{it}.$$

Testing if there is a serial correlation is equivalent to testing if H_0: $\rho=0.5$. Note here that we only use the observation at T and T-1 for Δu_{iT} and T-1 and T-2 for Δu_{iT-1}. We are not use information from other years to avoid the unnecessary panel data issue.

It is possible that Δu_{it} is actually iid. This occurs when u_{it} is random walk.

Comparing $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$:

Depends on error structure: If u_{it} is serially uncorrelated, the *FE* is more efficient. However, when u_{it} is a random walk, then *FD* is more efficient.

It is important to note that when there is a strong exogeneity, the difference between *FD* and *FE* is only the sampling error. Therefore, one can use Hausman-type test (to compare the difference between the *FE* estimates and *FD* estimates).

Prefer FD: 1. easier to estimate
Prefer FE: 1. more efficient (*FD* losses *N* observations)

Test of strict exogeneity: when T > 2, we can use the FE method.

$$y_{it} = x_{it}\beta + w_{it+1}\delta + c_i + u_{it},$$

where w_{it+1} is a subset of x_{it+1}. Testing $\delta=0$ is a test for strict exogeneity.

Test of strict exogeneity when T = 2. In this case, we may have to use FD method only.

$$\Delta y_{it} = \Delta x_{it}\beta + w_t\delta + \Delta u_{it},$$

where w_t is a subset of x_t. Again, testing $\delta=0$ is a test for strict exogeneity.

Comparing $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$

In some situations, it is often only possible to apply random effect model. The key variable we are interested in have no variation or very little variation over time.

Example: we are interested in how the size of the local labor market affect local the unemployment rate (Gan and Zhang, *Journal of Econometrics*, 2006, page 127-152.). We have panel data of 295 monthly city unemployment rates.

$$unemployment\ rate_{ct} = \alpha_c + X_{ct}\beta + \gamma \log(size_c) + u_{ct}$$

where X_{ct} includes: unemployment benefit, measurement of industry composition, percentage of people who are young (youth share), net migration rates, log of square miles of area (not changing over time), and finally the log of the size of the market (average employment), which does not change overtime.

There is a model developed in the paper that argues that a larger size of the market would yield a lower unemployment.

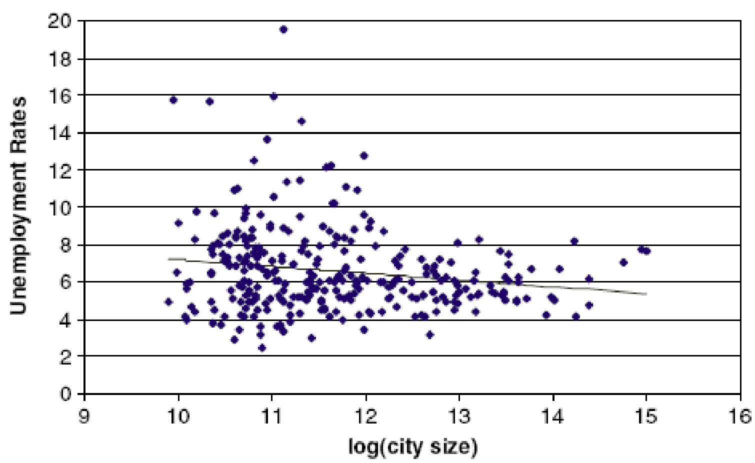

Fig. 2. Logarithm of city size and mean unemployment rates.

The paper finds that an increase in two standard deviation of the city size would result a decrease of 0.15 percentage points of unemployment rate.

Table 6
Unemployment rate mean regression results

Variables	(1)	(2)
time fixed effect	Yes	Yes
city random effect	Yes	Yes
constant	0.485	1.423
	(0.520)	(0.539)
Lagged unemployment rate	0.874	0.873
	(0.0020)	(0.0021)
INDCOM	−18.64	−18.66
	(1.033)	(1.033)
RISK	10.61	1.467
	(5.69)	(5.88)
INDCOM × RISK	360.8	424.9
	(163.8)	(163.4)
unemployment benefit	1.058	1.019
	(0.128)	(0.127)
youth share	−1.765	−2.147
	(0.528)	(0.531)
mean net migration rate	−0.0051	0.0080
	(0.011)	(0.011)
log(miles^2)	−0.141	−0.203
	(0.140)	(0.139)
$[\log(miles^2)]^2$	0.0105	0.0179
	(0.0095)	(0.0095)
log(size)		−0.0752
		(0.0129)
R^2	0.915	0.916
No. of obs.	50439	50439

Standard errors are in parentheses.

In this example, T is relatively large. Given that, we can show next that $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$ are close to each other.

In the case of varying x_{it}, we can compare $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$. These two estimators do exhibit some interesting relationships.

Again, define:

$$
\begin{aligned}
Q_T &\equiv I_T - j_T\left(j_T' j_T\right)^{-1} j_T' = I_T - P_T \\
&= I_T - \frac{1}{T} j_T j_T' \\
&= \begin{pmatrix} 1-\frac{1}{T} & & -1 \\ & \ddots & \\ -1 & & 1-\frac{1}{T} \end{pmatrix}
\end{aligned}
$$

Then:

$$Q_T j_T \equiv I_T j_T - j_T (j_T' j_T)^{-1} j_T' j_T = 0$$

For *FE* model, the transformed equation can be written as:

$$Q_T y_i = Q_T x_i \beta + Q_T u_i \qquad (14)$$

For *RE* model:

$$\begin{aligned}
\Omega &= \sigma_u^2 I_T + \sigma_c^2 j_T j_T' \\
&= \sigma_u^2 I_T + T\sigma_c^2 j_T (j_T j_T')^{-1} j_T' \\
&= \sigma_u^2 (P_T + Q_T) + T\sigma_c^2 P_T \\
&= (\sigma_u^2 + T\sigma_c^2) P_T + \sigma_u^2 Q_T \\
&= (\sigma_u^2 + T\sigma_c^2)\left(P_T + \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2} Q_T \right) \\
&= (\sigma_u^2 + T\sigma_c^2)(P_T + \eta Q_T)
\end{aligned}$$

where $\eta = \dfrac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2}$.

Define $S_T(\eta) \equiv P_T + \eta Q_T$. Note we have: $P_T P_T = I$, and $P_T Q_T = 0$, $P_T = I_T - Q_T$.
It is easy to show that $S_T^{-1} = P_T + \dfrac{1}{\eta} Q_T$ by showing $S_T S_T^{-1} = I$.
Again, one can also show that $S_T^{-1/2} = P_T + \dfrac{1}{\sqrt{\eta}} Q_T$, since $S_T^{-1/2} S_T^{-1/2} = S_T^{-1}$.

Given that, $\Omega = (\sigma_u^2 + T\sigma_c^2)(P_T + \eta Q_T)$ and the previous equation, we have:

$$\Omega^{-1/2} = (\sigma_u^2 + T\sigma_c^2)^{-1/2}\left(P_T + \frac{1}{\sqrt{\eta}} Q_T \right).$$

Again, define $\lambda = 1 - \sqrt{\eta} = 1 - \dfrac{\sigma_u}{(\sigma_u^2 + T\sigma_c^2)^{1/2}}$. Then,

$$
\begin{aligned}
\Omega^{-1/2} &= \frac{1-\lambda}{\sigma_u}\left(P_T + \frac{1}{1-\lambda}Q_T\right) \\
&= \frac{1}{\sigma_u}\left(P_T - \lambda P_T + Q_T\right) \\
&= \frac{1}{\sigma_u}\left(I_T - \lambda P_T\right)
\end{aligned}
$$

Further, define $G_T=I_T - \lambda P_T$. Consider a transformation of

$$G_T y_i = G_T x_i \beta + G_T v_i \tag{15}$$

It is easy to verify that the error term in (15) is iid:

$$E\left(G_T v_i v_i^{'} G_T\right) = \frac{1}{\sigma_u^2}\Omega^{-1}E\left(v_i v_i^{'}\right) = \frac{1}{\sigma_u^2}$$

Therefore, OLS of (14) yields an estimate:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \left(\sum_{i=1}^{N} x_i^{'} G_T G_T x_i\right)^{-1} \sum_{i=1}^{N} x_i^{'} G_T G_T y_i \\
&= \left(\frac{1}{\sigma_u^2}\sum_{i=1}^{N} x_i \Omega^{-1} x_i^{'}\right)^{-1} \frac{1}{\sigma_u^2}\sum_{i=1}^{N} x_i^{'} \Omega^{-1} y_i \\
&= \left(\sum_{i=1}^{N} x_i \Omega^{-1} x_i^{'}\right)^{-1} \sum_{i=1}^{N} x_i^{'} \Omega^{-1} y_i \\
&= \hat{\beta}_{RE}
\end{aligned}
$$

Rewrite equations (14) and (15) here:

$$Q_T y_i = Q_T x_i \beta + Q_T v_i \tag{14}$$
$$G_T y_i = G_T x_i \beta + G_T v_i \tag{15}$$

where $G_T=I_T - \lambda P_T$ and $Q_T=I_T - P_T$. The difference between the two transformation is:

$$\lambda = 1 - \frac{\sigma_u}{\left(\sigma_u^2 + T\sigma_c^2\right)^{1/2}}.$$

As T is large or σ_c/σ_u is large, $\lambda \to 1$, random effect model and the fixed effect is close to each other. Note whether T is large is not related to N. It has nothing to do with the N.

Hausman and Taylor-Type Models

Consider a model:

$$y_{it} = x_{it}\beta + z_i\gamma + c_i + u_{it}.$$

We allow $\text{Cov}(x_{it},c_i) \neq 0$, but assume $\text{Cov}(z_i,c_i) = 0$, and the strict exogeneity on u_{it}. The key coefficient here is γ. Since z_i do not vary with time, so the fixed-effect model cannot be used. A typical way to estimate this is to apply the random-effects model, where it is assumed that $\text{Cov}(x_{it},c_i) = 0$. The Hausman-Taylor type model has less strict assumption than the random-effects model by allowing x_{it} and c_i being correlated.

To estimate this model, first we estimate the fixed-effect model – note the both the $z_i\gamma$ term and c_i term are eliminated at this stage. From the estimates of the fixed-effects model, we can get residual. Note the residual includes three terms: $z_i\gamma$ term, c_i term, and the u_{it} term.

Second, we take the average (over time, for each i) to minimize the effect of the u_{it} term. The averaged residual now mostly consists of the z_i term and the c_i term. Because these two terms (z_i term and the c_i term) are uncorrelated, and z_i is observed, we can run regression of the averaged residual on z_i to obtain the consistent estimate of γ.

Example: we are interested in how the size of the local labor market affect local the unemployment rate (Gan and Zhang, *Journal of Econometrics*, 2006, page 127-152.). We have panel data of 295 monthly city unemployment rates.

$$\textit{unemployment rate}_{ct} = \alpha_c + X_{ct}\beta + \gamma \log(\textit{size}_c) + u_{ct}$$

It is possible to allow α_c and X_{ct} to be arbitrarily correlated, but assume that α_c and $\log(\textit{size}_c)$ to be uncorrelated.

How to estimate such a model:

Since $E(z_i'c_i) = 0$, we have an extra moment condition that can be used to identify γ.

Intuition: we first estimate β by fixed effect model. Then we use the residual to estimate γ, given

$$E(z_i'z_i)\gamma = E\left(z_i'(\bar{y}_i - \bar{x}_i\beta)\right)$$

The estimate is given by:

$$\hat{\gamma} = \left(\frac{1}{N}\sum_{i=1}^{N} z_i' z_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} z_i' \left(\bar{y}_i - \bar{x}_i \hat{\beta}_{FE}\right)\right)$$

STATA Deviation

Simple command for panel data model:

(1) xtreg

Example:

xtreg ln_w grade age* ttl_exp ttl_exp2 tenure tenure2 black not_smsa south, fe i(idcode)

fe – fixed effect model
"re" -- random effect model GLS
"be" – between-effects model
"mle" – random effect MLE

i(idcode), specifies the variable name that contains the unit to which the observation belongs

(2) Given the model:

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

We can allow u_{it} to be serially correlated: $u_{it} = \rho u_{it-1} + \varepsilon_{it}$

Since the model has a time series aspect, we need to tsset the data.

tsset
xtregar ln_w grade age* ttl_exp ttl_exp2 tenure tenure2 black not_smsa south, fe

(3) Hausman-Taylor

Consider the model: $y_{it} = x_{it}\beta + z_i\gamma + c_i + u_{it}$.

The STATA command on this type of models is `xthtaylor`. This command defines endogenous and exogenous variables. The endogenous variables are variables that are allowed to be correlated with c_i, while exogenous variables are variables that are not allowed to be correlated with c_i. In order to be able to identify γ, it is necessary to have at least one exogenous variable z_i to ensure the moment condition $E(z_i' c_i) = 0$. If all z_i are endogenous, then the model becomes the typical panel data fixed-effects model where γ cannot be identified.

The advantage of Hausman-Taylor model over the typical random-effects model is that it allows endogenous variables in x_{it} (allowing c_i and x_{it}) to be correlated. It even allows some of the z_i to be correlated.

Example 1: In Gan and Zhang, *Journal of Econometrics* 2006, page 127-152.)

In this example, the endogenous variables (that are allowed to be correlated with c_i) are X_{ct}, and the exogenous variables (that are uncorrelated with c_i) are log($size_c$).

`xthtaylor` *unemployment_rate* X_{ct} log($size_c$), `endog`(X_{ct})

Example 2: wage is a function of

How long this person has worked for the firm, wks
Binary variables if lives in a metropolitan area or in the south, smsa, and south
Marital status, *ms*;
Years of education, *ed*
A quadratic of work experience
In manufacturing or not,
Black or not, *blk*
Female or not, *fem*

It is expected that time-varying variables exp, exp2, wks, ms, and union are all correlated with the unobserved individual effect.

Assume exogenous variables *occ, south, smsa, ind, fem*, and *blk* are instruments for the endogenous, time-invariant variable *ed*.

xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp exp2 wks ms union ed)