Part II

Multivariate Linear Models

Yoosoon Chang
Department of Economics
Indiana University

and

Joon Y. Park
Department of Economics
Indiana University
and
Department of Economics
Sungkyunkwan University

January 2012

1. Multivariate Regression Models

1.1 The Model

The model we consider is given by

$$y_i' = x_i'B + u_i'$$

for $i = 1, \ldots, n$, where n is the sample size. We let $\{y_i\}$ and $\{u_i\}$ be ℓ-dimensional, and $\{x_i\}$ be m-dimensional vectors, that are interpreted similarly as in the univariate linear regression model studied previously. The $m \times \ell$ matrix B is the matrix of regression coefficients. The model thus specifies ℓ-linear relationships between y_i and x_i, each of which is given by the corresponding column of the coefficient matrix B. It is just a multiple of univariate regressions with common regressors, pulled together. In matrix form, the model is written as

$$Y = XB + U$$

where the matrices are defined as usual, i.e., observations along the rows and variables along the columns.

The errors $\{u_i\}$ are assumed to be $(0, \Sigma)$ and mutually uncorrelated. Therefore, $\mathbf{E}u_i u_j' = \Sigma$ for $i = j$ and 0 otherwise. Moreover, if we use the convention $\mathrm{var}(Z) = \mathrm{var}(\mathrm{vec}Z)$, then

$$\mathrm{var}\,(U) = I_n \otimes \Sigma$$

under our assumption.

1.2 Multivariate Least Squares

Since $\sum_{p,q} z_{pq}^2 = \mathrm{tr}Z'Z$ for a matrix $Z = (z_{pq})$, the least squares (LS) estimator $\widehat{B}$ of B is defined by

$$\widehat{B} = \underset{B}{\mathrm{argmin}}\ \mathrm{tr}\,(Y - XB)'(Y - XB) \tag{1}$$

If we denote by $q(B) = \mathrm{tr}\,(Y - XB)'(Y - XB)$, then we have

$$dq(B) = -2\,\mathrm{tr}\,(X'Y - X'XB)'\,dB$$

and the FOC $dq(B) = 0$ yields

$$\widehat{B} = (X'X)^{-1}X'Y$$

Similarly as in the univariate regression theory, let $\widehat{Y} = X\widehat{B}$ and $\widehat{U} = Y - X\widehat{B}$. Then

$$\widehat{Y} = P_X Y \quad \text{and} \quad \widehat{U} = (I - P_X)Y$$

where $P_X = X(X'X)^{-1}X'$, i.e., the projection on the range $\mathcal{R}(X)$ of X, as in the univariate regression.

It is easy to see that $\widehat{B}$ is indeed the solution for the minimization problem (1), since

$$q(B) = \mathrm{tr}\,(Y - X\widehat{B})'(Y - X\widehat{B}) + \mathrm{tr}\,(\widehat{B} - B)'X'X(\widehat{B} - B) \tag{2}$$

Notice that the LS estimate $\widehat{B}$ is nothing but the LS estimates of the ℓ-univariate regressions stacked together. That is, the k-th column of $\widehat{B}$ is precisely the LS estimate from the regression of y_k on X, where y_k is the k-th column of Y.

Given the LS estimate $\widehat{B}$, we may estimate Σ by

$$\begin{aligned}\widehat{\Sigma} &= \frac{1}{n}\sum_{t=1}^{n}\hat{u}_i\hat{u}_i' \quad \left(=\frac{1}{n}\widehat{U}'\widehat{U}\right)\\ &= \frac{1}{n}Y'(I-P_X)Y \quad \left(=\frac{1}{n}U'(I-P_X)U\right)\end{aligned}$$

1.3 ML Estimation

Under normality, $U \sim \mathbf{N}(0, I_n \otimes \Sigma)$ and its density is given by

$$p(U) = (2\pi)^{-\frac{n\ell}{2}}(\det\Sigma)^{-\frac{n}{2}}\operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}U'U\right)$$

where $\operatorname{etr}(\cdot) := \exp(\operatorname{tr}(\cdot))$. To obtain the density in the above form, we used

$$\begin{aligned}\det(I_n\otimes\Sigma) &= (\det\,\Sigma)^n\\ (\operatorname{vec}U)'(I_n\otimes\Sigma^{-1})(\operatorname{vec}U) &= \operatorname{tr}\Sigma^{-1}U'U\end{aligned}$$

The density of Y is obtained simply by replacing U with $Y - XB$, since the Jacobian of transformation is 1.

Ignoring the constant term, the loglikelihood function is therefore given by

$$\ell(B,\Sigma) = -\frac{n}{2}\log(\det\Sigma) - \frac{1}{2}\operatorname{tr}\Sigma^{-1}(Y-XB)'(Y-XB)$$

Totally differentiating $\ell(B,\Sigma)$ with respect to B and Σ yields the first order conditions

$$\begin{aligned}\operatorname{tr}\Sigma^{-1}(X'Y - X'XB)'\,dB &= 0\\ \operatorname{tr}\Sigma^{-1}\left(I - \frac{1}{n}(Y-XB)'(Y-XB)\Sigma^{-1}\right)d\Sigma &= 0\end{aligned}$$

To obtain the second FOC above, we use

(a) $d\log(\det\,\Sigma) = \operatorname{tr}\Sigma^{-1}\,d\Sigma$

(b) $d\Sigma^{-1} = -\Sigma^{-1}\,d\Sigma\,\Sigma^{-1}$

For (a), note that $(\partial/\partial\sigma_{pq})\det\,\Sigma = c_{pq}$, where $\Sigma = (\sigma_{pq})$ and c_{pq} is the cofactor of σ_{pq} so that $\operatorname{adj}\Sigma = (c_{qp})$. The result is then immediate from $\operatorname{adj}\Sigma/\det\,\Sigma = \Sigma^{-1}$. Part (b) can easily be obtained by totally differentiating the identity $\Sigma\Sigma^{-1} = I$.

It is now easy to see that the FOC's yield the LS estimators $\widehat{B}$ and $\widehat{\Sigma}$ which we obtained earlier. It is possible to show that these estimators indeed maximize the likelihood function of Y. That $\widehat{B}$ maximizes the likelihood function is obvious from (2). Moreover, we may apply the result in Problem 1 (with $A = \widehat{\Sigma}^{\frac{1}{2}}\Sigma^{-1}\widehat{\Sigma}^{\frac{1}{2}}$) to deduce that

$$\ell(\widehat{B},\Sigma) = -\frac{n}{2}\log(\det\,\Sigma) - \frac{n}{2}\operatorname{tr}\Sigma^{-1}\widehat{\Sigma}$$

is maximized when $\Sigma = \widehat{\Sigma}$. Consequently, the LS estimation is identical to the ML estimation under normality in the multivariate regression models as is in univariate regression models.

1.4 Statistical Properties of the Estimators

We derive various statistical properties of the estimators $\widehat{B}$ and $\widehat{\Sigma}$.

Theorem 1 *We have*

(a) $\quad \mathbf{E}\,(\widehat{B}) = B \quad$ and $\quad \mathrm{var}\,(\widehat{B}) = (X'X)^{-1} \otimes \Sigma$

(b) $\quad \mathbf{E}\,(\widehat{\Sigma}) = \dfrac{n-m}{n}\,\Sigma$

Proof Part (a) is straightforward from

$$\widehat{B} = B + (X'X)^{-1}X'U \quad \text{and} \quad \mathrm{vec}\,\widehat{B} = \mathrm{vec}\,B + ((X'X)^{-1}X' \otimes I_\ell)\,\mathrm{vec}\,U$$

For part (b), write

$$\widehat{\Sigma} = \frac{1}{n}U'(I - P_X)U = \frac{1}{n}V'V$$

with $V := H'U$, where H is an $n \times (n–m)$ matrix such that $I - P_X = HH'$ and $H'H = I_{n-m}$. Now, $\mathrm{vec}\,V = (H' \otimes I)\mathrm{vec}\,U$ and therefore $\mathrm{var}\,V = I_{n-m} \otimes \Sigma$, i.e., the $(n-m)$ rows v_i's of V are uncorrelated and each has covariance matrix Σ. Then it follows that

$$\mathbf{E}\,V'V = \mathbf{E}\left(\sum_{i=1}^{n-m} v_i v_i'\right) = (n-m)\,\Sigma$$

and the proof is now complete. ■

The $\widehat{B}$ is unbiased, but $\widehat{\Sigma}$ is not. However, we can easily construct an unbiased estimator for Σ as follows

$$\tilde{\Sigma} = \frac{1}{n-m}\,Y'(I - P_X)Y$$

For the models with normal errors, the distribution of $\widehat{B}$ is obviously normal with mean and variance given in part (a) of Theorem 1, since $\widehat{B}$ is a linear transformation of Y. To characterize the distribution of $\widehat{\Sigma}$, we introduce

Definition 1 *Let $z_i \sim$ i.i.d. $\mathbf{N}(0, \Sigma_p)$. Then*

$$\sum_{i=1}^{n} z_i z_i' \sim \mathcal{W}_p\,(n, \Sigma)$$

i.e., Wishart distribution with n degrees of freedom and covariance matrix Σ. The p is the dimensionality parameter.

Clearly, Wishart distribution is the multivariate generalization of chi-square distribution. Following the proof of part (b) in Theorem 1, we may readily show that the distribution of $\widehat{\Sigma}$ is Wishart, as we summarize below:

Theorem 2 *Under normality, we have*
(a) $\widehat{B} \sim \mathbf{N}\left(B, (X'X)^{-1} \otimes \Sigma\right)$
(b) $n\,\widehat{\Sigma} \sim \mathcal{W}_\ell\,(n-m, \Sigma)$

The asymptotic theory for the multivariate regression model is identical to that for the univariate regression model which we developed earlier. Under appropriate regularity conditions to ensure

$$\text{(a)}\quad \frac{X'X}{n} = \sum \frac{x_i x_i'}{n} \to_p M > 0 \quad \text{and} \quad \text{(b)}\quad \frac{X'U}{n} = \sum \frac{x_i u_i'}{n} \to_p 0$$

$\widehat{B}$ is consistent. Moreover, if the condition

$$\text{(c)}\quad \frac{X'U}{\sqrt{n}} = \sum \frac{x_i u_i'}{\sqrt{n}} \ \to_d\ \mathbf{N}(0, M \otimes \Sigma)$$

is satisfied, then we have

$$\sqrt{n}\,(\widehat{B} - B) \quad \to_d \quad \mathbf{N}\,\left(0, M^{-1} \otimes \Sigma\right) \tag{3}$$

since $\operatorname{vec}(\widehat{B}-B) = \left((X'X)^{-1} \otimes I_\ell\right) \operatorname{vec} X'U$. The consistency of $\widehat{\Sigma}$ can also be easily shown, exactly as in the proof of the consistency of $\hat{\sigma}^2$.

1.5 Hypothesis Testing

The theory of hypothesis testing in the multivariate regression is largely identical to that in the univariate regression model. Therefore, we only briefly summarize the results here. The general linear hypothesis on the coefficient matrix B can be formulated as

$$R \operatorname{vec} B = r$$

where R and r are known with dimensions $q \times m\ell$ and $q \times 1$, respectively, just as defined in the univariate case. Usually, the test is based upon the following Wald statistic W

$$W = (R\mathrm{vec}\widehat{B} - r)' \left(R((X'X)^{-1} \otimes \widehat{\Sigma})R'\right)^{-1} (R\mathrm{vec}\widehat{B} - r)$$

The limit distribution of W follows immediately from the asymptotic results established in (3) as

$$W \ \to_d\ \chi^2_q$$

The $\widehat{\Sigma}$ can of course be replaced by any other consistent estimate for Σ for the asymptotic chi-square tests.

The finite sample distribution theory of W under normality is, except for some very simple cases, overly complicated and certainly beyond the scope of this course. Therefore, the subject will not be developed any further.

1.6 Exercises

1. Show the following:
(a) $\overline{\text{vec}}(ABC) = (C' \otimes A)(\overline{\text{vec}}B)$.
(b) Let $x = \text{vec}\,X$ and $y = \text{vec}\,Y$. Then

$$\frac{\partial^2}{\partial x\,\partial y'}\text{tr}\,X'AYB' = A \otimes B$$

2. Show that the concentrated loglikelihood

$$\ell(\widehat{B}, \Sigma) = -\frac{n}{2}\log(\det\Sigma) - \frac{n}{2}\text{tr}\Sigma^{-1}\widehat{\Sigma},$$

derived in Section 1.3 of the lecture note is maximized when $\Sigma = \widehat{\Sigma}$.

3. Let $\hat{B}$ be the GLS estimator for B in the multivariate model

$$Y = XB + U$$

with $\text{var}(U) = \Sigma_1 \otimes \Sigma_2$ for some positive definite matrices Σ_1 and Σ_2. Show that $\hat{B}$ minimizes

$$\text{tr}\,\Sigma_2^{-1}(Y - XB)'\Sigma_1^{-1}(Y - XB)$$

and

$$\hat{B} = (X'\Sigma_1^{-1}X)^{-1}X'\Sigma_1^{-1}Y$$

4. Consider a multivariate regression model

$$y_i' = x_i'B + u_i'$$

where $y_i' = (y_{1i}, y_{2i}), x_i' = (x_{1i}, x_{2i}), u_i' = (u_{1i}, u_{2i})$,

$$B = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$$

and (u_i) are iid with mean 0 and variance

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Answer the following questions:
(a) Let $\beta_{21} = 0$ be known. Compare the single equation and system estimators of the parameters β_{11}, β_{12} and β_{22}.
(b) Let $\beta_{21} = 0$ *and* $\sigma_{12} = \sigma_{21} = 0$ be known. How would your answer to (a) change?
(c) Assume $\Sigma = \sigma^2 I$. Consider the hypothesis $\beta_{11} + \beta_{12} = 1$. Explain how to compute the

Wald statistic in two different ways: one using the OLS estimates for β_{11} and β_{12}, and the other based on the restricted and unrestricted sums of residuals.

5. Consider the regression models
   (A) $y_{ij} = x_{ij}'\beta + \varepsilon_{ij}$
   (B) $y_{ij} = x_{ij}'\beta_j + \varepsilon_{ij}$

where $i = 1, \ldots, n$ and $j = 1, 2$. Let $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})'$ be iid with covariance matrix given by one of

$$\Sigma_1 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}.$$

Answer the following questions:

(a) Explain how to find the feasible GLS estimate for β in regression (A) when the error covariance matrix is given by Σ_1.

(b) Find the GLS estimates of $\beta_j, j = 1, 2$, in regression (B) when the error covariance matrix is given by Σ_1 or Σ_2. Compare the GLS and OLS estimates in each case. What if $x_{i1} = x_{i2}$ in regression (B)?

2. Seemingly Unrelated Regressions

2.1 The Model

The seemingly unrelated regressions (SUR) is a system of regressions given by

$$y_k = X_k\beta_k + u_k$$

for $k = 1, \ldots, \ell$, where y_k, X_k, β_k and u_k are defined exactly as in the matrix representation for the univariate regression model and the subscript k denotes the k-th equation, for each of which we have n observations. To analyze such a system of regressions, we stack the ℓ-regressions and write it as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix} = \begin{pmatrix} X_1 & & \\ & \ddots & \\ & & X_\ell \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_\ell \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_\ell \end{pmatrix}$$

or as

$$y = X\beta + u$$

in matrix form. We let $\mathbf{E}\, u_p u_q' = \sigma_{pq} I$ and $\Sigma = (\sigma_{pq})$. Then it follows that $\operatorname{var} u = \Sigma \otimes I_n$. Note that for a random matrix U such that $u = \overline{\operatorname{vec}}\, U$ we have $\operatorname{var} U = \operatorname{var}(\operatorname{vec} U) = I_n \otimes \Sigma$.

The multivariate regression model considered in the previous section is just a special SUR for which $X_1 = \cdots = X_\ell$, i.e., the SUR with the same set of regressors. If we denote this common regressor by X_0, then we may write the multivariate regression model in SUR form, using

$$\overline{\text{vec}}\, Y = (I \otimes X_0)\, \overline{\text{vec}}\, B + \overline{\text{vec}}\, U$$

and defining $y = \overline{\text{vec}} Y, X = I \otimes X_0, \beta = \overline{\text{vec}} B$ and $u = \overline{\text{vec}} U$.

2.2 Estimation

The SUR system can of course be most efficiently estimated by GLS, which yields

$$\hat{\beta} = \Big(X'(\Sigma^{-1} \otimes I)X\Big)^{-1} X'(\Sigma^{-1} \otimes I)y \tag{4}$$

This is called the SUR estimator, which is nothing but the system GLS estimator. When Σ is unknown, as is usually the case, a feasible GLS estimator is used. To estimate Σ, we use any consistent estimate $\hat{\beta}_k$ of β_k (the OLS estimate, for instance) to get $\hat{u}_k = y_k - X_k\hat{\beta}_k$ for $k = 1, \ldots, \ell$. Using $\{\hat{u}_k\}$, we may obtain $\hat{\sigma}_{pq} = \hat{u}_p'\hat{u}_q/n$ for $p, q = 1, \ldots, \ell$ and construct a consistent estimate $\widehat{\Sigma} = (\hat{\sigma}_{pq})$.

Under normality, the density of u is precisely as in the multivariate regression model with U defined from u by $\overline{\text{vec}} U = u$. Moreover, the density of y can also be subsequently obtained from that of U by replacing the k-th column of U by $y_k - X_k\beta_k$. The SUR estimator defined in (4) is, of course, the ML estimator of β under normality. Given the ML estimate of β, the ML estimate of Σ is the same as that given for the multivariate regression. That is, if we define $\widehat{U}$ from $\hat{u}$, which is the ML (or SUR, equivalently) residuals, as we define U from u, then

$$\widehat{\Sigma} = \frac{1}{n}\widehat{U}'\widehat{U}$$

The ML estimates of β and Σ can therefore be obtained by the usual iterative scheme for the feasible GLS.

It should be noted that the system OLS for the SUR model is reduced to the single equation least squares (SELS). If we denote by $\tilde{\beta}$ the OLS estimator for β in the SUR system, then

$$\begin{aligned} \tilde{\beta} &= (X'X)^{-1}X'y \\ &= \begin{pmatrix} (X_1'X_1)^{-1}X_1'y_1 \\ \vdots \\ (X_\ell'X_\ell)^{-1}X_\ell'y_\ell \end{pmatrix} \end{aligned}$$

which is just the vector of the OLS estimates obtained by running each regression separately. For the SUR system, GLS is more efficient than OLS since the errors are nonspherical. This in turn implies that the SUR procedure is generally more efficient than the SELS, given the above result. Here comes the explanation for the name SUR: We may improve efficiency by pooling together regressions with nonoverlapping sets of regressors (and therefore seemingly unrelated).

2.3 Equivalence of SUR and SELS

There are two obvious cases that the SUR procedure becomes identical to the SELS (or the system OLS). First, this happens when $\sigma_{pq} = 0$ for all $p \neq q$, i.e., there are no correlations between regression errors. Secondly, such case arises when all the regressors are identical, as in the multivariate regression model, and X is given in the form $I \otimes X_0$. It is straightforward to show through direct computation that SUR is identical to SELS in these two cases.

The following theorem tells us exactly when the SUR procedure does not improve upon the SELS:

Theorem 3 *The* SUR *and* SELS *are identical if and only if*

$$\sigma_{pq} = 0 \quad \text{or} \quad \mathcal{R}(X_p) = \mathcal{R}(X_q)$$

for all $p \neq q$.

Proof By the Kruskal's theorem, the two are identical if and only if $\mathcal{R}\left((\Sigma \otimes I)X\right) = \mathcal{R}(X)$. Since $\operatorname{rank}\left((\Sigma \otimes I)X\right) = \operatorname{rank}(X)$, it suffices to show that

$$\mathcal{R}\left((\Sigma \otimes I)X\right) \subset \mathcal{R}(X)$$

which holds if and only if there exists a matrix T such that

$$(\Sigma \otimes I)X = XT$$

If we write $T = (T_{pq})$ with submatrices T_{pq}, then it implies

$$\sigma_{pq} X_q = X_p T_{pq}$$

for all p, q, which is true if and only if

$$\sigma_{pq} = 0 \quad \text{or} \quad \mathcal{R}(X_p) = \mathcal{R}(X_q)$$

for all $p \neq q$, as was to be shown. ■

2.4 Exercises

1. Consider a SUR model given by

$$\begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix} = \begin{pmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_\ell \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_\ell \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_\ell \end{pmatrix}$$

where y_k is $(n \times 1)$, X_k $(n \times m)$, β_k $(m \times 1)$ and u_k $(n \times 1)$ for all $k = 1, \ldots, \ell$. We assume that $\mathbf{E}u_k = 0$ and $\mathbf{E}u_j u_k' = \sigma_{jk} I_n$ for all $j, k = 1, \ldots, \ell$. We also assume X_k is non-random for all $k = 1, \ldots, \ell$ for simplicity. We may write the model as

$$y = X\beta + u,$$

where $y,\ X,\ \beta$ and u are defined as in the above equation. Answer the following:

(a) Derive the asymptotic distribution of SUR estimator for $\beta = (\beta_1', \ldots, \beta_\ell)'$.

(b) Show that the SUR estimator for β is the same as a stack of equation-by-equation OLS estimator when $\sigma_{jk} = 0$ for all $j \neq k$.

(c) Show that the SUR estimator for β is the same as a stack of equation-by-equation OLS estimator when $X_1 = \ldots = X_\ell$.

2. Consider the standard SUR system

$$\begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix} = \begin{pmatrix} X_1 & & \\ & \ddots & \\ & & X_\ell \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_\ell \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_\ell \end{pmatrix}$$

as specified in Question 1 above. We let $\mathbf{E}\, u_p u_q' = \sigma_{pq} I$ and $\Sigma = (\sigma_{pq})$. Also, let $\beta = (\beta_1, \ldots, \beta_\ell)'$. Explain in detail how to obtain the ML estimators of the parameters β and Σ under normality. Show rigorously, in particular, that the proposing estimators are indeed the ML estimators under normality. Recall that the density for n-dimensional multivariate normal distribution with mean μ and variance Ω is given by

$$p(x) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} (\det \Omega)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)'\Omega^{-1}(x-\mu)\right)$$

3. Consider a model

$$\begin{aligned} y_{1i} &= \alpha' x_{1i} + \varepsilon_{1i} \\ y_{2i} &= \alpha' x_{2i} + \beta' x_{3i} + \varepsilon_{2i} \end{aligned}$$

where $\{\varepsilon_i\}$, $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})'$, are iid with $\mathrm{var}(\varepsilon_i) = \sigma^2 I$. Let $\hat{\alpha}_1$ and $\hat{\alpha}_2$ be the OLS estimators based respectively on the first and second regressions, and $\hat{\alpha}$ be the system OLS estimator of α. Answer the following:

(a) Obtain $\hat{\alpha}$, and show that it may be written as

$$\hat{\alpha} = M_* \hat{\alpha}_1 + (I - M_*)\hat{\alpha}_2$$

where

$$M_* = (X_1'X_1 + X_2'(I - P_{X_3})X_2)^{-1} X_1'X_1$$

(b) Show that $\hat{\alpha}$ is uncorrelated with $(\hat{\alpha}_1 - \hat{\alpha}_2)$. Using this fact, prove that $\hat{\alpha}$ has smaller variance (in positive definite sense) than any estimator of the form

$$\tilde{\alpha} = M\hat{\alpha}_1 + (I - M)\hat{\alpha}_2$$

(Notice that $\tilde{\alpha} = \hat{\alpha} + (M - M_*)(\hat{\alpha}_1 - \hat{\alpha}_2)$).

3. Fixed Effects Models

3.1 The Model

The model we consider here is given by

$$y_{it} = \pi + \mu_i + \nu_t + x_{it}'\beta + \varepsilon_{it}$$

for $i = 1, \dots, a$, $t = 1, \dots, b$. The subscripts i and t denote individual and time, and the terms μ_i and ν_t capture individual and time effects, respectively. It is assumed that there is no interaction between them. The errors $\{\varepsilon_{it}\}$ are assumed to be uncorrelated $(0, \sigma_\varepsilon^2)$. We let $n = ab$ be the sample size.

In matrix form, the model can be written as

$$y = \iota_n \pi + (I_a \otimes \iota_b)\mu + (\iota_a \otimes I_b)\nu + X\beta + \varepsilon$$

where $y = (y_{11}, y_{12}, \dots, y_{1b}, \dots, y_{a1}, y_{a2}, \dots, y_{ab})'$, and X and ε are defined similarly from $\{x_{it}\}$ and $\{\varepsilon_{it}\}$. Moreover, $\mu = (\mu_1, \dots, \mu_a)'$ and $\nu = (\nu_1, \dots, \nu_b)'$. The identity matrix I and the vector of ones ι are written with subscripts to specify their dimensions. In the rest of this section, we use "$\bar{\ }$" with a subscript to denote the average over the other subscript. If no subscript is attached, then it implies the grand mean, i.e., the average over both i and t. For instance, $\bar{y}_i$, $\bar{y}_t$ and $\bar{y}$ denote, respectively, the averages of y_{it}'s over t, i and both i and t.

Let the individual and time effects terms μ and ν be fixed and nonrandom. Then we essentially have a model with dummy variables, and the model is indeed also called *dummy variables model*. The parameters μ and ν in the model are, however, unidentified since the regressors ι_n, $I_a \otimes \iota_b$ and $\iota_a \otimes I_b$ are linearly dependent. Therefore, we need identifying restrictions on these parameters. The most commonly used restrictions are

$$\sum_{i=1}^{a} \mu_i = \sum_{t=1}^{b} \nu_t = 0 \tag{5}$$

so that μ and ν represent the effects purely specific to individuals and times.

3.2 Estimation

We define

$$I - \frac{\iota_a \iota_a'}{a} = H_a H_a' \quad \text{and} \quad I - \frac{\iota_b \iota_b'}{b} = H_b H_b'$$

where H_a and H_b are orthogonal matrices such that $H_a' H_a = I_{a-1}$ and $H_b' H_b = I_{b-1}$, and let

$$H_a' \mu = \mu^* \quad \text{and} \quad H_b' \nu = \nu^*$$

Notice that $H_a H_a' \mu = \mu$ and $H_b H_b' \nu = \nu$ due to the identifying restrictions given in (5). We now write the model with unconstrained parameters μ^* and ν^* as

$$y = \iota_n \pi + (H_a \otimes \iota_b)\mu^* + (\iota_a \otimes H_b)\nu^* + X\beta + \varepsilon$$

This regression is very easy to handle, because the regressors ι_n, $H_a \otimes \iota_b$ and $\iota_a \otimes H_b$ are orthogonal as one can easily check. The projection P on these regressors is just the sum of projections on each regressor. More explicitly, we have

$$\begin{aligned} P &= \frac{\iota_n \iota_n'}{n} + \left(I_a - \frac{\iota_a \iota_a'}{a} \right) \otimes \frac{\iota_b \iota_b'}{b} + \frac{\iota_a \iota_a'}{a} \otimes \left(I_b - \frac{\iota_b \iota_b'}{b} \right) \\ &= I_a \otimes \frac{\iota_b \iota_b'}{b} + \frac{\iota_a \iota_a'}{a} \otimes I_b - \frac{\iota_n \iota_n'}{n} \end{aligned}$$

Let $Q = I - P$ and write the OLS estimate $\hat{\beta}$ of β as

$$\hat{\beta} = (X'QX)^{-1} X'Qy \tag{6}$$

Notice that

$$I_a \otimes \frac{\iota_b \iota_b'}{b} \ , \quad \frac{\iota_a \iota_a'}{a} \otimes I_b \ , \quad \frac{\iota_n \iota_n'}{n}$$

are the matrices that average over t, i and both i and t, respectively. It can therefore be easily seen that $\hat{\beta}$ is the OLS estimate of β in the regression

$$y_{it} - \bar{y}_i - \bar{y}_t + \bar{y} = (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})'\beta + e_{it}$$

The OLS estimates of π, μ^* and ν^* can be obtained from the regression

$$y - X\hat{\beta} = \iota_n \pi + (H_a \otimes \iota_b)\mu^* + (\iota_a \otimes H_b)\nu^* + \varepsilon$$

Since the regressors are orthogonal each other, the estimates are identical to those from three separate regressions. For the estimates of μ and ν, note that

$$\mu = H_a \mu^* \quad \text{and} \quad \nu = H_b \nu^*$$

We may explicitly write the OLS estimates of π, μ and ν as

$$\hat{\pi} = \bar{y} - \bar{x}'\hat{\beta} \ , \quad \hat{\mu}_i = (\bar{y}_i - \bar{y}) - (\bar{x}_i - \bar{x})'\hat{\beta} \ , \quad \hat{\nu}_t = (\bar{y}_t - \bar{y}) - (\bar{x}_t - \bar{x})'\hat{\beta}$$

3.3 Exercises

1. Consider the model

$$y_{it} = \mu_i + x_{it}'\beta_i + \varepsilon_{it}$$

for $i = 1, \ldots, a$ and $t = 1, \ldots, b$. Assume $\{x_{it}\}$ are m-dimensional and nonrandom, and $\{\varepsilon_{it}\}$ are i.i.d. $\mathbf{N}(0,\sigma^2)$. Construct the F-statistic to test the hypotheses $\beta_1 = \cdots = \beta_a$.

2. Consider the following regressions with dummy variables

$$\begin{aligned} y_{ij} &= \pi + \mu_j + x_{ij}'\beta + \varepsilon_{ij}, \\ y_{ij} &= \nu_j + x_{ij}'\beta + \varepsilon_{ij} \end{aligned}$$

where $\sum_j \mu_j = 0$.

(a) Show that the OLS estimates for β from the two regressions are identical.
(b) Let $\hat{\pi}$, $\hat{\mu}_j$ and $\hat{\nu}_j$ be the OLS estimates for π, μ_j and ν_j. Relate $\hat{\pi}$, $\hat{\mu}_j$ and $\hat{\nu}_j$.

3. Consider a model given by

$$y_{it} = \pi + \mu_t + \eta_i + \varepsilon_{it}$$

for $i = 1, \ldots, a$ and $t = 1, \ldots, b$, where (η_i) and (ε_{it}) are mutually uncorrelated, and iid with $\mathrm{var}(\eta_i) = \sigma_\eta^2$ and $\mathrm{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$. Answer the following:
(a) Obtain the GLS estimator for π and μ_t's for $t = 1, \ldots, b$.
(b) Compare the OLS and the GLS estimators.

4. Consider the fixed effect model

$$y_{it} = \pi + \mu_i + \nu_t + x_{it}'\beta + \varepsilon_{it}$$

for $i = 1, \ldots, a$ and $t = 1, \ldots, b$, where the errors (ε_{it}) are uncorrelated $(0, \sigma_\varepsilon^2)$.
(a) Discuss the identifiability of the parameters (μ_i) and (ν_t) representing the individual and time effects.
(b) Provide identifying restrictions on (μ_i) and (ν_t) and use them to reparameterize the given model. Discuss the identifiability of the parameters in the reparameterized model.
(c) Obtain the OLS estimators for the parameters π, (μ_i), (ν_t) and β in the original model.

4. Random Effects Models

4.1 The Model

We consider

$$y_{it} = \pi + x_t'\beta + \mu_i + \nu_t + \varepsilon_{it}$$

as in the previous section. However, we now let the terms μ and ν representing the individual and time effects be random, and included in unobserved errors. Assume $\{\mu_i\}$ and $\{\nu_t\}$ are uncorrelated $(0, \sigma_\mu^2)$ and $(0, \sigma_\nu^2)$, and are uncorrelated with $\{\varepsilon_{it}\}$. The variance of y_{it} is then the sum of variances of three components, i.e., $\sigma_\mu^2 + \sigma_\nu^2 + \sigma_\varepsilon^2$, and for this reason the model is also called *error components model*. Write the model in matrix form as

$$y = \iota_n \pi + X\beta + u$$

where

$$u = (I_a \otimes \iota_b)\mu + (\iota_a \otimes I_b)\nu + \varepsilon$$

The variance Σ of u is given by

$$\Sigma = \sigma_\mu^2 (I_a \otimes \iota_b \iota_b') + \sigma_\nu^2 (\iota_a \iota_a' \otimes I_b) + \sigma_\varepsilon^2 I_n$$

4.2 Estimation of Regression Coefficients

To analyze the GLS estimates of β and π, we define the following orthogonal pojections that are mutually orthogonal and sum up to I:

$$\begin{aligned} P_1 &= \left(I_a - \frac{\iota_a\iota_a'}{a}\right) \otimes \left(I_b - \frac{\iota_b\iota_b'}{b}\right) \\ P_2 &= \left(I_a - \frac{\iota_a\iota_a'}{a}\right) \otimes \frac{\iota_b\iota_b'}{b} \\ P_3 &= \frac{\iota_a\iota_a'}{a} \otimes \left(I_b - \frac{\iota_b\iota_b'}{b}\right) \\ P_4 &= \frac{\iota_a\iota_a'}{a} \otimes \frac{\iota_b\iota_b'}{b} \end{aligned}$$

Then we have

$$\Sigma = \sigma_\varepsilon^2 P_1 + (b\sigma_\mu^2 + \sigma_\varepsilon^2)P_2 + (a\sigma_\nu^2 + \sigma_\varepsilon^2)P_3 + (b\sigma_\mu^2 + a\sigma_\nu^2 + \sigma_\varepsilon^2)P_4$$

Write $\Sigma = \sigma_\varepsilon^2\Sigma_0$ and $\Sigma_0^{-1} = \sum_{k=1}^4 \lambda_k P_k$, where λ_k's are defined appropriately.

Conformably as in the previous section, we write

$$\hat{\beta} = (X'QX)^{-1}X'Qy \tag{7}$$

where Q is now given by

$$\begin{aligned} Q &= \Sigma_0^{-1} - \lambda_4 \frac{\iota_n\iota_n'}{n} \\ &= I_n - \varphi_1\left(I_a \otimes \frac{\iota_b\iota_b'}{b}\right) - \varphi_2\left(\frac{\iota_a\iota_a'}{a} \otimes I_b\right) + \varphi_3\frac{\iota_n\iota_n}{n} \end{aligned}$$

where $\varphi_1 = \lambda_1 - \lambda_2,\ \varphi_2 = \lambda_1 - \lambda_3,\ \varphi_3 = \lambda_1 - \lambda_2 - \lambda_3$, and are explicitly given by

$$\varphi_1 = \frac{b\sigma_\mu^2}{b\sigma_\mu^2 + \sigma_\varepsilon^2},\ \varphi_2 = \frac{a\sigma_\nu^2}{a\sigma_\nu^2 + \sigma_\varepsilon^2},\ \varphi_3 = \frac{ab\sigma_\mu^2\sigma_\nu^2 - \sigma_\varepsilon^4}{(a\sigma_\nu^2 + \sigma_\varepsilon^2)(b\sigma_\mu^2 + \sigma_\varepsilon^2)}$$

Moreover, the GLS estimate of π can be obtained from the GLS in $y - X\hat{\beta} = \iota_n\pi + u$, which is identical to the OLS estimate since ι_n is an eigenvector of Σ. It is therefore given by

$$\hat{\pi} = \bar{y} - \bar{x}'\hat{\beta} \tag{8}$$

As $a, b \to \infty$, $\varphi_1, \varphi_2, \varphi_3 \to 1$ and Q becomes the projection used in the definition of the fixed effect estimator given in (6) in the previous section. In this case, the fixed effects and random effects models yield the same estimate for β and π asymptotically.

4.3 Estimation of Error Components

The error components are typically unknown and need to be estimated. We let

$$\sigma_1^2 = \sigma_\varepsilon^2 \ , \ \sigma_2^2 = b\sigma_\mu^2 + \sigma_\varepsilon^2 \ , \ \sigma_3^2 = a\sigma_\nu^2 + \sigma_\varepsilon^2$$

and define

$$\hat{\sigma}_k^2 = \frac{(y - X\beta)' P_k (y - X\beta)}{d_k} \tag{9}$$

for $k = 1, 2, 3$, where d_k's are the dimensions of the projections P_k's which are explicitly given by $d_1 = (a-1)(b-1), d_2 = a-1$ and $d_3 = b-1$. It is easy to see that $\mathbf{E}\hat{\sigma}_k^2 = \sigma_k^2$. Notice that the π-term in the quadratic forms in (9) vanishes since P_k's are orthogonal to the constant term in the regression.

For the actual computation of $\hat{\sigma}_k^2$'s in (9), various consistent estimates of β can be used. The most obvious choice would be the OLS estimate. As an alternative, we may estimate β for each of $\hat{\sigma}_k^2$'s based on the regressions

$$\begin{aligned} y_{it} - \bar{y}_i - \bar{y}_t + \bar{y} &= (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})'\beta + e_{it} && (10) \\ \bar{y}_i - \bar{y} &= (\bar{x}_i - \bar{x})'\beta + e_i && (11) \\ \bar{y}_t - \bar{y} &= (\bar{x}_t - \bar{x})'\beta + e_t && (12) \end{aligned}$$

so that $(y - X\beta)' P_k (y - X\beta)$'s are just multiples of the RSS's of these regressions. It is also possible to estimate β from the regression (10) and use this estimate to compute the RSS's for the regressions (11) and (12). This is meaningful, since the regressions (11) and (12) are based only on a and b samples, respectively, whereas the first regression is based on the entire n samples.

In order to consider the ML estimates for the variances of the error components, note first that

$$\det \Sigma = \sigma_1^{2d_1} \sigma_2^{2d_2} \sigma_3^{2d_3} (\sigma_2^2 + \sigma_3^2 - \sigma_1^2)$$

in the notation defined above. Moreover, given the ML estimates $\hat{\beta}$ and $\hat{\pi}$ for β and π, that are just the GLS estimates given in (7) and (8), we have

$$(y - \iota_n \hat{\pi} - X\hat{\beta})' \Sigma^{-1} (y - \iota_n \hat{\pi} - X\hat{\beta}) = \sum_{k=1}^{3} \frac{(y - X\hat{\beta})' P_k (y - X\hat{\beta})}{\sigma_k^2}$$

Note that the fourth term vanishes since $P_4(y - \iota_n \hat{\pi} - X\hat{\beta}) = \bar{y} - \hat{\pi} - \bar{x}'\hat{\beta} = 0$. The presence of the term $\sigma_2^2 + \sigma_3^2 - \sigma_1^2$ in $\det \Sigma$ makes complicated the ML estimates for the variances of the error compontents. Were it not for this term, the estimates defined in (9) would yield the exact ML estimates. For the two error components model, this complication does not arise. Note also that the estimates for the σ_μ^2 and σ_ν^2 obtained from the estimates for σ_k^2's can obviously be negative and nonsensical. This problem is not solved as yet.

4.4 Exercises

1. Consider the random effect model

$$y_{it} = \pi + x'_{it}\beta + u_{it}$$

for $i = 1, \ldots, a$ and $t = 1, \ldots, b$, with the error term defined as

$$u_{it} = \mu_i + \nu_t + \varepsilon_{it}$$

including the random individual and time effects, (μ_i) and (ν_t), which are assumed to be uncorrelated $(0, \sigma_\mu^2)$ and $(0, \sigma_\nu^2)$, and also uncorrelated with (ε_{it}). Further assume that (ε_{it}) are uncorrelated $(0, \sigma_\varepsilon^2)$, and answer the following.
(a) Obtain the GLS estimators of π and β, which are the random effect estimators.
(b) Compare the random effect estimator of β obtained in (a) with the fixed effect estimator of β obtained in Question 1(c) above. When do they become identical?
(c) Explain in detail how you may obtain the feasible GLS estimator for β.

2. Suppose there are no time effects and $\nu_t = 0$. Define *within*- and *between*-group estimates $\hat{\beta}_w$ and $\hat{\beta}_b$ of β to be the LS estimates in the regressions

$$\begin{aligned} y_{it} - \bar{y}_i &= (x_{it} - \bar{x}_i)'\beta + \varepsilon_{it} \\ \bar{y}_i - \bar{y} &= (\bar{x}_i - \bar{x})'\beta + \varepsilon_i \end{aligned}$$

respectively. Show that
(a) The fixed effects model yields the within-group estimate $\hat{\beta}_w$ for β.
(b) Show using the identity $(A_1 + A_2)^{-1}(B_1 + B_2) = ((A_1 + A_2)^{-1}A_1)A_1^{-1}B_1 + ((A_1 + A_2)^{-1}A_2)A_2^{-1}B_2$ that the random effects model yields the estimate $\hat{\beta}$ of β which we may write as $\hat{\beta} = \Delta\hat{\beta}_w + (I - \Delta)\hat{\beta}_b$ for some Δ.

3. Consider the model given by

$$\begin{aligned} y_{it} &= \pi + \mu_i + x'_{it}\beta + \varepsilon_{it} \\ \mu_i &= \bar{x}'_i\alpha + \eta_i \end{aligned}$$

for $i = 1, \ldots, a$ and $t = 1, \ldots, b$, where $\{\varepsilon_{it}\}$ and $\{\eta_i\}$ are uncorrelated $(0,\sigma_\varepsilon^2)$ and $(0,\sigma_\eta^2)$, respectively. Write the model as

$$y_{it} = \pi + \bar{x}'_i\alpha + x'_{it}\beta + u_{it}$$

where $u_{it} = \eta_i + \varepsilon_{it}$. Show that the GLS estimators $\hat{\pi}, \hat{\alpha}$ and $\hat{\beta}$ of the parameters π, α and β are

$$\hat{\pi} = \bar{y} - \bar{x}'\hat{\beta}_b \ , \quad \hat{\alpha} = \hat{\beta}_b - \hat{\beta}_w \ , \quad \hat{\beta} = \hat{\beta}_w$$

where $\hat{\beta}_w$ and $\hat{\beta}_b$ are as defined in Problem 2.

4. Derive the ML estimates for the variances of the error components for the two error components model. Explain how to compute them.