Chapter 10

Multivariate Regression

10.1 Introduction

Multivariate regression is a system of regression equations. Multivariate regression is used as reduced form models for instrumental variable estimation (explored in Chaper 11), vector autoregressions (explored in Chapter 15), demand systems (demand for multiple goods), and other contexts.

Multivariate regression is also called by the name **systems of regression equations**. Closely related is the method of **Seemingly Unrelated Regressions** (SUR) which we introduce in Section 10.7.

Most of the tools of single equation regression generalize naturally to multivariate regression. A major difference is a new set of notation to handle matrix estimates.

10.2 Regression Systems

A system of linear regressions takes the form

$$y_{ji} = \boldsymbol{x}_{ji}'\boldsymbol{\beta}_j + e_{ji} \tag{10.1}$$

for variables $j = 1, ..., m$ and observations $i = 1, ..., n$, where the regressor vectors $\boldsymbol{x}_{ji}$ are $k_j \times 1$ and e_{ji} is an error. The coefficient vectors $\boldsymbol{\beta}_j$ are $k_j \times 1$. The total number of coefficients are $\overline{k} = \sum_{j=1}^{n} k_j$. The regression system specializes to univariate regression when $m = 1$.

It is typical to treat the observations as independent across observations i but correlated across variables j. As an example, the observations y_{ji} could be expenditures by household i on good j. The standard assumptions are that households are mutually independent, but expenditures by an individual household are correlated across goods.

To describe the dependence between the dependent variables, we can define the $m \times 1$ error vector $\boldsymbol{e}_i = (e_{1i}, ..., e_{mi})'$ and its $m \times m$ variance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}\left(\boldsymbol{e}_i \boldsymbol{e}_i'\right).$$

The diagonal elements are the variances of the errors e_{ji}, and the off-diagonals are the covariances across variables. It is typical to allow $\boldsymbol{\Sigma}$ to be unconstrained.

We can group the m equations (10.1) into a single equation as follows. Let $\boldsymbol{y}_i = (y_{1i}, ..., y_{mi})'$ be the $m \times 1$ vector of dependent variables, define the $\overline{k} \times m$ matrix of regressors

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{1i} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{x}_{2i} & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}_{mi} \end{pmatrix},$$

and define the $\overline{k} \times 1$ stacked coefficient vector

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_m \end{pmatrix}.$$

Then the m regression equations can jointly be written as

$$\boldsymbol{y}_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \boldsymbol{e}_i. \tag{10.2}$$

The entire system can be written in matrix notation by stacking the variables. Define

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}, \qquad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix}, \qquad \overline{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1' \\ \vdots \\ \boldsymbol{X}_n' \end{pmatrix}$$

which are $mn \times 1$, $mn \times 1$, and $mn \times \overline{k}$, respectively. The system can be written as

$$\boldsymbol{y} = \overline{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{e}.$$

In many (perhaps most) applications the regressor vectors $\boldsymbol{x}_{ji}$ are common across the variables j, so $\boldsymbol{x}_{ji} = \boldsymbol{x}_i$ and $k_j = k$. By this we mean that the same variables enter each equation with no exclusion restrictions. Several important simplifications occur in this context. One is that we can write (10.2) using the notation

$$\boldsymbol{y}_i = \boldsymbol{B}'\boldsymbol{x}_i + \boldsymbol{e}_i$$

where $\boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m)$ is $k \times m$. Another is that we can write the system in the $n \times m$ matrix notation

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_n' \end{pmatrix}, \qquad \boldsymbol{E} = \begin{pmatrix} \boldsymbol{e}_1' \\ \vdots \\ \boldsymbol{e}_n' \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}.$$

Another convenient implication of common regressors is that we have the simplification

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_i & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{x}_i & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}_i \end{pmatrix} = \boldsymbol{I}_m \otimes \boldsymbol{x}_i$$

where $\otimes$ is the Kronecker product (see Appendix A.16).

10.3 Least-Squares Estimator

Consider estimating each equation (10.1) by least-squares. This takes the form

$$\widehat{\boldsymbol{\beta}}_j = \left(\sum_{i=1}^n \boldsymbol{x}_{ji}\boldsymbol{x}_{ji}'\right)^{-1} \left(\sum_{i=1}^n \boldsymbol{x}_{ji} y_{ji}\right).$$

The combined estimate of $\boldsymbol{\beta}$ is the stacked vector

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \vdots \\ \widehat{\boldsymbol{\beta}}_m \end{pmatrix}.$$

It turns that we can write this estimator using the systems notation

$$\widehat{\boldsymbol{\beta}} = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}\left(\overline{\boldsymbol{X}}'\boldsymbol{y}\right) = \left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{y}_i\right). \tag{10.3}$$

To see this, observe that

$$\begin{aligned}
\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}} &= \begin{pmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_n \end{pmatrix}\begin{pmatrix} \boldsymbol{X}_1' \\ \vdots \\ \boldsymbol{X}_n' \end{pmatrix} \\
&= \sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i' \\
&= \sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{x}_{1i} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{x}_{2i} & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}_{mi} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}_{1i}' & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{x}_{2i}' & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}_{mi}' \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n}\boldsymbol{x}_{1i}\boldsymbol{x}_{1i}' & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \sum_{i=1}^{n}\boldsymbol{x}_{2i}\boldsymbol{x}_{2i}' & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \sum_{i=1}^{n}\boldsymbol{x}_{mi}\boldsymbol{x}_{mi}' \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
\overline{\boldsymbol{X}}'\boldsymbol{y} &= \begin{pmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_n \end{pmatrix}\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix} \\
&= \sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{y}_i \\
&= \sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{x}_{1i} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \boldsymbol{x}_{2i} & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}_{mi} \end{pmatrix}\begin{pmatrix} y_{1i} \\ \vdots \\ y_{mi} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n}\boldsymbol{x}_{1i}y_{1i} \\ \vdots \\ \sum_{i=1}^{n}\boldsymbol{x}_{mi}y_{mi} \end{pmatrix}.
\end{aligned}$$

Hence

$$\begin{aligned}
\left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}\left(\overline{\boldsymbol{X}}'\boldsymbol{y}\right) &= \left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{y}_i\right) \\
&= \begin{pmatrix} \left(\sum_{i=1}^{n}\boldsymbol{x}_{1i}\boldsymbol{x}_{1i}'\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_{1i}y_{1i}\right) \\ \vdots \\ \left(\sum_{i=1}^{n}\boldsymbol{x}_{mi}\boldsymbol{x}_{mi}'\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_{mi}y_{mi}\right) \end{pmatrix} \\
&= \widehat{\boldsymbol{\beta}}
\end{aligned}$$

as claimed.

The $m \times 1$ residual vector for the i^{th} observation is

$$\widehat{\boldsymbol{e}}_i = \boldsymbol{y}_i - \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}}$$

and the least-squares estimate of the $m \times m$ error variance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{e}}_i \widehat{\boldsymbol{e}}_i'. \tag{10.4}$$

In the case of common regressors, observe that

$$\widehat{\boldsymbol{\beta}}_j = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_{ji}\right).$$

We can set

$$\widehat{\boldsymbol{B}} = \left(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \cdots, \widehat{\boldsymbol{\beta}}_m\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Y}\right). \tag{10.5}$$

In Stata, multivariate regression can be implemented using the `mvreg` command.

10.4 Mean and Variance of Systems Least-Squares

We can calculate the finite-sample mean and variance of $\widehat{\boldsymbol{\beta}}$ under the conditional mean assumption

$$\mathbb{E}\left(\boldsymbol{e}_i \mid \boldsymbol{x}_i\right) = \mathbf{0} \tag{10.6}$$

where $\boldsymbol{x}_i$ is the union of the regressors $\boldsymbol{x}_{ji}$. Equation (10.6) is equivalent to $\mathbb{E}\left(y_{ji} \mid \boldsymbol{x}_i\right) = \boldsymbol{x}_{ji}'\boldsymbol{\beta}_j$, or that the regression model is correctly specified.

We can center the estimator as

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}\left(\overline{\boldsymbol{X}}'\boldsymbol{e}\right) = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{e}_i\right).$$

Taking conditional expectations, we find $\mathbb{E}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \boldsymbol{\beta}$. Consequently, systems least-squares is unbiased under correct specification.

To compute the variance of the estimator, define the conditional covariance matrix of the errors of the i^{th} observation

$$\mathbb{E}\left(\boldsymbol{e}_i \boldsymbol{e}_i' \mid \boldsymbol{x}_i\right) = \boldsymbol{\Sigma}_i$$

which in general is unrestricted. Observe that if the observations are mutually independent, then

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right) &= \mathbb{E}\left(\begin{pmatrix} \boldsymbol{e}_1\boldsymbol{e}_1 & \boldsymbol{e}_1\boldsymbol{e}_2 & \cdots & \boldsymbol{e}_1\boldsymbol{e}_n \\ \vdots & \ddots & & \vdots \\ \boldsymbol{e}_n\boldsymbol{e}_1 & \boldsymbol{e}_n\boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n\boldsymbol{e}_n \end{pmatrix} \mid \boldsymbol{X}\right) \\
&= \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_n \end{pmatrix}.
\end{aligned}$$

Also, by independence across observations,

$$\operatorname{var}\left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{e}_i \mid \boldsymbol{X}\right) = \sum_{i=1}^{n} \operatorname{var}\left(\boldsymbol{X}_i \boldsymbol{e}_i \mid \boldsymbol{x}_i\right) = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{\Sigma}_i \boldsymbol{X}_i'.$$

It follows that

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{\Sigma}_i \boldsymbol{X}_i'\right)\left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}.$$

When the regressors are common so that $\boldsymbol{X}_i = \boldsymbol{I}_m \otimes \boldsymbol{x}_i$ then the covariance matrix can be written as

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \left(\boldsymbol{I}_n \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)\left(\sum_{i=1}^{n}\left(\boldsymbol{\Sigma}_i \otimes \boldsymbol{x}_i\boldsymbol{x}_i'\right)\right)\left(\boldsymbol{I}_m \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right).$$

Alternatively, if the errors are conditionally homoskedastic

$$\mathbb{E}\left(\boldsymbol{e}_i\boldsymbol{e}_i' \mid \boldsymbol{x}_i\right) = \boldsymbol{\Sigma} \tag{10.7}$$

then the covariance matrix takes the form

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{\Sigma}\boldsymbol{X}_i'\right)\left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}.$$

If both simplifications (common regressors and conditional homoskedasticity) hold then we have the considerable simplication

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \boldsymbol{\Sigma} \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

10.5 Asymptotic Distribution

For an asymptotic distribution it is sufficient to consider the equation-by-equation projection model in which case

$$\mathbb{E}\left(\boldsymbol{x}_{ji}e_{ji}\right) = \boldsymbol{0}. \tag{10.8}$$

First, consider consistency. Since $\widehat{\boldsymbol{\beta}}_j$ are the standard least-squares estimators, they are consistent for the projection coefficients $\boldsymbol{\beta}_j$.

Second, consider the asymptotic distribution. Again by our single equation theory it is immediate that the $\widehat{\boldsymbol{\beta}}_j$ are asymptotically normally distributed. But our previous theory does not provide a joint distribution of the $\widehat{\boldsymbol{\beta}}_j$ across j. For this we need a joint theory for the stacked estimates $\widehat{\boldsymbol{\beta}}$, which we now provide.

Since the vector

$$\boldsymbol{X}_i\boldsymbol{e}_i = \begin{pmatrix} \boldsymbol{x}_{1i}e_{1i} \\ \vdots \\ \boldsymbol{x}_{mi}e_{mi} \end{pmatrix}$$

is i.i.d. across i and mean zero under (10.8), the central limit theorem implies

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{e}_i\right) \xrightarrow{d} \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{\Omega}\right)$$

where

$$\boldsymbol{\Omega} = \mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{X}_i'\right) = \mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{\Sigma}_i\boldsymbol{X}_i'\right).$$

The matrix $\boldsymbol{\Omega}$ is the covariance matrix of the variables $\boldsymbol{x}_{ji}e_{ji}$ across equations. Under conditional homoskedasticity (10.7) the matrix $\boldsymbol{\Omega}$ simplifies to

$$\boldsymbol{\Omega} = \mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{\Sigma}\boldsymbol{X}_i'\right) \tag{10.9}$$

(see Exercise 10.1). When the regressors are common then it simplies to

$$\boldsymbol{\Omega} = \mathbb{E}\left(\boldsymbol{e}_i\boldsymbol{e}_i' \otimes \boldsymbol{x}_i\boldsymbol{x}_i'\right) \tag{10.10}$$

(see Exercise 10.2) and under both conditions (homoskedasticity and common regressors) it simplifies to

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbb{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right) \tag{10.11}$$

(see Exercise 10.3).

Applied to the centered and normalized estimator we obtain the asymptotic distribution.

Theorem 10.5.1 *Under Assumption 7.1.2,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{Q}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}^{-1}$$

$$\boldsymbol{Q} = \mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right) = \begin{pmatrix} \mathbb{E}\left(\boldsymbol{x}_{1i}\boldsymbol{x}_{1i}'\right) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{E}\left(\boldsymbol{x}_{ni}\boldsymbol{x}_{ni}'\right) \end{pmatrix}$$

For a proof, see Exercise 10.4.

When the regressors are common then the matrix $\boldsymbol{Q}$ simplies as

$$\boldsymbol{Q} = \boldsymbol{I}_m \otimes \mathbb{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right) \tag{10.12}$$

(See Exercise 10.5).

If both the regressors are common and the errors are conditionally homoskedastic (10.7) then we have the simplication

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma} \otimes \left(\mathbb{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right)\right)^{-1} \tag{10.13}$$

(see Exercise 10.6).

Sometimes we are interested in parameters $\boldsymbol{\theta} = r(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m) = r(\boldsymbol{\beta})$ which are functions of the coefficients from multiple equations. In this case the least-squares estimate of $\boldsymbol{\theta}$ is $\widehat{\boldsymbol{\theta}} = r(\widehat{\boldsymbol{\beta}})$. The asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ can be obtained from Theorem 10.5.1 by the delta method.

Theorem 10.5.2 *Under Assumptions 7.1.2 and 7.10.1,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\theta}}\right)$$

where

$$\boldsymbol{V}_{\boldsymbol{\theta}} = \boldsymbol{R}'\boldsymbol{V}_{\boldsymbol{\beta}}\boldsymbol{R}$$

$$\boldsymbol{R} = \frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{r}\left(\boldsymbol{\beta}\right)'$$

For a proof, see Exercise 10.7.

Theorem 10.5.2 is an example where multivariate regression is fundamentally distinct from univariate regression. Only by treating the least-squares estimates as a joint estimator can we obtain a distributional theory for an estimator $\widehat{\boldsymbol{\theta}}$ which is a function of estimates from multiple equations and thereby construct standard errors, confidence intervals, and hypothesis tests.

10.6 Covariance Matrix Estimation

From the finite sample and asymptotic theory we can construct appropriate estimators for the variance of $\widehat{\boldsymbol{\beta}}$. In the general case we have

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i \widehat{\boldsymbol{e}}_i \widehat{\boldsymbol{e}}_i' \boldsymbol{X}_i'\right) \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}.$$

Under conditional homoskedasticity (10.7) an appropriate estimator is

$$\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}} = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i \widehat{\boldsymbol{\Sigma}} \boldsymbol{X}_i'\right) \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}.$$

When the regressors are common then these estimators equal

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} = \left(\boldsymbol{I}_n \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \left(\sum_{i=1}^{n} \left(\widehat{\boldsymbol{e}}_i \widehat{\boldsymbol{e}}_i' \otimes \boldsymbol{x}_i \boldsymbol{x}_i'\right)\right) \left(\boldsymbol{I}_n \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)$$

and

$$\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},$$

respectively.

Covariance matrix estimators for $\widehat{\boldsymbol{\theta}}$ are found as

$$\begin{aligned} \widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\theta}}} &= \widehat{\boldsymbol{R}}' \widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} \widehat{\boldsymbol{R}} \\ \widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\theta}}} &= \widehat{\boldsymbol{R}}' \widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}} \widehat{\boldsymbol{R}} \\ \widehat{\boldsymbol{R}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}\right)'. \end{aligned}$$

Theorem 10.6.1 *Under Assumption 7.1.2,*

$$n\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}$$

and

$$n\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{V}^0_{\boldsymbol{\beta}}$$

For a proof, see Exercise 10.8.

10.7 Seemingly Unrelated Regression

Consider the systems regression model under the conditional mean and conditional homoskedasticity assumptions

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{X}_i' \boldsymbol{\beta} + \boldsymbol{e}_i \qquad (10.14) \\ \mathbb{E}\left(\boldsymbol{e}_i \mid \boldsymbol{x}_i\right) &= \boldsymbol{0} \\ \mathbb{E}\left(\boldsymbol{e}_i \boldsymbol{e}_i' \mid \boldsymbol{x}_i\right) &= \boldsymbol{\Sigma} \end{aligned}$$