Lecture note 1

Reference: Davidson and MacKinnon Ch 2. In particular page 57-82.

Projection matrices

The matrix

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{1}$$

is often called the "residual maker". That nickname is easy to understand, since:

$$\begin{aligned}\mathbf{My} &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\varepsilon}}\end{aligned}$$

$\mathbf{M}$ plays a central role in many derivations. The following properties are worth noting (and showing for yourself) :

$$\mathbf{M} = \mathbf{M}', \text{ symmetric matrix} \tag{2}$$

$$\mathbf{M}^2 = \mathbf{M}, \text{ idempotent matrix} \tag{3}$$

$$\mathbf{MX} = \mathbf{0}, \text{ regression of } X \text{ on } X \text{ gives perfect fit} \tag{4}$$

Another way of interpreting $\mathbf{MX} = \mathbf{0}$ is that since $\mathbf{M}$ produces the OLS residuals, $\mathbf{M}$ is orthogonal to $\mathbf{X}$.

$\mathbf{M}$ does not affect the residuals:

$$\mathbf{M}\hat{\boldsymbol{\varepsilon}} = \mathbf{M}(\mathbf{y} - \mathbf{X}\,\hat{\boldsymbol{\beta}}) = \mathbf{My} - M\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\varepsilon}} \tag{5}$$

Since $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, we also have:

$$\mathbf{M}\,\boldsymbol{\varepsilon} = \mathbf{M}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \hat{\boldsymbol{\varepsilon}}. \tag{6}$$

Note that this gives:

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} \tag{7}$$

whihc show that SSR is a quadratic form in the theoretical disturbances (See lecture note 2 for use of this).

A second important matrix in regression analysis is:

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{8}$$

which is called the "prediction matrix", since

$$\hat{\mathbf{y}} = \mathbf{X}\,\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{Py} \tag{9}$$

$\mathbf{P}$ is also symmetric and idempotent.

In linear algebra, $\mathbf{M}$ and $\mathbf{P}$ are both known as projection matrices, Ch 2 in DM, page 57, in particular gives the geometric interpretation.

$\mathbf{M}$ and $\mathbf{P}$ are orthogonal:

$$\mathbf{MP} = \mathbf{PM} = \mathbf{0} \tag{10}$$

Since

$$\mathbf{M} = \mathbf{I} - \mathbf{P},$$

$\mathbf{M}$ and $\mathbf{P}$ are also complementary projections

$$\mathbf{M}+\mathbf{P}=\mathbf{I} \tag{11}$$

which gives

$$(\mathbf{M}+\mathbf{P})\mathbf{y}=\mathbf{I}\mathbf{y}$$

and therefore:

$$\mathbf{y}=\hat{\mathbf{y}}+\hat{\varepsilon}=\mathbf{P}\mathbf{y}+\mathbf{M}\mathbf{y}=\hat{\mathbf{y}}+\hat{\varepsilon} \tag{12}$$

The scalar product $\mathbf{y}'\mathbf{y}$ then becomes:

$$\begin{aligned}\mathbf{y}'\mathbf{y}&=(\mathbf{y}'\mathbf{P}+\mathbf{y}'\mathbf{M})(\mathbf{P}\mathbf{y}+\mathbf{M}\mathbf{y})\\ &=\hat{\mathbf{y}}'\hat{\mathbf{y}}+\hat{\varepsilon}'\hat{\varepsilon}\end{aligned}$$

Written out, this is:

$$\underbrace{\sum_{i=1}^{n}Y_i{}^2}_{\text{"}TSS\text{"}}=\underbrace{\sum_{i=1}^{n}\hat{Y}_i{}^2}_{\text{"}ESS\text{"}}+\underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}_{SSR} \tag{13}$$

You may be more used to write this famous decomposition as:

$$\underbrace{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}_{TSS}=\underbrace{\sum_{i=1}^{n}(\hat{Y}_i-\overline{\hat{Y}})^2}_{ESS}+\underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}_{SSR} \tag{14}$$

Question: Why is there no conflict between (13) and (14) as long as $\mathbf{X}$ contains a vector of ones (an intercept)?

Alternative ways of writing SSR that sometimes come in handy:

$$\hat{\varepsilon}'\hat{\varepsilon}=(\mathbf{y}'\mathbf{M}')\mathbf{M}\mathbf{y}=\mathbf{y}'\mathbf{M}\mathbf{y}=\mathbf{y}'\hat{\varepsilon}=\hat{\varepsilon}'\mathbf{y} \tag{15}$$

$$\begin{aligned}\hat{\varepsilon}'\hat{\varepsilon}&=\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y}=\mathbf{y}'\mathbf{y}-\mathbf{y}'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}\\ &=\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}=\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{y}'\mathbf{y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}.\end{aligned} \tag{16}$$

Partitioned regression

We now let $\boldsymbol{\beta}=\begin{bmatrix}\hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\beta}}_2\end{bmatrix}'$ where $\boldsymbol{\beta}_1$ has k_1 parameters and $\boldsymbol{\beta}_2$ has k_2 parameters. $k_1+k_2=k$. The corresponding partitioning of $\mathbf{X}$ is $\mathbf{X}=\begin{bmatrix}\underset{n\times k_1}{\mathbf{X}_1} & \underset{n\times k_2}{\mathbf{X}_2}\end{bmatrix}$. The OLS estimated model with this notation:

$$\begin{aligned}\mathbf{y}&=\mathbf{X_1}\hat{\boldsymbol{\beta}}_1+\mathbf{X_2}\hat{\boldsymbol{\beta}}_2+\hat{\varepsilon},\\ \hat{\varepsilon}&=\mathbf{M}\mathbf{y}.\end{aligned} \tag{17}$$

We begin by writing out the normal equations:

$$(\mathbf{X}'\mathbf{X})\,\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y}$$

in terms of the partitioning:

$$\begin{pmatrix}\mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\beta}}_\mathbf{1}\\ \hat{\boldsymbol{\beta}}_\mathbf{2}\end{pmatrix}=\begin{pmatrix}\mathbf{X}_1'\mathbf{y}\\ \mathbf{X}_2'\mathbf{y}\end{pmatrix}. \tag{18}$$

where we have used

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}\mathbf{X}_1'\\ \mathbf{X}_2'\end{bmatrix}\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$$

see page 20 and Exercise 1.9 in DM. The vector of OLS estimators can therefore be expressed as:

$$\begin{pmatrix} \hat{\boldsymbol\beta}_1 \\ \hat{\boldsymbol\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix}. \tag{19}$$

We need a result for the inverse of a partitioned matrix. There are several versions, but probably the most useful for us is:

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}_{11\cdot 2}^{-1} & -\mathbf{A}_{11\cdot 2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22\cdot 1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22\cdot 1}^{-1} \end{pmatrix} \tag{20}$$

where

$$\mathbf{A}_{11\cdot 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \tag{21}$$

$$\mathbf{A}_{22\cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}. \tag{22}$$

With the aid of (20)-(22) it can be shown that

$$\begin{pmatrix} \hat{\boldsymbol\beta}_1 \\ \hat{\boldsymbol\beta}_2 \end{pmatrix} = \begin{pmatrix} (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \\ (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y} \end{pmatrix}, \tag{23}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are the two residual-makers:

$$\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' \tag{24}$$

$$\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'. \tag{25}$$

As an application, start with:

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{X}_1 & \vdots & \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol\iota & \vdots & \mathbf{X}_2 \end{bmatrix}$$

where

$$\boldsymbol\iota = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1}$$

$$\mathbf{X}_2 = \begin{bmatrix} X_{12} & \dots & X_{1k} \\ X_{22} & \dots & X_{2k} \\ \vdots & & \vdots \\ X_{n2} & \dots & X_{nk} \end{bmatrix}$$

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{I} - \boldsymbol\iota(\boldsymbol\iota'\boldsymbol\iota)^{-1}\boldsymbol\iota' \\ &= \mathbf{I} - \boldsymbol\iota(\boldsymbol\iota'\boldsymbol\iota)^{-1}\boldsymbol\iota' \end{aligned}$$

$\boldsymbol\iota'\boldsymbol\iota = n$ so

$$\mathbf{M}_1 = \mathbf{I} - \frac{1}{n}\boldsymbol\iota\boldsymbol\iota' \equiv \mathbf{M}_\iota$$

the so-called centering matrix. Show that it is:

$$\mathbf{M}_\iota = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix}_{n\times n}$$

By using the general formula for partitioned regression above, we get

$$\begin{aligned}\hat{\boldsymbol{\beta}}_2 &= \left(\mathbf{X_2'}\mathbf{M}_1\mathbf{X_2}\right)^{-1}\mathbf{X_2'}\mathbf{M}_1\mathbf{y}\\ &= \left([\mathbf{M}_\iota\mathbf{X_2}]'\,[\mathbf{M}_\iota\mathbf{X_2}]\right)^{-1}[\mathbf{M}_\iota\mathbf{X_2}]'\,\mathbf{y}\\ &= \left[(\mathbf{X}_2-\bar{\mathbf{X}}_2)'(\mathbf{X}_2-\bar{\mathbf{X}}_2)\right]^{-1}(\mathbf{X}_2-\bar{\mathbf{X}}_2)'\mathbf{y} \qquad (26)\end{aligned}$$

where $\bar{\mathbf{X}}_2$ is the $n\times(k-1)$ matrix with the variable means in the $k-1$ columns.

In the exercise set to Seminar 1 you are invited to show the same result for $\hat{\boldsymbol{\beta}}_2$ more directly, without the use of the results from partitioned regression, by re-writing the regression model.

Frisch-Waugh-Lovell (FWL) theorem

The expression for $\hat{\boldsymbol{\beta}}_2$ in (23) suggests that there is another simple method for finding $\hat{\boldsymbol{\beta}}_2$ that involves $\mathbf{M}_1$: We premultiply the model (17), first with $\mathbf{M}_1$ and then with $\mathbf{X_2'}$:

$$\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\varepsilon} \qquad (27)$$

$$\mathbf{X}_2'\mathbf{M}_1\mathbf{y} = \mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 \qquad (28)$$

where we have used $\mathbf{M}_1\boldsymbol{X}_1 = \mathbf{0}$ and $\mathbf{M}_1\hat{\varepsilon} = \hat{\varepsilon}$ in (27) and $\mathbf{X}_2'\hat{\varepsilon} = \mathbf{0}$ in (27), since the influence of the k_2 second variables has already been regressed out. We can find $\hat{\boldsymbol{\beta}}_2$ from this normal equation:

$$\hat{\boldsymbol{\beta}}_2 = \left(\mathbf{X_2'}\mathbf{M}_1\mathbf{X_2}\right)^{-1}\mathbf{X_2'}\mathbf{M}_1\mathbf{y} \qquad (29)$$

It can be shown that the covariance matrix is

$$\widehat{Cov\left(\hat{\boldsymbol{\beta}}_2\right)} = \hat{\sigma}^2\left(\mathbf{X_2'}\mathbf{M}_1\mathbf{X_2}\right)^{-1}.$$

where $\hat{\sigma}^2$ denotes the unbiased estimator for the common variance of the disturbances:

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k}.$$

The corresponding expressions for $\hat{\boldsymbol{\beta}}_1$ is:

$$\hat{\boldsymbol{\beta}}_1 = \left(\mathbf{X_1'}\mathbf{M}_2\mathbf{X_1}\right)^{-1}\mathbf{X_1'}\mathbf{M}_2\mathbf{y} \qquad (30)$$

$$\widehat{Cov\left(\hat{\boldsymbol{\beta}}_1\right)} = \hat{\sigma}^2\left(\mathbf{X_1'}\mathbf{M}_2\mathbf{X}_1\right)^{-1}.$$

Interpretation of (29):Use symmetry and idempotency of $\mathbf{M}_1$:

$$\hat{\boldsymbol{\beta}}_2 = \left((\mathbf{M}_1\mathbf{X}_2)'(\mathbf{M}_1\mathbf{X}_2)\right)^{-1}(\mathbf{M}_1\mathbf{X}_2)'\mathbf{M}_1\mathbf{y}$$

Here, $\mathbf{M}_1\mathbf{y}$ is the residual vector $\hat{\boldsymbol{\varepsilon}}_{y|X_1}$ and $\mathbf{M}_1\mathbf{X}_2$ are the matrix with k_2 residuals obtained from regressing each variable in $\mathbf{X}_2$ on $\mathbf{X}_1$, $\hat{\boldsymbol{\varepsilon}}_{X_2|X_1}$. But this means that $\hat{\boldsymbol{\beta}}_2$ can be obtained by first running these two auxiliary regressions,saving the residuals in $\hat{\boldsymbol{\varepsilon}}_{Y|X_1}$ and $\hat{\boldsymbol{\varepsilon}}_{X_2|X_1}$and then running a second regression with $\hat{\boldsymbol{\varepsilon}}_{Y|X_1}$ as regressand, and with $\hat{\boldsymbol{\varepsilon}}_{X_2|X_1}$ as the matrix with regressors:

$$\hat{\boldsymbol{\beta}}_2 = \left(\underbrace{(\mathbf{M}_1\mathbf{X}_2)'}_{\hat{\boldsymbol{\varepsilon}}_{X_2|X_1}}\underbrace{(\mathbf{M}_1\mathbf{X}_2)}_{\hat{\boldsymbol{\varepsilon}}_{X_2|X_1}}\right)^{-1}\underbrace{(\mathbf{M}_1\mathbf{X}_1)'}_{\hat{\boldsymbol{\varepsilon}}_{X_2|X_1}}\underbrace{(\mathbf{M}_1\mathbf{y}_2)}_{\hat{\boldsymbol{\varepsilon}}_{Y|X_1}} \qquad (31)$$

We have shown the Frisch-Waugh(-Lovell) theorem. Followers of ECON 4150 spring 2013 will have seen a scalar version of this. The 4150 lecture note is still on that course page.

In sum: If we want to estimate the partial effects on Y of changes in the variables in $\mathbf{X}_2$ (i.e., controlled for the influence of $\mathbf{X}_1$), we can do that in two ways. Either by estimating the full multivariate model

$$\mathbf{y} = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\varepsilon}$$

or, by following the above two-step procedure of obtaining residuals with $\mathbf{X}_1$ as regressor and then obtaining $\hat{\boldsymbol{\beta}}_2$ from (31).

Other applications:

- Trend correction
- Seasonal adjustment by seasonal dummy variables (DM p 69)
- Leverage (or influential observations), DM p 76-77