London School of Economics and Political Science — Marcia Schafgans
Department of Economics — Michaelmas 2004

EC221: Principles of Econometrics
Handout 3: Classical Linear Regression Model
(Multiple Linear Regression)

1 Classical Linear Regression Model

1.1 Introduction

General form of the multiple linear regression model is:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i,\ i = 1, ..., n$$

$$y = X\beta + \varepsilon$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ & \ddots & \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Typically, the first column consists of ones, i.e., $(x_{11}, ..., x_{n1})' = (1, ..., 1)'$. This signifies the presence of an intercept in our regression.

1.2 Assumptions

(A.1) True model is $y = X\beta + \varepsilon$ *(functional form)*

(A.2) $E(\varepsilon) = 0$ *(zero mean)*

(A.3) $Var(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_n$ *(homoskedasticity and non-autocorrelation)*

(A.4) X is a non-stochastic $n \times k$ matrix, with rank $k \leq n$

(A.5) $\varepsilon \sim N(0, \sigma^2 I_n)$

Note: Assumption (A.4) guarantees that $X'X$ is non-singular. It ensures that no exact collinearity exists between the explanatory variables; i.e. there does not exist an exact linear relationship between them that holds for all i.

This assumption rules out the following behaviour. Suppose $k = 3$, $x_{i2} = 2x_{i3}$ and $x_{i1} = 1$ for all i. Then

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\ &= \beta_1 + \beta_2(2x_{i3}) + \beta_3 x_{i3} + \varepsilon_i \\ &= \beta_1 + \overline{\beta}_3 x_{i3} + \varepsilon_i, \end{aligned}$$

where $\overline{\beta}_3 = 2\beta_2 + \beta_3$. The parameter $\overline{\beta}_3$ can be easily estimated but we say that the parameters β_2 and β_3 cannot be separately identified.

2 OLS Estimation

An algebraic tool: Given a sample on y_i and characteristics $x_{i1}, .., x_{ik}$ we may ask: which linear combination of the characteristics would give a good approximation of y_i for all i. We can consider any arbitrary linear combination, which can be written as

$$b_1 x_{i1} + b_2 x_{i2} + ... + b_k x_{ik}$$

Our aim is to minimise the difference between an observed value y_i and its linear approximation, that is

$$\begin{aligned} &y_i - b_1 x_{i1} - b_2 x_{i2} - ... - b_k x_{ik}, \text{ or} \\ &y_i - x_i' b \end{aligned}$$

where

$$x_i = (x_{i1}, ..., x_{ik})' \text{ and } b = (b_1, ..., b_k)'.$$

Thus, we would like to choose values for $b_1, b_2, ..., b_k$ such that these differences are small for all i. Although different measures can be used to define what we mean by ‘small’, the most common approach is to choose b such that the sum of squared differences is as small as possible. This approach is referred to as the ordinary least squares or OLS approach.

Let $x_{i1} = 1$ for all i ($\forall i$) in the sequel (i.e., assume that the model contains an intercept)

$$\left(\widehat{\beta}_1, ..., \widehat{\beta}_k\right)' = \underset{\beta_1,\beta_2,...,\beta_k}{\arg\min} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_{i2} - ... - \beta_k x_{ik})^2, \text{ or}$$
$$\left(\widehat{\beta}_1, ..., \widehat{\beta}_k\right)' = \underset{\beta_1,\beta_2,...,\beta_k}{\arg\min} (y - X\beta)'(y - X\beta)$$

The Normal equations (FOC) are given by:

$$\begin{pmatrix} \sum_{i=1}^{n} \widehat{\varepsilon}_i \\ \sum_{i=1}^{n} x_{i2}\widehat{\varepsilon}_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}\widehat{\varepsilon}_i \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{12} & \cdots & x_{n2} \\ & \vdots & \\ x_{1k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \widehat{\varepsilon}_1 \\ \widehat{\varepsilon}_2 \\ \vdots \\ \widehat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where

$$\widehat{\varepsilon}_i = y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_{i2} - ... - \widehat{\beta}_k x_{ik},$$

Or,

$$\underset{\text{(kxn) (nx1) (kx1)}}{X'\widehat{\varepsilon} = 0} \qquad \left\{ \begin{array}{l} k \text{ equations to determine} \\ \widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_k \end{array} \right.$$

(**residuals orthogonal to regressors**)

The OLS estimator:

Since $\widehat{\varepsilon} = y - X\widehat{\beta}$

$$X'\widehat{\varepsilon} = 0 \text{ gives } \Rightarrow \quad X'(y - X\widehat{\beta}) = 0$$
$$X'y - X'X\widehat{\beta} = 0$$

$$X'X\widehat{\beta} = X'y \Rightarrow$$
$$\boxed{\widehat{\beta} = (X'X)^{-1}X'y}$$

provided $X'X$ is non-singular, i.e. $\det(X'X) \neq 0$.

If $X'X$ is singular, it cannot be inverted, so that the estimator $\widehat{\beta}$ cannot be computed. We mentioned before that in this case the true parameter vector β is not identified. (Analogy in the simple linear regression model, where $\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$ as long as $\Sigma(x_i - \overline{x})^2 \neq 0$)

Derive normal equations directly using matrix algebra

$$\widehat{\beta} = \arg\min_{\beta}(y - X\beta)'(y - X\beta) = \arg\min_{\beta} S(\beta)$$
$$\arg\min_{\beta}\{y'y - 2\beta'X'y + \beta'X'X\beta\} \text{ (Note } \beta'X'y = y'X\beta \text{ (scalar))}$$

$$FOC : \frac{\partial S(\widehat{\beta})}{\partial \beta} = \boxed{-2X'y + 2X'X\widehat{\beta} = 0} \quad \text{(See Problem Set 1 question 6)}$$
$$\left(\text{or } -2X'\widehat{\varepsilon} = 0 \text{ since } \widehat{\varepsilon} = y - X\widehat{\beta}\right)$$
$$\Rightarrow \widehat{\beta} = (X'X)^{-1}X'y$$

$$SOC \ : \text{is satisfied since } \boxed{X'X \text{ is positive definite}} \text{ (see Problem Set question 5)}$$

3 Finite Sample Properties of $\widehat{\beta}$

Using only A.1 through A.4, we can establish that the least squares estimators of the unknown parameters β

$$\boxed{\widehat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon}$$

have the following exact, finite sample properties

3.1 Unbiasedness of $\widehat{\beta}$

This means that in repeated sampling, we can expect that our estimator is on average equal to the true value β.

$$\begin{aligned} \widehat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'\varepsilon \end{aligned}$$

Take expectations (X is non-stochastic)

$$\begin{aligned} E(\widehat{\beta}) &= E(\beta + (X'X)^{-1}X'\varepsilon) \\ &= \beta + (X'X)^{-1}X'E(\varepsilon) = \beta \text{ b/c } E(\varepsilon) = 0 \end{aligned}$$

3.2 Variance of $\widehat{\beta}$

In addition to knowing that we are, on average, correct, we would also like to make statements about how (un)likely it is to be far off in a given sample. We would like to know the sampling distribution of $\widehat{\beta}$.

$$\begin{aligned}
\operatorname{Var}(\widehat{\beta}) &= E((\widehat{\beta} - E(\widehat{\beta}))(\widehat{\beta} - E(\widehat{\beta}))') \\
&= E((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)') \ ((k \times k) \quad \text{matrix}) \\
&= \begin{pmatrix} \operatorname{Var}(\widehat{\beta}_1) & \operatorname{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \cdots & \operatorname{Cov}(\widehat{\beta}_1, \widehat{\beta}_k) \\ & \operatorname{Var}(\widehat{\beta}_2) & & \\ & & \ddots & \\ & & & \operatorname{Var}(\widehat{\beta}_k) \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
&\widehat{\beta} - \beta = (X'X)^{-1}X'\varepsilon \\
&(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' = (X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}
\end{aligned}$$

Take expectations (X is non-stochastic)

$$\begin{aligned}
E((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)') &= E((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}) \\
&= (X'X)^{-1}X'(E\varepsilon\varepsilon')X(X'X)^{-1} \\
&= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

3.3 Best Linear Unbiased Estimator (BLUE) $\widehat{\beta}$

Gauss-Markov Theorem: Given our assumptions (A.1)–(A.4) the OLS estimator $\widehat{\beta}$ is the best linear unbiased estimator of β.

Proof of Gauss-Markov Theorem

Show $\widehat{\beta} = (X'X)^{-1}X'y$ is BLUE

- Let $\underset{(k\times 1)}{\widetilde{\beta}} \underset{}{=} \underset{(k\times n)}{C}\,\underset{(n\times 1)}{y}$ be another linear unbiased estimator. For $\widetilde{\beta}$ to be unbiased

$$E(\widetilde{\beta}) = E(Cy) = E(CX\beta + C\varepsilon) = CX\beta \equiv \beta$$

which implies $CX = I \Rightarrow \widetilde{\beta} = CX\beta + C\varepsilon = \beta + C\varepsilon$

- Want to show: $\text{Var}(\widehat{\beta}) \leq \text{Var}(\widetilde{\beta})$

$$\text{Var}(\widetilde{\beta}) = E((\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)') = E(C\varepsilon\varepsilon' C')$$
$$\text{Var}(\widetilde{\beta}) = CE(\varepsilon\varepsilon')C' = \sigma^2 CC' \text{ b/c } E(\varepsilon\varepsilon') = \sigma^2 I_n$$

$$\begin{aligned}
\text{Var}(\widetilde{\beta}) - \text{Var}(\widehat{\beta}) &= \sigma^2(CC' - (X'X)^{-1}) \\
&= \sigma^2(CC' - \underbrace{CX}_{\substack{=I \\ \text{unbiasedness}}} (X'X)^{-1} \underbrace{X'C'}_{\substack{=I \\ \text{unbiasedness}}}) \\
&= \sigma^2 C(I_n - X(X'X)^{-1}X')C' \\
&= \sigma^2 CMC' \qquad M \text{ idempotent and symmetric} \\
&= \sigma^2 CMM'C' \\
&= \sigma^2 DD' \qquad \text{positive semi-definite matrix,} \\
\text{b/c } \forall a,\ a'DD'a &= z'z = \sum z_i^2 \geq 0, \text{ where } z = D'a.
\end{aligned}$$

The result also applies to any linear combination of the elements of β, i.e.

$$\text{Var}(c'\widehat{\beta}) \leq \text{Var}(c'\widetilde{\beta})$$

where $c = (c_1, ..., c_k)'$ is a $k \times 1$ vector of arbitrary constants.

Corollary: For any vector of constants, c, the minimum variance linear unbiased estimator of $c'\beta$ in the classical regression model (given assumptions (A.1)–(A.4)) is $c'\widehat{\beta}$, where $\widehat{\beta}$ is the least squares estimator, e.g.

$$c = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i^{th} \text{ position}$$
$$\Rightarrow c'\beta = \beta_i$$
$$\text{Var}(\widehat{\beta}_i) \leq \text{Var}(\widetilde{\beta}_i)$$

Each coefficient is estimated at least as efficiently by $\widehat{\beta}$ as by any other linear unbiased estimator.

3.4 Exact Sampling Distribution of $\widehat{\beta}$

Recall: The precise way in which estimators reflect the population values defines the **Sampling Distribution** of the estimator. If another sample was drawn under identical conditions, different values would be obtained. The sampling distribution is used to make **Inferences** about the population.

Note that the estimator of $\widehat{\beta}$ is a linear combination of the errors, ε

$$\widehat{\beta} = \beta + (X'X)^{-1}X'\varepsilon.$$

Thus, assuming that the errors are normally distributed, $\widehat{\beta}$ will be normally distributed as well (any linear combination of normal random variables is normal again). Above we already showed that $\widehat{\beta}$ is unbiased (i.e. the mean of $\widehat{\beta}$ equals β), and that its variance equals $\sigma^2(X'X)^{-1}$. So,

$$\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

From this it also follows that each element in $\widehat{\beta}$ is normally distributed

$$\widehat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj}), \text{ where } c_{jj} = \left[(X'X)^{-1}\right]_{jj},$$

i.e. c_{jj} is the (j, j) element of the $k \times k$ matrix $(X'X)^{-1}$.

3.5 Minimum variance unbiased estimator of $\widehat{\beta}$

We showed in the simple linear regression model that under normality, our OLS estimator equals the Maximum Likelihood estimator. This holds for the multiple linear regression model as well, $\widehat{\beta} = \widehat{\beta}_{MLE}$. This implies that $\widehat{\beta}$ is not only the Best Linear Unbiased Estimator (BLUE), but it is the best estimator in the class of all unbiased estimators (MVU). $\Leftarrow$ Efficiency property of MLE.

4 Algebraic Aspects of the Solution

Model	$y = X\beta + \varepsilon$
Estimator	$\widehat{\beta} = (X'X)^{-1}X'y$
Residual	$\widehat{\varepsilon} = y - X\widehat{\beta} = y - X(X'X)^{-1}X'y$
	$= (I_n - X(X'X)^{-1}X')y$
Fitted values	$\widehat{y} = X\widehat{\beta} = X(X'X)^{-1}X'y$

We introduce the following two matrices

$$\boxed{\begin{array}{l} P = X(X'X)^{-1}X' \\ M = I_n - X(X'X)^{-1}X' = I_n - P \end{array}}$$

Properties of the P and M matrices

1) Symmetry $\quad P' = P, M' = M$

2) Idempotent $\quad PP = P, MM = M$

3) Eigenvalues are 0 or 1

4) Rank $(P) = tr(P) = tr(X(X'X)^{-1}X') = tr(X'X(X'X)^{-1}) = tr(I_k) = k$

   Rank $(M) = tr(M) = tr(I - P) = tr(I_n) - tr(P) = n - k$

5) $PX = X(X'X)^{-1}X'X = X$ The P matrix is also called the **Projection matrix**

   $\Rightarrow MX = 0 \quad$ (M is orthogonal to X)

6) $PM = 0$

Therefore, we can write

$$\boxed{\begin{array}{l} \text{Residuals: } \hat{\varepsilon} = My \\ \text{Fitted values: } \hat{y} = Py \\ \text{Residual sum of squares: } RSS = \hat{\varepsilon}'\hat{\varepsilon} = \varepsilon' M \varepsilon \end{array}}$$

$$\begin{aligned} RSS = \hat{\varepsilon}'\hat{\varepsilon} &= y'M\underbrace{My}_{\hat{\varepsilon}} \text{ (M symmetric)} \\ &= (X\beta + \varepsilon)'MM(X\beta + \varepsilon) \text{ (plug in } y = X\beta + \varepsilon) \\ &= \varepsilon' MM\varepsilon \qquad \text{(since } MX = 0,\ X'M = 0) \\ &= \varepsilon' M\varepsilon \qquad \text{(since M is idempotent)} \end{aligned}$$

5 Finite Sample Properties of s^2

Similarly as in the simple linear regression, we can propose the following estimator of σ^2

$$s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$$

Using only A.1 through A.4, we can establish the following exact, finite sample properties:

5.1 Unbiasedness of s^2

Proof that $E(s^2) = \sigma^2$,

$$\begin{aligned}
E(\widehat{\varepsilon}'\widehat{\varepsilon}) &= E(\varepsilon' M\varepsilon) \\
&= E(tr(\varepsilon' M\varepsilon)) && \text{trace of a scalar is the scalar itself} \\
&= E(tr(M\varepsilon\varepsilon')) \\
&= tr(ME(\varepsilon\varepsilon')) && X \text{ non-stochastic, } tr \text{ and } E \text{ are linear operators} \\
&= tr(M \cdot \sigma^2 I) \\
&= \sigma^2 tr(M) = \sigma^2(n-k)
\end{aligned}$$

s is the estimated **standard error of the regression**. $\widehat{\sigma}^2_{MLE} = \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{n}$ biased, but more efficient.

The variance of $\widehat{\beta}$ can thus be estimated by

$$\widehat{\text{Var}}(\widehat{\beta}) = s^2(X'X)^{-1}.$$

The estimated variance of an element $\widehat{\beta}_j$ is the j^{th} diagonal element of this estimated variance-covariance matrix. The square root of this estimated variance is usually referred to as the estimated **standard error of** $\widehat{\beta}_j$, $(\widehat{SE}(\widehat{\beta}_j))$.

5.2 Exact Sampling Distribution of s^2

It can be shown that the unbiased estimator s^2 (rescaled) has a χ^2 distribution with $n-k$ degrees of freedom

$$(n-k)\sigma^{-2}s^2 \sim \chi^2_{n-k}$$

Technical Aside: Quadratic Form Distributions: $Q = x'Ax$

$x \sim N(0, I_p)$

$$Q \sim \chi^2(k) \qquad k = \text{rank}(A) \Leftrightarrow A \text{idempotent}$$

$x \sim N(0, \sigma^2 I_p)$

$$\frac{Q}{\sigma^2} \sim \chi^2(k) \qquad k = \text{rank}(A) \Leftrightarrow A \text{idempotent}$$

$x \sim N(0, V)$

$$Q \sim \chi^2(k) \qquad k = \text{rank}(AV) \Leftrightarrow AV \text{idempotent}$$

Proof:

$$\begin{aligned}(n-k)\sigma^{-2}s^2 &= (n-k)\sigma^{-2}\cdot\frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{n-k}\\ &= \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{\sigma^2} = \frac{\varepsilon' M\varepsilon}{\sigma^2}\end{aligned}$$

This is a quadratic form of a $N(0,\sigma^2 I_n)$ vector, M idempotent with rank equal to $n-k \Rightarrow$

$$\frac{\varepsilon' M\varepsilon}{\sigma^2} \sim \chi^2_{n-k}$$

(See Problem Set 2, last question)

5.3 Zero-covariance between $\widehat{\beta}$ and $\widehat{\varepsilon}$

For testing purposes it is useful to note that $\widehat{\beta}$ and $\widehat{\varepsilon}$ have no correlation: under the additional assumption of normality (A.5) it implies that $\widehat{\beta}$ and s^2 independent (causality does not go the other way).
Proof:

$$\begin{aligned}\widehat{\beta} &= \beta + (X'X)^{-1}X'\varepsilon\\ \widehat{\varepsilon} &= M\varepsilon\end{aligned}$$

$$\begin{aligned}Cov(\widehat{\beta},\widehat{\varepsilon}) &= E\left[(\widehat{\beta}-\beta)(\widehat{\varepsilon}-0)'\right]\\ &= E\left((X'X)^{-1}X'\varepsilon(M\varepsilon)'\right)\\ &= E\left((X'X)^{-1}X'\varepsilon\varepsilon' M\right)\\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')M\\ &= \sigma^2(X'X)^{-1}X'M = 0.\end{aligned}$$

Both $\widehat{\beta}$ and $\widehat{\varepsilon}$ are linear in ε and therefore normaly distributed under (A.5). Because of this normality, the zero covariance implies that all the elements of the vector $\widehat{\beta}$ are independent of the elements of the vector $\widehat{\varepsilon} = M\varepsilon$. This implies that the elements of $\widehat{\beta}$ are independent of any function of the elements of $M\varepsilon$, for example of $\varepsilon' M' M\varepsilon = \varepsilon' M\varepsilon$. Thus, $\widehat{\beta}$ and s^2 are independent under the normality assumption.

6 Goodness of Fit

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{y'M_0 y} \qquad \text{where } M_0 = I_n - \iota(\iota'\iota)^{-1}\iota'$$

$$\text{where } \iota = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1}, \text{ or } M_0 = I_n - \frac{1}{n}\iota\iota' \text{ (See Problem Set 1, question 7)}$$

If the regression does not include an intercept we use instead:

$$R^2 = 1 - \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{y'y}.$$

Problem: R^2 can be made arbitrarily large (R^2 does not decline with k, see Problem Set)

$$\Rightarrow \text{ Adjusted } R^2 = \overline{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$$

7 Partitioning the Linear Regression Model

Say we partition the matrix (X) with the k explanatory variables into X_1 ($n \times k_1$) and X_2 ($n \times k_2$), $X = [X_1 : X_2]$, where $k_1 + k_2 = k$.

$$\begin{aligned} y &= X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon \\ &= \begin{bmatrix} \underset{(n\times k_1)}{X_1} \vdots \underset{(n\times k_2)}{X_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \begin{matrix} (k_1\times 1) \\ (k_2 \times 1) \end{matrix} + \varepsilon \end{aligned}$$

Then, from the FOC

$$X'X\widehat{\beta} = X'y \Rightarrow \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

or

$$\begin{aligned} X_1'X_1\widehat{\beta}_1 + X_1'X_2\widehat{\beta}_2 &= X_1'y \qquad (1) \\ X_2'X_1\widehat{\beta}_1 + X_2'X_2\widehat{\beta}_2 &= X_2'y \qquad (2) \end{aligned}$$

Solve (2) for $\widehat{\beta}_2$:

$$\widehat{\beta}_2 = (X_2'X_2)^{-1}X_2'y - (X_2'X_2)^{-1}X_2'X_1\widehat{\beta}_1$$

Substitute out in (1)

$$(X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1)\widehat{\beta}_1 = X_1'y - X_1'X_2(X_2'X_2)^{-1}X_2'y$$

or

$$(X_1'M_2X_1)\widehat{\beta}_1 = X_1'M_2y \qquad \text{where } M_2 = I_n - X_2(X_2'X_2)^{-1}X_2' = I - P_2$$

$$\Rightarrow \boxed{\widehat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2y}$$

subvector of $\widehat{\beta} = (X'X)^{-1}X'y$

Similarly, we can derive

$$\boxed{\widehat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1y}$$

Compare P_2 with $P = X(X'X)^{-1}X'$

- Recall $Py = \widehat{y}$, or Py gives the fitted values from a regression of y on X. We called P the projection matrix. The fitted value $\widehat{y}$ is the projection of y on the X space.
- Then P_2X_1 gives the projection of X_1 on the X_2 space, or the fitted values of a regression of X_1 on X_2.

Compare M_2 with $M = I_n - X(X'X)^{-1}X'$

- Recall $My = \widehat{\varepsilon}$, or My gives the residual vector from a regression of y on X.
- Then M_2X_1 (matrix $n \times k_1$) contains the residuals from the regression of X_1 on X_2

<u>Example of the partitioned regression result:</u>
In the simple linear regression model,

$$y = \beta_1\iota + \beta_2 x + \varepsilon, \text{ so } X_1 \equiv \iota = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1}, X_2 \equiv x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

or

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

Thus

$$\widehat{\beta}_2 = (x'M_1x)^{-1}x'M_1y = \left((M_1x)'(M_1x)\right)^{-1}(M_1x)'(M_1y)$$

where

$$M_1 = I_n - \iota(\iota'\iota)^{-1}\iota' = I_n - \frac{1}{n}\iota\iota'.$$

Recall from Problem Set 1 that for any $n \times 1$ vector z

$$M_1z = z - \overline{z}\iota = \begin{pmatrix} z_1 - \overline{z} \\ z_2 - \overline{z} \\ \vdots \\ z_n - \overline{z} \end{pmatrix}, \text{ where } \overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i.$$

Thus

$$\begin{aligned} \widehat{\beta}_1 &= ((x - \overline{x}\iota)'(x - \overline{x}\iota))^{-1}(x - \overline{x}\iota)'(y - \overline{y}\iota) \\ &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \text{ as expected} \end{aligned}$$

Finally, let's consider some properties of $\widehat{\beta}_1$

$$\begin{aligned} \widehat{\beta}_1 &= (X_1'M_2X_1)^{-1}X_1'M_2y \\ &= (X_1'M_2X_1)^{-1}X_1'M_2(X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= (X_1'M_2X_1)^{-1}X_1'M_2X_1\beta_1 + (X_1'M_2X_1)^{-1}X_1'M_2X_2\beta_2 \\ &+ (X_1'M_2X_1)^{-1}X_1'M_2\varepsilon \\ &= \beta_1 + (X_1'M_2X_1)^{-1}X_1'M_2\varepsilon \end{aligned}$$

as $M_2X_2 = 0$. Thus

$$\begin{aligned} E(\widehat{\beta}_1) &= E(\beta_1 + (X_1'M_2X_1)^{-1}X_1'M_2\varepsilon) = \beta_1 \\ Var(\widehat{\beta}_1) &= E\left((\widehat{\beta}_1 - \beta_1)(\widehat{\beta}_1 - \beta_1)'\right) \\ &= E\left((X_1'M_2X_1)^{-1}X_1'M_2\varepsilon\varepsilon'M_2X_1(X_1'M_2X_1)^{-1}\right) \\ &= (X_1'M_2X_1)^{-1}X_1'M_2E(\varepsilon\varepsilon')M_2X_1(X_1'M_2X_1)^{-1}, \; X \text{ nonstochastic} \\ &= \sigma^2(X_1'M_2X_1)^{-1}. \end{aligned}$$