Linear Regression

1 Single Explanatory Variable

Assume

$$\begin{aligned} y &\sim F, \\ Ey &= \alpha + \beta x \\ Vary &= \sigma^2 \end{aligned}$$

(y is not necessarily normal)

$$\Rightarrow y = \alpha + \beta x + u$$

where

$$\begin{aligned} Eu &= 0 \\ Varu &= \sigma^2 \\ Exu &= 0. \end{aligned}$$

Examples:

1. School performance as a function of SAT score
2. Consumption as a function of income
3. Murder rate as a function of crime policy

1.1 Curve fitting

1. Draw pictures
2. Options for fitting
   (a) Piecewise
   (b) Mean squared deviation
   (c) Mean absolute deviation
   (d) Mean orthogonal squared deviation

1.2 Estimation

1.2.1 Types of samples

Cross-section data:

$$y_i = \alpha + \beta x_i + u_i \quad i = 1, 2, .., n \tag{1}$$

Time Series data:

$$y_t = \alpha + \beta x_t + u_t \quad t = 1, 2, .., T$$

1.2.2 Mechanics

Consider the solution to

$$\left(\widehat{\alpha},\widehat{\beta}\right)=\arg\min_{\alpha,\beta}\sum_{i=1}^{n}\left[y_i-\alpha-\beta x_i\right]^2$$

The FOC's are

$$\begin{aligned} 0 &= -2\sum_{i=1}^{n}\left[y_i-\alpha-\beta x_i\right] \\ &\Rightarrow \widehat{\alpha}=\overline{y}-\widehat{\beta}\overline{x} \end{aligned}$$

and

$$\begin{aligned} 0 &= -2\sum_{i=1}^{n}\left[y_i-\alpha-\widehat{\beta}x_i\right]x_i \\ &\Rightarrow 0=\sum_{i=1}^{n}\left[\widetilde{y}_i-\widehat{\beta}\widetilde{x}_i\right]x_i \\ &\Rightarrow 0=\sum_{i=1}^{n}\left[\widetilde{y}_i-\widehat{\beta}\widetilde{x}_i\right](\widetilde{x}_i+\overline{x}) \\ &\Rightarrow 0=\sum_{i=1}^{n}\left[\widetilde{y}_i-\widehat{\beta}\widetilde{x}_i\right]\widetilde{x}_i+\sum_{i=1}^{n}\left[\widetilde{y}_i-\widehat{\beta}\widetilde{x}_i\right]\overline{x} \\ &\Rightarrow 0=\sum_{i=1}^{n}\left[\widetilde{y}_i-\widehat{\beta}\widetilde{x}_i\right]\widetilde{x}_i+\overline{x}\sum_{i=1}^{n}\widetilde{y}_i-\widehat{\beta}\overline{x}\sum_{i=1}^{n}\widetilde{x}_i \\ &\Rightarrow 0=\sum_{i=1}^{n}\left[\widetilde{y}_i-\widehat{\beta}\widetilde{x}_i\right]\widetilde{x}_i \end{aligned}$$

Note that, for any variable,

$$\sum_{i=1}^{n}\widetilde{z}_i=\sum_{i=1}^{n}(z_i-\overline{z})=\sum_{i=1}^{n}z_i-n\overline{z}=n\overline{z}-n\overline{z}=0.$$

Solving for $\widehat{\beta}$, one gets

$$\widehat{\beta}=\frac{\sum_{i=1}^{n}\widetilde{x}_i\widetilde{y}_i}{\sum_{i=1}^{n}\widetilde{x}_i^2}=\frac{\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i\widetilde{y}_i}{\frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_i^2}=\frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_{xx}}.$$

Note that one can take equation (1) and write it as

$$\widetilde{y}_i=\beta\widetilde{x}_i+\widetilde{u}_i,$$

set

$$\widehat{\beta}=\arg\min_{\beta}\sum_{i=1}^{n}\left[\widetilde{y}_i-\beta\widetilde{x}_i\right]^2,$$

and get the same $\widehat{\beta}$.

1.2.3 Properties of $\widehat{\beta}$

$$
\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^{n} \widetilde{x}_i \widetilde{y}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2} = \frac{\sum_{i=1}^{n} \widetilde{x}_i \left(\beta \widetilde{x}_i + \widetilde{u}_i\right)}{\sum_{i=1}^{n} \widetilde{x}_i^2} \\
&= \beta \frac{\sum_{i=1}^{n} \widetilde{x}_i^2}{\sum_{i=1}^{n} \widetilde{x}_i^2} + \frac{\sum_{i=1}^{n} \widetilde{x}_i \widetilde{u}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2} \\
&= \beta + \frac{\sum_{i=1}^{n} \widetilde{x}_i \widetilde{u}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2}.
\end{aligned}
$$

$$
\begin{aligned}
E\widehat{\beta} &= E\beta + E\frac{\sum_{i=1}^{n} \widetilde{x}_i \widetilde{u}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2} \\
&= \beta + \frac{\sum_{i=1}^{n} \widetilde{x}_i E\widetilde{u}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2} = \beta
\end{aligned}
$$

$\Rightarrow \widehat{\beta}$ is an unbiased estimator of β.

$$
\begin{aligned}
Var\widehat{\beta} &= E\left[\widehat{\beta} - E\widehat{\beta}\right]^2 = E\left[\widehat{\beta} - \beta\right]^2 \\
&= E\left[\frac{\sum_{i=1}^{n} \widetilde{x}_i \widetilde{u}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2}\right]^2 = \frac{E\left[\sum_{i=1}^{n} \widetilde{x}_i \widetilde{u}_i\right]^2}{\left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^2} \\
&= \frac{E\left[\sum_{i=1}^{n} \sum_{j=1}^{n} \widetilde{x}_i \widetilde{x}_j \widetilde{u}_i \widetilde{u}_j\right]}{\left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^2} = \frac{\left[\sum_{i=1}^{n} \sum_{j=1}^{n} \widetilde{x}_i \widetilde{x}_j E\widetilde{u}_i \widetilde{u}_j\right]}{\left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^2} \\
&= \frac{\sum_{i=1}^{n} \widetilde{x}_i^2 \sigma^2}{\left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^2} = \sigma^2 \left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^{-1}.
\end{aligned}
$$

Note that:

1. $Var\widehat{\beta}$ declines as n increases;
2. $Var\widehat{\beta}$ declines as variation in x increases;
3. $Var\widehat{\beta}$ increases as σ^2 increases.

So far, the assumptions are:

1. $Eu_i = 0$;
2. $Varu_i = \sigma^2 < \infty$;
3. $Cov\left(u_i, u_j\right) = 0 \quad \forall i \neq j$;
4. $Cov\left(u_i, x_i\right) = 0$.

If we also assume that $u_i \sim iidN\left(0,\sigma^2\right)$, then, since

$$\widehat{\beta} = \beta + \frac{\sum_{i=1}^{n} \widetilde{x}_i \widetilde{u}_i}{\sum_{i=1}^{n} \widetilde{x}_i^2}$$

is a linear combination of normal random variables, $\widehat{\beta}$ is normal

$$\Rightarrow \widehat{\beta} \sim N\left(\beta, \sigma^2 \left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^{-1}\right).$$

Alternatively (instead of assuming $u_i \sim iidN\left(0,\sigma^2\right)$), assume $n \to \infty$. Then, since $\widehat{\beta} - \beta$ is a weighted average of u_i, $i = 1, 2, .., n$, where each u_i has $Eu_i = 0$,

$$\begin{aligned}
E\left(\widehat{\beta} - \beta\right) &= 0; \\
Var\left(\widehat{\beta} - \beta\right) &= \sigma^2 \left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^{-1} \to 0; \\
E\sqrt{n}\left(\widehat{\beta} - \beta\right) &= 0; \\
Var\sqrt{n}\left(\widehat{\beta} - \beta\right) &= n\sigma^2 \left[\sum_{i=1}^{n} \widetilde{x}_i^2\right]^{-1} = \sigma^2 \left[\frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i^2\right]^{-1}
\end{aligned}$$

$$\Rightarrow (CLT) \quad \sqrt{n}\left(\widehat{\beta} - \beta\right) \sim N\left(0, \sigma^2 \left[\frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i^2\right]^{-1}\right).$$

In either case, since $\sqrt{n}\left(\widehat{\beta} - \beta\right)$ has a known distribution, one can construct hypothesis tests involving β. Consider

$$H_0 : \beta = \beta_0 \text{ vs. } H_A : \beta \neq \beta_0.$$

Under the null hypothesis,

$$\frac{\sqrt{n}\left(\widehat{\beta} - \beta_0\right)}{\sigma_\beta} \sim N(0,1)$$

where

$$\sigma_\beta = \sigma \Big/ \sqrt{\frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i^2},$$

or

$$\frac{\sqrt{n}\left(\widehat{\beta} - \beta_0\right)}{\widehat{\sigma}_\beta} \sim t_{n-2}$$

where

$$\widehat{\sigma}_\beta = s^2 / \sqrt{\frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i^2}$$
$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(\widetilde{y}_i - \widehat{\beta}\widetilde{x}_i\right)^2.$$

Why $n-2$?

We say that β is (statistically) significant if $\widehat{\beta}$ is significantly different than $\beta_0 = 0$.

1.2.4 Properties of Residuals

u_i is the error, and

$$\begin{aligned} \widehat{u}_i &= \widetilde{y}_i - \widehat{\beta}\widetilde{x}_i \\ &= y_i - \widehat{\alpha} - \widehat{\beta}x_i \\ &\neq u_i \end{aligned}$$

is the residual. Note that

$$\sum_{i=1}^{n} \widehat{u}_i = \sum_{i=1}^{n}\left(\widetilde{y}_i - \widehat{\beta}\widetilde{x}_i\right) = \sum_{i=1}^{n}\widetilde{y}_i - \widehat{\beta}\sum_{i=1}^{n}\widetilde{x}_i = 0;$$

$$\begin{aligned} \sum_{i=1}^{n} \widehat{u}_i\widetilde{x}_i &= \sum_{i=1}^{n}\left(\widetilde{y}_i - \widehat{\beta}\widetilde{x}_i\right)\widetilde{x}_i \\ &= \sum_{i=1}^{n}\widetilde{y}_i\widetilde{x}_i - \widehat{\beta}\sum_{i=1}^{n}\widetilde{x}_i^2 \\ &= \sum_{i=1}^{n}\widetilde{y}_i\widetilde{x}_i - \frac{\sum_{i=1}^{n}\widetilde{y}_i\widetilde{x}_i}{\sum_{i=1}^{n}\widetilde{x}_i^2}\sum_{i=1}^{n}\widetilde{x}_i^2 \\ &= \sum_{i=1}^{n}\widetilde{y}_i\widetilde{x}_i - \sum_{i=1}^{n}\widetilde{y}_i\widetilde{x}_i = 0. \end{aligned}$$

1.2.5 Goodness of Fit

$$\sum_{i=1}^{n}\widetilde{y}_i^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}\left[y_i - \widehat{y}_i + \widehat{y}_i - \overline{y}\right]^2$$

(where

$$\widehat{y}_i = \widehat{\alpha} + \widehat{\beta}x_i)$$

$$= \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 + \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 + 2\sum_{i=1}^{n} (y_i - \widehat{y}_i)(\widehat{y}_i - \overline{y}). \tag{2}$$

Note that

$$y_i - \widehat{y}_i = y_i - \widehat{\alpha} - \widehat{\beta} x_i = \widehat{u}_i$$

and

$$\widehat{y}_i - \overline{y} = \left(\widehat{\alpha} - \widehat{\beta} x_i\right) - \left(\widehat{\alpha} - \widehat{\beta}\overline{x}_i\right) = \widehat{\beta}\widetilde{x}_i.$$

Then

$$\sum_{i=1}^{n} (y_i - \widehat{y}_i)(\widehat{y}_i - \overline{y}) = \widehat{\beta} \sum_{i=1}^{n} \widehat{u}_i \widetilde{x}_i = 0.$$

So equation (2) becomes

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 + \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2.$$

Define

$$R^2 = \frac{\sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}.$$

Note that

$$0 \leq R^2 \leq 1.$$

When is $R^2 = 1$? When is $R^2 = 0$?

Now let

$$\begin{aligned} ESS &= \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2; \\ RSS &= \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2. \end{aligned}$$

Consider

$$H_0 : \beta = 0 \text{ vs. } H_A : \beta \neq 0.$$

It can be shown that

$$\left.\begin{array}{l} ESS/\sigma^2 \sim \chi^2_{n-2} \\ RSS/\sigma^2 \sim \chi^2_1 \end{array}\right\} \text{ independent}$$

$$\Rightarrow \frac{(n-2)\, RSS}{ESS} \sim F_{1,n-2}.$$

Note that

$$\begin{aligned} R^2 &= RSS/\left(RSS+ESS\right) \\ 1-R^2 &= ESS/\left(RSS+ESS\right) \end{aligned}$$

$$\Rightarrow \frac{(n-2)\, R^2}{1-R^2} = \frac{(n-2)\, RSS}{ESS} \sim F_{1,n-2}.$$

2 Multiple Regression

Now consider the more complex model

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{ti} + u_t. \tag{3}$$

Examples:

1. Consumption on interest rate, income, wealth, unemployment rate
2. Earnings on schooling, age, race, sex
3. Hours of work on children, sex, spouse's income, wage

2.1 Structure

We can write equation (3) as

$$y_t = \sum_{i=0}^{n} \beta_i x_{ti} + u_t$$

where $x_{t0} = 1$.

Assumptions:

1. Linear model with correct specification
2. x's are nonstochastic and linearly independent
3. $Eu_t = 0$
4.

$$Eu_t u_s = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

5. (maybe)

$$u_t \sim iidN\left(0, \sigma^2\right)$$

$$\Rightarrow y_t \sim indN\left(\sum_{i=0}^{n} \beta_i x_{ti}, \sigma^2\right).$$

Note that

$$\beta_i = \frac{\partial E y_t}{\partial x_{ti}}.$$

Discuss the significance of *partial* derivative.

2.2 Intro to Ordinary Least Squares

Consider minimizing the sum of squared residuals:

$$\min_{\beta_0, \beta_1, .., \beta_n} \sum_{t=1}^{T} \left[y_t - \sum_{i=0}^{n} \beta_i x_{ti} \right]^2 .$$

The FOC's are

$$\begin{aligned} 0 &= -2 \sum_{t=1}^{T} \left[y_t - \sum_{i=0}^{n} \beta_i x_{ti} \right] x_{ti} \qquad (4) \\ &\Rightarrow 0 = \sum_{t=1}^{T} \left[y_t - \sum_{i=0}^{n} \beta_i x_{ti} \right] x_{ti}. \end{aligned}$$

2.3 Matrix Notation

Let

$$\begin{aligned} \underset{T \times 1}{y} &= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}; \quad \underset{T \times (n+1)}{X} = \begin{pmatrix} x_{10} & x_{11} & \cdots & x_{1n} \\ x_{20} & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T0} & x_{T1} & \cdots & x_{Tn} \end{pmatrix}; \\ \underset{(n+1) \times 1}{\beta} &= \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}; \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}. \end{aligned}$$

Then, we can write equation (3) as

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

and we can write the FOC's in equation (4) as

$$X'\left(y - X\widehat{\beta}\right) = 0.$$

We can solve for $\widehat{\beta}$ to get

$$\widehat{\beta} = \left(X'X\right)^{-1}\left(X'y\right).$$

An alternative derivation is to let

$$\widehat{u} = y - X\widehat{\beta}$$

be the residuals and find $\widehat{\beta}$ so that $X'\widehat{u} = 0$ (this will generalize later to instrumental variables estimation):

$$\begin{aligned} X'\widehat{u} &= X'\left(y - X\widehat{\beta}\right) = 0 \\ &\Rightarrow \widehat{\beta} = \left(X'X\right)^{-1}\left(X'y\right). \end{aligned}$$

2.4 Statistical Properties of $\widehat{\beta}$

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

$$E\widehat{\beta} = E\beta + E\left(X'X\right)^{-1}X'u = \beta + \left(X'X\right)^{-1}X'Eu = \beta.$$

$$\begin{aligned} D\left(\widehat{\beta}\right) &= E\left(\widehat{\beta} - E\widehat{\beta}\right)\left(\widehat{\beta} - E\widehat{\beta}\right)' \\ &= E\left(\widehat{\beta} - \beta\right)\left(\widehat{\beta} - \beta\right)' \\ &= E\left(X'X\right)^{-1}X'uu'X\left(X'X\right)^{-1} \\ &= \left(X'X\right)^{-1}X'Euu'X\left(X'X\right)^{-1} \\ &= \left(X'X\right)^{-1}X'\sigma^2 IX\left(X'X\right)^{-1} \\ &= \sigma^2\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1} \\ &= \sigma^2\left(X'X\right)^{-1}. \end{aligned}$$

$\Rightarrow$

$$\begin{aligned} Var\widehat{\beta}_i &= \sigma^2\left(X'X\right)^{-1}_{ii}; \\ Cov\left(\widehat{\beta}_i, \widehat{\beta}_j\right) &= \sigma^2\left(X'X\right)^{-1}_{ij}. \end{aligned}$$

If $u \sim N\left(0, \sigma^2 I\right)$, then

$$\begin{aligned}
\widehat{\beta} &\sim N\left[\beta, \sigma^2 \left(X'X\right)^{-1}\right] \\
&\Rightarrow \left(\widehat{\beta}_i - \beta_i\right) / \sigma \sqrt{\left(X'X\right)_{ii}^{-1}} \sim N\left(0, 1\right).
\end{aligned}$$

Next, let

$$s^2 = \frac{\widehat{u}'\widehat{u}}{T - (n+1)}.$$

Later, we show that $Es^2 = \sigma^2$. Also, it is true that

$$\left[T - (n+1)\right] s^2 / \sigma^2 \sim \chi^2_{T-(n+1)}.$$

$\Rightarrow$

$$\frac{\left(\widehat{\beta}_i - \beta_i\right) / \sigma \sqrt{\left(X'X\right)_{ii}^{-1}}}{\sqrt{\left[T - (n+1)\right] s^2 / \sigma^2 \left[T - (n+1)\right]}} \sim t_{T-(n+1)}$$

which simplifies to

$$\frac{\left(\widehat{\beta}_i - \beta_i\right) / \sqrt{\left(X'X\right)_{ii}^{-1}}}{s} \sim t_{T-(n+1)}.$$

2.5 Asymptotics

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

$$\begin{aligned}
plim\ \widehat{\beta} &= plim\ \beta + plim\ \left(X'X\right)^{-1} X'u \\
&= \beta + plim\ \left(X'X\right)^{-1} X'u.
\end{aligned}$$

We would like to say that

$$plim\ \left(X'X\right)^{-1} X'u = \left(plim\ X'X\right)^{-1} plim\ X'u.$$

But $X'X$ explodes as $T \to \infty$, and $X'u$ doesn't converge. To see this note that

$$X'X = \begin{pmatrix}
T & \sum_t X_{t1} & \cdots & \sum_t X_{tn} \\
\sum_t X_{t1} & \sum_t X_{t1}^2 & \cdots & \sum_t X_{t1} X_{tn} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_t X_{tn} & \sum_t X_{t1} X_{tn} & \cdots & \sum_t X_{tn}^2
\end{pmatrix}$$

and

$$X'u = \begin{pmatrix} \sum_t u_t \\ \sum_t X_{t1} u_t \\ \vdots \\ \sum_t X_{tn} u_t \end{pmatrix}$$

which is random. However, if we rewrite

$$\left(X'X\right)^{-1} X'u = \left(\frac{X'X}{T}\right)^{-1} \frac{X'u}{T},$$

then, under some conditions (discuss), $\frac{X'X}{T}$ will converge to a positive definite finite matrix and $\frac{X'u}{T}$ will converge to 0 (by a Law of Large Numbers). Thus

$$\begin{aligned} plim\ \widehat{\beta} &= \beta + plim\ \left(\frac{X'X}{T}\right)^{-1} \frac{X'u}{T} \\ &= \beta + \left(plim\ \frac{X'X}{T}\right)^{-1} plim\ \frac{X'u}{T} \\ &= \beta + \left(plim\ \frac{X'X}{T}\right)^{-1} 0 = \beta. \end{aligned}$$

$$\begin{aligned} &\Rightarrow \quad plim\left(\widehat{\beta} - \beta\right) = 0 \\ &\Rightarrow \quad \sqrt{T}\left(\widehat{\beta} - \beta\right) \sim N\left[0, \sigma^2 \left(plim\ \frac{X'X}{T}\right)^{-1}\right] \end{aligned}$$

by a Central Limit Theorem. What would have happened if we had tried to derive the asymptotic distribution of $\sqrt{T}\widehat{\beta}$? What would have happened if we had tried to derive the asymptotic distribution of $\widehat{\beta} - \beta$?

2.6 Residual Analysis

$$\begin{aligned} \widehat{u} &= y - X\widehat{\beta} \\ &= y - X\left(X'X\right)^{-1} X'y \\ &= \left[I - X\left(X'X\right)^{-1} X'\right] y. \end{aligned}$$

Consider

$$P_X = X\left(X'X\right)^{-1} X'.$$

Note that

$$\begin{aligned} P_X X &= X\left(X'X\right)^{-1} X'X = X, \\ P_X Xb &= X\left(X'X\right)^{-1} X'Xb = Xb, \end{aligned}$$

and the same properties hold for $(Xb)' P_X$. P_X is called the projection matrix for X because it projects any vector into the space spanned by X.

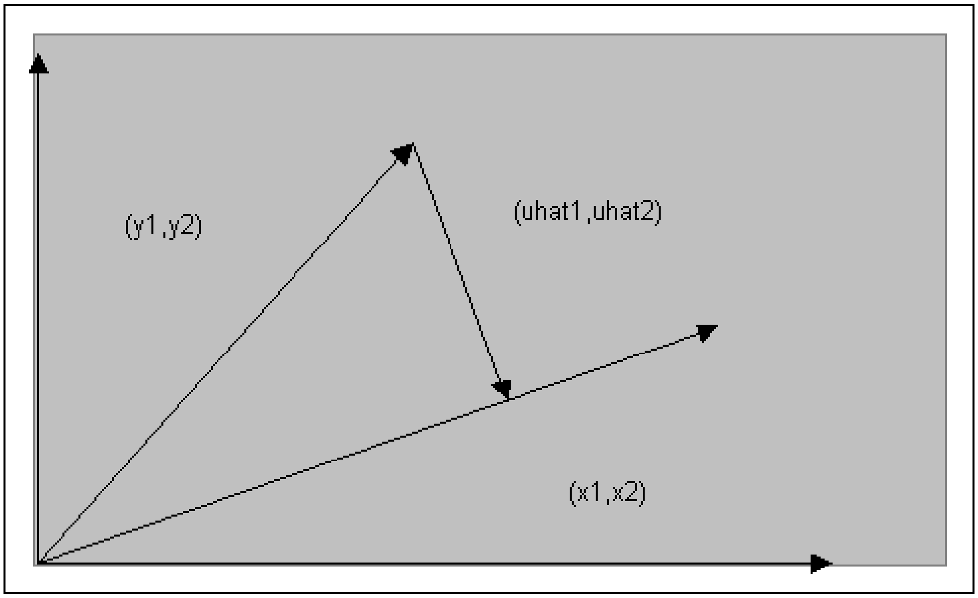

Figure 1:

Explain with a picture:

Show unbiasedness.

Explain with three dimensions.

Note that

$$P_X^2 = X\left(X'X\right)^{-1} X'X\left(X'X\right)^{-1} X' = X\left(X'X\right)^{-1} X' = P_X$$

$\Rightarrow P_X$ is idempotent. Consider $I - P_X$:

$$\left(I - P_X\right)\left(I - P_X\right) = I - P_X - P_X + P_X P_X = I - P_X$$

$\Rightarrow I - P_X$ is idempotent. Explain $I - P_X$ geometrically. Note that

$$\left(I - P_X\right) Xb = Xb - P_X Xb = Xb - Xb = 0.$$

Note that both P_X and $\left(I - P_X\right)$ are symmetric.

Back to residuals:

$$\widehat{u} = \left(I - P_X\right) y = \left(I - P_X\right)\left(X\beta + u\right) = \left(I - P_X\right) u.$$

Consider

$$b'X'\widehat{u} = b'X'\left(I - P_X\right) u = 0 \quad \forall b. \tag{5}$$

Let $\underset{(n+1)\times 1}{e_i}$ be a vector with 1 in the ith element and 0 everywhere else. Note that $Xe_i = X_i$ (the ith column of X).

$$\Rightarrow X_i'\widehat{u} = 0 \ \ \forall i$$

(set $b = e_i$ in equation (5)). When $i = 0$,

$$X_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

and

$$X_0'\widehat{u} = \sum_{t=1}^{T} \widehat{u}_t = 0.$$

Consider

$$\widehat{u}'\widehat{u} = \sum_{t=1}^{T} \widehat{u}_t^2 = [T - (n+1)]\, s^2.$$

$$\begin{aligned}
E\widehat{u}'\widehat{u} &= Eu'(I - P_X)'(I - P_X)u \\
&= Eu'(I - P_X)(I - P_X)u \\
&= Eu'(I - P_X)u \\
&= trEu'(I - P_X)u \\
&= Etru'(I - P_X)u \\
&= Etr(I - P_X)uu' \\
&= tr(I - P_X)Euu' \\
&= tr(I - P_X)\sigma^2 I \\
&= \sigma^2 tr(I - P_X) \\
&= \sigma^2 trI_T - trP_X \\
&= T\sigma^2 - \sigma^2 trX(X'X)^{-1}X' \\
&= T\sigma^2 - \sigma^2 tr(X'X)^{-1}X'X \\
&= T\sigma^2 - \sigma^2 trI_{n+1} \\
&= \sigma^2[T - (n+1)]
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad & Es^2 = E[T - (n+1)]^{-1}\widehat{u}'\widehat{u} \\
= \quad & [T - (n+1)]^{-1}E\widehat{u}'\widehat{u} \\
= \quad & \sigma^2.
\end{aligned}$$

2.7 More on Projection

Consider the model,

$$y = \underset{T\times k}{X}\beta + \underset{T\times m}{Z}\gamma + u,$$

and consider three different estimators for β:

1. Let

   $$X^* = (X \mid Z), \quad \beta^* = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$$

   $\Rightarrow$ we can write the model as

   $$y = X^*\beta^* + u.$$

   Define

   $$\widehat{\beta}^* = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}y$$

   and $\widehat{\beta}$ as the first k elements of $\widehat{\beta}^*$.

2. Let

   $$\begin{aligned} \widetilde{X} &= X - Z\widehat{\delta} \\ &= X - Z\left(Z'Z\right)^{-1} Z'X \\ &= \left[I - P_Z\right] X, \end{aligned}$$

   and define

   $$\widehat{\beta} = \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}'y.$$

3. Let

   $$\widetilde{y} = y - P_Z y = \left[I - P_Z\right] y,$$

   and define

   $$\widehat{\beta} = \left(\widetilde{X}'\widetilde{X}\right)^{-1} \widetilde{X}'\widetilde{y}.$$

All three of these estimators provide the exact same estimate of β. Why? Generalize for

$$y = X\beta + Z\gamma + Q\theta + u.$$

2.8 Significance

We might want to test the significance of the whole equation. Before we used R^2 and F. We can use the same statistics again. Let

$$\underbrace{\sum_{t=1}^{T}(y_t-\overline{y})^2}_{TSS}=\underbrace{\sum_{t=1}^{T}(y_t-\widehat{y}_t)^2}_{ESS}+\underbrace{\sum_{t=1}^{T}(\widehat{y}_t-\overline{y})^2}_{RSS}$$

where

$$\widehat{y}_t=X_t\widehat{\beta}.$$

Define

$$R^2=\frac{RSS}{TSS}=1-\frac{ESS}{TSS}.$$

Problems with R^2:

1. It's properties depend on having the correct specification of the model (implications for comparing R^2 across nonnested models).

2. It's properties depend on the number of X's (n). In particular, adding an extra column to X can *not* decrease R^2, and, if

$$X'_{n+2}\left(I-P_X\right)y\neq 0,$$

   then R^2 increases. $\Rightarrow T$ linearly independent X's will always produce $R^2=1$.

3. R^2 is only one measure of fit, and it is not a particularly good one. It corresponds to

$$H_0:\beta_1=\beta_2=..=\beta_n=0.$$

   What is the distribution of R^2?

Adjustments:

$$\overline{R}^2=1-\frac{\widehat{Var}\widehat{u}_t}{\widehat{Var}y_t}$$

where

$$\begin{aligned}\widehat{Var}\widehat{u}_t &= s^2;\\ \widehat{Var}y_t &= \frac{1}{T-1}\sum_{t=1}^{T}(y_t-\overline{y})^2.\end{aligned}$$

Note that

$$
\begin{aligned}
R^2 &= 1-\frac{s^2\left[T-(n+1)\right]}{\widehat{Vary_t}\left[T-1\right]} \\
&\Rightarrow\ 1-R^2=\frac{s^2\left[T-(n+1)\right]}{\widehat{Vary_t}\left[T-1\right]} \\
&\Rightarrow\ \overline{R}^2=1-\left(1-R^2\right)\frac{T-1}{T-(n+1)}.
\end{aligned}
$$

Note that $\overline{R}^2$ can be negative. It can be shown that, if the $|t-\text{statistic}|$ associated with a variable is greater than one, then the $\overline{R}^2$ will increase when that variable is included.

Consider the test

$$
\begin{aligned}
H_0 &: \beta_1=\beta_2=..=\beta_n=0 \\
H_A &: \beta_1\neq\beta_2\neq..\neq\beta_n\neq 0.
\end{aligned}
$$

Under the null hypothesis,

$$
\text{independent}\left\{\begin{array}{l}
ESS/\sigma^2=s^2\left[T-(n+1)\right]/\sigma^2\sim\chi^2_{T-(n+1)} \\
RSS/\sigma^2=\widehat{\beta}^{*\prime}X^{*\prime}X^*\widehat{\beta}^*/\sigma^2\sim\chi^2_n
\end{array}\right.
$$

where X^* is X without the first column and $\widehat{\beta}^*$ is the corresponding elements of $\widehat{\beta}$. $\Rightarrow$

$$
\begin{aligned}
\frac{RSS/n}{ESS/\left[T-(n+1)\right]} &\sim F_{n,T-(n+1)} \\
&= \frac{R^2/n}{\left(1-R^2\right)/\left[T-(n+1)\right]}.
\end{aligned}
$$

2.9 Multicollinearity

If two or more X_i's are linearly independent, we say they are colinear.

Example:

$$Ea=\beta_0+\beta_1 Sc+\beta_2 Ex+\beta_3 Ag+u$$

where

$$Ag=Sc+Ex+6.$$

In this example, $X'X$ is not invertible; it is singular. Where is the singularity? Assume you knew β. Consider the alternative model

$$Ea=\gamma_0+\gamma_1 Sc+\gamma_2 Ex+\gamma_3 Ag+u$$

where

$$\begin{aligned}
\gamma_0 &= \beta_0 + 6\beta_3; \\
\gamma_1 &= \beta_1 + \beta_3; \\
\gamma_2 &= \beta_2 + \beta_3; \\
\gamma_3 &= 0.
\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}
Ea &= \beta_0 + 6\beta_3 + (\beta_1 + \beta_3) Sc + (\beta_2 + \beta_3) Ex + u \\
&= \beta_0 + \beta_1 Sc + \beta_2 Ex + \beta_3 (6 + Sc + Ex) + u \\
&= \beta_0 + \beta_1 Sc + \beta_2 Ex + \beta_3 Ag + u
\end{aligned}$$

which is the same as the original model but with different implications.

In practice, even if there is not perfect colinearity, but near perfect colinearity (as measured by the correlation matrix of X or the eigenvalues of a standardized $X'X$), there will be problems: $D\left(\widehat{\beta}\right)$ will explode as $X'X$ approaches singularity.

What can we do? Drop variables, principal components, cry.

2.10 Linearity

Discuss which equations can be estimated with OLS:

1. $$y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + u$$

2. $$\log y = \alpha_0 + \alpha_1 \log X_1 + \alpha_2 \log X_2 + u$$

3. $$y = \gamma_0 X_1^{\gamma_1} X_2^{\gamma_2} u$$

4. $$y = \gamma_0 X_1^{\gamma_1} X_2^{\gamma_2} + u$$

2.11 Dummy Variables

Example: Wages

$$\log W_t = \beta_0 + \beta_1 Educ_t + \beta_2 Age_t + \beta_3 Sex_t + \beta_4 Race_t + \beta_5 Urban_t + \beta_6 South_t + u_t.$$

Explain each dummy variable.

1. What if you change definition?
2. With race, what if there are more than categories?
3. What if the constant is deleted?
4. How can you interact dummies?

Consider the equation below (from Byrne, Goeree, Hiedemann, and Stern 2001) and interpret each variable:

$$\begin{aligned}
\ln w_k &= \underset{(0.072)}{0.028} + \left[\underset{(0.006)}{0.035} * Educ_k * 1\left(Educ_k < 12\right)\right] \\
&+ \left[\underset{(0.052)}{0.540} * 1\left(Educ_k = 12\right)\right] + \left[\underset{(0.053)}{0.680} * 1\left(13 \leq Educ_k \leq 15\right)\right] \\
&+ \left[\underset{(0.053)}{0.978} * 1\left(Educ_k = 16\right)\right] + \left[\underset{(0.054)}{1.086} * 1\left(Educ_k = 17\right)\right] \\
&+ \left[\underset{(0.002)}{0.066} * Age_k\right] - \left[\underset{(0.00003)}{0.00069} * Age_k * Age_k\right] + \left[\underset{(0.030)}{0.099} * Male_k\right] \\
&+ \left[\underset{(0.029)}{0.028} * Marry_k\right] + \left[\underset{(0.022)}{0.066} * White_k\right] + \left[\underset{(0.042)}{0.090} * Male_k * Marry_k\right] \\
&+ \left[\underset{(0.033)}{0.022} * Male_k * White_k\right] - \left[\underset{(0.032)}{0.035} * Marry_k * White_k\right] \\
&+ \left[\underset{(0.045)}{0.093} * Male_k * Marry_k * White_k\right] + e_k
\end{aligned} \tag{6}$$

$$R^2 = 0.34$$

2.12 Omitted Variables

Consider the model

$$y = X\beta + Z\gamma + u.$$

Instead we estimate

$$y = Xb + e$$

with

$$\begin{aligned}
\widehat{b} &= \left(X'X\right)^{-1} X'y \\
&= \left(X'X\right)^{-1} X'\left(X\beta + Z\gamma + u\right) \\
&= \beta + \left(X'X\right)^{-1} X'Z\gamma + \left(X'X\right)^{-1} X'u,
\end{aligned}$$

and

$$E\widehat{b} = \beta + \left(X'X\right)^{-1} X'Z\gamma$$

which is equal to β iff $\left(X'X\right)^{-1} X'Z\gamma = 0$. This happens if $X'Z = 0$ or $\gamma = 0$; otherwise our estimator is biased.

2.13 Measurement Error

Consider the model

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

Instead of observing y, we observe

$$w = y + e$$

where

$$EX'e = 0.$$

Consider the properties of the OLS estimator

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

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

$$\Rightarrow E\widehat{\beta} = E(X'X)^{-1} X'y + E(X'X)^{-1} X'e = \beta.$$

Now, instead, assume we observe y but we observe

$$Z = X + e$$

instead of X with $Ee = 0$, $Ee'u = 0$, and $plim\ \frac{X'e}{T} = 0$. So we estimate b in

$$y = Zb + \varepsilon.$$

$$\begin{aligned} \widehat{b} &= (Z'Z)^{-1} Z'y \\ &= \left[(X+e)'(X+e)\right]^{-1} (X+e)'(X\beta+u). \end{aligned}$$

The $E\widehat{b}$ may not even exist. Consider

$$\begin{aligned} plim\widehat{b} &= plim\left[(X+e)'(X+e)\right]^{-1}(X+e)'(X\beta+u) \qquad (7) \\ &= plim\left[\frac{(X+e)'(X+e)}{T}\right]^{-1}\frac{(X+e)'(X\beta+u)}{T} \\ &= \left[plim\frac{X\prime X}{T} + plim\frac{X'e}{T} + plim\frac{e'X}{T} + plim\frac{e'e}{T}\right]^{-1} \cdot \\ &\quad \left[plim\frac{X'X\beta}{T} + plim\frac{X'u}{T} + plim\frac{e'X\beta}{T} + plim\frac{e'u}{T}\right] \\ &= \left[plim\frac{X\prime X}{T} + plim\frac{e'e}{T}\right]^{-1} plim\frac{X'X\beta}{T} \neq \beta. \end{aligned}$$

If $n = 1$, then equation (7) becomes

$$plim\widehat{b} = \left[plim\frac{\sum\widetilde{x}_t^2}{T} + plim\frac{\sum e_t^2}{T}\right]^{-1} plim\frac{\sum\widetilde{x}_t^2}{T}\beta$$

with the property that $\widehat{b}$ is biased towards 0 (attenuation bias).

2.14 Gauss-Markov Theorem

Theorem 1 *If*

$$y = X\beta + u$$

with

$$u \sim \left(0, \sigma^2 I\right),$$

then

$$\widehat{\beta} = \left(X'X\right)^{-1} X'y$$

is BLUE. (Define BLUE)

Proof. Consider an alternative linear unbiased estimator

$$Ly = L\left(X\beta + u\right).$$

Since $ELy = \beta$ (unbiased),

$$\begin{aligned} LX\beta &= \beta \ \ \forall\beta \\ &\Rightarrow LX = I. \end{aligned}$$

Note that

$$D\left(Ly\right) = \sigma^2 LL'$$

and

$$D\left(\widehat{\beta}\right) = \sigma^2 \left(X'X\right)^{-1}.$$

Define

$$Z = L - \left(X'X\right)^{-1} X',$$

and note that ZZ' is positive semidefinite. Also,

$$\begin{aligned} ZZ' &= \left[L - \left(X'X\right)^{-1} X'\right]\left[L - \left(X'X\right)^{-1} X'\right]' \\ &= LL' - LX\left(X'X\right)^{-1} - \left(X'X\right)^{-1} X'L' + \left(X'X\right)^{-1} X'X\left(X'X\right)^{-1} \\ &= LL' - \left(X'X\right)^{-1} - \left(X'X\right)^{-1} + \left(X'X\right)^{-1} \\ &= LL' - \left(X'X\right)^{-1}. \end{aligned}$$

This implies that

$$D\left(Ly\right) - D\left(\widehat{\beta}\right) \geq 0.$$

■

2.15 Testing

Consider the model

$$y = X\beta + u$$

and the test

$$H_0 : \beta_k = \beta_k^* \;\text{ vs } H_A : \beta_k \neq \beta_k^*.$$

Under H_0,

$$\sqrt{T}\left(\widehat{\beta}_k - \beta_k^*\right) \sim N\left[0, \sigma^2 \left(\frac{X'X}{T}\right)_{kk}^{-1}\right]$$

$\Rightarrow$

$$\frac{\sqrt{T}\left(\widehat{\beta}_k - \beta_k^*\right)}{\sqrt{\sigma^2 \left(\frac{X'X}{T}\right)_{kk}^{-1}}} \sim N\left[0,1\right].$$

We can construct a test statistic using a normal table. Do some examples.

If σ^2 is unknown, then

$$\frac{\sqrt{T}\left(\widehat{\beta}_k - \beta_k^*\right)}{\sqrt{s^2 \left(\frac{X'X}{T}\right)_{kk}^{-1}}} \sim t_{T-(n+1)}.$$

We can construct a test statistic using a t-statistic table. Do some examples. Consider a one-sided test.

Now consider

$$H_0 : a'\beta = c \;\text{ vs } H_A : a'\beta \neq c.$$

Give examples. Note that, under H_0,

$$\begin{aligned} plim\; a'\widehat{\beta} &= a'\beta = c; \\ D\left(a'\widehat{\beta}\right) &= a'D\left(\widehat{\beta}\right)a; \end{aligned}$$

and, therefore,

$$\sqrt{T}\left(a'\widehat{\beta} - c\right) \sim N\left[0, \sigma^2 a' \left(\frac{X'X}{T}\right)^{-1} a\right]$$

and

$$\frac{\sqrt{T}\left(a'\widehat{\beta} - c\right)}{\sqrt{s^2 a' \left(\frac{X'X}{T}\right)^{-1} a}} \sim t_{T-(n+1)}.$$

Now consider

$$H_0 : A\beta = c \text{ vs } H_A : A\beta \neq c.$$

Give examples (including significance of a subset of coefficients). Now

$$\sqrt{T}\left(A\widehat{\beta} - c\right) \sim N\left[0, \sigma^2 A \left(\frac{X'X}{T}\right)^{-1} A'\right].$$

Consider (in general)

$$Z \sim N[0, \Omega],$$

and define R such that

$$\begin{aligned} RR' &= \Omega^{-1} \\ &\Rightarrow \left(R^{-1}\right)' R^{-1} = \Omega. \end{aligned}$$

Consider $R'Z$:

$$\begin{aligned} ER'Z &= R'EZ = 0; \\ D(R'Z) &= ER'ZZ'R \\ &= R'EZZ'R \\ &= R'D(Z)R \\ &= R'\Omega R \\ &= R'\left(R^{-1}\right)' R^{-1} R = I \end{aligned}$$

$$\begin{aligned} &\Rightarrow RZ \sim N(0, I) \\ &\Rightarrow (RZ)'(RZ) \sim \chi^2_n. \end{aligned}$$

Now let

$$RR' = \left[A\left(\frac{X'X}{T}\right)^{-1} A'\right]^{-1} / \sigma^2$$

$$\begin{aligned} &\Rightarrow R'\sqrt{T}\left(A\widehat{\beta} - c\right) \sim N(0, I) \\ &\Rightarrow \left[R'\sqrt{T}\left(A\widehat{\beta} - c\right)\right]' \left[R'\sqrt{T}\left(A\widehat{\beta} - c\right)\right] \sim \chi^2_q \end{aligned}$$

where q is the number of rows in A $(= Rank(R))$

$$\begin{aligned} &\Rightarrow T\left(A\widehat{\beta} - c\right)' RR' \left(A\widehat{\beta} - c\right) \\ &= T\left(A\widehat{\beta} - c\right)' \left[A\left(\frac{X'X}{T}\right)^{-1} A'\right]^{-1} \left(A\widehat{\beta} - c\right) / \sigma^2 \sim \chi^2_q. \end{aligned}$$

If σ^2 is unknown, then

$$T\left(A\widehat{\beta}-c\right)'\left[A\left(\frac{X'X}{T}\right)^{-1}A'\right]^{-1}\left(A\widehat{\beta}-c\right)/qs^2 \sim F_{q,T-(n+1)}$$

(where did the q in the denominator come from?). This is a Wald test statistic.

Alternatively, we could construct a likelihood ratio test statistic. Consider the model

$$y = \underset{T\times n}{X}\beta + \underset{T\times k}{Z}\gamma + u$$

and the test

$$H_0 : \gamma = 0 \ \text{ vs } H_0 : \gamma \neq 0.$$

Estimate

$$\widehat{b} = \left(X'X\right)^{-1}X'y,$$

and define

$$\widehat{e} = y - X\widehat{b} = \left(I - P_X\right)y;$$

these are restricted residuals. Define

$$\frac{ESS_R}{\sigma^2} = \frac{\widehat{e}'\widehat{e}}{\sigma^2} \sim \chi^2_{T-n}.$$

Next, define

$$\widehat{\varepsilon} = \left(I - P_{XZ}\right)y;$$

these are unrestricted residuals. Define

$$ESS_U = \widehat{\varepsilon}'\widehat{\varepsilon}.$$

We can show that

$$\begin{aligned} &\text{Independent} \left\{ \begin{array}{l} \frac{ESS_R - ESS_U}{\sigma^2} \sim \chi^2_k \\ \frac{ESS_U}{\sigma^2} \sim \chi^2_{T-(n+k)} \end{array} \right. \\ \Rightarrow \quad & \frac{\left(ESS_R - ESS_U\right)/k}{ESS_U/\left[T-(n+k)\right]} \sim F_{k,T-(n+k)} \\ \Rightarrow \quad & \frac{\left(R^2_U - R^2_R\right)/k}{\left(1-R^2_U\right)/\left[T-(n+k)\right]} \sim F_{k,T-(n+k)}. \end{aligned}$$

We will discuss likelihood ratio tests (and why this is a likelihood ratio test) in more detail later.