Hayashi Chap 1
Greene 2-4

Linear Regression

$y_i \equiv$ dependent variable for the i^{th} observation
$x_{ij} =$ j^{th} explanatory variable for i^{th} obs.
these are r.v.'s

A1 Linearity

Each obs $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + \varepsilon_i \qquad i = 1, \dots, n$

in matrix notation
$$\underset{n\times 1}{Y} = \underset{n\times k}{X}\underset{k\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

but neither β nor ε is observed

Note: this is more versatile than it seems. Suppose the model is

$$y = \varepsilon A x^{\beta}$$
$$\log y = \log \varepsilon + \log A + \beta \log x$$

This is now in the desired form (called loglinear form or constant elasticity)

semilog: $\ln y_t = x_t'\beta + \varepsilon_t$

A2 Strict exogeneity

$$E(\varepsilon_i \mid X) = 0 \quad \forall i$$

equiv

$$E(\varepsilon_i \mid \underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n) = 0$$

x doesn't convey info about expected value of all ~~orthogonal/unrelated to regressors~~ ε

if the $1, \dots, n$ are t-s obs, this says the error is unrelated to all past and future regressors

Note that most regressions have a constant term

Implications of A2

unconditional $E(\varepsilon_i) = 0$ by law of iterated expectations

$$E(\varepsilon_i) = E_x\left[E(\varepsilon_i \mid X)\right] =$$
$$= E_x[0]$$

error term is orthogonal to all regressors — own and other

(2) $E(x_{jk}\varepsilon_i) = 0 \quad \forall i,j = 1,\dots,n \quad k = 1,\dots,K$

Pf.

~~By law of iterated expectations~~ $\cancel{E(\varepsilon_i \mid x_{jk}) = E\left[E(\varepsilon_i \mid X) \mid x_{jk}\right] = 0}$

$\cancel{E(\varepsilon_i \mid x_{jk}) = 0}$ ~~by strict exogeneity~~

$$
\begin{aligned}
E\left[x_{jk}\varepsilon_i\right] &= E\left[E(x_{jk}\varepsilon_i \mid x_{jk})\right] && \text{by law of iterated exp.} \\
&= E\left[x_{jk}\underbrace{E(\varepsilon_i \mid x_{jk})}\right] && \text{by linear cond'l exp} \\
&= 0 && \text{by strict exogeneity}
\end{aligned}
$$

Since $E(\varepsilon_i) = 0$, this also means $Cov(x_{jk}, \varepsilon_i) = 0$

Ex. failure of strict exogeneity

AR(1) $\quad y_i = \beta y_{i-1} + \varepsilon_i \qquad$ Suppose $E\left[y_{i-1}\varepsilon_i\right] = 0$

Then

$$
\begin{aligned}
E\left[y_i\varepsilon_i\right] &= E\left[(\beta y_{i-1} + \varepsilon_i)\varepsilon_i\right] \\
&= E\left[\beta \underbrace{y_{i-1}\varepsilon_i}_{0} + \underbrace{\varepsilon_i^2}\right] \\
&= E\left[\varepsilon_i^2\right] > 0 \quad \text{in gen'l}
\end{aligned}
$$

and y_i is the regressor for obs $i+1$, $\Rightarrow$ violates (A2)

A3 $\quad X$ has full column rank (rank K)

$\Rightarrow$ columns of X are linearly indep

$\Rightarrow$ at least K observations

Ex. collinearity

$$R_i = \beta_0 + \beta_1\,\text{CEOsalary}_i + \beta_2\,\text{CEObonus}_i + \beta_3\,\text{CEOtotalincome}_i + \varepsilon_i$$

where salary + bonus = total income

Define $\beta_1' = \beta_1 + a \qquad \beta_3' = \beta_3 - a \qquad$ so $\beta_1\ \beta_2\ \beta_3$

$\beta_2' = \beta_2 + a \qquad$ this works also $\qquad$ underid'd

$\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ but σ^2 not observed. what to do?

Recall that $e = My = M(X\beta + \varepsilon) = M\varepsilon$

so $e'e = \varepsilon' M \varepsilon$

Now $E(\varepsilon' M \varepsilon \mid X) = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} E(\varepsilon_i \varepsilon_j \mid X)$ because M is a fn of X

$= \sum_{i=1}^{n} m_{ii} \sigma^2$

$= \sigma^2 \sum_{i=1}^{n} m_{ii}$

$= \sigma^2 \, \text{trace}(M)$ (defn of trace for square matri

$\text{trace}(M) = \text{trace}(I - P) = \text{trace}(I) - \text{trace}(P)$ by linearity of trace

$= n - \text{trace}(P)$

$\text{trace}(P) = \text{trace}(X(X'X)^{-1}X')$

$= \text{trace}((X'X)^{-1}(X'X))$ $\text{trace}(AB) = \text{trace}(BA)$

$= \text{trace}(I_k) = k$

so $E(e'e \mid X) = \sigma^2 \cdot (n-k)$

$s^2 = \dfrac{e'e}{n-k}$ $E(s^2) = \sigma^2$

(also unbiased unconditionally)

and $\hat{\text{var}}(\hat{\beta} \mid X) = s^2 (X'X)^{-1}$

In order to conduct inference, we need a dist'n for $\hat{\beta}$

(A5) normal errors $\underset{\sim}{\varepsilon} \mid X \sim N(0, \sigma^2 I) \Rightarrow \hat{\beta} \sim N(\beta, \sigma^2 (X'X)$

ε, X are indep

why does this work?

note that we got the mean and var above.. Add normality and we're done

(A4) Spherical error variance

$$\mathrm{var}(\varepsilon_i^2 \mid X) = \sigma^2 \quad \forall i \qquad \text{homoskedasticity}$$

$$\mathrm{cov}(\varepsilon_i, \varepsilon_j \mid X) = 0 \qquad \text{no c-s or t-s correlation}$$

or equiv.

$$E[\varepsilon\varepsilon' \mid X] = \sigma^2 I$$

In experimental sciences, X is fixed or deterministic

In social sciences, X is often stochastic as well. So we view (y, X) as ~~a~~ r.v.'s

That's why we condition on X here.

If X is fixed, then

(A2): $E(\varepsilon_i) = 0$

(A4): $E(\varepsilon\varepsilon') = \sigma^2 I$

Sometimes we'll suppress the $\mid X$, but don't forget it's always there

ε's must be uncorr'd w/ every regressor

Now back to the OLS problem in matrix form

$$\hat{\beta}_{OLS} = \underset{\beta}{\mathrm{argmin}} \; (Y - X\beta)'(Y - X\beta) \qquad (\text{SSR})$$

$$y'y - \beta'X'y - y'X\beta + \beta'X'X\beta$$

$$y'y - 2\underbrace{\cancel{\beta}\cancel{X'X\beta}}_{\beta'X'y} + \beta'X'X\beta$$

FOC

$\partial/\partial\beta = 0$ $\qquad -2X'y + 2X'X\hat{\beta} = 0$

by (A2), $(X'X)^{-1}$ exists, so $\qquad \hat{\beta} = (X'X)^{-1}(X'y)$

Define $e = y - X\hat{\beta}$ $\qquad X'(y - X\hat{\beta}) = 0$ $\qquad$ sample orthogonality

$$X'e = 0$$

Projection

$$e = y - X\hat{\beta}$$
$$= y - X(X'X)^{-1}X'y$$
$$e = (I - X(X'X)^{-1}X')y$$
$$= \underset{n\times n}{M}y$$

symmetric, idempotent

M is called the residual-maker $My = e$

or the annihilator, as $MX = 0$

(if you regress X on X, you get a perfect fit)

Define Fitted value $\hat{y} = X\hat{\beta}$

Then

$$\hat{y} = y - e = (I - M)y = X(X'X)^{-1}X'y = \underset{n\times n}{P}y$$

give simple intuition

P is a projection matrix (also symmetric and idempotent)

P and M are orthogonal ($PM = MP = 0$)

$PX = X$

so $y = \underbrace{P\cancel{X}^{y}}_{\text{projection}} + \underbrace{M\cancel{X}^{y}}_{\text{residual}}$

note also that

$$e'e = y'M'My = y'My = y'e = e'y$$

Goodness of Fit

idea: magnitude of SSR depends on its units

want to scale by total variation in $y \equiv SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$

$$= \sum_{i=1}^{n}(\hat{y}_i - \bar{y} + e_i)^2$$
$$\underset{SST}{} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{SSR} + \underbrace{\sum_{i=1}^{n} e_i^2}_{SSE}$$

) these make up an analysis of variance

Def

$$R^2 \equiv \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

So as long as there's an intercept $R^2 \in [0,1]$

R^2 never goes down when you add a variable

Pf (outline)

SSE stays the same if the old coefs stay the same, new coef = 0
can only improve from there, & weakly lower SSE $\Rightarrow R^2 \uparrow$ weakly

Ajusted R^2 ($\bar{R}^2$) · provides a penalty for each new variable

Def $$\bar{R}^2 = 1 - \frac{n-1}{n-K}(1-R^2)$$

Note:

$\bar{R}^2$ declines if the |t-stat| on a new variable < 1

$\bar{R}^2$ can even be negative (see above w/ no explanatory power of regressors)

Computational note: if you're rolling your own in Matlab, careful w/ $(X'X)^{-1}$. if one variable is million, another is large then $(X'X)^{-1}$ can be ill-conditioned

scale or use a pkg. or in Matlab use $b = X \backslash y$

Properties of the OLS estimator

A1 – A3 $\Rightarrow E(\hat{\beta} | X) = \beta$ unbiased

A1 – A4 $\Rightarrow var(\hat{\beta} | X) = \sigma^2 (X'X)^{-1}$

A1 – A4 $\Rightarrow \hat{\beta}$ is BLUE

unbiasedness

$$\hat{\beta} = (X'X)^{-1}(X'y) = (X'X)^{-1}X'(X\beta + \varepsilon)$$
$$= \beta + (X'X)^{-1}X'\varepsilon$$
$$E(\hat{\beta} | X) = \beta + E[(X'X)^{-1}X'\varepsilon \mid X]$$
$$= \beta + (X'X)^{-1}X' E[\overset{0}{\varepsilon | X}]$$
$$= \beta$$

also unconditionally unbiased

$$E(\hat{\beta} \mid X) = \beta$$

$$E[E(\hat{\beta} \mid X)] = \beta$$

$$E(\hat{\beta}) = \beta$$

(b) $\mathrm{var}(\hat{\beta} \mid X) = \mathrm{var}(\hat{\beta} - \beta \mid X)$ $\quad$ β not random

$$= \mathrm{var}((X'X)^{-1}X'\varepsilon \mid X)$$

$$= (X'X)^{-1}X' \, \mathrm{var}(\varepsilon \mid X) \, X(X'X)^{-1}$$

$$= (X'X)^{-1}X' \, \sigma^2 I \, X(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

(c) $\hat{\beta}$ efficient (BLUE)

Consider another estimator $b = Cy$ that's linear in y

Define $D \equiv C - (X'X)^{-1}X'$

Then $b = (D + (X'X)^{-1}X')y$

$$= D(X\beta + \varepsilon) + \hat{\beta}$$

$$E(b \mid X) = DX\beta + D\underbrace{E(\varepsilon \mid X)}_{0} + E(\hat{\beta} \mid X)$$

for this other estimator to be unbiased, $DX\beta = 0 \Rightarrow DX = 0$

~~$b - \beta = D\varepsilon + (\hat{\beta} - \beta)$~~

now $\mathrm{var}(b) = \mathrm{var}[D + (X'X)^{-1}X'\varepsilon \mid X]$

$$= (D + (X'X)^{-1}X') \, \mathrm{var}(\varepsilon \mid X) \, (X(X'X)^{-1} + D')$$

$$= \sigma^2 [\underbrace{DX}_{0}(X'X)^{-1} + DD' + (X'X)^{-1} + (X'X)^{-1}\underbrace{X'D'}_{0}]$$

$DX = 0$

$$= \sigma^2 [DD' + (X'X)^{-1}]$$

$$\geq \sigma^2 (X'X)^{-1} \text{ because } DD' \text{ is p.s.d.}$$

Test $\beta_k = c$

If we knew σ^2, then we'd use a z-statistic

$$z_k = \frac{\beta_k - c}{\sqrt{(\sigma^2(X'X)^{-1})_{kk}}} \sim N(0,1)$$

$\leftarrow$ the kk^{th} element

Thm Under (A1)–(A5),

$$t_k = \frac{\beta_k - c}{\sqrt{[s^2(X'X)^{-1}]_{kk}}} \sim t_{n-k}$$

pf. see Greene or Hayashi

Linear hypotheses

H_0: $R\beta = r$ (R, r constant)

ex. $\beta_1 = 1$ $\beta_2 - \beta_3 = \beta_4$ $R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$ $r = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Thm Under (A1)–(A5), with rank$(R) = r^*$

$$F \equiv (R\hat{\beta} - r)' \left[\frac{R(X'X)^{-1}R'}{s^2} \right]^{-1} (R\hat{\beta} - r)' / r^* \sim F_{r^*, n-k}$$

Pf. see Hayashi

basic idea $\quad var(\hat{\beta} \mid X) = \sigma^2(X'X)^{-1}$

$var(R\hat{\beta} \mid X) = \sigma^2 R(X'X)^{-1}R'$

$E(R\hat{\beta}) = r$ under H_0

quadratic form $X \sim N(\mu, \Sigma)$

$$(X-\mu)' \Sigma^{-1} (X - \mu) \sim \chi^2_k$$

$$(R\beta - r)' [\sigma^2 R(X'X)^{-1}R']^{-1} (R\beta - r) \sim \chi^2_k$$

replace σ^2 w/ s^2 and get an F instead

Note: the t-test is a special case of the F test $\qquad F_{1,n-k} = T^2_{n-k}$