(without an intercept) of $\underset{\sim}{Z}$ on U.

c) For WLS, $Y_i = \underline{x}_i^T \underline{\beta} + \varepsilon_i$. The corresponding OLS model $\underline{Z} = U\underline{\beta} + \underset{\sim}{a}$ is equivalent to $Z_i = \underline{u}_i^T \underline{\beta} + a_i$ for $i = 1, \ldots, n$ where $\underline{u}_i^T$ is the ith row of U. Then $Z_i = \sqrt{w_i}\, Y_i$ and $\underline{u}_i = \sqrt{w_i}\, \underset{\sim}{x}_i$. Hence $\hat{\underline{\beta}}_{WLS}$ can be obtained from the OLS regression (without an intercept) of $Z_i = \sqrt{w_i}\, Y_i$ on $\underline{u}_i = \sqrt{w_i}\, \underline{x}_i$.

proof] a) $E(\underline{a}) = R^{-1} E(\underline{\varepsilon}) = \underline{0}$

$$\text{Cov}(\underline{a}) = R^{-1}\, \text{Cov}(\underline{\varepsilon})\, (R^{-1})^T = \sigma^2 R^{-1} V R^{-1}$$

$$= \sigma^2 R^{-1} R R R^{-1} = \sigma^2 I_n$$

OLS without an intercept needs to be used since the 1st column of U is $R^{-1} \underset{\sim}{1} \neq \underset{\sim}{1}$.

b) Let $\hat{\underset{\sim}{\beta}}_{ZU}$ denote the OLS estimator obtained by regressing $\underline{Z}$ on U. Then

$$\hat{\underline{\beta}}_{ZU} = (U^T U)^{-1} U^T \underline{Z} = [X^T (R^{-1})^T R^{-1} X]^{-1} X^T (R^{-1})^T R^{-1} \underset{\sim}{Y}$$

and the result follows since

$$V^{-1} = (RR)^{-1} = R^{-1}R^{-1} = (R^{-1})^T R^{-1}.$$

c) The result follows from b) if

$$z_i = \sqrt{w_i}\, Y_i \text{ , and } \underline{u}_i = \sqrt{w_i}\, \underline{x}_i .$$

But for WLS, $V = diag(v_1, ..., v_n)$ and

so $R = diag(\sqrt{v_1}, ..., \sqrt{v_n})$. Hence

$$R^{-1} = diag\left(\frac{1}{\sqrt{v_1}}, ..., \frac{1}{\sqrt{v_n}}\right) = diag\left(\sqrt{w_1}, ..., \sqrt{w_n}\right).$$

So $\underset{\sim}{Z} = R^{-1}\underset{\sim}{Y}$ has ith element $z_i = \sqrt{w_i}\, Y_i$

and $U = R^{-1}X$ has ith row $= \sqrt{w_i}\, \underset{\sim}{x}_i^T = \underline{u}_i^T$

4.4) Standard software will do WLS if supplied the weights.

4.5) The FGLS estimator can be found with the OLS regression (without intercept) of $\underset{\sim}{Z}$ on U where $V(\hat{e}) = RR$ and the FWLS estimator can be found with the OLS regression (without intercept) of $z_i = \sqrt{\hat{w}_i}\, Y_i$ on $\underset{\sim}{u}_i = \sqrt{\hat{w}_i}\, \underline{x}_i$. Now U

is a random matrix instead of a constant matrix. Hence FGLS and FWLS are highly nonlinear estimators.

4B) Under regularity conditions, the OLS estimator $\hat{\underline{\beta}}$ is a consistent estimator of $\underline{\beta}$ when the GLS model holds, but $\hat{\underline{\beta}}_{GLS}$ should generally be used because it generally has higher efficiency (use the Gauss Markov Th on the OLS model $\underline{Z} = U\underline{\beta} + \underline{\epsilon}$).

Skip §3.11, 3.12, 3.13

Ch4 Hypothesis Testing

Skim §4.1 skip §4.2

1) p99 Want to test Ho $\underset{q \times p}{A}\,\underline{\beta} = \underset{q \times 1}{\underline{c}}$.

If $\hat{\underline{\beta}} \sim N_p[\underline{\beta}, \sigma^2(X'X)^{-1}]$, then

$$A\hat{\underline{\beta}} - \underline{c} \sim N_q(\underbrace{A\underline{\beta} - \underline{c}}_{\underline{0} \text{ under Ho}}, \sigma^2 A(X'X)^{-1}A')$$

Hence $qF = \dfrac{(A\hat{\underset{\sim}{\beta}} - \underset{\sim}{c})^T [A(X'X)^{-1}A^T]^{-1} (A\hat{\beta} - \underset{\sim}{c})}{\sigma^2} \overset{H_0}{\sim} \chi^2_q.$ replace σ^2 by MSE

By the LS CLT, $\dfrac{\sigma^2}{MSE} qF \xrightarrow[D]{H_0} \chi^2_q$ if

$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$, X full rank, $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$ $Cov(\underset{\sim}{\epsilon}) = \sigma^2 I$,

for the large class of $\overset{iid}{\epsilon_i}$ error distributions.

2) 3 important tests have the form $A\underset{\sim}{\beta} = \underset{\sim}{0}$. Suppose the full model is $Y_i = \underbrace{\beta_0}_{\text{constant or intercept}} + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i$.

So the full model uses $Y, x_0 \equiv 1, x_1, \ldots, x_{p-1}$.

Let the reduced model use $Y, x_0 \equiv 1, x_{j_1}, x_{j_2}, \ldots, x_{j_K}$

where $\{j_1, \ldots, j_K\} \subset \{1, \ldots, p-1\}$, and $0 \le K < p-1$.

$Y_i = \beta_0 + \beta_{j_1} x_{i,j_1} + \cdots + \beta_{j_K} x_{i,j_K} + \epsilon_i$ $\qquad$ $K=0$ means ϕ.

i) The partial F test has $H_0: \beta_{j_{K+1}} = \cdots = \beta_{j_{p-1}} = 0$

or H_0 the reduced model is good.

Then the ith row of A has a 1 in the j_{K+i} position and 0's elsewhere.

In particular, if $\beta_0, \ldots, \beta_k$ are in the reduced model then

$$A = (0 \quad I_{p-k-1}) \quad \text{and}$$

$$\underset{(p-k-1)\times p}{A}\,\underset{\sim}{\beta} = (0 \quad I_{p-k-1}) \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \\ \hline \beta_{k+1} \\ \vdots \\ \beta_{p-1} \end{pmatrix} \begin{matrix} \}\, k+1 \\ \\ \}\, p-k-1 \end{matrix} = \begin{pmatrix} \beta_{k+1} \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

$q = p-k-1 = \#$ of predictors in full model but not in the reduced model

$p-1 = k + p-k-1$

ii) The <u>Anova F test</u> is the special case where the reduced model is $Y = \beta_0 + \varepsilon$ and tests Ho none of the nontrivial predictors are needed in the linear model

$$\underset{(p-1)\times p}{A}\,\underset{\sim}{\beta} = \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad I_{p-1} \right] \begin{pmatrix} \beta_0 \\ \hline \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}. \qquad H_0\ \beta_1 = \cdots \beta_{p-1} = 0$$

iii) The <u>Wald t test</u> is a test for whether x_j is needed in the model given that the other predictors are in the model where $x_0 \equiv 1$ is the trivial predictor.

$$\underset{\sim}{X} = \left[\overset{x_0}{\underset{\sim}{v_1}} \cdots \overset{x_{p-1}}{\underset{\sim}{v_p}} \right] \qquad H_0\ \beta_j = 0.$$

$\underset{1\times p}{A_j} = [0, \ldots, 0, 1, 0, \ldots, 0]$, $j = 0, \ldots, p-1$ and 1 is in the $j+1$ position.

Here the reduced model does not contain x_j
predictor x_j.

3] p100 A way to get a statistic for the partial F test is to fit the full model and the reduced model. Let $RSS = RSS(F)$ the RSS of the full model. Let $RSS(R) = RSS_H$ be the RSS $= RSS(F)$ the reduced model. Then the partial F test has test statistic

$$F_R = \frac{(RSS(R) - RSS(F))/q}{RSS(F)/(n-p)} = \frac{RSS(R) - RSS(F)}{q \underbrace{MSE(F)}_{\text{for full model}}}$$

Here $q = df_R - df_F = (n-k-1) - (n-p) = p-k-1$

$=$ # of terms in full model not in reduced model.

4] p100 Partial F Test Theorem Suppose H_0 $A\underline{B} = \underline{0}$ is true for the partial F test. Under the OLS full rank model,

a) if $\underline{e} \sim N_n(\underline{0}, \sigma^2 \underline{I})$ then a)

$$F_R = \frac{(A\hat{\beta})^T [A(X^TX)^{-1}A^T]^{-1} A\hat{\beta}}{q\, MSE}$$

b) If $\epsilon \sim N_n(0, \sigma^2 I)$, then $F_R \overset{H_0}{\sim} F_{q, n-p}$.

c) For a large class of zero mean error distributions, $q F_R \xrightarrow[H_0]{D} \chi^2_q$, by [1],

I don't show numerator = RSS(R) − RSS in this class, see Seber and Lee p100.

5) Recall that $W \sim F_{d_1, d_2}$ if

$W = \frac{X_1/d_1}{X_2/d_2}$ where $X_1 \sim \chi^2_{d_1} \perp X_2 \sim \chi^2_{d_2}$

also $X \sim \chi^2_n$ if $X = \sum_{i=1}^n U_i$ where

U_i are iid χ^2_1.

Sample mean of the U_i

Hence if $X_2 \sim \chi^2_{n-p}$, then $\frac{X_2}{n-p} = \frac{\sum_{i=1}^{n-p} U_i}{n-p} = \bar{U}_{n-p}$

$\xrightarrow{P} 1$. Hence if $W \sim F_{q, n-p}$,

then $q W \xrightarrow{D} \chi^2_q$

so $1-\delta = P(W \le F_{q, n-p, 1-\delta}) \approx P(q W \le \chi^2_{q, 1-\delta})$

So if $W \xrightarrow{D} \chi^2_q$

then $P\left(\frac{W}{q} \le F_{q,n-p,1-\delta}\right) \approx 1-\delta \approx P(W \le \chi^2_{q,1-\delta})$.

So $P\left(qF_R \le \chi^2_{q,1-\delta}\right) \approx P\left(\frac{F_R}{q} \le F_{q,n-p,1-\delta}\right)$.

6) ✓ A test with test statistic T_n is a <u>large sample right tailed δ test</u> if the test rejects H_0 if $T_n > a_n$ and $P_{H_0}(T_n > a_n) = \delta_n \to \delta$ as $n \to \infty$. If $T_n \xrightarrow{D} \chi^2_q$ when H_0 is true then the tests that reject H_0 when $T_n > \chi^2_{q,1-\delta}$ or when $\frac{T_n}{q} > F_{q,d_n,1-\delta}$ are large sample right tail δ tests if $d_n \to \infty$ as $n \to \infty$. Using an F distribution gives better results than using a χ^2 dist.

7) Analogy if the test statistic $T_n \xrightarrow{D} N(0,1)$ then the tests that reject H_0 if $T_n > z_{1-\delta}$ or if $T_n > t_{d_n,1-\delta}$ are large sample right tail δ tests if $d_n \to \infty$ as $n \to \infty$.