Elementary Regression Theory–Part 2

Richard E. Quandt
Princeton University

Corollary 3. *If $r - R\hat{\beta} = 0$, then $E(\tilde{\beta}) = \beta$ and $\tilde{\lambda} = 0$.*

Proof. Substituting $X\beta + u$ for Y and $(X'X)^{-1}X'Y$ for $\hat{\beta}$ in the expression for $\tilde{\beta}$ in Eq.(34a) and taking expectations yields the result for $E(\tilde{\beta})$. Using this by replacing $\tilde{\beta}$ by β in the formula for $\tilde{\lambda}$ yields the result for $E(\tilde{\beta})$.

Define

$$A = I - (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R. \tag{35}$$

We then have

Theorem 25. *The covariance matrix of $\tilde{\beta}$ is $\sigma^2 A(X'X)^{-1}$.*

Proof. Substituting $(X'X)^{-1}X'Y$ for $\hat{\beta}$ and $X\beta + u$ for Y in $\tilde{\beta}$ (Eq.(34a)), we can write

$$\begin{aligned} \tilde{\beta} - E(\tilde{\beta}) &= (X'X)^{-1}\big[X' - R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1}X'\big]u \\ &= \big[I - (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R\big](X'X)^{-1}X'u = A(X'X)^{-1}X'u. \end{aligned} \tag{36}$$

Multiplying Eq.(36) by its transpose and taking expectations, yields

$$\begin{aligned} \mathrm{Cov}(\tilde{\beta}) &= \sigma^2 A(X'X)^{-1}A' \\ &= \sigma^2\big[I - (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R\big](X'X)^{-1}\big[I - R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1}\big] \\ &= \sigma^2\big[(X'X)^{-1} - (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1} \\ &\qquad - (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1} \\ &\qquad + (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1}\big] \\ &= \sigma^2\big[(X'X)^{-1} - (X'X)^{-1}R'\big[R(X'X)^{-1}R'\big]^{-1}R(X'X)^{-1}\big] \\ &= \sigma^2 A(X'X)^{-1}. \end{aligned}$$

∎

We now consider the test of the null hypothesis $H_0 : R\beta = r$. For this purpose we construct an F-statistic as in Theorem 20 (see also Eq.(12)).

The minimum sum of squares subject to the restriction can be written as

$$
\begin{aligned}
S_r &= \left\{Y - X\left[\hat\beta + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat\beta)\right]\right\}' \\
&\quad \times \left\{Y - X\left[\hat\beta + (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat\beta)\right]\right\} \\
&= (Y - X\hat\beta)'(Y - X\hat\beta) - \left[X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat\beta)\right]'(Y - X\hat\beta) \\
&\quad - (Y - X\hat\beta)'\left[X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat\beta)\right] \\
&\quad + (r - R\hat\beta)'\left[R(X'X)^{-1}R'\right]^{-1}(X'X)^{-1}(X'X)(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat\beta) \\
&= S_u + (r - R\hat\beta)'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat\beta),
\end{aligned}
\tag{37}
$$

where S_u denotes the unrestricted minimal sum of squares, and where the disappearance of the second and third terms in the third and fourth lines of the equation is due to the fact that that $X'(Y - X\hat\beta) = 0$ by the definition of $\hat\beta$. Substituting the least squares estimate for $\hat\beta$ in (37), we obtain

$$
\begin{aligned}
S_r - S_u &= [r - R(X'X)^{-1}X'Y]'\left[R(X'X)^{-1}R'\right]^{-1}[r - R(X'X)^{-1}X'Y] \\
&= [r - R\beta - R(X'X)^{-1}X'u]'\left[R(X'X)^{-1}R'\right]^{-1}[r - R\beta - R(X'X)^{-1}X'u] \\
&= u'X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'u = u'B_1u,
\end{aligned}
\tag{38}
$$

since under H_0, $r - R\beta = 0$. The matrix B_1 is idempotent and of rank p because

$$
\begin{aligned}
X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}(X'X)(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X' \\
= X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{tr}(B_1) &= \operatorname{tr}(X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X') \\
&= \operatorname{tr}(R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}(X'X)(X'X)^{-1}) \\
&= \operatorname{tr}(\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}R') = \operatorname{tr}(I_p) = p.
\end{aligned}
$$

The matrix of the quadratic form S_u is clearly $B_2 = I - X(X'X)^{-1}X'$ which is idempotent and or rank $n - k$. Moreover, $B_1B_2 = 0$, since

$$
X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'(I - X(X'X)^{-1}X') = 0.
$$

Hence

$$
\frac{(S_r - S_u)/p}{S_u/(n-k)}
$$

is distributed as $F(p, n - k)$.

We now turn to the case in which the covariance matrix of u is Ω and we wish to test the hypothesis $H_0 : R\beta = r$. We first assume that Ω is known. We first have

Theorem 26. *If u is distributed as $N(0, \Omega)$, and if Ω is known, then the Lagrange Multiplier, Wald, and likelihood ratio test statistics are identical.*

Proof. The loglikelihood function is

$$\log L(\beta) = (2\pi)^{-n/2} + \frac{1}{2}\log|\Omega^{-1}| - \frac{1}{2}(Y - X\beta)'\Omega^{-1}(Y - X\beta),$$

where $|\Omega^{-1}|$ denotes the determinant of Ω^{-1}, and the score vector is

$$\frac{\partial \log L}{\partial \beta} = X'\Omega^{-1}(Y - X\beta).$$

By further differentiation, the Fischer Information matrix is

$$I(\beta) = X'\Omega^{-1}X.$$

The unrestricted maximum likelihood estimator for β is obtained by setting the score vector equal to zero and solving, which yields

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

Letting $\hat{u}$ denote the residuals $Y - X\hat{\beta}$, the loglikelihood can be written as

$$\log L = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\Omega^{-1}| - \frac{1}{2}\hat{u}'\Omega^{-1}\hat{u}.$$

To obtain the estimates restricted by the linear relations $R\beta = r$, we form the Lagrangian

$$L(\beta, \lambda) = \log L(\beta) + \lambda'(R\beta - r)$$

and set its partial derivatives equal to zero, which yields

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= X'\Omega^{-1}(Y - X\beta) + R'\lambda = 0 \\ \frac{\partial \log L}{\partial \lambda} &= R\beta - r = 0. \end{aligned} \tag{39}$$

Multiply the first equation in (39) by $(X'\Omega^{-1}X)^{-1}$, which yields

$$\tilde{\beta} = \hat{\beta} + (X'\Omega^{-1}X)^{-1}R'\tilde{\lambda}.$$

Multiplying this further by R, and noting that $R\tilde{\beta} = r$, we obtain

$$\tilde{\lambda} = -\big[R(X'\Omega^{-1}X)^{-1}R'\big]^{-1}(R\hat{\beta} - r) \tag{40}$$

$$\tilde{\beta} = \hat{\beta} - (X'\Omega^{-1}X)^{-1}R'\big[R(X'\Omega^{-1}X)^{-1}R'\big]^{-1}(R\hat{\beta} - r). \tag{41}$$

The loglikelihood, evaluated at $\tilde{\beta}$ is

$$\log L(\tilde{\beta}) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\Omega^{-1}| - \frac{1}{2}\tilde{u}'\Omega^{-1}\tilde{u}.$$

We now construct the test statistics. The Lagrange multiplier statistic is

$$\begin{aligned} LM &= \left[\frac{\partial \log L}{\partial \tilde{\beta}}\right]' I(\tilde{\beta})^{-1}\left[\frac{\partial \log L}{\partial \tilde{\beta}}\right] = \tilde{u}'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\tilde{u} \\ &= \tilde{\lambda}'R(X'\Omega^{-1}X)^{-1}R'\tilde{\lambda}. \end{aligned} \tag{42}$$

The Wald statistic is

$$W = (R\tilde{\beta} - r)'\big[R(X'\Omega^{-1}X)^{-1}R'\big]^{-1}(R\tilde{\beta} - r), \tag{43}$$

and since the covariance matrix of $(R\tilde{\beta} - r)$ is $R(X'\Omega^{-1}X)^{-1}R'$, W can be written as

$$\begin{aligned} W &= (R\tilde{\beta} - r)'\big[R(X'\Omega^{-1}X)^{-1}R'\big]^{-1}\big[R(X'\Omega^{-1}X)^{-1}R'\big]\big[R(X'\Omega^{-1}X)^{-1}R'\big]^{-1}(R\tilde{\beta} - r) \\ &= \tilde{\lambda}'\big[R(X'\Omega^{-1}X)^{-1}R'\big]\tilde{\lambda} = LM, \end{aligned}$$

where we have used the definition of $\tilde{\lambda}$ in (40). The likelihood ratio test statistic is

$$LR = -2\big[\log L(\tilde{\beta}) - \log L(\hat{\beta})\big] = \tilde{u}'\Omega^{-1}\tilde{u} - \hat{u}'\Omega^{-1}\hat{u}. \tag{44}$$

Since $\Omega^{-1/2}\tilde{u} = \Omega^{-1/2}(Y - X\tilde{\beta})$, and substituting in this for $\tilde{\beta}$ from its definition in (41), we obtain

$$\Omega^{-1/2}\tilde{u} = \Omega^{-1/2}\big[Y - X\hat{\beta} - X(X'\Omega^{-1}X)^{-1}R'\tilde{\lambda}\big]. \tag{45}$$

We multiply Eq.(45) by its transpose and note that terms with $(Y - X\hat{\beta})\Omega^{-1}X$ vanish; hence we obtain

$$\tilde{u}'\Omega^{-1}\tilde{u} = \hat{u}'\Omega^{-1}\hat{u} + \tilde{\lambda}'R(X'\Omega^{-1}X)^{-1}R'\tilde{\lambda}.$$

But the last term is the Lagrange multiplier test statistic from (42); hence comparing this with (44) yields $LR = LM$. ∎

We now consider the case when Ω is unknown, but is a smooth function of a p-element vector α, and denoted by $\Omega(\alpha)$. We then have

Theorem 27. *If u is normally distributed as $N(0, \Omega(\alpha))$, then $W \geqq LR \geqq LM$.*

Proof. Denote by θ' the vector (β', α'). The loglikelihood is

$$\log L(\theta) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\Omega^{-1}(\alpha)| + \frac{1}{2}(Y - X\beta)'\Omega^{-1}(\alpha)(Y - X\beta).$$

Denoting the unrestricted estimates by $\hat{\theta}$ and the restricted estimates by $\tilde{\theta}$, as before, and in particular, denoting by $\hat{\Omega}$ the matrix $\Omega(\hat{\alpha})$ and by $\tilde{\Omega}$ the matrix $\Omega(\tilde{\alpha})$, the three test statistics can be written, in analogy with Eqs.(42) to (44), as

$$\begin{aligned} LM &= \tilde{u}'\tilde{\Omega}^{-1}X(X'\tilde{\Omega}^{-1}X)^{-1}X'\tilde{\Omega}^{-1}\hat{u} \\ W &= (R\hat{\beta}-r)'\big[R(X'\hat{\Omega}^{-1}X)^{-1}R'\big]^{-1}(R\hat{\beta}-r) \\ LR &= -2\big(\log L(\tilde{\alpha},\tilde{\beta})-\log L(\hat{\alpha},\hat{\beta})\big). \end{aligned}$$

Now define

$$LR(\tilde{\alpha}) = -2\big(\log L(\tilde{\alpha},\tilde{\beta})-\log L(\tilde{\alpha},\tilde{\beta}_u)\big), \tag{46}$$

where $\tilde{\beta}_u$ is the unrestricted maximizer of $\log L(\tilde{\alpha},\beta)$ and

$$LR(\hat{\alpha}) = -2\big(\log L(\hat{\alpha},\hat{\beta}_r)-\log L(\hat{\alpha},\hat{\beta})\big), \tag{47}$$

where $\hat{\beta}_r$ is the maximizer of $\log L(\hat{\alpha},\beta)$ subject to the restriction $R\beta - r = 0$. $LR(\tilde{\alpha})$ employs the same Ω matrix as the LM statistic; hence by the argument in Theorem 26,

$$LR(\tilde{\alpha}) = LM.$$

It follows that

$$LR - LM = LR - LR(\tilde{\alpha}) = 2\big(\log L(\hat{\alpha},\hat{\beta}) - \log L(\tilde{\alpha},\tilde{\beta}_u)\big] \geqq 0,$$

since the $\hat{\alpha}$ and $\hat{\beta}$ estimates are unrestricted. We also note that W and $LR(\hat{\alpha})$ use the same Ω, hence they are equal by Theorem 26. Then

$$W - LR = LR(\hat{\alpha}) - LR = 2\big(\log L(\tilde{\alpha},\tilde{\beta}) - \log L(\hat{\alpha},\hat{\beta}_r)\big) \geqq 0, \tag{48}$$

since $\hat{\beta}_r$ is a restricted estimate and the highest value of the likelihood with the restriction that can be achieved is $\log L(\tilde{\alpha},\tilde{\beta})$. Hence $W \geqq LR \geqq LM$. ∎

We now prove a matrix theorem that will be needed subsequently.

Theorem 28. *If Σ is symmetric and positive definite of order p, and if H is of order $p \times q$, with $q \leqq p$, and if the rank of H is q, then*

$$\begin{bmatrix} \Sigma & H \\ H' & 0 \end{bmatrix}$$

is nonsingular.

Proof. First find a matrix, conformable with the first,

$$\begin{bmatrix} P & Q \\ Q' & R \end{bmatrix}$$

such that

$$\begin{bmatrix} \Sigma & H \\ H' & 0 \end{bmatrix} \begin{bmatrix} P & Q \\ Q' & R \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Performing the multiplication and equating the two sides, we obtain

$$\Sigma P + HQ' = I \tag{49}$$

$$\Sigma Q + HR = 0 \tag{50}$$

$$H'P = 0 \tag{51}$$

$$H'Q = I \tag{52}$$

From (49) we have

$$P + \Sigma^{-1}HQ' = \Sigma^{-1}. \tag{53}$$

Multiplying Eq.(53) on the left by H', and noting from Eq.(51) that $H'P = 0$, we have

$$H'\Sigma^{-1}HQ' = H'\Sigma^{-1}. \tag{54}$$

Since H is of full rank, $H'\Sigma^{-1}H$ is nonsingular by a straightforward extension of Lemma 1. Then

$$Q' = (H'\Sigma^{-1}H)^{-1}H'\Sigma^{-1}, \tag{55}$$

which gives us the value of Q. Substituting (55) in Eq.(53) gives

$$P = \Sigma^{-1} - \Sigma^{-1}H(H'\Sigma^{-1}H)^{-1}H'\Sigma^{-1}. \tag{56}$$

From Eq.(50) we have

$$\Sigma^{-1}HR = -Q,$$

and multiplying this by H' and using Eq.(52) yields

$$H'\Sigma^{-1}HR = -I$$

and

$$R = -(H'\Sigma^{-1}H)^{-1}, \tag{57}$$

which determines the value of R. Since the matrix

$$\begin{bmatrix} P & Q \\ Q' & R \end{bmatrix}$$

is obviously the inverse of the matrix in the theorem, the proof is complete. ∎

We now consider the regression mode $Y = X\beta + u$, where u is distributed as $N(0, \sigma^2 I)$, subject to the restrictions $R\beta = 0$; hence this is the same model as considered before with $r = 0$. Minimize the sum of squares subject to the restrictions by forming the Lagrangian

$$L = (Y - X\beta)'(Y - X\beta) + \lambda' R\beta. \tag{58}$$

The first order conditions can be written as

$$\begin{bmatrix} (X'X)^{-1} & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'Y \\ 0 \end{bmatrix}. \tag{59}$$

Denote the matrix on the left hand side of (59) by A, and write its inverse as

$$A^{-1} = \begin{bmatrix} P & Q \\ Q' & S \end{bmatrix}. \tag{60}$$

We can then write the estimates as

$$\begin{bmatrix} \tilde{\beta} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} PX'Y \\ Q'X'Y \end{bmatrix}, \tag{61}$$

and taking expectations, we have

$$E \begin{bmatrix} \tilde{\beta} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} PX'X\beta \\ Q'X'X\beta \end{bmatrix}. \tag{62}$$

From multiplying out $A^{-1}A$ we obtain

$$PX'X + QR = I \tag{63}$$

$$Q'X'X + SR = 0 \tag{64}$$

$$PR' = 0 \tag{65}$$

$$Q'R' = I \tag{66}$$

Hence we can rewrite Eq.(62) as

$$E \begin{bmatrix} \tilde{\beta} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} (I - QR)\beta \\ -SR\beta \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}, \tag{67}$$

since $R\beta = 0$ by definition. This, so far, reproduces Corollary 3.

Theorem 29. *Given the definition in Eq.(61), the covariance matrix of $(\tilde{\beta}, \tilde{\lambda})$ is*

$$\sigma^2 \begin{bmatrix} P & 0 \\ 0 & -S \end{bmatrix}.$$

Proof. It is straightforward to note that

$$\operatorname{cov}(\tilde{\beta}, \tilde{\lambda}) = E \left[\begin{pmatrix} \tilde{\beta} \\ \tilde{\lambda} \end{pmatrix} - \begin{pmatrix} \beta \\ 0 \end{pmatrix} \right] \left[\begin{pmatrix} \tilde{\beta} \\ \tilde{\lambda} \end{pmatrix} - \begin{pmatrix} \beta \\ 0 \end{pmatrix} \right]' = \sigma^2 \begin{bmatrix} PX'XP & PX'XQ \\ QX'XP & Q'X'XQ \end{bmatrix}. \tag{68}$$

From (65) and (66), multiplying the second row of A into the first column of A^{-1} gives

$$RP = 0,$$

and multiplying it into the second column gives

$$RQ = I.$$

Hence, multiplying Eq.(63) on the right by P gives

$$PX'XP + QRP = P$$

or, since $RP = 0$,

$$PX'XP = P.$$

Multiplying Eq.(63) by Q on the right gives

$$PX'XQ + QRQ = Q,$$

or, since $RQ = I$,

$$PX'XQ = 0.$$

Finally, multiplying (64) by Q on the right gives

$$Q'X'XQ + SRQ = 0,$$

which implies that

$$Q'X'XQ = -S.$$

∎

We now do large-sample estimation for the general unconstrained and constrained cases. We wish to estimate the parameters θ of the density function $f(x, \theta)$, where x is a random variable and θ is a parameter vector with k elements. In what follows, we denote the true value of θ by θ_0. The loglikelihood is

$$\log L(x, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta). \tag{69}$$

Let $\hat{\theta}$ be the maximum likelihood estimate and let D_θ be the differential operator. Also define $I_1(\theta)$ as $\mathrm{var}(D_\theta \log f(x, \theta))$. It is immediately obvious that $\mathrm{var}(D_\theta \log L(x, \theta)) = nI_1(\theta)$. Expanding in Taylor Series about θ_0, we have

$$0 = D_\theta \log L(x, \hat{\theta}) = D_\theta \log L(x, \theta_0) + (\hat{\theta} - \theta_0) D_\theta^2 \log L(x, \theta_0) + R(x, \theta_0, \hat{\theta}) \tag{70}$$

Theorem 30. *If $\hat{\theta}$ is a consistent estimator and the third derivative of the loglikelihood function is bounded, then $\sqrt{n}(\hat{\theta}-\theta_0)$ is distributed as $N(0, I_1(\theta_0)^{-1})$.*

Proof. From Eq.(70) we have

$$\sqrt{n}(\hat{\theta}-\theta_0) = -\frac{n^{-1/2} D_\theta \log L(x,\theta_0) + n^{-1/2} R(x,\theta_0,\hat{\theta})}{n^{-1} D_\theta^2 \log L(x,\theta_0)} \tag{71}$$

where R is a remainder term of the form $(\hat{\theta}-\theta_0)^3 D_\theta^3(\log L(x,\overline{\theta})/2$, $\overline{\theta}$ being between $\hat{\theta}$ and θ_0. The quantity $n^{-1/2} D_\theta \log L(x,\theta_0)$ is a sum of n terms, each of which has expectation 0 and variance $I_1(\theta_0)$; hence by the Central Limit Theorem, $n^{-1/2} D_\theta \log L(x,\theta_0)$ is asymptotically normally distributed with mean zero and variance equal to $(1/n)nI_1(\theta_0) = I_1(\theta_0)$. The remainder term converges in probability to zero. The denominator is $1/n$ times the sum of n terms, each of which has expectation equal to $-I_1(\theta_0)$; hence the entire denominator has the same expectation and by the Weak Law of Large Numbers the denominator converges to this expectation. Hence $\sqrt{n}(\hat{\theta}-\theta_0)$ converges in distribution to a random variable which is $I_1(\theta_0)^{-1}$ times an $N(0, I_1(\theta_0))$ variable and hence is asymptotically distributed as $N(0, I_1(\theta_0)^{-1})$.

We now consider the case when there are p restrictions given by $h(\theta)' = (h_1(\theta), \ldots, h_p(\theta)) = 0$. Estimation subject to the restrictions requires forming the Lagrangian

$$G = \log L(x,\theta) - \lambda' h(\theta)$$

and setting its first partial derivatives equal to zero:

$$\begin{aligned} D_\theta \log L(x,\tilde{\theta}) - H_\theta \tilde{\lambda} &= 0 \\ h(\tilde{\theta}) \qquad\qquad &= 0 \end{aligned} \tag{72}$$

where H_θ is the $k \times p$ matrix of the derivatives of $h(\theta)$ with respect to θ. Expanding in Taylor Series and neglecting the remainder term, yields asymptotically

$$\begin{aligned} D_\theta \log L(x,\theta_0) + D_\theta^2 \log L(x,\theta_0)(\tilde{\theta}-\theta_0) - H_\theta \tilde{\lambda} &= 0 \\ H_\theta'(\tilde{\theta}-\theta_0) \qquad\qquad &= 0 \end{aligned} \tag{73}$$

The matrix H_θ should be evaluated at $\tilde{\theta}$; however, writing $H_\theta(\tilde{\theta})\tilde{\lambda} = H_\theta(\theta_0)\tilde{\lambda} + H_\theta'(\theta_0)(\tilde{\theta}-\theta_0)$ and noting that if the restrictions hold, $\tilde{\theta}$ will be near θ_0 and $\tilde{\lambda}$ will be small, we may take H_θ to be evaluated at θ_0.

Theorem 31. *The vector*

$$\begin{bmatrix} \sqrt{n}(\tilde{\theta}-\theta_0) \\ \frac{1}{\sqrt{n}}\tilde{\lambda} \end{bmatrix}$$

is asymptotically normally distributed with mean zero and covariance matrix

$$\begin{bmatrix} P & 0 \\ 0 & S \end{bmatrix}$$

where

$$\begin{bmatrix} I_1(\theta_0) & H_\theta \\ H'_\theta & 0 \end{bmatrix}^{-1} = \begin{bmatrix} P & Q \\ Q' & S \end{bmatrix}.$$

Proof. Dividing the first line of (73) by $\sqrt{n}$ and multiplying the second line by $\sqrt{n}$, we can write

$$\begin{bmatrix} -\frac{1}{n} D^2_\theta \log L(x, \theta_0) & H_\theta \\ H'_\theta & 0 \end{bmatrix} \begin{bmatrix} \sqrt{n}(\tilde{\theta} - \theta_0) \\ \frac{1}{n}\tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} D_\theta \log L(x, \theta_0) \\ 0 \end{bmatrix}. \tag{74}$$

The upper left-hand element in the left-hand matrix converges in probability to $I_1(\theta_0)$ and the top element on the right hand side converges in distribution to $N(0, I_1(\theta_0))$. Thus, (74) can be written as

$$\begin{bmatrix} I_1(\theta_0) & H_\theta \\ H'_\theta & 0 \end{bmatrix} \begin{bmatrix} \sqrt{n}(\tilde{\theta} - \theta_0) \\ \frac{1}{n}\tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} D_\theta \log L(x, \theta_0) \\ 0 \end{bmatrix}. \tag{75}$$

Eq.(75) is formally the same as Eq.(59); hence by Theorem 29,

$$\begin{bmatrix} \sqrt{n}(\tilde{\theta} - \theta_0) \\ \frac{1}{\sqrt{n}}\tilde{\lambda} \end{bmatrix}$$

is asymptotically normally distributed with mean zero and covariance matrix

$$\begin{bmatrix} P & 0 \\ 0 & S \end{bmatrix}$$

where

$$\begin{bmatrix} I_1(\theta_0) & H_\theta \\ H'_\theta & 0 \end{bmatrix}^{-1} = \begin{bmatrix} P & Q \\ Q' & S \end{bmatrix}. \tag{76}$$

∎

We now turn to the derivation of the asymptotic distribution of the likelihood ratio test statistic. As before, $\hat{\theta}$ denotes the unrestricted, and $\tilde{\theta}$ the restricted estimator.

Theorem 32. *Under the assumptions that guarantee that both the restricted and unrestricted estimators ($\tilde{\theta}$ and $\hat{\theta}$ respectively) are asymptotically normally distributed with mean zero and covariance matrices $I_1(\theta_0)$ and P respectively, and if the null hypothesis $H_0 : h(\theta) = 0$ is true, the* *likelihood ratio test statistic,* $2\log\mu = 2(\log L(x,\hat\theta) - \log L(x,\tilde\theta))$ *is asymptotically distributed as* $\chi^2(p)$.

Proof. Expand $\log L(x,\tilde\theta)$ in Taylor Series about $\hat\theta$, which yields to an approximation

$$\log L(x,\tilde\theta) = \log L(x,\hat\theta) + D_\theta \log L(x,\hat\theta)(\hat\theta-\tilde\theta) + \frac{1}{2}(\hat\theta-\tilde\theta)'\big[D_\theta^2(\log L(x,\hat\theta)\big](\hat\theta-\tilde\theta). \tag{77}$$

Since the second term on the right hand side is zero by definition, the likelihood ratio test statistic becomes

$$2\log\mu = (\hat\theta-\tilde\theta)'\big[-D_\theta^2 \log L(x,\hat\theta)\big](\hat\theta-\tilde\theta). \tag{78}$$

Let v be a k-vector distributed as $N(0, I_1(\theta_0))$. Then we can write

$$\begin{aligned}\sqrt{n}(\hat\theta-\theta_0) &= I_1(\theta_0)^{-1}v\\ \sqrt{n}(\tilde\theta-\theta_0) &= Pv\end{aligned} \tag{79}$$

where P is the same P as in Eq.(76). Then, to an approximation,

$$\begin{aligned}2\log\mu &= v'(I_1(\theta_0)^{-1}-P)'I_1(\theta_0)(I_1(\theta_0)^{-1}-P)v\\ &= v'(I_1(\theta_0)^{-1}-P-P+PI_1(\theta_0)P)v\end{aligned}. \tag{80}$$

We next show that $P = PI_1(\theta_0)P$. From Eq.(56) we can write

$$P = I_1(\theta_0)^{-1} - I_1(\theta_0)^{-1}H(H'I_1(\theta_0)^{-1}H)^{-1}H'I_1(\theta_0)^{-1}. \tag{80}$$

Multiplying this on the left by $I_1(\theta_0)$ yields

$$I_1(\theta_0)P = I - H(H'I_1(\theta_0)^{-1}H)^{-1}H'I_1(\theta_0)^{-1},$$

and multiplying this on the left by P (using the right-hand side of (81)), yields

$$\begin{aligned}PI_1(\theta_0)P = I_1(\theta_0)^{-1} &- I_1(\theta_0)^{-1}H[H'I_1(\theta_0)^{-1}H]^{-1}H'I_1(\theta_0)^{-1}\\ &- I_1(\theta_0)^{-1}H[H'I_1(\theta_0)^{-1}H]^{-1}H'I_1(\theta_0)^{-1}\\ &+ I_1(\theta_0)^{-1}H[H'I_1(\theta_0)^{-1}H]^{-1}H'I_1(\theta_0)^{-1}H[H'I_1(\theta_0)^{-1}H]^{-1}H'I_1(\theta_0)^{-1}\\ = P\end{aligned} \tag{82}$$

Hence,

$$2\log\mu = v'(I_1(\theta_0)^{-1}-P)v. \tag{83}$$

Since $I_1(\theta_0)$ is symmetric and nonsingular, it can always be written as $I_1(\theta_0) = AA'$, where A is a nonsingular matrix. Then, if z is a k-vector distributed as $N(0, I)$, we can write

$$v = Az$$

and $E(v) = 0$ and $\mathrm{cov}(v) = AA' = I_1(\theta_0)$ as required. Then

$$\begin{aligned} 2\log\mu &= z'A'(I_1(\theta_0)^{-1} - P)Az \\ &= z'A'I_1(\theta_0)^{-1}Az - z'A'PAz \\ &= z'A'(A')^{-1}A^{-1}Az - z'A'PAz\,. \\ &= z'z - z'A'PAz \\ &= z'(I - A'PA)z \end{aligned} \tag{84}$$

Now $(A'PA)^2 = A'PAA'PA = A'PI_1(\theta_0)PA$, but from Eq.(82), $P = PI_1(\theta_0)P$; hence $A'PA$ is idempotent, and its rank is clearly the rank of P. But since the k restricted estimates must satisfy p independent restrictions, the rank of P is $k - p$. Hence the rank of $I - A'PA$ is $k - (k - p) = p$. ∎

We next turn to the Wald Statistic. Expanding $h(\hat{\theta})$ in Taylor Series about θ_0 gives asymptotically

$$h(\hat{\theta}) = h(\theta_0) + H'_\theta(\hat{\theta} - \theta_0)$$

and under the null hypothesis

$$h(\hat{\theta}) = H'_\theta(\hat{\theta} - \theta_0). \tag{85}$$

Since $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically distributed as $N(0, I_1(\theta_0)^{-1})$, $\sqrt{n}h(\hat{\theta})$, which is asymptotically the same as $H'_\theta\sqrt{n}(\hat{\theta} - \theta_0)$, is asymptotically distributed as $N(0, H'_\theta I_1(\theta_0)^{-1}H_\theta)$. Then the Wald Statistic, $h(\hat{\theta})'[\mathrm{cov}(h(\hat{\theta}))]^{-1}h(\hat{\theta})$ becomes

$$W = nh(\hat{\theta})'[H'_\theta I_1(\theta_0)^{-1}H_\theta]^{-1}h(\hat{\theta}). \tag{86}$$

Theorem 33. *Under $H_0 : h(\theta) = 0$, and if H_θ is of full rank r, W is asymptotically distributed as $\chi^2(p)$.*

Proof. Let z be distributed as $N(0, I)$ and let $I_1(\theta_0)^{-1} = AA'$, where A is nonsingular. Then AZ is distributed as $N(0, I_1(\theta_0)^{-1})$, which is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. Thus, when $h(\theta) = 0$,

$$\sqrt{n}h(\hat{\theta}) = H'_\theta\sqrt{n}(\hat{\theta} - \theta_0)$$

is asymptotically distributed as $H'_\theta Az$. The Wald Statistic can be written as

$$W = z'A'H_\theta[H'_\theta I_1(\theta_0)^{-1}H_\theta]^{-1}H'_\theta Az, \tag{87}$$

which we obtain by substituting in Eq.(86) the asymptotic equivalent of $\sqrt{n}h(\hat{\theta})$. But the matrix in Eq.(87) is idempotent of rank p, since

$$\begin{aligned} A'H_\theta[H'_\theta I_1(\theta_0)^{-1}H_\theta]^{-1}H'_\theta AA'H_\theta[H'_\theta I_1(\theta_0)^{-1}H_\theta]^{-1}H'_\theta A = \\ A'H_\theta[H'_\theta I_1(\theta_0)^{-1}H_\theta]^{-1}H'_\theta A \end{aligned}$$

where we have substituted $I_1(\theta_0)^{-1}$ for AA', $I_1(\theta_0)^{-1}$ is of rank k, H_θ is of full rank p, and A is nonsingular. ∎

We next turn to the Lagrange Multiplier test. If the null hypothesis that $h(\theta) = 0$ is true, then the gradient of the loglikelihood function is likely to be small, where the appropriate metric is the inverse covariance matrix for $D_\theta \log L(x,\theta)$. Hence the Lagrange Multiplier statistic is written generally as

$$LM = [D_\theta \log L(x,\tilde{\theta})]' \big[\mathrm{cov}(D_\theta \log L(x,\tilde{\theta}))\big]^{-1} [D_\theta \log L(x,\tilde{\theta})]. \tag{88}$$

Theorem 34. *Under the null hypothesis, LM is distributed as $\chi^2(p)$.*

Proof. Expanding $D_\theta \log L(x,\tilde{\theta})$ in Taylor Series, we have asymptotically

$$D_\theta \log L(x,\tilde{\theta}) = D_\theta \log L(x,\hat{\theta}) + D^2_\theta \log L(x,\hat{\theta})(\tilde{\theta}-\hat{\theta}). \tag{89}$$

$D^2_\theta \log L(x,\hat{\theta})$ converges in probability to $-nI_1(\theta_0)$, $D_\theta \log L(x,\hat{\theta})$ converges in probability to zero, and $D_\theta \log L(x,\tilde{\theta})$ converges in probability to $-nI_1(\theta_0)(\tilde{\theta}-\hat{\theta})$. But asymptotically $\hat{\theta} = \tilde{\theta}$ under the null; hence $n^{-1/2} D_\theta \log L(x,\tilde{\theta})$ is asymptotically distributed as $N(0, I_1(\theta_0))$. Hence the appropriate test is

$$LM = n^{-1}[D_\theta \log L(x,\tilde{\theta})] I_1(\theta_0)^{-1} [D_\theta \log L(x,\tilde{\theta})]$$

which by (88) is asymptotically

$$LM = n^{-1}\big[n(\tilde{\theta}-\hat{\theta})' I_1(\theta_0) I_1(\theta_0)^{-1} I_1(\theta_0)(\tilde{\theta}-\hat{\theta}) n\big] = n(\tilde{\theta}-\hat{\theta})' I_1(\theta_0)(\tilde{\theta}-\hat{\theta}). \tag{90}$$

But this is the same as Eq.(78), the likelihood ratio statistic, since the term $-D^2_\theta \log L(x,\tilde{\theta})$ in Eq.(78) is $nI_1(\theta_0)$. Since the likelihood ratio statistic has asymptotic $\chi^2(p)$ distribution, so does the LM statistic. ∎

We now illustrate the relationship among W, LM, and LR and provide arguments for their asymptotic distributions in a slightly different way than before with a regression model $Y = X\beta + u$, with u distributed as $N(0, \sigma^2 I)$, and the restrictions $R\beta = r$.

In that case the three basic statistics are

$$\begin{aligned} W &= (R\hat{\beta}-r)'\big[R(X'X)^{-1}R'\big]^{-1}(R\hat{\beta}-r)/\hat{\sigma}^2 \\ LM &= (R\hat{\beta}-r)'\big[R(X'X)^{-1}R'\big]^{-1}(R\hat{\beta}-r)/\tilde{\sigma}^2 \\ LR &= -\frac{n}{2}(\log\hat{\sigma}^2 - \log\tilde{\sigma}^2) \end{aligned} \tag{91}$$

where W is immediate from Eq.(45) when $\hat{\Omega}$ is set equal to $\hat{\sigma}^2 I$, LM follows by substituting (40) in to (42) and setting $\tilde{\Omega} = \tilde{\sigma}^2$, and $LR = -2\log\mu$ follows by substituting $\hat{\beta}$, respectively $\tilde{\beta}$ in the likelihood function and computing $-2\log\mu$. The likelihood ratio μ itself can be written as

$$\mu = \left(\frac{\hat{u}'\hat{u}/n}{\tilde{u}'\tilde{u}/n}\right)^{n/2} = \left[\frac{1}{1+\frac{1}{n\hat{\sigma}^2}(R\hat{\beta}-r)'(R\hat{\beta}-r)}\right]^{n/2} \tag{92}$$

where we have utilized Eq.(37) by dividing both sides by S_u and taking the reciprocal. We can also rewrite the F-statistic

$$\frac{(S_r - S_u)/p}{S_u/(n-k)}$$

as

$$F = \frac{(R\hat{\beta}-r)'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta}-r)/p}{S_u/(n-k)}. \tag{93}$$

Comparing (92) and (93) yields immediately

$$\mu = \left(\frac{1}{1+\frac{p}{n-k}F}\right)^{n/2} \tag{94}$$

and comparing W in (91) with (92) yields

$$\mu = \left(\frac{1}{1+W/n}\right)^{n/2}. \tag{95}$$

Equating (94) and (95) yields

$$\frac{W}{n} = \frac{p}{n-k}F$$

or

$$W = p\left(1+\frac{k}{n-k}\right)F. \tag{96}$$

Although the left-hand side is asymptotically distributed as $\chi^2(p)$ and F has the distribution of $F(p, n-k)$, the right hand side also has asymptotic distribution $\chi^2(p)$, since the quantity $pF(p, n-k)$ converges in distribution to that χ^2 distribution.

Comparing the definitions of LM and W in (91) yields

$$LM = \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}W\right) \tag{97}$$

and from Eq.(37) we have

$$\tilde{\sigma}^2 = \hat{\sigma}^2(1+W/n). \tag{98}$$

Hence, from (97) and (98) we deduce

$$LM = \frac{W}{1+W/n}, \tag{99}$$

and using (96) we obtain

$$LM = \frac{p\left(\frac{n}{n-k}\right)F}{1+\frac{p}{n}\frac{n}{n-k}F} = \frac{npF}{n-k+pF}, \tag{100}$$

which converges in distribution as $n \to \infty$ to $\chi^2(p)$. From (95) we obtain

$$-2\log\mu = LR = n\log\left(1+\frac{W}{n}\right). \tag{101}$$

Since for positive z, $e^z > 1+z$, it follows that

$$\frac{LR}{n} = \log\left(1+\frac{W}{n}\right) < \frac{W}{n}$$

and hence $W > LR$.

We next note that for $z \geqq 0$, $\log(1+z) \geqq z/(1+z)$, since (a) at the origin the left and right hand sides are equal, and (b) at all other values of z the derivative of the left-hand side, $1/(1+z)$ is greater than the slope of the right-hand side, $1/(1+z)^2$. It follows that

$$\log\left(1+\frac{W}{n}\right) \geqq \frac{W/n}{1+W/n}.$$

Using (99) and (101), this shows that $LR \geqq LM$.

Recursive Residuals.

Since least squares residuals are correlated, even when the true errors u are not, it is inappropriate to use the least squares residuals for tests of the hypothesis that the true errors are uncorrelated. It may therefore be useful to be able to construct residuals that are uncorrelated when the true errors are. In order to develop the theory of uncorrelated residuals, we first prove a matrix theorem.

Theorem 35 (Barlett's). *If A is a nonsingular $n \times n$ matrix, if u and v are n-vectors, and if $B = A + uv'$, then*

$$B^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1+v'A^{-1}u}.$$

Proof. To show this, we verify that pre- or postmultiplying the above by b yields an identity matrix. Thus, postmultiplying yields

$$\begin{aligned} B^{-1}B = I &= \left(A^{-1} - \frac{A^{-1}uv'A^{-1}}{1+v'A^{-1}u}\right)(A+uv') \\ &= I - \frac{A^{-1}uv'}{1+v'A^{-1}u} + A^{-1}uv' - \frac{A^{-1}uv'A^{-1}uv'}{1+v'A^{-1}u} \\ &= I + \frac{-A^{-1}uv' + A^{-1}uv' + A^{-1}uv'(v'A^{-1}u) - A^{-1}u(v'A^{-1}u)v'}{1+v'A^{-1}u} \\ &= I \end{aligned} \tag{102}$$

We consider the standard regression model $Y = X\beta + u$, where u is distributed as $N(0, \sigma^2 I)$ and where X is $n \times k$ of rank k. Define X_j to represent the first j rows of the X-matrix, Y_j the first j rows of the Y-vector, x'_j the j^{th} row of X, and y_j the j^{th} element of Y. It follows from the definitions, for example, that

$$X_j = \begin{bmatrix} X_{j-1} \\ x'_j \end{bmatrix} \qquad \text{and} \qquad Y_j = \begin{bmatrix} Y_{j-1} \\ y_j \end{bmatrix}.$$

Define the regression coefficient estimate based on the first j observations as

$$\hat{\beta}_j = (X'_j X_j)^{-1} X'_j Y_j. \tag{103}$$

We then have the following

Theorem 36.

$$\hat{\beta}_j = \hat{\beta}_{j-1} + \frac{(X'_{j-1}X_{j-1})^{-1} x_j (y_j - x'_j \hat{\beta}_{j-1})}{1 + x'_j (X'_{j-1}X_{j-1})^{-1} x_j}.$$

Proof. By Theorem 35,

$$(X'_j X_j)^{-1} = (X'_{j-1}X_{j-1})^{-1} - \frac{(X'_{j-1}X_{j-1})^{-1} x_j x'_j (X'_{j-1}X_{j-1})^{-1}}{1 + x'_j (X'_{j-1}X_{j-1})^{-1} x_j}.$$

We also have by definition that

$$X_j Y_j = X_{j-1} Y_{j-1} + x_j y_j.$$

Substituting this in Eq.(103) gives

$$\begin{aligned}
\hat{\beta}_j &= \left[(X'_{j-1}X_{j-1})^{-1} - \frac{(X'_{j-1}X_{j-1})^{-1} x_j x'_j (X'_{j-1}X_{j-1})^{-1}}{1 + x'_j (X'_{j-1}X_{j-1})^{-1} x_j} \right] (X_{j-1} Y_{j-1} + x_j y_j) \\
&= \hat{\beta}_{j-1} + (X'_{j-1}X_{j-1})^{-1} x_j y_j \\
&\qquad - \frac{(X'_{j-1}X_{j-1})^{-1} x_j x'_j [(X'_{j-1}X_{j-1})^{-1} X'_{j-1} Y_{j-1}] + (X'_{j-1}X_{j-1})^{-1} x_j x'_j (X'_{j-1}X_{j-1})^{-1} x_j y_j}{1 + x'_j (X'_{j-1}X_{j-1})^{-1} x_j} \\
&= \hat{\beta}_{j-1} + \frac{(X'_{j-1}X_{j-1})^{-1} x_j (y_j - x'_j \hat{\beta}_{j-1})}{1 + x'_j (X'_{j-1}X_{j-1})^{-1} x_j},
\end{aligned}$$

where, in the second line, we bring the second and third terms on a common denominator and also note that the bracketed expression in the numerator is $\hat{\beta}_{j-1}$ by definition. ∎

First define

$$d_j = \left[1 + x'_j (X'_{j-1}X_{j-1})^{-1} x_j \right]^{1/2} \tag{104}$$

and also define the recursive residuals $\tilde{u}_j$ as

$$\tilde{u}_j = \frac{y_j - x_j'\hat{\beta}_{j-1}}{d_j} \qquad j = k+1, \ldots, n \tag{105}$$

Hence, recursive residuals are defined only when the β can be estimated from at least k observations, since for j less than $k+1$, $(X_{j-1}'X_{j-1})^{-1}$ would not be nonsingular. Hence the vector $\tilde{u}$ can be written as

$$\tilde{u} = CY, \tag{106}$$

where

$$C = \begin{bmatrix} \frac{-x_{k+1}'(X_k'X_k)^{-1}X_k'}{d_{k+1}} & \frac{1}{d_{k+1}} & 0 & 0 & \ldots & 0 \\ \frac{-x_{k+2}'(X_{k+1}'X_{k+1})^{-1}X_{k+1}'}{d_{k+2}} & & \frac{1}{d_{k+2}} & 0 & \ldots & 0 \\ \vdots & & & & \ddots & \vdots \\ \frac{-x_n'(X_{n-1}'X_{n-1})^{-1}X_{n-1}'}{d_n} & & & & & \frac{1}{d_n} \end{bmatrix}. \tag{107}$$

Since the matrix X_j' has j columns, the fractions that appear in the first column of C are rows with increasingly more columns; hence the term denoted generally by $1/d_j$ occurs in columns of the C matrix further and further to the right. Thus, the element $1/d_{k+1}$ is in column $k+1$, $1/d_{k+2}$ in column $k+2$, and so on. It is also clear that C is an $(n-k) \times n$ matrix. We then have

, **Theorem 37.** *(1) $\tilde{u}$ is linear in Y; (2) $E(\tilde{u}) = 0$; (3) The covariance matrix of the $\tilde{u}$ is scalar, i.e., $CC' = I_{n-k}$; (4) For all linear, unbiased estimators with a scalar covariance matrix, $\sum_{i=k+1}^{n} \tilde{u}_i^2 = \sum_{i=1}^{n} \hat{u}_i^2$, where $\hat{u}$ is the vector of ordinary least squares residuals.*

Proof. (1) The linearity of $\tilde{u}$ in Y is obvious from Eq.(106).

(2) It is easy to show that $CX = 0$ by multiplying Eq.(107) by X on the right. Multiplying, for example, the $(p-k)^{\text{th}}$ row of C, $(p = k+1, \ldots, n)$, into X, we obtain

$$\frac{-x_p'(X_{p-1}'X_{p-1})^{-1}X_{p-1}'}{d_p} X_{p-1} + \frac{1}{d_p} x_p' = 0.$$

It then follows that $E(\tilde{u}) = E(CY) = E(C(X\beta + u)) = E(u) = 0.$ ∎

(3) Multiplying the $(p-k)^{\text{th}}$ row of C into the $(p-k)^{\text{th}}$ column of C', we obtain

$$\frac{1}{d_p^2} + \frac{x_p'(X_{p-1}'X_{p-1})^{-1}X_{p-1}'X_{p-1}(X_{p-1}'X_{p-1})^{-1}x_p}{d_p^2} = 1$$

by definition. Multiplying the $(p-k)^{\text{th}}$ row of C into the $(s-k)^{\text{th}}$ column of C', $(s > p)$, yields

$$\begin{bmatrix} \frac{-x_p'(X_{p-1}'X_{p-1})^{-1}X_{p-1}'}{d_p} & \frac{1}{d_p} & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} X_{p-1} \\ x_p' \\ x_{p+1}' \\ \vdots \end{bmatrix} \frac{(X_{s-1}'X_{s-1})^{-1}}{d_s} .$$

$$= \left[\frac{-x_p'}{d_p d_s} + \frac{x_p'}{d_p d_s} \right] (X_{s-1}'X_{s-1})^{-1} x_s = 0$$

$\blacksquare$

(4) We first prove that $C'C = I - X(X'X)^{-1}X'$. Since $CX = 0$ by (2) of the theorem, so is $X'C'$. Define $M = I - X(X'X)^{-1}X'$; then

$$MC' = (I - X(X'X)^{-1}X')C' = C'. \tag{108}$$

Hence,

$$MC' - C' = (M - I)C' = 0. \tag{109}$$

But for any square matrix A and any eigenvalues λ of A, if $(A - \lambda I)w = 0$, then w is an eigenvector of A. Since M is idempotent, and by Theorem 5 the eigenvalues of M are all zero or 1, the columns of C' are the eigenvectors of M corresponding to the unit roots (which are $n - k$ in number, becase the trace of M is $n - k$).

Now let G' be the $n \times k$ matrix which contains the eigenvectors of M corresponding to the zero roots. Then, since M is symmetric, the matrix of all the eigenvectors of M is orthogonal and

$$[\, C' \quad G' \,] \begin{bmatrix} C \\ G \end{bmatrix} = I.$$

Let Λ denote the diagonal matrix of eigenvalues for some matrix A and let W be the matrix of its eigenvectors. Then $AW = W\Lambda$; applying this to the present case yields

$$M[\, C' \quad G' \,] = [\, C' \quad G' \,] \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} = [\, C' \quad 0 \,].$$

Hence

$$M = MI = M[\, C' \quad G' \,] \begin{bmatrix} C \\ G \end{bmatrix} = [\, C' \quad 0 \,] \begin{bmatrix} C \\ G \end{bmatrix} = C'C.$$

But

$$\sum_{i=k+1}^{n} \tilde{u}^2 = Y'C'CY = Y'[I - X(X'X)^{-1}X']Y = \sum_{i=1}^{n} \hat{u}^2.$$

$\blacksquare$

Now define S_j by $S_j = (Y_j - X_j\hat{\beta}_j)'(Y_j - X_j\hat{\beta}_j)$; thus S_j is the sum of the squares of the least squares residuals based on the first j observations. We then have

Theorem 38. $S_j = S_{j-1} + \tilde{u}_j^2$.

Proof. We can write

$$\begin{aligned}
S_j &= (Y_j - X_j\hat{\beta}_j)'(Y_j - X_j\hat{\beta}_j) = Y_j'(I - X_j(X_j'X_j)^{-1}X_j)Y_j \\
&= Y_j'Y_j - Y_j'X_j(X_j'X_j)^{-1}X_j'X_j(X_j'X_j)^{-1}X_j'Y_j \\
&\quad \text{(where we have multiplied by } X_j'X_j(X_j'X_j)^{-1}\text{)} \\
&= Y_j'Y_j - \hat{\beta}_j'X_j'X_j\hat{\beta}_j + 2\hat{\beta}_{j-1}(-X_j'Y_j + X_j'X_j\hat{\beta}_j) \\
&\quad \text{(where we replaced } (X_j'X_j)^{-1}X_j'Y_j \text{ by } \hat{\beta}_j \text{ and where the third} \\
&\quad \text{term has value equal to zero)} \\
&= Y_j'Y_j - \hat{\beta}_j'X_j'X_j\hat{\beta}_j - 2\hat{\beta}_{j-1}'X_j'Y_j + 2\hat{\beta}_{j-1}'X_j'X_j\hat{\beta}_j \\
&\quad + \hat{\beta}_{j-1}'X_j'X_j\hat{\beta}_{j-1} - \hat{\beta}_{j-1}'X_j'X_j\hat{\beta}_{j-1} \\
&\quad \text{(where we have added and subtracted the last term)} \\
&= (Y_j - X_j\hat{\beta}_{j-1})'(Y_j - X_j\hat{\beta}_{j-1}) - (\hat{\beta}_j - \hat{\beta}_{j-1})'X_j'X_j(\hat{\beta}_j - \hat{\beta}_{j-1})
\end{aligned} \quad . \tag{110}$$

Using the definition of X_j and Y_j and the definition of regression coefficient estimates, we can also write

$$\begin{aligned}
X_j'X_j\hat{\beta}_j = X_j'Y_j = X_{j-1}'Y_{j-1} + x_jy_j &= X_{j-1}'X_{j-1}\hat{\beta}_{j-1} + x_jy_j \\
&= (X_j'X_j - x_jx_j')\hat{\beta}_{j-1} + x_jy_j \\
&= X_j'X_j\hat{\beta}_{j-1} + x_j(y_j - x_j'\hat{\beta}_{j-1})
\end{aligned} \quad ,$$

and multiplying through by $(X_j'X_j)^{-1}$,

$$\hat{\beta}_j = \hat{\beta}_{j-1} + (X_j'X_j)^{-1}x_j(y_j - x_j'\hat{\beta}_{j-1}). \tag{111}$$

Substituting from Eq.(111) for $\hat{\beta}_j - \hat{\beta}_{j-1}$ in Eq.(110), we obtain

$$S_j = S_{j-1} + (y_j - x_j'\hat{\beta}_{j-1})^2 - x_j'(X_j'X_j)^{-1}x_j(y_j - x_j'\hat{\beta}_{j-1})^2. \tag{112}$$

Finally, we substitute for $(X_j'X_j)^{-1}$ in Eq.(112) from Bartlett's Identity (Theorem 35), yielding

$$\begin{aligned}
&S_j = S_{j-1} + (y_j - x_j'\hat{\beta}_{j-1})^2 \times \\
&\times \left[\frac{1 + x_j'(X_{j-1}'X_{j-1})^{-1}x_j - x_j'(X_{j-1}'X_{j-1})^{-1}x_j - (x_j'(X_{j-1}'X_{j-1})^{-1}x_j)^2 + (x_j'(X_{j-1}'X_{j-1})^{-1}x_j)^2}{1 + x_j'(X_{j-1}'X_{j-1})^{-1}x_j}\right]
\end{aligned}$$

from which the Theorem follows immediately, since $\tilde{u}_j$ is defined as

$$(y_j - x_j'\hat{\beta}_{j-1})/(1 + x_j'(X_{j-1}'X_{j-1})^{-1}x_j).$$

∎

We now briefly return to the case of testing the equality of regression coefficients in two regression in the case of insufficient degrees of freedom (i.e., the Chow Test). As in Case 4, on p. 13, the number of observations in the two data sets is n_1 and n_2 respectively. Denoting the sum of squares from the regression on the first n_1 observations by $\hat{u}_u'\hat{u}_u$ and the sum of squares using all n_1+n_2 observations by $\hat{u}_r'\hat{u}_r$, where the $\hat{u}$s are the ordinary (not recursive) least squares residuals, the test statistic can be written as

$$\frac{(\hat{u}_r'\hat{u}_r - \hat{u}_u'\hat{u}_u)/n_2}{\hat{u}_u'\hat{u}_u/(n_1-k)}.$$

By Theorem 37, this can be written as

$$\frac{(\sum_{i=k+1}^{n_1+n_2}\tilde{u}_i^2 - \sum_{i=k+1}^{n_1}\tilde{u}_i^2)/n_2}{\sum_{i=k+1}^{n_1}\tilde{u}_i^2/(n_1-k)} = \frac{\sum_{i=n_1+1}^{n_1+n_2}\tilde{u}_i^2/n_2}{\sum_{i=k+1}^{n_1}\tilde{u}_i^2/(n_1-k)}.$$

It may be noted that the numerator and denominator share no value of $\tilde{u}_i$; since the $\tilde{u}$s are independent, the numerator and denominator are independently distributed. Moreover, each $\tilde{u}_i$ has zero mean, is normally distributed and is independent of every other $\tilde{u}_j$, and has variance σ^2, since

$$\begin{aligned} E(\tilde{u}_i^2) &= E\left[\frac{(x_i'\beta - u_i - x_i'\hat{\beta}_{i-1})^2}{1+x_i'(X_{i-1}'X_{i-1})^{-1}x_i}\right] \\ &= \frac{x_i'E\big[(\beta-\hat{\beta}_{i-1})(\beta-\hat{\beta}_{i-1})'\big]x_i + E(u_i^2)}{1+x_i'(X_{i-1}'X_{i-1})^{-1}x_i} = \sigma^2. \end{aligned}$$

Hence, the ratio has an F distribution, as argued earlier.

Cusum of Squares Test. We consider a test of the hypothesis that a change in the true values of the regression coefficients occured at some observation in a series of observations. For this purpose we define

$$Q_i = \frac{\sum_{j=k+1}^{i}\tilde{u}_j^2}{\sum_{j=k+1}^{n}\tilde{u}_j^2}, \tag{113}$$

where $\tilde{u}_j$ represent the recursive residuals.

We now have

Theorem 39. *On the hypothesis that the values of the regression coefficients do not change, the random variable* $1-Q_i$ *has Beta distribution, and* $E(Q_i) = (i-k)/(n-k)$.

Proof. From Eq.(113), we can write

$$Q_i^{-1} - 1 = \frac{\sum_{j=i+1}^{n}\tilde{u}_j^2}{\sum_{j=k+1}^{i}\tilde{u}_j^2}. \tag{114}$$

Since the numerator and denominator of Eq.(114) are sums of iid normal variables with zero mean and constant variance, and since the numerator and denominator share no common $\tilde{u}_j$, the quantity

$$z = (Q_i^{-1} - 1)\frac{i-k}{n-k}$$

is distributed as $F(n-i, i-k)$. Consider the distribution of the random variable w, where W is defined by

$$(n-i)z/(i-k) = w/(1-w)$$

. Then the density of w is the Beta density

$$\frac{\Gamma(\alpha+\beta-1)}{\Gamma(\alpha)\Gamma(\beta)}\alpha^{\alpha}(1-w)^{\beta},$$

with $\alpha = -1 + (n-i)/2$ and $\beta = -1 + (i-k)/2$. It follows that

$$E(1-Q_i) = \frac{\alpha+1}{\alpha+\beta+2} = \frac{n-i}{n-k},$$

and

$$E(Q_i) = \frac{i-k}{n-k}. \tag{115}$$

Durbin (*Biometrika*, 1969, pp.1-15) provides tables for constructing confidence bands for Q_i of the form $E(Q_i) \pm c_0$.