Chapter 20: Hypothesis Testing for Linear Regression

20.1 Hypotheses of Interest

We will work with the classical assumption that, conditional on X:

$$y \sim N(X\beta, \sigma^2 I)$$

This allows the derivation of exact tests. Later, we will relax this assumption and consider asymptotic approximations. We are interested in the following types of hypotheses:

1. **Single (Linear) Hypothesis**: $c^T\beta = \gamma$, e.g., $\beta_2 = 0$ (t-test).
2. **Multiple (Linear) Hypothesis**: $R_{q\times K}\beta_{K\times 1} = r_{q\times 1}, q \leq K$, e.g., $\beta_2 = \beta_3 = \cdots = \beta_K = 0$.
3. **Single Non-linear Hypothesis**: $\beta_1^2 + \beta_2^2 + \cdots + \beta_K^2 = 1$.

Note that these are all composite hypotheses, i.e., there are nuisance parameters like σ^2 that are not specified by the null hypothesis.

Example 20.1

a. The theoretical model is the Cobb-Douglas production function $Q = AK^\alpha L^\beta$. Empirical version: take logs and add an error term to give a linear regression.

$$q = a + \alpha k + \beta l + \epsilon$$

It is often of interest whether constant returns to scale operate, i.e., would like to test whether $\alpha + \beta = 1$ is true. We may specify the alternative as $\alpha + \beta < 1$, because we can rule out increasing returns to scale.

Intuition: The Cobb-Douglas production function is widely used in economics to represent the relationship between inputs (capital K and labor L) and output Q. The coefficients α and β represent the output elasticities of capital and labor, respectively. Constant returns to scale means that if we double both capital and labor, we double output. This translates mathematically to $\alpha + \beta = 1$.

Example 20.2

b. Market efficiency

$$r_t = \mu + \gamma^T I_{t-1} + \epsilon_t$$

where r_t are returns on some asset held between period $t-1$ and t, while I_t is public information at time t. Theory predicts that $\gamma = 0$; there is no particular reason to restrict the alternative here.

Intuition: The efficient market hypothesis states that asset prices fully reflect all available information. If past information (I_{t-1}) could predict returns (r_t), then the market would not be efficient because investors could use that information to earn abnormal profits. Therefore, the coefficient γ linking past information to current returns should be zero.

Example 20.3

c. Structural change

$$y = \alpha + \beta x_t + \gamma D_t + \epsilon_t$$

where

$$D_t = \begin{cases} 0, & t < 1974 \\ 1, & t \geq 1974 \end{cases}$$

Would like to test $\gamma = 0$.

Intuition: This model allows for a change in the relationship between y and x after 1974. The dummy variable D_t captures this change. If $\gamma = 0$, then there is no structural change, meaning the relationship between y and x is the same before and after 1974.

20.2 Test of a Single Linear Hypothesis

We wish to test the hypothesis $c^T\beta = \gamma$, e.g., $\beta_2 = 0$. Suppose that $y \sim N(X\beta, \sigma^2 I)$. Then,

$$\frac{c^T\hat{\beta} - \gamma}{\sigma\sqrt{c^T(X^TX)^{-1}c}} \sim N(0,1)$$

We don't know σ and must replace it by an estimate. There are two widely used estimates:

$$\hat{\sigma}^2_{mle} = \frac{\hat{\epsilon}^T\hat{\epsilon}}{n}$$

$$s_*^2 = \frac{\hat{\epsilon}^T\hat{\epsilon}}{n-K}$$

The first estimate is the maximum likelihood estimator of σ^2, which can be easily verified. The second estimate is a modification of the MLE, which happens to be unbiased. Now define the test statistic

$$T = \frac{c^T\hat{\beta} - \gamma}{s_*\sqrt{c^T(X^TX)^{-1}c}}$$

Theorem 20.1

Suppose that A4 and H_0 hold. Then, $T \sim t(n-K)$.

Proof:

We show that:

$$\frac{(n-K)s_*^2}{\sigma^2} \sim \chi^2(n-K) \qquad (20.2)$$

$$s_*, c^T\hat{\beta} - \gamma \text{ are independent.} \qquad (20.3)$$

This establishes the theorem by the defining property of a *t*-random variable. Recall that

$$\frac{\hat{\epsilon}^T\hat{\epsilon}}{\sigma^2} = \sum_{i=1}^{n} \frac{\epsilon_i^2}{\sigma^2} \sim \chi^2(n)$$

But $\hat{\epsilon}$ are residuals that use K parameter estimates. Furthermore, $\hat{\epsilon}^T\hat{\epsilon} = \epsilon^T M_X \epsilon$ and

$$\begin{aligned} E[\epsilon^T M_X \epsilon] &= E[trM_X\epsilon\epsilon^T] = trM_X E(\epsilon\epsilon^T) \\ &= \sigma^2 trM_X = \sigma^2(n - trP_X) \\ &= \sigma^2(n-K) \\ tr(X(X^TX)^{-1}X^T) &= trX^TX(X^TX)^{-1} = trI_K = K \end{aligned}$$

These calculations show that $E\hat{\epsilon}^T\hat{\epsilon} = n - K$, which suggests that $\hat{\epsilon}^T\hat{\epsilon}$ cannot be $\chi^2(n)$ [and incidentally that $Es_*^2 = \sigma^2$]. Note that M_X is a symmetric idempotent matrix, which means that it can be written $M_X = U\Lambda U^T$, where $UU^T = I$ and Λ is a diagonal matrix of eigenvalues, which in this case are either zero (K times) or one ($n-K$ times). Furthermore, by a property of the normal distribution, $U\epsilon = \epsilon^*$ has exactly the same distribution as ϵ [it has the same mean and variance, which is sufficient to determine the normal distribution]. Therefore,

$$\hat{\epsilon}^T\hat{\epsilon} = \sum_{i=1}^{n-K} \epsilon_i^{*2} \qquad (20.4)$$

for some i.i.d. standard normal random variables ϵ_i^*. Therefore, (20.4) is $\chi^2(n-K)$ by the definition of a chi-squared random variable.

Furthermore, under H_0, $c^T\hat{\beta} - \gamma = c^T(X^TX)^{-1}X^T\epsilon$ is uncorrelated with $\hat{\epsilon} = M_X\epsilon$, since

$$E[M_X\epsilon\epsilon^TX(X^TX)^{-1}c] = \sigma^2 M_X X(X^TX)^{-1}c = 0$$

Under normality, uncorrelatedness is equivalent to independence.

We can now base the test of H_0 on

$$T = \frac{c^T\hat{\beta} - \gamma}{s_*\sqrt{c^T(X^TX)^{-1}c}}$$

using the $t(n-k)$ distribution for an exact test under normality. Can test either one-sided and two-sided alternatives, i.e., reject if $|T| \geq t_{\alpha/2}(n-K)$ [two-sided alternative] or if $T \geq t_\alpha(n-K)$ [one-sided alternative].

Above is a general rule and would require some additional computations in addition to $\hat{\beta}$. Sometimes one can avoid this: if the computer automatically prints out results of the hypothesis for $\beta_i = 0$, and one can redesign the null regression suitably. For example, suppose that

$$H_0 : \beta_2 + \beta_3 = 1$$

Substitute the restriction into the regression $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \epsilon_i$, which gives the restricted regression $y_i - z_i = \beta_1 + \beta_2(x_i - z_i) + \epsilon_i$. Now test whether $\beta_3 = 0$ in the regression $y_i - z_i = \beta_1 + \beta_2(x_i - z_i) + \beta_3 z_i + \epsilon_i$.

20.3 Test of Multiple Linear Hypothesis

We now consider a test of the multiple hypothesis $R\beta = r$. Define the quadratic form

$$F=\frac{(R\hat{\beta}-r)^T[s_*^2R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta}-r)/q}{(n-K)s_*^2/(n-K)}=\frac{(R\hat{\beta}-r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta}-r)/q}{(n-K)s_*^2/(n-K)} \tag{20.5}$$

If $y\sim N(X\beta,\sigma^2I)$, then

$$F=\frac{\chi^2(q)/q}{\chi^2(n-K)/(n-K)}\sim F(q,n-K)$$

under H_0. The rule is that if

$$F\geq F_\alpha(q,n-K)$$

then reject H_0 at level α. Note that we can only test against a two-sided alternative $R\beta\neq r$ because we have squared value in (20.5).

Example 20.4

Standard F-test, which is outputted from the computer, is of the hypothesis

$$\beta_2=0,\ldots,\beta_K=0$$

where the intercept β_1 is included in the regression but exempt from the restriction. In this case, $q=K-1$, and $H_0:R\beta=0$, where

$$R=[0_{K-1}\quad I_{K-1}]$$

The test statistic is compared with the critical value from the $F(K-1,n-K)$ distribution.

Example 20.5

Structural Change. Null hypothesis is $y=X\beta+\epsilon$. Alternative is

$$\begin{aligned} y_1&=X_1\beta_1+\epsilon_1, \quad 1\leq i\leq n_1\\ y_2&=X_2\beta_2+\epsilon_2, \quad n_1<i\leq n\end{aligned}$$

where $n=n_1+n_2$. Let

$$y=\begin{bmatrix}y_1\\y_2\end{bmatrix}, X=\begin{bmatrix}X_1&0\\0&X_2\end{bmatrix}, \beta=\begin{bmatrix}\beta_1\\\beta_{2_{2K\times1}}\end{bmatrix}, \epsilon=\begin{bmatrix}\epsilon_1\\\epsilon_{2_{n\times1}}\end{bmatrix}$$

Then, we can write the alternative regression as

$$y=X\beta+\epsilon$$

Consider the null hypothesis $H_0:\beta_1=\beta_2$. Let $R_{K\times2K}=[I_K:-I_K]$. Compare with $F(K,n-2K)$.

A confidence interval is just a critical region centered not at H_0, but at a function of parameter estimates. For example,

$$c^T\hat{\beta}\pm t_{\alpha/2}(n-K)s_*\{c^T(X^TX)^{-1}c\}^{1/2}$$

is a two-sided confidence interval for the scalar quantity $c^T\beta$. One can also construct one-sided confidence intervals and multivariate confidence intervals, which are ellipses geometrically.

20.4 Test of Multiple Linear Hypothesis Based on Fit

The idea behind the F test is that under H_0, $R\beta - r$ should be stochastically small, but under the alternative hypothesis, it will not be so. An alternative approach is based on fit. Suppose we estimate β subject to the restriction $R\beta = r$, then the sum of squared residuals from that regression should be close to that from the unconstrained regression when the null hypothesis is true [but if it is false, the two regressions will have different fitting power]. To understand this we must investigate the restricted least squares estimation procedure.

1. **Unrestricted regression:**

$$\hat{\beta} = \arg\min_b (y - Xb)^T (y - Xb)$$

and let $\hat{\epsilon} = y - X\hat{\beta}$ and $Q = \hat{\epsilon}^T \hat{\epsilon}$.

2. **Restricted regression:**

$$\beta^* = \arg\min_b (y - Xb)^T (y - Xb) \text{ s.t. } Rb = r$$

and let $\epsilon^* = y - X\beta^*$ and $Q^* = \epsilon^{*T} \epsilon^*$.

To solve the restricted least squares problem we use the Lagrangian method. We know that β^* and λ^* solve the first-order condition of the Lagrangian

$$\mathcal{L}(b, \lambda) = \frac{1}{2}(y - Xb)^T (y - Xb) + \lambda^T (Rb - r)$$

The first-order conditions are:

$$-X^T y + X^T X\beta^* + R^T \lambda^* = 0 \qquad (20.6)$$
$$R\beta^* = r \qquad (20.7)$$

Now, from (20.6), $R^T \lambda^* = X^T y - X^T X\beta^* = X^T \epsilon^*$, which implies that

$$(X^T X)^{-1} R^T \lambda^* = (X^T X)^{-1} X^T y - (X^T X)^{-1} X^T X\beta^* = \hat{\beta} - \beta^*$$

and

$$R(X^T X)^{-1} R^T \lambda^* = R\hat{\beta} - R\beta^* = R\hat{\beta} - r$$

Therefore, $\lambda^* = [R(X^T X)^{-1} R^T]^{-1} (R\hat{\beta} - r)$, and

$$\beta^* = \hat{\beta} - (X^T X)^{-1} R^T [R(X^T X)^{-1} R^T]^{-1} (R\hat{\beta} - r) \qquad (20.8)$$

This gives the restricted least squares estimator in terms of the restrictions and the unrestricted least squares estimator. From this relation, we can derive the statistical properties of the estimator β^*.

We now return to the testing question. The idea is that under H_0, $Q^* \approx Q$, but under the alternative, the two quantities differ. The following theorem makes this more precise.

Theorem 20.2

Suppose that A4 and H_0 hold. Then,

$$\frac{Q^* - Q}{q}\frac{n-K}{Q} = F \sim F(q, n-K)$$

Proof:

We show that

$$Q^* - Q = (R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$$

Then, since $s_*^2 = Q/(n-K)$ the result is established.

First, write $\beta^* = \hat{\beta} + \beta^* - \hat{\beta}$ and

$$\begin{aligned}(y - X\beta^*)^T(y - X\beta^*) &= [y - X\hat{\beta} - X(\beta^* - \hat{\beta})]^T[y - X\hat{\beta} - X(\beta^* - \hat{\beta})] \\ &= (y - X\hat{\beta})^T(y - X\hat{\beta}) + (\hat{\beta} - \beta^*)^TX^TX(\hat{\beta} - \beta^*) - (y - X\hat{\beta})^TX(\beta^* - \hat{\beta}) \\ &= \hat{\epsilon}^T\hat{\epsilon} + (\hat{\beta} - \beta^*)^TX^TX(\hat{\beta} - \beta^*)\end{aligned}$$

using the orthogonality property of the unrestricted least squares estimator. Therefore,

$$Q^* - Q = (\hat{\beta} - \beta^*)^TX^TX(\hat{\beta} - \beta^*)$$

Substituting our formulae for $\hat{\beta} - \beta^*$ and λ^* obtained above and canceling out, we get

$$Q^* - Q = (R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$$

as required.

An intermediate representation is

$$Q^* - Q = \lambda^{*T}R(X^TX)^{-1}R^T\lambda^*$$

This brings out the use of the Lagrange Multipliers in defining the test statistic and leads to the use of this name.

Importance of the result: the fit version was easier to apply in the old days, before fast computers, because one can just do two separate regressions and use the sum of squared residuals.

Example 20.6

Zero restrictions

$$\beta_2 = \cdots = \beta_K = 0$$

Then restricted regression is easy. In this case, $q = K - 1$. Note that the R^2 can be used to do an F-test of this hypothesis. We have

$$R^2 = 1 - \frac{Q}{Q^*} = \frac{Q^* - Q}{Q^*}$$

which implies that

$$F = \frac{R^2/(K-1)}{(1-R^2)/(n-k)} \qquad (20.9)$$

Example 20.7

Structural change. Allow coefficients to be different in two periods. Partition

$$y = \begin{bmatrix} y_{1_{n_1}} \\ y_{2_{n_2}} \end{bmatrix} \quad y_1 = X_1\beta_1 + \epsilon_1 \quad \text{or } y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon$$

Null is of no structural change, i.e., $H_0 : \beta_1 = \beta_2, R = (I : -I)$.

Consider the more general linear restriction

$$\begin{aligned} \beta_1 + \beta_2 - 3\beta_4 &= 1 \\ \beta_6 + \beta_1 &= 2 \end{aligned}$$

Harder to work with. Nevertheless, can always reparameterize to obtain a restricted model as a simple regression.

Example 20.8

Chow Tests: Structural change with intercepts. The unrestricted model is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1 & 0 \\ 0 & i_2 & 0 & X_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

and let $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2) \in \mathbb{R}^{2K+2}$. Different slopes and intercepts allowed. The first null hypothesis is that the slopes are the same, i.e., for some $\beta \in \mathbb{R}^K$

$$H_0 : \beta_1 = \beta_2 = \beta \qquad (20.10)$$

The restricted regression is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & 0 & X_1 \\ 0 & i_2 & X_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

The test statistic is

$$F = \frac{(\epsilon^{*T}\epsilon^* - \hat{\epsilon}^T\hat{\epsilon})/K}{\hat{\epsilon}^T\hat{\epsilon}/(n - (2K + 2))}$$

which is compared with the quantiles from the $F(K, n - 2K - 2)$ distribution.

Example 20.9

The second null hypothesis is that the intercepts are the same, i.e.,

$$H_0 : \alpha_1 = \alpha_2 = \alpha \qquad (20.11)$$

Restricted regression $(\alpha, \beta_1, \beta_2)$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} i_1 & X_1 & 0 \\ i_2 & 0 & X_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

Now suppose that $n_2 < K$. The restricted regression is ok, but the unrestricted regression runs into problems in the second period because n_2 is too small. In fact, $\hat{\epsilon}_2 = 0$. In this case, we must simply acknowledge the fact that the degrees of freedom lost are n_2 not K. Thus

$$F = \frac{(Q^* - Q)/n_2}{Q/(n_1 - K)} \sim F(n_2, n_1 - K)$$

is a valid test in this case.

20.5 Likelihood Based Testing

We have considered several different approaches which all led to the F test in linear regression. We now consider a general class of test statistics based on the Likelihood function. In principle, these apply to any parametric model, but we shall at this stage just consider its application to linear regression.

The Likelihood is denoted $L(y, X; \theta)$, where y, X are the observed data and θ is a vector of unknown parameters. The maximum likelihood estimator can be determined from $L(y, X; \theta)$, as we have already discussed. This quantity is also useful for testing. Consider again the linear restrictions

$$H_0 : R\theta = r$$

where R is of full rank q. The maximum likelihood estimator of θ is denoted by $\hat{\theta}$, while the restricted MLE is denoted by θ^*, [this is maximizes L subject to the restrictions $R\theta - r = 0$]. Now define the following test statistics:

$$LR : 2\left[\log \frac{L(\hat{\theta})}{L(\theta^*)}\right] = 2[\log L(\hat{\theta}) - \log L(\theta^*)]$$

$$\text{Wald} : (R\hat{\theta} - r)^T \left[R \frac{\partial^2 \log L}{\partial\theta\partial\theta^T}\bigg|_{\hat{\theta}}^{-1} R^T \right]^{-1} (R\hat{\theta} - r)$$

$$LM : \frac{\partial \log L}{\partial\theta}\bigg|_{\theta^*}^{T} \left[-\frac{\partial^2 \log L}{\partial\theta\partial\theta^T}\bigg|_{\theta^*} \right]^{-1} \frac{\partial \log L}{\partial\theta}\bigg|_{\theta^*}$$

The Wald test only requires computation of the unrestricted estimator, while the Lagrange Multiplier only requires computation of the restricted estimator. The Likelihood ratio requires computation of both. There are circumstances where the restricted estimator is easier to compute, and there are situations where the unrestricted estimator is easier to compute. These computational differences are what has motivated the use of either the Wald or the LM test. The LR test has certain advantages, but computationally it is the most demanding. Under the null hypothesis, we may show that

$$T \overset{D}{\to} \chi^2(q) \qquad (20.12)$$

for all three test statistics, which yields an asymptotic test. In some cases, the exact distribution of T is known and we may perform an exact test.

In the linear regression case, $\theta = (\beta, \sigma^2)$. We consider the case where the restrictions only apply to β, so that $R\beta = r$. Furthermore, we can replace the derivatives with respect to θ by derivatives with respect to β only [this requires additional justification, which we will not discuss here]. The log-likelihood is repeated here

$$\log L(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\epsilon(\beta)^T\epsilon(\beta)$$

and its derivatives are

$$\begin{aligned}
\frac{\partial \log L}{\partial \beta} &= \frac{1}{\sigma^2}X^T\epsilon(\beta) \\
\frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon(\beta)^T\epsilon(\beta) \\
\frac{\partial^2 \log L}{\partial \beta \partial \beta^T} &= -\frac{1}{\sigma^2}X^TX \\
\frac{\partial^2 \log L}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\epsilon(\beta)^T\epsilon(\beta) \\
\frac{\partial^2 \log L}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4}X^T\epsilon(\beta)
\end{aligned}$$

The Wald test is

$$W = (R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T\hat{\sigma}^2]^{-1}(R\hat{\beta} - r) = \frac{Q^* - Q}{(Q/n)} \qquad (20.13)$$

where $\hat{\sigma}^2 = Q/n$ is the MLE of σ^2, and so is very similar to the F-test apart from the use of $\hat{\sigma}^2$ instead of s_*^2 and a multiplicative factor q. In fact,

$$W = qF\frac{n}{n-k}$$

This is approximately equal to qF when the sample size is large. The Lagrange Multiplier or Score or Rao test statistic is

$$LM = \frac{\epsilon^{*T}X}{\sigma^{*2}}\left[\frac{X^TX}{\sigma^{*2}}\right]^{-1}\frac{X^T\epsilon^*}{\sigma^{*2}} \qquad (20.14)$$

where $\sigma^{*2} = Q^*/n$. This can be rewritten in the form

$$LM = \frac{\lambda^{*T}R(X^TX)^{-1}R^T\lambda^*}{\sigma^{*2}}$$

where λ^* is the vector of Lagrange Multipliers evaluated at the optimum [Recall that $X^T\epsilon^* = R\lambda^*$]. Furthermore, we can write the score test as

$$LM = \frac{Q^* - Q}{(Q^*/n)} = n\left(1 - \frac{Q}{Q^*}\right)$$

When the restrictions are the standard zero ones, the test statistic is n times the R^2 from the unrestricted regression. The unrestricted and restricted Likelihoods are

$$
\begin{aligned}
\log L(\hat{\beta}, \hat{\sigma}^2) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2}\hat{\epsilon}^T\hat{\epsilon} \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \frac{\hat{\epsilon}^T\hat{\epsilon}}{n} - \frac{n}{2} \\
\log L(\beta^*, \sigma^{*2}) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^{*2} - \frac{1}{2\sigma^{*2}}\epsilon^{*T}\epsilon^* \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \frac{\epsilon^{*T}\epsilon^*}{n} - \frac{n}{2}
\end{aligned}
$$

These two lines follow because $\hat{\sigma}^2 = \hat{\epsilon}^T\hat{\epsilon}/n$ and $\sigma^{*2} = \epsilon^{*T}\epsilon^*/n$. Therefore,

$$
LR = 2\log\frac{L(\hat{\beta}, \hat{\sigma}^2)}{L(\beta^*, \sigma^{*2})} = n\left[\log\frac{Q^*}{n} - \log\frac{Q}{n}\right] = n[\log Q^* - \log Q]
$$

Note that W, LM, and LR are all monotonic functions of F, in fact

$$
W = F\frac{qn}{n-k}, \quad LM = \frac{W}{1+W/n}, \quad LR = n\log\left(1+\frac{W}{n}\right) \tag{20.15}
$$

If we knew the exact distribution of any of them we can obtain the distributions of the others and the test result will be the same. However, in practice one uses asymptotic critical values, which lead to differences in outcomes. We have

$$
LM \leq LR \leq W
$$

so that the Wald test will reject more frequently than the LR test and the LM tests, supposing that the same critical values are used.

20.6 Bayesian Approach

In the Bayesian approach, we require a prior distribution for the unknown parameters β, σ^2. Combining the prior and the likelihood we obtain the posterior for β, σ^2 from which we can do inference. We just consider the simplest case where we treat σ^2 as known for which the algebra is simple. We have the following result.

Theorem 20.3

Suppose that A4 holds with σ^2 known, and that the prior for β is $N(\beta_0, \Sigma_0)$ for some vector β_0 and covariance matrix Σ_0. The posterior distribution of $\beta|y, X$ is $N(b, \Omega)$ with

$$
\begin{aligned}
b &= \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}X^Ty + \Sigma_0^{-1}\beta_0\right) \\
\Omega &= \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}
\end{aligned}
$$

This result allows one to provide a Bayesian confidence interval or hypothesis test by just inverting the posterior distribution. For example, for $c^T\beta$, the posterior is normal with mean c^Tb and variance $c^T\Omega c$ and so the Bayesian confidence interval is $c^Tb \pm z_{\alpha/2}\sqrt{c^T\Omega c}$, which has Bayesian coverage probability $1-\alpha$.

It is possible to derive explicit results also for the case that σ^2 is unknown and has a prior distribution such as a Gamma distribution on $\mathbb{R}^+$, however, the algebra is more complicated. It is not clear where the normal prior comes from, nor for example why β_0, Σ_0 are themselves known and not subject to uncertainty like β, σ^2. It is possible to further “priorize” these quantities, but at some point one has to take something we might call a parameter as a fixed known quantity.

Exercises

Exercise 1

Solution 1

Consider the Cobb-Douglas production function $Q = AK^\alpha L^\beta$. Suppose you have collected data on output, capital, and labor for a sample of firms. Explain how you would test the hypothesis of constant returns to scale using a linear regression. Specifically, state the null and alternative hypotheses, and describe the regression you would run.

Exercise 2

Solution 2

In the context of market efficiency, you are given the following model:

$$r_t = \mu + \gamma^T I_{t-1} + \epsilon_t$$

where r_t represents stock returns, and I_{t-1} is a vector of publicly available information variables at time $t-1$. Explain what the efficient market hypothesis implies about the value of γ. How would you test this hypothesis?

Exercise 3

Solution 3

Consider the model with a structural change:

$$y_t = \alpha + \beta x_t + \gamma D_t + \epsilon_t$$

where $D_t = 1$ if $t \geq T$ and $D_t = 0$ if $t < T$. Explain in words what the null hypothesis $\gamma = 0$ means in this context.

Exercise 4

Solution 4

Derive the formula for the t-statistic for testing the hypothesis $c^T\beta = \gamma$ in the classical linear regression model $y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$. Start from the distribution of $c^T\hat{\beta}$.

Exercise 5

Solution 5

Explain why $s_*^2 = \frac{\hat{\epsilon}^T\hat{\epsilon}}{n-K}$ is an unbiased estimator of σ^2 in the classical linear regression model.

Exercise 6

Solution 6

Describe the steps involved in constructing a (1-α)% confidence interval for a single linear combination of regression coefficients, $c^T\beta$.

Exercise 7

Solution 7

Explain, in intuitive terms, why the F-statistic for testing multiple linear restrictions follows an F-distribution under the null hypothesis.

Exercise 8

Solution 8

Given a linear regression model, suppose you want to test the joint hypothesis that $\beta_2 = 0$ and $\beta_3 + \beta_4 = 1$. Write down the matrix R and the vector r for this test, such that the null hypothesis can be expressed as $R\beta = r$.

Exercise 9

Solution 9

Explain the difference between the restricted least squares estimator, β^*, and the unrestricted least squares estimator, $\hat{\beta}$, in the context of testing linear hypotheses.

Exercise 10

Solution 10

Derive the restricted least squares estimator, β^*, by minimizing the sum of squared errors subject to the linear constraint $R\beta = r$, using the method of Lagrange multipliers.

Exercise 11

Solution 11

Show that the difference between the restricted and unrestricted sum of squared residuals, $Q^* - Q$, can be expressed as $(R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$.

Exercise 12

Solution 12

Explain how the R^2 value from a regression can be used to conduct an F-test for the hypothesis that all slope coefficients are zero.

Exercise 13

Solution 13

In the context of structural change (Example 20.7), explain the null hypothesis being tested and how to set up the restricted regression under the null.

Exercise 14

[Solution 14](#)

Describe the difference between the Wald, Likelihood Ratio (LR), and Lagrange Multiplier (LM) tests. What are the computational advantages of each?

Exercise 15

[Solution 15](#)

For the linear regression model, write down the log-likelihood function and its first derivative with respect to β.

Exercise 16

[Solution 16](#)

Express the Wald statistic for testing $R\beta = r$ in terms of the difference in restricted and unrestricted sum of squared residuals, $Q^* - Q$ and other relevant quantities.

Exercise 17

[Solution 17](#)

Express the Lagrange Multiplier (LM) statistic in terms of Q and Q^*.

Exercise 18

[Solution 18](#)

Explain the relationship between the F-statistic and the Wald, LM, and LR statistics. Under what conditions are they approximately equal?

Exercise 19

[Solution 19](#)

What are the general steps in a Bayesian approach to inference in the linear regression model?

Exercise 20

[Solution 20](#)

If the prior for β is $N(\beta_0, \Sigma_0)$ and the likelihood is based on the classical linear regression model with known σ^2 , what is the posterior distribution of β?

Solutions

Solution 1

[Exercise 1](#)

Taking logs of the Cobb-Douglas production function, we get:

$$\log(Q) = \log(A) + \alpha \log(K) + \beta \log(L)$$

Let $q = \log(Q)$, $a = \log(A)$, $k = \log(K)$, and $l = \log(L)$. We can then write a linear regression model:

$$q_i = a + \alpha k_i + \beta l_i + \epsilon_i$$

Constant returns to scale implies that if we multiply both inputs by a constant, output is multiplied by the same constant. Mathematically, this means $\alpha + \beta = 1$.

- **Null Hypothesis (H_0)**: $\alpha + \beta = 1$
- **Alternative Hypothesis (H_1)**: $\alpha + \beta \neq 1$ (or $\alpha + \beta < 1$ if we rule out increasing returns to scale)

To test this, we would run the regression above and use a t-test or an F-test to test the restriction on the coefficients. We are testing a **single linear hypothesis** here. This relates to Section 20.1 of the text, where hypotheses of interest for linear regression are introduced.

Solution 2

Exercise 2

The **efficient market hypothesis** states that asset prices fully reflect all available information. This implies that past information, I_{t-1}, should not be able to predict future returns, r_t, beyond the average return, μ. Therefore, the efficient market hypothesis implies that $\gamma = 0$.

To test this hypothesis, we would run the given regression and perform a t-test (or a Wald test if γ is a vector) on the coefficient vector γ.

- **Null Hypothesis (H_0)**: $\gamma = 0$
- **Alternative Hypothesis (H_1)**: $\gamma \neq 0$

This test relates to Section 20.1 of the text as an example of a hypothesis of interest and to section 20.2, which is about testing a **single linear hypothesis**.

Solution 3

Exercise 3

The null hypothesis $\gamma = 0$ means that there is **no structural change** in the relationship between y_t and x_t at time T. In other words, the relationship between y_t and x_t is the same before and after time T, and is fully captured by α and β. If $\gamma \neq 0$, there *is* a structural break at time T. This test is an example of testing a **single linear hypothesis**.

Solution 4

Exercise 4

In the classical linear regression model, the OLS estimator $\hat{\beta}$ is normally distributed:

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

Therefore, any linear combination of $\hat{\beta}$ is also normally distributed:

$$c^T\hat{\beta} \sim N(c^T\beta, \sigma^2 c^T(X^TX)^{-1}c)$$

Subtracting the hypothesized value $c^T\beta = \gamma$ and dividing by the standard error, we get a standard normal variable under the null hypothesis:

$$\frac{c^T\hat{\beta} - \gamma}{\sqrt{\sigma^2 c^T(X^TX)^{-1}c}} \sim N(0,1)$$

However, we don't know σ^2, so we replace it with its unbiased estimator $s_*^2 = \frac{\hat{\epsilon}^T\hat{\epsilon}}{n-K}$. This gives us the t-statistic:

$$t = \frac{c^T\hat{\beta} - \gamma}{\sqrt{s_*^2 c^T(X^TX)^{-1}c}} = \frac{c^T\hat{\beta} - \gamma}{s_*\sqrt{c^T(X^TX)^{-1}c}} \sim t(n-K)$$

This derivation follows the steps outlined in Section 20.2 of the text. The resulting t-statistic follows a t-distribution with $n-K$ degrees of freedom under the null hypothesis.

Solution 5

Exercise 5

The residuals are given by $\hat{\epsilon} = y - X\hat{\beta} = y - X(X^TX)^{-1}X^Ty = (I - X(X^TX)^{-1}X^T)y = M_Xy = M_X\epsilon$. The sum of squared residuals is:

$$\begin{aligned}\hat{\epsilon}^T\hat{\epsilon} &= \epsilon^T M_X^T M_X \epsilon = \epsilon^T M_X \epsilon = tr(\epsilon^T M_X \epsilon) \\ &= tr(M_X \epsilon\epsilon^T)\end{aligned}$$

Taking expectations:

$$\begin{aligned}E[\hat{\epsilon}^T\hat{\epsilon}] &= E[tr(M_X\epsilon\epsilon^T)] = tr(M_X E[\epsilon\epsilon^T]) = tr(M_X\sigma^2 I) \\ &= \sigma^2 tr(M_X) = \sigma^2 tr(I_n - X(X^TX)^{-1}X^T) \\ &= \sigma^2(tr(I_n) - tr(X(X^TX)^{-1}X^T)) = \sigma^2(n - tr((X^TX)^{-1}X^TX)) \\ &= \sigma^2(n - tr(I_K)) = \sigma^2(n-K)\end{aligned}$$

Therefore, $E[\hat{\epsilon}^T\hat{\epsilon}] = (n-K)\sigma^2$. To obtain an unbiased estimator, we divide by $(n-K)$:

$$E[s_*^2] = E\left[\frac{\hat{\epsilon}^T\hat{\epsilon}}{n-K}\right] = \frac{1}{n-K}E[\hat{\epsilon}^T\hat{\epsilon}] = \frac{1}{n-K}(n-K)\sigma^2 = \sigma^2$$

This proof uses the properties of the trace operator and the idempotent matrix M_X, as shown in Section 20.2.

Solution 6

Exercise 6

1. **Estimate the regression model:** Obtain the OLS estimator $\hat{\beta}$ and the unbiased variance estimator $s_*^2 = \frac{\hat{\epsilon}^T\hat{\epsilon}}{n-K}$.

2. **Calculate the standard error:** The standard error of $c^T\hat{\beta}$ is $s.e.(c^T\hat{\beta}) = s_*\sqrt{c^T(X^TX)^{-1}c}$.

3. **Find the critical value:** Find the critical value $t_{\alpha/2}(n-K)$ from the t-distribution with $n-K$ degrees of freedom, corresponding to the desired confidence level (1-α).

4. **Construct the confidence interval:** The (1-α)% confidence interval is given by:

$$c^T\hat{\beta} \pm t_{\alpha/2}(n-K) \cdot s_*\sqrt{c^T(X^TX)^{-1}c}$$

This process is described at the end of Section 20.2 and the beginning of 20.3.

Solution 7

Exercise 7

The F-statistic is the ratio of two independent chi-squared random variables, each divided by their respective degrees of freedom. Under the null hypothesis, the numerator of the F-statistic, which represents the variation explained by the restrictions, follows a chi-squared distribution with degrees of freedom equal to the number of restrictions (q). The denominator, which represents the unexplained variation, follows a chi-squared distribution with degrees of freedom equal to $n-K$. The ratio of these two independent χ^2 variables, each divided by their respective degrees of freedom, defines the F-distribution. This relates to section 20.3 of the text, and the definition of F-statistic shown in equation (20.5).

Solution 8

Exercise 8

We can write the null hypothesis as a system of two linear equations:

$$\begin{aligned} \beta_2 &= 0 \\ \beta_3 + \beta_4 &= 1 \end{aligned}$$

In matrix form, $R\beta = r$, this becomes:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

So,

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 \end{bmatrix}, \quad r = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

This example illustrates how to express **multiple linear hypotheses** in matrix form, as discussed in Section 20.3.

Solution 9

Exercise 9

The **unrestricted least squares estimator**, $\hat{\beta}$, minimizes the sum of squared residuals without any constraints on the coefficients. The **restricted least squares estimator**, β^*, minimizes the sum of squared residuals *subject to* the linear constraints imposed by the null hypothesis, $R\beta = r$. If the null hypothesis is true, β^* will be close to $\hat{\beta}$. If the null hypothesis is false, β^* will be significantly different from $\hat{\beta}$. This corresponds to the approach of testing a **multiple linear hypothesis based on fit** described in section 20.4.

Solution 10

Exercise 10

We want to minimize $(y - Xb)^T(y - Xb)$ subject to $Rb = r$. Set up the Lagrangian:

$$\mathcal{L}(b, \lambda) = \frac{1}{2}(y - Xb)^T(y - Xb) + \lambda^T(Rb - r)$$

Take the derivatives with respect to b and λ and set them equal to zero:

$$\frac{\partial \mathcal{L}}{\partial b} = -X^T(y - Xb) + R^T\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = Rb - r = 0$$

From the first equation:

$$X^TXb = X^Ty - R^T\lambda$$

$$b = (X^TX)^{-1}X^Ty - (X^TX)^{-1}R^T\lambda = \hat{\beta} - (X^TX)^{-1}R^T\lambda$$

This b is β^*. Multiply by R:

$$R\beta^* = R\hat{\beta} - R(X^TX)^{-1}R^T\lambda$$

Since $R\beta^* = r$:

$$r = R\hat{\beta} - R(X^TX)^{-1}R^T\lambda$$

$$R(X^TX)^{-1}R^T\lambda = R\hat{\beta} - r$$

$$\lambda = [R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$$

Substituting back into the expression for $b = \beta^*$:

$$\beta^* = \hat{\beta} - (X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$$

This is the same derivation as in Section 20.4, equations (20.6), (20.7) and (20.8).

Solution 11

Exercise 11

We know that $Q = (y - X\hat{\beta})^T(y - X\hat{\beta})$ and $Q^* = (y - X\beta^*)^T(y - X\beta^*)$. From Solution 10, we also know $\beta^* = \hat{\beta} - (X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$. Let $A = (X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}$. Then

$\beta^* = \hat{\beta} - A(R\hat{\beta} - r)$.

$$\begin{aligned}
Q^* &= (y - X\beta^*)^T(y - X\beta^*) \\
&= (y - X(\hat{\beta} - A(R\hat{\beta} - r)))^T(y - X(\hat{\beta} - A(R\hat{\beta} - r))) \\
&= (y - X\hat{\beta} + XA(R\hat{\beta} - r))^T(y - X\hat{\beta} + XA(R\hat{\beta} - r)) \\
&= ((y - X\hat{\beta}) + XA(R\hat{\beta} - r))^T((y - X\hat{\beta}) + XA(R\hat{\beta} - r)) \\
&= (y - X\hat{\beta})^T(y - X\hat{\beta}) + 2(R\hat{\beta} - r)^T A^T X^T(y - X\hat{\beta}) + (R\hat{\beta} - r)^T A^T X^T XA(R\hat{\beta} - r)
\end{aligned}$$

Since $y - X\hat{\beta}$ is the vector of OLS residuals, it is orthogonal to the columns of X, so $X^T(y - X\hat{\beta}) = 0$. Thus,

$$Q^* = Q + (R\hat{\beta} - r)^T A^T X^T XA(R\hat{\beta} - r)$$

$$Q^* - Q = (R\hat{\beta} - r)^T A^T X^T XA(R\hat{\beta} - r)$$

Now, substitute the expression for A:

$$A^T X^T XA = [R(X^TX)^{-1}R^T]^{-1}R(X^TX)^{-1}X^TX(X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}$$

$$= [R(X^TX)^{-1}R^T]^{-1}R(X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1} = [R(X^TX)^{-1}R^T]^{-1}$$

So, $Q^* - Q = (R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$. This is the derivation shown in Section 20.4.

Solution 12

Exercise 12

The R^2 value measures the proportion of the total variation in y that is explained by the regression model. The F-test for the hypothesis that all slope coefficients are zero tests whether the model explains a significant portion of the variation in y.

The F-statistic for this test is given by:

$$F = \frac{R^2/(K-1)}{(1-R^2)/(n-K)}$$

where K is the total number of coefficients (including the intercept) and n is the number of observations. This F-statistic follows an $F(K-1, n-K)$ distribution under the null hypothesis. This corresponds to Example 20.6 and equation (20.9).

Solution 13

Exercise 13

The null hypothesis is that there is **no structural change**, meaning that the coefficients are the same for both subsamples: $\beta_1 = \beta_2$.

The unrestricted model is:

$$y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon$$

The restricted model, under the null hypothesis, imposes the constraint $\beta_1 = \beta_2 = \beta$. The restricted model is:

$$y = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \epsilon = X\beta + \epsilon$$

Where $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$.

Solution 14

Exercise 14

- **Wald Test:** Uses the **unrestricted** estimator $\hat{\theta}$. It measures the "distance" between $R\hat{\theta}$ and r. Its computational advantage is that it only requires estimating the unrestricted model.
- **Likelihood Ratio (LR) Test:** Compares the likelihood of the **unrestricted** model, $L(\hat{\theta})$, to the likelihood of the **restricted** model, $L(\theta^*)$. Its advantage is that it often has better small sample properties but requires estimating both the restricted and unrestricted models.
- **Lagrange Multiplier (LM) Test:** Uses the **restricted** estimator θ^*. It examines the slope of the log-likelihood function at the restricted estimator. Its computational advantage is that it only needs the restricted estimator. This is particularly useful when the restricted model is easier to estimate. This is covered in section 20.5.

Solution 15

Exercise 15

The log-likelihood function for the linear regression model $y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$, is:

$$\log L(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)$$

The first derivative with respect to β is:

$$\frac{\partial \log L}{\partial \beta} = -\frac{1}{2\sigma^2}[-2X^Ty + 2X^TX\beta] = \frac{1}{\sigma^2}X^T(y - X\beta) = \frac{1}{\sigma^2}X^T\epsilon$$

This is consistent with the content in Section 20.5.

Solution 16

Exercise 16

The Wald statistic is given by:

$$W = (R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T\hat{\sigma}^2]^{-1}(R\hat{\beta} - r)$$

where $\hat{\sigma}^2 = \frac{Q}{n}$ is the MLE of σ^2. We know from Exercise 11 that $Q^* - Q = (R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$. Therefore,

$$W = \frac{(R\hat{\beta} - r)^T [R(X^T X)^{-1} R^T]^{-1} (R\hat{\beta} - r)}{\hat{\sigma}^2} = \frac{Q^* - Q}{Q/n} = n\frac{Q^* - Q}{Q}$$

This corresponds to equation (20.13).

Solution 17

Exercise 17

The LM statistic can be written as (See solution to Exercise 16 for definition of Q and Q^*):

$$LM = n\frac{Q^* - Q}{Q^*}$$

This expression is derived in Section 20.5.

Solution 18

Exercise 18

The F-statistic is:

$$F = \frac{(Q^* - Q)/q}{Q/(n-K)}$$

The Wald statistic is $W = n\frac{Q^*-Q}{Q}$, the LM statistic is $LM = n\frac{Q^*-Q}{Q^*}$, and the Likelihood Ratio statistic is $LR = n \log \frac{Q^*}{Q}$. The relationships are

$$W = qF\frac{n}{n-k}, \quad LM = \frac{W}{1 + \frac{W}{n}}, \quad LR = n \log\left(1 + \frac{W}{n}\right)$$

As $n \to \infty$, $W/(n-K) \to 0$. Then $\dfrac{n}{n-k} \approx 1$, and we can use the approximation $\log(1+x) \approx x$ for small x to get $LM \approx LR \approx W \approx qF$. This corresponds to equation (20.15).

Solution 19

Exercise 19

1. **Specify a prior distribution:** Define a prior distribution for the parameters, $p(\beta, \sigma^2)$.
2. **Formulate the likelihood function:** Write down the likelihood function of the data, $L(y|X, \beta, \sigma^2)$.
3. **Calculate the posterior distribution:** Use Bayes' theorem to combine the prior and the likelihood to obtain the posterior distribution, $p(\beta, \sigma^2|y, X)$.
4. **Make inferences**: Base inferences (e.g., point estimates, confidence intervals, hypothesis tests) on the posterior distribution.

This is described briefly in section 20.6.

Solution 20

Exercise 20

The posterior distribution of β is also normal:

$$\beta|y, X \sim N(b, \Omega)$$

where

$$b = \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}X^Ty + \Sigma_0^{-1}\beta_0\right)$$
$$\Omega = \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}$$

This is stated in Theorem 20.3.

R scripts

R Script 1: Testing for Constant Returns to Scale in a Cobb-Douglas Production Function

```
# Load necessary libraries
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.2
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(car)
```

```
Loading required package: carData

Attaching package: 'car'

The following object is masked from 'package:dplyr':

    recode

The following object is masked from 'package:purrr':

    some
```

```
# Simulate data for a Cobb-Douglas production function
set.seed(123)
n <- 100
A <- 1
alpha <- 0.6
beta <- 0.4  # Set parameters such that alpha + beta = 1 (constant returns to scale)
K <- runif(n, 1, 10)
L <- runif(n, 1, 10)
epsilon <- rnorm(n, 0, 0.5)
Q <- A * K^alpha * L^beta * exp(epsilon)

# Take logs
```

```
q <- log(Q)
k <- log(K)
l <- log(L)

# Create a data frame
df <- data.frame(q, k, l)

# Estimate the linear regression model
model <- lm(q ~ k + l, data = df)

# Test the hypothesis of constant returns to scale (alpha + beta = 1)
# Using a linear restriction, we create the restriction matrix R and vector r
R <- matrix(c(0, 1, 1), nrow = 1)  # We are testing the restriction on coefficients of k and l.
r <- 1
car::linearHypothesis(model, R, r) #Ho: R*beta = r
```

```
Linear hypothesis test:
k  + l = 1

Model 1: restricted model
Model 2: q ~ k + l

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     98 23.143
2     97 23.108  1  0.034859 0.1463 0.7029
```

```
# Alternatively we can use
# car::linearHypothesis(model, "k + l = 1")


# # Visualize the data and the fitted plane in 3D (optional)
# # install.packages("plotly") # Uncomment to install plotly
# library(plotly)
# vis_3d <- plot_ly(df, x = ~k, y = ~l, z = ~q, type = "scatter3d", mode = "markers", name = "Data")
         %>%
#   add_trace(x = ~k, y = ~l, z = fitted(model), type = "mesh3d", name = "Fitted Plane")
```

Explanation:

1. **Simulate Data:** We simulate data from a Cobb-Douglas production function with constant returns to scale ($\alpha + \beta = 1$). We generate random values for capital (`K`), labor (`L`), and an error term (`epsilon`).

2. **Log Transformation:** We take the natural logarithm of the production function to linearize it: $q = \log(A) + \alpha k + \beta l + \epsilon$.

3. **Linear Regression:** We estimate the linear regression model `q ~ k + l` using `lm()`.

4. **Hypothesis Test:** We use the `linearHypothesis()` function from the `lmtest` package to test the null hypothesis $H_0 : \alpha + \beta = 1$. The `R` matrix and `r` vector represent the linear restriction. The function returns an F-statistic and p-value.

5. **(Optional) 3D Visualization:** We use the `plotly` package to create an interactive 3D plot showing the simulated data points and the fitted regression plane.

Connection to the text: This script directly implements Example 20.1(a) from the text, demonstrating how to test for constant returns to scale using a linear regression and hypothesis test. We are testing a **single linear hypothesis**.

R Script 2: Testing for Market Efficiency

```
# Load necessary libraries
library(tidyverse)
library(lmtest)
```

```
Loading required package: zoo


Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

```
# Simulate data for a market efficiency test
set.seed(456)
n <- 200
mu <- 0.01  # Average return
gamma <- 0   # Set gamma = 0 to simulate under the null hypothesis
It_minus_1 <- rnorm(n) # Simulate public information
epsilon <- rnorm(n, 0, 0.02)
rt <- mu + gamma * It_minus_1 + epsilon

# Create data frame
df <- data.frame(rt, It_minus_1)

# Estimate the linear regression model
model <- lm(rt ~ It_minus_1, data = df)

# Test the hypothesis of market efficiency (gamma = 0)
coeftest(model)  # t-test for gamma = 0
```

```

t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0123360  0.0013092  9.4228   <2e-16 ***
It_minus_1  0.0018201  0.0013239  1.3747   0.1708
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Visualize the relationship (or lack thereof)
ggplot(df, aes(x = It_minus_1, y = rt)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Market Efficiency Test Simulation",
       x = "Public Information (t-1)",
       y = "Return (t)")
```

```
`geom_smooth()` using formula = 'y ~ x'
```

Explanation:

1. **Simulate Data:** We simulate data under the null hypothesis of market efficiency ($\gamma = 0$). We generate random returns (`rt`), public information (`It_minus_1`), and an error term (`epsilon`).

2. **Linear Regression:** We estimate the linear regression `rt ~ It_minus_1`.

3. **Hypothesis Test:** We use `coeftest()` from `lmtest` to obtain the t-test for the coefficient of `It_minus_1` (which is γ). The output provides the t-statistic, p-value, and other relevant information.
4. **Visualization:** We create a scatter plot of `rt` against `It_minus_1` and add a regression line. Under the null hypothesis, we expect no clear relationship.

Connection to the text: This script implements Example 20.2(b), demonstrating the test for market efficiency. We are testing a **single linear hypothesis**, as in Section 20.2.

R Script 3: Testing for Structural Change

```
# Load necessary libraries
library(tidyverse)
library(lmtest)

# Simulate data with a structural change
set.seed(789)
n <- 150
alpha <- 2
beta <- 1
gamma <- 2  # Significant structural change (gamma != 0)
T_change <- 75
xt <- runif(n, 0, 5)
Dt <- ifelse(1:n >= T_change, 1, 0)
epsilon <- rnorm(n, 0, 1)
yt <- alpha + beta * xt + gamma * Dt + epsilon

# Create data frame
df <- data.frame(yt, xt, Dt)

# Estimate the linear regression model
model <- lm(yt ~ xt + Dt, data = df)

# Test for structural change (gamma = 0)
summary(model)  # Look at the t-test for the coefficient of Dt
```

```

Call:
lm(formula = yt ~ xt + Dt, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0109 -0.6494  0.0864  0.6415  3.1698 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.12835    0.18755   11.35   <2e-16 ***
xt           0.93244    0.06375   14.63   <2e-16 ***
Dt           1.86220    0.17599   10.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.066 on 147 degrees of freedom
Multiple R-squared:  0.7204,    Adjusted R-squared:  0.7166 
F-statistic: 189.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

```
# Visualize the structural change
ggplot(df, aes(x = xt, y = yt, color = factor(Dt))) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
```

```
  labs(title = "Structural Change Test Simulation",
       x = "x",
       y = "y",
       color = "Period")
```

```
`geom_smooth()` using formula = 'y ~ x'
```

Explanation:

1. **Simulate Data:** We simulate data *with* a structural change ($\gamma \neq 0$). We create a dummy variable `Dt` that indicates whether the observation is before or after the structural change point (`T_change`).

2. **Linear Regression:** We estimate the model `yt ~ xt + Dt`.

3. **Hypothesis Test:** We examine the t-test for the coefficient of `Dt` in the `summary(model)` output. A significant t-statistic suggests evidence of a structural change.

4. **Visualization:** We create a scatter plot of `yt` against `xt`, coloring the points by `Dt`. This visually highlights the different relationships before and after the structural change.

Connection to the text: This script illustrates Example 20.3(c) on testing for structural change. We are testing a **single linear hypothesis**.

R Script 4: F-test for Multiple Linear Restrictions

```
# Load necessary libraries
library(tidyverse)
library(car)

# Simulate data
set.seed(1011)
n <- 100
x1 <- rnorm(n)
x2 <- rnorm(n)
x3 <- rnorm(n)
beta0 <- 1
beta1 <- 0.5
beta2 <- 0 # restricting to zero under the null
beta3 <- 0 # restricting to zero under the null
epsilon <- rnorm(n, 0, 1)
y <- beta0 + beta1 * x1 + beta2 * x2 + beta3 * x3 + epsilon

# Create data frame
df <- data.frame(y, x1, x2, x3)

# Estimate the full model
full_model <- lm(y ~ x1 + x2 + x3, data = df)

# Estimate the restricted model (beta2 = beta3 = 0)
restricted_model <- lm(y ~ x1, data = df)

# Perform the F-test
anova(restricted_model, full_model) # Comparing restricted and unrestricted models.
```

```
Analysis of Variance Table

Model 1: y ~ x1
```

```
Model 2: y ~ x1 + x2 + x3
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     98 115.05                           
2     96 113.73  2    1.3188 0.5566  0.575
```

```
#Alternatively: using linearHypothesis()
R = rbind(c(0,0,1,0),
          c(0,0,0,1))
r = c(0,0)

linearHypothesis(full_model, R, r)
```

```
Linear hypothesis test:
x2 = 0
x3 = 0

Model 1: restricted model
Model 2: y ~ x1 + x2 + x3

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     98 115.05                           
2     96 113.73  2    1.3188 0.5566  0.575
```

Explanation:

1. **Simulate Data:** We simulate data where the true coefficients β_2 and β_3 are zero.
2. **Full Model:** We estimate the full model, including all variables.
3. **Restricted Model:** We estimate the restricted model, imposing the null hypothesis ($\beta_2 = \beta_3 = 0$).
4. **F-test:** We use the `anova()` function to compare the restricted and unrestricted models. This performs an F-test for the joint hypothesis that $\beta_2 = \beta_3 = 0$. The output provides the F-statistic and p-value. Alternatively, we can use the `linearHypothesis` function by specifying R and r appropriately.

Connection to the text: This script demonstrates how to test **multiple linear hypotheses** using an F-test, as discussed in Section 20.3.

R Script 5: Chow Test for Structural Break

```
# Load necessary libraries
library(tidyverse)
library(strucchange)
```

```
Loading required package: sandwich


Attaching package: 'strucchange'

The following object is masked from 'package:stringr':

    boundary
```

```
# Simulate data with a structural break in both intercept and slope
set.seed(1112)
n <- 200
breakpoint <- 100  # Breakpoint at observation 100
x <- runif(n, 0, 10)
y <- numeric(n)
```

```
# Before breakpoint
y[1:breakpoint] <- 1 + 0.5 * x[1:breakpoint] + rnorm(breakpoint, 0, 1)
# After breakpoint
y[(breakpoint + 1):n] <- 3 + 1.5 * x[(breakpoint + 1):n] + rnorm(n - breakpoint, 0, 1)

# Create a data frame
df <- data.frame(y, x)

# Perform the Chow test using the strucchange package
# sctest() tests for structural change using F-statistic
chow_test <- sctest(y ~ x, data = df, type = "Chow", point = breakpoint)
print(chow_test)
```

```
	Chow test

data:  y ~ x
F = 1478, p-value < 2.2e-16
```

```
# Visualize the data and the breakpoint
ggplot(df, aes(x = x, y = y)) +
  geom_point() +
  geom_vline(xintercept = median(x[1:breakpoint]), linetype = "dashed", color = "red") +
  geom_smooth(data = df[1:breakpoint, ], method = "lm", se = FALSE, color="blue") +
  geom_smooth(data = df[(breakpoint+1):n, ], method = "lm", se = FALSE, color = "green") +
  labs(title = "Chow Test Simulation",
       x = "x",
       y = "y")
```

```
`geom_smooth()` using formula = 'y ~ x'

`geom_smooth()` using formula = 'y ~ x'
```

Explanation:

1. **Simulate Data:** We simulate data with a clear structural break at observation 100. Both the intercept and slope change after the breakpoint.

2. **Chow Test:** We use the `sctest()` function from the `strucchange` package, specifying `type = "Chow"` and the `point` of the suspected breakpoint.

3. **Visualization:** The plot helps visualize the structural break and the different regression lines before and after the breakpoint.

Connection to the Text: This example demonstrates the **Chow test**, which is mentioned in examples 20.8 and 20.9, as a specific case of structural change analysis. The Chow test is a special type of F-test, used to check whether coefficients in two linear regressions on different data sets are equal.

YouTube Videos for Hypothesis Testing in Linear Regression

Here are some YouTube videos that explain the concepts mentioned in the attached text, along with their relevance and verification of availability:

1. Hypothesis Testing in the Multiple regression model

- **Video Title:** Hypothesis Testing in the Multiple regression model

- **Channel:** Ben Lambert
- **Link:** https://www.youtube.com/watch?v=k2jTkR0eB8M
- **Availability:** Verified (as of October 26, 2023).
- **Relevance:** This video focuses specifically on hypothesis testing within the multiple regression framework. It covers both t-tests for individual coefficients and F-tests for joint hypotheses, aligning well with Sections 20.2 and 20.3 of the text. It explains the concepts clearly and uses visual aids. The presenter uses matrix notation, as in the text.

2. F-statistic for joint significance

- **Video Title:** 1.2 - F-statistic for joint significance
- **Channel:** Pat Obi
- **Link:** https://www.youtube.com/watch?v=t_f5-bJqL5Y
- **Availability:** Verified (as of October 26, 2023).
- **Relevance:** This video is specifically dedicated to explaining the F-statistic and its use in testing the joint significance of multiple coefficients in a regression model. This directly relates to Section 20.3 (Test of Multiple Linear Hypothesis) and Example 20.4 (Standard F-test).

3. Hypothesis Tests and Confidence Intervals in Multiple Regression (FRM Part 1 – Book 2 – Chapter 8)

- **Video title**: Hypothesis Tests and Confidence Intervals in Multiple Regression (FRM Part 1 – Book 2 – Chapter 8)
- **Channel:** AnalystPrep
- **Link:** https://www.youtube.com/watch?v=FhLcpCkKk4s
- **Availability:** Verified (as of October 26, 2023).
- **Relevance:** This video clearly explains constructing confidence intervals for regression coefficients. It also discusses hypothesis tests, including t-tests, in the context of multiple regression. This connects to the discussion of confidence intervals near the end of Section 20.3.

4. Simple Linear Regression and Hypothesis Testing

- **Video Title:** Hypothesis Testing and p-values: Inferential statistics | Khan Academy
- **Channel:** Khan Academy
- **Link:** https://www.youtube.com/watch?v=KS6KEWaoOOE
- **Availability:** Verified (as of October 26, 2023)
- **Relevance:** While a more general statistics video, it gives a good foundational understanding of hypothesis testing, p-values, and their interpretation. It's a useful prerequisite for understanding the more specific hypothesis tests in the regression context. The concepts in this video are used throughout chapter 20.

5. Structural breaks and Chow test

- **Video Title:** Econometrics: Structural breaks and Chow test, F-test
- **Channel:** JDEConomics
- **Link:** https://www.youtube.com/watch?v=EQ5I2_eMR7w
- **Availability:** Verified (as of October 26, 2023)
- **Relevance:** The video discusses structural breaks and introduces the Chow test. The video complements examples 20.5, 20.7, 20.8 and 20.9.

6. Lagrange Multiplier, Likelihood Ratio and Wald tests (large sample tests)

- **Video title:** Lagrange Multiplier, Likelihood Ratio and Wald tests (large sample tests)
- **Channel:** Ben Lambert
- **Link:** https://www.youtube.com/watch?v=5JopxA_EdlQ
- **Availability:** Verified (as of October 26, 2023)
- **Relevance:** This video introduces the three asymptotically equivalent tests: Wald, LM and LR tests. It shows the intuition behind each of the tests. This video closely matches section 20.5.

These videos, taken together, provide a strong visual and auditory complement to the material presented in Chapter 20 of the text. They cover single and multiple hypothesis testing, the F-test, structural change analysis, and the principles behind different testing approaches.

Multiple Choice Exercises

MC Exercise 1

MC Solution 1

In the Cobb-Douglas production function $Q = AK^{\alpha}L^{\beta}$, constant returns to scale implies:

a. $\alpha = \beta$

b. $\alpha + \beta = 1$

c. $\alpha + \beta = 0$

d. $\alpha \cdot \beta = 1$

MC Exercise 2

MC Solution 2

In the market efficiency model $r_t = \mu + \gamma^T I_{t-1} + \epsilon_t$, the efficient market hypothesis predicts:

a. $\mu = 0$

b. $\gamma = 0$

c. $\epsilon_t = 0$

d. $I_{t-1} = 0$

MC Exercise 3

MC Solution 3

In the model $y_t = \alpha + \beta x_t + \gamma D_t + \epsilon_t$, where D_t is a dummy variable for a structural change, the null hypothesis $\gamma = 0$ implies:

a. There is no relationship between y and x.

b. There is no structural change.

c. The intercept is zero.

d. The slope is zero.

MC Exercise 4

MC Solution 4

The t-statistic for testing $c^T\beta = \gamma$ in the linear regression model is given by:

a. $\frac{c^T\hat{\beta}-\gamma}{\sigma\sqrt{c^T(X^TX)^{-1}c}}$

b. $\frac{c^T\hat{\beta}-\gamma}{s_*\sqrt{c^T(X^TX)^{-1}c}}$

c. $\frac{c^T\hat{\beta}-\gamma}{s_*^2c^T(X^TX)^{-1}c}$

d. $\frac{c^T\hat{\beta}-\gamma}{\sigma^2c^T(X^TX)^{-1}c}$

MC Exercise 5

MC Solution 5

$s_*^2 = \frac{\hat{\epsilon}^T\hat{\epsilon}}{n-K}$ is an unbiased estimator of σ^2 because:

a. $E[\hat{\epsilon}^T\hat{\epsilon}] = n\sigma^2$

b. $E[\hat{\epsilon}^T\hat{\epsilon}] = (n-K)\sigma^2$

c. $E[\hat{\epsilon}^T\hat{\epsilon}] = K\sigma^2$

d. $E[\hat{\epsilon}^T\hat{\epsilon}] = \sigma^2$

MC Exercise 6

MC Solution 6

A 95% confidence interval for $c^T\beta$ is given by:

a. $c^T\hat{\beta} \pm t_{0.05}(n-K) \cdot s_*\sqrt{c^T(X^TX)^{-1}c}$

b. $c^T\hat{\beta} \pm t_{0.025}(n-K) \cdot s_*\sqrt{c^T(X^TX)^{-1}c}$

c. $c^T\hat{\beta} \pm t_{0.95}(n-K) \cdot s_*\sqrt{c^T(X^TX)^{-1}c}$

d. $c^T\hat{\beta} \pm t_{0.975}(n-K) \cdot s_*\sqrt{c^T(X^TX)^{-1}c}$

MC Exercise 7

MC Solution 7

The F-statistic for testing multiple linear restrictions follows an F-distribution because, under the null hypothesis, it is the ratio of:

a. Two normal random variables.

b. A chi-squared random variable and a normal random variable.

c. Two chi-squared random variables, each divided by their degrees of freedom.

d. Two t-distributed random variables.

MC Exercise 8

MC Solution 8

To test the hypothesis $\beta_1 = 0$ and $\beta_2 = \beta_3$ in a regression with three independent variables, the R matrix in $R\beta = r$ would be:

a. $R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}$

b. $R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

c. $R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$

d. $R = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$

MC Exercise 9

MC Solution 9

The restricted least squares estimator, β^*, minimizes the sum of squared residuals:

a. Without any constraints.

b. Subject to the constraint $R\beta = r$.

c. Subject to the constraint $\beta = 0$.

d. Subject to the constraint that the residuals sum to zero.

MC Exercise 10

MC Solution 10

The formula for the restricted least squares estimator β^* is:

a. $\beta^* = \hat{\beta} + (X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$

b. $\beta^* = \hat{\beta} - (X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$

c. $\beta^* = (X^TX)^{-1}X^Ty$

d. $\beta^* = R(X^TX)^{-1}X^Ty$

MC Exercise 11

MC Solution 11

$Q^* - Q$, the difference between the restricted and unrestricted sum of squared residuals, can be expressed as:

a. $(R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$

b. $(R\hat{\beta} - r)^T(R\hat{\beta} - r)$

c. $(R\hat{\beta} - r)^T(X^TX)(R\hat{\beta} - r)$

d. $r^T[R(X^TX)^{-1}R^T]^{-1}r$

MC Exercise 12

MC Solution 12

To test if all slope coefficients are zero using R^2, the F-statistic is:

a. $\frac{R^2/K}{(1-R^2)/(n-K-1)}$

b. $\frac{R^2/(K-1)}{(1-R^2)/(n-K)}$

c. $\frac{R^2}{(1-R^2)}$

d. $\frac{(1-R^2)/(n-K)}{R^2/K}$

MC Exercise 13

MC Solution 13

In the context of structural change with two subsamples, the null hypothesis of no structural change implies:

a. The intercepts are different, but the slopes are the same.

b. The slopes are different, but the intercepts are the same.

c. Both the intercepts and slopes are the same across the two subsamples.

d. Both the intercepts and slopes are different across the two subsamples.

MC Exercise 14

MC Solution 14

Which of the following statements about the Wald, LR, and LM tests is FALSE?

a. The Wald test only requires estimating the unrestricted model.

b. The LM test only requires estimating the restricted model.

c. The LR test requires estimating both the restricted and unrestricted models.

d. The Wald, LR, and LM tests are asymptotically equivalent, but always produce identical results in finite samples.

MC Exercise 15

MC Solution 15

The first derivative of the log-likelihood function with respect to β in the linear regression model is:

a. $\frac{1}{\sigma^2}X^T(y - X\beta)$

b. $-\frac{1}{\sigma^2}X^T(y - X\beta)$

c. $X^T(y - X\beta)$

d. $\frac{1}{2\sigma^2}X^T(y - X\beta)$

MC Exercise 16

MC Solution 16

The Wald statistic for testing $R\beta = r$ can be expressed in terms of $Q^* - Q$ as:

a. $\frac{Q^*-Q}{q}$

b. $\frac{Q^*-Q}{Q/(n-K)}$

c. $n\frac{Q^*-Q}{Q}$

d. $\frac{Q^*-Q}{Q^*/(n-K)}$

MC Exercise 17

MC Solution 17

The LM statistic can be expressed in terms of Q and Q* as:

a. $n\frac{Q}{Q^*}$

b. $n\frac{Q^*}{Q}$

c. $n\frac{Q^*-Q}{Q^*}$

d. $n\frac{Q^*-Q}{Q}$

MC Exercise 18

MC Solution 18

Which of the following statements about the relationship between the F, Wald (W), LM, and LR statistics is generally TRUE?

a. $F = W = LM = LR$ for all sample sizes.

b. $LM \leq LR \leq W$ and they are asymptotically equivalent as $n \to \infty$.

c. $W \leq LR \leq LM$ and they are asymptotically equivalent as $n \to \infty$.

d. The F-statistic is unrelated to the W, LM, and LR statistics.

MC Exercise 19

MC Solution 19

The Bayesian approach to inference in the linear regression model involves:

a. Only using the likelihood function.

b. Combining a prior distribution with the likelihood to obtain a posterior distribution.

c. Only using the prior distribution.

d. Maximizing the sum of squared residuals.

MC Exercise 20

MC Solution 20

If the prior for β is $N(\beta_0, \Sigma_0)$ and the likelihood is from the classical linear regression model with known σ^2, the posterior distribution of β is:

a. t-distributed

b. F-distributed

c. Normal

d. Chi-squared

Multiple Choice Solutions

MC Solution 1

MC Exercise 1

(b) $\alpha + \beta = 1$

Constant returns to scale means that doubling all inputs doubles the output. This is mathematically represented as $\alpha + \beta = 1$. This is directly from Example 20.1.

MC Solution 2

MC Exercise 2

(b) $\gamma = 0$

The efficient market hypothesis implies that past information (I_{t-1}) should not predict returns (r_t). Thus, the coefficient vector γ should be zero. This is directly from Example 20.2.

MC Solution 3

MC Exercise 3

(b) There is no structural change.

If $\gamma = 0$, the dummy variable D_t has no effect, meaning the relationship between y and x is the same regardless of the value of D_t (i.e., before and after the potential structural break). This is directly from Example 20.3.

MC Solution 4

MC Exercise 4

(b) $\frac{c^T\hat{\beta}-\gamma}{s_*\sqrt{c^T(X^TX)^{-1}c}}$

We use the unbiased estimator of the variance, s_*, in the denominator. This is derived and presented in Section 20.2.

MC Solution 5

MC Exercise 5

(b) $E[\hat{\epsilon}^T\hat{\epsilon}] = (n-K)\sigma^2$

As shown in the derivation in Section 20.2, the expected value of the sum of squared residuals is $(n-K)\sigma^2$.

MC Solution 6

MC Exercise 6

(b) $c^T\hat{\beta} \pm t_{0.025}(n-K) \cdot s_*\sqrt{c^T(X^TX)^{-1}c}$

For a 95% confidence interval, we use the critical value from the t-distribution with $n-K$ degrees of freedom that corresponds to a two-tailed test with $\alpha/2 = 0.025$ in each tail. This concept is described at the end of Section 20.2 and in the beginning of Section 20.3.

MC Solution 7

MC Exercise 7

(c) Two chi-squared random variables, each divided by their degrees of freedom.

This is the definition of an F-distributed random variable, as discussed in Section 20.3.

MC Solution 8

MC Exercise 8

b. $R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

Assuming the regression model includes an intercept, $\beta = [\beta_0, \beta_1, \beta_2, \beta_3]^T$. The first row of R corresponds to $\beta_1 = 0$ and the second row of R to $\beta_2 - \beta_3 = 0$ (or $\beta_2 = \beta_3$). This illustrates setting up a matrix for testing **multiple linear hypotheses** as in section 20.3.

MC Solution 9

MC Exercise 9

(b) Subject to the constraint $R\beta = r$.

The restricted least squares estimator imposes the restrictions of the null hypothesis. This is the core concept of Section 20.4.

MC Solution 10

MC Exercise 10

(b) $\beta^* = \hat{\beta} - (X^TX)^{-1}R^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$

This is the formula derived using Lagrange multipliers in Section 20.4, equation (20.8).

MC Solution 11

MC Exercise 11

(a) $(R\hat{\beta} - r)^T[R(X^TX)^{-1}R^T]^{-1}(R\hat{\beta} - r)$

This is shown in the proof of Theorem 20.2 in Section 20.4.

MC Solution 12

MC Exercise 12

(b) $\frac{R^2/(K-1)}{(1-R^2)/(n-K)}$

This is the F-statistic for testing the overall significance of the regression, where $K - 1$ is the number of slope coefficients (excluding the intercept) and $n - K$ is the degrees of freedom for the residuals. This is discussed in Example 20.6.

MC Solution 13

MC Exercise 13

(c) Both the intercepts and slopes are the same across the two subsamples.

No structural change implies that the entire regression relationship is the same for both groups. This relates to Examples 20.7, 20.8 and 20.9.

MC Solution 14

MC Exercise 14

(d) The Wald, LR, and LM tests are asymptotically equivalent, but always produce identical results in finite samples.

The tests are asymptotically equivalent, but their values and resulting decisions can differ in finite samples. The inequalities $LM \leq LR \leq W$ are generally true. This is discussed in Section 20.5.

MC Solution 15

MC Exercise 15

(a) $\frac{1}{\sigma^2} X^T (y - X\beta)$

This is derived in Section 20.5.

MC Solution 16

MC Exercise 16

(c) $n \frac{Q^* - Q}{Q}$

This comes from expressing the Wald statistic in terms of the MLE of σ^2 which is Q/n, as shown in Equation (20.13).

MC Solution 17

MC Exercise 17

(c) $n \frac{Q^* - Q}{Q^*}$

This is the form of the LM statistic derived in Section 20.5.

MC Solution 18

MC Exercise 18

(b) $LM \leq LR \leq W$ and they are asymptotically equivalent as $n \to \infty$.

This is the general relationship discussed in Section 20.5, and equation (20.15).

MC Solution 19

MC Exercise 19

(b) Combining a prior distribution with the likelihood to obtain a posterior distribution.

This is the fundamental principle of Bayesian inference, as mentioned in Section 20.6.

MC Solution 20

MC Exercise 20

(c) Normal

With a normal prior and a normal likelihood (from the classical linear regression model), the posterior distribution is also normal. This is stated in Theorem 20.3.

Author: Peter Fuleky