ECOM40006/90013 Econometrics 3

Week 9 Extras

Background: Hypothesis Testing Building Blocks

Before launching into hypothesis testing, this question aims to give some foundation for those who might need it.

Hypothesis testing has two main goals, which concern the concepts of **size** and **power**. More generally, for a specified hypothesis test, we define

$$\text{Size} = \mathbf{P}(\text{reject } H_0 \mid H_0 \text{ true})$$

and

$$\text{Power} = \mathbf{P}(\text{reject } H_0 \mid H_0 \text{ false}).$$

Accompanying a test will be a testing rule, which will be the conditions under which the null hypothesis is rejected. Generally, this is measured by some sort of *critical value*: namely, if your test statistic exceeds this critical value, then the null is rejected. You can think of it as a measure of distance, or "how far is far enough to reject".

(a) Does the size of a test represent the probability of making a correct decision or a wrong decision? How about for the power of a test? Given your answers to the above, what would the ideal hypothesis test look like in terms of size and power?

(b) Consider the hypotheses

(i.) $H_0 : \beta_1 = 0$

(ii.) $H_0 : \beta_1 = 4$

(iii.) $H_0 : \beta_1 = c$ for some $c > 0$

against the alternative hypothesis that β_1 is not equal to the constant on the right-hand side of the null hypothesis. Further suppose that the true value of β_1 is 0. For each of the cases, does a rejection of the null contribute to the size or to the power of the test? Be very explicit in your answer.

(c) All standard statistics packages test against a null hypothesis of $\beta_1 = 0$. Intuitively, what do you expect to happen to the rejection rate of the test if

(i.) the true value β_1 increases?

(ii.) the number of observations n increases?

In your answer, consider two cases: where H_0 is true and where H_0 is not true. In particular, be careful with the answer to part (ii) for when the null is true.

(d) The standard rejection rule for a hypothesis test is to reject when $|t| > \alpha$, where α is the critical value of the test.

(i.) Suppose that the null hypothesis $H_0 : \beta_1 = 0$ is true. Derive the size of the test $\mathbf{P}(|t| > \alpha)$. (Two hints: (i) is it the case that t is t-distributed? (ii) $|t| > \alpha$ means that either one of $t > \alpha$ or $t < -\alpha$.)

(ii.) Suppose that the null hypothesis $H_0 : \beta_1 = 0$ is **false**. Derive the power of the test $\mathbf{P}(|t| > \alpha)$ under these conditions. (*Hint*: the t-statistic is not t-distributed here. Convince yourself this is true and find something else that is t-distributed, then work with that instead.)

Throughout, let $F_{n-k}(.)$ denote the CDF of the t distribution with $n-k$ degrees of freedom.

(e) For each of the expressions you derived in part (e), take the derivative with respect to α. Are size and power increasing or decreasing in the critical value α? What do your results say about your ability to come up with a "ideal" hypothesis test? (*Hint*: you'll need to compare to your answers from part (a) for this one.)

You may denote $f_{n-k}(.)$ to be the PDF of the t distribution with $n-k$ degrees of freedom.

(f) Not all hypotheses are meant to be tested. Suppose you wanted to test the hypothesis

$$H_0 : \beta_1 = 1 \quad \text{and} \quad \beta_1 = -1,$$

against the alternative that the hypothesis is false. Something goes wrong with our ability to calculate one of either the size or the power of the test. What goes wrong and what does this tell you about our ability to test this hypothesis?

Question 1: Linear restrictions

Let's get started with some basic foundations of constructing hypotheses in matrix form. For each part, the regression model is provided. Using this information, along with the hypotheses provided, find R and r. The testing rule is not included here as we are looking at just constructing restrictions. Note that this question can get a bit repetitive, but fortunately the working is not as long as it looks.

Univariate restrictions. *For this part, the model is* $y_i = \beta_0 + u_i$.

(a) $H_0 : \beta_0 = 6$ and $H_1 : \beta_0 \neq 6$

For the following parts, the model is now $y_i = \beta_0 + \beta_1 x_{1,i} + u_i$.

(b) $H_0 : \beta_0 = 4$ and $H_1 : \beta_0 \neq 4$

(c) $H_0 : \beta_1 = 3$ and $H_1 : \beta_1 \neq 3$

(d) $H_1 : \beta_0 = 0,\ \beta_1 = 0$ and H_1 : at least one of the restrictions in H_0 does not hold. (The null hypothesis may be more recognisable in the form $\beta_0 = \beta_1 = 0$.)

For the following parts, the model is now $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i$.

(e) $H_0 : \beta_0 = 6$ and $H_1 : \beta_0 \neq 6$

(f) $H_0 : \beta_0 + \beta_1 = 3,\ \beta_2 = 4$ and H_1 : at least one of the restrictions in H_0 does not hold.

(g) $H_0 : 3\beta_0 + 7\beta_1 = -6,\ 2\beta_0 - \beta_2 = 7$ and H_1 : at least one of the restrictions in H_0 does not hold.

Multivariate restrictions. Define the linear regression model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u_i$$

with the following hypotheses

$$\begin{aligned} H_0 : \quad & \beta_2 = 0, \\ & \beta_3 = 0 \end{aligned}$$

and H_1 : at least one of $\beta_2 = 0$ or $\beta_3 = 0$ is not satisfied.

(h) Find R and r so that the above restrictions can be written in the form $R\beta = r$.

(i) What are the dimensions of R?

(j) Let Σ be the 4×4 covariance matrix associated with $\hat{\beta}$, which is the OLS estimator. What are the dimensions of the matrix $R\Sigma R'$?

(k) Given your answer in (c), what does $R\Sigma R'$ represent? (i.e. how does it look relative to the original matrix Σ?)

Question 2: Linear hypotheses, continued

You might be familiar with joint tests from previous courses. In this question, we'll look at what happens when we test multiple restrictions in a matrix format. In particular, consider the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u_i \qquad \text{or} \qquad y = X\beta + u,$$

where $u \sim N(0, \sigma^2 I_N)$ is i.i.d. (where all the classic assumptions apply). Suppose that we want to test the hypotheses

$$\begin{aligned} & H_0 : \beta_0 = 3 \text{ and } 2\beta_1 + \beta_2 = 1, \\ & H_1 : \text{at least one restriction above fails.} \end{aligned}$$

(a) Write the null hypothesis in the form $H_0 : R\beta = r$. State R, β and r explicitly. Also state the *size* (i.e. number of rows and columns) of all three expressions.

(b) What is the mean and variance of $R\hat{\beta} - r$, assuming that H_0 is true?

(c) Let $g = R\beta - r$ and hence $\hat{g} = R\hat{\beta} - r$. When H_0 is true, we can construct a Wald test statistic

$$W = \hat{g}'\text{Var}(\hat{g})^{-1}\hat{g}.$$

Derive the asymptotic distribution of this test statistic. What does this suggest about the critical values you should use when doing a joint hypothesis test? (*Hint: It's not as bad as it sounds if you use a certain theorem – perhaps we might have seen a useful theorem in one of the earlier extras?*)

Question 3: The Delta method, revisited

Up to this point, the hypotheses we've been playing with generally involve linear combinations of our model coefficients. We're going to look at what happens in the case where these hypotheses are nonlinear. Suppose that we have a $K \times 1$ vector β and that we wish to test the hypotheses

$$H_0 : f(\beta) = c$$
$$H_1 : f(\beta) \neq c$$

with the usual testing rule of rejecting if the test statistic exceeds the critical value $\alpha > 0$ in absolute magnitude. The first step in such a test is to create a new function

$$g(\beta) = f(\beta) - c$$

such that it is equal to zero under the null hypothesis. We also know that the vector of parameter estimates $\hat{\beta}$ is a random vector with the asymptotic distribution

$$\sqrt{N}(\hat{\beta} - \beta) \sim N(0, \Sigma),$$

or

$$\hat{\beta} \sim N\left(\beta, \frac{\Sigma}{N}\right).$$

(a) Suppose that $g(\hat{\beta})$ is a continuous function of a consistent estimator $\hat{\beta}$ (for β). Restate the Delta method for the univariate and multivariate cases.

(b) A test statistic can be constructed under H_0 as

$$\frac{g(\hat{\beta}) - g(\beta)}{\text{sd}(g(\hat{\beta}))} = \frac{g(\hat{\beta})}{\text{sd}(g(\hat{\beta}))}.$$

What is the asymptotic distribution of this test statistic? What does this imply about the critical values we can use?

(c) Using the Delta method, explain how you would test the hypotheses

$$H_0 : \beta_1 \exp(\beta_2) = 1$$
$$H_1 : \beta_1 \exp(\beta_2) \neq 1$$

where $\beta = (\beta_1, \beta_2)'$ is the vector of parameters to test.

Question 4: A trinity of tests (univariate)

Consider $n = 3$ i.i.d. observations $\{y_1, y_2, y_3\} = \{1, 2, 3\}$ from the exponential density

$$f(y_i; \theta) = \frac{1}{\theta} \exp\left(-\frac{y_i}{\theta}\right)$$

with log-likelihood

$$\log L(\theta; y_i) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^{n} y_i$$

Suppose that we were interested in testing the following univariate hypothesis:

$$H_0 : \theta = 1$$
$$H_1 : \theta \neq 1$$

(a) Give an intuitive, or heuristic, description of the difference between a restricted model (under H_0) and an unrestricted model.

(b) Give a description of three possible hypothesis tests that you could employ to evaluate the hypotheses above. What are the limiting distributions of these three hypothesis tests?

(c) Derive expressions for the score $S(\theta)$ and the outer product of the gradients (OPG) matrix, $J(\theta)$..

(d) The OPG matrix can be used as an estimator for $\text{Var}(\hat{\theta})$. Define $\hat{\theta}_0$ to be the *restricted* parameter estimate of θ and $\hat{\theta}_1$ to be the unrestricted parameter estimate of θ. Use each of the three tests you identify in (b) to test the hypotheses in this question at the 5% level of significance. You may also use the 5% critical value for a χ^2_1 distribution, which is 3.841.

Question 5: A trinity of tests (multivariate)

Consider $n = 3$ i.i.d. observations $\{y_1, y_2, y_3\} = \{1, 2, 3\}$ from the normal distribution with parameters $\theta = \{\mu, \sigma^2\}$ and density

$$f(y_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

with log-likelihood

$$\log L(\theta; y) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{n}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

with gradient at observation i (note this is not the score) given by

$$g_i(\theta) = \begin{bmatrix} \dfrac{\partial \log f(y_i; \theta)}{\partial \mu} \\[2ex] \dfrac{\partial \log f(y_i; \theta)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}(y_i - \mu) \\[2ex] -\dfrac{1}{2\sigma^2} + \dfrac{1}{2\sigma^4}(y_i - \mu)^2 \end{bmatrix}$$

and maximum likelihood estimators

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2.$$

In this question, our interest lies in testing the following hypotheses

$$H_0 : \mu = 1 \text{ and } \sigma^2 = 1$$
$$H_1 : \text{at least one restriction fails to hold}$$

Throughout, you may use the fact that the 5% critical value for a χ^2_2 distribution is 5.991. Use the inverse OPG matrix $J(\hat{\theta})^{-1}$ as an estimate for $\text{Var}(\hat{\theta})$ in all cases.

(a) Test the hypotheses above at the 5% level, using a Likelihood Ratio (LR) test.

(b) Repeat (a) using a Wald test.

(c) Repeat (a) using a Lagrange Multiplier (LM) test.

It is worth noting that if you have the information matrix available, that is a good choice to use for estimating Var($\hat{\theta}$) in general. The main point of this question is to give you some practice at calculating the OPG matrix manually so that you have a better idea of what it entails.