

ECON 5350 Class Notes

Hypothesis Testing and Prediction

1 Introduction

Our primary goal in this chapter is develop a systematic method for testing restrictions, which allow us to distinguish between **nested models**. Nested models are such that one model can be written as a special case of the other. For example, if one wished to test $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$ versus $y_i = \beta_1 + \epsilon_i$, this would be considered a nested hypothesis test because the second model is nested ($\beta_2 = 0$) within the first model. If one wished to test the first model against $y_i = \beta_1 + \beta_2 z_i + \epsilon_i$, this would be considered a non-nested test because one model cannot be written as a special case of the other.

2 Testing Linear Restrictions

Start by reintroducing Classical assumption #6: $\epsilon \sim N(0, \sigma^2 I)$. Because b is a linear function of ϵ , we then know that $b \sim N(\beta, \sigma^2 (X'X)^{-1})$. In this first section, we will take a first pass at testing some simple hypotheses regarding the population regression equation. The following sections develop a more general approach for hypothesis testing.

2.1 Single Coefficient Tests

Consider two different cases – σ^2 is known to the econometrician, and more realistically, σ^2 is not known to the econometrician. We wish to test whether a single element of the β vector is equal to zero or not.

- σ^2 known. Assume we are testing the null hypothesis, $H_0: \beta_k = 0$, against the alternative hypothesis, $H_A: \beta_k \neq 0$. We can form the statistic

$$z_k = \frac{b_k - \beta_k}{\sigma \sqrt{s^{kk}}},$$

where s^{kk} is the k^{th} diagonal element of $(X'X)^{-1}$. Under the null hypothesis, z_k will have a standard normal $N(0, 1)$ distribution.

- σ^2 unknown. Testing the same hypothesis, we want to replace σ with s . Doing so, means z_k no longer has a standard normal distribution. Instead, we form the statistic

$$t_k = \frac{b_k - \beta_k}{s_{bk}}$$

where $s_{bk} = s\sqrt{s^{kk}}$, which has a student's t distribution with $n - k$ degrees of freedom. This is motivated by the fact that the ratio of z_k (standard normal) to the square root of $(n - k)s^2/\sigma^2$ (chi-squared distribution) divided by its degrees of freedom has a student's t distribution. This also relies on independence between the z_k and $(n - k)s^2/\sigma^2$.

2.2 Tests of Overall Significance

Now assume that we wish to assess the overall significance of the regression model. That is, we want to test whether or not the explanatory variables explain a significant amount of the variation in the dependent variable. The null hypothesis in this case will be $H_0: \beta_2 = \dots = \beta_k = 0$. The alternative hypothesis is that the null is false. Not surprisingly, we can use the R^2 to execute the test. The test statistic is

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

which under the null hypothesis has an F distribution with $k - 1$ and $n - k$ degrees of freedom in the numerator and denominator, respectively. A couple of notes.

- If the true model is $y_i = \beta_1 + \epsilon_i$, then no variation in y_i around its mean of β_1 is explained and $R^2 = 0$.
- It is possible for all individual coefficients to be insignificant while jointly they are significant.

2.3 Asymptotic Behavior of Test Statistics

We know that if the errors are normal (i.e., $\epsilon \sim N(0, \sigma^2 I)$), then

- $t = (b - \beta)/s_b \sim t(n - k)$.
- $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k - 1, n - k)$.

If instead we did not impose normality (i.e., $\epsilon \sim (0, \sigma^2 I)$), then we can show

- $t = (b - \beta)/s_b \stackrel{asy}{\sim} N(0, 1)$.
- $(k - 1)F = \frac{R^2}{(1-R^2)/(n-k)} \stackrel{asy}{\sim} \chi^2(k - 1)$.

2.4 General Testing Approach (Unrestricted)

This is called the "unrestricted approach" because we will only estimate the unrestricted model (i.e., without imposing the restriction in H_0). We will represent a set of J linear testable restrictions on $Y = X\beta + \epsilon$ as

$$R\beta = q$$

where R is a $(J \times k)$ restriction matrix with full row rank and q is a $(J \times 1)$ vector of constants. Here are some examples:

- $H_0: \beta_1 = 0$

$$\begin{aligned} R &= [1 \ 0 \ \dots \ 0]_{1 \times k} \\ \beta' &= [\beta_1 \ \beta_2 \ \dots \ \beta_k]_{1 \times k} \\ q &= 0 \end{aligned}$$

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

$$\begin{aligned} R &= [0 \ 1 \ 1 \ 0 \ \dots \ 0]_{1 \times k} \\ \beta' &= [\beta_1 \ \beta_2 \ \dots \ \beta_k]_{1 \times k} \\ q &= 1 \end{aligned}$$

- $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$

$$\begin{aligned} R &= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{J \times k} \\ \beta' &= [\beta_1 \ \beta_2 \ \dots \ \beta_k]_{1 \times k} \\ q &= [0 \ 0 \ \dots \ 0]_{1 \times J} \end{aligned}$$

2.4.1 Motivating the Test Statistic

Assume that $\epsilon \sim N(0, \sigma^2 I)$. What is the sampling distribution of Rb ?

- $E(Rb) = R\beta$.
- $\text{var}(Rb) = E[(Rb - E[Rb])(Rb - E[Rb])'] = E[R(b - \beta)(b - \beta)'R'] = R\text{var}(b)R' = \sigma^2 R(X'X)^{-1}R'$.
- $Rb \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$.
- If H_0 is true, $Rb - q = m \sim N(0, \sigma^2 R(X'X)^{-1}R')$.
- From (Greene, Theorem B.11), $(Rb - q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(Rb - q) = m'\text{var}(m)^{-1}m \sim \chi^2(J)$.

- Replace σ^2 with $s^2 \implies F = \frac{(Rb-q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(Rb-q)/J}{(n-k)s^2/\sigma^2(n-k)}$ is the ratio of two independent chi-squared random variables. Therefore, we know that

$$F = \frac{(Rb - q)'(R(X'X)^{-1}R')^{-1}(Rb - q)}{s^2 J} \sim F(J, n - k). \quad (1)$$

2.4.2 Examples Continued

- $H_0: \beta_1 = 0$

$$F = \frac{(Rb - q)'(R(X'X)^{-1}R')^{-1}(Rb - q)}{s^2 J} = \frac{b_1'(s^{11})^{-1}b_1}{s^2} = \frac{b_1^2}{\widehat{\text{var}}(b_1)} \sim F(1, n - k)$$

or taking square roots...

$$t = \frac{b_1}{\sqrt{\widehat{\text{var}}(b_1)}} = \frac{b_1}{\text{se}(b_1)} = \frac{b_1}{s_{b_1}} \sim t(n - k).$$

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

$$F = \frac{(Rb - q)'(R(X'X)^{-1}R')^{-1}(Rb - q)}{s^2 J} = \frac{(b_2 + b_3 - 1)'(b_2 + b_3 - 1)}{s^2(s^{22} + 2s^{23} + s^{33})} \sim F(1, n - k)$$

or taking square roots...

$$t = \frac{b_2 + b_3 - 1}{\text{se}(b_2 + b_3)} \sim t(n - k).$$

2.4.3 A Few Notes

1. It would be simple to test these last two restrictions jointly ($J = 2$)

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 \end{bmatrix} \text{ and } q = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

2. It is possible to calculate joint confidence regions analogous to confidence intervals.
3. Recall, if the errors are not normal but $\epsilon \sim (0, \sigma^2 I)$, then the Wald, Lagrange multiplier and likelihood ratio statistics are asymptotically distributed chi-squared and can be used to test linear restrictions.

2.5 General Testing Approach (Restricted)

In the "restricted approach", we estimate the model with the restriction imposed and then compare the change in the goodness-of-fit of the model with and without the restriction imposed. Turn now to the problem of a restricted regression.

2.5.1 Restricted Regression

The problem is to

$$\min_{\beta} (Y - X\beta)'(Y - X\beta) \text{ subject to } R\beta = q.$$

Next, form the Lagrangian

$$L^* = (Y - X\beta)'(Y - X\beta) + 2\lambda'(R\beta - q).$$

The first-order conditions are

$$\begin{aligned} \frac{\partial L^*}{\partial b_*} &= -2X'Y + 2(X'X)b_* + 2R'\lambda = 0 \\ \frac{\partial L^*}{\partial \lambda} &= 2(Rb_* - q) = 0. \end{aligned}$$

Written in matrix form gives

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} b_* \\ \lambda \end{bmatrix} = \begin{bmatrix} X'Y \\ q \end{bmatrix} \implies \begin{bmatrix} b_* \\ \lambda \end{bmatrix} = \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'Y \\ q \end{bmatrix}. \quad (2)$$

Solving (2) for b_* and λ is straightforward, although tedious, using (Greene, A-74). This produces

$$b_* = b - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(Rb - q). \quad (3)$$

So, as expected, the restricted and unrestricted estimates of β are equal if the restriction is exactly true in the sample data. Otherwise, b_* and b will be different (note that because we are dealing with a random sample, this will generally be true even if the restriction holds in the population, i.e., $R\beta = q$).

2.5.2 Test Based on Loss of Fit

Let $e = Y - Xb$ and $e_* = Y - Xb_*$ be the unrestricted and restricted residuals, respectively. We can relate them according to

$$\begin{aligned} e_* &= Y - Xb_* \\ &= Y - Xb - X(b_* - b) \\ &= e - X(b_* - b). \end{aligned}$$

Now taking the inner product of e_* (i.e., sum of squared restricted residuals) gives

$$e_*'e_* = e'e + (b_* - b)'(X'X)(b_* - b).$$

Substituting in (3) and simplifying gives

$$(e'_*e_* - e'e) = (Rb - q)'(R(X'X)^{-1}R')^{-1}(Rb - q). \quad (4)$$

Finally, substituting (4) into (1) gives

$$F = \frac{(e'_*e_* - e'e)/J}{(e'e)/(n - k)} \sim F(J, n - k).$$

This shows that the statistic used to test $R\beta = q$ can be interpreted as the relative loss in fit caused by imposing the restriction. If the restriction is true, the loss in fit should be small, the F statistic will be small, and you will fail to reject the null. If the restriction is false, the loss in fit will be large, the F statistic will be large, and you will reject the null. Alternatively, if one divides through by the SST , then the F statistic can be written in terms of the unrestricted and restricted R^2 s

$$F = \frac{((1 - e'e/y'M^0y) - (1 - e'_*e_*/y'M^0y))/J}{(e'e/y'M^0y)/(n - k)} = \frac{(R^2 - R_*^2)/J}{(1 - R^2)/(n - k)} \sim F(J, n - k). \quad (5)$$

Note that equations (4) and (5) are used to produce alternative interpretations of the hypothesis test. Generally, researchers continue to use equation (1), which only requires calculation of the unrestricted estimator.

2.6 MATLAB Example

This example uses cross-sectional data from the 1998 Current Population Survey. The data are for $n = 1000$ males. The regression equation is

$$\ln(\text{wage}_i) = \beta_1 + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 \text{grade}_i + \beta_5 \text{married}_i + \beta_6 \text{union}_i + \epsilon_i.$$

We wish to test two hypotheses. The first is whether schooling has an impact on wages (let's hope it does!) and the second is whether these five variables jointly explain a significant amount of the variation in wages across the 1000 males.

1. Schooling hypothesis. $H_0: \beta_4 = 0$ versus $H_A: \beta_4 \neq 0$. In our notation, we set $R = (0 \ 0 \ 0 \ 1 \ 0 \ 0)$ and $q = 0$. At the 5% significance level, the critical F value with 1 degree of freedom in the numerator and 994 degrees of freedom in the denominator is 3.84.
2. Overall significance. $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ versus the $H_A: H_0$ is false. We set $R = (0 \mid I_5)$ and $q = 0$, the latter being a (5×1) vector. The critical F value with 5 degrees of freedom in the numerator and 994 degrees of freedom in the denominator is 2.21.

See [MATLAB example 11](#) to calculate the statistics and complete the tests.

3 Prediction

In this section, we will use our regression model to predict values of the dependent variable given observations on the regressors, X^0 . These observations may either be in-sample or out-of-sample. The predicted value is

$$\hat{Y}^0 = X^0 b.$$

The Gauss-Markov theorem implies that \hat{Y}^0 is the best linear unbiased estimator of $E[Y^0|X^0] = X^0\beta$. The prediction error, $e^0 = Y^0 - \hat{Y}^0$, has the following variance

$$\begin{aligned} \text{var}(e^0) &= \text{var}(Y^0 - \hat{Y}^0) = \text{var}(X^0\beta + \epsilon^0 - X^0b) \\ &= \text{var}(X^0(\beta - b) + \epsilon^0) \\ &= X^0 \text{var}(b) X^{0'} + \sigma^2 I \\ &= X^0 \sigma^2 (X'X)^{-1} X^{0'} + \sigma^2 I \\ &= \sigma^2 (I + X^0 (X'X)^{-1} X^{0'}). \end{aligned} \tag{6}$$

Equation (6) highlights that there are two sources of uncertainty associated with predicting Y^0 – the first is associated with the random error term ϵ^0 and the second is associated with estimating the population parameters β . Note also that X , as opposed to X^0 , appears inside equation (6). This occurs because all n observations are used in calculating the least squares estimator b , not just the n^0 observations in X^0 .

It is often desirable to present a confidence interval around the predicted value. In this case, the $(1 - \lambda) * 100$ percent confidence interval for Y^0 is

$$\hat{Y}^0 \pm t_{\lambda/2} se(e^0)$$

where the standard error of e^0 is given by square root of (6), with σ^2 replaced by s^2 .

After predictions are made, we often wish to evaluate their accuracy. Two common measures are

$$\begin{aligned} \text{Root mean square error (RMSE)} &= \sqrt{\frac{1}{n^0} \sum_{i=1}^{n^0} (Y_i^0 - \hat{Y}_i^0)^2} \text{ and} \\ \text{Theil's U statistic} &= \frac{RMSE}{\sqrt{\frac{1}{n^0} \sum_{i=1}^{n^0} (Y_i^0)^2}} \end{aligned}$$

where the latter measure removes the units of measurement (and hence potential scaling problems).