ECON 5340 Class Notes
Chapter 6. Inference 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. Non-nested hypotheses test will be covered later.

Next, we proceed with two alternative (yet equivalent) approaches to testing linear restrictions.

2 Testing Linear Restrictions

2.1 Unrestricted Approach

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: $\beta_1 = 0$

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

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

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

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

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

2.1.1 Motivating the Test Statistic

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

- $E(Rb) = R\beta$.
- $var(Rb) = E[(Rb - E[Rb])(Rb - E[Rb])'] = E[R(b-\beta)(b-\beta)'R'] = Rvar(b)R' = \sigma^2 R(X'X)^{-1}R'$.
- $Rb \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$.
- If H_0 is true, $Rb - q = m \sim N(0, \sigma^2 R(X'X)^{-1}R')$.
- From (Greene, Theorem B.11), $(Rb-q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(Rb-q) = m'var(m)^{-1}m \sim \chi^2(J)$.
- Replace σ^2 with $s^2 \Longrightarrow F = \frac{(Rb-q)'(\sigma^2 R(X'X)^{-1}R')^{-1}(Rb-q)/J}{(n-k)s^2/\sigma^2(n-k)}$ is the ratio of two independent chi-squared random variables. Therefore, we know that

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

2.1.2 Examples Continued

- H_0: $\beta_1 = 0$

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

or taking square roots...

$$t = \frac{b_1}{\sqrt{\widehat{var}(b_1)}} = \frac{b_1}{se(b_1)} = \frac{b_1}{s_{b1}} \sim t(n-k).$$

- 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...

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

2.1.3 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 analagous 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.

2.2 Restricted Approach

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.

2.2.1 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

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} b_* \\ \lambda \end{bmatrix} = \begin{bmatrix} X'Y \\ q \end{bmatrix} \Longrightarrow \begin{bmatrix} b_* \\ \lambda \end{bmatrix} = \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'Y \\ q \end{bmatrix}. \tag{2}$$

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

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

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., $R\beta = q$).

2.2.2 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

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

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^2s

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

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.

2.3 Gauss Example

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

$$\ln(wage_i) = \beta_1 + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 grade_i + \beta_5 married_i + \beta_6 union_i + \epsilon_i.$$

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 Gauss example 6.1 to calculate the statistics and complete the tests.

3 Testing Nonlinear Restrictions

Testing nonlinear restrictions requires a fairly simple modification of the linear testing procedure. Consider testing the null hypothesis H_0: $C(\beta) = q$, where $C()$ is a possibly nonlinear function in β. The first step is to linearize the restriction using a first-order Taylor series approximation around β

$$C(\hat{\beta}) \simeq C(\beta) + \tilde{C}'(\hat{\beta} - \beta)$$

where $\tilde{C} = \partial C(\beta)/\partial \beta$. The variance of $C(\hat{\beta})$ is then approximately

$$var(C(\hat{\beta})) \simeq \tilde{C}' var(\hat{\beta}) \tilde{C}.$$

This implies that we can form the Wald-like statistic

$$W = (C(\hat{\beta}) - q)'(var(C(\hat{\beta})))^{-1}(C(\hat{\beta}) - q) \overset{asy}{\sim} \chi^2(J).$$

4 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

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

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).

4.1 Gauss Example

In this example, we will predict the wage of a particular male and calculate the 95% confidence interval for the true wage. The male is 34 years old, has 22 years of schooling, is married and is not a union member (i.e., $X^0 = (1, 34, 34^2, 22, 1, 0)$). See Gauss example 6.2 for more details.