Hypothesis Tests and Model Selection

Henrique Veras

PIMES/UFPE

Intro

Purposes of linear regression models:

1. Estimation
2. Prediction
3. Hypothesis testing

Hypothesis Testing Methodology

We start by providing a general framework for conducting hypothesis testing.

The general approach to testing a hypothesis is to formulate a statistical model that contains the hypothesis as a restriction on the model's parameters.

Consider a model of investment:

$$\ln I_t = \beta_1 + \beta_2 i_t + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t$$

This theory states that investment is a function of the *nominal* interest rate, inflation rate, income, and factors that evolve together over time.

Let's call this model 1.

Nested Models

An alternative theory suggests that investors care about *real* interest rates:

$$\ln I_t = \beta_1 + \beta_2(i_t - \Delta p_t) + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t$$

This theory, however, does not have a testable implication for our model.

If, instead, we impose a **restriction** on model 1 stating that "investors care *only* about real interest rates, the equation would become model 2 below:

$$\ln I_t = \beta_1 + \beta_2(i_t - \Delta p_t) + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t$$

In this case, model 2 is the same as model 1 with the added restriction that $\beta_2 = -\beta_3$ or $\beta_2 + \beta_3 = 0$.

Nested Models

We, then, say that models 1 and 2 are nested.

The parameter space of the *restricted* model (model 2) is smaller than the parameter space of the *unrestricted* model (model 1).

Size, Power, and Consistency of a Test

Prior to conducting a hypothesis testing, such as whether investors care only about real interest rates, we define a rejection (and acceptance) region, which indicates whether we have gathered enough evidence against a given theory.

For instance, supoose we estimate

$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \varepsilon$$

and impose a restriction $\beta_3 = 0$. The classical hypothesis testing procedure suggests to define two (mutually exclusive) hypotheses:

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

Size, Power, and Consistency of a Test

In performing a test, there are two ways in which we can incur in a error (given the randomness of process):

1. Type I error: the null hypothesis is incorrectly rejected
2. Type II error: the null hypothesis is incorrectly retained.

The probability of a type I error is called the **size of the test**.

1 – Prob(Type II error) is called the **power of the test**. This is the probability of correctly rejecting the null hypothesis.

A test procedure is **consistent** if its power goes to 1 as the sample size goes to infinity. If a test is based on consistent estimators, then the test itself is consistent.

Three Approaches to Testing Hypotheses

1. Wald tests: based on estimation of the unrestricted model – the test measures how close the estimated unrestricted model is to the hypothesis restrictions;

2. Fit based tests: comparing R^2s when we estimate the unrestricted model and when we incorporate the restriction.

3. Lagrange multiplier tests: based on correlation between variables from restriction and the residuals from the regression of restricted model.

General Approach

The general linear hypothesis is a set of J restrictions on the linear regression model

$$y = \mathbf{X}\beta + \varepsilon$$

The restrictions are

$$r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1K}\beta_K = q_1$$

$$r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2K}\beta_K = q_2$$

$$\vdots$$

$$r_{J1}\beta_1 + r_{J2}\beta_2 + \cdots + r_{JK}\beta_K = q_J$$

In matrix form:

$$\mathbf{R}\beta = \mathbf{q}$$

General Approach

The hypotheses implied by the restrictions are written

$$H_0 : \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$$

$$H_1 : \mathbf{R}\beta - \mathbf{q} \neq \mathbf{0}$$

Example: Assume you have a model with an intercept and 5 independent variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

What are **R** and **q** in each case below?

1. $\beta_3 = 0$
2. $\beta_2 = \beta_3$
3. $\beta_1 + \beta_2 + \beta_3 = 1$
4. $\beta_1 = 0,\ \beta_2 = 0,\ \beta_3 = 0$
5. $\beta_1 + \beta_2 = 1,\ \beta_3 + \beta_4 = 1,\ \beta_5 + \beta_6 = 0$
6. All coefficients in the model except the intercept are zero

Some Information about $\mathbf{R}$

1. Full row rank ($J \leq K$) but we assume strict inequality
2. Rows must be linearly independent

Wald Tests Based on the Distance Measure

Let's start by testing a hypothesis about a coefficient.

Suppose we are interested in testing

$$H_0 : \beta_k = \beta_k^0$$

The **Wald distance** of a coefficient estimate from a hypothesized value is the distance measured in standard deviation units

$$W_k = \frac{b_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}}$$

W_k follows a standard normal distribution if the null hypothesis is true.

Notice, however, that σ^2 is unknown. We estimate it with s^2:

$$t_k = \frac{b_k - \beta_k^0}{\sqrt{s^2 S^{kk}}}$$

Wald Tests Based on the Distance Measure

To construct a testing procedure, we define the rejection/acceptance region by determining the *confidence* with which we would like to draw the conclusion.

Using a 95% confidence interval, we can say that

$$\text{Prob}[-t^*_{(1-\alpha/2),[n-K]} < t_k < t^*_{(1-\alpha/2),[n-K]}]$$

That is, if the null hypothesis is true, finding a value of t_k outside this range is highly unlikely.

It must be that the mean value β_k^0 is not the one hypothesized.

The F Statistic

We now consider testing a set of J linear restrictions

$$H_0 : \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$$
$$H_1 : \mathbf{R}\beta - \mathbf{q} \neq \mathbf{0}$$

Given the LS estimator $\mathbf{b}$, we are interested in the **discrepancy vector** $\mathbf{Rb} - \mathbf{q} = \mathbf{m}$.

Our question is how far $\mathbf{m}$ is from zero.

Notice that $\mathbf{m}$ is normally distributed, since it is a linear combination of $\mathbf{b}$, which is normally distributed.

The F Statistic

Let's find **m**'s expected value and variance:

$$E[\mathbf{m}|\mathbf{X}] = \mathbf{R}E[\beta|\mathbf{X}] - \mathbf{q} = 0$$
$$Var[\mathbf{m}|\mathbf{X}] = \mathbf{R}Var[\beta|\mathbf{X}]\mathbf{R}' = \mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$$

The test statistic is given by

$$\begin{aligned} W &= \mathbf{m}'Var[\mathbf{m}|\mathbf{X}]^{-1}\mathbf{m} \\ &= (\mathbf{Rb} - \mathbf{q})'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{Rb} - \mathbf{q}) \sim \chi^2[J] \end{aligned}$$

Thus, W follows a chi-squared distribution with J degrees of freedom if the null hypothesis is correct.

The F Statistic

By estimating σ^2 with s^2 and dividing the result by J, we obtain a usable F statistic with J and $n - K$ degrees of freedom:

$$F[J, n-K|X] = \frac{(\mathbf{Rb}-\mathbf{q})'[\mathbf{R}s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb}-\mathbf{q})}{J}$$

If we are interested in testing *one* linear restriction of the form

$$H_0 : r_1\beta_1 + r_2\beta_2 + \cdots + r_K\beta_K = \mathbf{r}'\beta = q,$$

the F statistic is

$$F[1, n-K] = \frac{\sum (r_j\beta_j - q)^2}{\sum_j \sum_k (r_j r_k Est.Cov(b_j, b_k))}$$

If the test if for $\beta_j = q$, the F statistic is then

$$F(1, n-K) = \frac{(b_j - q)^2}{Est.Var(b_j)}$$

The F Statistic

Consider an alternative approach: find the sample estimate of $\mathbf{r}'\beta$

$$r_1 b_1 + r_2 b_2 + \cdots + r_K b_K = \hat{\mathbf{q}}$$

The test statistic is

$$t = \frac{\hat{q} - q}{se(\hat{q})}$$

where $Est.Var(\hat{q}|\mathbf{X}) = \mathbf{r}'[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{r}$

It follows, therefore, that for testing a single restriction, the t statistic is the square root of the F statistic that would be used to test that hypothesis.

Tests based on the fit of the regression

Another approach to hypothesis testing is asking whether choosing some other value for the slopes of the regression leads to a significant loss of fit.

That is, we can compare R^2s from both unrestricted and restricted models to test for a given hypothesis.

The Restricted LS Estimator

If we explicitly impose the restrictions of the general linear hypothesis, in the regression, we solve

$$Min_{\mathbf{b}_0} S(\mathbf{b}_0) = (\mathbf{y} - \mathbf{X}\mathbf{b}_0)'(\mathbf{y} - \mathbf{X}\mathbf{b}_0)$$

subject to $\mathbf{R}\mathbf{b}_0 = \mathbf{q}$
We can show that the solution to the restricted LS coefficient $\mathbf{b}_*$ is given by

$$\mathbf{b}_* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

$$\lambda_* = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

The Restricted LS Estimator

Moreover, Greene and Seaks (1991) show that $Var[\mathbf{b}_*|X]$ is simply σ^2 times the upper left block of $\mathbf{A}^{-1}$:

$$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{I} - \mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}$$

Thus,

$$\begin{aligned} Var[\mathbf{b}_*|X] &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \\ &= Var[\mathbf{b}|X] - \text{a nonnegative matrix} \end{aligned}$$

One intuition for the difference $Var[\mathbf{b}|X] < Var[\mathbf{b}_*|X]$ is the value of the new information brought by the restrictions.

The Loss of Fit from Restricted LS

Let's first consider the case of a restriction on a single coefficient.

Recall from Theorem 3.6, which shows the change in the R^2 following the introduction of a variable z to the model:

$$R^2_{Xz} = R^2_X + (1 - R^2_X) r^{*2}_{yz}$$

A convenient result is that

$$r^{*2}_{yz} = \frac{t_z^2}{t_z^2 + (n - K)},$$

where t_z^2 is the square of the t ratio for testing the hypothesis that the coefficient on z is zero in the multiple regression of $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{z}$.

The Loss of Fit from Restricted LS

Substituting one equation into the other, we obtain

$$t_z^2 = \frac{(R_{Xz}^2 - R_X^2)/1}{(1 - R_{Xz}^2)/(n-K)}$$

As we saw before, $t^2(n-K) = F[1, n-K]$. Then, the statistic above follows an F distribution with 1 and $n-K$ d.f.

The Loss of Fit from Restricted LS

Let's now find a general representation for the above procedure.

First, notice that the fit of the restricted LS cannot be better than the restricted solution:

Let $\mathbf{e}_* = \mathbf{y} - \mathbf{X}\mathbf{b}_*$. Adding and subtracting $\mathbf{X}\mathbf{b}$, we can write

$$\mathbf{e}_* = \mathbf{e} - \mathbf{X}(\mathbf{b}_* - \mathbf{b})$$

The new sum of squared deviations is

$$\mathbf{e}_*'\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\mathbf{b}_* - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{b}_* - \mathbf{b}) \geq \mathbf{e}'\mathbf{e}$$

The Loss of Fit from Restricted LS

We define the loss of fit as the difference in the sum of squared residuals from the restricted model and the sum of squared residuals from the restricted model:

$$\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e} = (\mathbf{b}_* - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{b}_* - \mathbf{b})$$

Recall:

$$\mathbf{b}_* - \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

Thus, substituting this in the previous expression, we find

$$\mathbf{e}_*'\mathbf{e}_* = (\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

The Loss of Fit from Restricted LS

This expression appears in the numerator of the F statistic. Inserting the remaining parts:

$$F[J, n-K] = \frac{(\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e})/J}{\mathbf{e}'\mathbf{e}/(n-K)}$$

Dividing both numerator and denominator by $\sum_i (y_i - \bar{y})^2$, we have

$$F[J, n-K] = \frac{(R^2 - R_*^2)/J}{(1-R^2)/(n-K)}$$

This means that large values of F give evidence against the validity of the hypothesis.

Testing the Significance of the Regression

A relevant question is whether the regression equation as a whole is significant (that's question 6 from our previous example).

This is a special case with $R_*^2 = 0$, so that the F statistic become

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

Lagrange Multiplier Tests

Recall the solution for the Lagrange multiplier:

$$\lambda_* = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

Lagrange multiplier tests rely on the idea that testing whether $\lambda_* = 0$ should be equivalent to testing $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ (why?)

The chi-square test statistic is computed as

$$W_{LM} = (\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

A feasible version of this test statistic is obtained by using s^2 (from the restricted model) in place of the unknown σ^2.

Large Sample Tests

Without the normality assumption, the exact distributions of the test statistics that we calculated before depend on the data and the parameters.

They are no longer t, F, and chi-squared.

Recall:

THEOREM 4.3 Asymptotic Distribution of b with IID Observations
If $\{\varepsilon_i\}$ are independently distributed with mean zero and finite variance σ^2 and x_{ik} is such that the Grenander conditions are met, then

$$\mathbf{b} \overset{a}{\sim} N\left[\boldsymbol{\beta}, \frac{\sigma^2}{n}\mathbf{Q}^{-1}\right]. \qquad \textbf{(4-33)}$$

and plim $s^2 = \sigma^2$

Large Sample Tests

The test statistic for testing some hypothesis about a given β_k is

$$t_k = \frac{\sqrt{n}(b_k - \beta_k^0)}{\sqrt{s^2(\mathbf{X}'\mathbf{X}/n)^{-1}_{kk}}}$$

Notice that this is the same as before, as $\sqrt{n}$ cancels out.

Without normality, the exact distribution of this statistic is unknown.

The denominator of t_k converges to $\sqrt{\sigma^2 \mathbf{Q}^{-1}_{kk}}$, where $\mathbf{Q} = \text{plim } \frac{\mathbf{X}'\mathbf{X}}{n}$.

Large Sample Tests

Therefore,

$$\tau_k = \frac{\sqrt{n}(b_k - \beta_k^0)}{\sqrt{\sigma^2 \mathbf{Q}_{kk}^{-1}}}$$

Thus, τ_k follows (asymptotically) a standard normal distribution under the null hypothesis.

Large Sample Tests

As for the F test, we can multiply numerator and denominator by σ^2 and rearreange it to find

$$F = \frac{(\mathbf{Rb} - \mathbf{q})'[\mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})}{J(s^2/\sigma^2)}$$

If F has a limiting distribution, then it is the same as the limiting distribution of

$$\begin{aligned} W^* &= \frac{1}{J}(\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\sigma^2/n)\mathbf{Q}^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q}) \\ &= \frac{1}{J}(\mathbf{Rb} - \mathbf{q})'[Asy.Var(\mathbf{Rb} - \mathbf{q})]^{-1}(\mathbf{Rb} - \mathbf{q}) \end{aligned}$$

Recall that $Var[\mathbf{Rb} - \mathbf{q}|\mathbf{X}] = \mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$.

This expression is $1/J$ multiplied by a Wald statistic, based on the asymptotic distribution.

Limiting Distribution of the Wald Statistic

THEOREM 5.1 Limiting Distribution of the Wald Statistic
If $\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \boldsymbol{\Sigma}]$ *and if* H_0: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ *is true, then*

$$W = (\mathbf{Rb} - \mathbf{q})'\{\mathbf{R}\boldsymbol{\Sigma}\mathbf{R}'\}^{-1}(\mathbf{Rb} - \mathbf{q}) = JF \xrightarrow{d} \chi^2[\mathrm{J}].$$

Testing Nonlinear Restrictions

Suppose now we are interested in testing a hypothesis that involves a nonlinear function of the regression coefficients:

$$H_0 : c(\beta) = \mathbf{q}$$

The test statistic is

$$z = \frac{c(\hat{\beta}) - q}{\text{estimated st. error}}$$

The numerator can be easily obtained. To find the estimated variance, however, we rely on a Taylor series approximation around the true parameter β:

$$c(\hat{\beta}) \approx c(\beta) + \left(\frac{\partial c(\beta)}{\partial \beta}\right)' (\hat{\beta} - \beta)$$

Thus,

$$Var[c(\hat{\beta})] \approx \left(\frac{\partial c(\beta)}{\partial \beta}\right)' Asy.Var[\hat{\beta}] \left(\frac{\partial c(\beta)}{\partial \beta}\right)$$

.

Testing Nonlinear Restrictions

We use $\mathbf{b}$ as an estimator of β and estimate $c(\beta)$ with $c(\mathbf{b})$.

According to the Slutsky theorem, if plim $b = \beta$, then plim $c(b) = c(\beta)$.

Choosing Between Nonnested Models

Suppose that, for example, we are now interested in testing whether a linear or a log-linear model is appropriate, or determining which of two possible sets of regressors is more appropriate.

We are interested in comparing two competing linear models:

$$H_0 : \mathbf{y} = \mathbf{X}\beta + \varepsilon_1$$

$$H_1 : \mathbf{y} = \mathbf{Z}\gamma + \varepsilon_2$$

Notice how the models are nonnested. In this case, how do we perform a test? Is the usual size of the test in this case useful?

An Encompassing Model

We say Model 0 *encompasses* Model 1 if the features of Model 1 can be explained by model 0, but the reverse is not true.

Let $\bar{\mathbf{X}}$ denote the set of variables in $\mathbf{X}$ that are not in $\mathbf{Z}$ and define $\bar{\mathbf{Z}}$ likewise with respect to $\mathbf{X}$.

Moreover, let $\mathbf{W}$ be the set of variables that the models have in common.

Then, H_0 and H_1 can be combined in a supermodel:

$$\mathbf{y} = \bar{\mathbf{X}}\bar{\beta} + \bar{\mathbf{Z}}\bar{\gamma} + \mathbf{W}\delta + \varepsilon$$

An Encompassing Model

H_0 can be rejected if it is found that $\bar{\beta} = 0$ and H_1 likewise for $\bar{\gamma} = 0$.

This test can be performed applying a conventional F test.

Problems with this approach:

1. We cannot test whether β or γ are zero. why?
2. This compound model can have an extremely large set of regressors.

An Alternative Approach

Suppose H_0 is correct. Then, we can estimate γ by regressing $\mathbf{y}$ on $\mathbf{Z}$, which yields an estimator vector, say, $\mathbf{c}$.

If we regress $\mathbf{X}\beta$ on $\mathbf{Z}$, we should find the exact same coefficients, say $\mathbf{c}_0$ (assuming ε's are random noises). Why?

A test for the proposition that Model 0 encompasses Model 1 is whether $E[\mathbf{c} - \mathbf{c}_0] = 0$

Comprehensive Approach – The J Test

The J test, proposed by Davidson and MacKinnon (1981) suggests an alternative:

$$\mathbf{y} = (1 - \lambda)\mathbf{X}\beta + \lambda\mathbf{Z}\gamma + \varepsilon$$

A test of $\lambda = 0$ would be against H_1. However, this cannot be separately estimated.

The J test consists of estimating γ by LS regression of $\mathbf{y}$ on $\mathbf{Z}$ followed by a LS regression of $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{Z}\gamma$

Asymptotically, $\hat{\lambda}/se(\hat{\lambda})$ is distributed as standard normal.

A Specification Test

Finally, the idea now is to consider a particular null model and alternatives that are not explicitly given in the form of restrictions on the regression equation.

Ramsey's (1967) RESET test:

$$\begin{aligned} H_0 : \mathbf{y} =& \mathbf{X}\beta + \varepsilon \\ H_1 : \mathbf{y} =& \mathbf{X}\beta + \text{higher order powers of } x_k \text{ and other terms} + \varepsilon \end{aligned}$$

Approach: add second- and third-order powers of x_k and cross-products of the regressors in H_1.

Issue: With a large number of regressors, these additional terms might increase the number of estimated parameters at large rates.

A Specification Test

Solution: use LS predictions

1. Fit the null model
2. Include higher order and other terms

Table of Contents