Econometrics notes round 1.

**Chapter 4: hypothesis testing in linear regression model**

We can test to what extend do we trust that is a reliable estimate of the true parameter. There are two ways of doing so: hypothesis testing and confidence intervals.

For hypothesis testing, we wish to test a null hypothesis that if the parameter estimates are of specific values, e.g., , which we call for short. In order to test if is true, we need to calculate a test statistic, which is a r.v that has a known distribution if is true and some other distributions if is false. On the other hand, if the value of the test statistic is an extreme value which would rarely be encountered by chance under the null hypothesis, then the test does provide evidence against . If we believe this evidence is strong/convincing enough, we reject .

Imagine that we concern the population mean from which the sample was drawn. Suppose the DGP is:

,

(4.01)

Thus, the LS estimator of and its variance, given a sample of size *n*, are:

(4.02)

(4.02) can be calculated from on the formula for OLS estimator and its variance where is just an *n*-vector of 1’s.

Now we may make two assumptions about (4.01). The first is that is normally distributed, and the second is that is known. Thus, we can construct a test statistic:

It turns out that, as is an unbiased estimator of , if (e.g., ) is true, follows a standard normal distribution, e.g., . Also, we have unity variance for :

Note that must be normally distributed because is a linear combination of each of which is normally distributed corresponding to the error terms (e.g., 4.02). We can also see this in the general form:

where is the vector of test statistics for the estimate parameters.

For every , we have at least one alternative hypothesis, e.g., . This is what we are testing against, e.g., under , does not follow a standard normal distribution. Suppose the true value of is , thus, we have the estimate as , (e.g., see (3.05)) where has a mean of zero and variance of , and is normally distributed as the error terms. Thus, we can find out that:

As we know that , we can rewrite it as:

(4.04)

It is obvious that and , thus, .

(4.04) suggests that, when *n* is sufficiently large, we would expect the mean of to be positive and large if , and large and negative if . Thus, we reject whenever is sufficiently far from zero.

If we want to test an and , we must conduct a two-tail test and reject whenever is far from zero. In contrast, if we want to test an and , we must perform a one-tail test and reject whenever is large and positive (and reject whenever is large and negative if we want to test an and ). In general, test of equality restrictions are two-tail tests, and test of inequality restrictions are one-tail tests.

“Since is a r.v and can takes on any value on the real line, no value of is absolutely incompatible with the null hypothesis, and thus we can never be absolutely certain that the hull hypothesis is false. One way to deal with this is to decide in advance a **rejection rule**, according to which we choose to reject the null if and only if the value of fall into the **reject region** of the rule. For two-tailed tests, the appropriate reject region is the union of two sets, one containing all values of greater than some positive value, the other all values of less than some negative value. For a one-tailed test, the rejection region would consist of just one set, containing either sufficiently positive or sufficiently negative values of , according to the sign of the inequality we wish to test.”

A test statistic combined with a reject rule is sometimes called a **test**. If a test incorrectly leads us to reject a hull that is true, we are said to make a **Type I error**. The probability of making such an error is the probability that, under the null, falls into the rejection region. This probability is also called the **level of significance** of the test. Popular values are .05 and .01.

So far, we assumed that the distribution of the test statistic is known exactly, so that we have what is called an **exact test**. In reality, the distribution of the test statistic is only known approximately. In this case, we refer to what we call ‘**nominal** level of the test’, e.g., the probability of making a Type I error according to whatever approximate distribution we are using to determine the rejection region – this is different from the **actual rejection probability** (which we usually do not know in practice).

The probability that a test rejects a null is called the **power** of the test. If the null is true, the power of an exact test equals to the significance level. In general, power depends on precisely the DGP and on the sample size, e.g., in (4.04), we have:

Thus, the power is proportional to and the square root of the sample size, and inversely proportional to .

If we look at (4.04), we notice that values of different from zero would move the probability mass of distribution away from the centre of . As a result, with , we would be more likely to have the calculated test statistic falling into the reject region of .

Wrongfully failing to reject a false null is called making a **Type II error**. Its probability is 1 minus the power of test.

In order to construct the reject region, we need to calculate the critical value associated with the significance level, e.g., (see details in p126 in Davison and Mackinnon, 2003).

**P values**

The result of a test is only yes or no – a more sophisticated /informative way is to decide if we reject the null is to calculate the p-value or marginal significance level associated with the observed test statistic . One advantage for using p-value is to preserve information. e.g., two large test statistics may both lead to the rejection of the null, but when they are transformed to p-values, we may observe that one to be more confidence (e.g., of more evidence) for the rejection of the null.

**[see contents for common distributions in chapter 0]**

**4.4 exact tests in the classical normal linear model**

So far, we show how we calculate a test statistic which is distributed as under the null. Tests based on this statistic are ‘exact’ because we assume that we know the exact distribution of the test statistic. In practice, this is rare. One example is where we test linear restrictions on classical normal linear model. e.g.,

(4.20)

where is an *n* x k matrix of regressors. We assume the error terms are independently and normally distributed. We also assume exogeneity, e.g., .

It can be proven that we can find a test statistic for the parameter of a specific regressor, say, , which follows a student’s t-distribution.

**4.5 large sample tests in linear regression models**

The *t* test and *F* test are derived under strong assumptions of normally distributed errors and exogeneity and we call them ‘exact tests’. In practice, these assumptions may not hold but we can still calculate those corresponding test statistics – they do not follow their namesake distributions but tend to do so when the sample size gets large. We call these tests (e.g., the same tests without the assumptions being hold but based on large samples) **asymptotic** tests.

**Law of large numbers (LLN)**

There are two types of results on which the asymptotic theory is based. The first is LLN. Suppose that we have a r.v which is the average of *n* random r.v, e.g.,

where are independent r.v, each with its own bounded finite variance and with a common mean . Then the LLN assures us that, when *n* approaches infinity, tends to be .

**Central limit theorem (CLT)**

The other type of results on which the asymptotic theory is based is CLT which suggests that in many cases times the sum of *n* centred r.v approximately follows a normal distribution when *n* is sufficiently large.

Suppose that r.v are independently and identically distributed with mean and variance , we can have the following quantity:

The quantity of is asymptotically distributed as . This is called the Lindeberg-Levy CLT. There are other forms of CLTs which usually have weaker assumptions (e.g., compared to independence and identical distribution, they only require that the r.v are not too much dependence or too much heterogeneity).

We can write the multivariate version of CLTs. e.g., suppose that we have a sequence of uncorrelated random *m*-vector , for some fixed *m*, with . Then the multivariate CLT tells us that:

where is multivariate normal, and each is an *m* x *m* matrix.

**Asymptotic tests**

The *t*-test and *F*-test are asymptotically valid under much weaker conditions compared to those which were required for them to have namesake distributions in finite samples. Suppose that we have:

We may not know the distribution of the error terms and we may have lagged dependent variables as regressors. It can be proven that as long as we can make the following two assumptions, then the *t*-test and *F*-test are asymptotically valid (even the assumptions of exogeneity does not hold and the distribution of the error term is unknown). The two assumptions are:

From the point of view of the explanatory variables, the explanatory variables are said to be **predetermined**. From the point of view of the error terms, the error terms are also called **innovations**. An innovation is a r.v of which the mean is zero conditional on the information in the explanatory variables, and so knowledge of the values taken by the latter is of no use in predicting the man of the innovation.

Thus, we can still use *t*-test and *F*-test without the assumption of exogeneity and known distribution for the error term as long as we make the two assumptions (e.g., mean independence and variance independence).

**The *t* test with predetermined regressors**

If we relax of the assumption of exogeneity, the analysis becomes more sophisticated. However, we can still use the *t*-test with asymptotic properties.

**Asymptotic F test**

Whatever distribution we use, the p values we have are approximate – they are different from those obtained from exact tests and we do not know how accurate or inaccurate they are when we have finite sample. If we decide to use a nominal level of for a test, we reject if the approximate p value is less than the significance level (e.g., ). We may sometimes over reject or under reject.

**4.6 simulation-based tests**

A pivotal variable is a r.v whose distribution does not depend on the unknown parameters. e.g., suppose we have . We can form a score as:

Although we need to know the value of to calculate score, we know that always follow a standard normal distribution, e.g., . The distribution which follows does not depend on the unknown parameters (e.g., of ).

There are many pivotal random variables in statistics, e.g.,

And we know that , follows a standard normal distribution which does not dependent the unknown parameters (e.g., .

Until now, we have introduced the test statistic which we assume to follow a specific (known at least approximately) distribution under the null. In fact, we also implicitly assume that, when the null is not true, the test statistic follows an alternative distribution. However, in practice, when the null is not true, there would possibly be many alternative GDPs which all could generate the data and as a result, the test statistic may not only follow one alternative distribution but many and we do not know which distribution does the test statistic follows.

For example, the test statistics for the *t*-test and the *F*-test depends on the distribution of the error term. That is, the distributions of their test statistics depend on the distribution of the error term. However, fortunately, the test statistics for the *t*-test and the *F*-test are pivotal asymptotically.

**Simulated p-values**

**Random number generator**

**Bootstrap tests**