Likelihood Ratio Tests

Previously, we have employed a program (Jtest) to test an hypothesis involving many parameters (multiple constraints). This test, which involves comparing parameter estimates with hypothesized values, is known as a Wald test. We also used the Jtest program to compare two different estimators in a Hausman test. There is an alternative method for conducting joint tests, which can be implemented easily both from within and outside of Gauss. In this handout, we will first discuss this method in general and then see how it works in the familiar context of a linear regression model.

1 Motivation

We have already discussed that in large samples it is desirable to estimate models by selecting estimates so as to maximize the likelihood function. Intuitively, since the true parameter values generated the data, in large samples, the data are most likely to have been generated by the true parameter values. According, if we view the likelihood function as the probability of the data, as previously discussed it should seem reasonable that we should select estimates so as to maximize the likelihood.

With estimates of the parameters in a model selected to maximize the likelihood, it is natural to employ the maximized likelihood as a measure of how well the model fits the data. For example, suppose that we postulate a model, and select parameters to maximize the likelihood (the "probability of the data"). However, suppose that the maximized likelihood is quite small. In other words, while we selected parameter estimates to maximize it, the probability of the data coming from the postulated model is small. In this case, one would naturally say that the model provides a "poor" fit to the data. On the other hand, if the maximized likelihood is very large, then the data are very likely to have come from the postulated model. In this case, the model provides a "good" fit to the data.

With the maximized likelihood as our measure of data-fit, it should seem intuitive that we might compare models on the basis of how well they fit the data. In this handout, we will develop a test based on this principle. While this principle will apply to a very wide range of models, here we will examine it in terms of the probit model . With $V \equiv X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \beta_4$, write this model as:

$$Q = \begin{cases} 1: & V + u > 0 \\ 0: & V + u \leq 0 \end{cases}$$

Here, for example, X_1 might be the price of the product, X_2 the price of a close substitute, and X_3 an individual's income, etc.

Employing the probit model above, suppose that we want to test whether or not X_2 and X_3 jointly belong in the model. Here, we are testing:

$$H_0 : \beta_2 = 0,\ \ \beta_3 = 0 \text{ vs. } H_1 : Not\ H_0.$$

To formulate a test, suppose that we first maximize the likelihood under the null hypothesis, H_0, by imposing the constraints under the null hypothesis. To do this, we would maximize the likelihood with X_2 and X_3 left out of the model. For this constrained case, write the likelihood (omitting X_2 and X_3 from the model) as:

$$L_0 \left(\beta_1, \beta_4\right) \equiv Likelihood \text{ under } H_{0.}$$

With $\left(\hat{\beta}_1, \hat{\beta}_4\right)$ as the maximum likelihood estimates under H_0, the maximized likelihood under H_0 is given as:

$$\hat{L}_0 \equiv L_0 \left(\hat{\beta}_1, \hat{\beta}_4\right)$$

Similarly, for the unconstrained case (where all variables are included) , define:

$$L_1 \left(\beta_1, \beta_2, \beta_3, \beta_4\right) \equiv Likelihood \text{ under } H_1$$

With $\left(\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4\right)$ as the maximum likelihood estimates under H_1, the maximized likelihood under $H_1 is$ given as:

$$\hat{L}_1 \equiv\ L_0 \left(\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4\right).$$

In maximizing a likelihood, it is equivalent to maximize the log likelihood instead. With the likelihood being a product of terms, its log will be more tractable as a sum of terms. In testing, it also terns out to be easier to work with log likelihoods. Define the likelihood-ratio test-statistic as:

$$LRT = 2 * \left[Ln\left(\hat{L}_1\right) - Ln\left(\hat{L}_0\right)\right] = 2 * Ln\left(\hat{L}_1 / \hat{L}_0\right).$$

Here, LRT has two properties that we require for test statistics. First, it naturally measures the difference in how the two models fit the data, where the likelihood is used to measure data-fit. We would then reject H_0 if LRT is very large, which would mean that the model fits the data much better under H_1 than under H_0. As a second required property, we must know the distribution of the test statistic under the null-hypothesis. Here, an amazing result is true in large samples. Namely, under the null hypothesis and when the sample size is large, LRT has a chi-squared distribution with r degrees of freedom. The degrees of freedom, r, is equal to the number of restrictions under the null hypothesis (2 in the example above).

With the rejection region now given by:

$$LRT > c,$$

we could now find the P-value, which is that significance level such that c=LRT in the sample. Alternatively, at any given significance level, we could look of c in a $\chi^2(r)$ table. In the remainder of this handout, we will return to the more familiar linear model and show that likelihood ratio tests in such models are equivalent to the same types of tests you have done previously.

2 Likelihood Ratio Tests: Linear Models

The tests outlined above apply whenever we estimate models by maximum likelihood estimation. In this manner, these tests apply to the probit, ordered, and censored models that we have examined. To further motivate and explain these tests, in this section we will apply them to the linear model. Consider the following demand model without censoring:

$$Q \equiv X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \beta_4 + u,$$

where the error, u, is distributed as N(0, σ^2). In what follows, for simplicity we will first assume that the error variance, σ^2, is known. Subsequently, we will argue that in large samples it does not matter whether or not this variance is known.

To simplify the forms of the likelihoods in this case, it is useful to employ sums of squared residuals under the null and alternative hypotheses.

Accordingly, let:

$$
\begin{aligned}
R_0 &\equiv \sum \left[Q - X_1\hat{\beta}_1 + \hat{\beta}_4\right]^2 \\
R_1 &\equiv \sum \left[Q - X_1\bar{\beta}_1 + X_2\bar{\beta}_2 + X_3\bar{\beta}_3 + \bar{\beta}_4\right]^2
\end{aligned}
$$

In the linear model with normal errors, OLS estimates (which by definition minimize a sum of squared residuals) are also maximum likelihood estimates. From discussions we have had in class, we may then write the maximized likelihoods as:

$$
\begin{aligned}
\hat{L}_0 &= C \exp\left(-\frac{1}{2}\left[R_0/\sigma^2\right]\right) \\
\hat{L}_1 &= C \exp\left(-\frac{1}{2}\left[R_1/\sigma^2\right]\right),
\end{aligned}
$$

where C is a constant. **You should make sure that you understand why this is the case.** The statistic LRT defined above is given in this case as:

$$
\begin{aligned}
LRT &= 2 * \left[Ln\left(L_1^*\right) - Ln\left(L_0^*\right)\right] \\
&= 2 * \left[\left(-\frac{1}{2}\left[R_1/\sigma^2\right]\right) - \left(-\frac{1}{2}\left[R_0/\sigma^2\right]\right)\right] \\
&= \left[R_0 - R_1\right]/\sigma^2, \quad R_0 \succeq R_1
\end{aligned}
$$

We reject the null hypothesis if LRT is large, which means that the sum of squared residuals under H_0 is much larger than that under H_1. In this case, the model would fit the data much better under H_1, which would lead us to reject H_0. As to the distribution of LRT, it can be shown that:

$$R_0/\sigma^2 \text{ is distributed as } \varkappa^2\left(N - K_0\right),$$

where N is the sample size and K_0 is the number of parameters estimated under H_0. Notice that we subtract one degree of freedom for each parameter that we estimate. Similarly,

$$R_1/\sigma^2 \text{ is distributed as } \varkappa^2\left(N - K_1\right),$$

where K_1 is the number of parameters estimated in the unconstrained case. Again, we subtract one degree of freedom for every parameter we estimation (accounting for the degree of "estimation uncertainty"). Employing a property of Chi-square variables:

$$
\begin{aligned}
& [R_0 - R_1]/\sigma^2 \text{ is distributed as } \varkappa^2(r), \\
r \quad = \quad & (N - K_0) - (N - K_1) = K_1 - K_0.
\end{aligned}
$$

The above discussion assumed that we knew the disturbance variance, σ^2. Denote $\hat{\sigma}^2$ as a consistent estimator for σ^2. In Econ 322, when σ^2 was not known, we replaced it with an estimate. With a minor adjustment (dividing the numerator of LRT by its degrees of freedom), the resulting test statistic had an F distribution. Here, we will assume the sample size is large, in which case $\hat{\sigma}^2$ is probably close to σ^2. As a result, in large samples it can be shown that the above test is unaffected if we $\hat{\sigma}^2$ replaces σ^2 above. We can still use the χ^2 tables; no further adjustments are required.

In large samples, you would find that the F-test would give the same results as the $\varkappa^2(r)$ test described above. Namely, when we reject under one test, we reject under the other. When we fail to reject under one test, we fail to reject under the other. The P-values under both tests are equivalent. It is in this sense that we say that the two tests are equivalent. The F-tets you did in the past may be viewed as versions of the Likelihood-Ratio test.