Economics 241B
Hypothesis Testing: Large Sample Inference

Statistical inference in large-sample theory is based on test statistics whose distributions are known under the truth of the null hypothesis. Derivation of these distributions is easier than in finite-sample theory because we are only concerned with the large-sample approximation to the exact distribution.
In what follows we assume that a consistent estimator of S exists, which we term $\hat{S}$. Recall that $S = E\left(g_t g_t'\right)$, where $g_t = X_t U_t$.

Testing Linear Hypotheses

Consider testing a hypothesis regarding the k-th coefficient β_k. Proposition 2.1, which established the asymptotic distribution of the OLS estimator, implies that under $H_0 : \beta_k = \bar{\beta}_k$,

$$\sqrt{n}\left(B_k - \bar{\beta}_k\right) \underset{d}{\rightarrow} N\left(0, Avar\left(B_k\right)\right) \quad \text{and} \quad \widehat{Avar\left(B_k\right)} \underset{p}{\rightarrow} Avar\left(B_k\right).$$

Here B_k is the k-th element of the OLS estimator B and $Avar\left(B_k\right)$ is the (k,k) element of the $K \times K$ matrix $Avar\left(B\right)$. The key issue here is that we have not assumed conditional homoskedasticity, hence

$$\widehat{Avar\left(B_k\right)} = S_{XX}^{-1}\,\hat{S}\,S_{XX}^{-1},$$

which is the (heteroskedasticity-consistent) robust asymptotic variance. Under the Slutsky result (Lemma 2.4c), the resultant robust t-ratio

$$t_k \equiv \frac{\sqrt{n}\left(B_k - \bar{\beta}_k\right)}{\sqrt{\widehat{Avar\left(B_k\right)}}} = \frac{\left(B_k - \bar{\beta}_k\right)}{SE^*\left(B_k\right)} \underset{d}{\rightarrow} N\left(0,1\right),$$

where the robust standard error is $SE^* = \sqrt{\frac{1}{n}\widehat{Avar\left(B_k\right)}}$. Note this robust t-ratio is distinct from the t-ratio introduced under the finite-sample assumptions in earlier lectures.

To test $H_0 : \beta_k = \bar{\beta}_k$, simply follow these steps:

Step 1: Calculate the robust t-ratio

Step 2: Obtain the critical value from the $N\left(0,1\right)$ distribution

Step 3: Reject the null hypothesis if $|t_k|$ exceeds the critical value

There are several differences from the finite-sample test that relies on conditional homoskedasticity.

- The standard error is calculated in a different way, to accommodate conditional heteroskedasticity.
- The normal distribution is used to obtain critical values, rather than the t distribution.
- The actual (or empirical) size of the test is not necessarily equal to the nominal size. The difference between the actual size and the nominal size is the size distortion. Because the asymptotic distribution of the robust t ratio is standard normal, the size distortion shrinks to zero as the sample size goes to infinity.

To summarize these results, together with the behavior of the Wald statistic let us briefly recall the assumptions required for Proposition 2.1:

Assumption 2.1 (linearity):

$$Y_t = X_t'\beta + U_t \qquad (t = 1, \ldots, n,)$$

where X_t is a K-dimensional vector of regressors, β is a K-dimensional vector of coefficients and U_t is the latent error.

Assumption 2.2 (ergodic stationarity): *The $(K+1)$-dimensional vector stochastic process $\{Y_t, X_t\}$ is jointly stationary and ergodic.*

Assumption 2.3 (predetermined regressors): *All regressors are predetermined, in the sense that they are orthogonal to the contemporaneous error: $E(X_{tk}U_t) = 0$ for all t and $k(= 1, 2, \ldots, K)$. This can be written as*

$$E(g_t) = 0 \qquad \text{where } g_t \equiv X_t \cdot U_t.$$

Assumption 2.4 (rank condition): *The $K \times K$ matrix $E(X_t X_t')$ is nonsingular (and hence finite). We denote this matrix by Σ_{XX}.*

Assumption 2.5 (g_t is a martingale difference sequence with finite second moments): *$\{g_t\}$ is a martingale difference sequence (so by definition $E(g_t) = 0$). The $K \times K$ matrix of cross moments, $E(g_t g_t')$, is nonsingular. Let S denote Avar $(\bar{g})$ (the variance of the asymptotic distribution of $\sqrt{n}\bar{g}$, where $\bar{g} = \frac{1}{n}\sum_t g_t$).*

By Assumption 2.2 and the Ergodic Stationary Martingale Differences CLT, $S = E(g_t g_t')$.

Proposition 2.3 (robust t-ratio and Wald statistic): *Given a consistent estimator $\hat{S}$ of S, if Assumptions 2.1 to 2.5 hold, then*

a) *Under the null hypothesis* $H_0 : \beta_k = \bar{\beta}_k$,

$$t_k \underset{d}{\rightarrow} N(0,1)$$

b) *Under the null hypothesis* $H_0 : R\beta = r$, *where* R *is a* $\#r \times K$ *matrix (where* $\#r$*, the dimension of* r*, is the number of restrictions) of full row rank,*

$$W \equiv n \cdot (Rb - r)' \left\{ R\,\widehat{Avar(B)}\,R' \right\}^{-1} (Rb - r) \underset{d}{\rightarrow} \chi^2(\#r).$$

Proof: We have already established part a. Part b is a straightforward application of Lemma 2.4(d). Write W as

$$W = c_n' Q_n^{-1} c_n \quad \text{where } c_n = \sqrt{n}(Rb - r) \text{ and } Q_n = R\,\widehat{Avar(B)}\,R'.$$

Under H_0, $c_n = R\sqrt{n}(b - \beta)$, so Proposition 2.1 implies

$$c_n \underset{d}{\rightarrow} c \quad \text{where } c \sim N(0, RAvar(B)R').$$

Also by Proposition 2.1

$$Q_n \underset{p}{\rightarrow} Q \quad \text{where } Q \equiv RAvar(B)R'.$$

Because R is full row rank and $Avar(B)$ is positive definite, Q is invertible. Therefore, Lemma 2.4(d) implies

$$W \underset{d}{\rightarrow} c'Q^{-1}c.$$

Because c is normally distributed with dimension $\#r$, and because Q equals the variance of c, $c'Q^{-1}c = \chi^2(\#r)$. *QED*

The statistic W is a Wald statistic because it is constructed from unrestricted estimators (B and $\widehat{Avar(B)}$) that are not constrained by the null hypothesis. To test $H_0 :: R\beta = r$, simply follow these steps:

Step 1: Construct W.

Step 2: To find the critical value for a test with size 5%, find the point of the $\chi^2(\#r)$ distribution that gives 5% to the upper tail.

Step 3: If W exceeds the critical value, then reject the null hypothesis.

Consistent Test

The power of a test is the probability of rejecting a false null hypothesis. The power of a test (for a given size) depends on the alternative DGP. For example, consider any DGP $\{Y_t, X_t\}$ that satisfies Assumptions 2.1-2.5 but for which $\beta_k \neq \bar{\beta}_k$. The power of a test based on the t ratio is

$$\text{power} = \Pr(|t_k| > cv_\alpha),$$

where cv_α is the critical value associated with a size of α. Because the DGP controls the distribution of t_k, the power depends upon the DGP. The test is consistent against a set of DGP's (none of which satisfy the null) if the power against any member of the set of DGP's approaches unity as $n \to \infty$ (for any assumed significance level). To see that the test based on the t ratio is consistent, note that for

$$t_k = \frac{\sqrt{n}\left(B_k - \bar{\beta}_k\right)}{\sqrt{\widehat{Avar\left(B_k\right)}}}$$

the denominator converges to $\sqrt{Avar\left(B_k\right)}$ despite the fact that DGP does not satisfy the null (Proposition 2.1 requires only Assumptions 2.1-2.5 and does not depend on the truth of the null hypothesis). In contrast, the numerator tends to either $+\infty$ or $-\infty$, because B_k converges to the value of β_k for the DGP, which differs from $\bar{\beta}_k$. Hence the power tends to unity as the sample size tends to infinity, implying that the test is consistent for all members of the set of DGP's. The same is true for the Wald test.

Asymptotic Power

The power of the t test approaches unity as the sample size increases for any fixed alternative DGP. If, however, the DGP is not held fixed but allowed to get closer and closer to the null hypothesis as the sample size increases, then the power may not converge to unity. A sequence of such DGP's is called a sequence of local alternatives. For the regression model with the null of $H_0 : \beta_k = \bar{\beta}_k$, a sequence of local alternatives is a sequence of DGP's such that (i) the n-th DGP $\left\{Y_t^{(n)}, X_t^{(n)}\right\}$ $(t = 1, 2, \ldots)$ satisfies Assumptions 2.1-2.5 and converges in a certain sense to a fixed DGP $\{Y_t, X_t\}$, and (ii) the value of β_k for the n-th DGP, $\beta_k^{(n)}$, converges to $\bar{\beta}_k$.

Suppose, further, that $\beta_k^{(n)}$ satisfies

$$\beta_k^{(n)} = \bar{\beta}_k + \frac{\gamma}{\sqrt{n}}$$

for some given $\gamma \neq 0$. Such a sequence of local alternatives, in which $\beta_k^{(n)}$ approaches $\bar{\beta}_k$ at the rate of $1/\sqrt{n}$, is called a Pitman drift or Pitman sequence. Under this sequence of local alternatives, the t ratio is

$$t_k = \frac{\sqrt{n}\left(B_k - \beta_k^{(n)}\right)}{\sqrt{\widehat{Avar\left(B_k\right)}}} + \frac{\gamma}{\sqrt{\widehat{Avar\left(B_k\right)}}}.$$

If a sample of size n is generated by the n-th DGP of a Pitman sequence, does t_k converge to a nontrivial distribution? Because the n-th DGP satisfies Assumptions 2.1-2.5, the first term on the right side converges in distribution to $N(0,1)$ by Proposition 2.3(a). By part (c) of Proposition 2.1 ($\widehat{Avar} \underset{p}{\rightarrow} Avar$) and the fact that $\left\{Y_t^{(n)}, X_t^{(n)}\right\}$ "converges" to a fixed DGP, the second term converges in probability to

$$\mu := \frac{\gamma}{\sqrt{Avar\left(B_k\right)}}$$

where $Avar\left(B_k\right)$ is evaluated at the fixed DGP. Therefore, $t_k \underset{d}{\rightarrow} N(\mu, 1)$ along this sequence of local alternatives. If the significance level is α, the power converges to

$$\Pr\left(|x| > cv\right),$$

where cv is the critical value (i.e. 1.96) and $x \sim N(\mu, 1)$. This probability is termed the asymptotic power. Evidently, the larger is $|\mu|$ the higher is the asymptotic power for any given size α. By a similar argument, the Wald statistic converges to a non-central chi-square distribution along a Pitman sequence.

Testing Nonlinear Hypotheses

The Wald test can be generalized to a test of nonlinear restrictions on β. Consider a null hypothesis of the form

$$H_0 : a(\beta) = 0.$$

Here a is a vector-valued function with continuous first derivatives. Let $\#a$ be the dimension of a (so the null hypothesis has $\#a$ restrictions) and let $A(\beta)$ be the $\#a \times K$ matrix of first derivatives evaluated at β: $A(\beta) = \partial a(\beta)/\partial \beta'$. For the hypothesis to be well defined, we assume that $A(\beta)$ is full row rank (this is the generalization of the requirement for linear hypotheses $R\beta = r$ that R is of full row rank). By Lemma 2.5 (the delta method) we have

$$\sqrt{n}\left(a(B) - a(\beta)\right) \underset{d}{\rightarrow} c, \quad c \sim N\left(0, A(\beta)\, Avar(B)\, A(\beta)'\right).$$

Because $B \underset{p}{\rightarrow} \beta$, Lemma 2.3(a) implies that $A(B) \underset{p}{\rightarrow} A(\beta)$. Because (Proposition 2.1(c)) $\widehat{Avar(B)} \underset{p}{\rightarrow} Avar(B)$, Lemma 2.3(a) implies

$$A(B)\, \widehat{Avar(B)} A(B)' \underset{p}{\rightarrow} A(\beta)\, Avar(B)\, A(\beta)' = Var(c).$$

Because $A(B)$ has full row rank and $Avar(B)$ is positive definite, $Var(c)$ is invertible. Upon noting that, under the null hypothesis $a(\beta) = 0$, Lemma 2.4(d) implies

$$\sqrt{n} a(B)' \left\{A(B)\, \widehat{Avar(B)} A(B)'\right\}^{-1} \sqrt{n} a(B) \underset{d}{\rightarrow} c' Var(c)^{-1} c \sim \chi^2(\#a).$$

If we combine the two $\sqrt{n}$'s into n we have proven

Proposition 2.3 (continued - nonlinear hypotheses):

c) Under the null hypothesis $H_0 : a(\beta) = 0$ (which contains $\#a$ restrictions) if $A(\beta)$ (the $\#a \times K$ matrix of continuous first derivatives of $a(\beta)$) is of full row rank, then

$$W \equiv n \cdot a(B)' \left\{A(B)\, \widehat{Avar(B)} A(B)'\right\}^{-1} a(B) \underset{d}{\rightarrow} \chi^2(\#a).$$

Part (c) is a generalization of part (b), if we set $a(B) = RB - r$, then W reduces to the Wald statistic for linear restrictions.

For a given set of restrictions, $a(\cdot)$ may not be unique. For example, the restriction $\beta_1 \beta_2 = 1$ could be written as $a(\beta) = 0$ with $a(\beta) = \beta_1\beta_2 - 1$ or with $a(\beta) = \beta_1 - 1/\beta_2$. While part c of the Proposition guarantees that the outcome of the Wald test is the same (for either expression of the restriction) in large samples, the numerical value of W does depend on the way the restriction is written. Hence, the outcome of the test can depend on the expression in finite samples. In this example, the representation $a(\beta) = \beta_1 - 1/\beta_2$ does not satisfy the requirement of continuous derivatives at $\beta_2 = 0$. Indeed, a 1985 Monte Carlo study by Gregory and Veal reports that, when β_2 is close to zero, the Wald test based on the second representation rejects the null too often in small samples. (Actual size exceeds nominal size.)