Econometrics 1, Class Week 4

Back-up Video: Class week 4 video (click here)

Learning Outcomes

(a) Hypothesis testing in the *normal* linear regression model.

(b) Application: individual and joint hypotheses.

Prerequisites

1. Distributional results on distributional handout (DH).

2. Concepts in Linear Algebra:
   - Bilinearity of variance-covariance matrices (AMN p.105);
   - Cholesky decomposition of square, symmetric matrices (AMN p.45).

Setting: Normal Linear Regression Model

Linear Regression Model

$$
\begin{aligned}
\underset{T\times 1}{\mathbf{y}} &= \underset{T\times k}{\mathbf{X}}\underset{k\times 1}{\beta_0} + \underset{T\times 1}{\mathbf{u}} && (1)\\
\mathbf{u} &\sim N(\mathbf{0}, \sigma_0^2 \mathbf{I}_T). && (2)
\end{aligned}
$$

I.e. assume in addition (2) that errors $\mathbf{u}$ are multivariate normal, with mean zero, and homoskedastic.

Can treat $\mathbf{X}$ as exogenous and non-stochastic for now.

Estimate β_0 by OLS estimator

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X'X})^{-1}\mathbf{X'y}\\
&= \beta_0 + (\mathbf{X'X})^{-1}\mathbf{X'u}.
\end{aligned}
$$

So OLS estimator $\hat{\beta}$ is a linear function of $\mathbf{u}$ and hence, as a consequence of (2), also multivariate normal:

$$
\hat{\beta} \sim N(\beta_0, \sigma_0^2(\mathbf{X'X})^{-1}), \qquad (3)
$$

where $\text{var}(\hat{\beta}) = \sigma_0^2(\mathbf{X'X})^{-1}$, as shown in lecture.

Hypothesis Testing: Main Ideas

Null hypothesis (H_0) about the true, but unknown vector of regression coefficients: In this exercise, $\beta_0 = \mathbf{q}$, for some known and given vector $\mathbf{q}$ (perhaps stipulated by theory about the data generating process).

Do the data support this hypothesis?

Since $\hat{\beta}$ is continuously distributed (under (2)), with probability zero $\hat{\beta} = \mathbf{q}$.

Is the sampling variation of $\hat{\beta}$ consistent with H_0?

- If $\hat{\beta}$ is a very precise estimator of β_0, then under H_0 even small deviations from $\mathbf{q}$ are evidence against H_0.
- If $\hat{\beta}$ is a very imprecise estimator of β_0, then under H_0 small deviations from $\mathbf{q}$ are still consistent with H_0.

So, to test H_0, need to evaluate discrepancy between $\hat{\beta}$ and $\mathbf{q}$ in the metric defined by the variance-covariance matrix of $\hat{\beta}$.

.

Hypothesis Testing: Main Ideas (cont'd)

Under H_0,

$$\hat{\beta} \sim N(\mathbf{q}, \sigma_0^2(\mathbf{X'X})^{-1}). \tag{4}$$

Discrepancy $\hat{\beta} - \mathbf{q}$ is a k-vector. Need to determine its length in the metric defined by the inverse of its variance-covariance matrix: Elements that are (im)precisely estimated are given a (low) high weight. So, under H_0,

$$\hat{\beta} - \mathbf{q} \sim N(\mathbf{0}, \sigma_0^2(\mathbf{X'X})^{-1}). \tag{5}$$

Using the Cholesky decomposition of $\mathbf{X'X} = \mathbf{X'X}^{\frac{1}{2}}(\mathbf{X'X}^{\frac{1}{2}})'$,

$$\mathbf{w} := \frac{1}{\sigma_0}(\mathbf{X'X}^{\frac{1}{2}})'(\hat{\beta} - \mathbf{q}) \sim N(\mathbf{0}, \mathbf{I}_k). \tag{6}$$

i.e. under H_0 the k-vector $\mathbf{w}$ is distributed multivariate standard normal.

Wald Test and Wald Test Statistic

The Euclidean squared length of **w** in (6) is the non-negative scalar

$$W = \mathbf{w}'\mathbf{w}. \tag{7}$$

By the definition of a χ^2_k random variable,

$$W \sim \chi^2_k \quad \text{under } H_0. \tag{8}$$

So, the Wald test statistic W satisfies

$$W = (\hat{\beta} - \mathbf{q})'(\mathbf{X'X})(\hat{\beta} - \mathbf{q})/\sigma_0^2 \sim \chi^2_k \quad \text{under } H_0. \tag{9}$$

I.e. W weighs each component of the discrepancy between $\hat{\beta}$ and **q** inversely proportional to its variance.

If W is large (larger than the 95th percentile of the χ^2_k), then this is evidence against H_0 (with a 5 percent probability of making a type-1 error, i.e. of rejecting H_0 even though it is true).

Problem: Unknown σ_0^2

W depends on σ_0^2 which is typically unknown, but can be estimated by s^2.

What is the distribution of s^2?

Recall from Week 2,

$$
\begin{aligned}
s^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-k) \\
&= \mathbf{u}'M\mathbf{u}/(T-k) \\
&\sim \frac{\sigma_0^2}{T-k}\chi^2_{T-k} \\
&\quad \text{b/c } \mathrm{rk}(M) = T-k \text{, Theorem 1 on DH} \\
V = \frac{(T-k)s^2}{\sigma_0^2} &\sim \chi^2_{T-k} \qquad (10)
\end{aligned}
$$

If σ_0^2 were replaced by s^2 in W, should no longer expect resulting statistic to be distributed χ_k^2 because more randomness is introduced.

Combining Statistics

Could divide W/k (9) by $V/(T-k)$ (10) to eliminate σ_0^2:

$$\begin{aligned}\mathcal{W} &= \frac{W/k}{s^2/\sigma_0^2} \\ &= \frac{(\hat{\beta}-\mathbf{q})'(\mathbf{X'X})(\hat{\beta}-\mathbf{q})/k\sigma_0^2}{s^2/\sigma_0^2} \\ &= (\hat{\beta}-\mathbf{q})'(\mathbf{X'X})(\hat{\beta}-\mathbf{q})/ks^2. \end{aligned} \quad (11)$$

This is ratio of two χ^2 distributions, divided by their respective d.f. (with k numerator and $T-k$ denominator d.f.).

$\mathcal{W}$ is feasible, unlike W, even if σ_0^2 is unknown.

It is distributed $F_{k,T-k}$ by Theorem 2 on DH *if* the two χ^2-distributed random variables are independent.

Independence of W and V

1. $\hat{\beta}$ and $\hat{\mathbf{y}}$ are one-to-one:

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\hat{\beta} \\ \hat{\beta} &= (\mathbf{X'X})^{-1}\mathbf{X'}\hat{\mathbf{y}}.\end{aligned}$$

2. $\hat{\mathbf{u}} \perp \hat{\mathbf{y}}$:

$$\hat{\mathbf{u}}'\hat{\mathbf{y}} = \mathbf{u}' M P_X \mathbf{y} = 0.$$

So variation in s^2 (induced by $\hat{\mathbf{u}}$) is orthogonal to variation in $\hat{\beta}$ (equivalent to variation induced by $\hat{\mathbf{y}}$).

Hence, W and V are statistically independent.

Therefore, by Theorem 2 on DH, the Wald test statistic $\mathcal{W}$ satisfies

$$\mathcal{W} = (\hat{\beta} - \mathbf{q})'(\mathbf{X'X})(\hat{\beta} - \mathbf{q})/ks^2 \sim F_{k,T-k}. \qquad (12)$$

Note: Due to added randomness,

95th percentile of $F_{k,T-k} >$ 95th percentile of χ^2_k.

Representation in terms of Sums of Squared Residuals

Denominator s^2 uses unrestricted sum of squared residuals $\hat{\mathbf{u}}'\hat{\mathbf{u}}$.

Numerator W is

$$\begin{aligned} W &= (\hat{\beta} - \mathbf{q})'(\mathbf{X'X})(\hat{\beta} - \mathbf{q})/\sigma_0^2 \\ &= (\mathbf{X}\hat{\beta} - \mathbf{Xq})'(\mathbf{X}\hat{\beta} - \mathbf{Xq})/\sigma_0^2 \\ &= (\mathbf{y} - \mathbf{Xq} - (\mathbf{y} - \mathbf{X}\hat{\beta}))'(\mathbf{y} - \mathbf{Xq} - (\mathbf{y} - \mathbf{X}\hat{\beta}))/\sigma_0^2 \\ &= (\hat{\mathbf{u}} - \tilde{\mathbf{u}})'(\hat{\mathbf{u}} - \tilde{\mathbf{u}})/\sigma_0^2 \\ &= (\hat{\mathbf{u}}'\hat{\mathbf{u}} + \tilde{\mathbf{u}}'\tilde{\mathbf{u}} - 2\hat{\mathbf{u}}'\tilde{\mathbf{u}})/\sigma_0^2, \end{aligned}$$

where $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{Xq}$ is vector of restricted residuals.

Notice that

$$\begin{aligned}\tilde{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y}-\mathbf{Xq})'M\mathbf{y} \\ &= \mathbf{y}'M\mathbf{y} \\ &= \hat{\mathbf{u}}'\hat{\mathbf{u}}.\end{aligned}$$

Therefore,

$$W = (\tilde{\mathbf{u}}'\tilde{\mathbf{u}} - \hat{\mathbf{u}}'\hat{\mathbf{u}})/\sigma_0^2. \tag{13}$$

And hence, the Wald test statistic can be re-expressed as

$$\mathcal{W} = \frac{(\tilde{\mathbf{u}}'\tilde{\mathbf{u}} - \hat{\mathbf{u}}'\hat{\mathbf{u}})/k}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-k)}. \tag{14}$$

(b) Application: Dividends and Earnings

Two specifications: (1) static, and (2) dynamic.

(i)-(ii) To test hypotheses about single coefficient (in (1)), form t-ratios.

$$
\begin{aligned}
H_0 : \text{coeff.}{=}1 \qquad t_1 &= \frac{0.4-1}{0.1} = -6 \\
H_0 : \text{coeff.}{=}0 \qquad t_2 &= \frac{0.4}{0.1} = 4
\end{aligned}
$$

and compare against critical values (2.5th and 97.5th percentile) of a t-distribution with 1 d.f. (± 1.96).

So reject both hypotheses.

Joint tests

(i) Individual test of hypothesis that coefficients on lags are zero

$$\begin{aligned} t_1 &= \frac{-0.1}{0.1} = -1 \qquad \text{cannot reject } H_0 \\ t_2 &= \frac{0.7}{0.14} = 5 \qquad \text{reject } H_0. \end{aligned}$$

(ii) Test joint hypothesis

$$\begin{aligned} F\text{-test} \qquad \mathcal{W} &= \frac{(2.1849 - 0.84821)/2}{0.84821/(24-4)} = \frac{0.668}{0.042} = 15.90 \\ \text{LLR test} \qquad LR &= 2(LL_U - LL_R) = 2(6.0576 - (-5.297)) = 22.71 \end{aligned}$$

and compare to 95th percentile of $F_{2,T-4}$ (3.49) and χ^2_2 (5.99), respectively. So reject in both cases.

Note: It may be misleading that in (a) restrictions on all k coefficients were imposed, so that the numerator d.f. of the F-statistic are k. In general, the numerator d.f. equal the number of restrictions, so here this is $m = 2 < k = 4$; in general, $F_{m,T-k}$ is the reference distribution.

Linear Restrictions

Want $R\beta = r$ for $\beta' = (\beta_1, \cdots, \beta_4)$ and joint hypothesis that $\beta_3 = \beta_4 = 0$:

$$\underset{2\times 4}{R} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\underset{2\times 1}{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$