Inference and forecasting in OLS

Susan Thomas
IGIDR, Bombay

13 October, 2008

Recap

- The linear regression model is one of the form:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_J X_{J,i} + u_i$$

- Or, $Y = X\beta + U$ where $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_J)'$
- Where β minimise the SSE:

$$\sum \epsilon_i = \sum (Y_i - \beta_0 - \ldots - \beta_J X_{J,i}) = \sum u_i = 0$$
$$\min_{\beta,\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_{1,i} - \ldots - \beta_J X_{J,i})^2 = \min_{\beta,\sigma^2} \sum u_i^2$$

- The solution for this is

$$\beta = (X'X)^{-1}(X'Y)$$

  where $\beta \sim N(0, \sigma^2 / S_{XX})$
- Next goal: How do we form and test null hypothesis in this framework?

Hypothesis testing

- We know that $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Or,
$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$$
- Test 1: Is $\beta_1 = 0$? Ie, is there any effect of X_i on Y_i?
- Alternatively, the test can be whether $\beta_1 = 0.1$? When X_i increases by 1, Y_i increases by 0.1.
- Generally, we want to test if $\beta = \beta_{H_0}$. (This is a test with population parameters against our null.)
- However, $\hat{\beta}_1$ is a good estimator for β_1
- So, we would like to test if $\hat{\beta} = \beta_{H_0}$, or is $\hat{\beta} - \beta_{H_0}$ small?
- Since $\hat{\beta}$ is rv, this would depend upon the distribution of $\hat{\beta}$.
- If we can make a statement like:
$$\frac{\hat{\beta} - \beta_{H_0}}{\sigma_{\hat{\beta}}} \sim N(0, 1)$$
then the difference should be a small number with a very high probability where the probability is well-defined.

Recap:The hypothesis testing framework, 3 steps

1. Form H_0, and H_A
2. Form the distribution of the estimator begin tested.
3. Form the decision rule.

Step 1: Formulating two opposing hypothesis

1. Null Hypothesis, H_0
2. Alternative Hypothesis, H_A

$$H_0: \quad \beta = \beta_0$$
$$H_A: \quad \beta \neq \beta_0 \quad \text{(Two tailed test)}$$

$$H_0: \quad \beta = \beta_0$$
$$H_A: \quad \beta > \beta_0 \quad \text{(One tailed test)}$$

$$H_0: \quad \beta = \beta_0$$
$$H_A: \quad \beta < \beta_0 \quad \text{(One tailed test)}$$

Step 2: Distribution of β

- Assumption: For given X_i

$$u_i \sim N(0, \sigma^2)$$

- Since $Y_i = \beta_0 + \beta_1 X_i + u_i$

$$\Rightarrow \quad Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

- $\hat{\beta}_1 = \sum \omega_i Y_i$, where

$$\omega_i = \frac{(X_i - \bar{X})}{S_{XX}}$$

$$\Rightarrow \hat{\beta}_1 \quad \text{Linear in Y}$$

$$\Rightarrow \hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\beta_1}) \qquad \sigma^2_{\beta_1} = \frac{\sigma^2}{S_{XX}}$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0, 1)$$

Distribution of $\hat{\beta}_1$

- We have $(\hat{\beta}_0, \hat{\beta}_1)$ are normally distributed as:

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\beta_0}) \quad \Rightarrow \quad \frac{(\hat{\beta}_0 - \beta_0)}{\sigma^2_{\beta_0}} \sim N(0, 1)$$
$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\beta_1}) \quad \Rightarrow \quad \frac{(\hat{\beta}_1 - \beta_1)}{\sigma^2_{\beta_1}} \sim N(0, 1)$$

- We need $\sigma_{\hat{\beta}_1}, \sigma_{\hat{\beta}_0}$ – only have an estimator.
- Estimator for $\sigma_{\hat{\beta}_1}$:

$$\sigma_{\hat{\beta}_1} = S^2_{\hat{\beta}_1} = \frac{\sum \hat{u}_i^2}{(N-2)}$$

- Distribution for $\hat{\beta}_1$:

$$(\frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}})$$

  Is the distribution same as N(0,1)?

Deriving the distribution of $\hat{\beta}_1$

- Observation: $\sum \frac{\hat{u}_i^2}{\sigma^2} \sim \chi^2_{(N-2)}$
- Observation: $\hat{\beta}_0, \hat{\beta}$, and $\sum \frac{\hat{u}_i^2}{\sigma^2}$ are independent.
- Fact: If $W \sim N(0,1)$ and $Z \sim \chi^2(N)$, then

$$W/\sqrt{(Z/N)} \sim t(N)$$

- This means we can write:

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma^2_{\hat{\beta}_1}} / \sqrt{\frac{\sum \hat{u}_i^2}{\sigma^2} \cdot \frac{1}{N-2}} \sim t(N-2)$$

- What does this "reduce" to?

Deriving the distribution of $\hat{\beta}_1$

$$\frac{\hat{\beta}-\beta}{\sigma_{\hat{\beta}}} \Big/ \sqrt{\frac{\sum \hat{u}_{t^2}}{N-2} \cdot \frac{1}{\sigma^2}} \sim t(N-2)$$

$$\frac{\hat{\beta}-\beta}{\sigma_{\hat{\beta}}} \Big/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} \sim t(N-2)$$

$$\frac{\hat{\beta}-\beta}{\sigma_{\hat{\beta}}} \Big/ \frac{\hat{\sigma}}{\sigma} \quad \text{or} \quad \frac{\hat{\beta}-\beta}{\sigma_{\hat{\beta}}} . \frac{\sigma}{\hat{\sigma}} \sim t(N-2)$$

$$\text{But, } \frac{\sigma}{\sigma_{\hat{\beta}}\hat{\sigma}} = \frac{\not{\sigma}}{\frac{\not{\sigma}}{\sqrt{S_{xx}}}.\hat{\sigma}} = \frac{1}{\frac{\hat{\sigma}}{\sqrt{S_{xx}}}} = \frac{1}{S_{\hat{\beta}}}$$

$$\text{Therefore, } \frac{\hat{\beta}-\beta}{S_{\hat{\beta}}} \sim t(N-2)$$

Under $H_0 : \beta = \beta_0$ and the test is:

$$t_c = \frac{\hat{\beta}-\beta_0}{S_{\hat{\beta}}} \sim t(N-2)$$

Recap: Hypothesis testing, the decision rule

- Fix a critical level, α
- If H_0 is true, then the t_c will be less than $t(N-2)(\alpha)$ value 95% of the time.
- If $t_c > t(N-2)(\alpha)$, then a very unlikely value has happened and we do not accept H_0 and consider the alternative H_A.
- Decision rule:

$$\begin{array}{lcll} t_c & > & t(N-2)(\alpha) & \text{Do not accept } H_0 \\ t_c & \leq & t(N-2)(\alpha) & \text{Accept } H_0 \end{array}$$

- If $H_0 : \beta = 0$, then

$$t_c = \frac{\hat{\beta}}{S_{\hat{\beta}}} = \text{t-statistic}$$

A numerical example

- $\hat{\beta} = 0.13875, S_{\hat{\beta}} = 0.01873, N = 14, K = 2$
- Test $H_0 : \beta = 0, H_1 : \beta > 0$ at $\alpha = 0.01$ (99% confidence)
- $t(12)(0.01) = 2.681$

- $t_c = \frac{0.13875}{0.01873} = 7.41$
- $t_c > t(12)(0.01)$,
  Therefore do not accept H_0, consider $H_A : \beta > 0$

A numerical example, 2

- $\hat{\beta}_0 = 52.351, S_{\hat{\beta}_0} = 37.285, N = 14, K = 2$
- Test $H_0 : \beta_0 = 0, H_1 : \beta_0 < 0$ at $\alpha = 0.01$ (99% confidence)
- $t(12)(0.01) = 2.681$

- $t_c = \frac{52.351}{37.285} = 1.404$
- $t_c < t(12)(0.01)$,
  Therefore cannot reject H_0, ie, accept $H_0 : \beta_0 = 0$

Testing goodness of fit using R^2

Step 1: Forming H_0, H_A

- There is no absolute value of R^2 for "good fit".
- However, we can test $H_0 : R^2 = 0$ against $H_A : R^2 > 0$.
- If we cannot reject $H_0 : R^2 = 0$, then the model does not fit.
- Alternative POV:

$$\begin{array}{lcl} H_0 & : & R^2 = 0 \\ H_A & : & R^2 > 0 \end{array} \Rightarrow \begin{array}{lcl} H_0 & : & \rho_{XY} = 0 \\ H_A & : & \rho_{XY} \neq 0 \end{array}$$

Step 2: Deriving the test-statistic

$$
\begin{aligned}
F &= \frac{R^2(n-2)}{(1-R^2)} \sim F_{1,(N-2)} \\
F &= \frac{\frac{RSS}{TSS}(N-2)}{(1-\frac{RSS}{TSS})} \\
F &= \frac{RSS(N-2)}{ESS} \sim F_{1,(N-2)}
\end{aligned}
$$

Step 3: Decision Rule

$$F > F^*_{1,(n-2)}(0.05) \quad \Rightarrow \quad \text{reject } H_0 : R^2 = 0$$
$$\text{Otherwise accept } H_0$$

Example

- $R^2 = 0.82052, N - 2 = 12$
- Remember that R^2 is (ESS/TSS).
- What is the value of F for this example?
- F is:

$$F = \frac{0.82052(14-2)}{1-0.82052} = 54.86$$

- At $\alpha = 0.05$, $F_{1,12} = 4.75$
- Do you accept or not accept the H_0?
- $F > 4.75$, therefore Reject H_0.
  Ie, the model with this R^2 has some explanatory power.

Predicting Y_i

Confidence intervals for $\hat{\beta}_1$

- If $Y_i = \beta_0 + \beta_1 X_i + u_i$, what is our estimate of $\beta \Rightarrow \hat{\beta}$?
  1. $\hat{\beta}$ is a point-estimate. $\hat{\beta}$ is a random variable from different samples. Therfore it has distribution
  2. Instead of a point estimate for β, we can estimate a confidence interval for β
  3. We know $t_c = \frac{\hat{\beta}-\beta}{S_{\hat{\beta}}} \sim t(N-2)$
     and therefore $-t(n-2)(\alpha/2)$ and $+t(n-2)(\alpha/2)$
  4. Such that $P[-t(n-2)(\alpha/2) \leq t_c \leq t(n-2)(\alpha/2)] = 1-\alpha$
  5. This means:

$$[-t(N-2)(\alpha/2) \leq \frac{\hat{\beta}-\beta}{S_{\hat{\beta}}} \leq t(N-2)(\alpha/2)] = 0.95$$

$$\Rightarrow P[\hat{\beta} - t(N-2)(\alpha/2)S_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t(N-2)(\alpha/2)S_{\hat{\beta}}] = 0.95$$

A numerical example

- $\hat{\beta} = 0.13875, S_{\hat{\beta}} = 0.01873, N = 14, K = 2$
- At $\alpha = 0.05, \alpha/2 = 0.025$, $t(12)(0.025) = 2.179$

- 95% CI for $\hat{\beta} = 0.13875 \underline{+} (2.179 \times 0.18373)$

$$\Rightarrow [0, 099, 0.179]$$

Forecasting Y_i

- $Y_i = \beta_0 + \beta_1 X_i + u_i$
- Suppose $X_i = X_0$. Want to estimate Y_0.
- If we know β_0, β_1 surely, then

$$\begin{aligned} Y_0 &= \beta_0 + \beta_1 X_0 + u_0 \\ \Rightarrow E(Y_0|X_0) &= \beta_0 + \beta_1 X_0 \end{aligned}$$

  This is the conditional mean of Y_0 given X_0.
- We know a sample estimate $\hat{\beta}_0, \hat{\beta}_1$. Then:

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

- $\hat{Y}_0$ is an unbiased predictor of the average of Y, given X_0:

$$\begin{aligned} \Rightarrow E(\hat{Y}_0|X_0) &= E(\hat{\beta}_0) + E(\hat{\beta}_1 X_0) \\ &= \beta_0 + \beta_1 X_0 \end{aligned}$$

- We see $\hat{Y}_0$ is a rv –what is it's variance?

$$\begin{aligned}
\text{var}(\hat{Y}_0) &= \text{var}(\hat{\beta}_0 + \hat{\beta}_1 X_0) \\
&= \text{var}(\hat{\beta}_0) + \text{var}(\hat{\beta}_1 X_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1 X_0) \\
&= \frac{\sum X_i^2}{N} \cdot \frac{\sigma^2}{S_{XX}} + X_0^2 \cdot \frac{\sigma^2}{S_{XX}} - 2X_0 \frac{\bar{X}, \sigma^2}{S_{XX}} \\
\text{var}(\hat{Y}_0) &= \frac{\sum X_i^2}{N} \cdot \frac{\sigma^2}{S_{XX}} + X_0^2 \cdot \frac{\sigma^2}{S_{XX} - 2X_0 \bar{X} \frac{\sigma^2}{S_{XX}}}
\end{aligned}$$

Estimate for variance:

$$S^2_{\hat{Y}_0} = \hat{\sigma}^2 \left[\frac{1}{N} + \frac{(X_0 - \bar{X})^2}{S_{XX}} \right]$$

- The variance of $\hat{Y}_0$ is a function of $(X_0 - \bar{X})$!
  The further X_i is away from $\bar{X}$, the larger the variance of $E(Y_i)$.

Forecast of $\hat{Y}_0$

- $E(\hat{Y}_0|X_0) = \hat{\beta}_0 + \hat{\beta}_1 X_0$
- $\text{var}(\hat{Y}_0) = S^2_{\hat{Y}_0} = \hat{\sigma}^2 \left[\frac{1}{N} + \frac{(X_0 - \bar{X})^2}{S_{XX}}\right]$
- 95% confidence interval:

$$\hat{Y}_0 \pm t(N-2)(\alpha/2).S_{\hat{Y}_0}$$

  Where $\alpha/2$ is the critical value for a two-tailed test.
  And $t(N-2)(\alpha/2)$ is the critical region

- At $\alpha = 0.05$, 95% CI $\Rightarrow t(12)(0.025) = 2.179$

HW: Numerical example

1. These are the estimation results for a model: $Y_i = \beta_0 + \beta_1 X_i + u_i$.
   - $\hat{Y}_i = 52.351 + 0.13875 X_i$
   - Std. Error of coefficients are: $37.29, 0.019$
   - $R^2 = 0.821, (N-2) = 12, \sigma^2 = 39.023$
   - What is $(\hat{Y}_0 | X_0 = 2000)$?
   - Can you calculate the $\text{var}(\hat{Y}_0 | X_0 = 2000)$?
2. What is the mathematical form of $\text{var}(\hat{u}_0)$?