Lecture notes to Stock and Watson chapter 5

Inference in linear regression with a single regressor

Tore Schweder

September 2009

SW Excercises

4.3

4.4 d (an extra question)
Show that $\beta = 1$ for all assets in a CAPM model with equal weights, $R_m = \frac{1}{N}\sum_{i=1}^{N} R_i$, and equally variable and symmetrically correlated excess earnings $X_i = R_i - R_f \, : \, var\left(X_i\right) = \sigma^2$, $corr\left(X_i, X_j\right) = \rho \;\; i \neq j$

Inference in regression

Model: $E[Y|X = x] = \beta_0 + \beta_1 x$
Typical questions referring to the model, and empirical answers:

- What is the marginal expected effect on Y of a change in X? Answer: β_1
  - Estimate or confidence interval for β_1
- What is the expected effect of increasing X from x_1 to $x_1 + \Delta x$?Answer: $\beta_1 \cdot \Delta x$
  - Estimate or confidence interval for $\theta = \beta_1 \cdot \Delta x$
- Is Y affected by X in the mean (is there correlation)?
  - Test $H_0 : \beta_1 = 0 \quad vrs. \quad H_1 : \beta_1 \neq 0$
- What is the mean response of Y when $X = x$?
  - Estimate or confidence interval for $\theta = \beta_0 + \beta_1 x$

Methods

- Linear parameters $\theta = a\beta_0 + b\beta_1$ are estimated by $\widehat{\theta} = a\widehat{\beta}_0 + b\widehat{\beta}_1$, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the OLS estimators, *a* and *b* are numbers.
- $var\widehat{\theta} = a^2 var\left(\widehat{\beta}_0\right) + b^2 var\left(\widehat{\beta}_1\right) + 2ab \cdot cov\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$
- Estimates of $\sqrt{var\left(\widehat{\beta}_0\right)}$ and $\sqrt{var\left(\widehat{\beta}_1\right)}$, $SE\left(\widehat{\beta}_0\right)$ and $SE\left(\widehat{\beta}_1\right)$ are given in computer output, along with an estimate of $\rho = cor\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$
  - compute $SE\left(\widehat{\theta}\right)$ from these numbers.
  - STATA provides inference on linear parameters

Methods cont.

- For large samples, and when the basic assumptions are satisfied (Key Concepts 4.3 in SW: Linear regression, random sample, large outliers are unlikely):
  - Two-sided confidence interval of degree $1-2\alpha$ for θ is

$$\widehat{\theta} \pm z_{1-\alpha} SE\left(\widehat{\theta}\right) = \left(\widehat{\theta} - z_{1-\alpha} SE\left(\widehat{\theta}\right), \widehat{\theta} + z_{1-\alpha} SE\left(\widehat{\theta}\right)\right)$$

    where $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ is the $1-\alpha$ quantile of the standard normal distribution $z_{.95} = 1.64$
  - Reject $H_0 : \theta = \theta_0$ versus the two-sided alternative, at level α, if

$$\left|\frac{\widehat{\theta} - \theta_0}{SE\left(\widehat{\theta}\right)}\right| > z_{1-\alpha} \Leftrightarrow \theta_0 \text{ is not in } \widehat{\theta} \pm z_{1-\alpha} SE\left(\widehat{\theta}\right)$$

  - the p-value is $p = 2 \cdot \min\left\{\Phi\left(\frac{\widehat{\theta}-\theta_0}{SE\left(\widehat{\theta}\right)}\right), 1 - \Phi\left(\frac{\widehat{\theta}-\theta_0}{SE\left(\widehat{\theta}\right)}\right)\right\}$

Methods cont.

One-sided confidence interval of degree $1-\alpha$ for θ is $\left(\widehat{\theta} - z_{1-\alpha} SE\left(\widehat{\theta}\right), \infty\right)$

- the p-value for the one-sided test problem $H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0$ is

$$p = 1 - \Phi\left(\frac{\widehat{\theta} - \theta_0}{SE\left(\widehat{\theta}\right)}\right).$$

- Reject H_0 at level α if

$$\begin{aligned} p &< \alpha \\ &\Leftrightarrow \frac{\widehat{\theta} - \theta_0}{SE\left(\widehat{\theta}\right)} > z_{1-\alpha} \\ &\Leftrightarrow \theta_0 \text{ is not in } \left(\widehat{\theta} - z_{1-\alpha} SE\left(\widehat{\theta}\right), \infty\right) \end{aligned}$$

Methods for estimating standard errors for OLS regression parameters

Calculate empirical residuals

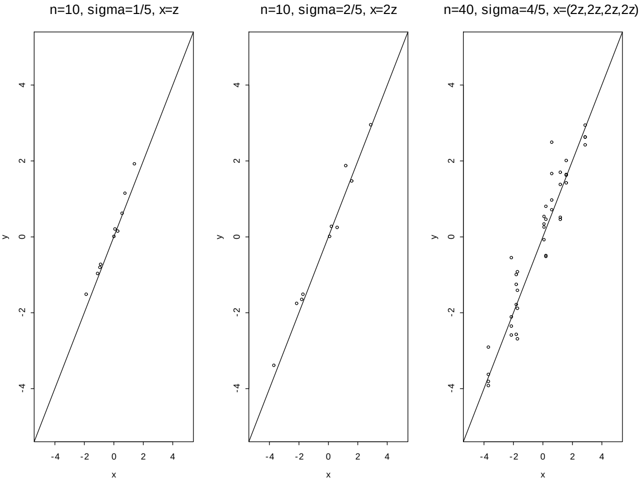
$\widehat{u}_i = Y_i - \widehat{Y}_i = Y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_i\right) \quad i = 1, \cdots, n$

- Heteroskedasticity-robust standard error:

$$SE\left(\widehat{\beta}_1\right) = \frac{1}{n} \frac{\frac{1}{n-2} \sum \left(X_i - \overline{X}\right)^2 \widehat{u}_i^2}{\left[\frac{1}{n} \sum \left(X_i - \overline{X}\right)^2\right]^2}$$

- The robust standard errors are obtained by STATA command `regress...,robust`
- $SE\left(\widehat{\beta}_1\right)$ is small when the empirical residuals $\widehat{u}_i$ are small and/or when $\sum \left(X_i - \overline{X}\right)^2$ is large!

$var\left(\widehat{\beta}_1\right)$ is identical in the three graphs



Properties of OLS

- Linearity, $E\left[Y_i|X_i\right] = \widehat{\beta}_0 + \widehat{\beta}_1 X_i \Rightarrow$linear parameters $\widehat{\theta} = a\widehat{\beta}_0 + b\widehat{\beta}_1$ are unbiased, i.e. $E\widehat{\theta} = \theta$
- Linearity + homoskedasticity+independent normally distributed residuals, $u_1, \cdots, u_n$ are iid $N(0, \sigma^2) \Rightarrow \frac{\widehat{\theta}-\theta}{SE\left(\widehat{\theta}\right)}$ is t-distributed with $n-2$ degrees of freedom
- Linearity+random sample+large outliers unlikely $\Rightarrow \frac{\widehat{\theta}-\theta}{SE\left(\widehat{\theta}\right)}$ is approximately standard normally distributed
- approximation better the larger n is and/or the closer to normality the residuals are distributed.

STATA

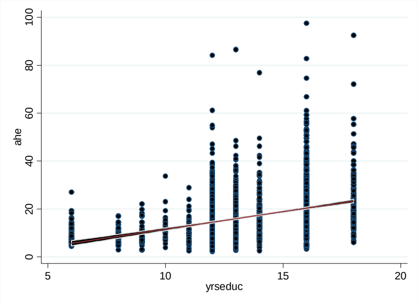
The Economic Value of a Year of Education (SW p165)

use D:\Schweder\undervisning\Master\ECON4135_2009\ch5_cps_box.dta
reg ahe yrseduc , robust

```
Linear regression                Number of obs  =    2950
                                 F(  1,  2948)  =  415.76
                                 Prob > F       =  0.0000
                                 R-squared      =  0.1304
                                 Root MSE       =  8.7694
```

ahe	Robust Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
yrseduc	1.466925	.0719423	20.39	0.000	1.325863	1.607988
_cons	-3.134371	.9258849	-3.39	0.001	-4.949818	-1.318925

twoway (lfitci ahe yrseduc) (scatter ahe yrseduc)



Problems to be done in class

- SW: 5.4
- SW: 5.8
- SW: 5.15