Computer Handout 5: Importance of Graphics for Forecasting

Diego Escobari

Econ 3342

This Computed Handout 5 will show the importance of using graphical tool before engaging into sophisticated statistical forecasting.

Consider the following four data sets:

obs	X1	Y1	X2	Y2	X3	Y3	X4	Y4
1	10.00000	8.040000	10.00000	9.140000	10.00000	7.460000	8.000000	6.580000
2	8.000000	6.950000	8.000000	8.140000	8.000000	6.770000	8.000000	5.760000
3	13.00000	7.580000	13.00000	8.740000	13.00000	12.74000	8.000000	7.710000
4	9.000000	8.810000	9.000000	8.770000	9.000000	7.110000	8.000000	8.840000
5	11.00000	8.330000	11.00000	9.260000	11.00000	7.810000	8.000000	8.470000
6	14.00000	9.960000	14.00000	8.100000	14.00000	8.840000	8.000000	7.040000
7	6.000000	7.240000	6.000000	6.130000	6.000000	6.080000	8.000000	5.250000
8	4.000000	4.260000	4.000000	3.100000	4.000000	5.390000	19.00000	12.50000
9	12.00000	10.84000	12.00000	9.130000	12.00000	8.150000	8.000000	5.560000
10	7.000000	4.820000	7.000000	7.260000	7.000000	6.420000	8.000000	7.910000
11	5.000000	5.680000	5.000000	4.740000	5.000000	5.730000	8.000000	6.890000

Each data set has one variable x, with it corresponding variable y.

If you estimate the model:

$$Y1_i = \beta_0 + \beta_1 X1_i + \varepsilon_i$$

you get the following computer output:

Dependent Variable: Y1
Method: Least Squares
Date: 09/26/10 Time: 22:33
Sample: 1 11
Included observations: 11

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.000091	1.124747	2.667348	0.0257
X1	0.500091	0.117906	4.241455	0.0022

R-squared	0.666542	Mean dependent var	7.500909
Adjusted R-squared	0.629492	S.D. dependent var	2.031568
S.E. of regression	1.236603	Akaike info criterion	3.425579
Sum squared resid	13.76269	Schwarz criterion	3.497924
Log likelihood	-16.84069	F-statistic	17.98994
Durbin-Watson stat	3.212290	Prob(F-statistic)	0.002170

That corresponds to the following estimated equation:

$$\widehat{Y1}_i = 3 + 0.5\, X1_i$$

What is interesting, is that this is the case for any pair of y and x:

$$\widehat{Y2}_i = 3 + 0.5\,X2_i; \quad \widehat{Y3}_i = 3 + 0.5\,X3_i; \quad \widehat{Y4}_i = 3 + 0.5\,X4_i$$

You will also get the same R^2 as well as the same standard errors, t-statistics and p-values.

What's the problem with this? There doesn't seem to be any problem, you may think, as different pairs of x and y can give exactly the same regression equation. The problem becomes clear when you graph the data: