FEC 522: Financial Econometrics II

Spring 2012

Problem Sheet

(to be discussed Saturday, March 10, 2012)

Problem 1: File `data_dji_fchi_gdaxi_daily.csv` contains, among others, daily returns in percent on stock indices DJIA (variable `dji.ret`), CAC 40 (`fchi.ret`), and German DAX (`gdaxi.ret`) for the year 2011.

a) Using function `plot3d` of R package `rgl`, make a 3-dimensional scatterplot of the returns.

b) Try out different perspectives and find out how the perspective of the plot can be changed with the mouse. For example, move the mouse forward and backward while keeping the Ctrl key and the middle mouse key (the scroll wheel) pressed. This will zoom the plot like a tele or wide angle lens.

c) Shift to a "tele" perspective and rotate the scatterplot such that you have a vertical view of one plane (for example, with `fchi.ret` as abscissa and `gdaxi.ret` as ordinate). Which of the three pairs has the highest correlation, according to a visual inspection?

d) Would it make sense to analyze the associations between the return series with a multiple linear regression model? Why or why not?

Problem 2: In an effort to investigate the determinants of four-star hotel room rates (price per night) in Prague in June 2011, the values of several variables were recorded for each hotel in a sample taken from the website www.hotels.com. The following table gives variable names and their explanations, minimum and maximum values observed in the sample, and sample means. For the data, see file `hotels_prague_v2012-03-09.csv`.

variable	explanation	minimum	maximum	average
`price`	room rate = price for 1 night (€)	52	166	104
`rating`	average customer rating	7.2	9.2	8.45
`number.ratings`	number of customer ratings	10	263	47.7
`ln.number.ratings`	natural logarithm of number.ratings	2.303	5.572	3.518
`distance`	distance from the city center (km)	0.1	11.0	2.6
`ln.distance`	natural logarithm of distance	−2.303	2.398	0.433

A higher rating indicates a higher degree of customer satisfaction. The number of customer ratings is the number of customers who give a rating on www.hotels.com.
Computer output for two alternative regression models fitted to the data, with `price` as dependent variable:

Model 1:

```
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          -62.753     48.936  -1.282  0.20482
rating                16.859      5.652   2.983  0.00417 **
log(distance)        -10.602      3.614  -2.933  0.00479 **
log(number.ratings)    9.057      4.470   2.026  0.04734 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.96 on 58 degrees of freedom
Multiple R-squared: 0.3802, Adjusted R-squared: 0.3481
F-statistic: 11.86 on 3 and 58 DF,  p-value: 3.686e-06
```

Model 2:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    113.391      3.860  29.379  < 2e-16 ***
log(distance)  -15.417      3.716  -4.149 0.000107 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.37 on 60 degrees of freedom
Multiple R-squared: 0.2229, Adjusted R-squared:  0.21
F-statistic: 17.22 on 1 and 60 DF,  p-value: 0.0001068
```

a) From file `hotels_prague_v2012-03-09.csv`, extract those hotels which have 4 stars, at least 10 ratings, and price for one nights at most €180.

b) Using R, reproduce Models 1 and 2.

c) Write the regression equation of Model 1.

d) Explain the meaning of `Multiple R-squared` (R^2) in Model 1. What precisely is explained by this model?

e) What is the meaning of the intercept in Model 2? Does it have a meaning at all? (Hint: The independent variable is the natural logarithm of distance, and $\ln(1) = 0$.)

f) When writing up a report for a hotel association, which model would you prefer?

g) Moods Boutique Hotel has four stars with customer rating 9.0, 17 customer ratings, and a distance of 0.7 km from the city center. Write the mathematical expression of the point forecast for the price for one night when using Model 1.

h) The value of the point forecast in (f) is €118.42, while Moods Boutique Hotel actually asks €147 for one night. What might be the reasons for this difference?

i) Using Model 1, the 95% prediction interval for the room price of the four-star hotels with rating 8.0, 50 ratings, located 10 km from the city center is [33, 134]. What does this mean in this case? Now suppose Euphony (out-of-sample, i.e. not used for estimating Model 1) is such a hotel, asking €111. What is your conclusion?

j) It can be shown that distance from the city center is not significant in a regression model with price as dependent variable for *five-star* hotels. What are possible implications when you think of advertising strategies for four- and five-star hotels?