Problem Set 6

Tao Wu

November 21, 2023
Professor Suleyman Taspinar
ECON 387: Advanced Econometrics

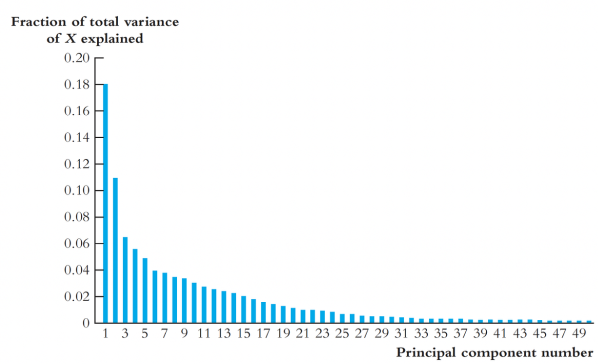
1 Question 1

(20 pts) Define the following terms in your own words.

(a) In big datasets, we often have many highly correlated variables in a regression model. To eliminate the multicollinearity problem, we group these correlated variables into units called principal components which are uncorrelated among each other. The original variables are also picked for the purpose of maximizing the variances captured in the data. Once done, we run regression on principle components as the new independent variables.

(b) A scree plot shows the ratio between the sample variance of each principle component and the total sample variance of all regressors which consist of 2nd-degree, 3rd-degree, and interaction terms of the original variables. Plotted values on the graph indicate the percentage of total variance explained by each principal component.

FIGURE 14.6 Scree Plot for the 817-Variable School Data Set (First 50 Principal Components)



Plotted values are the fraction of the total variance of the 817 regressors explained by the indicated principal component. The first principal component explains 18% of the total variance of the 817 X's, and the first 10 principal components together explain 63% of the total variance.

(c) After deciding the number of principal components, based on their share of the total variance (higher, better), to use for the regression analysis, we regress the original outcome variable on these uncorrelated principal components.

(d) Time series data is a type of data that is collected at a defined time interval. The ordering of the observations must be chronological.

$$Y_t = Y_1, Y_2, Y_3, ..., Y_t$$

A sequence of successive data points in time.

(e) The first lag of a time series Y_t is Y_{t-1}. The jth lag is Y_{t-j}

(f) The first difference of a time series Y_t is $\Delta Y_t = Y_t - Y_{t-1}$. In terms of logarithm-form difference, $\Delta \ln Y_t = \ln Y_t - \ln Y_{t-1}$

(g) The percentage change of a time series Y_t between periods $t-1$ and t is approximately $100 \times \Delta \ln Y_t$.

2 Question 2

Suppose a data set with 10 variables produces a scree plot that is flat. What does this tell you about the correlation of the variables? What does this suggest about the usefulness of using the first few principal components of these variables in a predictive regression? First of all, if I get a flat scree plot, I think that the fraction of the sample variance of the X's explained by every principal component is the same, which is not very good news. In this situation, it is not at all clear which principal components to select for regression analysis. Even if we pick the first few principal components, it is not 100% useful in terms of regression prediction because the lack of up and down in the scree plot indicates redundancy in the information captured by each principal component thus defeating the purpose of constructing principal components in the first place.

3 Question 3

Consider the time series given in Figure 1 below. Which of these series appears to be non-stationary? Explain. The series in graph a) is stationary. It might look like it has a seasonality pattern at first glance, but with careful examination, it does not. Trends and changing levels rule out graph b) which has different means in different time periods. Trends and changing levels rule out graph c) which has a constant variance but lacks a constant mean. Finally, graph d) has a constant mean, and no seasonality pattern, but its variance fluctuates in all the time periods, thus it is non-stationary.

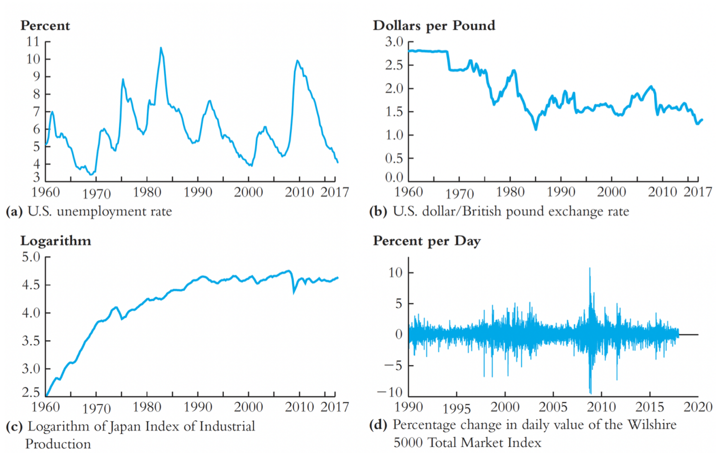


Figure 1: Four economic time series

4 Question 4

Consider the AR(1) autoregression model $Y_t = \beta_0 + \beta_1 Y_{t\text{-}1} + u_t$. Suppose the process is stationary

(a) In a stationary time series dataset, its properties such as constant mean and finite variance of the data do not change over time. Since this is given information for 1st order autoregressive model, stationarity tells us $E(Y_t) = \mu_t$, where μ_t is the mean of of Y_t at time t.

$$E(Y_t) = \mu_t = \mu_{t-1} = E(Y_{t\text{-}1})$$

(b) In a stationary time series, zero conditional mean $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$ is assumed. Based on the result of a), we get

$$Y_t = \beta_0 + \beta_1 Y_{t\text{-}1} + u_t$$

$$E(Y_t) = E(\beta_0 + \beta_1 Y_{t\text{-}1} + u_t)$$

$$E(Y_t) = \beta_0 + \beta_1 E(Y_{t\text{-}1}) + E(u_t)$$

$$E(Y_t) = \beta_0 + \beta_1 E(Y_t) + 0$$

$$E(Y_t) - \beta_1 E(Y_t) = \beta_0$$

$$(1 - \beta_1)E(Y_t) = \beta_0$$

$$(1 - \beta_1)E(Y_t) = \beta_0$$

$$E(Y_t) = \frac{\beta_0}{1 - \beta_1}$$

5 Question 5

(a) We can use the method of Lagrange multipliers. The Lagrangian for this optimization problem is given by:

$$\mathcal{L}(w_1, w_2, \gamma) = Var(w_1 X_1 + w_2 X_2) + \gamma(w_1^2 + w_2^2 - 1)$$

$$\mathcal{L}(w_1, w_2, \gamma) = w_1^2 Var(X_1) + w_2^2 Var(X_2) + 2Cov(w_1 X_1, w_2 X_2) + \gamma(w_1^2 + w_2^2 - 1)$$

$$\mathcal{L}(w_1, w_2, \gamma) = w_1^2 + w_2^2 + 2w_1 w_2 Cov(X_1, X_2) + \gamma(w_1^2 + w_2^2 - 1)$$

$$\mathcal{L}(w_1, w_2, \gamma) = w_1^2 + w_2^2 + 2w_1 w_2 \rho + \gamma w_1^2 + \gamma w_2^2 - \gamma$$

Next, we take the derivative of the Lagrangian with respect to w_1 and w_2 and set it to 0 yields:

$$\frac{\partial \mathcal{L}(w_1, w_2, \gamma)}{\partial w_1} = 2w_1 + 2w_2\rho + 2\gamma w_1 = 0$$

$$\frac{\partial \mathcal{L}(w_1, w_2, \gamma)}{\partial w_2} = 2w_2 + 2w_1\rho + 2\gamma w_2 = 0$$

$$\frac{\partial \mathcal{L}(w_1, w_2, \gamma)}{\partial \gamma} = w_1^2 + w_2^2 - 1 = 0$$

Let's do our algebra in a system of equation and calculate the gradient.

$$\nabla w_1, w_2, \gamma \mathcal{L}(w_1, w_2, \gamma) = (\frac{\partial \mathcal{L}}{\partial w_1}, \frac{\partial \mathcal{L}}{\partial w_2}, \frac{\partial \mathcal{L}}{\partial \gamma}) = 0$$

$$\nabla w_1, w_2, \gamma \mathcal{L}(w_1, w_2, \gamma) = \begin{cases} 2w_1 + 2w_2\rho + 2\gamma w_1 = 0 \\ 2w_2 + 2w_1\rho + 2\gamma w_2 = 0 \\ w_1^2 + w_2^2 - 1 = 0 \end{cases}$$

And... I am stuck again! Okay, even though I can't get answer through calculation. I will attach a picture here to show that the optimization or maximization occurs at the point $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ along the diagonal lines (with slope -1) in the unit circle.

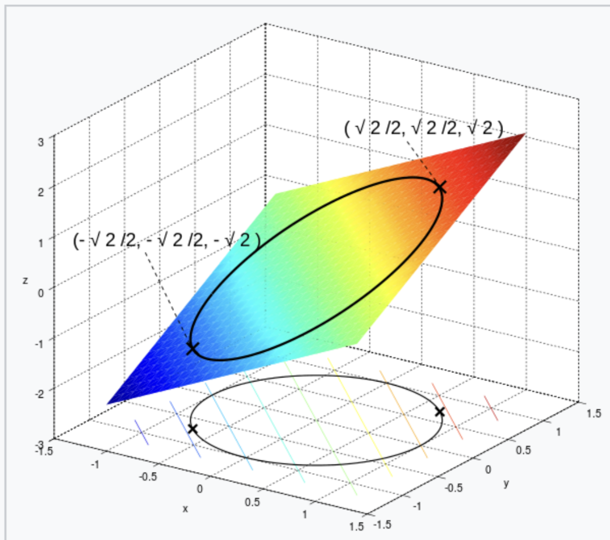


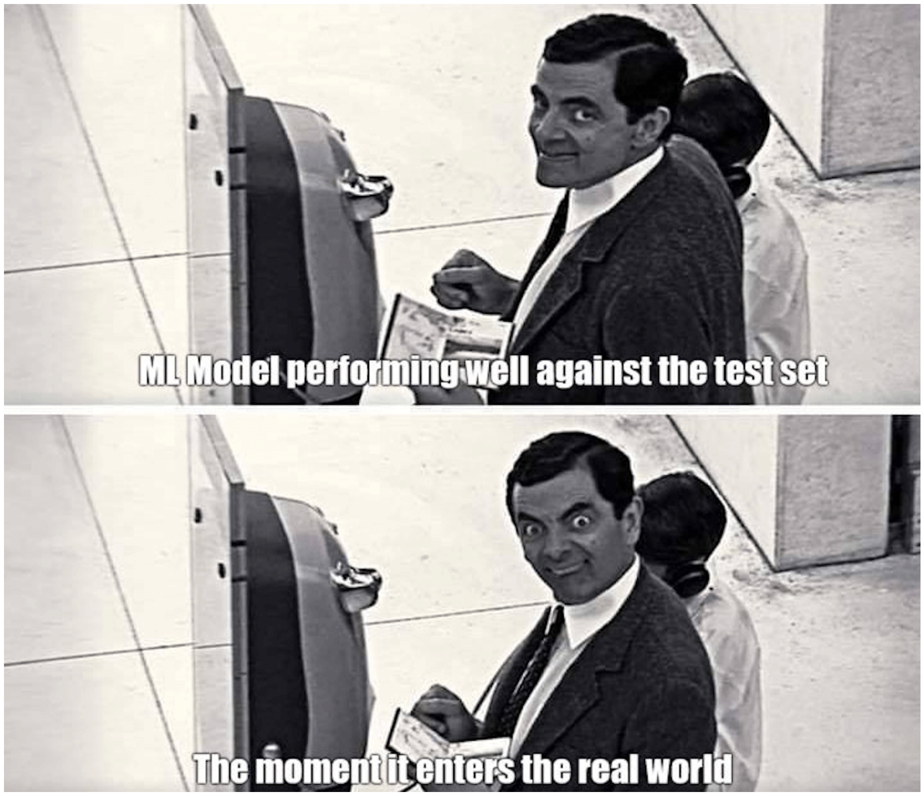
Illustration of the constrained optimization

Now let's pretend I've got w_1 and w_2 both to be $\frac{1}{\sqrt{2}}$.

$$PC_1 = w_1 X_1 + w_2 X_2 = \frac{(X_1 + X_2)}{\sqrt{2}}$$

(b) B

(c) C


ML Model performing well against the test set
The moment it enters the real world