Macro-Finance:
Class 2 (Regression Background, and Applications)

Imperial College London
Department of Finance

Today's Outline

- Import data and visualise, matrices, loops, etc. in R
- Regression (fit, properties, R^2, etc.)
- Rolling regression (finance examples, implications)
- Business Cycles

Multivariate Regression

Suppose a dataset provides returns $r_{t,j}$ over $t = 1, \ldots, T$ periods and for $j = 1, \ldots, k$ assets. One multivariate linear regression analysis based on this dataset can be expressed as:

$$y_t = \beta_0 + \beta_1 r_{t,1} + \beta_2 r_{t,2} + \ldots + \beta_k r_{t,k} + u_t \tag{1}$$

we can arrange the return data as the explanatory variables into a $T \times (k+1)$ matrix $\boldsymbol{x}$ and the $T \times 1$ outcome (dependent) variable as vector y:

$$\underset{T\times 1}{\boldsymbol{y}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ r_{T-1,1} \\ r_{T,1} \end{pmatrix}, \quad \underset{T\times(k+1)}{\boldsymbol{x}} = \begin{array}{c} \\ \texttt{t=1} \\ \texttt{t=2} \\ \texttt{t=3} \\ \vdots \\ \texttt{t=T} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Intercept} & \text{Asset 1} & \text{Asset 2} & \ldots & \text{Asset } k \end{array} \\ \begin{pmatrix} 1 & r_{1,1} & r_{1,2} & \ldots & r_{1,k} \\ 1 & r_{2,1} & r_{2,2} & \ldots & r_{2,k} \\ 1 & r_{3,1} & r_{3,2} & \ldots & r_{3,k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & r_{T,1} & r_{T,2} & \ldots & r_{T,k} \end{pmatrix} \end{array}$$

- each row is collection of asset returns within the same period
- whereas each column is the collection of observations of the same assets.

The regression can be written as:

$$
\begin{aligned}
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} &= \begin{bmatrix} 1 & r_{11} & r_{12} & \dots & r_{1k} \\ 1 & r_{21} & r_{22} & \dots & r_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & r_{T1} & r_{T2} & \dots & r_{Tk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \\
\underset{T\times 1}{\boldsymbol{y}} &= \begin{bmatrix} \underset{T\times 1}{\mathbf{1}} & \underset{T\times 1}{\boldsymbol{r}_1} & \underset{T\times 1}{\boldsymbol{r}_2} & \cdots & \underset{T\times 1}{\boldsymbol{r}_k} \end{bmatrix} \boldsymbol{\beta} + \underset{T\times 1}{\boldsymbol{u}} \\
\boldsymbol{y} &= \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{u}
\end{aligned}
$$

where

$$
\underset{(k+1)\times 1}{\widehat{\boldsymbol{\beta}}} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} \tag{2}
$$

- full rank property of $\boldsymbol{x}$ is required for the inverse to exist
- matrix multiplications directly gives the estimates in computational packages e.g. R
- vector $\widehat{\boldsymbol{\beta}}$ includes $k+1$ scalar estimated parameters
- the first element of which is the regression intercept
- subsequent k elements are attributed to k assets used in the original regression in (1).

Autoregressive Specification

An autoregressive framework models variations of a time-series e.g. returns, as a function of its past. A simple example of this model is an AR(1) model that considers only the most recent observation to be an explanatory variable to predict the variations of the outcome variable:

$$r_t \quad = \quad \beta_0 + \beta_1 r_{t-1} + u_t \tag{3}$$

in this set up, matrices $\boldsymbol{x}$ and $\boldsymbol{y}$ can be written as:

$$\underset{(T-1)\times 2}{\boldsymbol{x}} \quad = \quad \begin{bmatrix} 1 & r_1 \\ 1 & r_2 \\ \vdots & \vdots \\ 1 & r_{T-1} \end{bmatrix}, \quad \underset{(T-1)\times 1}{\boldsymbol{y}} \quad = \quad \begin{bmatrix} r_2 \\ r_3 \\ \vdots \\ r_T \end{bmatrix} \tag{4}$$

note that the length of the observations in each side of the regression is reduce from T to $T-1$ because of the similarly, the estimated quantities for the parameters are:

$$\underset{2\times 1}{\widehat{\boldsymbol{\beta}}} \quad = \quad \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} \quad = \quad (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} \tag{5}$$

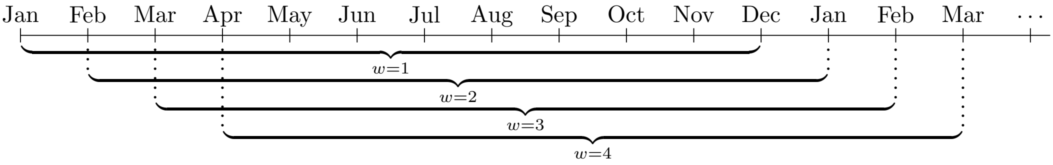
Rolling Regression

Suppose the sample size is T. We wish to estimate a parameter β_w:

$$y_t = x_t \beta_w + u_t$$

within smaller subsamples of size N with $N < T$ such as monthly returns from January to December

- denote each regression with w
- there exists $T - N + 1$ (overlapping) subsamples
- first rolling regression ($w = 1$) uses observations 1 to 12 (January to December)
- second ($w = 2$) uses 2 to 13 (February to January next year), and so on
- till the end of the sample $(T - 11)$ to T
- each regression gives a $\widehat{\beta}_{1,w}$



Example 1 (Persistence)

Consider an AR(1) specification below:

$$r_t = \beta_{0,w} + \beta_{1,w} r_{t-1} + u_t \tag{6}$$

1. Each parameter: autoregressive characteristic, within that window
2. Collectively: time-varying autoregressive characteristics

For instance, the post-war quarterly macroeconomic data

- 1945:Q1 to 2018:Q4
- $T = 296$ quarters
- $N = 16$
- $w = T - N + 1 = 281$

Implication:

1. Each $\widehat{\beta}_{1,w}$ is a measure of persistence of the returns with four years
2. When $\widehat{\beta}_{2,w}$ is high, the series demonstrates high degree of persistence

Example 2 (CAPM)

Consider an example from the capital asset pricing model with the following specification used to interrelate the excess return, on a given asset $r_t - r_{f,t}$ where $r_{f,t}$ is the risk-free rate, to the market return denoted by $r_{m,t}$

$$\underbrace{r_t - r_{f,t}}_{\text{risk premium}} = \alpha_w + \beta_w(\underbrace{r_{M,t} - r_{f,t}}_{\text{market premium}}) + u_t \tag{7}$$

note that the object of interest is the time-varying feature of the coefficient $\widehat{\beta}_w$. Particularly, $\widehat{\beta}_w$ summarizes the conditional relationship (given the rolling window) between the market risk premium $r_{M,t} - r_{f,t}$ and the excess return.

- expected excess returns' sensitivity to the expected excess market returns ($\widehat{\beta}_w$)
- when market risk premium increases, how much the asset return should be?
- zero value classifies investments without (low) risk, e.g. money market funds with a constant share value of $1, certificates of deposit backed by federal deposit insurance, and cash (inflation erodes the purchasing power of money, meaning your zero-beta investment actually loses if it pays interest at less than the rate of inflation.)
- negative beta correlation means an investment moves in the opposite direction from the stock market. Can you give examples?

Example 3 (Portfolio Allocation)

Suppose we run the following regression of 'one' on two risky asset returns:

$$1 = \alpha_{1,w} r_{1,t} + \alpha_{2,w} r_{2,t} + u_t \tag{8}$$

Let $\boldsymbol{\alpha}_w = (\alpha_{1,w}, \alpha_{2,w})'$ and $\mathbf{1}_{2\times 1} = (1,1)'$, then:

$$\underbrace{\boldsymbol{\Omega}_w}_{2\times 1} := \frac{\widehat{\boldsymbol{\alpha}}_w}{\mathbf{1}'_{2\times 1}\widehat{\boldsymbol{\alpha}}_w} \tag{9}$$

where $\boldsymbol{\Omega}_w = (\Omega_{1,w}, \Omega_{2,w})'$ is a 2×1 (estimated) allocation weights on assets 1 and 2.

Question

The rolling regression is a method to split the full sample into rolling subsamples. Can you think of a different way of splitting a dataset to derive additional implications?

Business Cycles

Real Economic Indicators

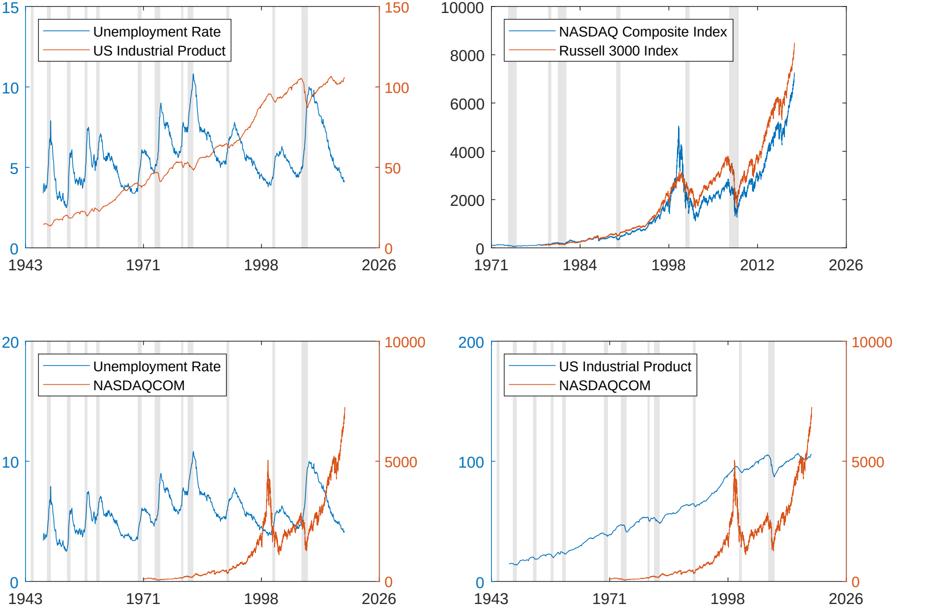
Use the library '`Quandl`' in R to download data directly from the FRED

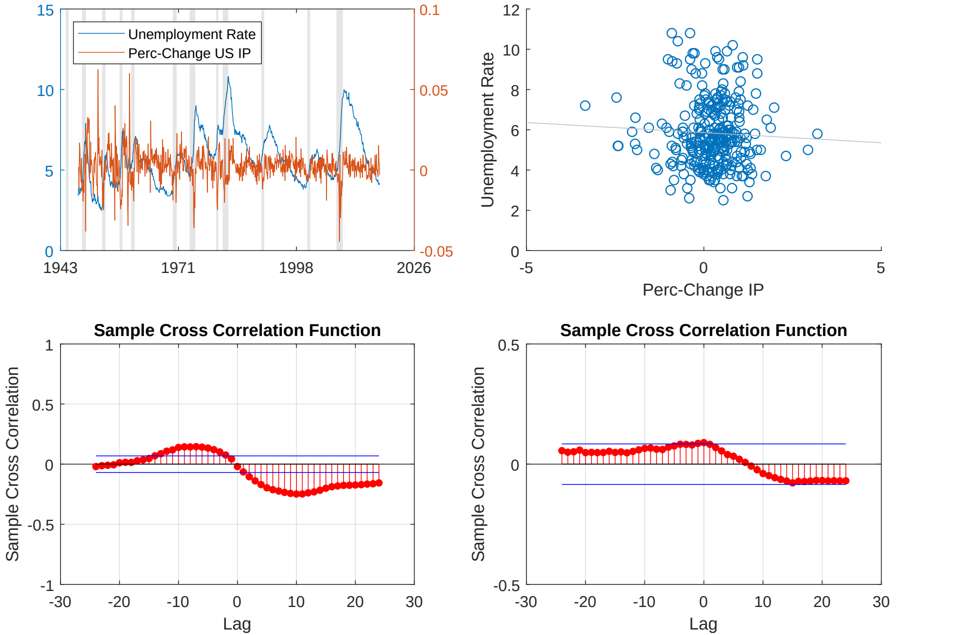
- `install.packages('Quandl')`
- `library(Quandl)`
- unemployment rate, industrial production, NBER-dated recession
- (U.S. series at a monthly frequency

Figures depict the following variables:

1. '`UNRATE`': Civilian Unemployment Rate
2. '`INDPRO`': Industrial Production Index
3. '`USRECM`': NBER-based Recession Indicators for the United States
4. '`NASDAQCOM`': NASDAQ Composite Index
5. '`DJIA`': Dow Jones Industrial Average
6. '`RU3000TR`': Russell 3000 Total Market Index

Plot the series and construct a cross-correlation function between the unemployment rate, the industrial production growth rate series and the stock market return index. Figures shows that the unemployment rate rises during recessions and industrial production falls during recessions.


15
10
5
0
150
100
50
0
1943
1971
1998
2026
Unemployment Rate
US Industrial Product
10000
8000
6000
4000
2000
0
1971
1984
1998
2012
2026
NASDAQ Composite Index
Russell 3000 Index
20
10
0
10000
5000
0
1943
1971
1998
2026
Unemployment Rate
NASDAQCOM
200
100
0
10000
5000
0
1943
1971
1998
2026
US Industrial Product
NASDAQCOM


Unemployment Rate
Perc-Change US IP
15
10
5
0
0.1
0.05
0
-0.05
1943
1971
1998
2026
12
10
8
6
4
2
0
Unemployment Rate
-5
0
5
Perc-Change IP
Sample Cross Correlation Function
Sample Cross Correlation
1
0.5
0
-0.5
-1
-30
-20
-10
0
10
20
30
Lag
Sample Cross Correlation Function
Sample Cross Correlation
0.5
0
-0.5
-30
-20
-10
0
10
20
30
Lag

CRSP Dataset

The center for research and security pricing at the University of Chicago:

1. Access these data through the WRDS account in the library under business data, company and financial data information
2. `wrds-web.wharton.upenn.edu/wrds/`
   Create an account and download variables `vwretd` and `vwretx`
3. The intended variables are returns on the US Total Market Index produced by the CRSP that comprises nearly 4,000 constituents across mega, large, small and micro capitalizations, representing nearly 100% of the U.S. investable equity market. Symbols `vwretd` and `vwretx` denote including distribution or dividends and excluding dividends, respectively.
4. These are both equally weighted stocks (1/N if there are N stocks in the index) and value-weighted stocks (the index is calculated by using as weights the value of each stock relative to the total stock market valuation).

Cyclical Properties

Implications

- Unemployment is a lagging indicator (falls after we leave the recession period) since for positive k it has a negative correlation (less than -0.108) after lead k=2. This means that unemployment rate lags industrial production growth which is consistent with the stylized fact that unemployment tends to fall after an economy exits a recession.
- Stock returns seem to rise before industrial production growth rises (for negative k the cross correlation is positive and slightly statistically significant (above 0.1 approximately in this case)).
- Stock returns and the unemployment rate do not seem to be very highly correlated at the monthly frequency (all cross-correlations are less than 0.1 which could be somewhat statistically significant). Why might this be the case? Stock returns might fall in a recession but it takes a lot longer for unemployment to fall than the more volatile stock market. Using the change in unemployment rate might help. Last worksheet indicates that the change in unemployment rate is negatively correlated with the previous 4-10 month increase in the stock market returns.

Exercises

$\mathbb{Q}$ — The unemployment rate is defined as the ratio of:

1. all adults not working to the total population.
2. unemployed to employed members of the labor force.
3. unemployed members of the labor force to the total labor force.
4. discouraged workers to the total population.
5. unemployed members of the labor force to the total population.

$\mathbb{Q}$ — If the U.S. real output is growing, and labor income accounts for about two-thirds of this:

1. the unemployment rate is falling.
2. on average, capital is getting poorer over time.
3. income inequality is decreasing.
4. on average, workers are getting richer over time.
5. we are not getting any better off.

$\mathbb{Q}$ — The natural rate of unemployment is the unemployment rate that would prevail:

1. during changes in the business cycle.
2. if the economy were in neither a boom nor a recession.
3. if people voluntarily left work.
4. during seasonal changes in the economy.
5. if the unemployment rate were zero.

$\mathbb{Q}$ — Cyclical unemployment is the unemployment that results from:

1. prevailing labor market institutions.
2. workers losing jobs during recession.
3. workers changing jobs in a dynamic economy.
4. workers losing jobs during seasonal changes.
5. workers leaving the labor force.

$\mathbb{Q}$ — The present discounted value equation, $\$386 = \$1,000/(1.1)^{10}$, means you:

1. would prefer to receive \$386 today rather than \$1,000 in 10 years.
2. are indifferent between receiving \$386 today and \$1,000 in 10 years.
3. would prefer to receive \$1,000 in 10 years rather than \$386 today.
4. are indifferent between receiving \$386 today and \$1,000 in 100 years.
5. Not enough information is given.

$\mathbb{Q}$ — According to the study conducted by Piketty discussed in the references, between 1970 and 2010, income inequality:

1. fell in the United Kingdom.
2. increased in the United States relative to France.
3. fell in France.
4. fell in the United States faster than in France.
5. was about the same in the United States and France.

$\mathbb{Q}$ — For the profit-maximizing firm, if the real interest rate is greater than the marginal product of capital, the firm should:

1. invest in more capital.
2. get rid of some capital.
3. keep its capital stock the same, as there is a risk premium attached to the real interest rate.
4. hire more workers.
5. buy stocks.