Empirical Finance: Study Questions Week 3
Imperial College London

Hormoz Ramian

Question 1 This question requires working with CRSP dataset with the following series:

— Monthly returns data on a value-weighted market portfolio including dividends reinvested (variables `vwretd`) and excluding dividends (variable: `vwretx`)[1]. This is based 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[2].

— Quarterly short term interest rates (US one-year maturity t-bills, variable: `b1ret`) and quarterly CPI returns (inflation, variable: `cpiret`).

Further variables description is provided by the CRSP[3]. Use this data to answer the following questions:

(1.1) Explain briefly how to construct the dividend-yield from the variables provided. Implement a code to construct the dividend-yield from the series on quarterly and annual frequencies. Plot both constructed series (clearly label axes, years, etc.).

(1.2) Implement a code to use the dividend-yield series to examine whether it can forecast excess returns in the stock market over the full sample. Include the regression specification you suggest to use in this analysis and briefly explain each variable and their units. Interpret the value of the adjusted R-squared and the statistical properties of different estimated coefficients. What is your conclusion about the ability of this regression to forecast excess returns?

(1.3) Implement a code to run rolling regressions using all overlapping 15-year consecutive subsamples, and the annual frequency series. What do you notice about the adjusted R-squared obtained from the rolling regressions? Comment.

(1.4) Perform an out-of-sample forecast by predicting one-step ahead excess return, given a window size equal to 60 quarters.

(1.5) How does adding extra lags and additional explanatory variables change the forecasting?

Question 2 The model is:

$$y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + u_i \tag{1}$$

Assume that,

$$\widehat{\beta} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' + \lambda \boldsymbol{I}_k \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i \right)$$

Can you think of a case where $\widehat{\beta}$ takes the form above, and what would be the main purpose of such regression?

[1] Visit `crsp.com/products/documentation/stock-file-indexes`
[2] You can download the data from WRDS `wrds-web.wharton.upenn.edu/wrds/`
[3] Visit `crsp.com/products/documentation/file-specifications`

Question 3 Consider the following date:

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}
1	0	—	—	—
2	1	0	—	—
3	2	1	0	—
4	0	2	1	0
5	1	0	2	1
6	1	1	0	2
7	3	1	1	0
8	3	3	1	1
9	2	3	3	1
10	0	2	3	3

(3.1) Suppose we wish to choose one of the following models using the data during only $t = 4, \ldots, 7$. Discuss what method you would use to decide which model is the best specification.

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + u_{t1} \\ y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_1 y_{t-2} + u_{t2} \\ y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + u_{t3} \end{aligned}$$

(3.2) Show that $\mathrm{MSE}[\widehat{y}_{t+h}] = \mathrm{var}[\widehat{y}_{t+h}] + \mathrm{bias}[\widehat{y}_{t+h}]^2$ for a forecast value $\widehat{y}_{t+h}$ where h is the forecast horizon.

Answers

Question 1

(1.1) Dividends represent one of the primary means by which invested capital is returned to common stockholders, and dividend-yield is the financial ratio that measures the dividends payout, in cash or equivalent, to shareholders relative to the market value per share[4]. It is computed by dividing the dividend per share by the market price per share and multiplying the result by 100. A company with a high dividend yield pays a substantial share of its profits in the form of dividends. Dividend yield of a company is always compared with the average of the industry to which the company belongs.

Companies distribute a portion of their profits as dividends, while retaining the remaining portion to reinvest in the business. Dividends are paid out to the shareholders of a company. Dividend yield measures the quantum of earnings by way of total dividends that investors make by investing in that company. It is normally expressed as a percentage. The formula for computing the dividend yield is

$$\text{dividend-yield} \quad = \quad \frac{\text{cash dividend per share}}{\text{market price per share}} \times 100$$

Suppose a company with a stock price of \$100 declares a dividend of \$10 per share. In that case, the dividend yield of the stock will be $\frac{10}{100} \times 100 = 10\%$. High dividend yield stocks are good investment options during volatile times, as these companies offer good payoff options. They are suitable for risk-averse investors. The caveat is, investors need to check the valuation as well as the dividend-paying track record of the company. Companies with high dividend yield normally do not keep a substantial portion of profits as retained earnings. Their stocks are called income stocks. This is in contrast to growth stocks, where the companies retain a major portion of the profit in the form of retained earnings and invest that to grow the business.

Construction from the Data Dividend-yield is a real variable constructed by the dividend per share, divided by the share price. Let P_t and D_t denote the share price and dividend payout, then:

$$R_{t+1} \quad = \quad \frac{P_{t+1}}{P_t} \tag{2}$$

$$R^d_{t+1} \quad = \quad \frac{P_{t+1} + D_{t+1}}{P_t} \tag{3}$$

where R_{t+1} and R^d_{t+1} are the simple (gross) stock return without and with dividend, respectively. Dividing equation (3) by (2) gives:

$$\frac{R^d_{t+1}}{R_{t+1}} \quad = \quad \frac{P_{t+1} + D_{t+1}}{P_t} / \frac{P_{t+1}}{P_t} \quad = \quad 1 + \frac{D_{t+1}}{P_{t+1}}$$

[4]Most firms pay cash dividends on a quarterly basis. The dividend is declared by the firm's board of directors on a date known as the announcement date. The board's announcement states that a cash payment will be made to stockholders who are registered owners on a given record date. The dividend checks are mailed to stockholders on the payment date, which is usually about two weeks after the record date. Stock exchange rules generally dictate that the stock is bought or sold with the dividend until the ex-dividend date, which is a few business days before the record date. After the ex-dividend date, the stock is bought and sold without the dividend.

then

$$\frac{D_{t+1}}{P_{t+1}} = \frac{R^d_{t+1} - R_{t+1}}{R_{t+1}} = \frac{R^d_{t+1}}{R_{t+1}} - 1 \tag{4}$$

where D_{t+1}/P_{t+1} is the dividend yield (numerator and denominator are nominal). The CRSP dataset provides the US total market index comprising nearly 4,000 constituents across mega, large, small and micro capitalizations, representing nearly 100% of the U.S. investable equity market. Monthly returns data on a value-weighted market portfolio including dividends reinvested (variables `vwretd`) and excluding dividends (variable: `vwretx`)[5]. Thus re-writing equation (4) gives:

$$\frac{D_{t+1}}{P_{t+1}} = \frac{\texttt{vwretd}_{t+1}}{\texttt{vwretx}_{t+1}} - 1 \tag{5}$$

where the subscript t denotes a monthly observation. We often need to convert monthly data to quarterly or annual when working with larger datasets, including multiple variables with various sampling frequencies, to provide regression analysis[6].

An observable time-series with monthly basis such as returns can be re-arranged to construct multi-period returns over a longer horizon such as quarterly:

$$\begin{aligned} R_{t,t+3} &= \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}} \\ &= \underbrace{(1 + r_{t+1})}_{\text{monthly}} \times \underbrace{(1 + r_{t+2})}_{\text{monthly}} \times \underbrace{(1 + r_{t+3})}_{\text{monthly}} = \prod_{i=1}^{3}(1 + r_{t+i}) \end{aligned}$$

where the notation $R_{t,t+3}$ indicates return between 3 consecutive months. The same principle applies to returns with dividend:

$$\begin{aligned} R^d_{t,t+3} &= \frac{P_{t+1} + D_{t+1}}{P_t} \times \frac{P_{t+2} + D_{t+2}}{P_{t+1}} \times \frac{P_{t+3} + D_{t+3}}{P_{t+2}} \\ &= \underbrace{(1 + r^d_{t+1})}_{\text{monthly}} \times \underbrace{(1 + r^d_{t+2})}_{\text{monthly}} \times \underbrace{(1 + r^d_{t+3})}_{\text{monthly}} = \prod_{i=1}^{3}(1 + r^d_{t+i}) \end{aligned}$$

other conversions follow this logic, except that the number of multiplicative returns should adjust according to the final sampling frequency of interest e.g. for annual, twelve consecutive monthly returns should be considered.

Re-arranging equation (3) gives:

$$R^d_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \underbrace{\frac{D_t}{P_t}}_{\text{Dividend Yield}_t} \cdot \underbrace{\frac{D_{t+1}}{D_t}}_{\text{Dividend Growth}} + \underbrace{\frac{P_{t+1}}{P_t}}_{\text{capital gain}}$$

(1.2) Consider the *nominal* price of a risk-free asset (e.g. Treasuries), B_t and the percentage change

[5]Visit `crsp.com/products/documentation/stock-file-indexes`

[6]More sophisticated mixed-data regression methods such as a MIDAS regression is able to incorporate variables with different sampling frequencies but this remains beyond the scope of this course (interested readers see for example https://mpiktas.github.io/midasr//) and https://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/midas-regressions.pdf.

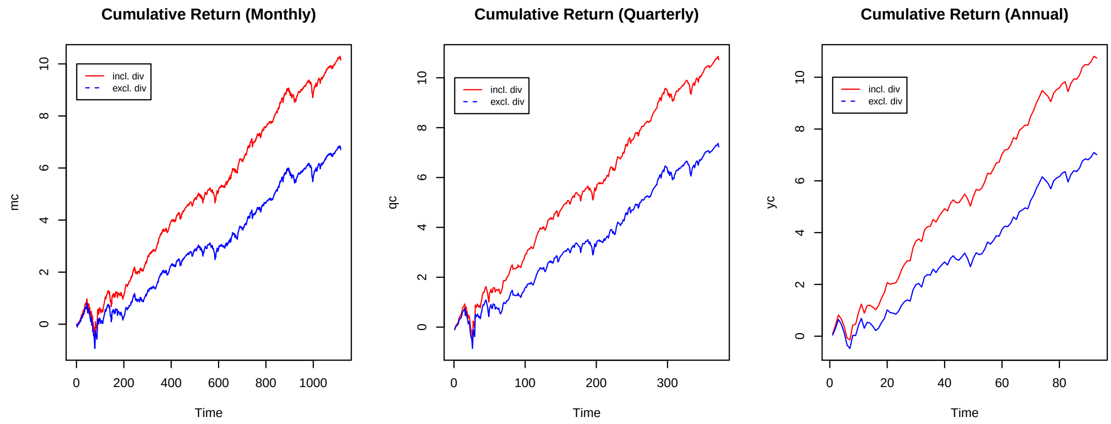
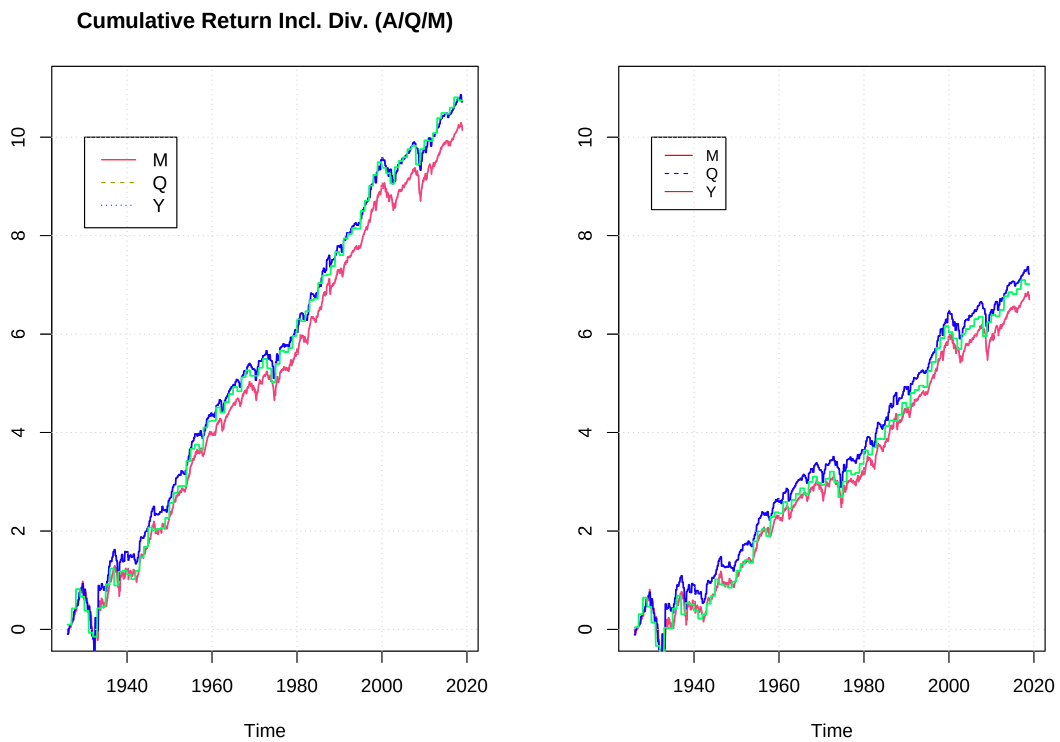


Figure 1: The figure on the left shows cumulative returns during 1926 to 2018 including 93 years of total market index returns on monthly bases. Figures in the middle and on the right show the same quantities but on quarterly and annual basis. Each figure shows cumulative returns including and excluding dividend.



consumer price index[7], $\pi_{t+1} = \log \frac{Q_{t+1}}{Q_t}$, then:

$$\log R_{f,t+1} \quad = \quad \log \frac{B_{t+1}}{B_t} - \log \frac{Q_{t+1}}{Q_t}$$

[7]Macroeconomic data are published by different agencies at different calendar dates e.g. Bureau of Labor Statistics or the Federal Reserves in the US or the Office for National Statistics or the Bank of England in the UK each releases various macroeconomic indicators such as CPI, growth, etc. along different dates during a year. Staggered publication lags potentially affect statistical inference mainly because an earlier release of one indicator may affect the economy in a certain way that results in a real change in a subsequent release of some other indicators in the same quarter. However, often data releases in the same quarter regardless of publication dates are considered concurrent observations. Further statistical models e.g. nowcasting exploit such lags to provide further corrections. This topic is not part of this exercise

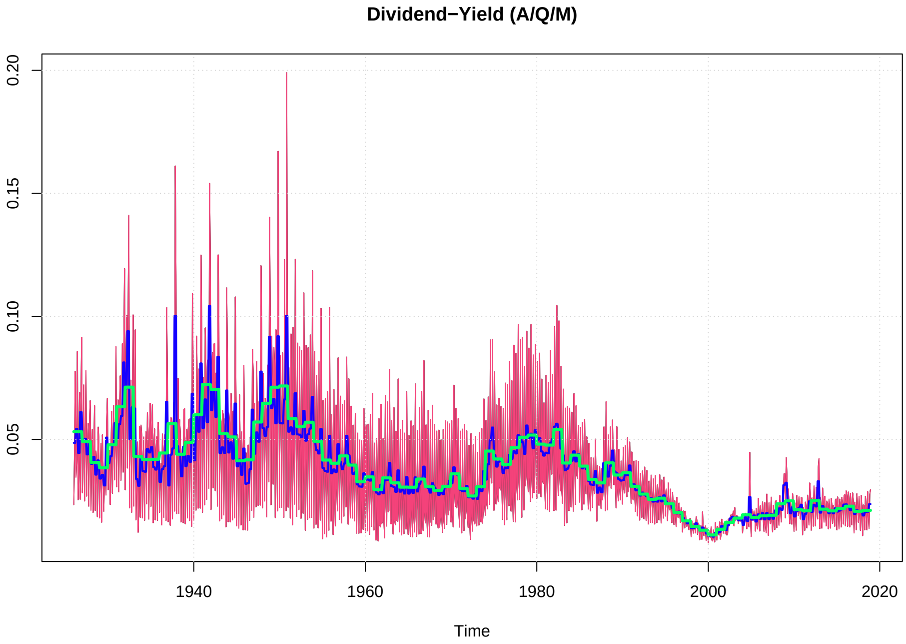


Figure 2: The figure depicts dividend-yield on monthly basis (red), quarterly basis (blue) and annual basis green) from 1926 to 2018.

is the (real, log) risk-free rate, and:

$$\begin{aligned}
\log X_{t+1} &= \log R_{t+1} - \log R_{f,t+1} \\
&= \left(\log \frac{P_{t+1}}{P_t} - \log \frac{Q_{t+1}}{Q_t}\right) - \left(\log \frac{B_{t+1}}{B_t} - \log \frac{Q_{t+1}}{Q_t}\right) \\
&= \log \frac{P_{t+1}}{P_t} - \log \frac{B_{t+1}}{B_t} \\
&= \log R^{\$}_{t+1} - \log R^{\$}_{f,t+1}
\end{aligned}$$

is the (real, log) excess return where $R^{\$}_{t+1}$ is the return .... We wish to run the following regression to examine how dividend yield (a real variable) is able to explain variations of excess returns (measured in log units) in the next period:

$$\underbrace{\log R^{\$}_{t+1} - \log R^{\$}_{f,t+1}}_{\text{excess return}} = \beta_0 + \beta_1 \left(\frac{\texttt{vwretd}_t}{\texttt{vwretx}_t} - 1\right) + u_{t+1}$$

note that the left hand side can be approximated as:

$$\log R^{\$}_{t+1} - \log R^{\$}_{f,t+1} \approx \frac{R^{\$}_{t+1}}{\pi_{t+1}} - \frac{R^{\$}_{f,t+1}}{\pi_{t+1}} \tag{6}$$

denote $x_t = \log X_{t+1}$ and re-write as:

$$x_{t+1} = \beta_0 + \beta_1 \frac{D_t}{P_t} + u_{t+1}$$

Theory suggests that some amount of predictability is likely. If expected returns vary over time with some degree of persistence, predictability is expected. Studies find that predictabil-

or this course but interested readers see e.g. https://www.economics.ox.ac.uk/materials/papers/13019/paper674.pdf.

ity using lagged variables is largely explained by asset pricing models if they allow the risk premiums to vary over time.

(1.3) Rolling Regression

(1.4) Forecasting Statistical tests of a model's forecast performance are commonly conducted by splitting a given data set into an in-sample period, used for the initial parameter estimation and model selection, and an out-of-sample period, used to evaluate forecasting performance. Empirical evidence based on out-of-sample forecast performance is generally considered more trustworthy than evidence based on in-sample performance, which can be more sensitive to outliers and data mining. Out-of-sample forecasts also better reflect the information available to the forecaster in real time.

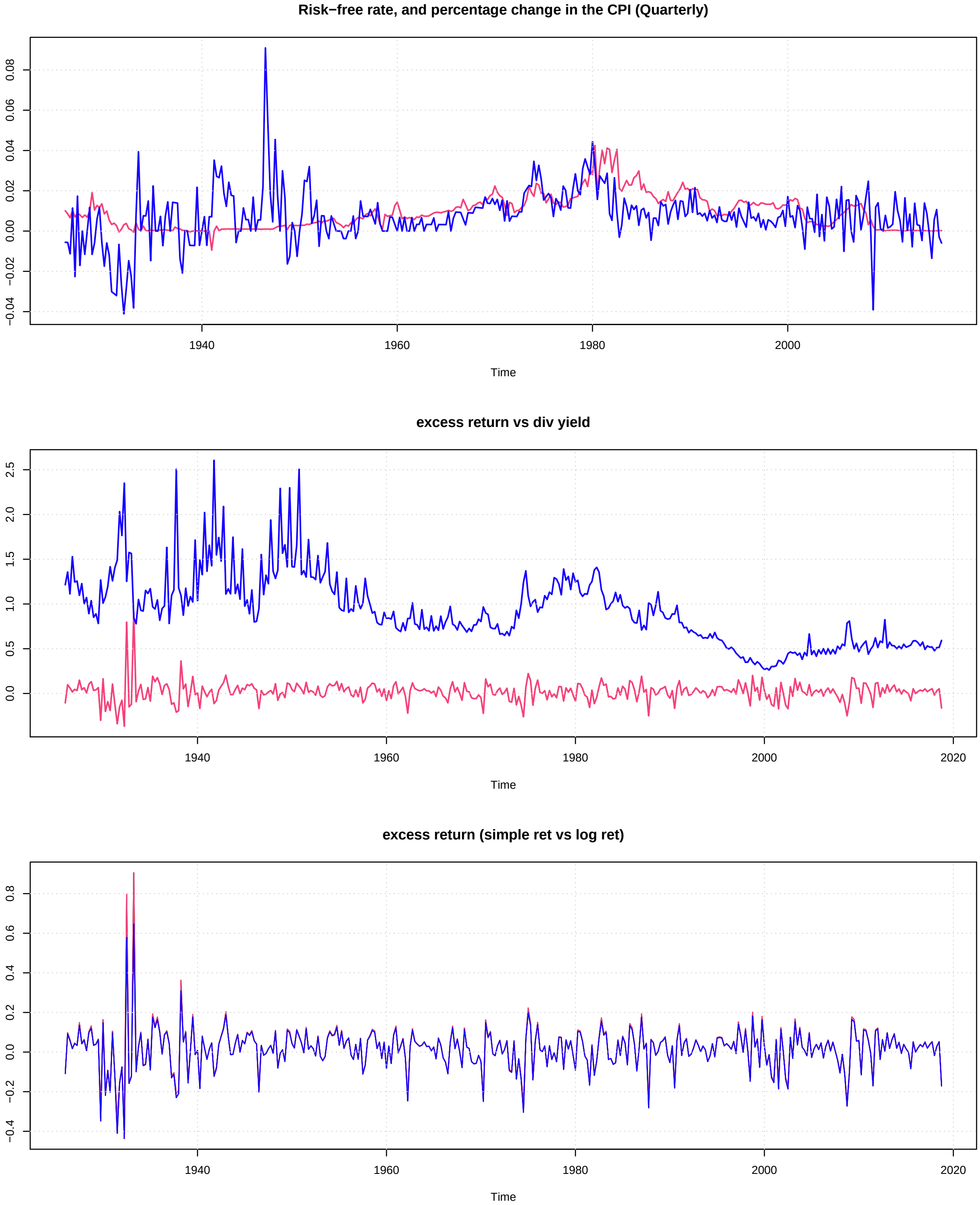


Figure 3: The figure on the first row shows percentage change in CPI or inflation rate, and the percentage change in the Treasury bills or the risk-free rate. The figure in the middle row shows the dividend-yield on quarterly basis and the real excess returns. The figure on the last row shows that log returns and simple return are almost overlapping over the entire length of the data. However, during the periods of high volatility specifically when prices deviate from their previous values, log returns and simple return can be poor approximations. This can be seen during 1930s where the red and blue series are different when returns show large spikes.

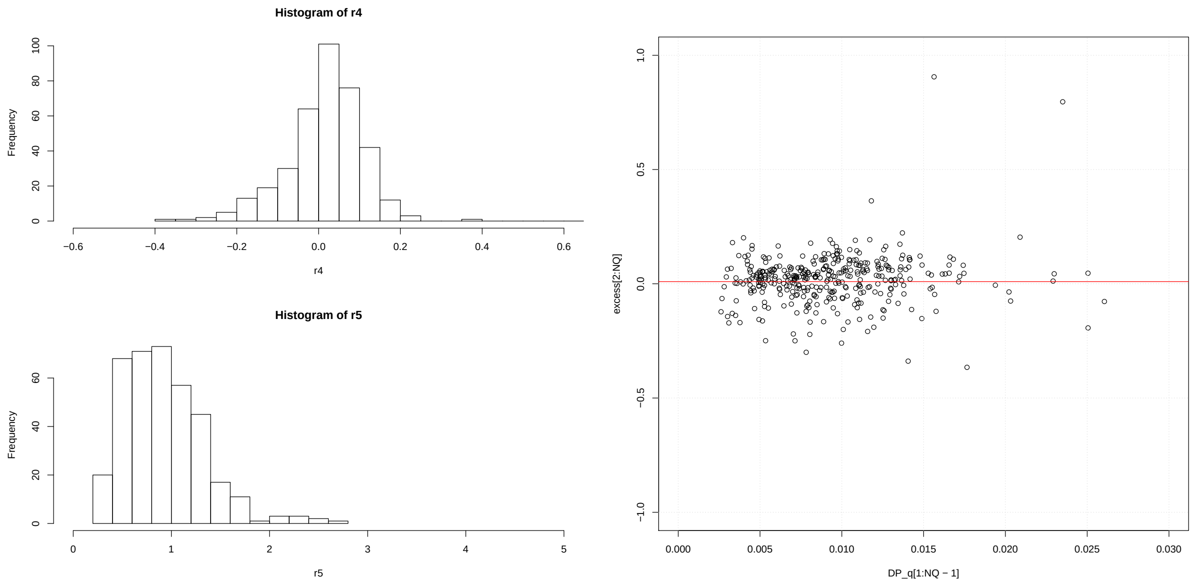


Figure 4: The figure on the top-left shows the empirical histogram of real excess returns, and the figure on bottom-left shows the same diagram for dividend-yield which is always a positive quantity as stocks are protected by limited liability. The figure on the right shows scatter plot of lagged dividend yield on the horizontal axis and present real excess returns on the vertical axis.

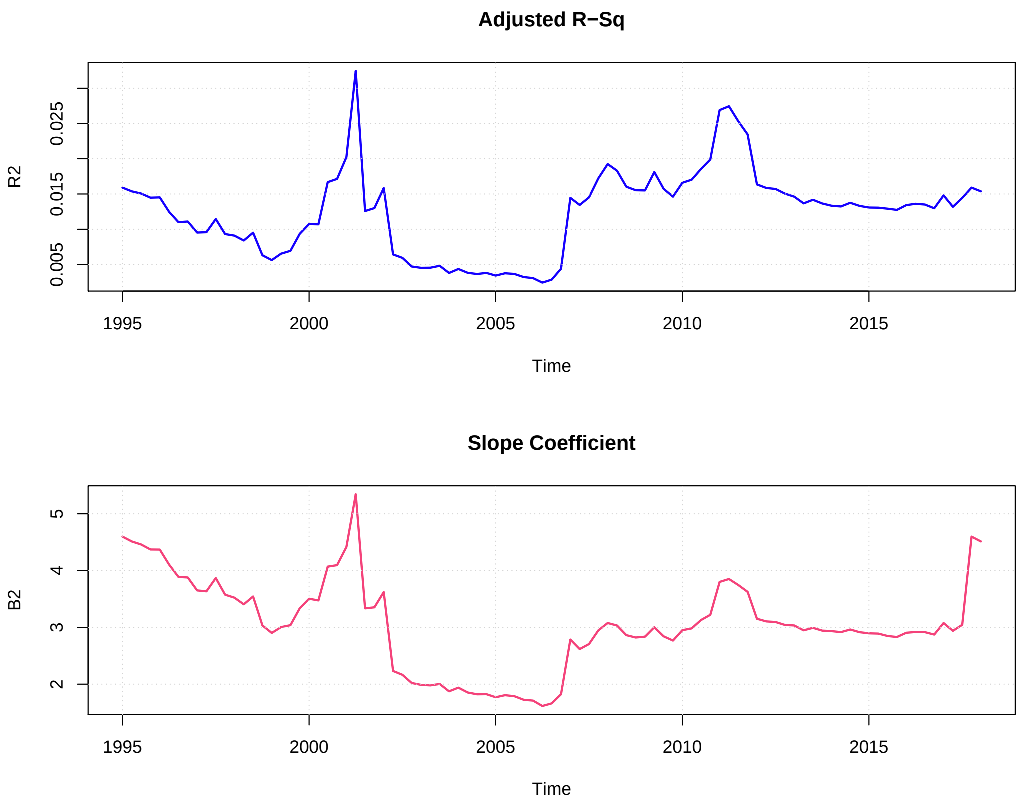
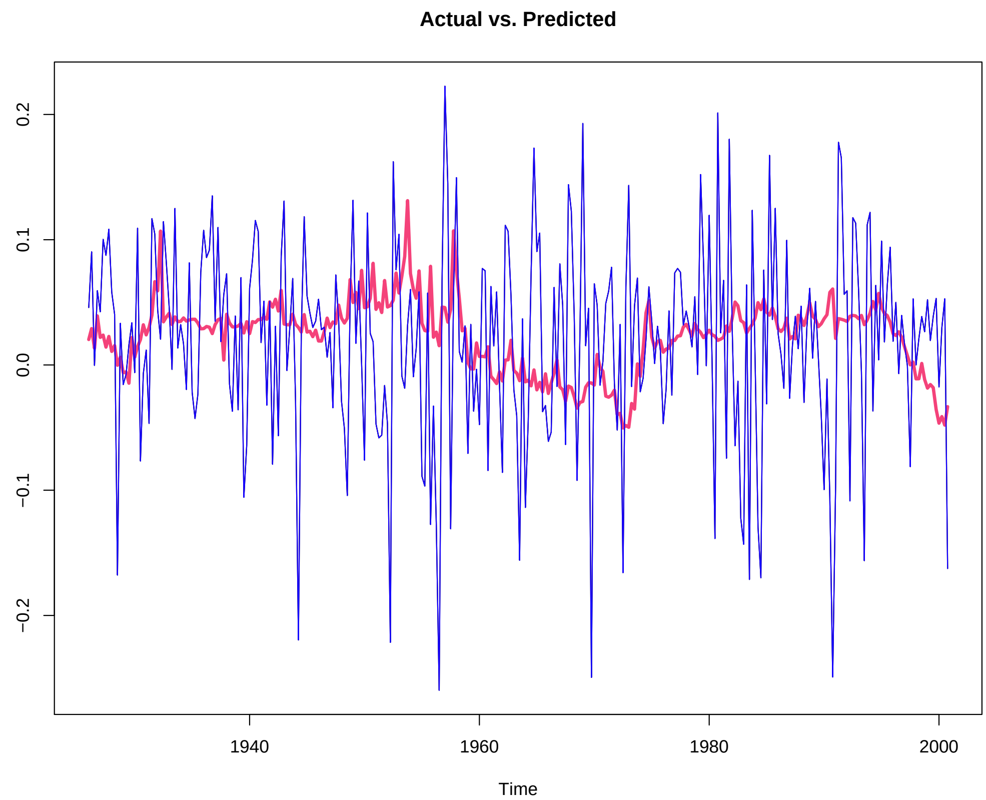
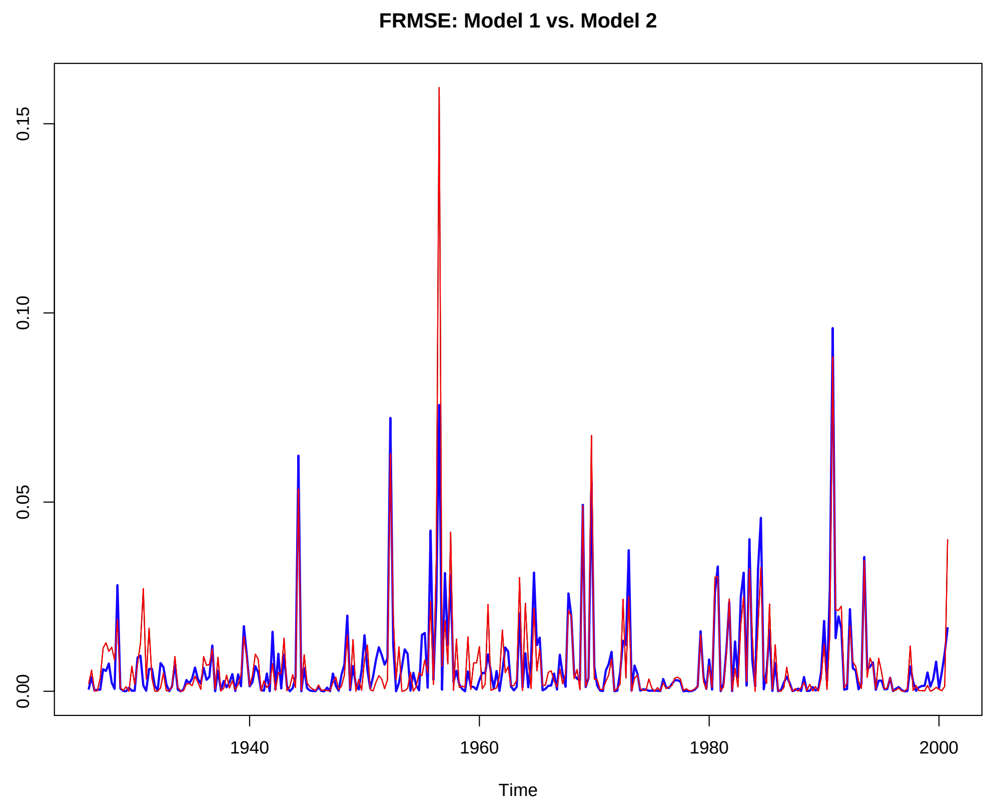
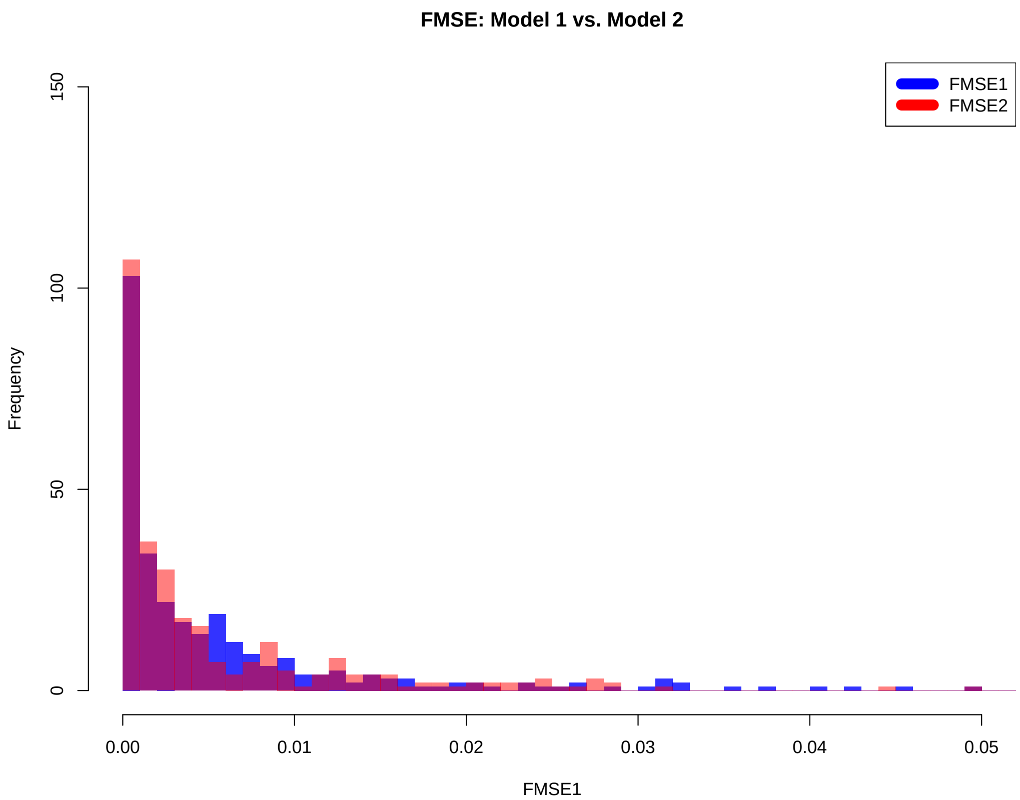


Figure 5: The figure on the first row shows the adjusted R^2 with a falling value over recent years and the second figure shows the estimated slope coefficient from the univariate regression on rolling basis.


Actual vs. Predicted
0.2
0.1
0.0
−0.1
−0.2
1940
1960
1980
2000
Time


FRMSE: Model 1 vs. Model 2
0.15
0.10
0.05
0.00
1940
1960
1980
2000
Time



FMSE: Model 1 vs. Model 2
FMSE1
FMSE2
Frequency
0
50
100
150
0.00
0.01
0.02
0.03
0.04
0.05
FMSE1

Question 2 The reason to introduce this additional term is that the inverse form $\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}$ only exists if $\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)$ is positive-definite and full rank. However, even if we have a full rank matrix, still this may be ill-conditioned, indicating that the number observations are very close to the number of explanatory variables. In this case the usual OLS estimator is poorly estimated. In particular, confidence intervals are very large and estimates largely depend on individual observation perturbation. The ridge regression estimator overcomes this issue via introducing the additional term. As we see in the derivations above the impact fades as sample size increases.

Optional: Find probability limit of $\widehat{\boldsymbol{\beta}}$ as $n \to \infty$. Assuming the $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{y}$ $(n \times 1)$ and $\boldsymbol{X}$ $(n \times k)$ and the property that $\mathbb{E}(\boldsymbol{x}_i e_i) = 0$. The ridge regression estimator:

$$\widehat{\beta} = \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_i y_i\right)$$

Find probability limit of $\widehat{\boldsymbol{\beta}}$ as $n \to \infty$

$$\begin{aligned}
\widehat{\beta} &= \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_i y_i\right) \\
&= \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{x}_i(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{e})\right) \\
&= \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\beta} + \left(\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \lambda\boldsymbol{I}_k\right)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{e}
\end{aligned}$$

Dividing and multiplying by $\dfrac{1}{n}$ yields:

$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)\boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{e}\right)$$

Let $\widehat{\boldsymbol{Q}}_{xx\lambda}^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}$, $\widehat{\boldsymbol{Q}}_{xx} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$ and $\widehat{\boldsymbol{Q}}_{xe} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{e}$. First, because $\mathbb{E}(\boldsymbol{x}_i e_i) = 0$, we can show that, $\widehat{\boldsymbol{Q}}_{xe} \xrightarrow{p} \boldsymbol{0}$ therefore, we can simplify the following as:

$$\widehat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)\boldsymbol{\beta} + \boldsymbol{0}$$

In the first term, we can show that asymptotically $\frac{1}{n}\lambda\boldsymbol{I}_k$ disappears and we have:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \frac{1}{n}\lambda\boldsymbol{I}_k\right)^{-1} \xrightarrow{p} \boldsymbol{Q}_{xx}^{-1} \text{ because } \frac{1}{n}\lambda\boldsymbol{I}_k \to \boldsymbol{0}_k \text{ as } n \to \infty$$

Re-writing the expression for $\widehat{\beta}$ yields,

$$\begin{aligned}
\widehat{\beta} &\xrightarrow{p} \boldsymbol{Q}_{xx}^{-1}\boldsymbol{Q}_{xx}\boldsymbol{\beta} \\
&\xrightarrow{p} \boldsymbol{\beta}
\end{aligned}$$

which shows that the regression estimator is a consistent estimator for $\boldsymbol{\beta}$. Intuitively, this result holds since the impact of particular form of constraint on the regression model, becomes less important as the sample size increases. Adding the term $\lambda\boldsymbol{I}_k$ is particularly important when the sample size is small which enables the inverse form to be well-conditioned.