QBUS6830 Financial Time Series and Forecasting
S1, 2019

Solutions to Lab Sheet 4

Q1 (Stress testing CAPM)

Get the Kenneth French data on factors and 5 industry portfolios (value weighted) on a monthly frequency. The link is on Blackboard.

(a) Fit the multi-factor CAPM model to each of the five industry group portfolio excess return series.

	α	95% CI for α	β_1	95% CI for β_1	β_2	95% CI for β_2	β_3	95% CI for β_3
Cnsmr	0.396	0.28, 0.52	0.917	0.89, 0.94	-0.003	-0.04, 0.03	0.011	-0.03, 0.05
Manuf	0.328	0.24, 0.42	0.977	0.96, 0.99	0.162	0.14, 0.19	-0.098	-0.13, -0.07
Hi-tech	0.423	0.29, 0.55	0.982	0.96, 1.01	-0.323	-0.36, -0.29	0.035	-0.007,0.08
Health	0.581	0.37, 0.79	0.886	0.85, 0.93	-0.186	-0.24, -0.13	-0.097	-0.16, -0.03
Other	0.088	-0.03, 0.21	1.049	1.03, 1.07	0.374	0.34, 0.41	0.065	0.03, 0.10

Discussed in lab 3.

(b) Estimate the average (and standard deviation of) industry return, for each industry, for each combination of the following cases:

1. *The excess market return is -5%, -10%.*
2. *HML is 0%, -2%*
3. *SMB is 0%, -2%*

For averages, simply insert the values above into the estimated regression equations, with parameter estimates above, e.g. for the Consumer portfolio:

$$E(cnsmr_t | Mkt_t, HML_t, SMB_t) = 0.396 + 0.917Mkt_t - 0.003HML_t + 0.011SMB_t$$

Results are:

Mkt	HML	SMB	Cnsmr	Manuf	Hi-Tech	Health	Other
-5	0	0	-4.1871	-4.5575	-4.4845	-3.8490	-5.1561
-5	0	-2	-4.2081	-4.3608	-4.5543	-3.6559	-5.2854
-5	-2	0	-4.1816	-4.8819	-3.8391	-3.4779	-5.9047
-5	-2	-2	-4.2026	-4.6852	-3.9089	-3.2848	-6.0339
-10	0	0	-8.7700	-9.4426	-9.3920	-8.2790	-10.4002
-10	0	-2	-8.7910	-9.2459	-9.4618	-8.0859	-10.5294
-10	-2	0	-8.7645	-9.7669	-8.7466	-7.9079	-11.1487
-10	-2	-2	-8.7855	-9.5703	-8.8164	-7.7148	-11.2779

Average return estimates are strongly affected by market returns and weakly affected by the other two factors. The Health portfolio has the lowest negative returns, i.e. best predicted performance, at all combinations of the three factor variables; whilst Other has the worst predicted performance at all combinations of the three factor variables, and is predicted to be more extreme than the market in each case.

(c) *For each combination in (b), estimate a Value at Risk and expected shortfall at the 1% level, for each asset, using the Gaussian distribution and a sample percentile approach. Discuss.*

The VaR at 1% is the point in the distribution, say of $cnsmr_t \mid Mkt_t, HML_t, SMB_t$ that, given values for the three factors, only 1% of cnsmr returns would lie below. Mathematically, VaR is the solution to:

$$\Pr\left(cnsmr_t \leq VaR_{0.01} \mid Mkt_t, HML_t, SMB_t\right) = 0.01$$

The ES at 1% is the average of the distribution of $cnsmr_t \mid Mkt_t, HML_t, SMB_t$, given that the cnsmr return is below the 1% VaR level. Mathematically, this is:

$$E\left(cnsmr_t \mid Mkt_t, HML_t, SMB_t, cnsmr_t \leq VaR_{0.01}\right) = ES_{0.01}$$

For the Gaussian distribution, the $ES_{0.01}$ point occurs at the 0.0038 quantile level. i.e. the average beyond the 1% quantile in a Gaussian distribution occurs exactly at the 0.38% quantile. (For a 5% level, the $ES_{0.05}$ point occurs at the 0.0196 quantile level, or the 1.96% quantile.)

Under the Gaussian distribution then

$$VaR_{0.01} = \alpha + \beta_1 Mkt_t + \beta_2 HML_t + \beta_3 SMB_t - \Phi^{-1}(0.01)\sigma$$

$$ES_{0.01} = \alpha + \beta_1 Mkt_t + \beta_2 HML_t + \beta_3 SMB_t - \Phi^{-1}(0.0038)\sigma$$

The first part of the expression above is the regression conditional mean for $cnsmr_t \mid Mkt_t, HML_t, SMB_t$. The second part reflects the choice of Gaussian distribution and homoskedastic, constant variance. We estimate $\hat{\sigma} = SER$, where the values for each asset are below:

SER values

Cnsmr	Manuf	HiTech	Health	Other
1.9237	1.4263	2.1314	3.3549	1.9617

For the sample percentile method, I utilise the estimated residuals to estimate the 1% VaR and ES values. Since, e.g.

$$cnsmr_t = \alpha + \beta_1 Mkt_t + \beta_2 HML_t + \beta_3 SMB_t + \varepsilon_t$$

Estimating the regression gives us a large sample of estimate residuals, $\hat{\varepsilon}_t = cnsmr_t - \hat{\alpha} + \hat{\beta}_1 Mkt_t + \hat{\beta}_2 HML_t + \hat{\beta}_3 SMB_t$, for $t = 1, \ldots, n$. I simply estimate the 1% quantile and then the mean of the residuals from the sample percentile and sample mean beyond that quantile. Thus, under this non-parametric approach:

$$VaR_{0.01} = \alpha + \beta_1 Mkt_t + \beta_2 HML_t + \beta_3 SMB_t + Q\left(\hat{\varepsilon}; 0.01\right)$$

$$ES_{0.01} = \alpha + \beta_1 Mkt_t + \beta_2 HML_t + \beta_3 SMB_t + \bar{\varepsilon}\left(\varepsilon_t \leq Q\left(\hat{\varepsilon}; 0.01\right)\right)$$

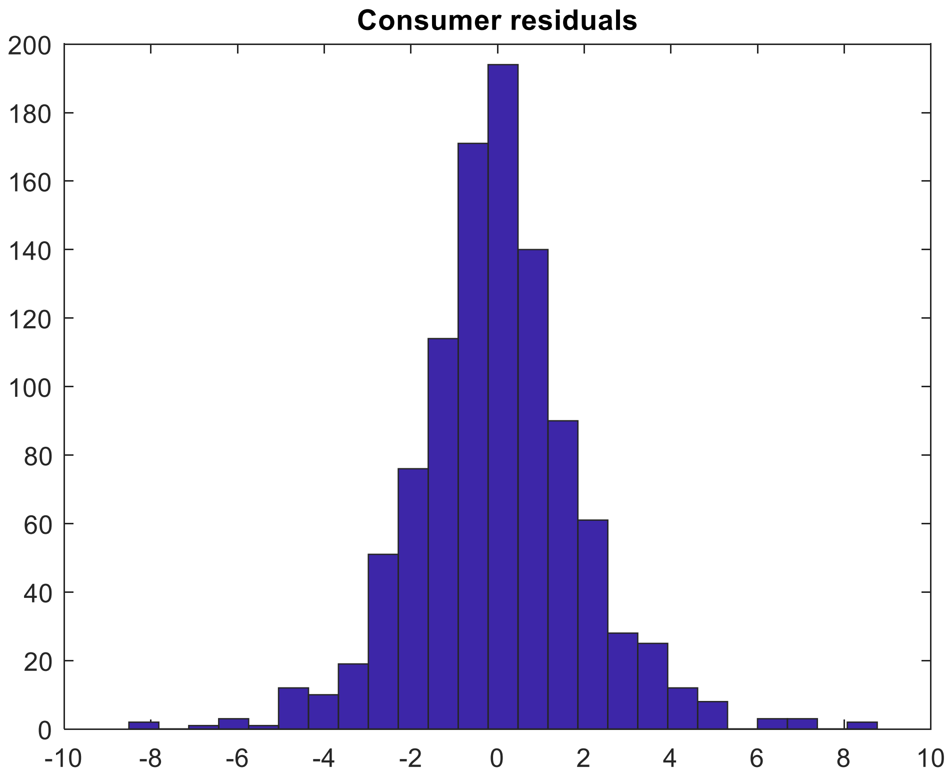
Where Q represents the sample percentile and $\bar{\varepsilon}$ the sample mean (with brackets indicating that only those residuals beyond the sample percentile estimate were considered)

The following estimated results were obtained for each industry portfolio:

Consumer

Mkt	HML	SMB	Mean	Gaussian VaR	Gaussian ES	Sample percentiles VaR	Sample percentiles ES
-5	0	0	-4.187	-8.662	-9.322	-9.034	-10.379
-5	0	-2	-4.208	-8.683	-9.343	-9.055	-10.400
-5	-2	0	-4.182	-8.656	-9.316	-9.029	-10.374
-5	-2	-2	-4.203	-8.677	-9.337	-9.050	-10.395
-10	0	0	-8.770	-13.245	-13.905	-13.617	-14.962
-10	0	-2	-8.791	-13.266	-13.926	-13.638	-14.983
-10	-2	0	-8.764	-13.239	-13.899	-13.612	-14.957
-10	-2	-2	-8.785	-13.260	-13.920	-13.633	-14.978

The variation in HML and SMB has almost no effect on any of these quantities. The variation in Market premium from -5 to -10% has a very large effect on all the quantities. Further, whether assuming a Gaussian distribution, or using sample percentiles, also has a reasonably large effect on VaR and ES estimates. The VaR estimates are more than 0.5% lower, while the ES estimates are more than 1% lower, when using the sample percentile method compared to the Gaussian distribution. This is because the estimated residuals have fat tails and are not close to a Gaussian distribution. For consumer portfolio, a histogram of these residuals is:

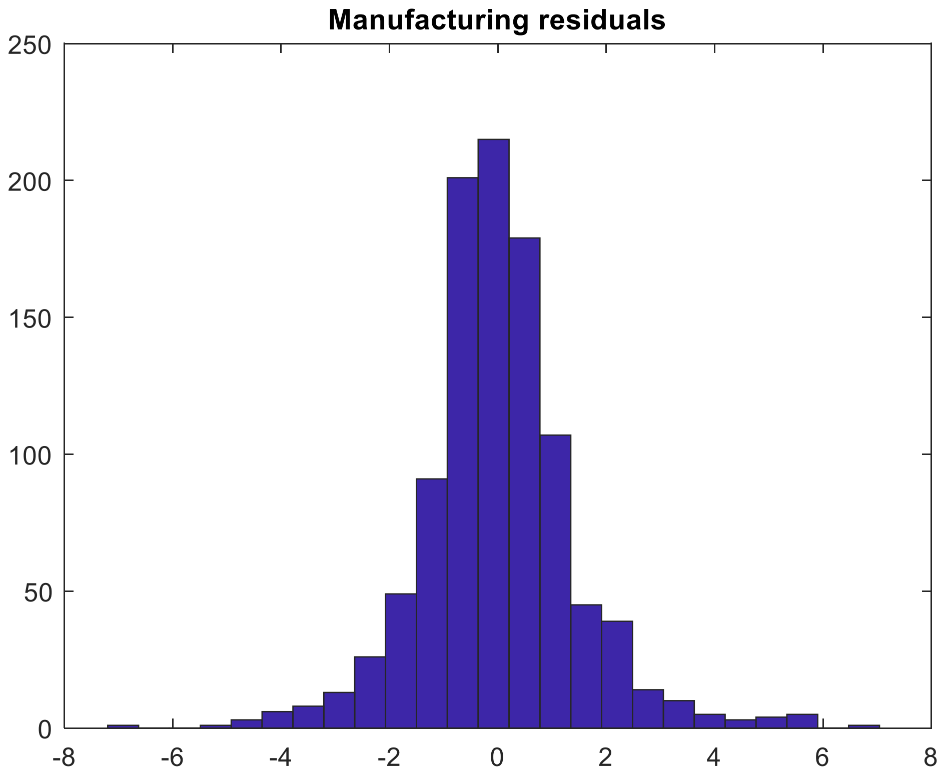


Fat tails are clear here. The Jarque-Bera test has a p-value < 0.001, indicating that Gaussianity can be strongly rejected. The sample skewness is 0.16 while the sample kurtosis is 5.2. Thus, the sample percentile method's VaR and ES estimates are likely more accurate than those for the Gaussian.

<u>Manufacturing</u>

Mkt	HML	SMB	Mean	Gaussian		Sample percentiles	
				VaR	ES	VaR	ES
-5	0	0	-4.557	-7.875	-8.364	-8.414	-9.256
-5	0	-2	-4.360	-7.678	-8.168	-8.217	-9.060
-5	-2	0	-4.881	-8.199	-8.689	-8.738	-9.581
-5	-2	-2	-4.685	-8.003	-8.492	-8.541	-9.384
-10	0	0	-9.442	-12.760	-13.249	-13.299	-14.141
-10	0	-2	-9.245	-12.563	-13.053	-13.102	-13.945
-10	-2	0	-9.766	-13.085	-13.574	-13.623	-14.466
-10	-2	-2	-9.570	-12.888	-13.377	-13.426	-14.269

Very similar comments to those for Consumer industry apply to the Manufacturing industry group. The Manufacturing group would average a loss slightly worse than the Consumer portfolio if the market dropped 5% and between 0.5-1% higher loss if the market dropped by 10%, compared to the Consumer portfolio. However, due to the lower SER (1.43 vs 1.92), the tail risk loss measures for Manufacturing are uniformly lower, by close to 1%, compared to the consumer portfolio, under the Gaussian distribution and sample percentile methods. Again tail risk measures are more extreme under the sample percentile method, as is expected from this residual histogram:



Fat tails are again clear here. The Jarque-Bera test has a p-value < 0.001, indicating that Gaussianity can be strongly rejected. The sample skewness is 0.36 while the sample kurtosis is 6.1. Thus, the sample percentile method's VaR and ES estimates are likely more accurate than those for the Gaussian.

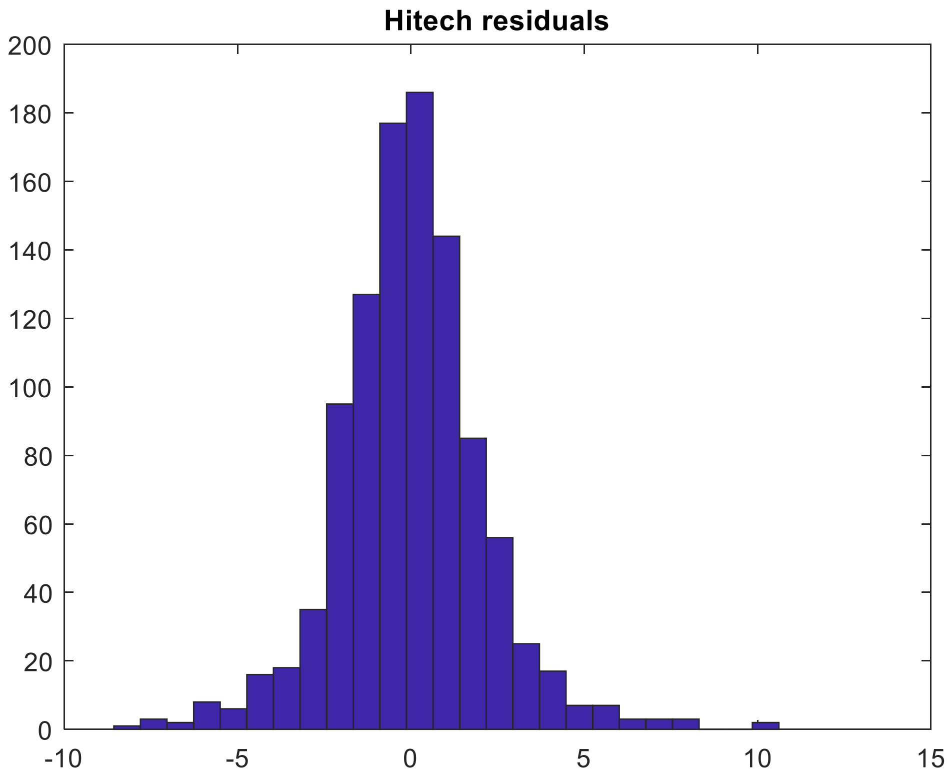
We could say that regarding average loss, the higher market beta for Manufacturing makes it more sensitive to market drops than Consumer portfolio. However, if extreme market drops coincide with extreme industry drops, then the Consumer portfolio is subject to higher tail risk and larger prospective tail losses compared to the Manufacturing portfolio.

Hi-Tech

Mkt	HML	SMB	Mean	Gaussian		Sample percentiles	
				VaR	ES	VaR	ES
-5	0	0	-4.484	-9.442	-10.173	-10.299	-11.304
-5	0	-2	-4.554	-9.512	-10.243	-10.369	-11.374
-5	-2	0	-3.839	-8.797	-9.528	-9.654	-10.659
-5	-2	-2	-3.908	-8.867	-9.598	-9.724	-10.729
-10	0	0	-9.392	-14.350	-15.081	-15.207	-16.212
-10	0	-2	-9.461	-14.420	-15.151	-15.276	-16.282
-10	-2	0	-8.746	-13.704	-14.435	-14.561	-15.567
-10	-2	-2	-8.816	-13.774	-14.505	-14.631	-15.636

Hi-Tech has a comparatively large effect for HML, (-0.323) in the regression equation. Thus, HML dropping by 2% is associated with an expected loss decrease of about 0.75% for the Hi-tech portfolio and tail loss also decreasing by a similar amount for VaR and ES. Average losses are about the same as those for Manufacturing when HML=0, but are less than both Manuf and Cnsmr when HML=-2%.

Mainly since HiTech has a comparatively large SER (2.1%), the tail risk losses are generally higher than those for both Manuf and Cnsmr, for both VaR and ES. Also, assuming a Gaussian distribution reduces estimated losses by about 1%.

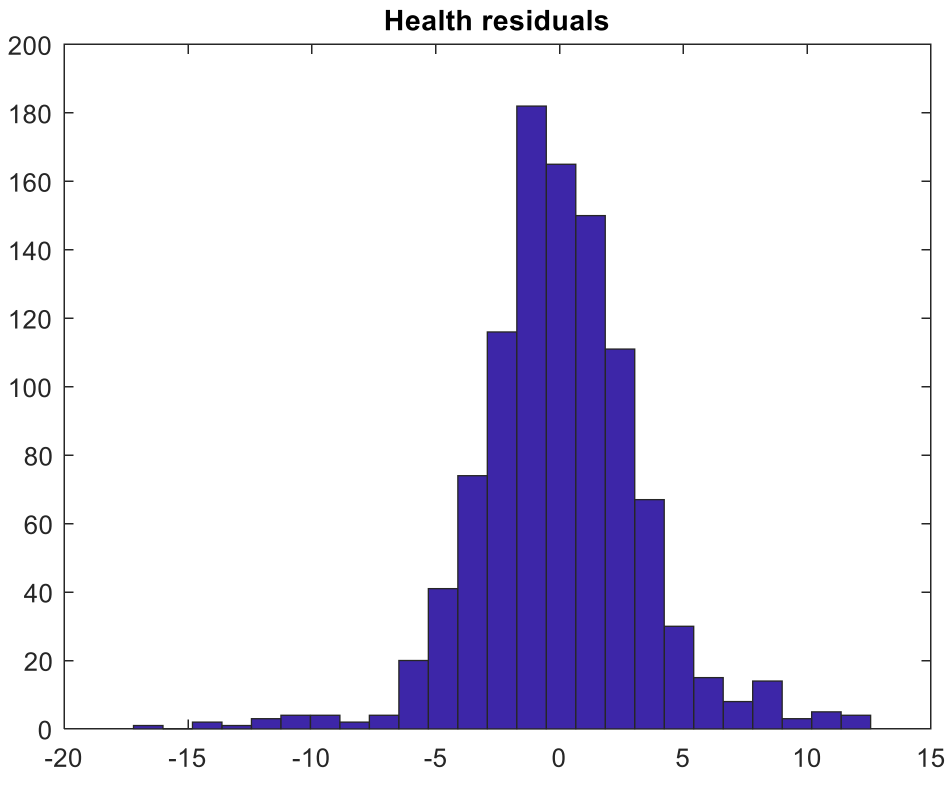


Gaussianity is again strongly rejected for the residuals, so the sample percentile method's VaR and ES estimates are likely more accurate than those for the Gaussian.

Health

Mkt	HML	SMB	Mean	Gaussian VaR	Gaussian ES	Sample percentiles VaR	Sample percentiles ES
-5	0	0	-3.849	-11.653	-12.804	-14.322	-16.370
-5	0	-2	-3.655	-11.460	-12.611	-14.129	-16.177
-5	-2	0	-3.477	-11.282	-12.433	-13.951	-15.999
-5	-2	-2	-3.284	-11.089	-12.240	-13.758	-15.806
-10	0	0	-8.279	-16.083	-17.234	-18.752	-20.800
-10	0	-2	-8.085	-15.890	-17.041	-18.559	-20.607
-10	-2	0	-7.907	-15.712	-16.863	-18.381	-20.429
-10	-2	-2	-7.714	-15.519	-16.670	-18.188	-20.236

The average estimated losses are the lowest for the Health portfolio in each case above, compared to Consumer, Manufacturing and Hi-Tech portfolios. Looking below, they are the lowest estimated losses among all five industry groupings. However, the opposite effect occurs with tail risk, mainly since SER is the highest for this industry (3.35). In the event that the market drops by 5 or 10%, if the Health industry also drops by an extreme amount, 1% or below, then the losses for that are by far the highest, up to 20% in fact, for the Health portfolio. Further, the difference between the Gaussian and sample percentile methods here is as high as 4% and typically about 1-2%.



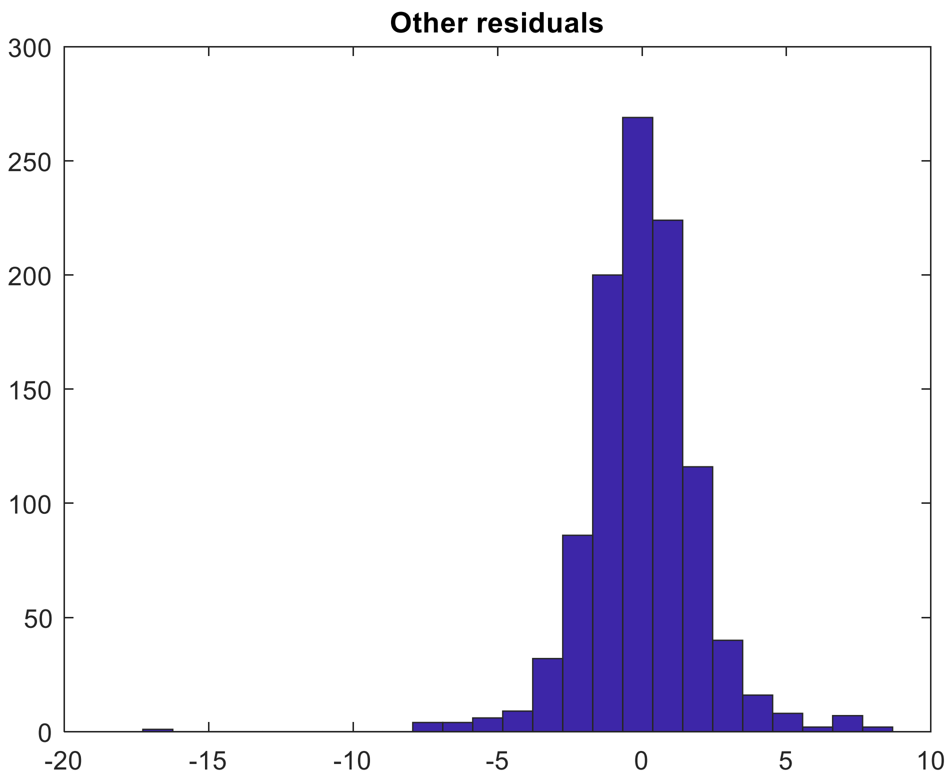
We can see that this is due to quite a few outliers on the negative side of the distribution, around the -12, -15% area. Gaussianity is again strongly rejected for the residuals, so the sample percentile method’s VaR and ES estimates are likely more accurate than those for the Gaussian.

Other

Mkt	HML	SMB	Mean	Gaussian VaR	Gaussian ES	Sample percentiles VaR	Sample percentiles ES
-5	0	0	-5.156	-9.719	-10.392	-10.617	-12.747
-5	0	-2	-5.285	-9.849	-10.521	-10.747	-12.877
-5	-2	0	-5.904	-10.468	-11.141	-11.366	-13.496
-5	-2	-2	-6.033	-10.597	-11.270	-11.495	-13.625
-10	0	0	-10.400	-14.963	-15.636	-15.861	-17.991
-10	0	-2	-10.529	-15.093	-15.766	-15.991	-18.121
-10	-2	0	-11.148	-15.712	-16.385	-16.610	-18.740
-10	-2	-2	-11.277	-15.841	-16.514	-16.739	-18.869

The Other industry grouping has the highest, or worst, average estimated losses for all cases above, compared to the other four portfolios. Losses are higher when HML = -2 (since the estimated HML beta is 0.37).

The tail risk measures are consistently the 2nd most extreme across the industry groupings for the Other industry.



The large negative outlier will have a big influence on the sample percentile method for ES, but not VaR. Think about why this is.

(d) *Using your answers in (b) and (c), discuss the risk properties of each industry portfolio and also how you could protect your investment in each against the situations in part (b).*

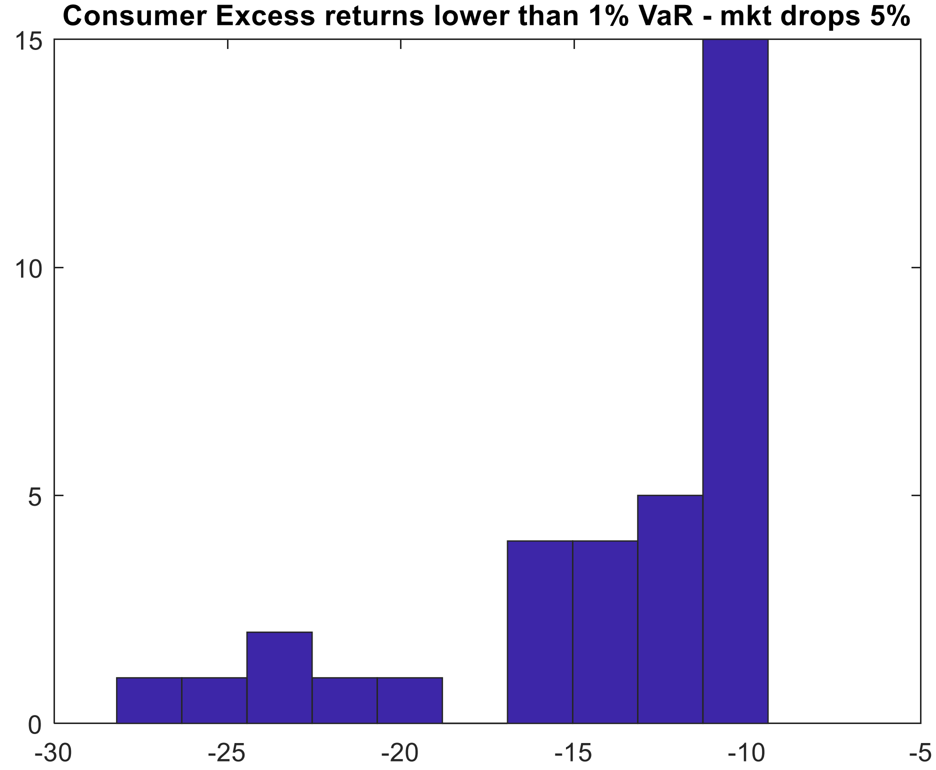
Other has highest market beta (1.05), which leads to the highest average losses when the market drops by 5% or 10%, which are even higher when HML drops by 2%. The Health portfolio has the lowest predicted average losses when the market drops by 5% or 10%; unsurprisingly it has the lowest market beta (0.88). We can see that market beta is a good ranking tool for average losses at the time of extreme market movements.

However, Health has the highest SER and many outlying residuals beyond -10%. Then, regarding tail measures, Health is at the highest risk of loss when the market drops and an extreme Health return results, both for VaR and ES. The Other portfolio tends to have the 2nd most extreme risk measures and predicted tail losses. The Manufacturing portfolio has the lowest and least extreme predicted tail losses, closely followed by the consumer portfolio. Market beta is clearly not at all a good ranking tool for predicted tail losses at the time of extreme market movements; this is well evidenced by Health having the lowest beta but highest predicted tail losses. Tail losses are more dependent on the spread of the portfolio return distribution, with SER being a good ranking tool here.
The Basel accords on banking supervision (Basel I, Basel II, Basel III) recommend capital requirements for banks as buffers against negative market movements. The capital requirement of 4.5% of common equity is recommended by Basel III (up from 2% in Basel II after huge losses seen over the GFC). While the capital requirements help, they certainly do not completely offset the losses seen in part (b).
An alternative strategy would be to protect your portfolio via the use of some form of derivatives (e.g. purchase puts options for all assets you are holding). Many financial instiutions wil do this for you. The cost of the derivatives, however, will reduce any potential profits from the investments.
More generally, diversifying your portfolio by holding cash, bonds, property etc will help avoid losses when the markets are in turmoil.

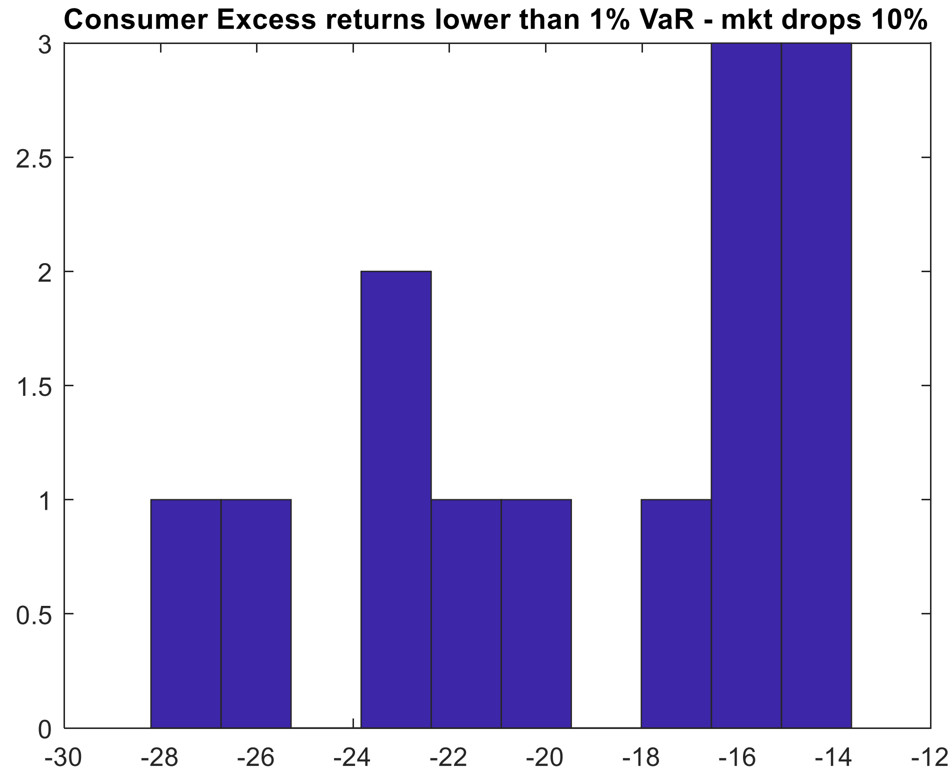
(e) *How likely are the scenarios in part (b)?*

You might well ask how likely are these losses to occur? Good question.
The market dropped by 5% or more in 128 months in this sample of 1026 months, i.e. in 12.5% of the months in this period. The market dropped by 10% or more in 34 months, or at a rate of 3% per month over this period. Thus, these are not very unlikely events at all: being about once a year (drop by 5%) and about once every 4 years (drop by 10%)! Clearly, analysts should factor in such drops into their risk calculations.

In the 128 months that the market dropped by 5% or more, the Consumer portfolio also dropped by more than its 1% VaR estimate (using Gaussian distribution) in 34 of those months, or 27% of those months. Thus the combined chance, from this sample, of the market dropping by 5% or more and consumer returns being lower than their 1% VaR is 0.125*0.266 = 0.033 or 3.3% (i.e. 34/1026)! Note that 34/128 * 128/1026 = 34/1026 = 3.3%. Thus, the event of the market dropping by 5% or more AND Consumer return being lower than its 1% VaR did happen in 3.3% (i.e. 34) of the months in this sample period. The average Consumer return for these 34 months was -13.9%. A histogram of these 34 Consumer returns is:



In the 34 months that the market dropped by 10% or more, the Consumer portfolio also dropped by more than its 1% VaR estimate in 13 of those months, or 38% of those months! Thus the combined chance, from this sample, of the market dropping by 10% or more and consumer returns being lower than their 1% VaR is (13/1026=) 0.03*0.382 = 0.0128 or 1.3%! The average Consumer return for these 13 months was -19%. A histogram shows:



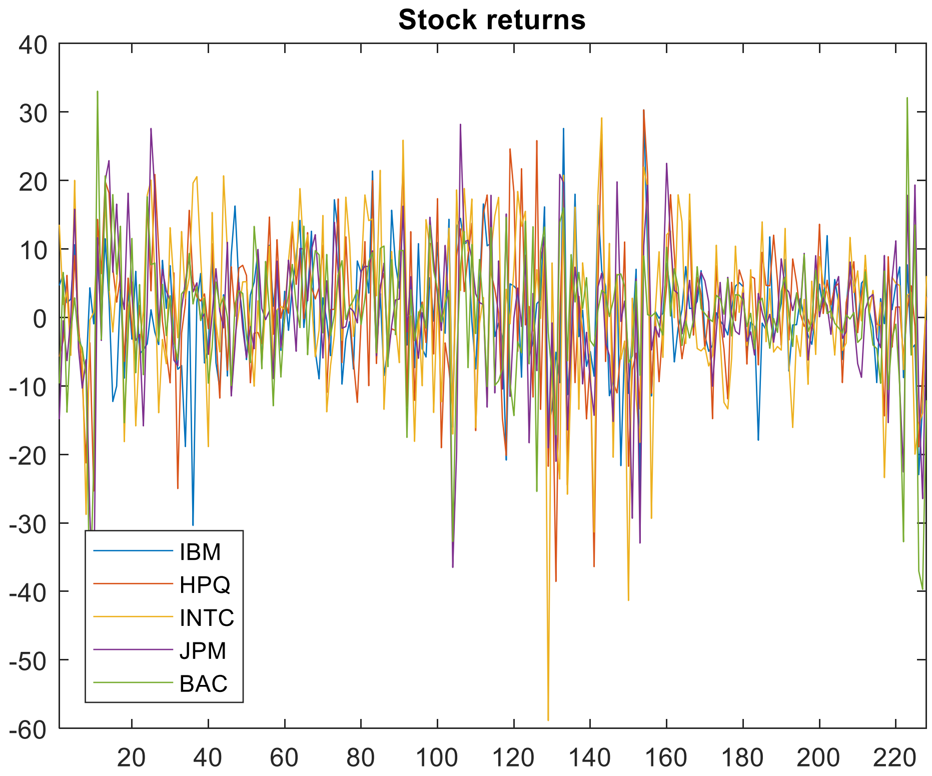
These events are not particularly unlikely at all it seems, and they are indeed associated with large, damaging portfolio losses, as predicted (!).

The table following shows these figures for each industry:

Industry	Mkt	No. < 1% VaR	Prob<VaR|Mkt	Prob<VaR & Mkt	Avg. ret
Consumer	< -5%	34	0.266	0.033	-13.95
	< -10%	13	0.384	0.013	-19.05
Manufact	< -5%	45	0.352	0.044	-13.26
	< -10%	16	0.471	0.016	-18.37
HiTech	< -5%	36	0.281	0.035	-14.74
	< -10%	11	0.324	0.011	-20.26
Health	< -5%	11	0.086	0.011	-20.91
	< -10%	6	0.176	0.006	-24.42
Other	< -5%	39	0.305	0.038	-16.21
	< -10%	16	0.471	0.016	-20.76

Similar comments apply to all industries, as they did to Consumer. Events as or more extreme as defined in part (b) (only relating to Market returns) did occur quite regularly (12.5% for market drops of 5% or more and 3% for market drops of 10% or more). Further, between 0.6% and 4.4% of the time, across industries, large market drops were accompanied by large drops in the industry return, below its 1% percentile VaR estimate. These events were associated with quite large monthly industry return losses, from -14% to -24% on average (and hence much larger individual month losses occurred too). Also, note that although the ES estimates for each portfolio are reasonably accurate, they all under-estimate what actually happened in the data!

Q2 (PCA and Factor modelling)
We use the data from the textbook by Tsay, in Chapter 9, being monthly returns on IBM, HPQ, Intel, JP Morgan and Bank of America, from January, 1990 to December, 2008.



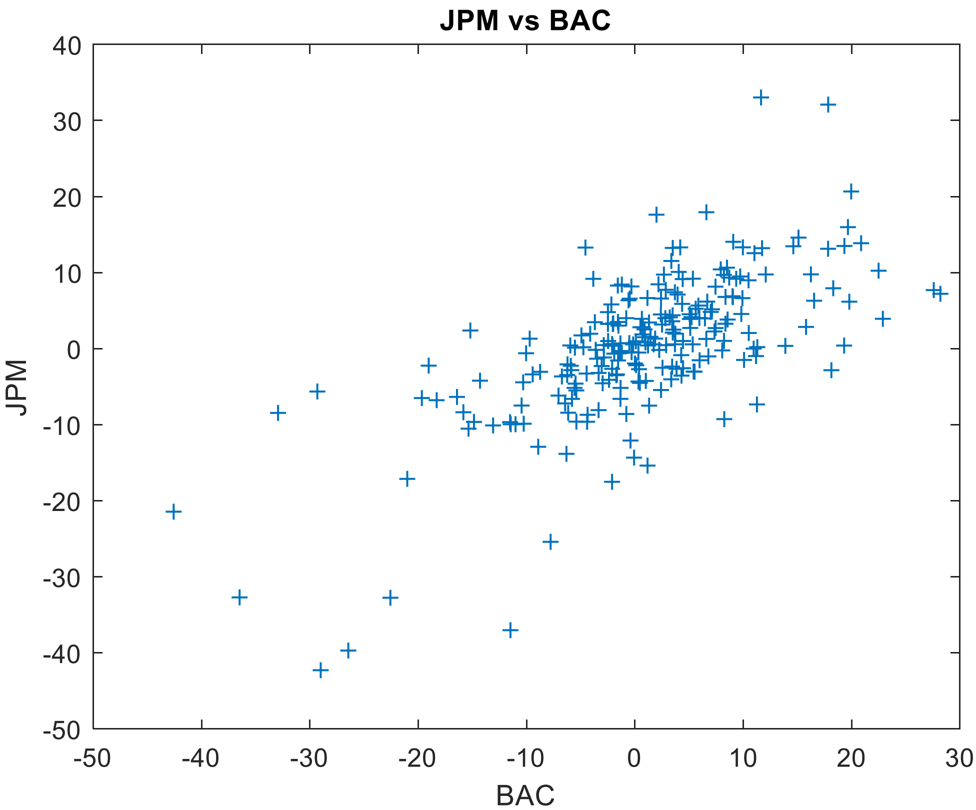
The low volatility period leading up to 2008 is readily apparent, as is the increase in volatility marking the GFC period in the last part of the data.

(a) *Form the correlation matrix from these asset return series and comment on the apparent relationships observed. Does it make sense to do a PCA on these variables?*

	IBM	HPQ	INTC	JPM	BAC
IBM	1.0000	0.4620	0.4593	0.3384	0.2545
HPQ	0.4620	1.0000	0.5495	0.3889	0.2591
INTC	0.4593	0.5495	1.0000	0.3578	0.2521
JPM	0.3384	0.3889	0.3578	1.0000	0.6836
BAC	0.2545	0.2591	0.2521	0.6836	1.0000

Pairwise correlations range from 0.25 up to 0.68. Clearly these variables are not independent and seem to show some relationships.

A plot of the highest correlated variables indicates that they appear to show a roughly linear relationship. These variables are clearly not independent of each other, so it makes sense to do a PCA.



(b) Perform a PCA on this data and report the results.

The table below shows the weights applied to each individual asset series to form each PC. The PCs are listed in order of importance regarding how much overall variability they capture in the data as a whole.

Principal components

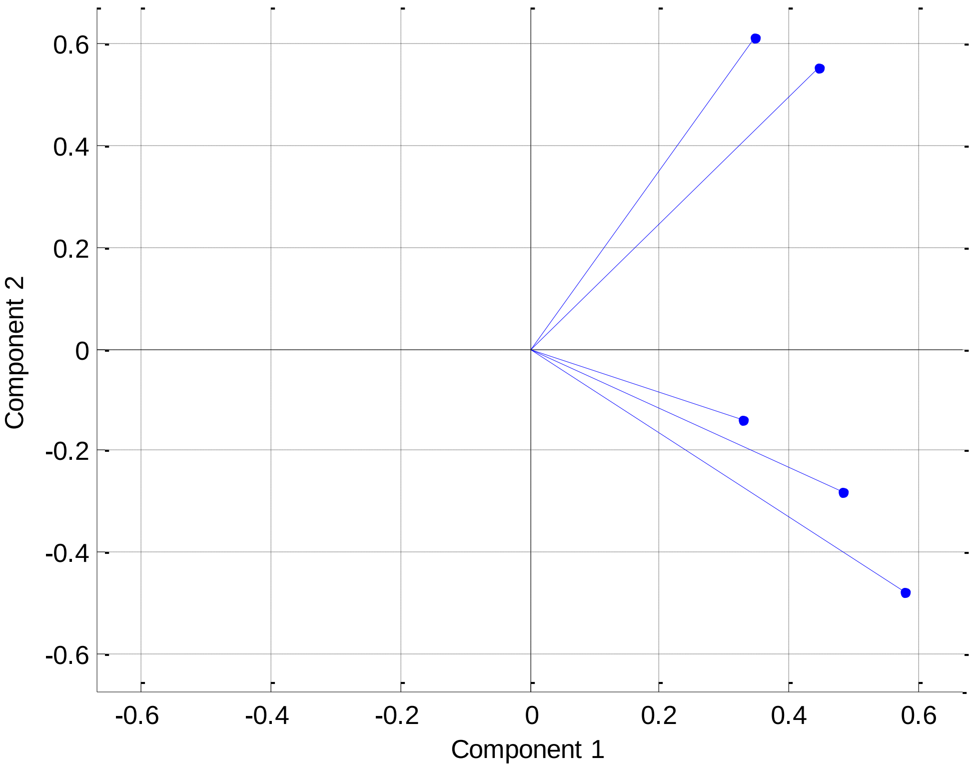
	1	2	3	4	5
IBM	0.3298	-0.1393	0.2643	0.8954	0.0144
HPQ	0.4826	-0.2786	0.7009	-0.4298	0.1159
INTC	0.5808	-0.4781	-0.6516	-0.0962	0.0163
JPM	0.4476	0.5502	-0.0128	-0.0642	-0.7019
BAC	0.3474	0.6097	-0.1188	-0.0093	0.7024

Lambdas

	1	2	3	4	5
Value	284.168	112.932	57.437	46.806	29.874
% variance	0.5349	0.2126	0.1081	0.0881	0.0562
Cumulative	0.5349	0.7475	0.8557	0.9438	1.0000

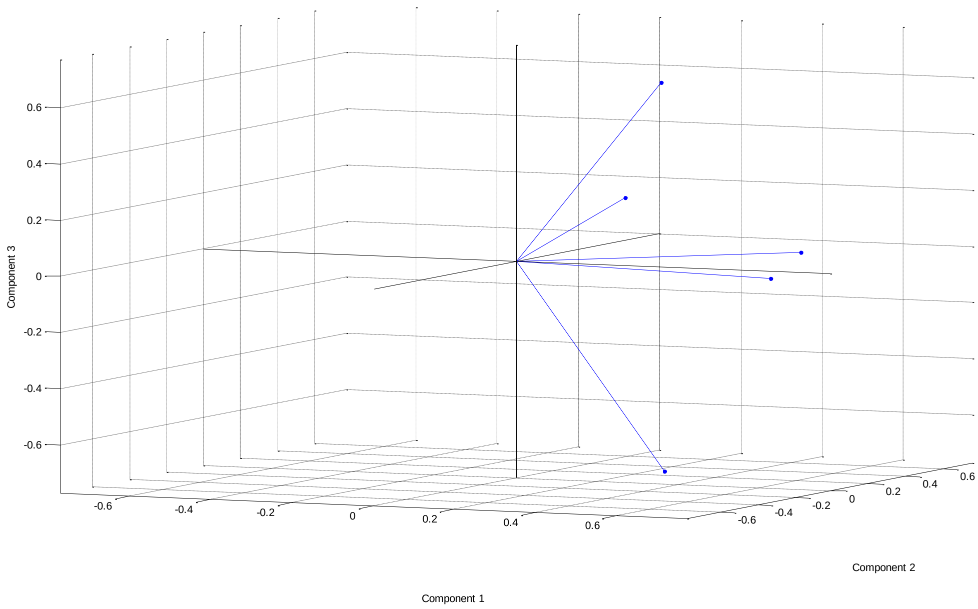
The 1^{st} component captures 53.5% of the total variance and has a variance of 284.17. The 2^{nd} component captures a further 21.3% of total variance; the 3^{rd} captures 10.8% etc.
The first three PCs capture 86% of total variance; the 1^{st} 4 components capture 94%.

A PC biplot of the 1^{st} 2 component weights shows:



Which reflects the positive similar weights on all assets in PC1 and the contrasting weights on two sectors, positive on financials and negative on technical stocks, in PC2.

A biplot of the first 3 PCs is:



Which shows only 2 large weights, contrasting HPQ and INTC. In matlab you can rotate this plot in any direction to make it clearer or to highlight any particular component (or component pairs). Make sure you try this by clicking on the little round arrow at the top of the matlab figure and using your mouse to drag the plot in different directions as you like.

(c) How many principal components do you think are adequate to explain these variables? Describe the PCs found, do they make intuitive sense and/or have a relevant or useful interpretation?

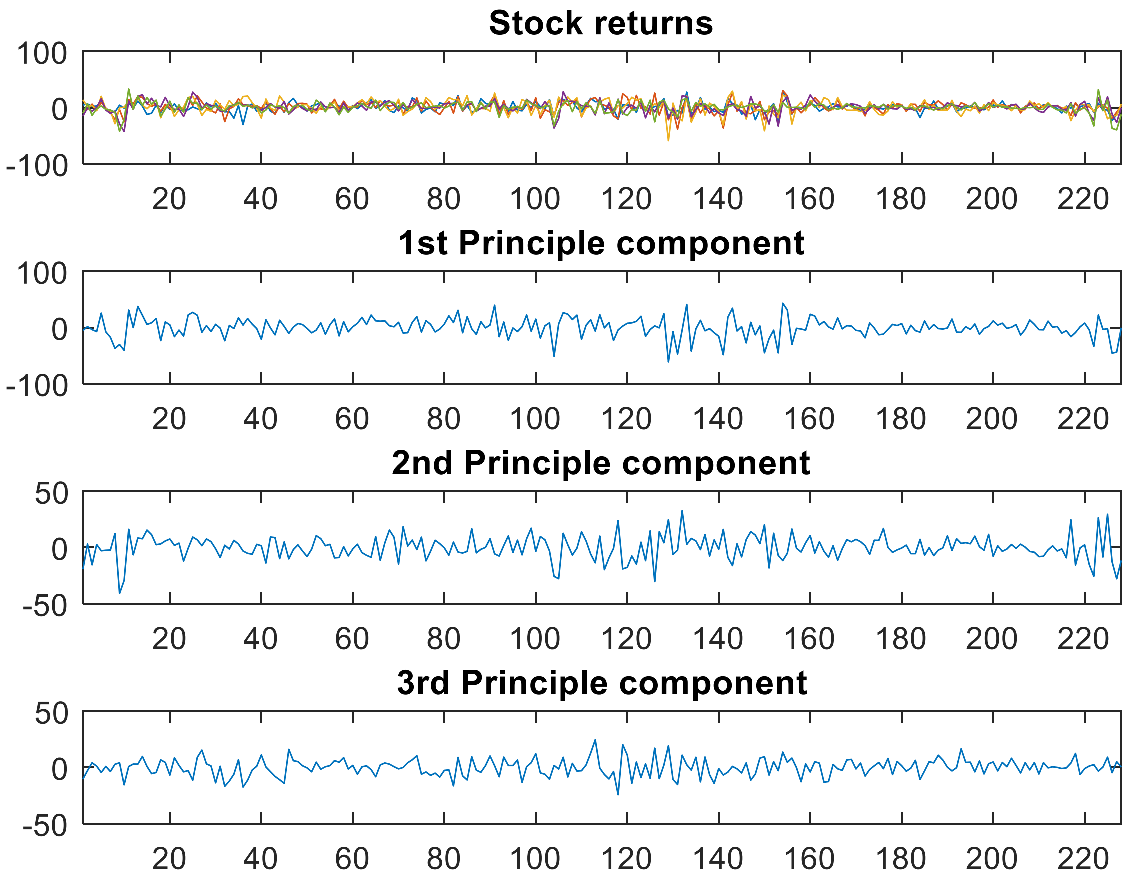
3 or 4 components seem needed; most analysts would likely choose one of these two options. I'd likely prefer 3 in this case.

The 1st component seems like a market factor, since it weights close to equally on all assets, though lowest on IBM. This may indicate that IBM is least affected by the market OR it may be because IBM has the lowest variance (which it does here; remember this component wants to maximise the variance, so down-weghting low varying return series is natural and intuitive) OR a bit of both. Note that the sign of the weights is irrelevant. The highest weighted component is INTC, which also has the highest sample variance among these assets (again this component maximises the variance, so highest weight on highest varying return series makes sense).

The 2nd component seems to contrast the returns from IBM, HPQ and INTC against JPM and BAC: i.e. this seems like a tech-stock vs financial stock type of component.

The 3rd component seems to contrast HPQ (and perhaps IBM) with INTC: a within tech-stock contrasting component. While the 4th component seems to directly contrast IBM and HPQ. Finally the 5th PC seems to contrast JPM and BAC: a financial stock contrasting component.

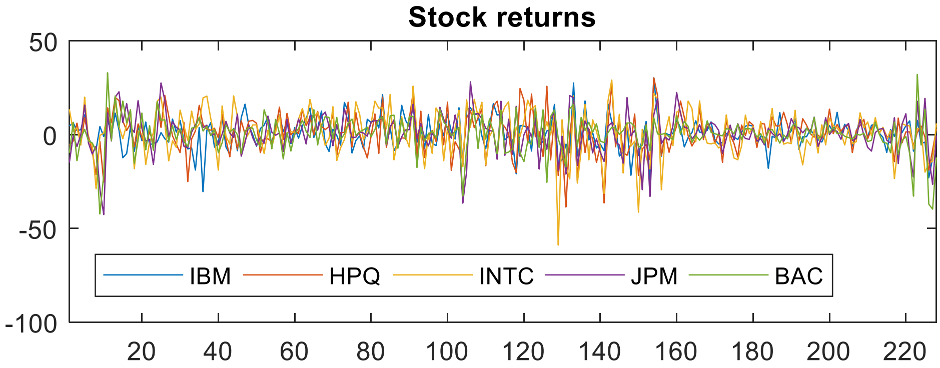
The plot below shows the 5 asset return series on the same plot. Below that are the 1st three PCs over time, PC1, 2 and 3. Remember that these PCs are uncorrelated with other

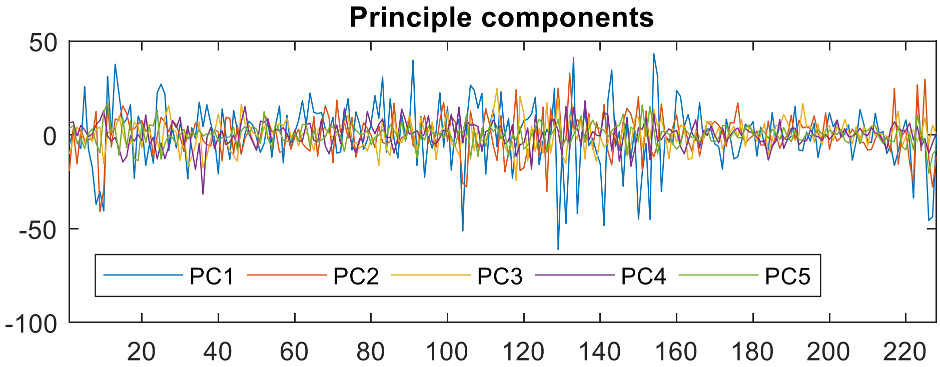


Clearly the first component does capture the major variance component across these series, being volatile whenever any of the 5 series are volatile.

The 2nd PC seems similar to the first, though it is less prevalent during the middle high volatility period (say obs 125-150), whilst perhaps more prevalent in obs 100-125 and in the brief high volatility periods at the end and start of the data. It seems similar to PC1 in low volatility periods.

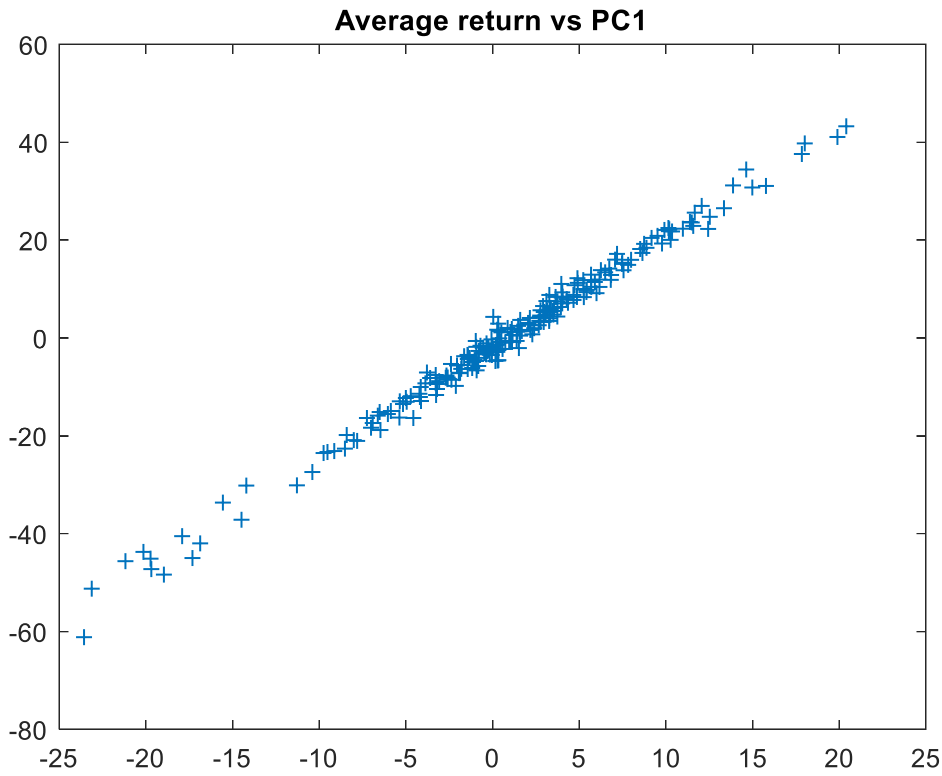
The 3rd PC seems to show little volatility at the start and end of the data, but does capture some variance in the middle of the data (say obs 110-135).

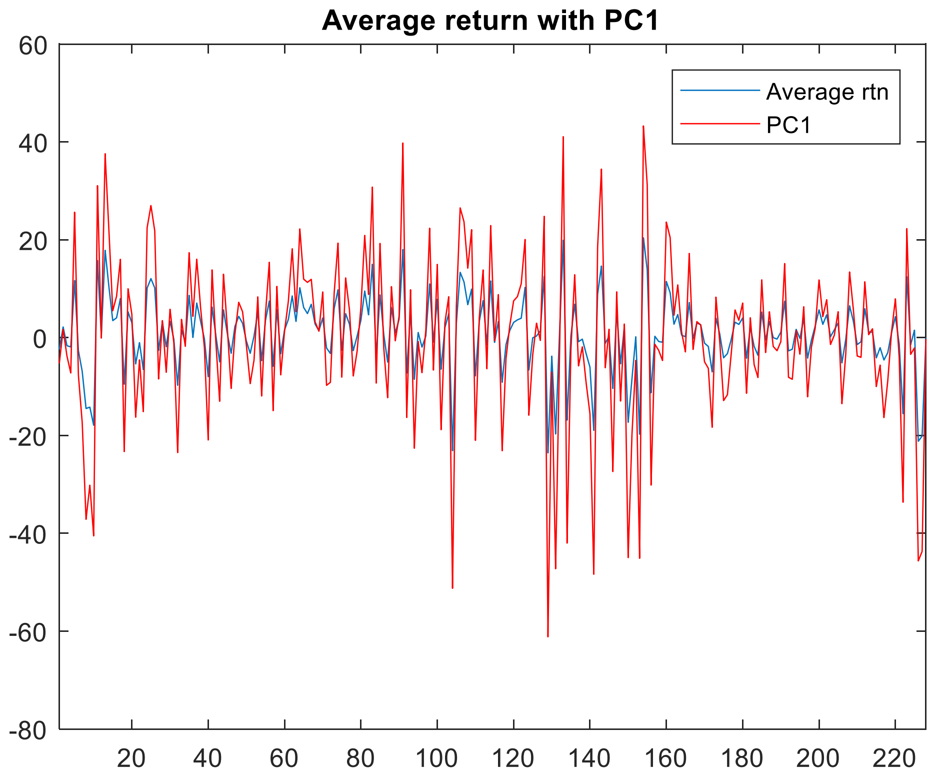




The plot above again shows the five asset returns series on top and then below shows the 5 PCs over time, overlaid. At the start of the data PC 1 captures the high returns in INTC and HPQ, then the subsequent low returns for INTC, JPM and BAC and the following high returns of INTC and BAC, then the high returns for all assets except IBM at about month 25. PC 1 captures high returns from INTC whilst PC 4 captures a simultaneous drop in IBM, at about month 35. From obs 100-150 PC1 captures most of the up and down movements. At the end of the data, PC2 seems to capture the upswings in returns whilst PC1 captures the downswings. The upswings were dominated by JPM and BAC (financials), whilst all the assets had downswings together (captured in PC 1).

Finally, the plot below shows the average over the 5 assets for each month versus the 1st PC. The correlation here is 0.995. The final plot shows the monthly mean of asset returns and the 1st PC (in red).





Clearly PC1 captures the average or mean monthly across the assets as well as most of the variance of these assets. We'll discuss these results further in the 5th lab session.