# QBUS6830 Financial Time Series and Forecasting
S1, 2019

Solutions to Lab Sheet 3

Q1 (Single factor CAPM)
Get the Kenneth French data on factors and 5 industry portfolios on a monthly frequency. The link is on Blackboard.

(a)Plot the industry portfolio, market index plus the HML and SMB factor returns over time. Find summary statistics for each of the eight return series. Comment and discuss.
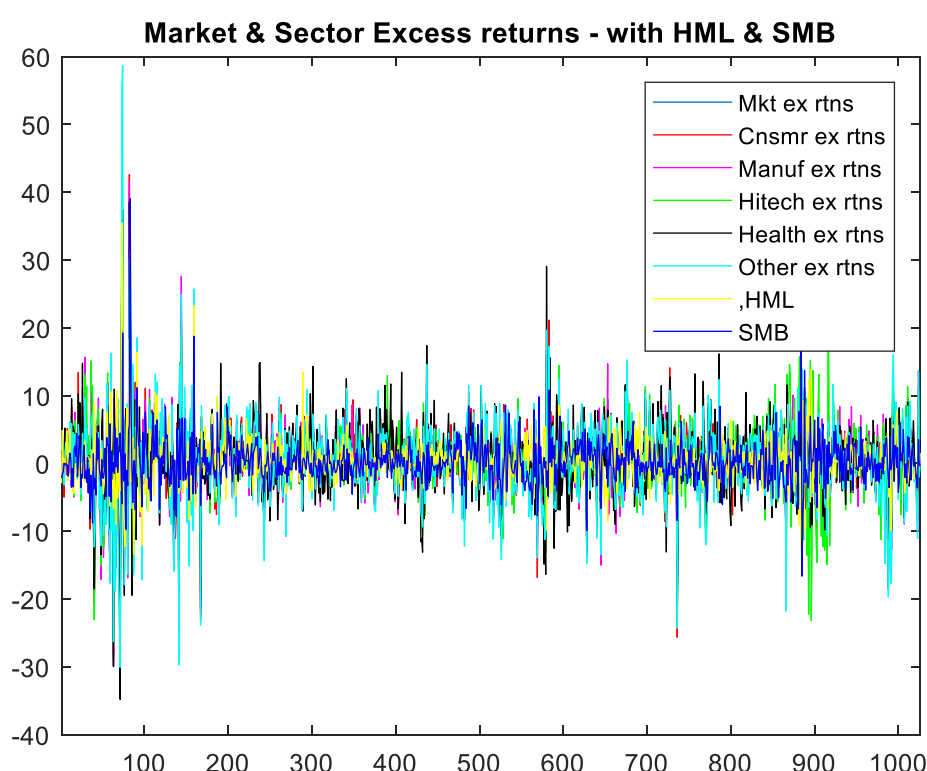

No real need to have eight different plots. Financial returns all group and tend and stay around 0, even those on the artificial hml and smb portfolios. Same frequency usually means similar scale, though SMB and HML seem to have slightly less variation than the other series, in most time periods. Same comments as last week apply.

From the summary statistics below, all industry portfolios and the market had positive, though small, average and median monthly returns over this period. Their standard deviations were almost all between 5 and 6%, except the Other group which was ~6.6%. Most monthly returns ranged between -30% and 30-40%, though again Other had a large outlying return of nearly 60% in one month.

	Mean	Median	Std	Min	Max	skew	kurtosis
Market	0.6175	0.9550	5.4572	-29.0400	38.2700	0.1685	10.3983
Consumer	0.6940	0.9100	5.3856	-28.2000	42.5900	0.1153	10.1566

|							
Manufacturing	0.6820	0.9100	5.5966	-29.8400	41.5300	0.3579	11.0530
HiTech	0.6257	0.8900	5.6908	-26.7700	33.8000	-0.1695	6.5058
Health	0.7735	0.7350	5.7350	-34.8000	38.5600	0.1798	10.1047
Other	0.5860	0.9350	6.5587	-30.0500	58.7100	0.8929	15.8557
HML	0.3822	0.2250	3.5674	-13.4500	35.4800	1.8328	18.5977
SMB	0.2415	0.0600	3.3151	-16.6200	39.0400	2.1781	25.1477

Further, the kurtosis of each series is above 3, indicating these all have fat-tails compared to a Gaussian distribution. The kurtosis is highest for the HML and SMB series. The skewness of most of the series is positive, though HiTech has a negative skewness. The skewness is again highest and most positive for the HML and SMB series.
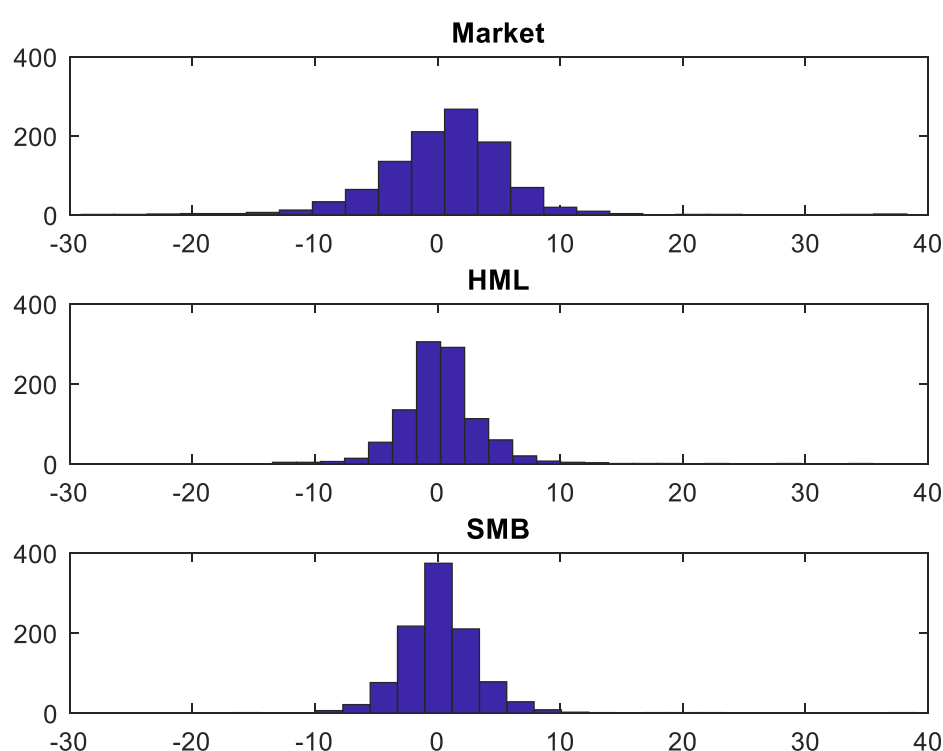

HML and SMB have high positive sample skewness, mainly due to their larger positive outliers. i.e. they appear symmetric in shape, except for a small number of large, positive outliers. These outliers are likely biasing the skewness statistics upwards here.

(b) *Form the excess returns on each industry portfolio and the excess returns on the market. Construct a scatterplot of each excess industry return against the HML and SMB series respectively. Calculate the pairwise correlations between each industry and the hml and smb series respectively. Test whether each is different to 0 (or not) at the 5% level.*
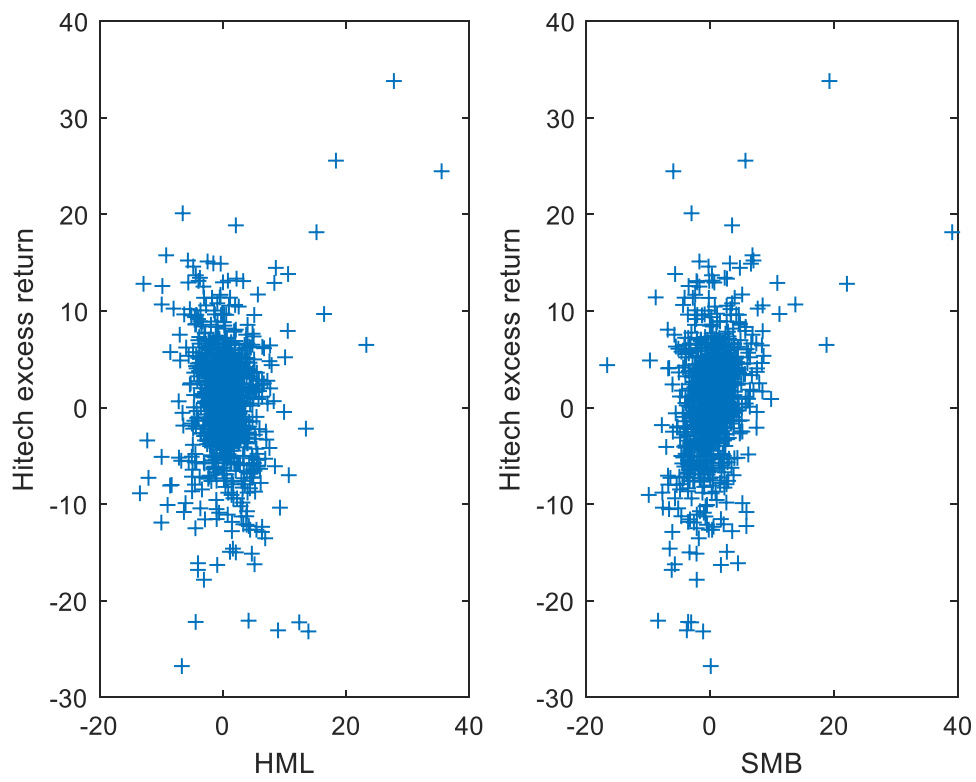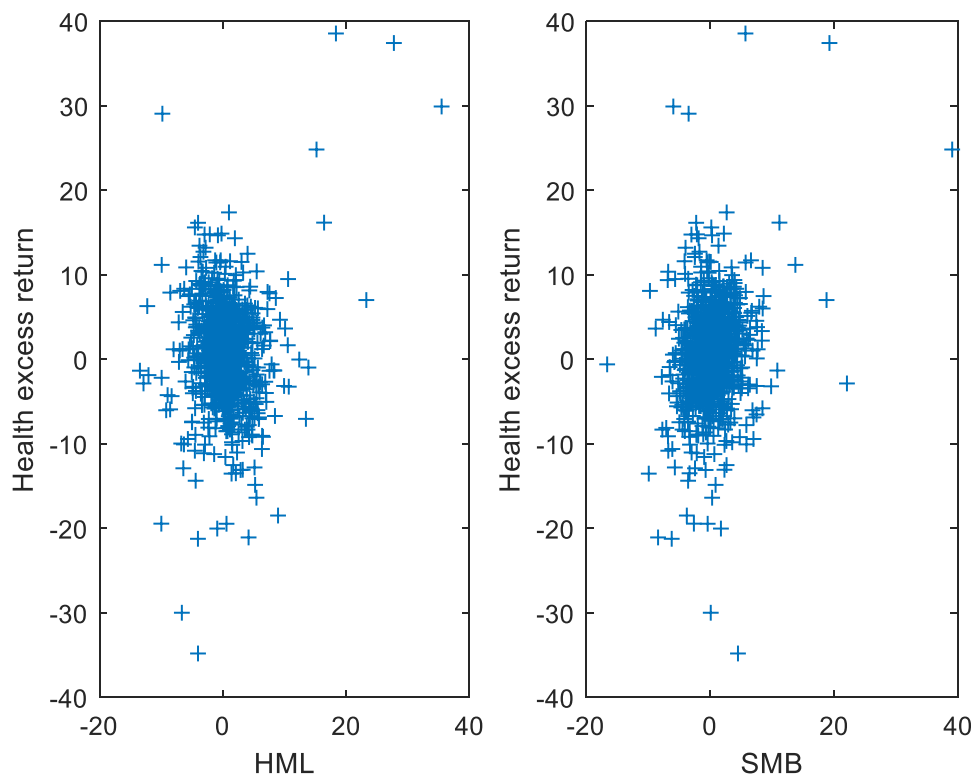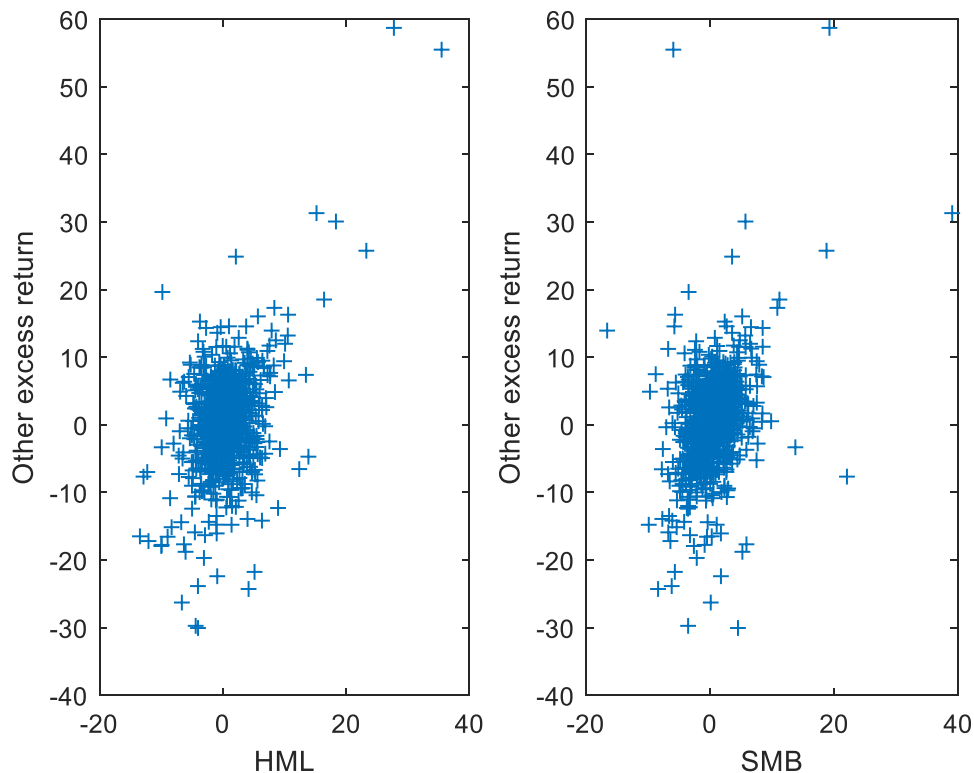
Consumer
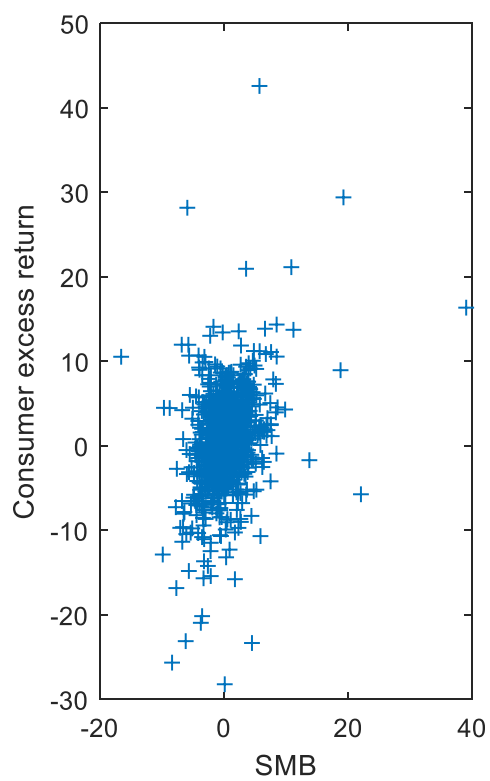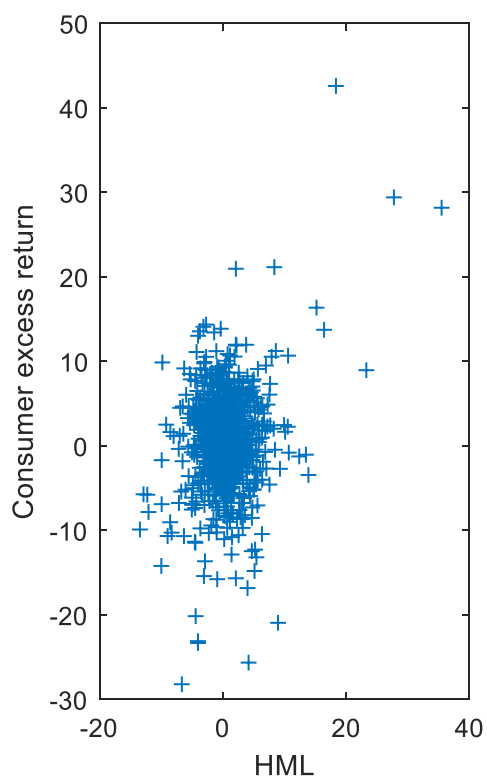
Manufacturing
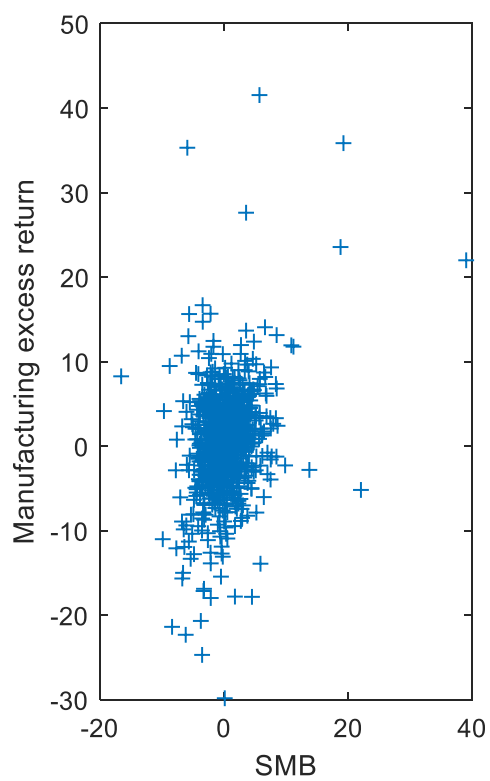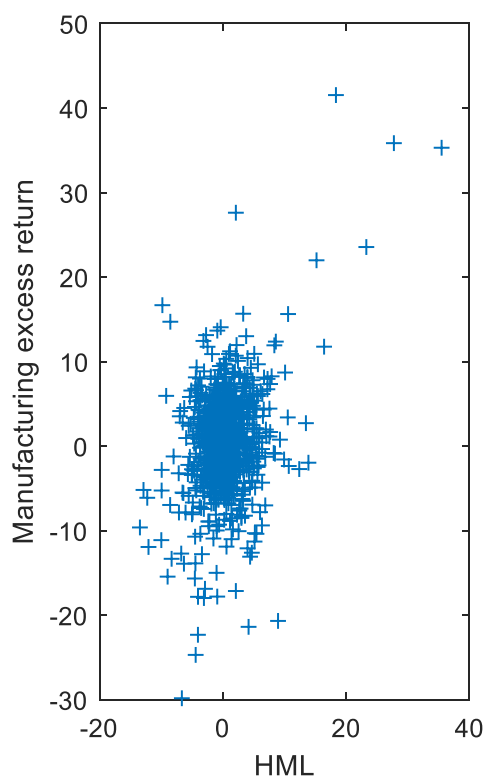
HiTech

Health

Other

These plots show the marginal or unconditional relationships between each industry return series and HML (left) and SMB (right). These relationships all seem to be pretty weak and it seems unclear whether they are positive or negative in general. Actually they all look like residual plots! However, one thing apparent is that generally the small number of very high monthly HML and/or SMB returns are associated with the small number of high industry group portfolio returns; see the right hand top corner of each plot. However, since this does not seem to hold for low HML or SMB returns, where associated industry returns seem close to average, the marginal relationships here may be NONLINEAR. Regarding outliers, all scatterplots show quite a few return pairs to worry about. More on this below.

As the relationships above may not be close to linear, correlations may make little sense to evaluate. I will do so for exposition only, noting that these results could be ignored or discarded.

Correlation	Cnsmr	Manuf	HiTech	Health	Other
HML	0.2088	0.3130	0.0125	0.0697	0.4041
p-value	0.0000	0.0000	0.0257	0.6889	0.0000
SMB	0.3160	0.2697	0.3142	0.2139	0.3438
p-value	0.0000	0.0000	0.0000	0.0000	0.0000

All correlations with the HML and SMB returns are weakly positive, though HiTech is lowest at 0.0125 with HML. Most p-values are 0, indicating strong rejection of the null and that most of these correlations are highly significantly different to, and greater than 0. Exceptions are HiTech and Health with HML; with a p-val of 0.69, HiTech correlation with

HML is not significantly greater than 0. However, Health's correlation with HML is significantly different to 0 at 5% (but not at 1%).

These results are clearly strongly influenced by the small groups of large returns in the top RH corner of each scatterplot above. They should be taken with a grain of salt and perhaps even simply ignored.

(c) Fit a multi-factor CAPM to each excess industry return series, using all three factors supplied, and report the estimates as well as 95% intervals for each estimate. Briefly comment and assess whether each industry could be classed as high, medium or low market risk. Provide a test or interval to support your conclusions.

Sector	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	0.124	0.006, 0.24	0.921	0.90, 0.94	-0.006	-0.04, 0.03	0.012	-0.025, 0.05
Manuf	0.038	-0.05, 0.13	0.982	0.96, 0.999	0.038	-0.05, 0.13	0.982	0.96, 0.999
Health	0.132	0.001, 0.26	0.987	0.96, 1.01	-0.326	-0.36, -0.29	0.037	-0.005, 0.08
HiTech	0.318	0.11, 0.52	0.892	0.85, 0.93	-0.189	-0.25, -0.13	-0.095	-0.16, -0.03
Other	-0.223	-0.34, -0.10	1.055	1.03, 1.08	0.371	0.34, 0.40	0.066	0.03, 0.10

All alphas are positive, except for Other. Consumer, HiTech and Health have significantly positive alphas. Other has a significantly negative alpha.

All market betas are positive and reasonably close to 1, minimum is Hitech which has estimated market beta of 0.89. Consumer, Manufacturing and Health all have betas significantly lower than 1 at the 5% level (their CIs are all below 1); indicating these are low risk portfolios compared to the market. Other has a market beta significantly greater than 1, indicating it is a high risk portfolio compared to market. Finally Health's beta is not significantly different to 1, indicating it is of moderate risk and has a similar risk level to the market portfolio.

(d) Discuss whether the assumptions made in these analyses might be valid or not. In particular, are outliers an issue or problem in any of these data sets?

The assumptions of OLS regression are:
1. The population residuals and the X variables are uncorrelated. In other words, $E(\varepsilon \mid X) = 0$
2. The data sample are iid
3. The 4th moments of both Y and each X are finite, i.e. $E(X^4 < \infty); E(Y^4 < \infty)$. This implies that the mean and variance of each of Y and each X are also finite.
4. The X variables, if there are more than 1, are not perfectly correlated with each other, and none is a perfect linear combination of any of the others: i.e. there is no perfect multi-collinearity.

A1: Errors and each explanatory (X) variable are uncorrelated.
For assumption 1 to be satisfied, any variable not contained in the model that influences the excess returns in the three industries should not be correlated with the excess market return, nor the IA and ROE variables. Industry-specific news or events that did not

affect market overall and did not affect the IA and ROE factors. The assumption is not satisfied if e.g. a variable (like general macro-economic news) affects the industry portfolio return and affects either the market return, IA or ROE. Because there are so many economic and financial variables that can influence stock returns for a range of industries concurrently this assumption is unlikely to hold perfectly in practice. Each industry should be assessed separately here. We can only hope that their influences are not too large.

If assumption 1 is not satisfied, then the estimates of the slope coefficients are biased as are t-statistics (same direction) and hence the p-value and conclusions of the tests could be incorrect.

A2 data are an iid sample.
The second assumption, that data sample being iid, requires that there is no autocorrelation in both the dependent variables, i.e. the industry excess returns, as well the explanatory variables, i.e. the excess market returns, IA and ROE. While many financial models assume such correlation does not exist in asset and portfolio returns, we have seen that often a weak level of correlation can be estimated. Without prior knowledge, it could probably be thought quite likely that financial factors such as Investment-to-Assets and the Return of Equity for companies and portfolios, could well be autocorrelated (and indeed they are!)

A3 Outliers are rare, i.e. fourth moments are finite.
An informal way to assess this is via residual plots. Some below:

Consumer
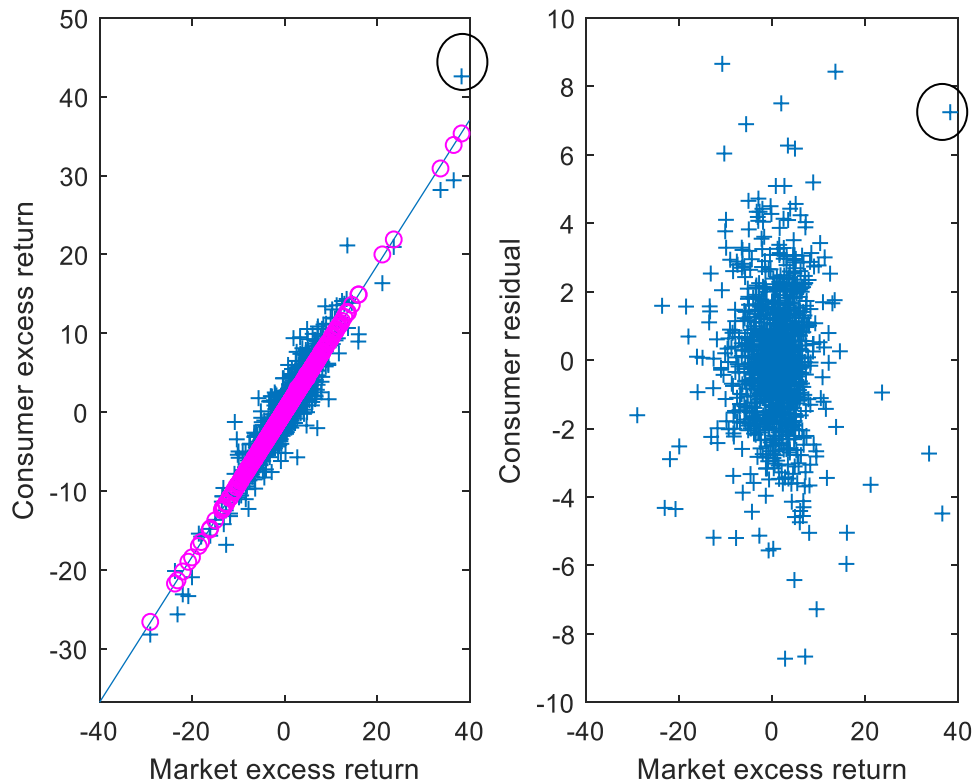
The only point that really clearly stands out in any way might be in the top RH corner of the residual plot. However, since this is not even the largest residual in magnitude, it is by definition not an outlier. I have no concerns here re outliers, since none of the points seems clearly or well separated from the rest of the data.

Note also that there is no clear pattern in the residual plot. The average residual seems to be ~0 for each value of the market excess return. This suggests the model is a good fit. IT DOES NOT PROVE THAT ASSUMPTION 1 IS SATISFIED, though it does at least suggest that it MAY be satisfied. Assumption 1 is about population errors, not estimated residuals: omitted variable bias would also bias the residuals away from the population errors.

Manufacturing

Comments as above apply. Good fitting model, no real concerns re outliers or extreme observations.
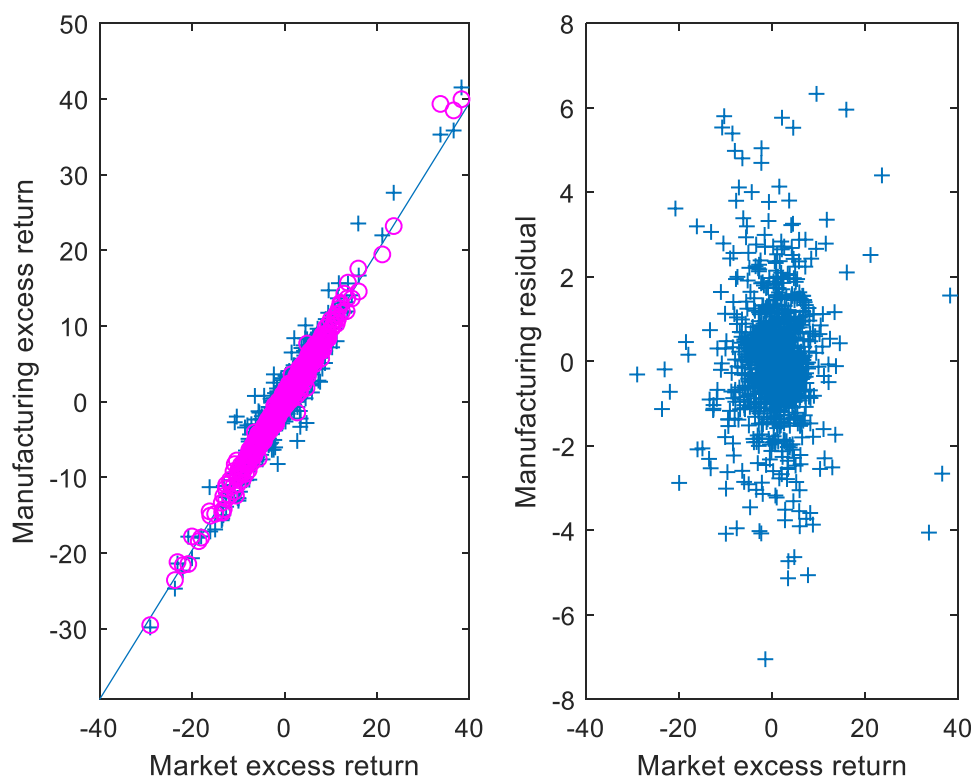
HiTech
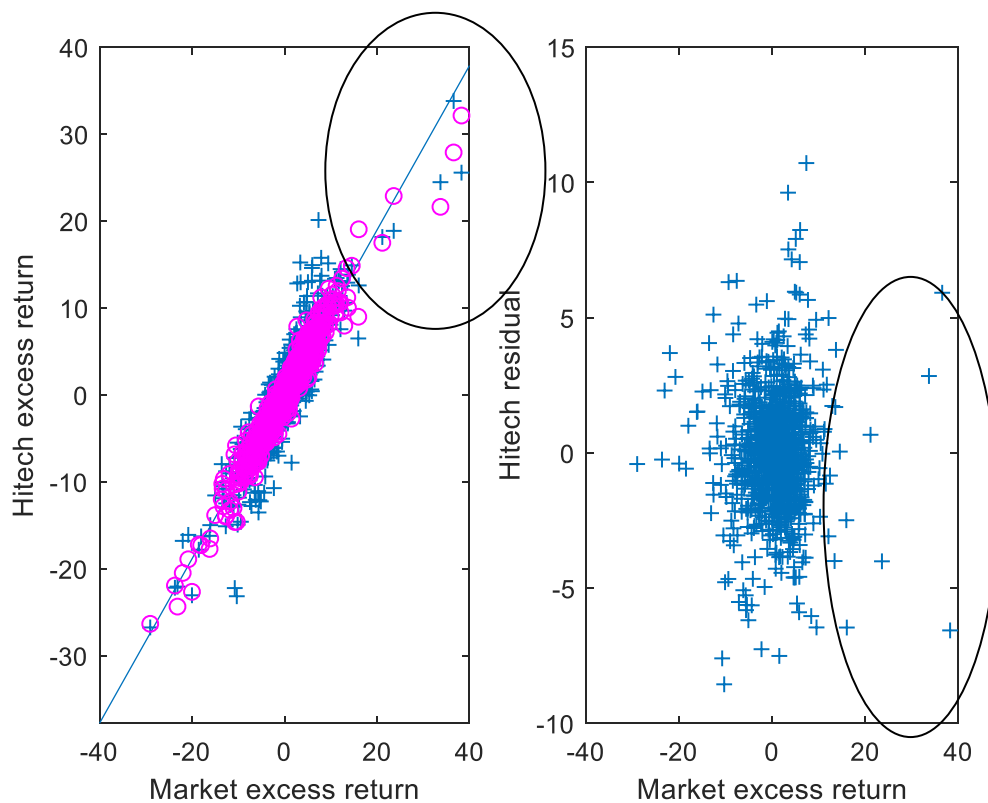
Here's where adding the factors HML and SMB really seems to have paid off in terms of model fit. The single factor CAPM model was rejected (by me) last week for this

data since the points circled above were all negative residuals and below the straight line, causing too many negative residuals for high market returns. However, the predictions, magenta circles on the left plot above, are mostly also below the CAPM line and are much closer to the actual returns for these circled points. On the right, we see that the residuals that were all negative for the CAPM are now averaging close to 0 and have both positive and negative values. This is the better model that I called for last week!

So, this seems to be a good fitting model, with no clear residual patterns and no outlier problems.

Health

Similar comments to those for Consumer, Manufacturing apply. The model seems to fit the data well. No real concerns about outliers or ill-fitting areas. Three positive residuals for the three highest market returns is a slight concern, but not too unlikely if the model fit well.
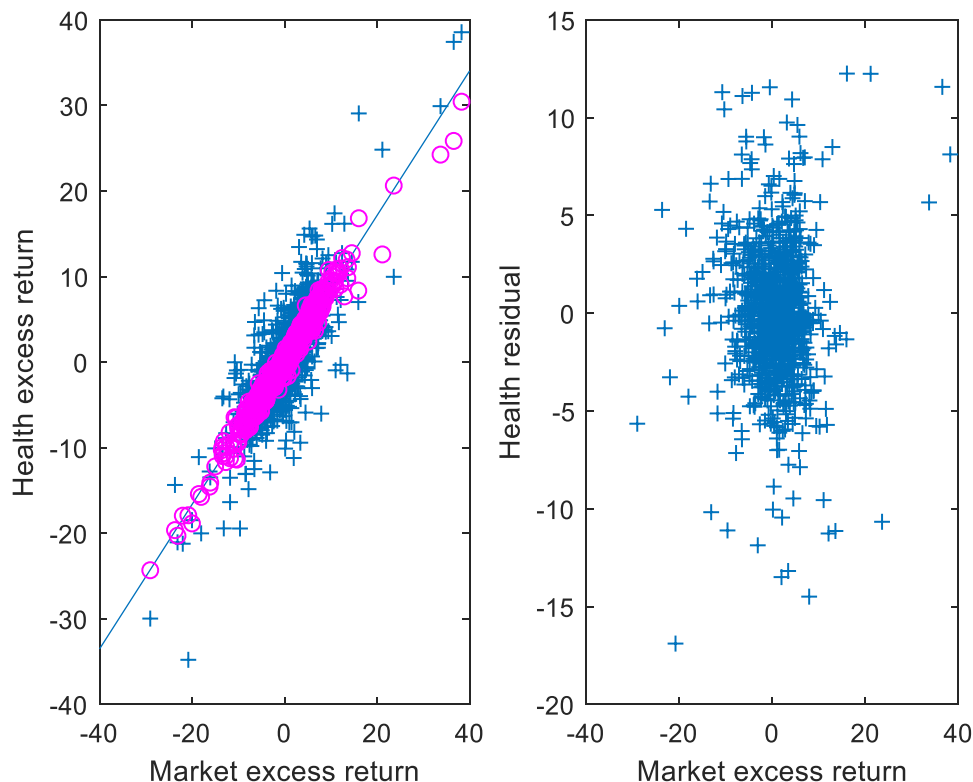
Other

This was another dataset I was also concerned about last week. However, once again the model seems to fit reasonably well for all types of market returns: the line goes nicely through the centre and the residuals average 0 at all points on the market return x-axis.
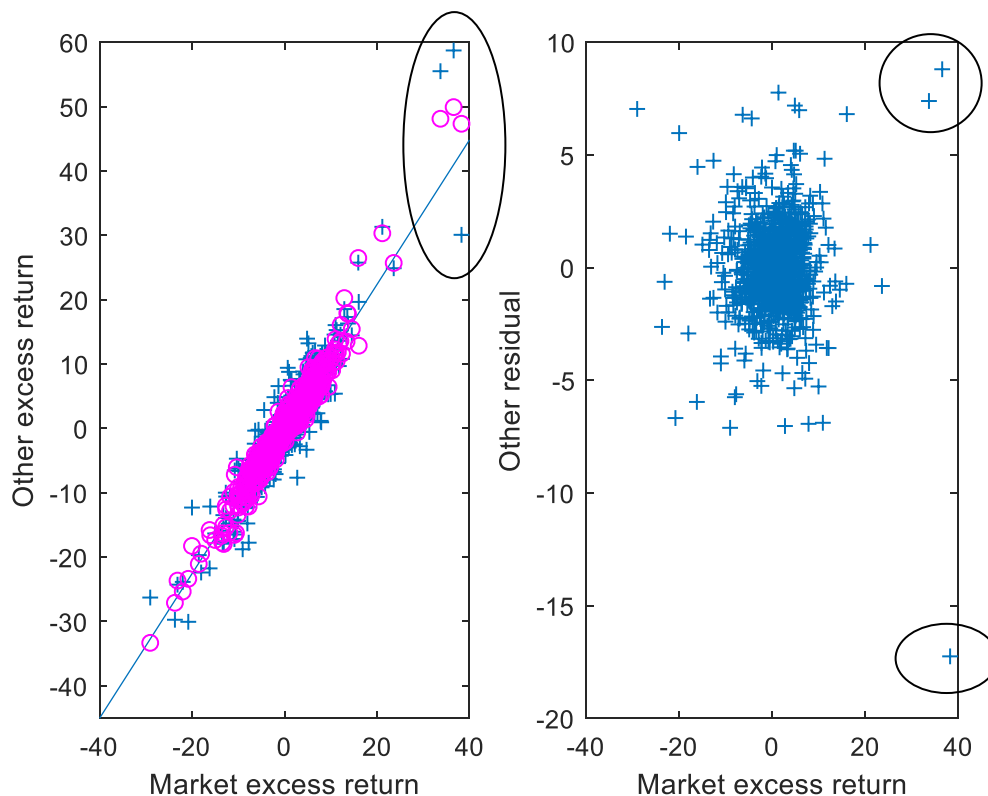
However, the concern that has again been upheld, and not completely corrected by the 3 factor model here, concerns the three outliers occurring at very high market excess returns (see plot on left). These three points are the three furthest from the CAPM line and have the three largest residual values in magnitude. Now, the two high positive ones that were of most concern for the CAPM, seem to be better explained by the 3 factor model: the residual plot on the right shows these are now not really extreme outliers at all, since several other residuals are about the same magnitude as these two. However, the negative outlier here is now the big concern. It is more than double the magnitude of the next highest in magnitude negative residuals. This suggests the possibility of a (error) distribution that may not have finite 4th moment or even finite variance.

If assumption 3 does not hold, then the central limit theorem will be violated so that the confidence intervals and p-values for the parameters are not accurate. Further, the estimates of the standard errors of the parameter estimates will also not be accurate and will not tend to their true values in large samples, making the t-statistics inaccurate too.

For this last dataset, I would not trust the result of the CAPM or the 3 factor model analysis regarding inference and confidence intervals for each parameter. A robust method may give much more accurate and useful results in this case.

(e) Report on how well and how strongly the Multi-factor CAPM fits each industry *portfolio data set. Compare results to the single factor CAPMs from lab 2.*

The first part is done in (d) already. The strength of each model can be assessed by R^2 and SER for each CAPM model.

	Multi factor			Single factor CAPM	
	R^2	Adjusted R^2	SER	R^2	SER
Consumer	0.8746	0.8742	1.9101	0.8745	1.9087
Manufacture	0.9363	0.9361	1.4145	0.9239	1.5443
HiTech	0.8625	0.8621	2.1134	0.8227	2.3976
Health	0.6619	0.6609	3.3396	0.6458	3.4148
Other	0.9133	0.9131	1.9336	0.8735	2.3340

The residual plots above show that the residuals can be very high and very low, e.g. up to -15% or 15%. This is still an enormous error to make in a prediction of investment returns, even on a monthly frequency! This is reflected in SER, showing typical errors in prediction range from 1.41% (Manuf) to 3.34% (Health). I would consider these somewhat large typical errors to make, even for monthly data (especially those above 2%). However, they are all slightly smaller than the SER from the single factor CAPM model, except for Consumer.

Again, when compared to the original standard deviations (around 5 or 6%) without accounting for market return effect, the error standard deviations (SER) have reduced by between 66% (Health) and 94% (Manuf). This illustrates that the three factor CAPM is a great improvement over just predicting industry group returns with their sample mean alone, and a smaller improvement over the CAPM model in most sectors.

When comparing two regression models, adjusted R^2 is more sensible to use. For all portfolios except Consumer, the adjusted R^2 for the three factor model is higher than the R^2 for the single factor CAPM. Thus, the three factor should be preferred for all sectors except the Consumer sector.

So it seems these three factor CAPM models are strong fits to the data after all and mostly slightly better than the single factor CAPM.

(f) Are there any changes in conclusions from using daily data, as in lecture, to monthly data? Any changes from conclusions in lab 2? Other issues?

Yes, a few.
The multi-factor model improved all the single factor CAPM models and allowed valid analysis of the Health industry portfolios, where the single factor CAPM did not. It did not solve the issue for the Other sector portfolio.