FM320 Summative 2 - Analysis

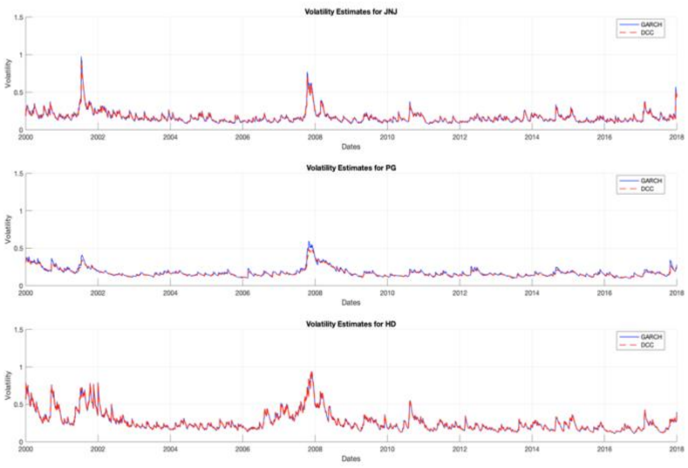
Question (b) - PCA analysis with securities and industries include JNJ (consumer products), WMT (retail), PG (consumer products), T (telecommunications), HD (retail) and VZ (telecommunications).

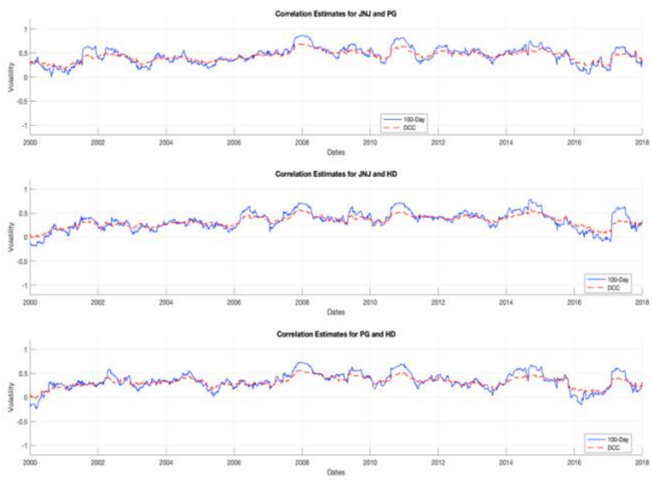
	1	2	3	4	5	6	7
1	'JNJ'	0.2484	0.0941	−0.4107	0.2184	0.8413	−0.0726
2	'WMT'	0.3902	−0.2623	−0.2743	−0.8386	−2.0183e−04	0.0218
3	'PG'	0.2768	0.0961	−0.7213	0.3362	−0.5274	0.0520
4	'T'	0.4535	0.5028	0.2530	−0.0466	−0.1132	−0.6801
5	'HD'	0.5659	−0.6644	0.3215	0.3657	−0.0320	−0.0139
6	'VZ'	0.4282	0.4679	0.2616	−0.0137	0.0152	0.7272
7	'Factor Variance'	7.2932e−04	2.6607e−04	1.6525e−04	1.2050e−04	9.0950e−05	7.4549e−05
8	'Fraction Total'	0.5041	0.1839	0.1142	0.0833	0.0629	0.0515

- The following analysis will show that for this PCA result a factor structure with one market factor (common factor) and three industry factors (specifically consumer products, retail and telecom industries) and with 6 idiosyncratic risk sources is appropriate, this is more complicated than a simple market model.
- An equation of excess returns is: $r(n) \;=\; \beta(n) * r(market) + \sum_{i=1}^{3} \beta(n,i) * r(i) + \epsilon(n)$, where $\beta(n)$ is the exposure of each security to the market factor; $\beta(n,i)$ is the exposure to i-th industry risk of stock n, and is 1 if stock n is in industry i, 0 otherwise; r(i) is the return to the i-th industry factor; and $\epsilon(n)$ is the idiosyncratic risk for each stock. And the reason and drawbacks for choosing this factor structure is given below:
- *Reasoning:*
1) For the 2nd column, the first source of risk which results in half of the total variance is identified as the market factor, to which all the stocks have positive exposure. This common factor can be interpreted as the main axis in the six-dimensional scatter plot of daily returns, and when the market volatility fluctuates, all the securities move in the same direction. The 2nd column suggests that when we invest in a portfolio composing of weights being the same as vectors in column 2, the portfolio would result in the largest volatility. Furthermore, the returns from this source of risk is recorded as r(market) in the equation.
2) As there would be more sources of risk in a factor structure than the risk factors that PCA would return to us, some sources of risk are combined in one factor. In this case, the 2nd risk factor is identified as some things that move Walmart and Home Depot (retail companies) in one direction, and move AT&T and Verizon Communications (US telecom giants) in another direction, while do not seem to affect Johnson & Johnson and Procter & Gamble much (as 0.09 ≈ 0), and also it is not a common factor. Thus, I conclude that 2nd factor is composed of two industry risk dummies, namely retail and telecom industry risks, which accounts for more than 18% of the total variance. It is an interesting fact that Home Depot is more exposed to the retail risk than Walmart.
3) The third risk factor indicates the industry risk for Johnson & Johnson and P&G, which is the consumer product industry. In this case P&G is much more sensitive than J&J to this industry factor, as -0.7213 has much larger magnitude than -0.4107. The market factor and the 3 industry factors mentioned above have incorporated slightly over 80% of the total variance, and it can be proved by statistical testing that 80% is a statistically significant number for these 4 sources of risk to be the main factors in our factor structure. The six idiosyncratic risks (account for 20% of variance) for each of the six securities are not that important.
4) By the functioning of PCA, the six idiosyncratic risks are incorporated in factors 4, 5 and 6 in our PCA analysis. In factor 4 (column #5), it is capturing residual risks for Walmart and Home Depot; factor 5 is capturing the idiosyncratic risks of Johnson& Johnson and P&G, and since for Johnson & Johnson the sign of the weight is different from the others (Verizon has negligible weight and all the others have negative sign), it might be because of the case that J&J have negative stock beta; and factor 6 is incorporating idiosyncratic risks of AT&T and Verizon, as the weights of the other stocks in the sixth portfolio are negligible. In conclusion, the factors 4,5, and 6 are capturing idiosyncratic risks of stocks within each industry.
5) Note that overall PCA analysis returns weights in each portfolio having Euclidean sum of 1, and since each portfolio/ factor is orthogonal to each other, each source of risk will only occur once in each factor.
- *Tradeoff between a clear factor structure and modelling difficulty*:
In the factor structure I have chosen above, I have separated the second principal factor component into two sub-industry dummies, and this 4-factor model has brought the benefit of effectively clearer interpretation and communications. However, this has made the factor model estimation (which we will see in question c, to

use PCA results and O-GARCH to generate time-varying volatilities and correlation estimates) a more difficult process, because the factors (in particular 2 of the industry factors) are no longer orthogonal portfolios and so the covariance matrices for each of the sources of risk are hard to compute (covariances is no longer 0).

Question (c) - DCC and O-GARCH models for JNJ (consumer products), PG (consumer product) and HD (retail)
Comparing DCC volatility and correlation estimates with GARCH model:



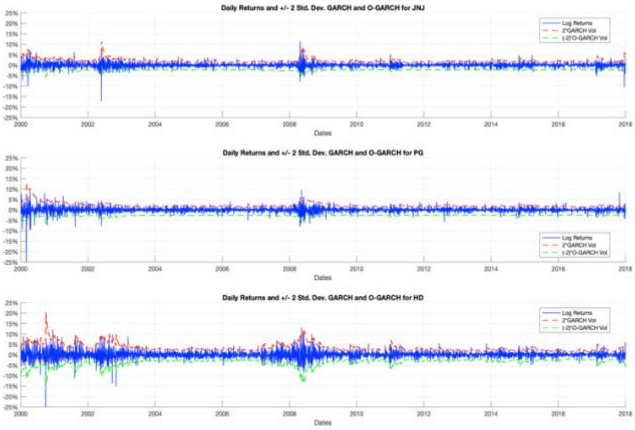


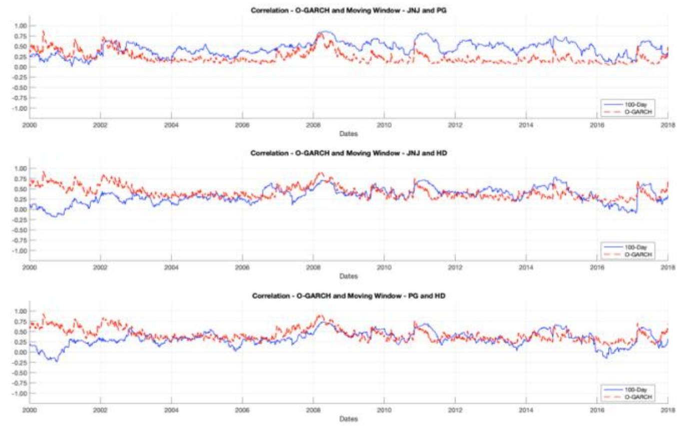
- The formula for DCC is $\sum_t = D(t)R(t)D(t)$, in which the variance-covariance matrix of the three securities is separated into two parts D(t) and R(t). For the volatility part (Dt) we use a GARCH (1,1) model to estimate the univariate volatilities of each securities. Hence this results in the time-varying volatilities for each of the three securities well-fitting into the GARCH-generated volatilities, and indeed, from the left figure above, the red line (DCC volatility) almost coincides with the blue line (GARCH volatility). However, I notice that for stock PG the DCC estimated volatility is not capturing some of the extreme events as GARCH does, this is shown in the fact that there is some blank space between the two lines in some of the peaks, such as the volatility shock in 2008 due to the financial crisis. An explanation for this could be the GARCH estimated alpha for P&G (0.0486) is larger than its DCC estimated alpha (0.0264), probably results from the calculating mechanism within Matlab, so that the DCC volatility for PG may not have such sensitive adaptation to yesterday's shock as pure GARCH does.
- The R(t) is the correlation matrix which has 1 on its diagonals after normalization and correlations between pairs of stocks on its off diagonals. For the dynamic conditional correlation model, the R(t) is time-varying and I choose a DCC(1,1) model to estimate R(t). It can be seen from the right graph that the DCC correlations for three pairs of stocks are capturing the median trends as GARCH, but is not capturing extremely high correlations during extreme periods, so the DCC correlation estimate is more stable and less sensitive. We can see that in the long-term, correlation between JNJ and PG is larger than the correlation between JNJ and HD, which remains in a similar level with that between PG and HD, and indeed in our DCCRbar, which is the estimate of long-term correlation matrix, the correlation between JNJ and PG is 0.4246 which is larger than 0.3154 (JNJ and HD) and 0.3105 (PG and HD). This could be reasoning from the fact that Johnson & Johnson and P&G are in the same industry (consumer products), thus they have higher correlation with each other than pairs in separate industries.

	1	2	3
1	1	0.4246	0.3154
2	0.4246	1	0.3105
3	0.3154	0.3105	1

- Furthermore, we expect correlations between stocks to rise during extreme periods, and this is indeed true, for example, in 2002, correlation between Johnson & Johnson and P&G rise to 0.6; and in 2008 the correlations between all pairs of stocks have risen due to the financial crisis, with JNJ and PG rises the most to 0.85; and there are also short-term peaks for three pairs in 2017 due to the US-China trade war. It is also remarkable that for most of the times in the two decades, the 3 pairs of stocks have positive correlations, which could be explained by having positive exposure to some risk factors and that both consumer products and retail industries are selling products to households and individuals rather than manufacturers and industries, so their sales revenues and manufacturing costs are tightly correlated and are exposed to same fluctuating factors such as oil price and interest rates.
- In general, the DCC method is applicable to a small number of stocks (<10 securities), otherwise it will have large estimation error, but compared to CCC method, DCC method gives accurate estimations for a small number of securities.

Comparing O-GARCH volatility and correlation estimates with GARCH model:





- I first compute a PCA analysis for JNJ, PG and HD stocks only, and the results are shown in the below table.

	1	2	3	4
1	'JNJ'	0.2999	0.4975	0.8140
2	'PG'	0.3455	0.7387	-0.5788
3	'HD'	0.8892	-0.4548	-0.0496
4	'Factor V...	4.3762e-04	1.8407e-04	9.1826e-05
5	'Fraction ...	0.6133	0.2580	0.1287

The first factor shown in the 2nd column is the market factor, which is a common factor that all of the three stocks have positive exposure to, and the 2nd and 3rd factor are incorporating the idiosyncratic risks of the 3 mstocks, hence this is a simple market model with one market factor and 3 idiosyncratic risks, and the formula is given by $r(n) = \beta(n)r(market) + \epsilon(n)$, with assumption made to be an equilibrium market and each idiosyncratic term are themselves uncorrelated and uncorrelated to the market factor. Thus, the variance covariance matrix of stock returns is $Var(r) = \beta Var(rm)\beta' + Var(\epsilon)$
- There are 6 parameters and the variance of return to be computed by GARCH. The variance of return to market factor is time-varying, while the stocks' exposures to market risk and the variance of each idiosyncratic risk factors are held constant in Matlab. In this Orthogonal-GARCH method to estimate factor structure, univariate volatilities and time-varying correlations between pairs of stocks are returned to us simultaneously.
- The left chart above compares volatilities estimated by O-GARCH and GARCH methods, and the red and green lines are the 95% confidence intervals of daily returns. While for Home Depot, the O-GARCH estimated volatilities perfectly fit into the actual daily returns and GARCH volatilities, the O-GARCH estimations of volatilities for JNJ and P&G are constantly overestimating the stock volatilities, as there is always white space between log returns and the green lines. Thus, the O-GARCH technique does not estimate risk as well as GARCH does, and one way to improve this is to compute variance of idiosyncratic risks to be time-varying (just as variance of return to the market factor does), but Matlab limits us in doing so.
- For the correlation chart comparing O-GARCH and moving window correlations of the three pairs of stocks on the top right, the O-GARCH estimation seem to be capturing the medium trend in pairs JNJ&HD and PG&HD, but the O-GARCH technique is consistently underestimating the actual correlations by 0.25. Moreover, for pairs of JNJ&HD and PG&HD, the O-GARCH correlation is substantially underestimating the actual correlations during the 2000s. Note that we expect to see close to 0 correlation when the market is calm because the idiosyncratic risk is very high for each stock, and close to 1 correlation when the market is volatile as market risk is very high at these periods.
- This may suggest that in this case the O-GARCH may not be a very good model for volatility and correlation estimation, because it does not capture the time-varying idiosyncratic risks of each securities as GARCH does, but the common factor between stocks only. To consider for improvements, we could use GARCH to estimate $Var(\epsilon)$ instead of a stagnant stock-specific risk.