REHM Rémi

Several insights about financial econometrics

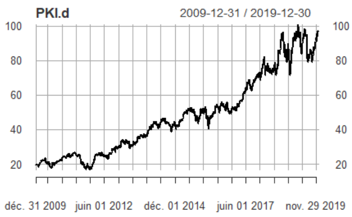
8 stylized facts

The chosen company is PerkinElmer, Inc. which is an American global corporation focused in the business areas of diagnostics, life science research, food, environmental and industrial testing. The company is listed on the NYSE (PKI).

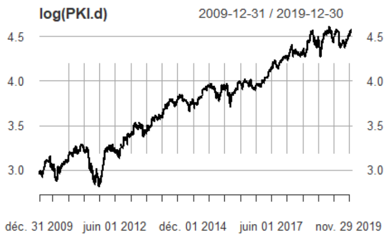
Summary table	Daily log-returns	Monthly log-returns
Mean	0.0006442	0.01350
St. deviation	0.01551177	0.06322579
Diameter C.I. Mean	0.001251991	0.02347346
Skewness	-0.2473808	-0.2266731
Kurtosis	8.802099	3.120867
Excess. Kurtosis	8.8021	3.1209
Min	-0.1272930	-0.17439
Quant. 5%	-0.02324018	-0.09982656
Quant.25%	-0.0069993	-0.02370
Median	0.0008088	0.01927
Quant.75%	0.0087134	0.05266
Quant.95%	0.0234929	0.1085633
Max	0.1059092	0.18458
Jarque.Bera.stat	3553.4	1.1007
Jarque.Bera.pvalue	<2.2e-16	0.5768
Lillie.test.stat	0.06778	0.062639
Lillie.test.pvalue	<2.2e-16	0.2957
N.obs	2515	120

8 STYLIZED FACTS

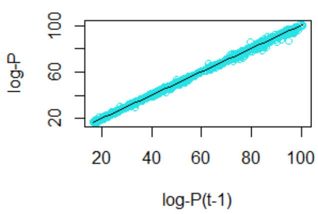
1) Prices are non-stationary



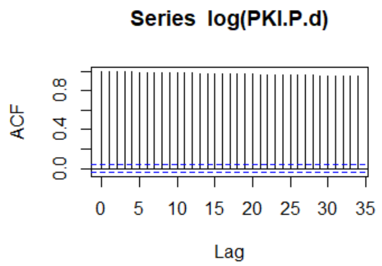
Let’s show that prices of stock PKI have stochastic trend as the actual price constantly depends on past history up to time t. It seems that the price of the stock has exponentially increased and in an irregular manner since 2009, it has definitely not a stationary trend. Indeed, considering the prices as realisations of r.v, the expected value of P_t seems to depend on time. Same for the volatility of the stock (the standard deviation of the price): it looks higher during the period of 2018-2019 than during the rest of the time map.



We can have the same reflexion on the log-prices. The increase of the price looks linear over time which is in tune with the random walk with drift theory. We could write that $p_t = p_0 + t\mu + \sum \varepsilon_t$. (linearity with t). The following graph shows that we can write $p_t = p_{t-1} + \mu + \epsilon$, which implies that parameters depend on time.

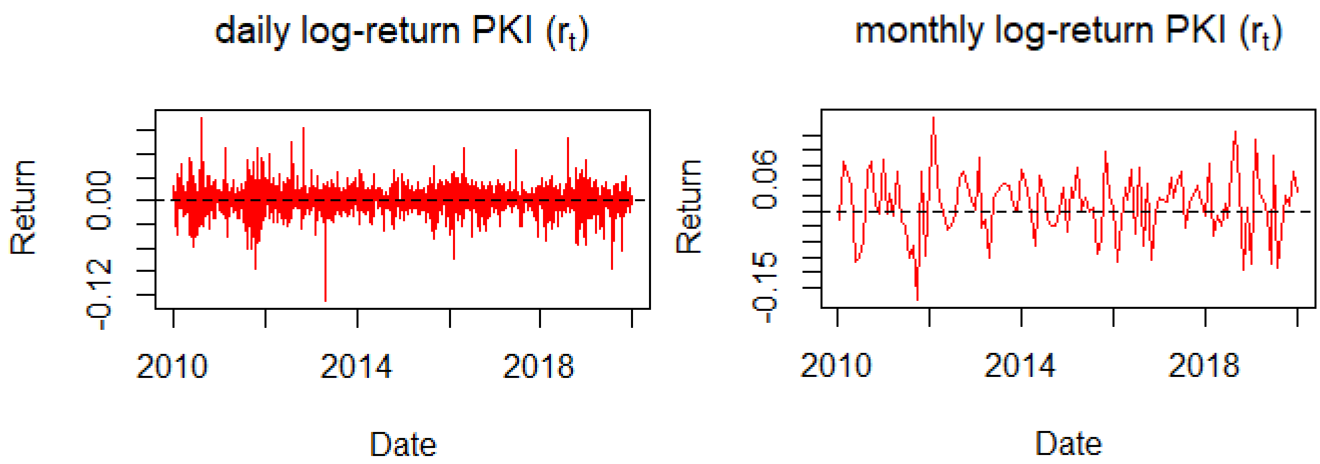


We now consider 2 different time intervals, let's say I_1=[dec. 31 2009, dec. 31 2017] and I_2=[jan. 1 2018, dec. 31 2019]. The empirical mean of the log-price in I_1 is 3.581 and the standard deviation is 0.394 whereas we have 4.436 and 0.093 in I_2. We can then measure if the empirical mean 3.581 of the prices in I_1 sample is significantly different from 4.436 at a level of confidence of 95%. Confidence interval: [3.563671, 3.598105] does not include 4.436. This is not a proof that parameters of the law followed by the r.v log-price statistically depend on time because 4.436 is an estimation, but it is a further evidence as the estimation 4.436 is considerably far from the mean during I_1.

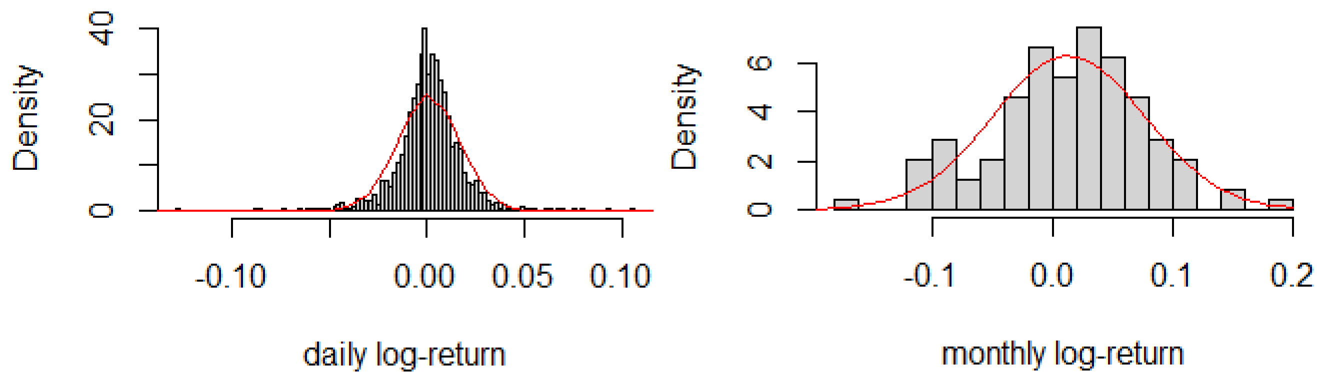


Finally, the graphic at the left presents the ACF applied to the log daily prices. It has very large values (near 1) and slowly decreasing as k increases which is a property of long memory.

2) Returns are stationary



We can see that the log-return fluctuate around the constant value 0. The following histograms compares the daily and monthly log-returns with a normal distribution (red curve) with the same parameters as the estimated ones for our sample:



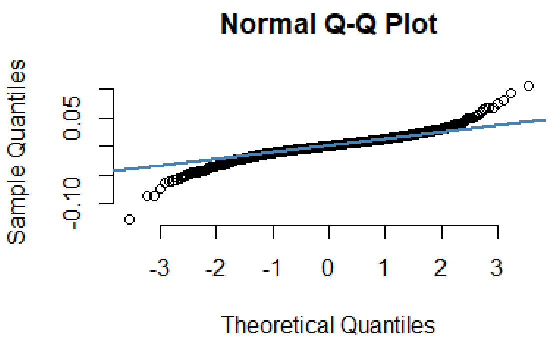
Our distributions seem to be centred: the sample means are very close to 0 but are they significantly equal to 0? If we realize the test with $\hat{\sigma} = 0.01551177$and consider that the sample size is big enough, we have $\frac{\hat{\mu}-0}{\sqrt{\frac{\hat{\sigma}^2}{2515}}} = 2.0827$ which is >1.96, hence we reject H_0 : μ=0 with a significance level of α=5%. This is because even if the sample mean is near 0, the standard deviation is big. If we do the test for the monthly returns, with only 120 observation, we compute with the t-table and we obtain $\frac{\hat{\mu}-0}{\sqrt{\frac{\hat{\sigma}^2}{120}}} = 2.339$ which is >1.984 = $t_{.975}^{100} > t_{.975}^{120}$. Then the real mean of the monthly log-returns is significantly different from 0 with a significance level of α=5%.

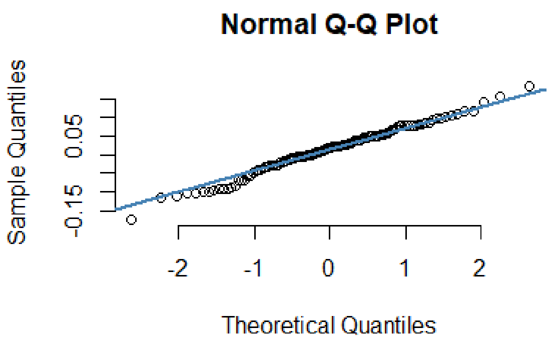
3) Asymmetry

These histograms above are consistent with the values found in the summary table, especially for the kurtosis (leptokurtic) but also for the light negative skewness as the shape of the histogram is sticking out on the left of the red curve (normal distribution). It shows that the distributions of returns are asymmetric and generally negatively skewed.

4) Heavy Tails

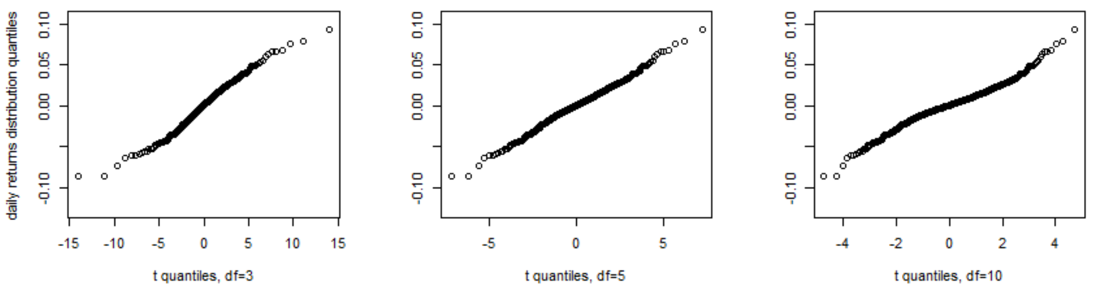
The distribution of returns often exhibits heavy tails. Indeed, in the previous histograms we could see that the distribution was leptokurtic. It is confirmed by the summary table: the kurtosis values are respectively 8.80 and 3.12 for the daily and monthly returns.





These QQ plots are further evidence of the heavy tails. They are used to compare the log-return distributions and the normal distribution N ($\mu = \bar{r}, \sigma^2 = \hat{V}(r_T)$) by comparing the quantiles of both distributions. Following interpretation works for both graphics even if for monthly distribution it looks less extreme: we can see that the points on the left of the graphic which represent the lower quantiles are below the straight line. Therefore, the left tail of our return distribution is heavier than the normal distribution as the lower quantiles of our distribution are smaller than what we expected from normal distribution. Symmetrically, we could do the same reasoning for the right tail.

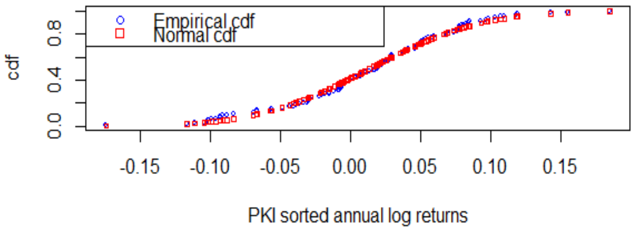
However, we can see that daily returns look far more compatible with Student-t, especially with degree of freedom 5:

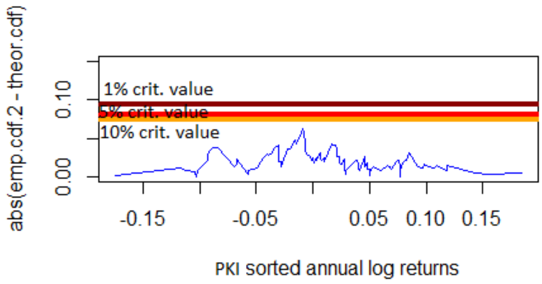


5) High frequency non-gaussianity

Performing the Jarque-Bera test (see summary table), we reject the null hypothesis at very low level of significance (p-value very close to 0, cf summary table). Then, we can reject the normal distribution for both daily returns at very high level of confidence.

Performing the Lilliefors test on monthly returns:





Here, we can see that for all the values of KS absolute difference between G_T and Φ of monthly returns are above the three different significance level lines. Indeed, the critical value for the significance level α = 10% is 0.886/ $\sqrt{T}$ = 0.886/ $\sqrt{120}$ = 0.08088036 which is smaller than the computed value of Lilliefors test statistics KS_L = 0.0964. Hence, we cannot reject the null hypothesis of normality at a 10% confidence level. Further confirmation is in the summary table where we can see that monthly returns distribution has a Lilliefors test p-value of 0.2957 which is higher than 10%.

Concerning aggregational gaussianity, when we calculate the mean of monthly returns divided by the mean of daily return, we obtain 20.95 which is closed to 21 the number of trading days in months.

QQ plots and Lilliefors test have shown that even for the same stock, the return distribution is affected by the frequency of the data and lower frequency return distributions tend to get close to normal distribution.

6) Returns not autocorrelated

Daily returns:

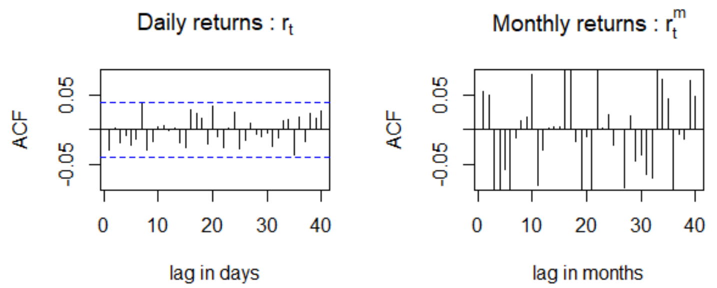
lag	acf	acf diam.	acf test	Box-Pierce stat	BP pval	LB stat	LB pval	crit
1	-0.030	0.039	-1.503	2.259	0.133	2.262	0.133	3.841
2	0.002	0.039	0.118	2.273	0.321	2.276	0.320	5.991
3	-0.019	0.039	-0.973	3.220	0.359	3.224	0.358	7.815
4	-0.008	0.039	-0.384	3.367	0.498	3.372	0.498	9.488
5	-0.022	0.039	-1.115	4.610	0.465	4.619	0.464	11.070
6	-0.013	0.039	-0.643	5.024	0.541	5.033	0.540	12.592
7	0.040	0.039	2.011	9.067	0.248	9.091	0.246	14.067
8	-0.029	0.039	-1.443	11.149	0.193	11.181	0.192	15.507
9	-0.016	0.039	-0.811	11.806	0.224	11.842	0.222	16.919
10	0.004	0.039	0.187	11.841	0.296	11.877	0.293	18.307
11	0.006	0.039	0.319	11.943	0.368	11.979	0.365	19.675
12	0.000	0.039	-0.024	11.943	0.450	11.980	0.447	21.026
13	0.003	0.039	0.131	11.961	0.531	11.997	0.528	22.362
14	-0.019	0.039	-0.967	12.895	0.535	12.937	0.531	23.685

Monthly returns:

lag	acf	acf diam.	acf test	Box-Pierce stat	BP pval	LB stat	LB pval	crit
1	0.055	0.179	0.600	0.360	0.548	0.369	0.543	3.841
2	0.051	0.179	0.555	0.669	0.716	0.688	0.709	5.991
3	-0.275	0.179	-3.014	9.753	0.021	10.161	0.017	7.815
4	-0.155	0.179	-1.693	12.617	0.013	13.173	0.010	9.488
5	-0.058	0.179	-0.635	13.020	0.023	13.601	0.018	11.070
6	-0.094	0.179	-1.035	14.092	0.029	14.747	0.022	12.592
7	-0.012	0.179	-0.127	14.108	0.049	14.765	0.039	14.067
8	0.012	0.179	0.134	14.126	0.079	14.784	0.063	15.507
9	0.018	0.179	0.198	14.165	0.117	14.827	0.096	16.919
10	0.080	0.179	0.876	14.933	0.135	15.679	0.109	18.307
11	-0.080	0.179	-0.875	15.698	0.153	16.536	0.122	19.675
12	-0.029	0.179	-0.317	15.799	0.201	16.649	0.163	21.026
13	0.002	0.179	0.026	15.799	0.260	16.650	0.216	22.362
14	0.004	0.179	0.043	15.801	0.326	16.652	0.275	23.685

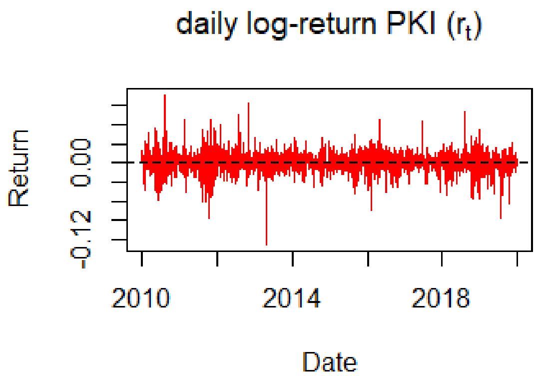
Interpreting results:

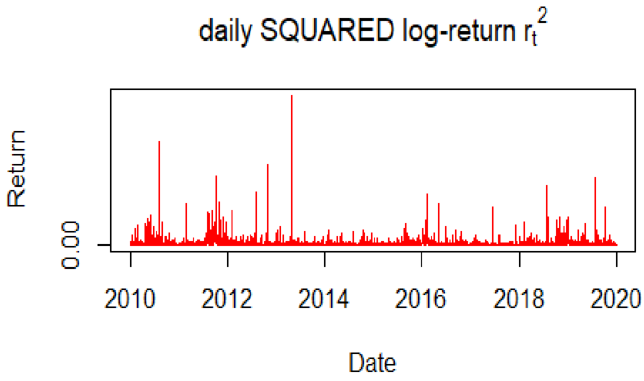
Here, daily and monthly returns ACF are very closed to zero. In both cases, whatever the lag, the ACF are individually different from zero. However, Box-Pierce and Ljung-Box tests exhibit that H_0 (which is the ACF are all equal to zero 'at the same time') is not rejected. These conclusions reveal **that returns are not autocorrelated**. Nonetheless, we can not say that they are independent (it is not an equivalent relation).

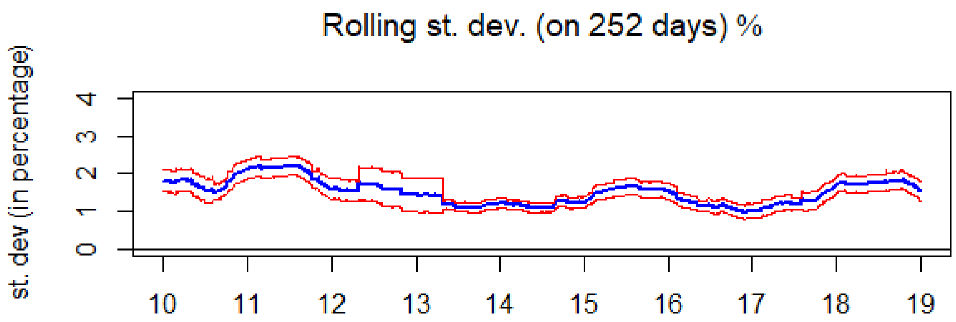


These graphs confirm what we have seen above. For daily returns, the correlations are very low and monthly ACF are slightly higher. In both cases, the ACF estimates seems to randomly fluctuate around 0 which is in tune with the stylized fact.

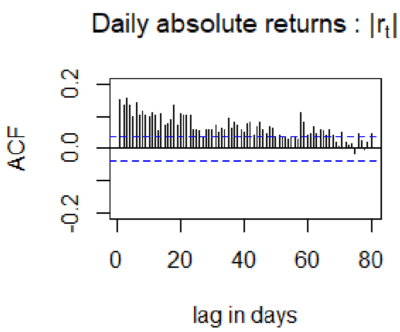
7) Volatility clustering and long range dependence on squared returns

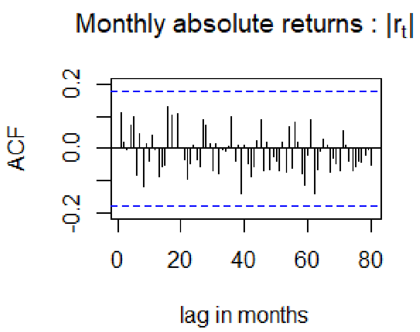


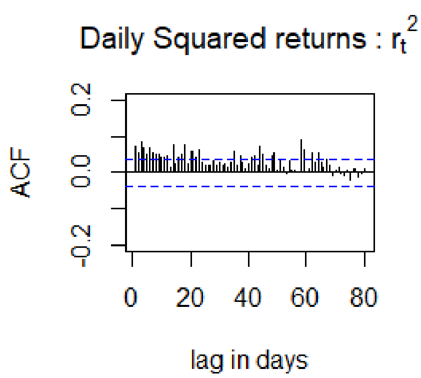


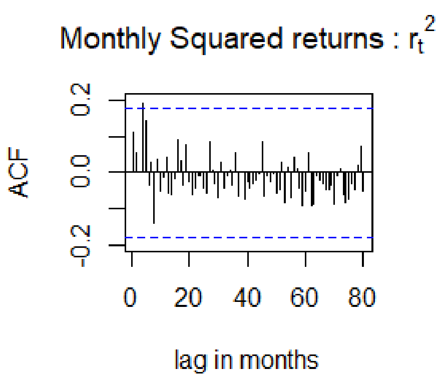


We can see on these graphs the volatility clustering effect on daily prices. Indeed, we can see on the graph on the right side the daily squared log-returns over time and we can see that in general, high squared returns are clustered (for example the periods between 2018 and 2020 or 2010 and 2011) and as well as low squared returns (for example between 2014 and 2015). Well, squared returns are high when returns are far from the mean i.e. when there is high volatility. Hence, the graphs justify volatility clustering. Here are graphs of ACF for absolute returns and squared returns:



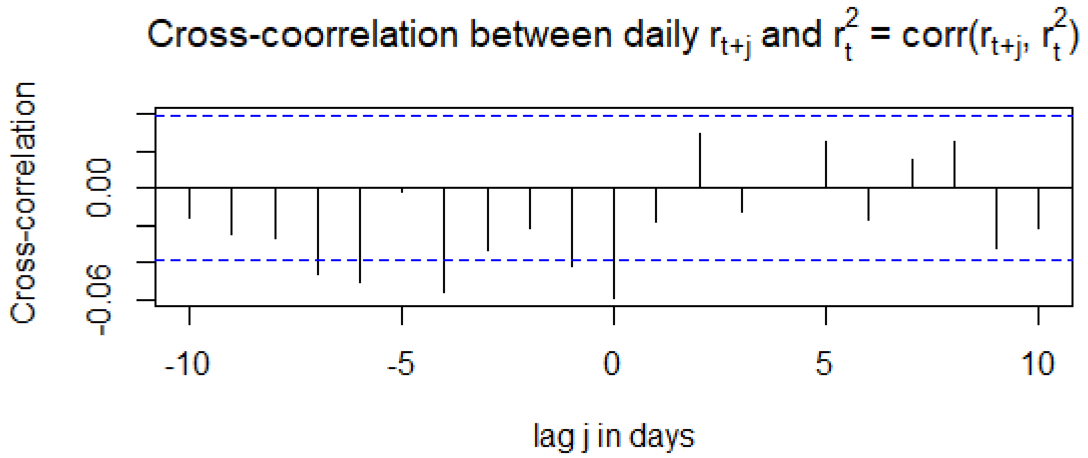






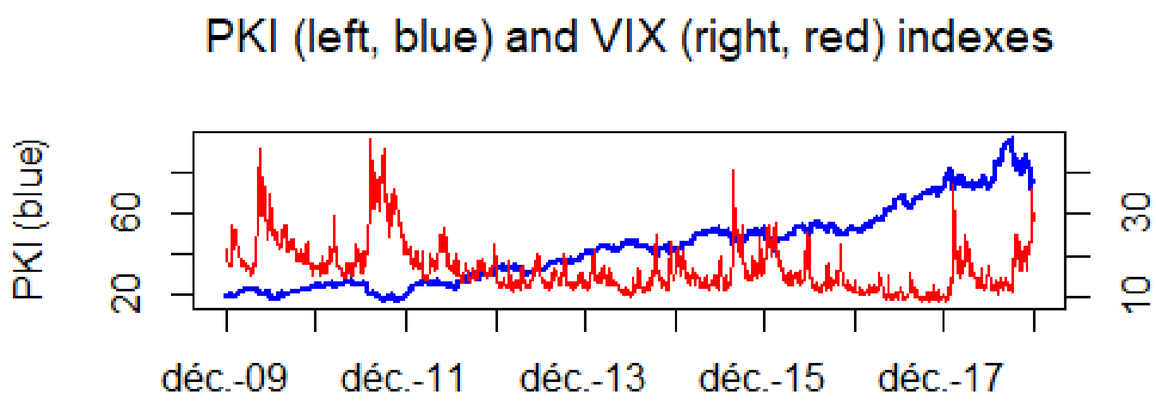
Same interpretation of both |r| and r^2 returns can be done: we can notice that for daily returns, the ACF values slowly decays to zero as the lag increases which indicates possible long-memory properties (long-range dependence). It clearly appears that the daily squared returns are autocorrelated, it is called ARCH effect. It does not work anymore for monthly periods.

8) Leverage effect

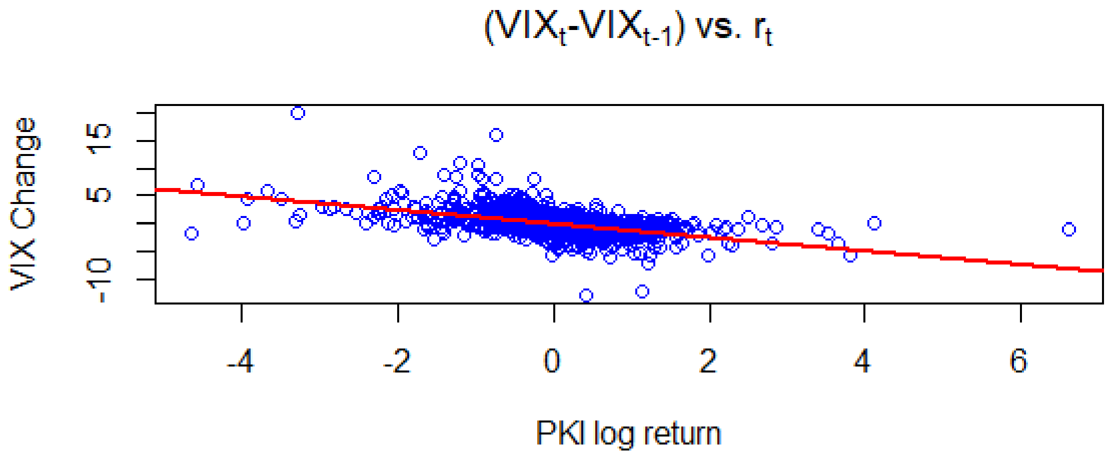


We can observe on this graph that we have a mainly negative correlations between r_{t+j} and r^2_t where j is the lag. It is obvious especially for negative lag: we have corr($r_{t\text{-lag}}$,r_t^2) < 0 for positive lag. This is a first evidence of the leverage effect.

The second evidence comes from the negative correlation between our asset and the changes in the VIX index which is an imperfect measure of expected future volatility of the S&P 500.



If we look carefully at the variations of returns and VIX, we can see that when VIX index skyrockets, there is a drop in PKI returns.



On this graph, we can clearly see that the asset's return is negatively correlated with the change in VIX, ie with a change in approximation of future volatility of the S&P 500. As S&P 500 is one of the most important indexes in the US and PerkinElmer Inc. is listed on the NYSE, we can say that change in VIX is correlated to change in expected PKI's future volatility. Hence, the conclusion is that we have both evidence of leverage effect.

ARCH & GARCH

Introduction

ARCH and GARCH are type models allowing to forecast volatility. More precisely, we use them to compute the realisation of the estimator $\hat{\sigma}_{t+1|t}$ (ie $\forall\omega \in \Omega, \hat{\sigma}_{t|t-1}(\omega)$, where Ω is the set of events provided with $\mathfrak{F}$ the σ-algebra which represents the information available). This estimator $\hat{\sigma}_{t+1|t}$(it is a random variable) is the conditional variance forecast $\mathbb{V}(R_{t+1}|\mathfrak{F}_t)$ *where* $\mathfrak{F}_t$ *is the* σ-algebra generated by $R_t^{-1}(B_t), R_{t-1}^{-1}(B_{t-1}), \dots, R_1^{-1}(B_1)$,($B_1$, … , B_t borelians of $\mathbb{R}$, R_t the return at time t) or the 'available information at time t'.

Presentation of the models and how they fit

ARCH type models are **AutoRegressive Conditional Heteroskedasticity** models. A process (X_t , t an integer) is said to be an ARCH(1(p)), where 1 (p) is the parameter of the model, if $X_t=Z_t\sigma_t$ where (Z_t) iid reduced and centred and $\sigma_t^2=\alpha_0 + \alpha_1 X_{t-1}^2$ ($\sigma_t^2=\alpha_0 + \sum_{1\le i\le p} \alpha_i X_{t-i}^2$).

GARCH type models are **Generalized AutoRegressive Heteroskedasticity** models. A GARCH process (ε_t , t an integer) is different from ARCH process as it has 2 parameters (p,q) and we have(we still have $\varepsilon_t = Z_t\sigma_t$ where (Z_t) iid reduced and centred) $\sigma_t^2 = \omega + \sum_1^p \alpha_i \varepsilon_{t-i}^2 + \sum_1^q \beta_j \sigma_{t-j}^2$ with conditions on ω (>0), alpha and beta (>=0) and the sum of alpha plus the sum of beta <1. GARCH models are more general and the GARCH(1,1) is very useful as it has only 3 parameters to be estimated and fits as well as high-order ARCH (~10).

Given data, we can estimate the parameters of the models as well as their relevance with the **Maximum Likelihood Estimation**. Then, we can as well check the model with a Ljung-Box test on Z_t with $\hat{Z}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$ to check the adequacy of the mean equation and on Z_t^2 to check the volatility.

The objective is now to see how good the following models fits the Nasdaq index returns.

$$R_t = \mu + \phi_1 R_{t-1} + \varepsilon_t$$
$$\varepsilon_t = Z_t\sigma_t, \quad Z_t \sim iidN(0,1)$$
$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2$$

```
Conditional Variance Dynamics
-----------------------------------
GARCH Model      : sGARCH(1,0)
Mean Model       : ARFIMA(1,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000768    0.000253  3.03116 0.002436
ar1    -0.016814    0.035643 -0.47173 0.637117
omega   0.000072    0.000004 18.78231 0.000000
alpha1  0.297212    0.049398  6.01672 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000768    0.000260  2.95154 0.003162
ar1    -0.016814    0.043945 -0.38262 0.702003
omega   0.000072    0.000008  8.78051 0.000000
alpha1  0.297212    0.070077  4.24123 0.000022

LogLikelihood : 4055.01
```

```
Akaike       -6.4404
Bayes        -6.4241
Shibata      -6.4404
Hannan-Quinn -6.4343

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.2011  0.6538
Lag[2*(p+q)+(p+q)-1][2]    0.2749  0.9959
Lag[4*(p+q)+(p+q)-1][5]    1.0239  0.9432
d.o.f=1
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic   p-value
Lag[1]                     0.9665 3.256e-01
Lag[2*(p+q)+(p+q)-1][2]    7.4578 9.009e-03
Lag[4*(p+q)+(p+q)-1][5]   43.3133 5.225e-12
d.o.f=1

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale   P-Value
ARCH Lag[2]     12.94 0.500 2.000 3.214e-04
ARCH Lag[4]     51.99 1.397 1.611 4.374e-14
ARCH Lag[6]     77.09 2.222 1.500 0.000e+00
```

We can see that AR(1) seems not to be significant in the conditional mean equation of the AR(1)-ARCH(1).
If we try now for ARCH(4), excluding AR coef, we obtain :

```
GARCH Model      : sGARCH(4,0)
Mean Model       : ARFIMA(0,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000851    0.000229   3.7179 0.000201
omega   0.000040    0.000003  11.4321 0.000000
alpha1  0.190990    0.039581   4.8252 0.000001
alpha2  0.119047    0.033997   3.5017 0.000462
alpha3  0.134807    0.036234   3.7204 0.000199
alpha4  0.180513    0.037431   4.8226 0.000001
```

Which is better as all the parameters are relevant.

Much more complex models can be used to improve the fit and the forecasts results. These models are for example EGARCH or IGARCH where 'I' means 'Integrated' and 'E' means 'Exponential'. These are asymmetric GARCH. The difference is that for Integrated-GARCH $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + (1 - \alpha)\varepsilon_{t-1}^2 \; where \; \alpha \in [0,1[$, and for Exponential-GARCH, $ln(\sigma_t^2) = \omega + \alpha Z_{t-1} + \gamma(|Z_{t-1}| - \mathbb{E}(|Z_{t-1}|)) + \beta ln(\sigma_{t-1}^2)$. If we specify the IGARCH model for example to estimate the volatility of Nasdaq Index:

```
GARCH Model      : iGARCH(1,1)
Mean Model       : ARFIMA(0,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000794    0.000233   3.4093 0.000651
omega   0.000004    0.000001   3.0034 0.002669
alpha1  0.205951    0.028412   7.2487 0.000000
beta1   0.794049          NA       NA       NA

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000794    0.000259 3.07293 0.002120
omega   0.000004    0.000004 0.99347 0.320482
alpha1  0.205951    0.058820 3.50137 0.000463
beta1   0.794049          NA       NA       NA

LogLikelihood : 4127.96
```

We can notice that all the parameters are relevant and less parameters are required. We can also see that beta1 has no standard error.

Another model is the GJR-GARCH type and it features this expression for sigma squared on t.

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\mathbb{I}_{\varepsilon_{t-1}<0}\,\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

$$\frac{\partial\sigma_t^2}{\partial\varepsilon_{t-1}^2} = \begin{cases} \alpha + \gamma & \text{if } \varepsilon_{t-1} < 0 \\ \alpha & \text{otherwise} \end{cases}$$

Here is the result of the specification :

```
GARCH Model      : gjrGARCH(1,1)
Mean Model       : ARFIMA(0,0,0)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error    t value Pr(>|t|)
mu      0.000394    0.000216   1.820400 0.068698
omega   0.000006    0.000000 124.728019 0.000000
alpha1  0.000000    0.006524   0.000003 0.999998
beta1   0.818751    0.013167  62.180804 0.000000
gamma1  0.236182    0.034066   6.933162 0.000000

Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.000394    0.000213  1.851320 0.064123
omega   0.000006    0.000000 59.601985 0.000000
alpha1  0.000000    0.005834  0.000003 0.999998
beta1   0.818751    0.018023 45.429222 0.000000
gamma1  0.236182    0.045666  5.171916 0.000000
```

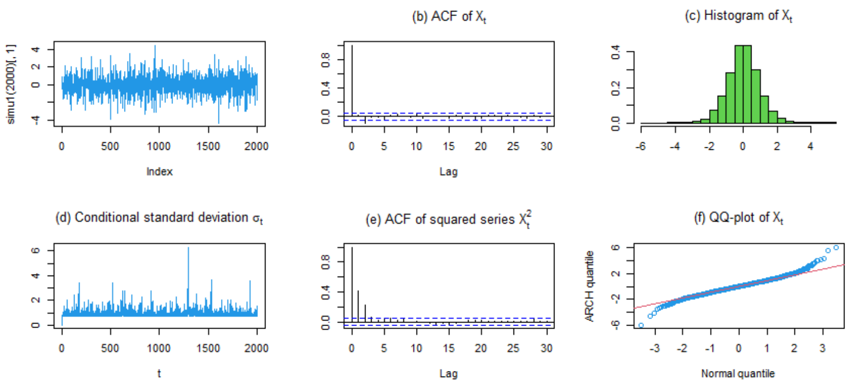
We can see that not all the parameters are relevant. The most interesting is that gamma is strictly positive which is a consequence of leverage effect which is the **8th stylized fact**: asset returns are negatively correlated with the changes in their volatility. It makes a good transition for the next part.

Stylized Facts

Actually, these models are very useful to model asset returns or their error as they feature many stylized facts observable on asset returns.

In order to show these stylized facts, I simulate an ARCH(1) and a GARCH type process:

$$X_t = Z_t\sigma_t$$
$$Z_t \sim iid\,\mathcal{N}(0,1)$$
$$\sigma_t^2 = 0.5 + 0.5X_{t-1}^2$$

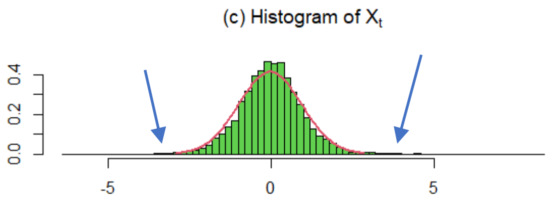


We can already see that this stochastic process has the same visual aspects of the asset returns we have seen in the first question.

Stylized fact 7: the daily squared returns often exhibits significant correlation due to ARCH effect. We can indeed see on the middle bottom graph that for small values of lag k, X_t^2 and X_{t-k}^2 have significant correlation. It can also be proved by calculus by writing the expression of X_t^2 and X_{t-k}^2 in the covariance and then simplify the expression and deduce correlation with the variances.

Stylized Fact 2: autocorrelations of asset returns are often insignificant except for very small intraday time scales. This is exactly what we can observe on the top middle graph: acf values are all equal to zero except for the first lag which would correspond to the 'very small intraday time scale' return. It can be proved showing that X_t is a martingale difference.

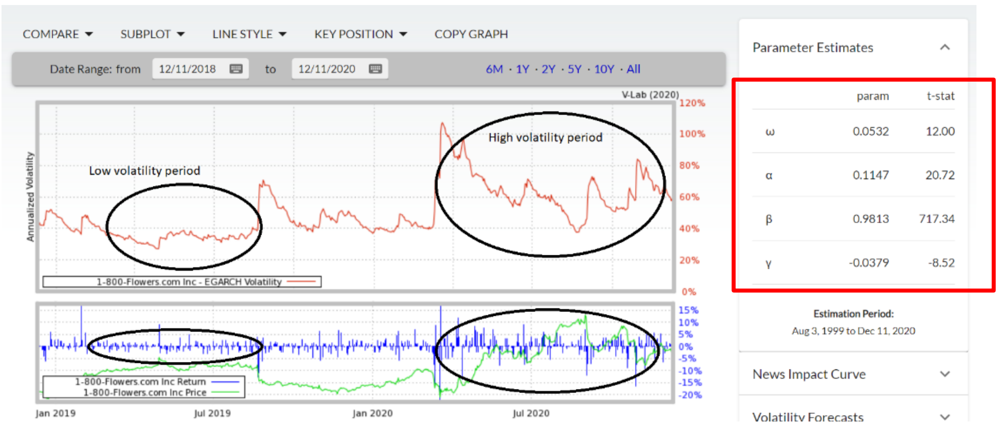
Stylized Fact 1: prices are not stationary whereas returns are stationary. It can be proved that ARCH(1) is weak stationary by checking that unconditional variance is constant, that the process is conditionally heteroskedastic and its two first moment are finite and constant. It is in tune with the first graph which shows that the process is centred.

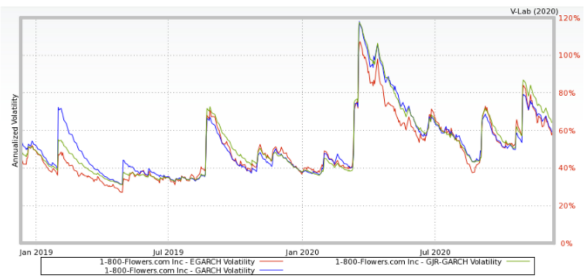


Stylized Fact 3: Heavy tails. We can prove that the kurtosis coefficient is strictly bigger than 3 which implies directly heavy tails. See also QQ-plots with extrem values out of the line.

Stylized Fact 5: volatility clustering. We can notice this phenomenon on the above bottom left graph where we can see that high volatility is grouped by periods.

Here is a specification of an EGARCH model to perform a volatility analysis of 1-800-Flowers.com Inc..





We can see the annualized volatility which confirm the volatility clustering stylized fact. Stationarity of returns and non-stationarity of prices can also be noticed. We can also see the parameters estimations with a negative gamma!

Finally, we can also compare the fit of the EGARCH (red) model with the GJR-GARCH (green) and GARCH (blue) model and see that the results of the analysis are quite the same with lower volatility of EGARCH and generally GJR-GARCH and GARCH look very closed to each other.