Quantitative Techniques in Finance
Variance estimation

SID: 829283

Introduction

It is very well known concept, that markets behave unexpectedly, often rallying, or what is more harmful, plunge within very short periods of time. Size of those movements are a measure of uncertainty in the market. Low volatility, is a feature of stable conditions, while high volatility, occurs during panic, when irrational behaviors are more common. Based on current conditions, market participants have to adjust their positions to match actual situation. Pricing models, play important role in this process. But they are dependent on actual measures of market volatility. Therefore, accurate estimation of this variable, is a key to accurate pricing of derivatives, options especially. This work has been prepared to present key issues of estimation process for Volatility and Correlation coefficients between assets. Number of methods have been employed (MA, EWMA, ARCH(1) and GARCH(1,1)) and compared, as well as their advantages and disadvantages briefly outlined. Whole report is based on a case study of two companies: BAT and Prudential both listed on FTSE100.

Theory and assumptions

Before estimation results will be presented, few assumptions had to be made. In this section, they will be described and briefly analyzed.

Stationarity

Estimating fundamental parameters for data is only possible if data are stationary. Calculating returns transforms data from non-stationary price value to stationary returns.

Normality

Returns are assumed to be stationary random variable with normal distribution. Because of this assumption, one can estimate expected returns (as a mean value of returns) and volatility (expressed as standard deviation of returns). Those two variables fully describe shape of normal curve.

Assumptions for variance estimate (Hull, 2002)

$$\sigma^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\mu_{n-i} - \bar{\mu})^2$$

where

$$\bar{\mu} = \frac{1}{m}\sum_{i=1}^{m}\mu_{n-i}$$

1. $\mu_i = \frac{S_i - S_{i-1}}{S_{i-1}}$
2. $\bar{\mu}$ is assumed to be zero
3. $m-1$ is replaced by m

As a result of above assumptions, following equation is formed for **equally weighted** variance and covariance

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} {\mu_{n-i}}^2 \qquad cov_n = \frac{1}{m}\sum_{i=1}^{m} \mu_{n-i}\nu_{n-i} \tag{1}$$

Which can be modified to assign **different weights** resulting in **weighted variance**

$$\sigma^2 = \sum_{i=1}^{m} \alpha_i {\mu_{n-i}}^2 \qquad cov_n = \sum_{i=1}^{m} \alpha_i \mu_{n-i}\nu_{n-i} \tag{2}$$

where

$\sum_{i=1}^{m} \alpha_i = 1$ and for exponentially weighted $\alpha_i = (1-\lambda)\lambda^{i-1}$

Assuming weighting and introducing long term variance and covariance, **ARCH(m)** model is formed

$$\sigma^2 = \varpi + \sum_{i=1}^{m} \alpha_i {\mu_{n-i}}^2 \qquad cov_n = \varpi + \sum_{i=1}^{m} \alpha_i \mu_{n-i}\nu_{n-i} \tag{3}$$

where

$\varpi = \gamma V$

Assuming exponential decrease of weights, formula for **Exponentially Weighted Moving Average** is formed

$$\sigma^2 = \lambda\sigma_{n-1}^2 + (1-\lambda){\mu_{n-i}}^2 \qquad cov_n = \lambda cov_{n-1} + (1-\lambda)\mu_{n-i}\nu_n \tag{4}$$

where

$0 < \lambda < 1$

Expanding EWMA by long-run average variance/covariance, **GARCH(1,1)** model is formed

$$\sigma^2 = \varpi + \alpha{\mu_{n-i}}^2 + \beta\sigma_{n-1}^2 \qquad \sigma^2 = \varpi + \alpha\mu_{n-i}\nu_{n-i} + \beta cov_{n-1} \tag{5}$$

Where

$\gamma = 1 - \alpha - \beta, \alpha + \beta < 1, \varpi = \gamma V$

GARCH model can be judged based on its ability to remove autocorrelation from returns μ_i

In this work, every mentioned method for estimating volatility will be used, but in different order than they were presented.

Estimation case study

Data Description

Data employed for this study are prices for two companies listed on London Stock Exchange, constituents of FTSE100: British American Tobacco (BAT) and Prudential Inc. (PRU). There are 1817 days of raw prices (Figure 1. Price over time), which have been transformed into simple returns.

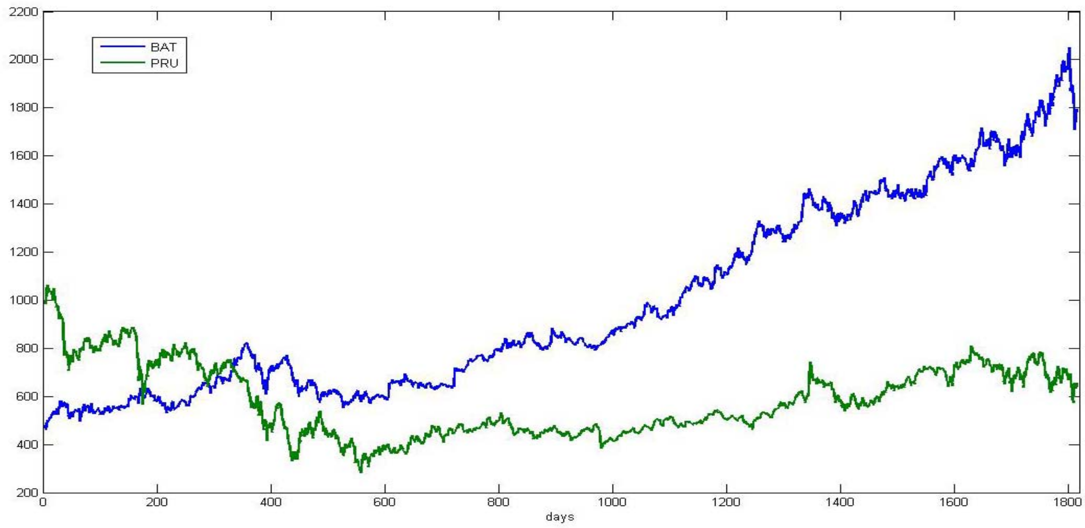

Figure 1. Price over time

Theoretically obtained returns should be normally distributed. This has been tested using asymptotic Jarque-Berra test for normality and confirmed by Lilliefors tests for small samples. Although histograms form bell shaped curve, applied tests suggest rejection of H_0, that returns are normally distributed (at 5% significance level).

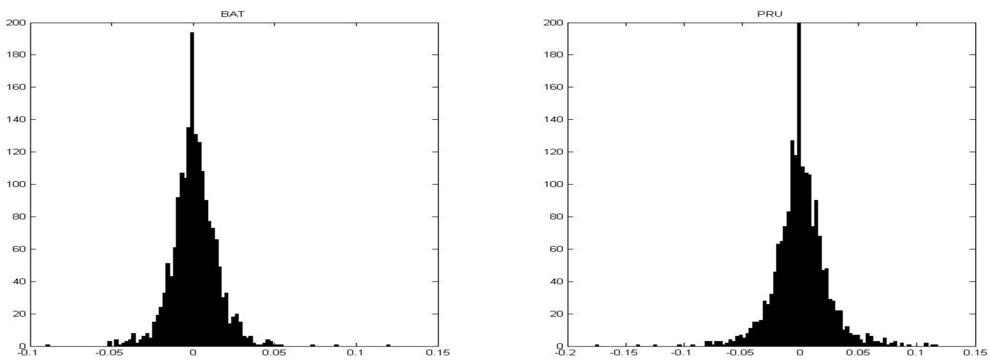

Figure 2. Histograms of BAT and PRU returns

Pre-estimation Analysis

At first, Autocorrelation (ACF) and Partial Autocorrelation (PACF)of raw returns and squared returns have been examined. Both are presented on Figure 3 and Figure 4. Apart from individual values, correlation stays low thus it is not required to use methods which eliminate high autocorrelation. Only squared returns of PRU indicate higher autocorrelation than standard deviation confidence bounds. To prove visual conclusions, Ljung-Box Q-test has been applied to test departure from randomness. Results confirm (Table 1, where H_0 - no serial correlation, H_1 – serial correlation exists) confirm that there exists autocorrelation effect within returns.

Table 1. Results of Q-test

	Lag	H	p-Value	Stat	Critical Value
BAT	10	1	0.000	30.00	18.30
	15	1	0.004	33.05	24.99
	20	1	0.005	39.91	31.41
PRU	10	1	0.002	27.66	18.30
	15	1	0.000	50.80	24.99
	20	1	0.000	60.13	31.41

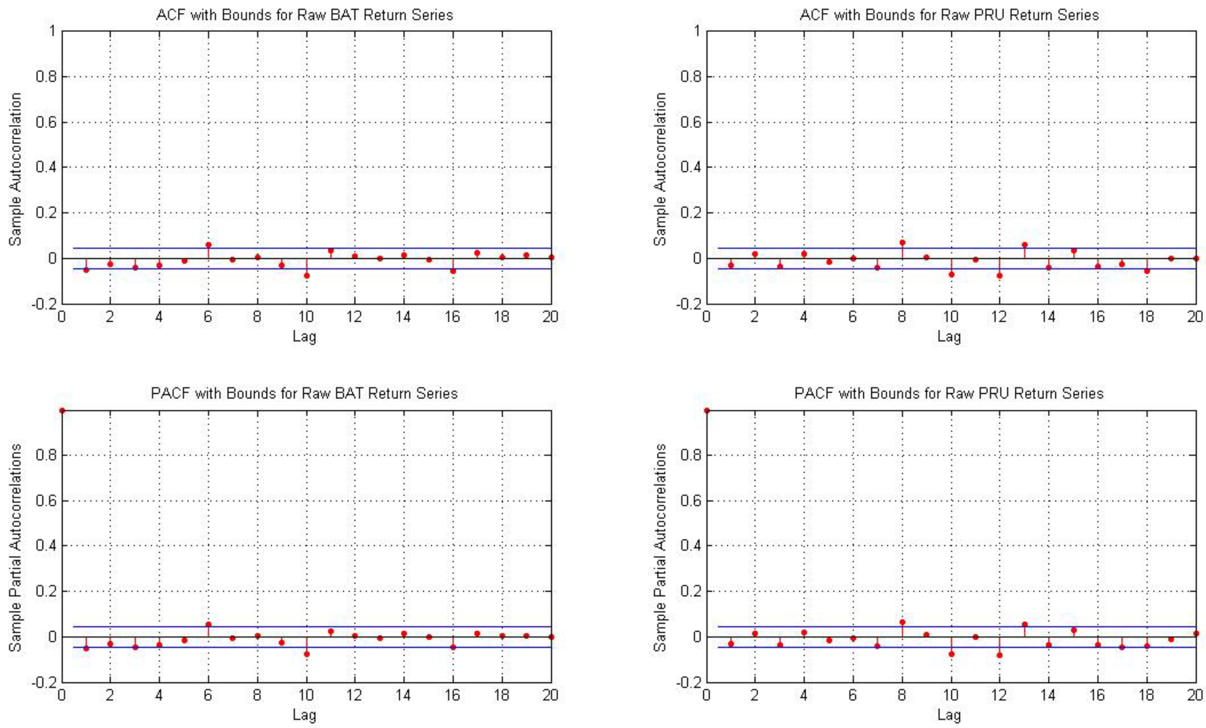

Figure 3. ACF and PACF of raw returns

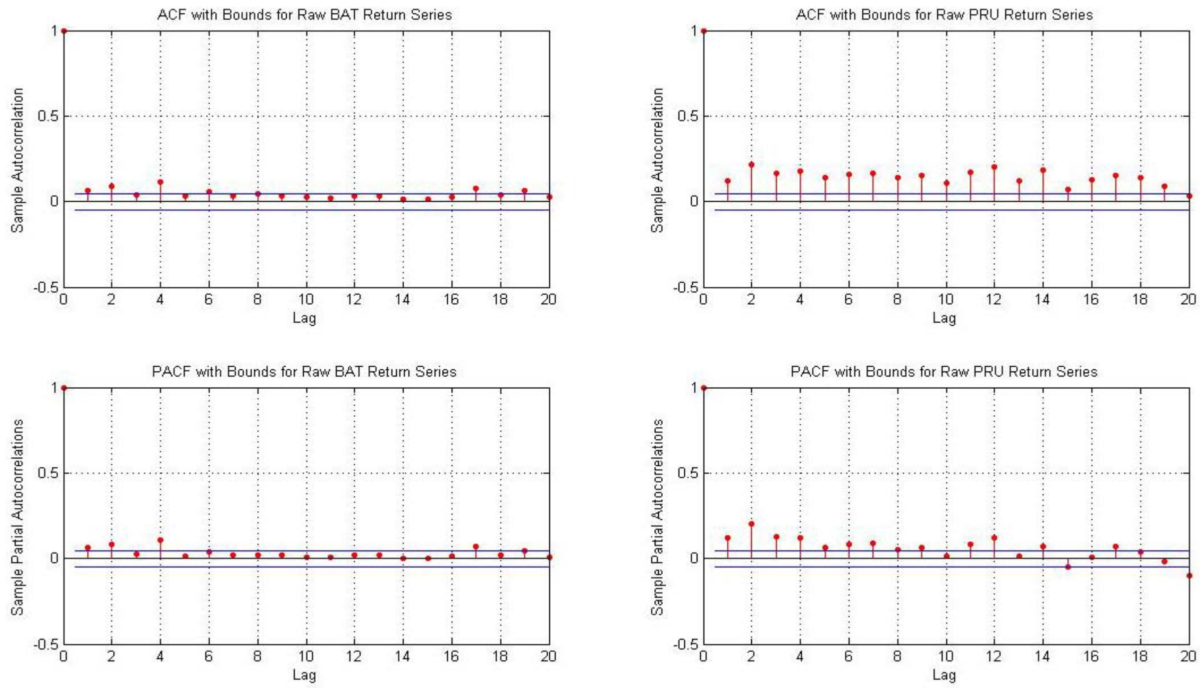

Figure 4. ACF and PACF of squared returns

Note: Presented in this report values relate to last obtained value for given returns.

Estimation methodology

Experiments described in this work, have been designed in Matlab environment as this is very flexible and user friendly programming and scripting platform. Algorithms (with source code published in appendix), have been launched and compared for different estimation period size. This was to investigate influence of estimation period size for the final value. As presented, output from different methods is similar, but due to method specific characteristics, they are not exactly the same.

Table 2. Equally weighted estimates (eq.1)

Est. per.	Volatility: $\sqrt{\sigma^2}$		Covariance	Correlation
	BAT	PRU	$x\ 10^{-3}$	
Full - 1815	0.0140	0.0234	0.0749	0.2289
1000	0.0111	0.0174	0.0551	0.2860
500	0.0118	0.0206	0.0823	0.3402
250	0.0131	0.0230	0.1176	0.3921
100	0.0155	0.0278	0.1139	0.2646
50	0.0180	0.0321	0.1525	0.2642
25	0.0193	0.0317	0.0249	0.0407
10	0.0231	0.0415	-0.0022	-0.0023

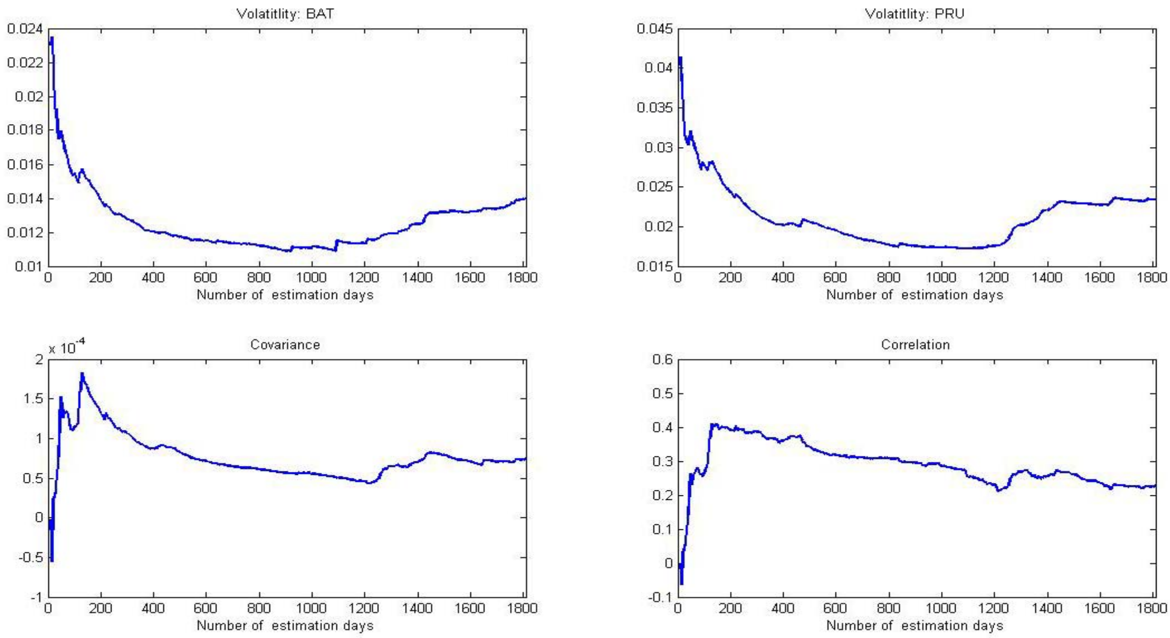

Figure 5. Equally weighted estimate (eq. 1)

Equally weighted method (eq 1) is influenced equally by most recent as well as very past data. This makes it completely variable, whatever period length is used for estimation. It is clearly shown in the results on Figure 5 and Table 2.

Comments: This method is easy and straightforward to calculate, requires no parameters estimation, but careful period selection is needed. It is also slightly expensive in terms of computing resources as full data has to be kept for calculation.

Table 3. Exponentially Weighted Moving Average estimate (eq 2)

Est. per.	Volatility: $\sqrt{\sigma^2}$		Covariance	Correlation
	BAT	PRU	$x\ 10^{-3}$	
Full - 1815	0.0197	0.0325	0.5647	0.0883
1000	0.0197	0.0325	0.5647	0.0883
500	0.0197	0.0325	0.5647	0.0883
250	0.0197	0.0325	0.5647	0.0883
100	0.0197	0.0324	0.5591	0.0875
50	0.0195	0.0321	0.5367	0.0855
25	0.0184	0.0296	0.1936	0.0354
10	0.0150	0.0257	-0.0005	-0.0001

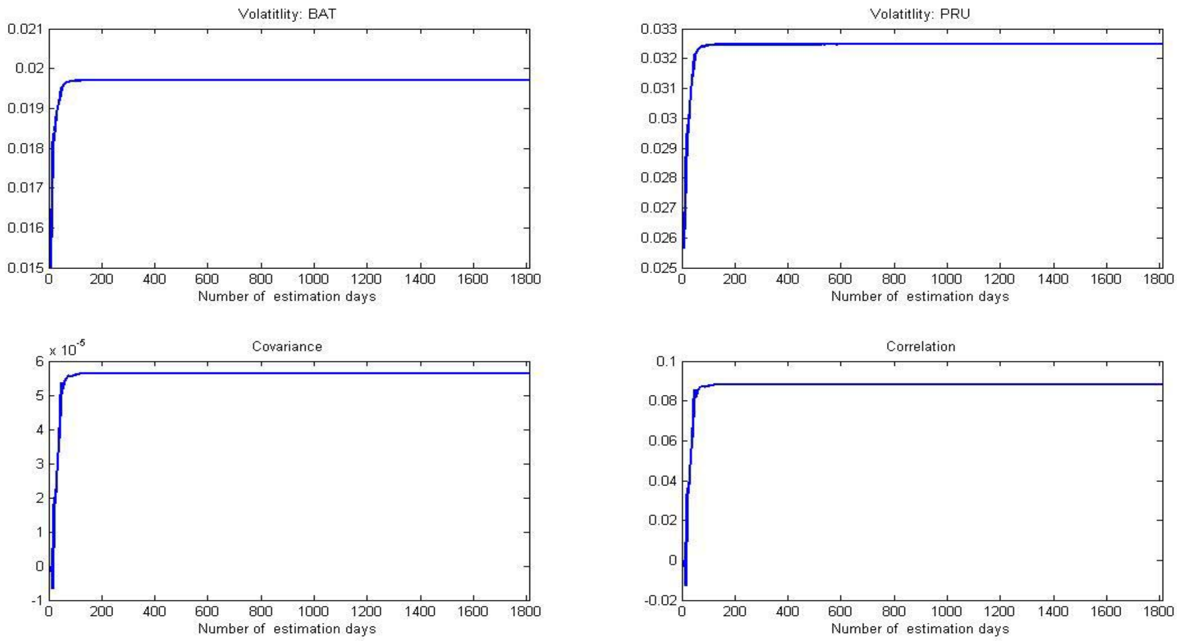

Figure 6. Exponentially Weighted Moving Average (eq 2)

Exponentially weighted method has been developed to overcome one of the biggest disadvantage of Equally weighted method. It assigns different weights, with respect to how recent data are. In this way, most recent values are more important than older, thus it is expected that this estimate better reflects real value of volatility. Table 3 and Figure 6. Exponentially Weighted Moving Average (eq 2) shows results of estimation with respect to estimation period size. As clearly seen, data which are older than approx 80days have such small weight, that it does not influence final estimate at all.

Comments: Exponentially Weighted Moving Average is less computationally expensive than equally weighted method. Here, two algorithms for its calculations have been compared (eq 2 and eq 4 – simplified). As presented on Figure 6 and Figure 7, calculated value is not much different from each other, but simplicity in calculation using equation 4 is tremendous. It is enough to keep only last calculated value, thus full sample history is not required (as opposed to equally weighted method).

Table 4. Exponentially Weighted Moving Average (eq 4)

Est. per.	Volatility: $\sqrt{\sigma^2}$		Covariance	Correlation
	BAT	PRU	$x\ 10^{-3}$	
Full - 1815	0.0201	0.0331	0.7642	0.1151
1000	0.0201	0.0331	0.7642	0.1151
500	0.0201	0.0331	0.7642	0.1151
250	0.0201	0.0331	0.7642	0.1151
100	0.0201	0.0331	0.7583	0.1143
50	0.0199	0.0327	0.7345	0.1128
25	0.0188	0.0301	0.3695	0.0654
10	0.0152	0.0259	0.1630	0.0415

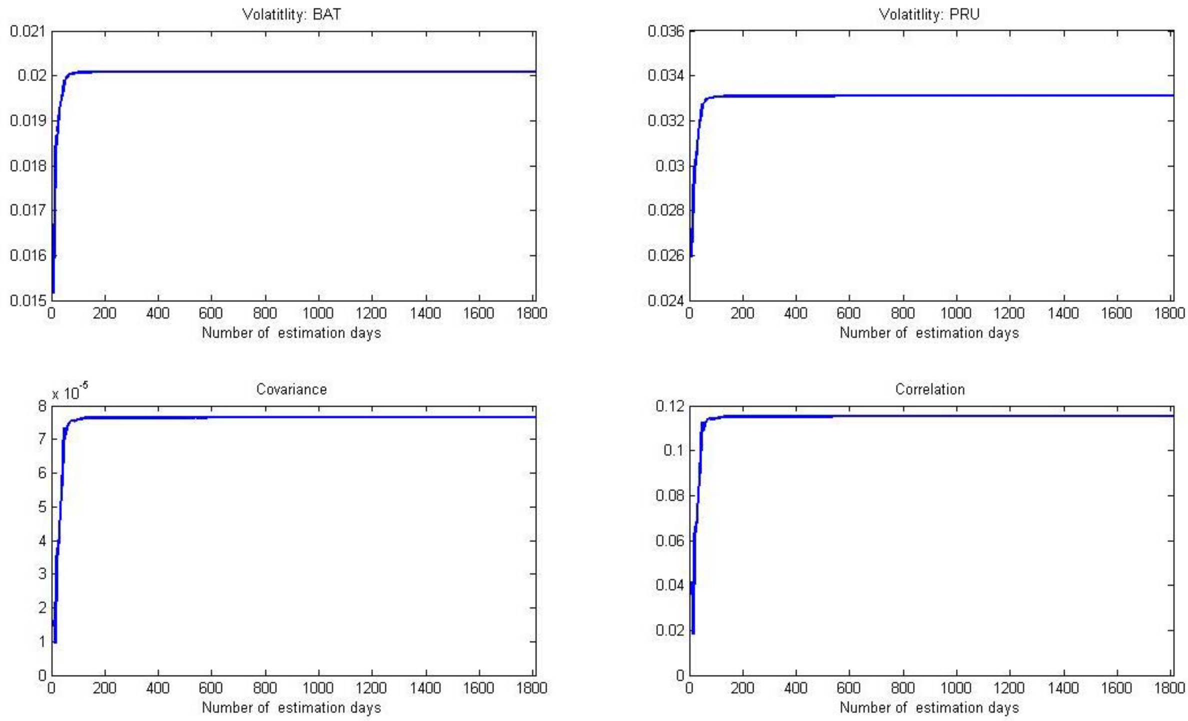

Figure 7. Exponentially Weighted Moving Average (eq 4)

Exponentially weighted moving average requires parameter λ (decay factor) to be set up. In this study, it has been assumed that $\lambda = 0.94$ as suggested in study by JP Morgan in Risk Metrics (JPMorgan 1995).

Table 5. ARCH(1)

Est. per.	Volatility: $\sqrt{\sigma^2}$		Covariance	Correlation
	BAT	PRU	$x\ 10^{-3}$	
Full - 1815	0.0172	0.0285	0.1805	0.3670
1000	0.0172	0.0285	0.1805	0.3670
500	0.0172	0.0285	0.1805	0.3670
250	0.0172	0.0285	0.1805	0.3670
100	0.0172	0.0285	0.1801	0.3667
50	0.0171	0.0283	0.1789	0.3689
25	0.0164	0.0267	0.1595	0.3632
10	0.0143	0.0243	0.1486	0.4269

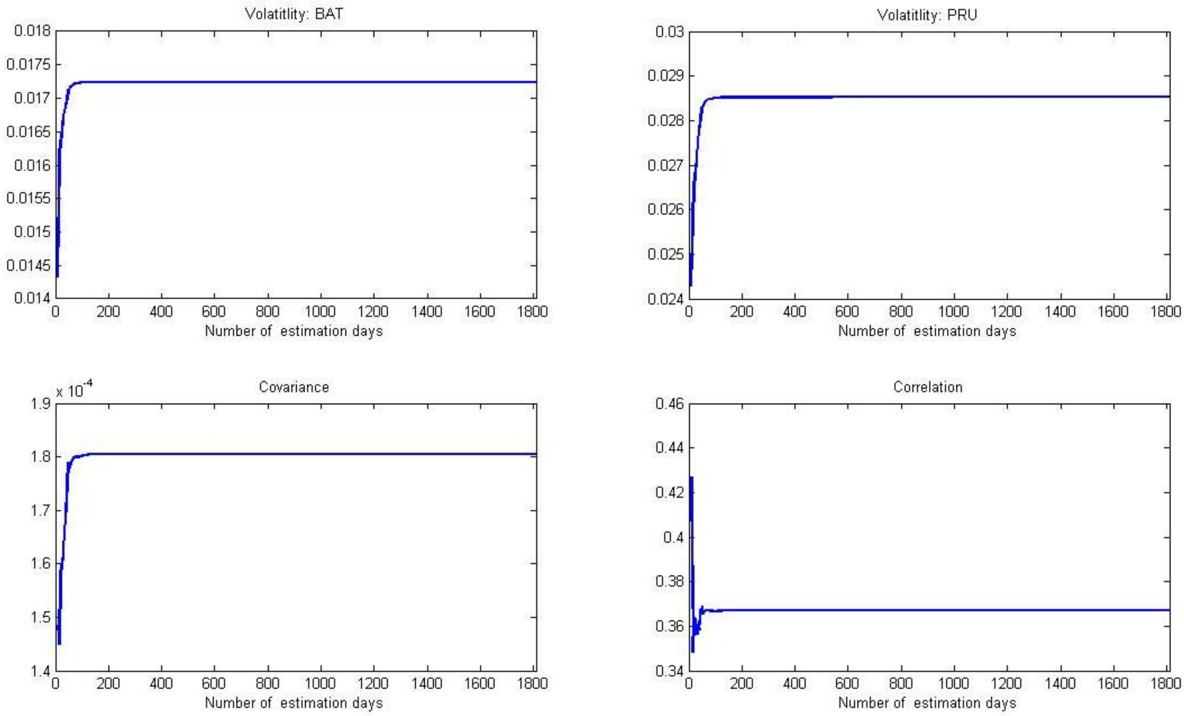

Figure 8. ARCH(1)

Estimate of ARCH(1), similarly to EWMA converges to a constant value, as weights for further periods become very close to 0. It is also influenced by long term value of variance, therefore estimate differ from other methods. Parameters for this model (eq 3) have been set the following way: decay factor $\lambda = 0.94$, long term variance weight $\gamma = 0.4$.

Table 6. GARCH (1,1)

Est. per.	Volatility: $\sqrt{\sigma^2}$		Covariance	Correlation
	BAT	PRU	$x\ 10^{-3}$	
Full - 1815	0.0276	0.0361	0.3815	0.3822
1000	0.0276	0.0361	0.3329	0.3335
500	0.0276	0.0361	0.2759	0.2764
250	0.0276	0.0361	0.3602	0.3609
100	0.0276	0.0361	0.3347	0.3353
50	0.0276	0.0361	0.3837	0.3844
25	0.0276	0.0356	0.3178	0.3231
10	0.0264	0.0324	0.5257	0.6146

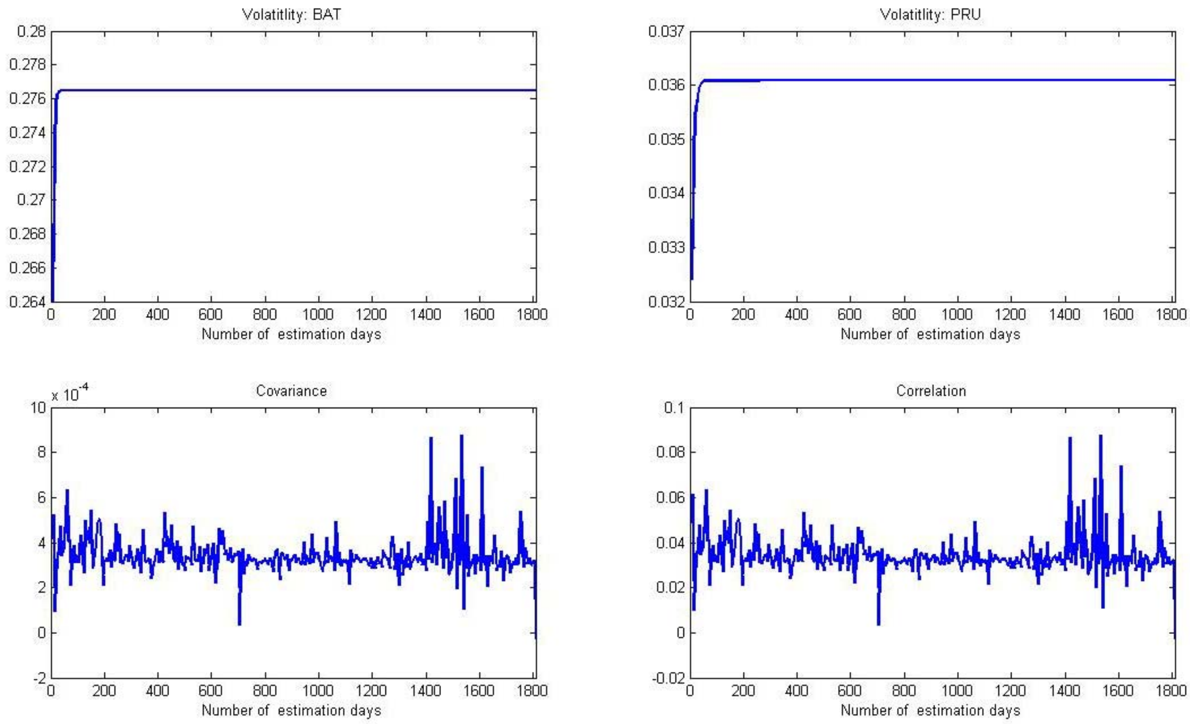

Figure 9. GARCH(1,1)

Garch(1,1) is an extension to EWMA in a simplified way. The difference is in including long term constant volatility and assigning it some weight. Therefore, model (eq. 5) has 3 parameters: α, β, ϖ. The task of estimating those is done using Logarithm Maximum Likelihood Estimator. Based on eq. 6, parameters have been estimated using Excel Solver (Appendix). Results based on MLE for each of the series are presented in Table 7. These estimates has been obtained for the full sample (Example spreadsheet attached in appendix).

$$\sum_{i=1}^{m}\left[-\ln(v_i)-\frac{u_i^2}{v_i}\right] \quad (6)$$

Table 7. GARCH(1,1) Parameters

	$\varpi\ [\cdot 10^{-3}]$	α	β
BAT	1.8000	0.2185	0.7635
PRU	0.0063	0.1245	0.8755
Covariance	0.5000	0.5894	0.4101

Comments: GARCH(1,1) requires parameters estimation, but after this process is performed, it is less demanding while calculating volatility for each day. It is important to state, that in case of GARCH model, estimation of best parameters is a step, which might be laborious. Although it requires highest number of parameters, it has desired characteristic of variance estimator, i.e. accounts for mean reversion or clustering existent in real data.

Conclusions

Results obtained for full period estimation have been presented in

	Equally weighted	Exponentially weighted (2)	Exponentially weighted (4)	ARCH(1)	GARCH(1)
BAT	0.0140	0.0197	0.0201	0.0172	0.0276
PRU	0.0234	0.0325	0.0331	0.0285	0.0361
Correlation	0.2289	0.0883	0.1151	0.3670	0.3822

Each of presented methods generated similar answers. Obtained differences are the reason of application of different parameters. Also, methods differ between each other, thus they tend to produce varying results. The task of selecting the best set of those parameters is therefore very important. Especially, it is possible to measure how good is the model, by comparing calculated values with realized values which might work as a benchmark (i.e. realized implied volatility). Only if differences are minimal, assets can be priced accurately and arbitrage opportunity is maximally reduced what is the goal of institutions trading at derivatives market.

Main conclusion after performing this project is, that simple average might be the worst estimate for a short period of time. In tends to true value, but huge data set is required. Weighted moving average tend to be able to limit impact of estimation window size choice, but in this case, it is suggested to perform pre-estimation in order to choose best estimation size, rather than too small, because then errors are much higher.

In general, including more parameters in the model, tend to make it complicated, what does not guarantee, that it will be better. Simple methods are easier to understand and calculate, often producing solution very similar to the one which is desired and not significantly different from more complicated methods.

Bibliography

Brooks, C. (2002). *Introductory Econometrics for Finance.* Cambridge Univesrity Press.

Hull, J. C. (2002). *Options, Futures, and Other Derivatives (5th Edition).* Prentice Hall.

JP Morgan. (1995). Risk Metrics.

MathWorks Inc. (2007). Matlab Help File.

Appendix – Matlab source code prepared for calculations

Listing 1. Main source file

```
clear all;

[FileName,PathName,FilterIndex] = uigetfile('*.xls')
[price, txt]=xlsread(strcat(PathName,FileName), -1);

ret = tick2ret(price);

fl = size(ret,1);
perds = [1815 1000 500 250 100 50 25 10];

sBat = [];
sPru = [];
sCov = [];

gpar = [0.017973175 0.218502001 0.763525824;
        6.35059E-06 0.124520878 0.875472771;
        0.000499833 0.589449761 0.410051406];

% perds = seqreverse([ 10 : 5 : 1815]);
perds = [ 10 : 5 : 1815];
for i= 1 : size(perds,2)
```

```
    %Standard, equally weighted formula
    stB = std(ret(fl-perds(i)+1:end,1));
    stP = std(ret(fl-perds(i)+1:end,2));
    sBat = [ sBat;  stB];
    sPru = [ sPru;  stP];
    cot = cov(ret(fl-perds(i)+1:end, 1:2));
    sCov = [sCov; cot(1,2)];

    % Simple exponentially weighted formula
    stB = ewma(ret(fl-perds(i)+1:end,1), 0.94)^0.5;
    stP = ewma(ret(fl-perds(i)+1:end,2), 0.94)^0.5;
    sBat = [ sBat;  stB];
    sPru = [ sPru;  stP];
    cot = ewmacov(ret(fl-perds(i)+1:end, 1:2), 0.94);
    sCov = [ sCov; cot];

    % Simplified exponentially weighted formula
    stB = ewmas(ret(fl-perds(i)+1:end,1), 0.94)^0.5;
    stP = ewmas(ret(fl-perds(i)+1:end,2), 0.94)^0.5;
    sBat = [ sBat;  stB];
    sPru = [ sPru;  stP];
    cot = ewmacovs(ret(fl-perds(i)+1:end, 1:2), 0.94);
    sCov = [ sCov; cot];

    % GARCH weighted formula
    stB = mygarch(ret(fl-perds(i)+1:end,1), gpar(1,:))^0.5;
    stP = mygarch(ret(fl-perds(i)+1:end,2), gpar(2,:))^0.5;
    sBat = [ sBat;  stB];
    sPru = [ sPru;  stP];
     cot = ewmacovs(ret(fl-perds(i)+1:end, 1:2), gpar(3,:));
    sCov = [ sCov; cot];
end

rap = [1815; 1000; 500; 250; 100; 50; 25; 10];
disp('sBat')
sBat(rap/5-1)
disp('sPru')
sPru(rap/5-1)
disp('sCov')
sCov(rap/5-1)
sCor = sCov./(sBat.*sPru);
disp('sCor')
sCor(rap/5-1)

%% plot
lt = size(perds, 2)  ;
figure('Name', 'EWMA (eq 1)', 'Color', 'White')
```

```
subplot(2,2,1); plot(perds(1:lt), sBat(1:lt), 'LineWidth', 2); title('Volatitlity: BAT'), xlim([0 lt * 5]);
xlabel('Number of  estimation days')
subplot(2,2,2); plot(perds(1:lt), sPru(1:lt), 'LineWidth', 2); title('Volatitlity: PRU'), xlim([0 lt * 5]);
xlabel('Number of  estimation days')
subplot(2,2,3); plot(perds(1:lt), sCov(1:lt), 'LineWidth', 2); title('Covariance'), xlim([0 lt * 5]); xlabel('Number
of  estimation days')
subplot(2,2,4); plot(perds(1:lt), sCor(1:lt), 'LineWidth', 2); title('Correlation'), xlim([0 lt * 5]); xlabel('Number
of  estimation days')
```

Listing 2. Exponentially Weighted

```
function v = ewmas( ret, la );

vt = 0;
sr = size(ret,1);

for i = 2 : sr
   vt = [ vt la * vt( i - 1) + (1 - la)*ret(i - 1)^2 ];
end

v = vt(end);
```

Listing 3. Exponentially Weighted – Simplified

```
function v = ewma( ret, la )

sr = size(ret,1);
v = 0;
for i = 1 : sr
   v = v +  la^( sr-i ) * ret( i )^2;
end
v = ( 1 - la ) * v;
```

Listing 4. ARCH(1)

```
function v = myarch( ret, om, lw )

al = 0.94;
sr = size(ret,1);
v = 0;

for i = 1 : sr
   v = v +  al^(sr-i+1) * ret( i )^2;
end

v = (1-om)*lw + om*( 1 - al ) * v;
```

Listing 5. GARCH(1, 1)

```
function v = mygarch(ret, par)

O = par(1);
A = par(2);
B = par(3);

vt = 0;
sr = size(ret,1);
for i = 2 : sr
   vt = [ vt; O + A*ret(i - 1)^2 + B * vt( i - 1)];
end

 v = vt(end);
```

Appendix – Sample GARCH excel parameters estimation.

Price		ui	vi	MLE	w	a	B	sum
470	1				0.0018	0.2185	0.7635	=SUM(F2:H2)
468	2	=(A3-A2)/A2						
474.5	3	=(A4-A3)/A3	=C3^2	=-LN(D4)-(C4^2)/D4		Oldest on Top		
474.5	4	=(A5-A4)/A4	=$F$2+$G$2*C4^2+$H$2*D4	=-LN(D5)-(C5^2)/D5				
460.75	5	=(A6-A5)/A5	=$F$2+$G$2*C5^2+$H$2*D5	=-LN(D6)-(C6^2)/D6				
460.75	6	=(A7-A6)/A6	=$F$2+$G$2*C6^2+$H$2*D6	=-LN(D7)-(C7^2)/D7				
1985	1803	=(A1804-A1803)/A1803	=$F$2+$G$2*C1803^2+$H$2*D1803	=-LN(D1804)-(C1804^2)/D1804				
1945	1804	=(A1805-A1804)/A1804	=$F$2+$G$2*C1804^2+$H$2*D1804	=-LN(D1805)-(C1805^2)/D1805				
1871	1805	=(A1806-A1805)/A1805	=$F$2+$G$2*C1805^2+$H$2*D1805	=-LN(D1806)-(C1806^2)/D1806				
1895	1806	=(A1807-A1806)/A1806	=$F$2+$G$2*C1806^2+$H$2*D1806	=-LN(D1807)-(C1807^2)/D1807				
1882	1807	=(A1808-A1807)/A1807	=$F$2+$G$2*C1807^2+$H$2*D1807	=-LN(D1808)-(C1808^2)/D1808				
1891	1808	=(A1809-A1808)/A1808	=$F$2+$G$2*C1808^2+$H$2*D1808	=-LN(D1809)-(C1809^2)/D1809				
1832	1809	=(A1810-A1809)/A1809	=$F$2+$G$2*C1809^2+$H$2*D1809	=-LN(D1810)-(C1810^2)/D1810				
1802	1810	=(A1811-A1810)/A1810	=$F$2+$G$2*C1810^2+$H$2*D1810	=-LN(D1811)-(C1811^2)/D1811				
1708	1811	=(A1812-A1811)/A1811	=$F$2+$G$2*C1811^2+$H$2*D1811	=-LN(D1812)-(C1812^2)/D1812				
1738	1812	=(A1813-A1812)/A1812	=$F$2+$G$2*C1812^2+$H$2*D1812	=-LN(D1813)-(C1813^2)/D1813				
1759	1813	=(A1814-A1813)/A1813	=$F$2+$G$2*C1813^2+$H$2*D1813	=-LN(D1814)-(C1814^2)/D1814				
1742	1814	=(A1815-A1814)/A1814	=$F$2+$G$2*C1814^2+$H$2*D1814	=-LN(D1815)-(C1815^2)/D1815				
1773	1815	=(A1816-A1815)/A1815	=$F$2+$G$2*C1815^2+$H$2*D1815	=-LN(D1816)-(C1816^2)/D1816				
1795	1816	=(A1817-A1816)/A1816	=$F$2+$G$2*C1816^2+$H$2*D1816	=-LN(D1817)-(C1817^2)/D1817				
			MLE	**=SUM(E4:E1817)**				