Lecture 5a: ARCH Models

Big Picture

1. We use ARMA model for the conditional mean
2. We use ARCH model for the conditional variance
3. ARMA and ARCH model can be used together to describe both conditional mean and conditional variance

Price and Return

Let p_t denote the price of a financial asset (such as a stock). Then the return of "buying yesterday and selling today" (assuming no dividend) is

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \approx \log(p_t) - \log(p_{t-1}).$$

The approximation works well when r_t is close to zero.

Continuously Compounded Return

Alternatively, r_t measures the continuously compounded rate

$$
\begin{aligned}
& r_t = \log(p_t) - \log(p_{t-1}) && (1)\\
\Rightarrow \quad & e^{r_t} = \frac{p_t}{p_{t-1}} && (2)\\
\Rightarrow \quad & p_t = e^{r_t} p_{t-1} && (3)\\
\Rightarrow \quad & p_t = \lim_{n\to\infty} \left(1 + \frac{r_t}{n}\right)^n p_{t-1} && (4)
\end{aligned}
$$

Why conditional variance?

1. An asset is risky if its return r_t is volatile (changing a lot over time)

2. In statistics we use variance to measure volatility (dispersion), and so the risk

3. We are more interested in conditional variance, denoted by

$$\texttt{var}(r_t|r_{t-1}, r_{t-2}, \ldots) = E(r_t^2|r_{t-1}, r_{t-2}, \ldots),$$

   because we want to use the past history to forecast the variance. The last equality holds if $E(r_t|r_{t-1}, r_{t-2}, \ldots) = 0$, which is true in most cases.

Volatility Clustering

1. A stylized fact about financial market is "volatility clustering". That is, a volatile period tends to be followed by another volatile period, or volatile periods are usually clustered.
2. Intuitively, the market becomes volatile whenever big news comes, and it may take several periods for the market to fully digest the news
3. Statistically, volatility clustering implies time-varying conditional variance: big volatility (variance) today may lead to big volatility tomorrow.
4. The ARCH process has the property of time-varying conditional variance, and therefore can capture the volatility clustering

ARCH(1) Process

Consider the first order autoregressive conditional heteroskedasticity (ARCH) process

$$\begin{aligned} r_t &= \sigma_t e_t && (5) \\ e_t &\sim \texttt{white noise}(0,1) && (6) \\ \sigma_t &= \sqrt{\omega + \alpha_1 r_{t-1}^2} && (7) \end{aligned}$$

where r_t is the return, and is assumed here to be an ARCH(1) process. e_t is a white noise with zero mean and variance of one. e_t may or may not follow normal distribution.

ARCH(1) Process has zero mean

The conditional mean (given the past) of r_t is

$$\begin{aligned} E(r_t|r_{t-1}, r_{t-2}, \ldots) &= E(\sigma_t e_t|r_{t-1}, r_{t-2}, \ldots) \\ &= \sigma_t E(e_t|r_{t-1}, r_{t-2}, \ldots) \\ &= \sigma_t * 0 = 0 \end{aligned}$$

Then by the law of iterated expectation (LIE), the unconditional mean is

$$E(r_t) = E[E(r_t|r_{t-1}, r_{t-2}, \ldots)] = E[0] = 0$$

So the ARCH(1) process has zero mean.

ARCH(1) process is serially uncorrelated

Using the LIE again we can show

$$\begin{aligned} E(r_t r_{t-1}) &= E[E(r_t r_{t-1}|r_{t-1}, r_{t-2}, \ldots)] \\ &= E[r_{t-1} E(r_t|r_{t-1}, r_{t-2}, \ldots)] \\ &= E[r_{t-1} * 0] = 0 \end{aligned}$$

Therefore the covariance between r_t and r_{t-1} is

$$\mathtt{cov}(r_t, r_{t-1}) = E(r_t r_{t-1}) - E(r_t)E(r_{t-1}) = 0$$

In a similar fashion we can show $\mathtt{cov}(r_t, r_{t-j}) = 0, \forall j \geq 1$

Because of the zero covariance, r_t cannot be predicted using its history $(r_{t-1}, r_{t-2}, \ldots)$. This is the evidence for the efficient market hypothesis (EMH).

However, r_t^2 can be predicted

To see this, note the conditional variance of r_t is given by

$$\begin{aligned} \texttt{var}(r_t|r_{t-1}, r_{t-2}, \ldots) &= E(r_t^2|r_{t-1}, r_{t-2}, \ldots) \\ &= E(\sigma_t^2 e_t^2|r_{t-1}, r_{t-2}, \ldots) \\ &= \sigma_t^2 E(e_t^2|r_{t-1}, r_{t-2}, \ldots) \\ &= \sigma_t^2 * 1 = \sigma_t^2 \end{aligned}$$

So σ_t^2 represents the conditional variance, which by definition is function of history,

$$\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2$$

and so can be predicted by using history r_{t-1}^2.

OLS Estimation

- Note that we have

$$E(r_t^2 | r_{t-1}, r_{t-2}, \ldots) = \omega + \alpha_1 r_{t-1}^2 \tag{8}$$

- This implies that we can estimate ω and α_1 by regressing r_t^2 onto an intercept term and r_{t-1}^2.
- It also implies that r_t^2 follows an AR(1) Process.

Unconditional Variance and Stationarity

- The unconditional variance of r_t is obtained via LIE

$$\begin{aligned}
\texttt{var}(r_t) &= E(r_t^2) - [E(r_t)]^2 = E(r_t^2) && (9)\\
&= E[E(r_t^2|r_{t-1}, r_{t-2}, \ldots)] && (10)\\
&= E[\omega + \alpha_1 r_{t-1}^2] && (11)\\
&= \omega + \alpha_1 E[r_{t-1}^2] && (12)\\
\Rightarrow\ & E(r_t^2) = \frac{\omega}{1-\alpha_1}\ (\texttt{if}\ 0 < \alpha_1 < 1) && (13)
\end{aligned}$$

Along with the zero covariance and zero mean, this proves that the ARCH(1) process is stationary.

Unconditional and Conditional Variances

Let $\sigma^2 = \texttt{var}(r_t)$. We just show

$$\sigma^2 = \frac{\omega}{1-\alpha_1}$$

which implies that

$$\omega = \sigma^2(1-\alpha_1)$$

Plugging this into $\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2$ we have

$$\sigma_t^2 = \sigma^2 + \alpha_1(r_{t-1}^2 - \sigma^2)$$

So conditional variance is a combination of the unconditional variance, and the deviation of squared error from its average value.

ARCH(p) Process

We obtain the ARCH(p) process if r_t^2 follows an AR(p) Process:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2$$

GARCH(1,1) Process

- It is not uncommon that p needs to be very big in order to capture all the serial correlation in r_t^2.
- The generalized ARCH or GARCH model is a parsimonious alternative to an ARCH(p) model. It is given by

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{14}$$

  where the ARCH term is r_{t-1}^2 and the GARCH term is σ_{t-1}^2.
- In general, a GARCH(p,q) model includes p ARCH terms and q GARCH terms.

Stationarity

- The unconditional variance for GARCH(1,1) process is

$$\texttt{var}(r_t) = \frac{\omega}{1 - \alpha - \beta}$$

if the following stationarity condition holds

$$0 < \alpha + \beta < 1$$

- The GARCH(1,1) process is stationary if the stationarity condition holds.

IGARCH effect

- Most often, applying the GARCH(1,1) model to real financial time series will give

$$\alpha + \beta \approx 1$$

- This fact is called integrated-GARCH or IGARCH effect. It means that r_t^2 is very persistent, and is almost like an integrated (or unit root) process

ML Estimation for GARCH(1,1) Model (Optional)

- ARCH model can be estimated by both OLS and ML method, whereas GARCH model has to be estimated by ML method.
- Assuming $e_t \sim \texttt{i.i.d.n}(0, 1)$ and $r_0^2 = \sigma_0^2 = 0$, the likelihood can be obtained in a recursive way:

$$
\begin{aligned}
\sigma_1^2 &= \omega \\
\frac{r_1}{\sigma_1} &\sim N(0,1) \\
\dots &= \dots \\
\sigma_t^2 &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\frac{r_t}{\sigma_t} &\sim N(0,1)
\end{aligned}
$$

- ML method estimates ω, α, β by maximizing the product of all likelihoods.

Warning

- Because the GARCH model requires ML method, you may get highly misleading results when the ML algorithm does not converge.
- Lesson: always check convergence occurs or not.
- You may try different sample or different model specification when there is difficulty of convergence

Heavy-Tailed or Fat-Tailed Distribution

- Another stylized fact is that financial returns typically have "heavy-tailed" or "outlier-prone" distribution (histogram)
- Statistically heavy tail means kurtosis greater than 3
- The ARCH or GARCH model can capture part of the heavy tail
- Even better, we can allow e_t to follow a distribution with tail heavier than the normal distribution, such as Student T distribution with a very small degree of freedom

Asymmetric GARCH

Let 1(.) be the indicator function. Consider a threshold GARCH model

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma r_{t-1}^2 1(r_{t-1} < 0) \tag{15}$$

So the effect of previous return on conditional variance depends on its sign. It is α when r_{t-1} is positive, while $\alpha + \gamma$ when r_{t-1} is negative. We expect $\gamma > 0$ if the respond of the market to bad news (which cause negative return) is more than the good news.

GARCH-in-Mean

- If investors are risk-averse, risky assets will earn higher returns (risk premium) than low-risk assets
- The GARCH-in-Mean model takes this into account:

$$
\begin{aligned}
r_t &= \mu + \delta\sigma_{t-1}^2 + u_t && (16)\\
u_t &\sim \sigma_t e_t && (17)\\
\sigma_t &= \sqrt{\omega + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2} && (18)
\end{aligned}
$$

We expect the risk premium will be captured by a positive δ.

ARMA-GARCH Model

- Finally we can combine the ARMA with the GARCH.
- For instance, consider the AR(1)-GARCH(1,1) combination

$$
\begin{aligned}
r_t &= \phi_0 + \phi_1 r_{t-1} + u_t && (19)\\
u_t &\sim \sigma_t e_t && (20)\\
\sigma_t &= \sqrt{\omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2} && (21)
\end{aligned}
$$

Now we allow the return to be predictable, both in level and in squares.

Lecture 5b: Examples of ARCH Models

Get data

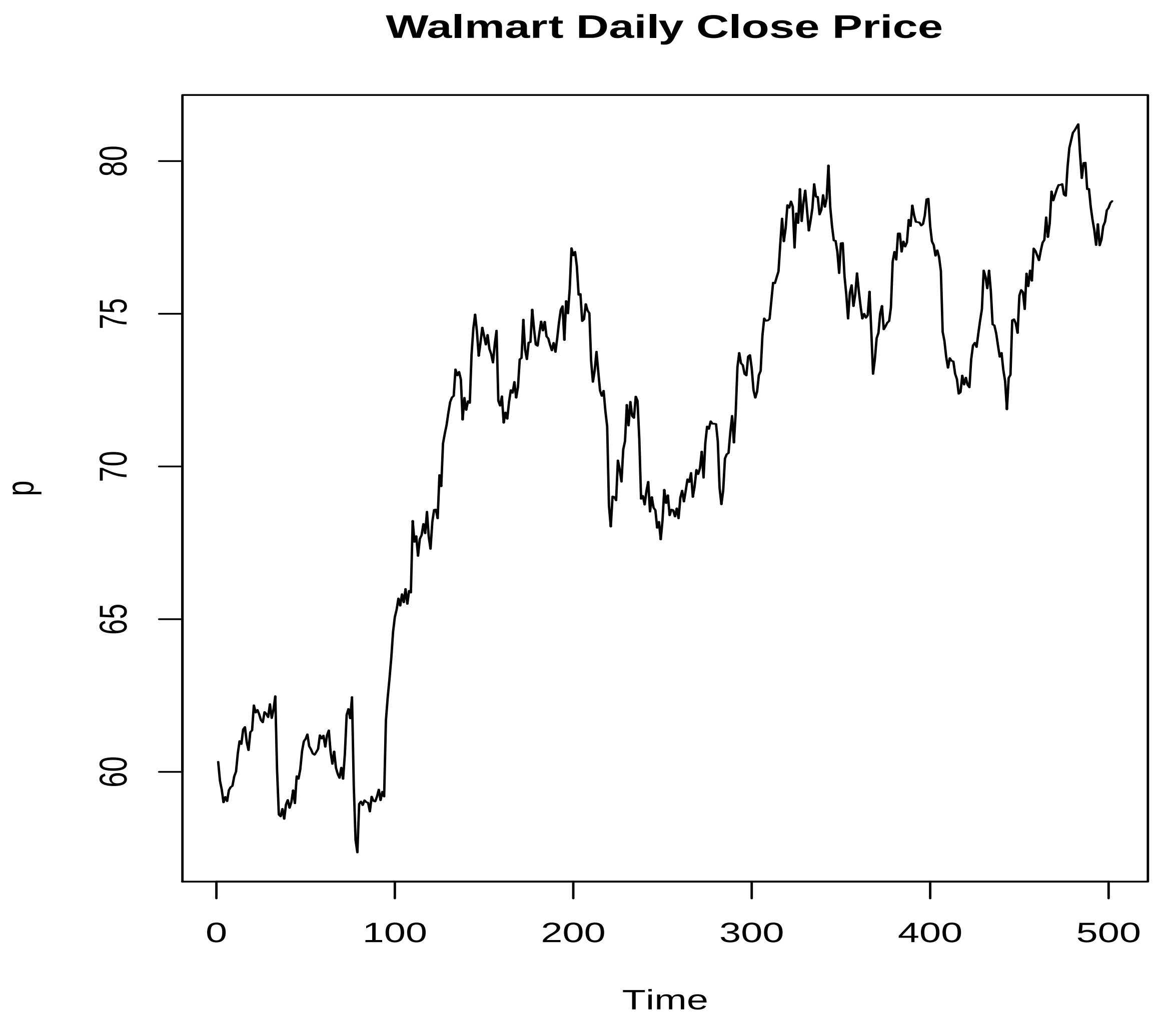
- We download the daily close stock price in year 2012 and 2013 for Walmart (WMT) from Yahoo finance.
- The original data are in Excel format. We can sort the data (so the first observation is the earliest one) and resave it as (tab delimited) txt file
- The first column of the txt file is date; the second column is the daily close price

Generate the return

- We then generate the return by taking log of the price, and take difference of the log price
- We also generate the squared return
- The R commands are

```
p = ts(data[,2])                # price
r = diff(log(p))                # return
r2 = r^2                        # squared return
```

Price



Walmart Daily Close Price
p
60
65
70
75
80
0
100
200
300
400
500
Time

Remarks

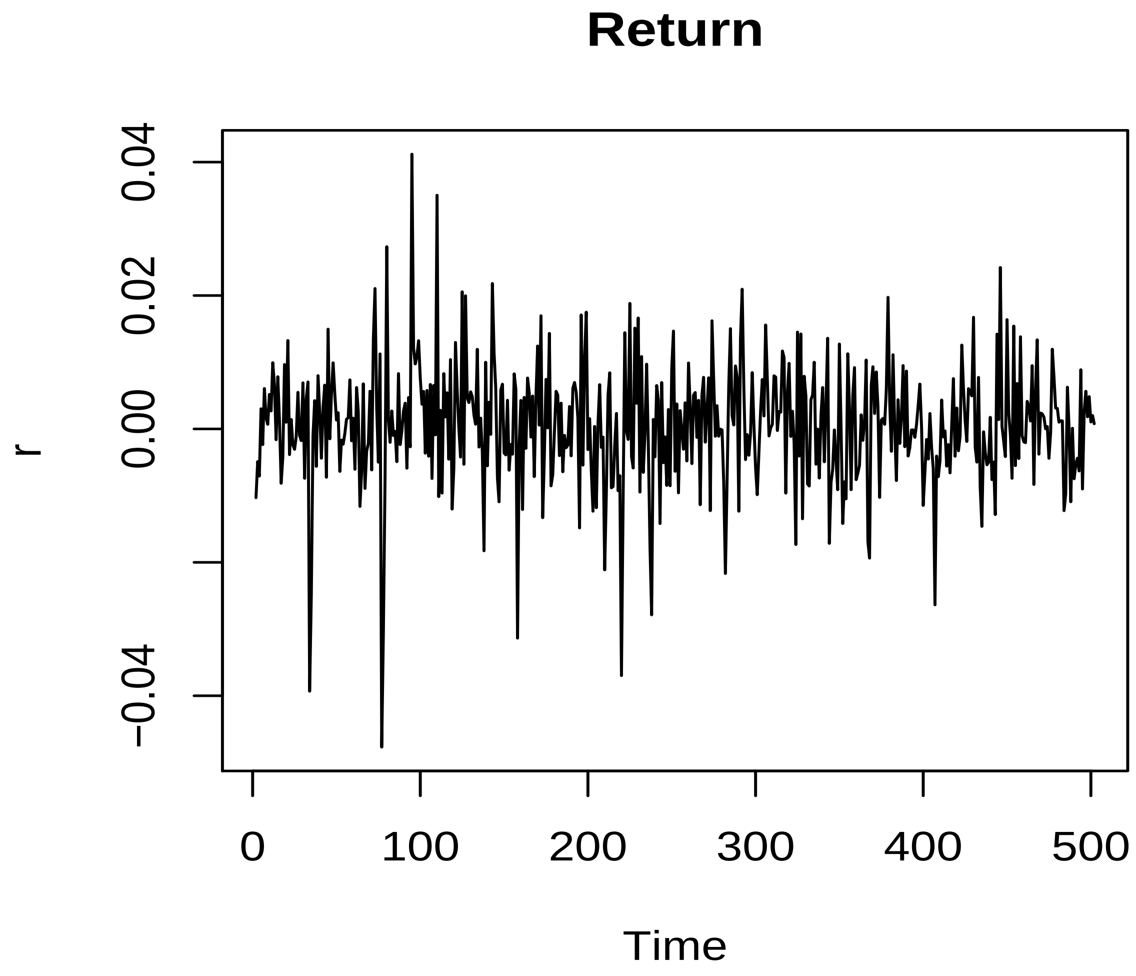
- The WMT stock price is upward-trending in this sample. The trend is a signal for nonstationarity.
- Another signal is the smoothness of the series, which means high persistence. The AR(1) model applied to the price is

```
arima(x = p, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.9964    70.7369
s.e.  0.0034     5.4126
```

  Note that the autoregressive coefficient is 0.9964, very close to one.

Return



Remarks

- One way to achieve stationarity is taking (log) difference. That is also how we obtain the return series
- The return series is not trending. Instead, it seems to be mean-reverting (choppy), which signifies stationarity.
- The sample average for daily return is almost zero

  ```
  mean(r)
  [1] 0.0005303126
  ```

  So on average, you can not make (or lose) money by using the "buying yesterday and selling today" strategy for this stock in this period.

Is return predictable?

- First, the Ljung-Box test indicates that the return is like a white noise, which is serially uncorrelated and unpredictable:

  ```
  Box.test (r, lag = 1, type="Ljung")
          Box-Ljung test
  data:  r
  X-squared = 0.8214, df = 1, p-value = 0.3648
  ```

  Note the p-value is 0.3648, greater than 0.05. So we cannot reject the null that the series is a white noise.

Is return predictable?

- Next, the AR(1) model applied to the price is

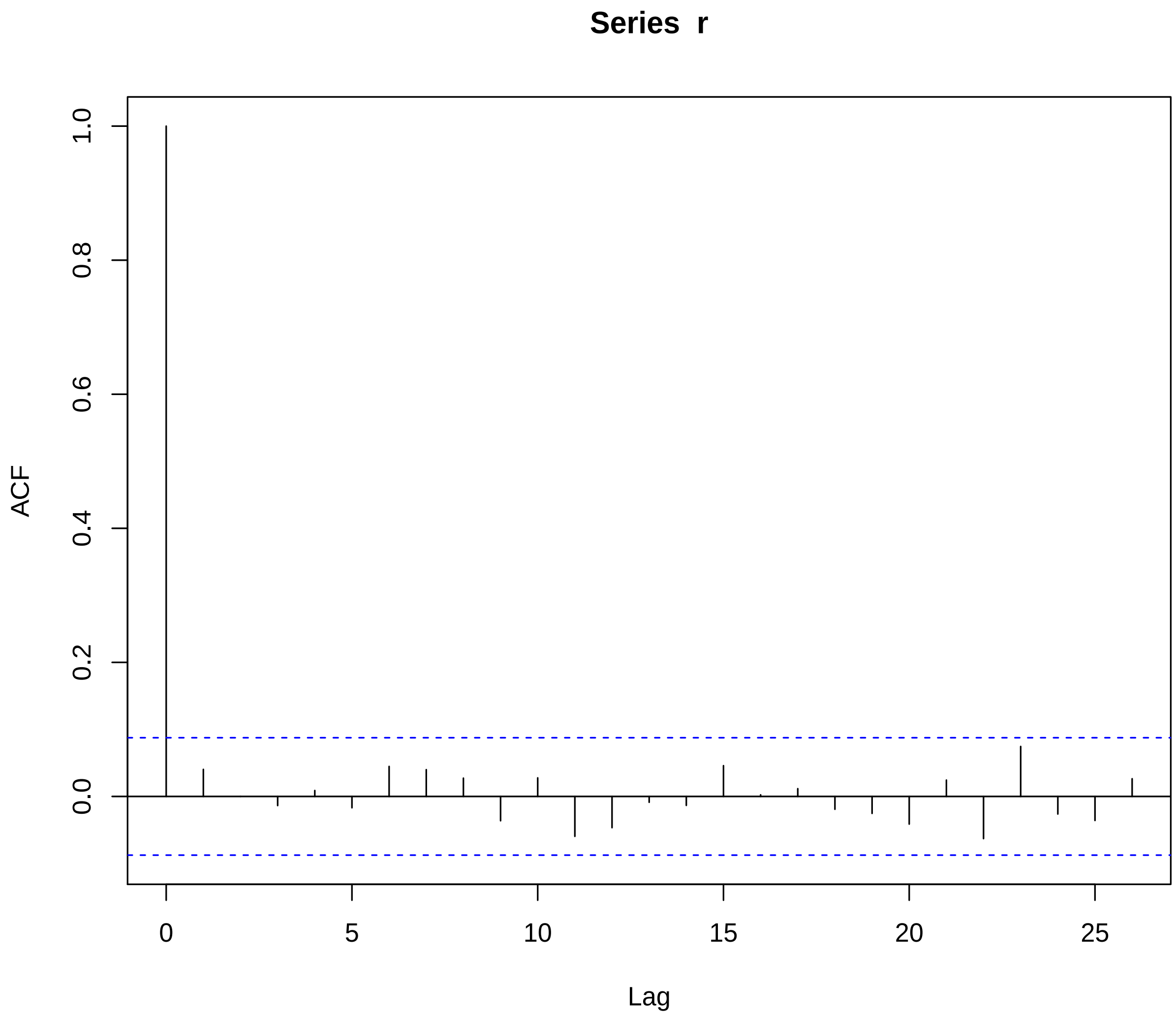
  ```
  arima(x = r, order = c(1, 0, 0))

  Coefficients:
           ar1  intercept
        0.0404      5e-04
  s.e.  0.0447      4e-04
  ```

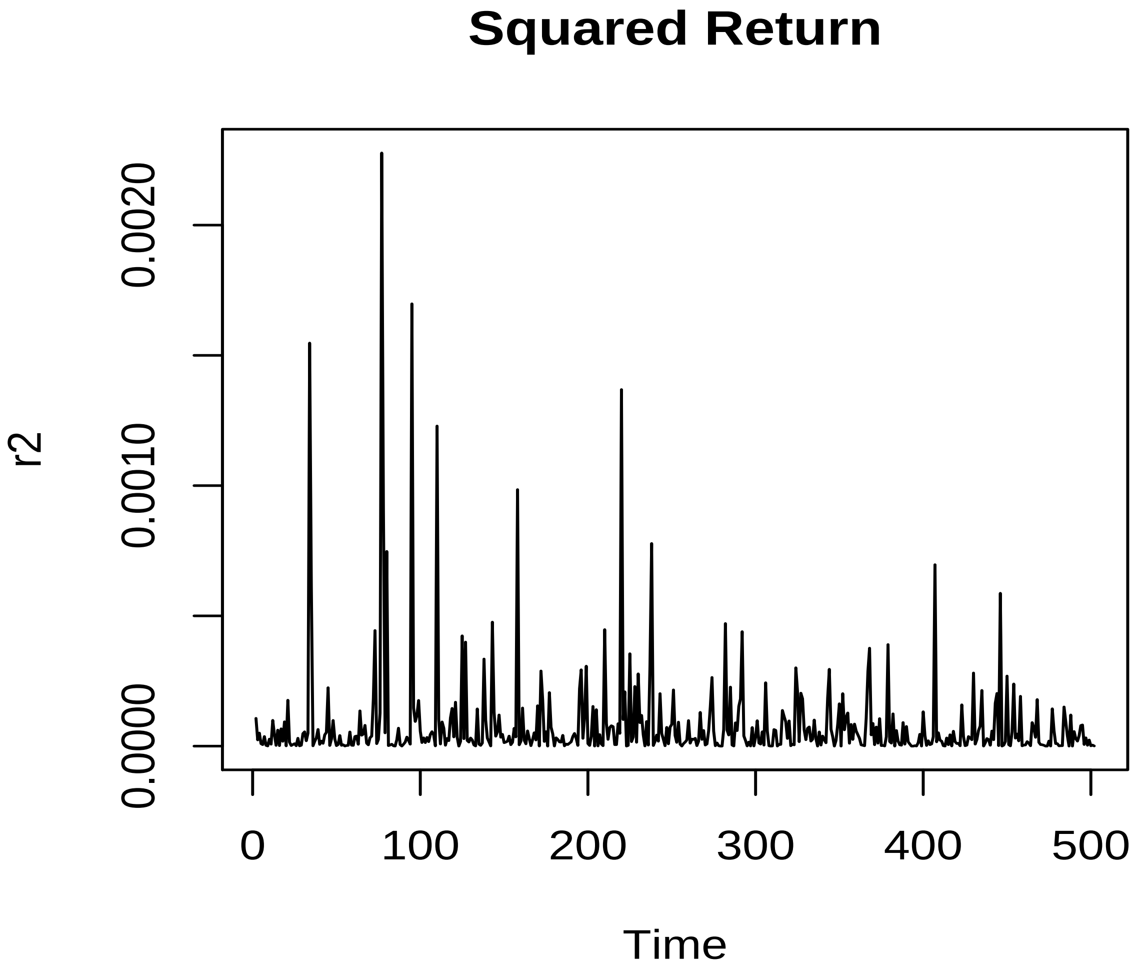
  where both the intercept and autoregressive coefficients are insignificant

- The last evidence for unpredictable return is its ACF function

ACF of return



How about squared return



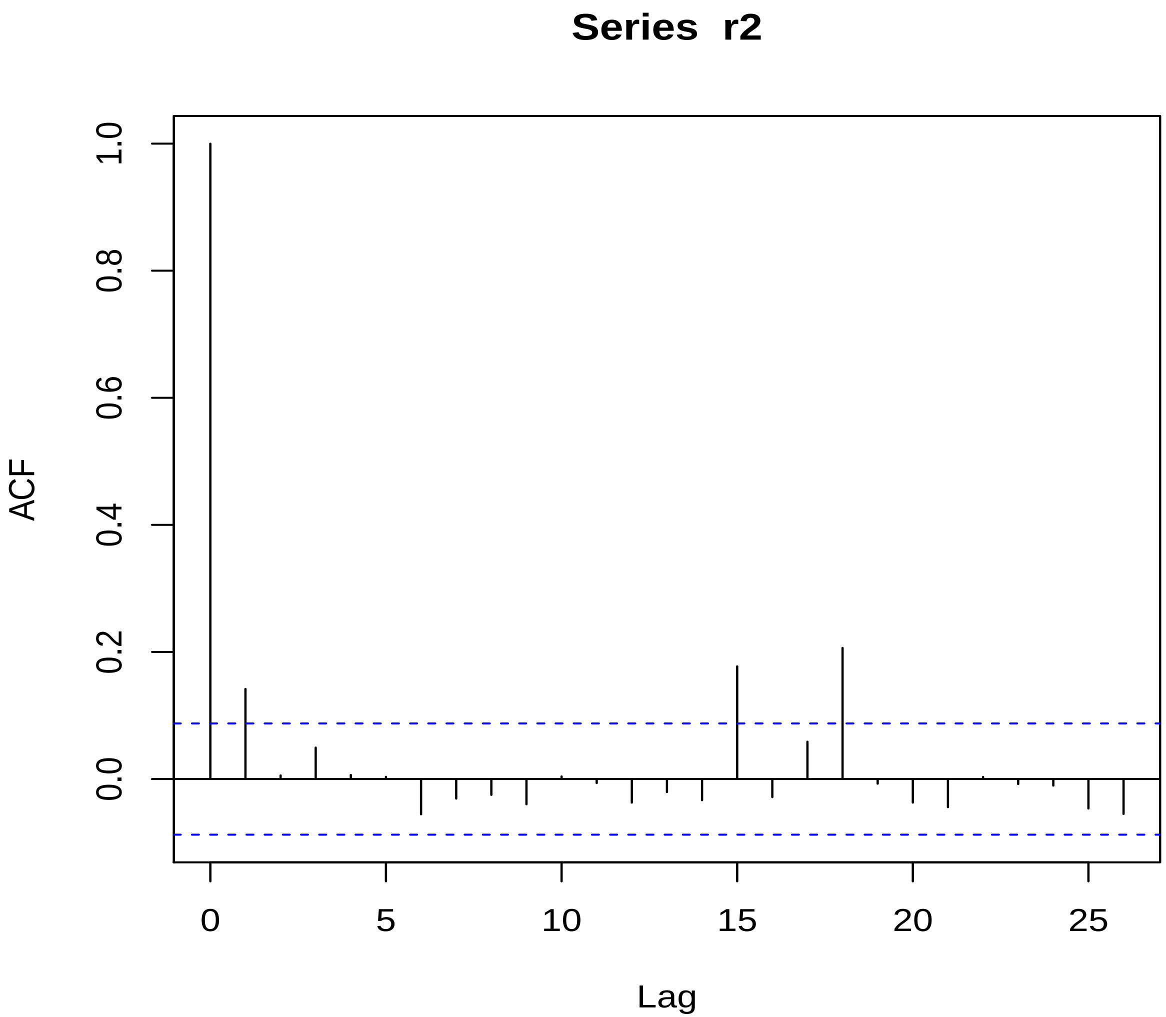
Remarks

- We see that volatile periods are clustered; so volatility in this period will affect next period's volatility.
- The Ljung-Box test applied to squared return is

```
> Box.test (r2, lag = 1, type="Ljung")
        Box-Ljung test
data:  r2
X-squared = 10.1545, df = 1, p-value = 0.001439
```

  Now we can reject the null hypothesis of squared return being white noise at 1% level (the p-value is 0.001439, less than 0.01)

ACF of squared return



ACF of squared return

We can see significant autocorrelation at the first and 15th lags. This is evidence that the squared return is predictable.

ARCH(1) Model: OLS estimation

- We first try OLS estimation of the ARCH(1) model, which essentially regresses r_t^2 onto its first lag

```
> arima(x = r2, order = c(1, 0, 0), method = "CSS")

Coefficients:
         ar1  intercept
      0.1420      1e-04
s.e.  0.0442      0e+00
```

Both the intercept and arch coefficient are significant.

ARCH(1) Model: ML estimation

```
garch(x = r, order = c(0, 1))
Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 7.463e-05   3.799e-06    19.64   <2e-16 ***
a1 9.873e-02   4.592e-02     2.15   0.0315 *
Diagnostic Tests:
        Jarque Bera Test
data:  Residuals
X-squared = 319.4852, df = 2, p-value < 2.2e-16
        Box-Ljung test
data:  Squared.Residuals
X-squared = 0.0416, df = 1, p-value = 0.8383
```

Remarks

- The algorithm converges!
- The ARCH coefficient estimated by ML is 0.09873, close to the OLS estimate 0.1420
- The Jarque Bera Test rejects the null hypothesis that the conditional distribution of the return is normal distribution
- The Box-Ljung test indicates that the ARCH(1) model is dynamically adequate with white noise error.

GARCH(1,1) Model

```
garch(x = r, order = c(1, 1))
Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 5.680e-05   1.553e-05    3.658 0.000254 ***
a1 9.657e-02   4.569e-02    2.113 0.034570 *
b1 2.179e-01   2.044e-01    1.066 0.286403
Diagnostic Tests:
        Jarque Bera Test
data:  Residuals
X-squared = 332.4991, df = 2, p-value < 2.2e-16
        Box-Ljung test
data:  Squared.Residuals
X-squared = 0.0723, df = 1, p-value = 0.7881
```

Remarks

- The algorithm converges!
- The GARCH coefficient is 0.2179, and is insignificant.
- The ARCH coefficient is 0.09657, similar to the ARCH(1) model
- Because $a1 + b1 \ll 1$ the squared return series is stationary (there is no IGARCH effect for WMT stock)
- Overall, we conclude that the return of Walmart stock price follows an ARCH(1) process.