Lecture 12
Heteroscedasticity

• Assumption (A3) is violated in a particular way. \( \varepsilon \) has unequal variances, but \( \varepsilon_i \) and \( \varepsilon_j \) are still not correlated with each other. Some observations (lower variance) are more informative than others (higher variance).
Heteroscedasticity

- Now, we have the CLM regression with hetero-(different) scedastic (variance) disturbances.
- **DGP:** \( y = X \beta + \varepsilon \) is correctly specified.
- \( E[\varepsilon | X] = 0 \)
- \( \text{Var}[\varepsilon_i] = \sigma^2 \omega_i, \ \omega_i > 0. \) (CLM \( \Rightarrow \omega_i = 1, \) for all \( i \))
- \( X \) has full column rank - \( \text{rank}(X) = k \), where \( T \geq k \).

- Popular normalization: \( \sum_{i=1}^{T} \omega_i = 1. \) (A scaling, absorbed into \( \sigma^2 \))

- A characterization of the heteroscedasticity: Well defined estimators and methods for testing hypotheses will be obtainable if the heteroscedasticity is “well behaved” in the sense that
  \[
  \frac{\omega_i}{\sum_{i=1}^{T} \omega_i} \to 0 \text{ as } T \to \infty. \quad \text{i.e., no single observation becomes dominant.}
  \]
  \[
  \frac{1}{T} \sum_{i=1}^{T} \omega_i \to \text{some stable constant.} \quad \text{(Not a plim!)}
  \]

GR Model and Testing

- Implications for conventional OLS and hypothesis testing:
  1. \( b \) is still unbiased.
  2. Consistent? We need the more general proof. Not difficult.
  3. If \( \text{plim } b = \beta \), then \( \text{plim } s^2 = \sigma^2 \) (with the normalization).
  4. Under usual assumptions, we have asymptotic normality.

- Two main problems with OLS estimation under heterocedasticity:
  1. The usual standard errors are not correct. (They are biased!)
  2. OLS is not BLUE.

- Since the standard errors are biased, we cannot use the usual \( t \)-statistics or \( F \)-statistics or \( LM \) statistics for drawing inferences. This is a serious issue.
Heteroscedasticity: Inference Based on OLS

• Q: But, what happens if we still use $s^2(X'X)^{-1}$?
A: It depends on $X'\Omega X - XX'$. If they are nearly the same, the OLS covariance matrix will give OK inferences.

But, when will $X'\Omega X - XX'$ be nearly the same? The answer is based on a property of weighted averages. Suppose $\omega_i$ is randomly drawn from a distribution with $E[\omega_i] = 1$. Then,

$$\frac{1}{T} \sum_{i=1}^{T} \omega_i x_i^2 \overset{p}{\rightarrow} E[x^2] \quad \text{just like} \quad \frac{1}{T} \sum_{i=1}^{T} x_i^2.$$

• Remark: For the heteroscedasticity to be a significant issue for estimation and inference by OLS, the weights must be correlated with $x$ and/or $x_i^2$. The higher correlation, heteroscedasticity becomes more important ($b$ is more inefficient).

Finding Heteroscedasticity

• There are several theoretical reasons why the $\sigma_i^2$ may be related to some explanatory variables $z_1, \ldots, z_j$ and/or $z_i^2, \ldots, z_j^2$.

Examples:
1. Following the error-learning models, as people learn, their errors of behavior become smaller over time. Then, $\sigma_i^2$ is expected to decrease.
2. As data collecting techniques improve, $\sigma_i^2$ is likely to decrease. Companies with sophisticated data processing techniques are likely to commit fewer errors in forecasting customer's orders.
3. As companies grow, companies expand and tend to be more diversified and, thus, safer. Hence, $\sigma_i^2$ is likely to decrease with size.
4. Companies with larger profits tend to have greater variability in their dividend/buyback policies than companies with lower profits.
Finding Heteroscedasticity

• Heteroscedasticity can also be the result of model misspecification.

• It can arise as a result of the presence of outliers (either very small or very large). The inclusion/exclusion of an outlier, especially if \( T \) is small, can affect the results of regressions.

• Violations of (A1) — model is correctly specified — can produce heteroscedasticity, due to omitted variables from the model.

• Skewness in the distribution of one or more regressors included in the model can induce heteroscedasticity. Examples are economic variables such as income, wealth, and education.

• David Hendry notes that heteroscedasticity can also arise because of
  - (1) incorrect data transformation (e.g., ratio or first difference).
  - (2) incorrect functional form (e.g., linear vs log-linear models).

Finding Heteroscedasticity

• Heteroscedasticity is usually modeled using one of the following specifications:
  - \( H_1: \sigma_t^2 \) is a function of past \( \varepsilon_t^2 \) and past \( \sigma_t^2 \) (ARCH models).
  - \( H_2: \sigma_t^2 \) increases monotonically with one (or several) exogenous variable(s) \( (z_1, \ldots, z_j) \).
  - \( H_3: \sigma_t^2 \) decreases monotonically with \( \text{Size} \) (Market Cap).
  - \( H_4: \sigma_t^2 \) is the same within \( p \) subsets of the data but differs across the subsets (grouped heteroscedasticity). This specification allows for structural breaks.

• These are the usual alternatives hypothesis (\( H_1 \)) in the heteroscedasticity tests.
Finding Heteroscedasticity

- **Visual test**
  In a plot of residuals against dependent variable or other variable will often produce a fan shape.

![Graph showing visual test for heteroscedasticity](image)

Testing for Heteroscedasticity

- We can use OLS along the White estimator for consistent inferences. Then, why do we want to test for heteroscedasticity?

- The GLS/FGSL estimator has a lower asymptotic variance.

- We want to test: $H_0: \mathbb{E}(\varepsilon_i^2 | x_1, x_2, \ldots, x_k) = \sigma^2$

- $H_1$ and the structure of the test depend on what we consider the drivers of $\sigma_i^2$ — i.e., in the previous examples: $H_1, H_2, H_3, H_4$, etc.

- The key is whether $\mathbb{E}[\varepsilon_i^2] = \sigma_i^2$ is related to $X$ and/or $X^2$. Suppose we suspect a particular independent variable, say $x_j$, is driving $\sigma_i^2$:
  $$\sigma_i^2 = f(x_j)$$

- Then, a simple test: Check the RSS for large values of $x_j$, and the RSS for small values of $x_j$. This is the Goldfeld-Quandt (GQ) test.
The Goldfeld-Quandt test

GQ tests $H_0: \sigma_i^2 = \sigma^2$

$H_1: \sigma_i^2 = f(x_j)$

Easy to compute:

- **Step 1.** Arrange the data from small to large values of the independent variable suspected of causing heteroscedasticity, $x_j$.

- **Step 2.** Run two separate regressions, one for small values of $x_j$ and one for large values of $x_j$, omitting $d$ middle observations ($d \approx 20\%$). Get the RSS for each regression: $RSS_1$ for small values of $x_j$ and $RSS_2$ for large $x_j$'s.

- **Step 3.** Calculate the F ratio

\[
GQ = \frac{RSS_2}{RSS_1} \sim F_{df, df}, \text{ with } df = [(T - d) - 2(k + 1) ]/2 \quad (A5 \text{ holds})
\]

Testing for Heteroscedasticity

- **The Goldfeld-Quandt test**

If (A5) does not hold, the F distribution becomes an approximation. Other tests may be preferred.

**Note:** When we suspect more than one variable is driving the $\omega_i$'s, this test is not very useful.

- But, the GQ test is a popular test for structural breaks (two regimes) in variance. For these tests, we rewrite step 3 to allow for a different sample size in the sub-samples 1 and 2.

- **Step 3.** Calculate the F-test ratio

\[
GQ = [RSS_2/ (T_2 - k)]/[RSS_1/ (T_1 - k)]
\]
Testing for Heteroscedasticity: GQ Test

**Example:** We test if the 3-factor FF model for IBM and GE returns shows heteroscedasticity with a GQ test, using `gqtest` in package `lmtest`.

- **IBM returns**

```r
> library(lmtest)
> gqtest(ibm_x ~ Mkt_RF + SMB + HML, fraction = .20)

Goldfeld-Quandt test
data:  ibm_x ~ Mkt_RF + SMB + HML
GQ = 1.1006, df1 = 224, df2 = 223, p-value = 0.2371
alternative hypothesis: variance increases from segment 1 to 2
```

- **GE returns**

```r
gqtest(ge_x ~ Mkt_RF + SMB + HML, fraction = .20)

Goldfeld-Quandt test
data:  ge_x ~ Mkt_RF + SMB + HML
GQ = 2.744, df1 = 281, df2 = 281, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2
```

Testing for Heteroscedasticity: LR Test

- **The Likelihood Ratio Test**

Let’s define the likelihood function, assuming normality, for a general case, where we have $g$ different variances:

$$
\ln L = -\frac{T}{2} \ln (2\pi) - \frac{1}{2} \sum_{i=1}^{g} T_i \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{g} \frac{1}{\sigma_i^2} (y_i - \bar{X}_i \beta)'(y_i - \bar{X}_i \beta)
$$

We have two models:

- **(R)** Restricted under $H_0$: $\sigma_i^2 = \sigma^2$. From this model, we calculate $\ln L$

  $$
  \ln L_R = -\frac{T}{2} \left[ \ln (2\pi) + 1 \right] - \frac{T}{2} \ln (\hat{\sigma}^2)
  $$

- **(U)** Unrestricted. From this model, we calculate the log likelihood.

  $$
  \ln L_U = -\frac{T}{2} \left[ \ln (2\pi) + 1 \right] - \sum_{i=1}^{g} \frac{T_i}{2} \ln \hat{\sigma}_i^2; \quad \hat{\sigma}_i^2 = \frac{1}{T_i} (y_i - \bar{X}_i \beta)'(y_i - \bar{X}_i \beta)
  $$
Testing for Heteroscedasticity: LR Test

- Now, we can estimate the Likelihood Ratio (LR) test:

\[
LR = 2(\ln L_0 - \ln L_R) = T \ln \hat{\sigma}^2 - \sum_{i=1}^g T_i \ln \hat{\sigma}_i^2 \rightarrow \chi^2_{g-1}
\]

Under the usual regularity conditions, LR is approximated by a \( \chi^2_{g-1} \).

- Using specific functions for \( \sigma_i^2 \), this test has been used by Rutemiller and Bowers (1968) and in Harvey's (1976) groupwise heteroscedasticity paper.

Testing for Heteroscedasticity

- **Score LM tests**

  - We want to develop tests of \( H_0: \mathbb{E}(\epsilon_i^2 | x_1, x_2, \ldots, x_k) = \sigma^2 \) against an \( H_1 \) with a general functional form.

  - Recall the central issue is whether \( \mathbb{E}[\epsilon_i^2] = \sigma_i^2 \) is related to \( X \) and/or \( X^2 \). Then, a simple strategy is to use OLS residuals to estimate disturbances and look for relationships between \( \epsilon_i^2 \) and \( x_i \) and/or \( x_i^2 \).

  - Suppose that the relationship between \( \epsilon^2 \) and \( X \) is linear:

    \[
    \epsilon_i^2 = x_i' \alpha + v_i
    \]

    Then, we test: \( H_0: \alpha = 0 \) against \( H_1: \alpha \neq 0 \).

  - We base the test on how the squared OLS residuals, \( e \), correlate with \( X \).
Testing for Heteroscedasticity

- Popular heteroscedasticity LM tests:
  - Breusch and Pagan (1979)'s LM test (BP).
  - White (1980)'s general test.

- Both tests are based on OLS residuals. That is, calculated under H0: No heteroscedasticity.

- The BP test is an LM test, based on the score of the log likelihood function, calculated under normality. It is a general test designed to detect any linear forms of heteroskedasticity.

- The White test is an asymptotic Wald-type test, normality is not needed. It allows for nonlinearities by using squares and crossproducts of all the x's in the auxiliary regression.

Testing for Heteroscedasticity: BP Test

- Let's start with a general form of heteroscedasticity:
  \[ h(z_0 + z_1 \alpha_1 + z_2 \alpha_2 + \ldots + z_m \alpha_m) = \sigma_i^2 \]

- We want to test: \( H_0: E(\varepsilon_i^2 | z_1, z_2, \ldots, z_m) = h(z_1 \alpha) = \sigma^2 \)
  or \( H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0 \) (m restrictions)

- Assume normality. That is, the log likelihood function is:
  \[ \log L = \text{constant} + \frac{1}{2} \sum \log \sigma_i^2 - \frac{1}{2} \sum \varepsilon_i^2 / \sigma_i^2 \]

Then, construct an LM test:

\[ LM = S(\theta)^\prime I(\theta)^{-1} S(\theta) \quad \theta = (\beta, \alpha) \]

\[ S(\theta) = \partial \log L / \partial \theta = [-\Sigma \sigma_i^{-2} \varepsilon_i^2 - \frac{1}{2} \sum \partial h / \partial \alpha] z_1 \sigma_1^{-2} + \frac{1}{2} \sum \sigma_1 \varepsilon_i^2 \partial h / \partial x z_1 \]

\[ I(\theta) = E[\partial^2 \log L / \partial \theta \partial \theta^\prime] \]

- We have block diagonality, we can rewrite the LM test, under \( H_0: \)
  \[ LM = S(\alpha_0, 0)^\prime [I_{22} - I_{24} I_{44} I_{42}]^{-1} S(\alpha_0, 0) \]
Testing for Heteroscedasticity: BP Test

• We have block diagonality, we can rewrite the LM test, under $H_0$:

$$LM = S(\alpha_0)' [I_{22} - I_{21} I_{11}^{-1} I_{21}] S(\alpha_0)$$

$$S(\alpha_0,0) = -\frac{1}{2} \sum_i (\partial h/\partial \alpha | \alpha_{0,R,0})' \frac{\sigma_R^{-2}}{} \sum_i e_i^2 (\partial h/\partial \alpha | \alpha_{0,R,0})' z_i$$
$$= \frac{1}{2} \sum_i \frac{e_i^2}{\sigma_R^{-2}} (\partial h/\partial \alpha | \alpha_{0,R,0})' \sum_i z_i e_i^2 / \sigma_R^2$$

$$I_{22}(\alpha_0,0) = E[\partial^2 \log L/\partial \alpha \partial \alpha'] = \frac{1}{2} [\sigma_R^{-2} (\partial h/\partial \alpha | \alpha_{0,R,0})' \sum_i z_i z_i']$$

$$\sigma_R^2 = (1/T) \sum_i e_i^2 \quad (\text{MLE of } \sigma \text{ under } H_0).$$

Then,

$$LM = \frac{1}{2} (\sum_i z_i w_i)' [\sum_i z_i z_i']^{-1} (\sum_i z_i w_i)$$

$$= \frac{1}{2} W'Z (Z'Z)^{-1} Z W \sim \chi_m^2$$

Note: Recall $R^2 = [y'X (X'X)^{-1} X'y - T \bar{y}^2]/[y'y - T \bar{y}^2] = \text{ESS/TSS}$

Testing for Heteroscedasticity: BP Test

• Since under $H_0$: $E[w_i] = 0$, $E[e_i^2] = 1$

$$LM = \frac{1}{2} W'Z (Z'Z)^{-1} Z W = \frac{1}{2} \text{ESS}$$

$\text{ESS} = \text{Explained SS in regression of } w_i (= \frac{e_i^2}{\sigma_R^2 - 1}) \text{ against } z_i.$

• Under the usual regularity conditions, and under $H_0$,

$$\sqrt{T} (\alpha_{ML} - \alpha) \overset{d}{\rightarrow} N(0, 2 \sigma^4 (Z'Z/T)^{-1})$$

Then,

$$\text{LM-BF} = (2 \sigma_R^4)^{-1} \text{ESS}_e \overset{d}{\rightarrow} \chi_m^2.$$

$\text{ESS}_e = \text{ESS in regression of } e_i^2 \text{ against } z_i.$

Since $\sigma_R^4 \overset{p}{\rightarrow} \sigma^4 \quad \Rightarrow \quad \text{LM-BF} \overset{d}{\rightarrow} \chi_n^2$
Testing for Heteroscedasticity: BP Test

Note: Recall $R^2 = [y'X(X'X)^{-1}X'y - T\bar{y}^2]/[y'y - T\bar{y}^2]$.

Under $H_0: E[\omega_i]=0$, $E[\omega_i^2]=1$, the LM test is equivalent to a $TR^2$.
(Think of $\bar{y} = 0$ & $y'y/T=1$ above).

• Variations:
  (1) Glesjer (1969) test. Use absolute values instead of $e_i^2$ to estimate the varying second moment. Following our previous example,
  $$|e_i| = \omega_0 + \omega_1 z_{i,1} + \omega_2 z_{i,2} + ... + \omega_m z_{i,m} + v_i$$

  (2) Harvey-Godfrey (1978) test. Use $\ln(e_i^2)$. Then, the implied model for $\sigma_i^2$ is an exponential model.
  $$\ln(e_i^2) = \omega_0 + \omega_1 z_{i,1} + \omega_2 z_{i,2} + ... + \omega_m z_{i,m} + v_i$$

  Note: Implied model for $\sigma_i^2 = \exp\{\omega_0 + \omega_1 z_{i,1} + ... + \omega_m z_{i,m} + v_i\}$.

Testing for Heteroscedasticity: BP Test

• Variations:
  (3) Koenker's (1981) studentized LM test. A usual problem with statistic LM is that it crucially depends on the assumption that $\varepsilon$ is normal. Koenker (1981) proposed studentizing the statistic LM-BP by
  $$LM-S = (2 \sigma_R^2) LM-BP / [(\sum_{i=1}^T (e_i^2 - \sigma_R^2))^2 / T] \xrightarrow{d} \chi^2_m$$

  The studentized version of the test is asymptotically equivalent to a $T^*R^2$ test, where $R^2$ is calculated from a regression of $e_i^2 / \sigma_R^2$ on the variables $Z$. (Omitting $\sigma_R^2$ from the denominator is OK.)
Testing for Heteroscedasticity: BP Test

- **Step 1.** Run OLS on DGP:
  \[ y = X \beta + \varepsilon. \]
  - Keep \( \varepsilon \) and compute \( \sigma_R^2 = \frac{RSS}{T} \)

- **Step 2.** (Auxiliary Regression). Run the regression of \( \varepsilon^2 \) on the \( m \) explanatory variables, \( z \). In our example,
  \[ \frac{\varepsilon_i^2}{\sigma_R^2} = \alpha_0 + \alpha_1 z_{1,i} + \ldots + \alpha_m z_{m,i} + \nu_i \]
  - Keep \( R^2 \)

- **Step 3.** Use the \( R^2 \) from Step 2. Let's call it \( R_{e2}^2 \). Calculate
  \[ LM = T \cdot R_{e2}^2 \rightarrow \chi_m^2 \]

Testing for Heteroscedasticity: Example – IBM

**Example:** We suspect that squared Mkt_RF (x1) – a measure of the overall market’s variance – drives heteroscedasticity. We do a studentized LM-BP test for IBM in the 3-factor FF model:

```r
fit_ibm_ff3 <- lm (ibm_x ~ Mkt_RF + SMB + HML) # Step 1 – OLS in DGP (3-factor FF model)
e_ibm <- fit_ibm_ff3$residuals # Step 1 – keep residuals
e_ibm2 <- e_ibm^2 # Step 1 – squared residuals
Mkt_RF2 <- Mkt_RF^2
fit <- lm (e_ibm2 ~ Mkt_RF2) # Step 2 – Auxiliary regression
Re_2 <- summary(fit)$r.squared # Step 2 – keep R^2
LM_BP_test <- Re2 * T # Step 3 – Compute LM-BP test: R^2 * T
[1] 0.25038
p_val <- 1 - pchisq(LM_BP_test, df = 1) # p-value of LM_test
[1] 0.6168019
LM-BP Test: 0.25028 \( \Rightarrow \) cannot reject \( H_0 \) at 5% level (\( \chi^2_{1,.05} \approx 3.84 \)), with a \( p\text{-value} = .6168 \).
Testing for Heteroscedasticity: Example – IBM

**Example (continuation):** The `bptest` in the `lmtest` package performs a studentized LM-BP test for the same variables used in the model (Mkt, SMB and HML). For IBM in the 3-factor FF model:

```r
> bptest(ibm_x ~ Mkt_RF + SMB + HML) #bptest only allows to test H_0: \sigma^2 = \sigma(x,model variables)

studentized Breusch-Pagan test

data:  ibm_x ~ Mkt_RF + SMB + HML
BP = 4.1385, df = 3, p-value = 0.2469

LM-BP Test: 4.1385 \Rightarrow cannot reject H_0 at 5% level (\chi^2_{[3],0.05} \approx 7.815); with a p-value = 0.2469.
```

**Note:** Heteroscedasticity in financial time series is very common. In general, it is driven by squared market returns or squared past errors.

---

Testing for Heteroscedasticity: Example – DIS

**Example:** We suspect that squared Market returns drive heteroscedasticity. We do an LM-BP (studentized) test for Disney:

```r
lr_dis <- log(x_dis[-1]/x_dis[-T]) # Log returns for DIS
dis_x <- lr_dis – RF # Disney excess returns
fit_dis_ff3 <- lm (dis_x ~ Mkt_RF + SMB + HML) # Step 1 – OLS in DGP (3-factor FF model)
e_dis <- fit_dis_ff3$residuals # Step 1 – keep residuals
e_dis2 <- e_dis^2 # Step 2 – squared residuals
fit <- lm (e_dis2 ~ Mkt_RF2) # Step 2 – Auxiliary regression
Re_e2 <- summary(fit)$r.squared # Step 2 – Keep R^2 from Auxiliary regression
LM_BP_test <- Re_e2 * T # Step 3 – Compute LM Test: R^2 * T
> LM_BP_test
[1] 14.15224
> p_val <- 1 - pchisq(LM_BP_test, df = 1)  # p-value of LM_test
> p_val
[1] 0.0001685967

LM-BP Test: 14.15 \Rightarrow reject H_0 at 5% level (\chi^2_{[1],0.05} \approx 3.84), with a p-value = .0001.
```
Testing for Heteroscedasticity: Example – DIS

Example (continuation): We do the same test, but with SMB squared for Disney:

\[ \text{SMB}^2 \leftarrow \text{SMB}^2 \]
\[ \text{fit} \leftarrow \text{lm}(\text{e_dis}^2 \sim \text{SMB}^2) \]
\[ \text{Re}_2 \leftarrow \text{summary(fit)} \text{r.squared} \]
\[ \text{LM_BP_test} \leftarrow \text{Re}_2 \times T \]

> \text{LM_BP_test}

[1] 7.564692

> \text{p_val} \leftarrow 1 - \text{pchisq(LM_BP_test, df = 1)} \quad \# \text{p-value of LM_test}

> \text{p_val}

[1] 0.005952284

LM-BP Test: 7.56 \implies \text{reject H}_0 \text{ at 5\% level (} \chi^2_{0.05} \approx 3.84 \text{), with a p-value= .006.}
Testing for Heteroscedasticity: White Test

- **White heteroscedasticity test:**
  
  \[
  W = \left( \frac{1}{T} \sum_{i=1}^{T} (e_i^2 - s_i^2) \right) \psi_i' D_T^{-1} \left( \frac{1}{T} \sum_{i=1}^{T} (e_i^2 - s_i^2) \right) \psi_i' \xrightarrow{d} \chi_m^2
  \]

  where
  
  \[
  D_T = \text{Var} \left( \frac{1}{T} \sum_{i=1}^{T} (e_i^2 - s_i^2) \right) \psi_i'
  \]

  **Note:** \( W \) is asymptotically equivalent to a \( TR^2 \) test, where \( R^2 \) is calculated from a regression of \( e_i^2 / \sigma^2 \) on the \( \psi_i \)’s.

Testing for Heteroscedasticity: White Test

- **Usual calculation of the White test**
  - **Step 1.** Run OLS on DGP:
    
    \[
    y = X \beta + \varepsilon \quad \text{--Keep residuals, } e_i.
    \]

  - **Step 2.** (Auxiliary Regression). Regress \( e^2 \) on all the explanatory variables \( (x_j) \), their squares \( (x_j^2) \), & all their cross products \( (x_j * x_l) \).

  For example, when the model contains \( k = 2 \) explanatory variables, the test is based on:

  \[
  e_i^2 = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i}^2 + \beta_4 x_{2,i}^2 + \beta_5 x_{1,i} x_{2,i} + v_i
  \]

  Let \( m \) be the number of regressors in auxiliary regression (in the above example, \( m = 5 \)). Keep \( R^2 \), say \( R^2_{e_2} \).

  - **Step 3.** Compute the LM statistic:
    
    \[
    \text{LM} = T * R^2_{e_2} \xrightarrow{d} \chi_m^2.
    \]
Testing for Heteroscedasticity: White Test

**Example:** White Test for 3-factor FF model residuals for IBM:

\[
\begin{align*}
\text{HML}^2 & \leftarrow \text{HML}^2; \\
\text{Mkt} \times \text{HML} & \leftarrow \text{Mkt} \times \text{HML} \\
\text{Mkt} \times \text{SMB} & \leftarrow \text{Mkt} \times \text{SMB} \\
\text{SMB} \times \text{HML} & \leftarrow \text{SMB} \times \text{HML} \\
xx2 & \leftarrow \text{cbind(Mkt} \times \text{RF, SMB2, HML2, Mkt} \times \text{HML, Mkt} \times \text{SMB, SMB} \times \text{HML)} \\
\text{fit} \_ \text{ibm} \_ \text{W} & \leftarrow \text{lm(e} \_ \text{ibm}^2 \sim \text{xx2)} & \# \text{Not including original variables OK} \\
r2 \_ c2 & \leftarrow \text{summary(fit} \_ \text{ibm} \_ \text{W}$\text{r.squared} & \# \text{Keep R}^2 \text{from Auxiliary regression} \\
> r2 \_ c2 \\
[1] \ 0.0165492 \\
\text{lm} \_ t & \leftarrow \text{T} \times r2 \_ c2 & \# \text{Compute LM test: R}^2 \times \text{sample size (T)} \\
> \text{lm} \_ t \\
[1] \ 10.93483 \\
\text{df} \_ \text{lm} & \leftarrow \text{ncol(xx2)} \\
\text{qchisq(0.95, df} = \text{df} \_ \text{lm)} \\
\text{LM-White Test: 10.93} \Rightarrow \text{cannot reject H}_0 \ \text{at 5\% level (}\chi^2_{[6],.05} \approx 12.59\text{).}
\end{align*}
\]

Example (continuation): Now, we do a White Test for the 3 factor F-F model for DIS and GE returns.

- For DIS, we get:

\[
\begin{align*}
\text{fit} \_ \text{dis} \_ \text{W} & \leftarrow \text{lm(e} \_ \text{dis}^2 \sim \text{xx2)} \\
\text{Re}_2 \_ \text{W} & \leftarrow \text{summary(fit} \_ \text{dis} \_ \text{W}$\text{r.squared} \\
\text{LM}_\text{W-test} & \leftarrow \text{Re}_2 \_ \text{W} \times \text{T} \\
> \text{LM}_\text{W-test} \\
[1] \ 25.00148 \\
> \text{qchisq(0.95, df} = \text{df} \_ \text{lm)} \\
[1] \ 12.59159 \\
> p \_ \text{val} & \leftarrow 1 - \text{pchisq(LM}_\text{W-test, df} = 6) & \# \text{p-value of LM_test} \\
> p \_ \text{val} \\
[1] \ 0.0003412389 \\
\text{LM-White Test: 20.15} (p\,-value=0.0026) \Rightarrow \text{reject H}_0 \ \text{at 5\% level.}
\end{align*}
\]

- For GE, we get:

\[
\begin{align*}
\text{LM-White Test: 20.15} (p\,-value=0.0026) \Rightarrow \text{reject H}_0 \ \text{at 5\% level.}
\end{align*}
\]
Testing for Heteroscedasticity: White Test

Example: We do a White Test for the residuals in the encompassing (IFE + PPP) model for changes in the USD/GBP ($T=363$):

```r
fit_gbp <- lm(lr_usdgbp ~ inf_dif + int_dif)
e_gbp <- fit_gbp$residuals
e_gbp2 <- e_gbp^2
int_dif2 <- int_dif^2
inf_dif2 <- inf_dif^2
int_inf_dif <- int_dif*inf_dif

fit_W <- lm(e_gbp2 ~ int_dif2 + inf_dif2 + int_inf_dif)
Re_e2W <- summary(fit_W)$r.squared
LM_W_test <- Re_e2W * T
p_val <- 1 - pchisq(LM_W_test, df = 3)  # p-value of LM_test

> LM_W_test
[1] 15.46692
> p_val
[1] 0.001458139  ⇒ reject H_0 at 5% level
```

Testing for Heteroscedasticity: Remarks

• Drawbacks of the Breusch-Pagan test:
  - It has been shown to be sensitive to violations of the normality assumption.
  - Three other popular LM tests: the Glejser test, the Harvey-Godfrey test, and the Park test, are also sensitive to such violations.

• Drawbacks of the White test
  - If a model has several regressors, the test can consume a lot of df's.
  - In cases where the White test statistic is statistically significant, heteroscedasticity may not necessarily be the cause, but model specification errors.
  - It is general. It does not give us a clue about how to model heteroscedasticity to do FGLS. The BF test points us in a direction.
Testing for Heteroscedasticity: Remarks

- Drawbacks of the White test (continuation)

- In simulations, it does not perform well relative to others, especially, for time-varying heteroscedasticity, typical of financial time series.

- The White test does not depend on normality; but the Koenker's test is also not very sensitive to normality. In simulations, Koenker's test seems to have more power—see, Lyon and Tsai (1996) for a Monte Carlo study of the heteroscedasticity tests presented here.

Testing for Heteroscedasticity: Remarks

- General problems with heteroscedasticity tests:

  - The tests rely on the first four assumptions of the CLM being true.

  - In particular, (A2) violations. That is, if the zero conditional mean assumption, then a test for heteroskedasticity may reject the null hypothesis even if \( \text{Var}(y|X) \) is constant.

  - This is true if our functional form is specified incorrectly (omitted variables or specifying a log instead of a level). Recall David Hendry's comment.

- Knowing the true source (functional form) of heteroscedasticity may be difficult. A practical solution is to avoid modeling heteroscedasticity altogether and use OLS along the White heteroskedasticity-robust standard errors.
Estimation: WLS form of GLS

• While it is always possible to estimate robust standard errors for OLS estimates, if we know the specific form of the heteroskedasticity, we can obtain more efficient estimates than OLS: GLS.

• GLS basic idea: Efficient estimation through the transform the model into one that has homoskedastic errors – called WLS.

• Suppose the heteroskedasticity can be modeled as:
  \[ \text{Var}[\varepsilon|x] = \sigma^2 h(x) \]

• The key is to figure out what \( h(x) \) looks like. Suppose that we know \( h_i \). For example, \( h_i(x) = x_i^2 \). (Make sure \( h_i \) is always positive.)

• Then, use \( \frac{1}{\sqrt{x_i^2}} \) to transform the model.

Estimation: WLS form of GLS

• Suppose that we know \( h_i(x) = x_i^2 \). Then, use \( 1/\sqrt{x_i^2} \) to transform the model:
  \[ \text{Var}[\frac{\varepsilon_x}{\sqrt{x_i^2}} | x] = \sigma^2 \]

• Thus, if we divide our whole equation by \( \sqrt{h_i(x)} \) we get a (transformed) model where the error is homoskedastic.

• Assuming weights are known, we have a two-step GLS estimation:
  - Step 1: Use OLS, then the residuals to estimate the weights.
  - Step 2: Weighted least squares using the estimated weights.

• Greene has a proof based on our asymptotic theory for the asymptotic equivalence of the second step to true GLS.
Estimation: FGLS

- More typical is the situation where we do not know the form of the heteroskedasticity. In this case, we need to estimate $h_i(x)$.

- Typically, we start by assuming a fairly flexible model, such as
  \[ \text{Var}[\varepsilon | x] = \sigma^2 \exp(X\delta) \]
  - make sure $\text{Var}[\varepsilon_i | x] > 0$.

But, we don't know $\delta$, it must be estimated. By our assumptions:

\[ \varepsilon^2 = \sigma^2 \exp(X\delta) \nu \]

with $E(\nu | X) = 1$.

Then, if $E(\nu) = 1$

\[
\ln(\varepsilon^2) = X\delta + u \quad (\ast)
\]

where $E(u) = 0$ and $u$ is independent of $X$.

We know that $e$ is an estimate of $\varepsilon$, so we can estimate ($\ast$) by OLS.

---

Estimation: FGLS

- Now, an estimate of $h$ is obtained as $\hat{h} = \exp(\hat{\delta})$, and the inverse of this is our weight. Now, we can do GLS as usual.

- Summary of FGLS

  1. Run the original OLS model, save the residuals, $e$. Get $\ln(e^2)$.
  2. Regress $\ln(e^2)$ on all of the independent variables. Get fitted values, $\hat{\delta}$.
  3. Do WLS using $1/\sqrt{\exp(\hat{\delta})}$ as the weight.
  4. Iterate to gain efficiency.

- Remark: We are using WLS just for efficiency – OLS is still unbiased and consistent. Sandwich estimator gives us consistent inferences.
Estimation: MLE

• ML estimates all the parameters simultaneously. To construct the likelihood, we assume a distribution for ε. Under normality (A5):

\[
\ln L = -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^{T} \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{T} \left( \frac{1}{\sigma_i^2} (y_i - X_i\beta)'(y_i - X_i\beta) \right)
\]

• Suppose \( \sigma_i^2 = \exp(\alpha_0 + \alpha_1 z_{1,i} + \ldots + \alpha_m z_{m,i}) = \exp(z_i'\alpha) \)

Then, the first derivatives of the log likelihood wrt \( \theta = (\beta, \alpha) \) are:

\[
\frac{\partial \ln L}{\partial \beta} = -\sum_{i=1}^{T} x_i e_i / \sigma_i^2 = X'\Sigma^{-1}e
\]

\[
\frac{\partial \ln L}{\partial \alpha_i} = -\frac{1}{2} \sum_{i=1}^{T} 1 / \sigma_i^2 \exp(\alpha_i)z_i - (-\frac{1}{2}) \sum_{i=1}^{T} e_i^2 / \sigma_i^4 \exp(\alpha_i)z_i = -\frac{1}{2} \sum_{i=1}^{T} z_i (e_i^2 / \sigma_i^2 - 1)
\]

• Then, we get the f.o.c. We get a non-linear system of equations.

Estimation: MLE

• We take second derivatives to calculate the information matrix:

\[
\frac{\partial^2 \ln L}{\partial \beta^2} = -\sum_{i=1}^{T} x_i x_i'/ \sigma_i^2 = X'\Sigma^{-1}X
\]

\[
\frac{\partial^2 \ln L}{\partial \beta \alpha_i} = -\frac{1}{2} \sum_{i=1}^{T} x_i z_i e_i / \sigma_i^2
\]

\[
\frac{\partial^2 \ln L}{\partial \alpha_i \alpha_j} = -\frac{1}{2} \sum_{i=1}^{T} z_i z_j e_i^2 / \sigma_i^2
\]

• Then,

\[
I(\theta) = E\left[ -\frac{\partial \ln L}{\partial \theta \theta'} \right] = \begin{bmatrix} X'\Sigma^{-1}X & 0 \\ 0 & \frac{1}{2} Z'Z \end{bmatrix}
\]

• We can estimate the model using Newton’s method:

\[
\theta_{j+1} = \theta_j - H_j^{-1} g_j \quad \text{with} \quad g_j = \partial \log L_j / \partial \theta'
\]
Estimation: MLE

- We estimate the model using Newton’s method:
  \[ \theta_{j+1} = \theta_j - H_j^{-1} g_j \]
  \[ g_j = \partial \log L_j / \partial \theta' \]
  where subscript \( j \) stands for values computed under iteration \( j \).

Since \( H \) is block diagonal, then
\[ \beta_{j+1} = \beta_j - (X' \Sigma_j^{-1} X)^{-1} X' \Sigma_j^{-1} \varepsilon_j \]
\[ \alpha_{j+1} = \alpha_j - (\frac{1}{2} Z' Z)^{-1} \{ \frac{1}{2} \sum_{i=1}^{T} z_i (\tilde{e}_{i,j}^2 / \sigma_{i,j}^2 - 1) \} = \alpha_j - (Z' Z_j^{-1} Z' \nu, \]
where
\[ \nu = (\tilde{e}_{i,j}^2 / \sigma_{i,j}^2 - 1) \]

Convergence will be achieved when \( g_j = \partial \log L_j / \partial \theta' \) is close to zero.

- We have an iterative algorithm \( \implies \) Iterative FGLS = MLE!

Heteroscedasticity: Log Transformations

- A log transformation of the data, can eliminate (or reduce) a certain type of heteroskedasticity.
  - Assume - \( \mu_t = E[Z_t] \)
  - \( \text{Var}[Z_t] = \delta \mu_t^2 \) (Variance proportional to the squared mean)

- We log transformed the data: \( \log(Z_t) \). Then, we use the delta method to approximate the variance of the transformed variable. Recall: \( \text{Var}[f(X)] \) using delta method:
  \[ \text{Var}[f(X)] \approx f'(\theta)^2 \text{Var}[X] \]

- Then, the variance of \( \log(Z_t) \) is roughly constant:
  \[ \text{Var}[\log(Z_t)] \approx (1/\mu_t)^2 \text{Var}[Z_t] = \delta \]
ARCH Models

• Until the early 1980s econometrics had focused almost solely on modeling the conditional (on information set at time $t$, $I_t$) mean:

$$y_t = E[y_t | I_t] + \varepsilon_t,$$

$\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

Suppose we have an AR(1) process:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t.$$

Then, the conditional mean, conditioning on $I_{t-1}$, is:

$$E_{t-1}[y_t | I_{t-1}] = \alpha + \beta y_{t-1}$$

• Recall the distinction between conditional moments and unconditional ones. The unconditional mean is:

$$E[y_t] = \frac{\alpha}{1-\beta} = \text{constant}$$

The conditional mean is time varying; the unconditional mean is not!

ARCH Models

• Similar idea for the variance. For the AR(1) process:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t.$$

Unconditional variance:

$$\text{Var}[y_t] = E[(y_t - E[y_t])^2] = \frac{\sigma^2}{1-\beta^2} = \text{constant}$$

Conditional variance:

$$\text{Var}_{t-1}[y_t | I_{t-1}] = E_{t-1}[y_t - E[y_t])^2] = E_{t-1}[\varepsilon_t^2]$$

Remark: Conditional moments are time varying; unconditional moments are not!
ARCH Models

- The unconditional variance measures overall uncertainty. In the AR(1) example, the information set $I_t$ plays no role: $\text{Var}[y_t] = \frac{\sigma^2}{1 - \beta^2}$

- The conditional variance, $\text{Var}_t[y_{t+1} | I_t]$, is a better measure of uncertainty at time $t$. It is a function of information at time $t$, $I_t$.

ARCH Models: Stylized Facts of Asset Returns

- **Thick tails** - Mandelbrot (1963): leptokurtic (thicker than Normal)

- **Volatility clustering** - Mandelbrot (1963): “large changes tend to be followed by large changes of either sign.”

- **Leverage Effects** — Black (1976), Christie (1982): Tendency for changes in stock prices to be negatively correlated with changes in volatility.

- **Non-trading Effects, Weekend Effects** — Fama (1965), French and Roll (1986): When a market is closed information accumulates at a different rate to when it is open —for example, the weekend effect, where stock price volatility on Monday is not three times the volatility on Friday.
ARCH Models: Stylized Facts of Asset Returns

- Expected events — Cornell (1978), Patell and Wolfson (1979), etc: Volatility is high at regular times such as news announcements or other expected events, or even at certain times of day—for example, less volatile in the early afternoon.

- Volatility and serial correlation — LeBaron (1992): Inverse relationship between the two.


• We need a model that accommodates all these facts.

ARCH Models: Stylized Facts of Asset Returns

• Easy to check leptokurtosis (Stylized Fact #1)

Figure: Descriptive Statistics and Distribution for Monthly S&P500 Returns

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (%)</td>
<td>0.626332</td>
</tr>
<tr>
<td>(p-value: 0.0004)</td>
<td></td>
</tr>
<tr>
<td>Standard Dev (%)</td>
<td>4.37721</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.43764</td>
</tr>
<tr>
<td>Excess Kurtosis</td>
<td>2.29395</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>145.72</td>
</tr>
<tr>
<td>(p-value: &lt;0.0001)</td>
<td></td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.0258</td>
</tr>
<tr>
<td>(p-value: 0.5249)</td>
<td></td>
</tr>
</tbody>
</table>
ARCH Models: Stylized Facts of Asset Returns

- Easy to check Volatility Clustering (Stylized Fact #2)


- We start with assumptions (A1) to (A5), but with a specific (A3'):
  \[ y_t = \gamma X_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) \]
  \( (A3') \sigma_t^2 = \text{Var}t_{-1}[\varepsilon_t] = \text{E}_{t-1}[\varepsilon_t^2] = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2, \]
  which we can write, using the \( L \) operator, as:
  \[ \sigma_t^2 = \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \omega + \alpha(L)\varepsilon^2 \]

  - We can write the model in terms of an AR(q) for \( \varepsilon_t^2 \). Define
    \[ \nu_t \equiv \varepsilon_t^2 - \sigma_t^2, \]
    - an error term for the variance.
    Then,
    \[ \varepsilon_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \nu_t \]

  - Correlated \( \varepsilon_t^2 \)'s. High (low) past \( \varepsilon_t^2 \)'s produce a high (low) \( \varepsilon_t^2 \) today.

- The model
  \[ \sigma_t^2 = \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \omega + \alpha(L)\varepsilon^2 \]
is an AR(q) model for squared innovations, \( \varepsilon_t^2 \). We have the ARCH model: **Auto-Regressive Conditional Heteroskedasticity**.

- The ARCH(q) model estimates the unobservable (latent) variance.

- Non-negative constraints: Since we are dealing with a variance, we usually impose \( \omega > 0 \) and \( \alpha_i > 0 \) for all \( i \).

**Notation**: \( E_{t-1}[\varepsilon_t^2] = E[\varepsilon_t^2 | I_t] \)

**Useful result**: Since \( E[\varepsilon_t] = 0 \), then \( E_{t-1}[\varepsilon_t^2] = \sigma_t^2 \)

ARCH Models: Unconditional Variance

- The unconditional variance is determined by:
  \[ \sigma^2 = E[\sigma_t^2] = \omega + \sum_{i=1}^{q} \alpha_i E[\varepsilon_{t-i}^2] = \omega + \sum_{i=1}^{q} \alpha_i \sigma^2 \]

That is,
\[ \sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i} \]

To obtain a positive \( \sigma^2 \), we impose another restriction: \( (1 - \sum_{i=1}^{q} \alpha_i) > 0 \)

**Example**: ARCH(1)
\[
Y_t = \beta X_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) \\
\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \quad \Rightarrow \sigma^2 = \frac{\omega}{1 - \alpha_1}
\]

- We need to impose restrictions: \( \omega > 0, \alpha_1 > 0, \) & \( (1 - \alpha_1) > 0 \).
ARCH Models: Leptokurtosis

• Errors may be serially uncorrelated, but they are not independent: There will be volatility clustering, which produces fat tails.

We want to calculate the kurtosis of the errors:

\[ \kappa(\varepsilon_t) = \frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} \]

• We define standardized errors:

\[ z_t = \frac{\varepsilon_t}{\sigma_t} \]

They have conditional mean zero and a time invariant conditional variance equal to 1. That is, \( z_t \sim \mathcal{N}(0, 1) \).

• From the definition of \( z_t \) we have: \( \varepsilon_t = z_t \sigma_t \)

Now, we compute the fourth (also central, since \( E[\varepsilon_t]=0 \)) moment:


Then, using Jensen’s inequality:


\[ \kappa(\varepsilon_t) = \frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} > 3. \]

where we have used the fact that since \( E[\varepsilon_t] = 0 \), then \( E[\varepsilon_t^2] = E[\sigma_t^2] \)

Technical point: It can be shown that for an ARCH(1), the 4th moment for an ARCH(1):

\[ \kappa(\varepsilon_t) = \frac{3(1 - \alpha^2)}{1 - 3\alpha^2} \quad \text{if} \ 3\alpha^2 < 1. \]
ARCH Models: Alternative Representation

• More convenient, but less intuitive, presentation of the ARCH(1) model:

\[ \varepsilon_t = \sigma_t^2 z_t \]

where \( v_t \) is i.i.d. with mean 0, and \( \text{Var}[z_t]=1 \). Since \( z_t \) is i.i.d., then:

\[ E_{t-1}[\varepsilon_t^2] = E_{t-1}[\sigma_t^2 z_t^2] = E_{t-1}[\sigma_t^2]E_{t-1}[z_t^2] = \omega + \alpha_1 \varepsilon_{t-1}^2 \]

• It turns out that \( \sigma_t^2 \) is a very persistent process. Such a process can be captured with an ARCH(\( q \)), where \( q \) is large. This is not efficient.

• In practice, \( q \) is often large. A more parsimonious representation is the Generalized ARCH model or GARCH(\( q, p \)):

\[ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \]

\[ = \omega + \alpha(L)\varepsilon^2 + \beta(L)\sigma^2 \]

GARCH: Bollerslev (1986)

• A more parsimonious representation is the GARCH(\( q, p \)):

\[ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \]

which is an ARMA(max(\( p, q \)), \( p \)) model for the squared innovations.

• Popular GARCH model: GARCH(1,1):

\[ \sigma_{t+1}^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 \]

with an unconditional variance:

\[ \text{Var}[\varepsilon_t^2] = \sigma_t^2 = \frac{\omega}{1 - \alpha_1 - \beta_1} \]

\[ \Rightarrow \text{Restrictions: } \omega > 0, \alpha_1 > 0, \beta_1 > 0, (1 - \alpha_1 - \beta_1) > 0. \]

• Technical details: This is covariance stationary if all the roots of \( \alpha(L) + \beta(L) = 1 \) lie outside the unit circle. For the GARCH(1,1) this amounts to \( \alpha_1 + \beta_1 < 1 \).
GARCH: Bollerslev (1986)

- Bollerslev (1986) showed that if $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, the second and 4th (unconditional) moments of $\varepsilon_t$ exist:

$$E[\varepsilon_t^2] = \frac{\omega}{(1 - \alpha_1 - \beta_1)}$$

$$E[\varepsilon_t^4] = \frac{3\omega^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)} \quad \text{if } (1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2) \neq 0$$

GARCH-X

- In the GARCH-X model, exogenous variables are added to the conditional variance equation.

Consider the GARCH(1,1)-X model:

$$\sigma_{t+1}^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 + \delta \, f(X_{t-1})$$

where $f(X_t, \Theta)$ is strictly positive for all $t$. Usually, $X_t$, is an observed economic variable or indicator, say liquidity index, and $f(.)$ is a non-linear transformation, which is non-negative.

• Recall the technical detail: The standard GARCH model:

\[
\sigma_t^2 = \omega + \alpha(L)e_t^2 + \beta(L)\sigma_t^2
\]

is covariance stationary if \( \alpha(1) + \beta(1) < 1 \)

• But strict stationarity does not require such a stringent restriction (That is, that the unconditional variance does not depend on \( t \)).

In the GARCH(1,1) model, if \( \alpha_1 + \beta_1 = 1 \), we have the Integrated GARCH (IGARCH) model. In practice, in many applications, \( \alpha_1 + \beta_1 \) are close to one.

• Implication: Shocks to the conditional variance are "persistent."

This can be seen in the variance forecasts: \( E_t[\sigma_{t+j}^2] = \sigma_t^2 + j\omega \)

\[ \Rightarrow \text{Today's variance remains important for all future forecasts} \]

• In practice, it is often found that \( \alpha_1 + \beta_1 \) are close to 1.

• It is often argued that IGARCH is a product of omitted variables, for example, structural breaks.
GARCH: Variations – GARCH-in-mean

- The time-varying variance affects mean returns:
  Mean equation: \( y_t = X_t \gamma + \delta \sigma_t^2 + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2) \)
  Variance equation: \( \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \)

- We have a dynamic mean-variance relations. It describes a specific form of the risk-return trade-off.

- Finance intuition says that \( \delta \) has to be positive and significant. However, in empirical work, it does not work well: \( \delta \) is not significant or negative.

GARCH: Variations – Asymmetric GJR

- GJR-GARCH model – Glosten, Jagannathan & Runkle (JF, 1993):
  \[ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}^2 I_{t-i} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \]
  where \( I_{t-i} = 1 \) if \( \epsilon_{t-i} < 0 \),
  \( = 0 \) otherwise.

- Using the indicator variable \( I_{t-i} \), this model captures sign (asymmetric) effects in volatility. Negative news \( (\epsilon_{t,i} < 0) \) increase the conditional volatility (leverage effect).

- The GARCH(1,1) version:
  \[ \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2 I_{t-1} + \beta_1 \sigma_{t-1}^2 \]
  where \( I_{t-1} = 1 \) if \( \epsilon_{t-1} < 0 \),
  \( = 0 \) otherwise.
GARCH: Variations – Asymmetric GJR

- The GARCH(1,1) version:
  \[ \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma \epsilon_{t-1}^2 I_{t-1} + \beta \sigma_{t-1}^2 \]

When \( \epsilon_{t-1} < 0 \) \( \Rightarrow \sigma_t^2 = \omega + (\alpha_1 + \gamma) \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \)

When \( \epsilon_{t-1} > 0 \) \( \Rightarrow \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \)

- This is a very popular variation of the GARCH models. The leverage effect is significant.

- There is another variation, the Exponential GARCH, or EGARCH, that also captures the asymmetric effect of negative news on the conditional variance.

GARCH: Variations – EGARCH

  It models an exponential function for the time-varying variance:
  \[ \log(\sigma_t^2) = \omega + \sum_{j=1}^{q} \alpha_j (z_{t-j}^2 + \gamma |z_{t-j}|) + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2) \]
  where \( z \) is a standardized i.i.d. \( \mathcal{N}(0, 1) \) innovation.

- By design, we have the variance follows an exponential function. Thus, no non-negative restrictions on the parameters are imposed.

- Negative news \( (z_{t-1} < 0) \) increase \( \sigma_t^2 \) (leverage effect).

Note: Nelson provides formulas of the unconditional moments, under the GED. But, under leptokurtic distributions such as the Student-t the unconditional variance does not exist. (Intuition: we have an exponential formulation, with a large shock it can explode.)
GARCH: Variations – NARCH


These models apply the Box-Cox-type transformation to the conditional variance:

$$\sigma_t^\gamma = \omega + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i} - \kappa|^\gamma + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\gamma$$

Special case: $\gamma = 2$ (standard GARCH model).

Note: The variance depends on both the size and the sign of the variance which helps to capture leverage type (asymmetric) effects.

GARCH: Variations – TARCH

• Threshold ARCH (TARCH) – Rabemananjara & Zakoian (1993)

Large events – i.e., large errors- have a different effect from small events. We use 2 indicator variables, $I(\varepsilon_{t-i} > \kappa)$ & $I(\varepsilon_{t-i} < \kappa)$: one for “large events,” ($\varepsilon_{t-i} > \kappa$), & one for “small events;” ($\varepsilon_{t-i} < \kappa$):

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \{ \alpha_i^+ I(\varepsilon_{t-i} > \kappa) + \alpha_i^- I(\varepsilon_{t-i} < \kappa) \} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

There are two variances:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i^+ \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \quad \text{if } (\varepsilon_{t-i} > \kappa)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i^- \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \quad \text{if } (\varepsilon_{t-i} < \kappa)$$

• We can modify the model in many ways. For example, we can allow for the asymmetric effects of negative news.
GARCH: Variations – SWARCH

• Switching ARCH (SWARCH) – Hamilton and Susmel (JE, 1994).

**Intuition:** $\sigma_t^2$ depends on the state of the economy – regime. It’s based on Hamilton’s (1989) time series models with changes of regime:

$$\sigma_t^2 = \omega_{s_t,s_{t-1}} + \sum_{i=1}^{q} \alpha_i e_{i-1}^2$$

The key is to select a parsimonious representation:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i e_{i-1}^2$$

For a SWARCH(1) with 2 states (1 and 2) we have 4 possible $\sigma_t^2$:

- $\sigma_t^2 = \omega_{1,1} + \alpha_1 e_{t-1}^2 / \gamma_{1}$, $s_t = 1$, $s_{t-1} = 1$
- $\sigma_t^2 = \omega_{1,2} + \alpha_1 e_{t-1}^2 / \gamma_{2}$, $s_t = 1$, $s_{t-1} = 2$
- $\sigma_t^2 = \omega_{2,1} + \alpha_1 e_{t-1}^2 / \gamma_{1}$, $s_t = 2$, $s_{t-1} = 1$
- $\sigma_t^2 = \omega_{2,2} + \alpha_1 e_{t-1}^2 / \gamma_{2}$, $s_t = 2$, $s_{t-1} = 2$

GARCH: Forecasting and Persistence

• Consider the forecast in a GARCH(1,1) model

$$\sigma_{t+1}^2 = \omega + \alpha_1 e_t^2 + \beta_1 \sigma_t^2 = \omega + \sigma_t^2 (\alpha_1 z_t^2 + \beta_1)$$

$$e_t^2 = \sigma_t^2 z_t^2$$

Taking expectation at time $t$

$$E_t[\sigma_{t+1}^2] = \omega + \sigma_t^2 (\alpha_1 z_t^2 + \beta_1)$$

Then, by repeated substitutions:

$$E_t[\sigma_{t+j}^2] = \omega + \sigma_t^2 (\alpha_1 z_t^2 + \beta_1)^j = \omega [1 + (\alpha_1 + \beta_1)] + \sigma_t^2 (\alpha_1 + \beta_1)^j$$

$$= \omega \sum_{i=0}^{j-1} (\alpha_1 + \beta_1)^i + \sigma_t^2 (\alpha_1 + \beta_1)^j$$

Assuming $(\alpha_1 + \beta_1) < 1$, as $j \to \infty$, the forecast reverts to the unconditional variance:

$$\sigma_t^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}$$
GARCH: Forecasting and Persistence

• Today’s forecast in a GARCH(1,1) model for horizon \( j \):
  \[
  E_t[\sigma_{t+j}^2] = \omega \sum_{l=0}^{j-1} (\alpha_1 + \beta_1)^l + \sigma_t^2 (\alpha_1 + \beta_1)^j
  \]

When \( \alpha_1 + \beta_1 = 1 \), today’s volatility affect future forecasts for all horizons:
  \[
  E_t[\sigma_{t+1}^2] = \sigma_t^2 + j\omega
  \]

A shock to \( \sigma_t^2 \) stays forever in the forecast!

ARCH Estimation: MLE

• All of these models can be estimated by maximum likelihood. First we need to construct the sample likelihood.

• Since we are dealing with dependent variables, we use the conditioning trick to get the joint distribution:
  \[
  f(y_1, y_2, \ldots, y_T; \theta) = f(y_1 | x_1; \theta)f(y_2 | y_1, x_2; \theta)f(y_3 | y_2, y_1, x_3, x_2; \theta) \ldots f(y_T | y_{T-1}, \ldots, y_1, x_T-1, \ldots, x_1; \theta).
  \]

Taking logs:
  \[
  L = \log(f(y_1, y_2, \ldots, y_T; \theta))) = \log(f(y_1 | x_1; \theta)) + \log(f(y_2 | y_1, x_2; \theta)) + \ldots + \log(f(y_T | y_{T-1}, \ldots, y_1, x_T-1, \ldots, x_1; \theta))
  \]
  \[
  = \sum_{t=1}^{T} \log(f(y_t | y_{t-1}, x_t; \theta))
  \]

• We maximize this function with respect to the \( k \) mean parameters (\( \gamma \)) and the \( m \) variance parameters (\( \omega, \alpha_1, \beta_1 \)). Let \( \theta = (\gamma, \omega, \alpha_1, \beta_1) \).
ARCH Estimation: MLE

- We compute f.o.c.'s: \( \frac{\partial L}{\partial \theta} = S(y_t, 0) = 0 \) -S(\( )) = Score vector.

- We have a \((k+m \times k+m)\) system. But, it is a non-linear system. We use numerical optimization. Gauss-Newton or BHHH (approximates \( H \) also by the outer product of \( S(y_t, \theta) \)) can be easily implemented.

- Given the AR structure, we will need to make assumptions about \( \sigma_0 \) (and \( \varepsilon_0, \varepsilon_1, ..., \varepsilon_p \) if we assume an AR(\( p \)) process for the mean).

- Alternatively, we take \( \sigma_0 \) (\& \( \varepsilon_0, \varepsilon_1, ..., \varepsilon_p \)) as parameters to estimate.

**Note:** The \( \frac{\partial L}{\partial Y} = 0 \) \((k \times 1\) vector of f.o.c.'s) will give us GLS (or WLS).

\[
\frac{\partial L}{\partial Y} = -\sum_{t=1}^{T} X_t' \varepsilon_t / \sigma_t^{MLE} = \sum_{t=1}^{T} X_t' (y_t - X_t \mu_{MLE}) / \sigma_t^{MLE} = 0
\]

\[
= \sum_{t=1}^{T} \frac{X_t}{\sigma_t^{MLE}} \left( \frac{y_t}{\sigma_t^{MLE}} - \frac{X_t}{\sigma_t^{MLE}} \mu_{MLE} \right) = 0
\]

ARCH Estimation: MLE

- MLE = GLS for mean parameters, \( y_t \) \& \( X_t \) weighted by \( 1/ \sigma_t^{MLE} \).

- If the conditional density is well specified and \( \theta_0 \) belongs to \( \Omega \), then

\[
T^{1/2} (\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, A_0^{-1}), \quad \text{where} \quad A_0^{-1} = T^{-1} \sum_{t=1}^{T} \frac{\partial S_t(y_t, \theta_0)}{\partial \theta}
\]

- Under the correct specification assumption, \( A_0 = B_0 \), where

\[
B_0 = T^{-1} \sum_{t=1}^{T} E[S_t(y_t, \theta_0), S_t(y_t, \theta_0)']
\]

We estimate \( A_0 \) and \( B_0 \) by replacing \( \theta_0 \) by its estimated MLE value.

- The estimator \( B_0 \) has a computational advantage over \( A_0 \): Only first derivatives are needed. But \( A_0 = B_0 \) only if the distribution is correctly specified. This is very difficult to know in practice.
ARCH Estimation: MLE – ARCH(1)

Example: ARCH(1) model.

Mean equation: \( y_t = X_t \gamma + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2) \)

Variance equation: \( \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \)

We write the pdf for the normal distribution,

\[
f(\varepsilon_t | \gamma, \omega, \alpha_1) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[ -\frac{\varepsilon_t^2}{2\sigma_t^2} \right] = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[ -\frac{(y_t - X_t\gamma)^2}{2\sigma_t^2} \right]
\]

We form the likelihood \( L \) (the joint pdf):

\[
L = \prod_{t=1}^{T} \exp \left( \frac{-\varepsilon_t^2}{2\sigma_t^2} \right) = (2\pi)^{-T/2} \prod_{t=1}^{T} \frac{1}{\sqrt{\sigma_t^2}} \exp \left( -\frac{\varepsilon_t^2}{2\sigma_t^2} \right)
\]

We take logs to form the log likelihood, \( L = \log L \):

\[
L = \sum_{t=1}^{T} \log(f_t) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t^2 / \sigma_t^2
\]

Then, we maximize \( L \) with respect to \( \theta = (\gamma, \omega, \alpha_1) \) the function \( L \).
ARCH Estimation: MLE – Example (in R)

- Log likelihood of AR(1)-GARCH(1,1) Model:

```r
log_lik_garch11 <- function(theta, data) {
  mu <- theta[1]; rho1 <- theta[2]; alpha0 <- abs(theta[3]); alpha1 <- abs(theta[4]); beta1 <- abs(theta[5]);
  chk0 <- (1 - alpha1 - beta1)
  r <- ts(data)
  n <- length(r)
  u <- vector(length=n), u <- ts(u)
  for (t in 2:n) {
    u[t] = r[t] - mu - rho1*r[t-1] # this setup allows for ARMA in mean
  }
  h <- vector(length=n), h <- ts(h)
  h[1] = alpha0/chk0 # set initial value for h[t] series
  if  (chk0==0) {h[1]=.000001} # check to avoid dividing by 0
  for (t in 2:n) {
    h[t] = abs(alpha0 + alpha1*(u[t-1]^2) + beta1*h[t-1])
    if  (h[t]==0) {h[t]=.00001} #check to avoid log(0)
  }
  return(-sum(- 0.5*log(abs(h[2:n])) - 0.5*(u[2:n]^2)/abs(h[2:n]))) # ignore constants
}

Example 1: AR(1)-GARCH(1,1) model for changes in CHF/USD.
We will use R function `optim` (min also OK) to maximize the function.

```r
PPP_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/ppp_2020_m.csv",head=TRUE,sep=",")
x_chf  <- PPP_da$CHF_USD # CHF/USD 1971-2020 monthly data
T <- length(x_chf)
z <- log(x_chf[-1]/x_chf[-T])
theta0 = c(-0.002, 0.026, 0.001, 0.19, 0.71) # initial values
ml_2 <- optim(theta0, log_lik_garch11, data=z, method="BFGS", hessian=TRUE)
logL_g11 <- log_lik_garch11(ml_2$par, z) # value of log likelihood
ml_2$par # estimated parameters
I_Var_m2 <- ml_2$hessian
eigen(I_Var_m2) # check if Hessian is pd.
sqrt(diag(solve(I_Var_m2))) # parameters SE
chf_usd <- ts(z, frequency=12, start=c(1971,1)) # time series plot of data
plot.ts(chf_usd)
```

ARCH Estimation: MLE – Example (in R)
ARCH Estimation: MLE – Example (in R)

Example 1 (continuation):

```r
> logL_g11 # Log likelihood value
[1] -1745.197

> ml_2$par # Extract from ml_2 function parameters
[1] -0.0021051742  0.0260003610  0.00012375 0.1900276519  0.7100718082

> I_Var_m2 <- ml_2$hessian # Extract Hessian (matrix of 2nd derivatives)

> eigen(I_Var_m2) # Check if Hessian is pd to invert.
$eigenvalues:
[1] 1.687400e+08 6.954454e+05 7.200084e+03 5.120984e+02 2.537958e+02
$eigenvalues: if positives => Hessian is pd

> sqrt(diag(solve(I_Var_m2))) # Invert Hessian: Parameters Var on diag
[1] 1.203690e-03 4.419049e-02 7.749756e-05 5.014454e-02 3.955411e-02

> t_stats <- ml_2$par/sqrt(diag(solve(I_Var_m2)))

> t_stats
[1] -1.7489333  0.5883701  1.5967743  3.7895984 17.9519078
```
ARCH Estimation: MLE – Example (in R)

Example 1 (continuation): Summary for CHF/USD changes

\[ e_{f,t} = \log(S_t) - \log(S_{t-1}) = a_0 + a_1 e_{f,t-1} + \epsilon_t, \quad \epsilon_t \mid I_{t-1} \sim N(0, \sigma_t^2) \]

\[ \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \]

- \( T \): 562 (January 1971 - July 2020, monthly).

The estimated model for \( e_{f,t} \) is given by:

\[ e_{f,t} = -0.00211 + 0.02600 e_{f,t-1}, \]

\[ \sigma_t^2 = 0.00012 + 0.19003 \epsilon_{t-1}^2 + 0.71007 \sigma_{t-1}^2. \]

Unconditional \( \sigma^2 = \frac{0.00012}{1 - 0.19003 - 0.71007} = 0.001201201 \)

Log likelihood: 1745.197

Note: \( \alpha_1 + \beta_1 = 0.90 < 1. \) (Persistent.)
ARCH Estimation: MLE – Example (in R)

**Example 2:** Using Robert Shiller’s monthly data set for the S&P 500 (1871:Jan - 2020:Aug, T=1,795), we estimate an AR(1)-GARCH(1,1) model:

\[ r_t = \log(P_t) - \log(P_{t-1}) = \alpha_0 + \alpha_1 r_{t-1} + \epsilon_t, \quad \epsilon_t | I_{t-1} \sim N(0, \sigma^2_t) \]

\[ \sigma^2_t = \omega + \alpha_1 \epsilon^2_{t-1} + \beta_1 \sigma^2_{t-1} \]

The estimated model for \( s_t \) is given by:

\[ r_t = 0.338 + 0.278 r_{t-1}, \]

\[ \sigma^2_t = 0.756 + 0.126 \epsilon^2_{t-1} + 0.826 \sigma^2_{t-1}. \]

Unconditional \( \sigma^2 = 0.756 / (1 - 0.126 - 0.826) = 15.4630 \)

Log likelihood: 4795.08

Note: \( \alpha_1 + \beta_1 = 0.952 < 1. \) (Very persistent.)

 ARCH Estimation: MLE – Example (in R)

**Example 2:** Below, we plot the time-varying variance. Certain events are clearly different, for example, the 1930 great depression, with a peak variance of 282 (18 times unconditional variance!). The covid-19 volatility similar to the 2008-2009 financial crisis recession:
ARCH Estimation: MLE – Regularity Conditions

**Note:** The appeal of MLE is the optimal properties of the resulting estimators under ideal conditions.

- Crowder (1976) gives one set of sufficient regularity conditions for the MLE in models with dependent observations to be consistent and asymptotically normally distributed.

- Verifying these regularity conditions is very difficult for general ARCH models - proof for special cases like GARCH(1,1) exists.

**Example:** For the GARCH(1,1) model: if $E[\ln(\alpha_1 z_t^2 + \beta_2)] < 0$, the model is strictly stationary and ergodic. See Nelson (1990) & Lumsdaine (1996).

**ARCH Estimation: MLE – Regularity Conditions**

- **Block-diagonality**
  In many applications of ARCH, the parameters can be partitioned into mean parameters, $\theta_1$, and variance parameters, $\theta_2$.

  Then, $\delta \mu(\theta)/\delta \theta_2 = 0$ and, although, $\delta \sigma^2(\theta)/\delta \theta_2 \neq 0$, the Information matrix is *block-diagonal* (under general symmetric distributions for $z_t$ and for particular ARCH specifications).

  Not a bad result:
  - Regression can be consistently done with OLS.
  - Asymptotically efficient estimates for the ARCH parameters can be obtained on the basis of the OLS residuals.
ARCH Estimation: MLE – Remarks

- But, block diagonality cannot buy everything:
  - Conventional OLS standard errors could be terrible.

- When testing for serial correlation, in the presence of ARCH, the conventional Bartlett s.e. – $T^{1/2}$ could seriously underestimate the true standard errors.

ARCH Estimation: Non-Normality

- The basic GARCH model allows a certain amount of leptokurtosis. It is often insufficient to explain real world data.

Solution: Assume a distribution other than the normal which help to allow for the fat tails in the distribution.

- $t$ Distribution - Bollerslev (1987)

  The $t$ distribution has a degrees of freedom parameter which allows greater kurtosis. The $t$ likelihood function is

  $$l_i = \ln(\Gamma(0.5(v + 1))\Gamma(0.5v)^{-1}(v - 2)^{-1/2}(1 + \frac{z_i}{v}(v - 2)^{-1})^{(v+1)/2}) - 0.5 \ln(\sigma_i^2)$$

  where $\Gamma$ is the gamma function and $v$ is the degrees of freedom. As $v \to \infty$, this tends to the normal distribution.

- GED Distribution - Nelson (1991)
ARCH Estimation: GMM

- Suppose we have an ARCH(\(q\)). We need moment conditions:

\[
\begin{align*}
(1) & \quad E[m_1] = E[x'_t(y_t - x_t\gamma)] = 0 \\
(2) & \quad E[m_2] = E[e_i^2(\sigma_i^2 - \sigma^2)] = 0 \\
(3) & \quad E[m_3] = E[e_i^2 - \omega / (1 - \alpha_1 - \ldots - \alpha_q)] = 0
\end{align*}
\]

Note: (1) refers to the conditional mean, (2) refers to the conditional variance, and (3) to the unconditional mean.

- GMM objective function:

\[
Q(X, y; \theta) = E[m(\theta; X, y)]' W E[m(\theta; X, y)]
\]

where

\[
E[m(\theta; X, y)] = [E[m_1]' E[m_2]' E[m_3]']'
\]

ARCH Estimation: GMM

- \(\gamma\) has \(k\) free parameters, \(\alpha\) has \(q\) free parameters. Then, we have \(J = k + q + 1\) parameters. Note that:

\(m(\theta, X, y)\) has \(J = k + q + 1\) equations.

**Dimensions:** \(Q\ is \ 1x1\); \(E[m(\theta, X, y)]\ is \ Jx1\); \(W\ is \ JxJ\).

- Problem is over-identified: more equations than parameters so cannot solve \(E[m(\theta, X, y)] = 0\), exactly.

- Choose a weighting matrix \(W\) for objective function and minimize using numerical optimization.

- Optimal weighting matrix:

\[
W = (E[m(\theta, X, y)] E[m(\theta, X, y)]')^{-1}
\]

\[\text{Var}(\theta) = (1/T)[DW^{-1}D']^{-1},\]

where \(D = \delta E[m(\theta, X, y)]/\delta \theta\') \quad \text{expressions evaluated at } \theta_{GMM}\]
ARCH Estimation: Testing

- Standard BP test, with auxiliary regression given by:
  \[ e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_q e_{t-q}^2 + v_t \quad \text{-Keep } R^2. \]

  \( H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_q = 0 \) (No ARCH). It is not possible to do GARCH test, since we are using the same lagged squared residuals.

  Then, the LM test is \( (T - q) R^2 \xrightarrow{d} \chi^2_q \) — Engle’s (1982).

- In ARCH Models, testing as usual: LR, Wald, and LM tests.

  Reliable inference from the LM, Wald and LR test statistics generally does require moderately large sample sizes of at least two hundred or more observations.

ARCH Estimation: Testing

- Issues:
  - Non-negative constraints must be imposed. \( \theta_0 \) is often on the boundary of \( \Omega \). (Two sided tests may be conservative.)
  - Lack of identification of certain parameters under \( H_0 \) creates a singularity of the Information matrix under \( H_0 \). For example, under \( H_0: \alpha_1 = 0 \) (No ARCH), in the GARCH(1,1), \( \omega \) and \( \beta_1 \) are not jointly identified. See Davies (1977).

- Ignoring ARCH
  - You suspect \( y_t \) has an AR structure: \( y_t = \gamma_0 + \gamma_1 y_{t-1} + \epsilon_t \)

    Hamilton (2008) finds that OLS t-test with no correction for ARCH spuriously reject \( H_0: \gamma_1 = 0 \) with arbitrarily high probability for sufficiently large \( T \). White’s (1980) SE help. NW SE help less.
ARCH Estimation: Testing

**Figure.** From Hamilton (2008). Fraction of samples in which OLS $t$-test leads to rejection of $H_0: \gamma_1 = 0$ as a function of $T$ for regression with Gaussian errors (solid line) and Student’s $t$ errors (dashed line). **Note:** $H_0$ is actually true and test has nominal size of 5%.

![Graph showing fraction of samples leading to rejection of $H_0$](image)

Testing for Heteroscedasticity: ARCH

- ARCH Test for the 3 factor F-F model for IBM returns ($T=320$), with one lag:
  \[
  r_{i,t} - r_f = \alpha_i + \beta_1 (r_{m,t} - r_f) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_{i,t}
  \]

  ```
  > b <- solve(t(x)%*% x)%*% t(x)%*%y #OLS regression
  > e <- y - x%*%b
  > e2 <- e^2
  > xx1 <- e2[1:T-1]
  > fit2 <- lm(e2[2:T]~xx1)
  > r2_e2 <- summary(fit2)$r.squared
  > r2_e2
  [1] 0.2656472
  > lm_t <- (T-1)*r2_e2
  > lm_t
  [1] 84.74147
  LM-ARCH Test: 84.74 \Rightarrow reject H_0 at 5% level ($\chi^2_{[1]} \approx 3.84$), the usual result for financial time series.
  ```
**GARCH: Forecasting and Persistence (Again)**

- Consider the forecast in a GARCH(1,1) model:

\[
\sigma_{t+1}^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 (\alpha_1 z_t^2 + \beta_1) \quad (\varepsilon_t^2 = \sigma_t^2 z_t^2)
\]

Taking expectation at time \( t \):

\[
E_t[\sigma_{t+1}^2] = \omega + \sigma_t^2 (\alpha_1 + \beta_1)
\]

Then, by repeated substitutions:

\[
E_t[\sigma_{t+j}^2] = \omega \left[ \sum_{i=0}^{j-1} (\alpha_1 + \beta_1)^i \right] + \sigma_t^2 (\alpha_1 + \beta_1)^j
\]

- When \( \alpha_1 + \beta_1 = 1 \), today's volatility affects future forecasts forever:

\[
E_t[\sigma_{t+j}^2] = \sigma_t^2 + j\omega
\]

---

**GARCH: Forecasting and Persistence**

**Example 1:** We want to forecast next month (September 2020) variance for CHF/USD changes. Recall we estimated \( \sigma_t^2 \):

\[
\sigma_t^2 = 0.00012 + 0.19003 \varepsilon_{t-1}^2 + 0.71007 \sigma_{t-1}^2.
\]

Getting \( \sigma_{2020:9}^2 = 0.003672220 \) \( (=\sigma_{2020:9}^2 = \sqrt{0.00367} = 6.1\%) \)

We based the \( \sigma_{2020:10}^2 \) forecast on:

\[
E_t[\sigma_{t+j}^2] = \omega \left[ \sum_{i=0}^{j-1} (\alpha_1 + \beta_1)^i \right] + \sigma_t^2 (\alpha_1 + \beta_1)^j
\]

Then, \( (\alpha_1 + \beta_1) = 0.190 + 0.710 = 0.900 \)

\[
E_{2020:9}[\sigma_{2020:10}^2] = 0.00012 + 0.00367 \times (0.9) = 0.003423
\]

We also forecast \( \sigma_{2020:12}^2 \)

\[
E_{2020:9}[\sigma_{2020:12}^2] = 0.00012 \times (1 + (0.9) + (0.9)^2) + 0.00367 \times (0.9)^3
\]

\[
= 0.00300063
\]
Example 1 (continuation):
We forecast volatility for March 2021:
\[
E_{2020:6}[\sigma^2_{2021:03}] = 0.00012 \times (1 + 0.9) + (0.9)^2 + \ldots + (0.9)^5 + \\
+ 0.00367 \times (0.9)^6 = 0.002512659
\]

Remark: We observe that as the forecast horizon increases \((j \to \infty)\),
the forecast reverts to the unconditional variance:
\[
\frac{\omega}{(1 - \alpha - \beta)} = \frac{0.00012}{1 - 0.9} = 0.0012
\]
\[
\Rightarrow \sigma = \sqrt{0.0012} = 0.0346 \quad (3.46\% \approx \text{close to sample SD = 3.36\%})
\]

Example 2: On August 2020, we forecast the December’s variance for the S&P 500 changes. Recall we estimated \(\sigma^2_t\):
\[
\sigma^2_t = 0.756 + 0.125 \varepsilon^2_{t-1} + 0.826 \sigma^2_{t-1}
\]
getting \(\sigma^2_{2020:8} = 43.037841\)

We based the \(\sigma^2_{2020:12}\) forecast on:
\[
E_t[\sigma^2_{t+j}] = \omega \times [\sum_{i=0}^{j-1}(\alpha_1 + \beta_1)i] + \sigma^2 (\alpha_1 + \beta_1)^j
\]

Then, since \((\alpha_1 + \beta_1) = 0.952\)
\[
E_{2020:8}[\sigma^2_{2020:12}] = 0.756 \times (1 + (0.952) + (0.952)^2 + (0.952)^3) + \\
+ 43.037841 \times (0.952)^4 = 38.02797
\]

Lower variance forecasted for the end of the year, but still far from the unconditional variance of 15.4.
ARCH: Which Model to Use

- Questions
  1) Lots of ARCH models. Which one to use?
  2) Choice of p and q. How many lags to use?

- Hansen and Lunde (2004) compared lots of ARCH models:
  - It turns out that the GARCH(1, 1) is a great starting model.
  - Add a leverage effect for financial series and it's even better.
  - A t-distribution is also a good addition.

RV Models: Intuition

- The idea of realized volatility is to estimate the latent (unobserved) variance using the realized data, without any modeling. Recall the definition of sample variance:

\[ s^2 = \frac{1}{(T - 1)} \sum_{i=1}^{T} (x_i - \bar{x})^2 \]

- Suppose we want to calculate the daily variance for stock returns. We know how to compute it: we use daily information, for T days, and apply the above definition.

- Alternatively, we use hourly data for the whole day (with k hours). Since hourly returns are very small, ignoring \( \bar{x} \) seems OK. We use \( r_{t,i}^2 \) as the i\(^{th}\) hourly variance on day t. Then, we add \( r_{t,i}^2 \) over the day:

\[ Variance_t = \sum_{i=1}^{k} r_{t,i}^2 \]
RV Models: Intuition

• In more general terms, we use higher frequency data to estimate a lower frequency variance:

\[ RV_t = \sum_{i=1}^{k} r_{t,i}^2 \]

where \( r_{t,i} \) is the realized returns in (higher frequency) interval \( i \) of the (lower frequency) period \( t \). We estimate the \( t \)-frequency variance, using \( k \) \( i \)-intervals. If we have daily returns and we want to estimate the monthly variance, then, \( k \) is equal to the number of days in a month.

• It can be shown that as the interval \( i \) becomes smaller (\( i \to 0 \)),
\[ RV_t \to \text{Return Variation} \ [t - 1, t]. \]

That is, with an increasing number of observations we get an accurate measure of the latent variance.

RV Models: High Frequency

• Note that RV is a model-free measure of variation — i.e., no need for ARCH-family specifications. The measure is called realized variance (RV). The square root of the realized variance is the realized volatility (RVol, RealVol):

\[ RVol_t = \sqrt{RV_t} \]

• Given the previous theoretical result, RV is commonly used with intra-daily data, called high frequency (HF) data.

• It lead to a revolution in the field of volatility, creating new models and new ways of thinking about volatility and how to model it.

• We usually associate realized volatility with an observable proxy of the unobserved volatility.
RV Models: High Frequency – Tick Data

• The theory behind realized variation measures dictates that the sampling frequency, or $k$ in the $RV_t$ formula above, goes to $\infty$. Then, use highest frequency available, say millisecond to millisecond returns.

• Intra-daily data applications are the most common. But, when using intra-daily data, RV calculations are affected by microstructure effects: bid-ask bounce, infrequent trading, calendar effects, etc. $r_{t,i}$ does not look uncorrelated.

Example: The bid-ask bounce induces serial correlation in intra-day returns, which biases $RV_t$.

• As the sampling frequency increases, the “noise” (microstructure effects) becomes more dominant and swallows the “signal” (true volatility).

RV Models: High Frequency – Tick Data

• In practice, sampling a typical stock price every few seconds can overestimate the true volatility by a factor of two or more.

• The usual solutions:
(1) Filter data using an ARMA model to get rid of the autocorrelations and/or dummy variables to get rid of calendar effects.

Then, used the filtered data to compute $RV_t$.

(2) Sample at frequencies where the impact of microstructure effects is minimized and/or eliminated.

We follow solution (2).
RV Models: High Frequency – Practice

- In intra-daily RV estimation, it is common to use 10' intervals. They have good properties. However, there are estimations with 1' intervals.

- Some studies suggest using an optimal frequency, where optimal frequency is the one that minimizes the MSE.

- Hansen and Lunde (2006) find that for very liquid assets, such as the S&P 500 index, a 5' sampling frequency provides a reasonable choice. Thus, to calculate daily RV, we need to add 78 five-minute intervals.

RV Models: High Frequency – TAQ

**Example**: Based on TAQ (Trade and Quote) NYSE data, we use 5' realized returns to calculate 30' variances –i.e., we use six 5' intervals. Then, the 30' variance, or $RV_{t=30-min}$, is:

$$ RV_{t=30-min} = \sum_{j=1}^{k=6} r_{t,j}^2, \quad t = 1, 2, \ldots, T = 13 $$

$r_{t,j}$ is the 5' return during the $j^{th}$ interval on the half hour $t$. Then, we calculate 30' variances for the whole day –i.e., we calculate 13 variances, since the trading day goes from 9:30 AM to 4:00 PM.

The Realized Volatility, $RVol_t$, is:

$$ RVol_{t=30-min} = \sqrt{RV_{t=30-min}} $$
RV Models: High Frequency – TAQ

Example: Below, we show the first transaction of the SPY TAQ (Trade and Quote) data (tick-by-tick trade data) on January 2, 2014.

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>DATE</th>
<th>TIME</th>
<th>PRICE</th>
<th>SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>500</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>500</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>200</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>1000</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>1000</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>800</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>100</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>100</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.97</td>
<td>200</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.97</td>
<td>1000</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.97</td>
<td>100</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>1000</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.97</td>
<td>2500</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.98</td>
<td>1000</td>
</tr>
<tr>
<td>SPY</td>
<td>20140102</td>
<td>9:30:00</td>
<td>183.97</td>
<td>400</td>
</tr>
</tbody>
</table>

RV Models: High Frequency – TAQ

Example: Below, we show the first transaction of the AAPL TAQ (Trade and Quote) data (tick-by-tick quote data) on January 2, 2014: 4 AM.

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>DATE</th>
<th>TIME</th>
<th>BID</th>
<th>OFR</th>
<th>BIOSIZ</th>
<th>OFRSIZ</th>
<th>MODE</th>
<th>EX</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:00</td>
<td>455.39</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:00</td>
<td>553.5</td>
<td>558</td>
<td>2</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:01</td>
<td>455.39</td>
<td>561.02</td>
<td>1</td>
<td>2</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:45</td>
<td>551.21</td>
<td>558</td>
<td>1</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:51</td>
<td>552.1</td>
<td>558.4</td>
<td>1</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:51</td>
<td>551.21</td>
<td>558.8</td>
<td>1</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:00:51</td>
<td>552.1</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:01:14</td>
<td>553</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:01:30</td>
<td>553.01</td>
<td>561.02</td>
<td>1</td>
<td>2</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:01:43</td>
<td>553.01</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:01:44</td>
<td>553.05</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:01:49</td>
<td>455.39</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:01:49</td>
<td>553.61</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:02:02</td>
<td>553.05</td>
<td>559</td>
<td>1</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:02:04</td>
<td>455.39</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:02:04</td>
<td>548.28</td>
<td>559</td>
<td>1</td>
<td>1</td>
<td>12T</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:02:33</td>
<td>553.05</td>
<td>558.83</td>
<td>1</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:02:33</td>
<td>555.17</td>
<td>558.83</td>
<td>2</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>20140102</td>
<td>4:03:50</td>
<td>555.2</td>
<td>558.83</td>
<td>5</td>
<td>2</td>
<td>12P</td>
<td></td>
</tr>
</tbody>
</table>
Example (continuation): We read SPY trade data for 2014 Jan.

```r
> HF_da <- read.csv("c:/Financial Econometrics/SPY_2014.csv", head=TRUE, sep=";")
> summary(HF_da)

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>DATE</th>
<th>TIME</th>
<th>PRICE</th>
<th>SIZE</th>
<th>G127</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPY:6800865</td>
<td>20140102 9:30:00</td>
<td>21436</td>
<td>.1766</td>
<td>1</td>
<td>Min 0</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>20140110 16:00:00</td>
<td>11352</td>
<td>.1789</td>
<td>1st Qu.</td>
<td>100 1st Qu.</td>
</tr>
<tr>
<td>Median</td>
<td>20140121 9:30:01</td>
<td>5922</td>
<td>.1826</td>
<td>Median</td>
<td>100 Median</td>
</tr>
<tr>
<td>Mean</td>
<td>20140119 15:59:59</td>
<td>4090</td>
<td>.1814</td>
<td>Mean</td>
<td>337 Mean</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>20140128 15:59:55</td>
<td>3198</td>
<td>.1835</td>
<td>3rd Qu.</td>
<td>300 3rd Qu.</td>
</tr>
<tr>
<td>Max.</td>
<td>20140131 15:50:00</td>
<td>2916</td>
<td>.1892</td>
<td>Max</td>
<td>4715350 Max</td>
</tr>
<tr>
<td>(Other)</td>
<td>6751951</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CORR</th>
<th>COND</th>
<th>EX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.0e+00 @</td>
<td>3351783 T</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.0e+00 F</td>
<td>2888182 P</td>
</tr>
<tr>
<td>Median</td>
<td>0.0e+00 Z</td>
<td>524409</td>
</tr>
<tr>
<td>Mean</td>
<td>1.9e-04 O</td>
<td>18057</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>4.8e-04 K</td>
<td>9098</td>
</tr>
<tr>
<td>Max.</td>
<td>1.2e+01 6</td>
<td>356539</td>
</tr>
<tr>
<td>(Other)</td>
<td>1194</td>
<td>(Other) 777625</td>
</tr>
</tbody>
</table>

Example (continuation): Using the SPY trade data, we calculate using 5'-returns a daily realized volatility for the first 4 days in 2014 (2014:01:02 - 2014:01:07). Originally, we have $T = 1,048,570$.

```r
summary(HF_da)
pt <- as.POSIXct(paste(HF_da$DATE, HF_da$TIME), format="%Y%m%d %H:%M:%S")
library(xts)
hf_1 <- xts(x=HF_da, order.by = pt) # Define a specific time series data set
spy_p <- as.numeric(hf_1$PRICE) # Read price data as numeric
T <- length(spy_p)
spy_ret <- log(spy_p[-1]/spy_p[-T])
plot(spy_ret, type="l", ylab="Return", main="Tick by Tick Return (2014:01:02 - 2014:01:07)"
mean(spy_ret)
sd(spy_ret)
```

RV Models: High Frequency – TAQ
Example (continuation): We plot the tick-by-tick data.

Very noisy data, with lots of “jumps”:
Mean tick by tick return: -3.7365e-09
Tick-by-tick SD: 6.3163e-05

Example (continuation): For the whole month of January 2014:

```r
> mean(spy_ret)
[1] -4.796933e-09
> sd(spy_ret)
[1] 7.804991e-05
```
Example (continuation): We plot the autocorrelogram for the TAQ SPY data:

```r
> acf_spy_raw

Autocorrelations of series 'spy_ret', by lag

                         0      1   2   3   4   5   6    7   8   9  10
 1.000 -0.469 -0.013 -0.010 0.014 -0.008 0.000 -0.002 -0.001 0.000 0.000

Note: We have only a significant autocorrelation, the 1st-order autocorrelation: -0.459.
```

Example (continuation): We aggregate the tick-by-tick data in 5’ intervals using the function `aggregateTrades` in the R package `highfrequency`. It needs as an input an xts object (hf_1, for us).

```r
library(highfrequency)
spy_5 <- aggregateTrades(hf_1, on = "minutes", k = 5,
marketOpen = "09:30:00", marketClose = "16:00:00", tz = "GMT")

spy_5_p <- as.numeric(spy_5$PRICE)
T <- length(spy_5_p)
spy_5_ret <- log(spy_5_p[-1]/spy_5_p[-T])
plot(spy_5_ret, type="l", ylab="Return", main="5-minute Return (2014:01:02 - 2014:01:07)")
```
RV Models: High Frequency – TAQ

Example (continuation): We plot the 5-minute return data. Smoother, easier to read.

5-minute Return (2014:01:02 - 2014:01:07)

RVolt\textsubscript{2014:01:02} = 0.0053344
RVolt\textsubscript{2014:01:03} = 0.0043888
RVolt\textsubscript{2014:01:04} = 0.0059836
RVolt\textsubscript{2014:01:05} = 0.0052772

Example (continuation): We plot the autocorrelogram for the 5' TAQ SPY data:

> acf_spy_5 <- acf(spy_5_ret, main = "5-minute SPY Data: January 2014")
> acf_spy_5

Autocorrelations of series 'spy_ret', by lag

0 1 2 3 4 5 6 7 8 9 10
1.000 -0.105 -0.024 -0.104 0.018 0.147 0.016 -0.024 -0.088 0.048 0.037

Note: We have a negative 1st-order autocorrelation: -0.105, thought not significant. However, the autocorrelation of order 5 is significant.
**Example (continuation):** We plot the 10-minute return data. Smoothing increases.

![10-minute Return (2014:01:02 - 2014:01:07)](image)

\[
\begin{align*}
RV_{\text{Vol}}_{2014:01:02} &= 0.005478294 \\
RV_{\text{Vol}}_{2014:01:03} &= 0.004256046 \\
RV_{\text{Vol}}_{2014:01:04} &= 0.006190508 \\
RV_{\text{Vol}}_{2014:01:05} &= 0.005145601
\end{align*}
\]

---

**Example (continuation):** We plot the autocorrelogram for the 10’ TAQ SPY data:

![10-minute SPY Data: January 2014](image)

**Note:** Now, none of the autocorrelations is significant. The 10-minute returns look independent.
RV Models: R Script

**Example:** R script to compute realized volatility

```r
MSCI_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/MSCI_daily.csv", head=TRUE, sep=",")
x_us <- MSCI_da$USAT <- length(x_us)
us_r <- log(x_us[-1]/x_us[-T])
x <- us_r # US log returns from MSCI USA Index
T <- length(x)
rvs=NULL # create vector to fill with RV
i <- 1
k <- 21 # k: observations per period
while (i < T-k) {
s2 <- sum(x[i:(i+k)]^2) # realized variance
i <- k + i
rvs <- rbind(rvs,s2)
}
rvol <- sqrt(rvs) # realized volatility
mean(rvol) # mean
sd(rvol) # variance
```

RV Models: Monthly RV From Daily Data

**Example:** Using daily data we calculate 1-mo Realized Volatility ($k = 21$ days) for log returns for the MSCI (1970: Jan – 2020: Oct).

US MSCI Monthly Stock Return RVol

> mean(cvol) # average monthly Rvol in the sample
[1] 0.04326531

⇒ very close to monthly S&P Volatility: 4.49%

> sd(cvol) # standard deviation of monthly Rvol in the sample
[1] 0.02592653

⇒ dividing by $\sqrt{T}$ we get the SE = 0.001 (very small)
RV Models: Log Rules

- The log approximations rules for the variance and SD are used to change frequencies for the RV and RVOL. For example, suppose we are calculating RV based on frequency $j$, $RV_t=j$. Using the $j$-period $RV_t=j$, we can compute the annualized variance as:

$$RV_t^j = J \times RV_t=j$$

The $RVOL_t^j$ is the square root of $RV_t=j$.

Example: We calculated using 10’ data the daily realized variance, $RV_{t=daily}$. Then, the annual variance can be calculated as

$$RV_{t=annual} = 260 \times RV_{t=daily}$$

where 260 is the number of trading days in the year. The annualized RVOL is the squared root of $RV_{annual}$:

$$RVOL_{t=annual} = sqrt(260) \times RVOL_{t=daily}$$

We can use time series models —say, an ARIMA model— for $RV_t$ to forecast daily volatility.
Example: Using daily data we calculate 3-mo Realized Volatility \( (k = 66 \text{ days}) \) for log returns for the MSCI (1970: March – 2020: Oct).

\[
\text{mean(rvol)} \quad \# \text{average monthly Rvol in the sample}
\]

\[[1] \quad 0.07725361
\]

\[
\text{sd(rvol)} \quad \# \text{standard deviation of monthly Rvol in the sample}
\]

\[[1] \quad 0.02592653
\]

RV Models: Quarterly RV From Daily Data

- Under some conditions (bounded kurtosis and autocorrelation of squared returns less than 1), \( RV \) is consistent and m.s. convergent.
- Realized volatility is a measure. It has a distribution.
- For returns, the distribution of RV is non-normal (as expected). It tends to be skewed right and leptokurtic. For log returns, the distribution is approximately normal.
- Daily returns standardized by RV measures are nearly Gaussian.
- RV is highly persistent.
- The key problem is the choice of sampling frequency (or number of observations per day).
Realized Volatility (RV) Models - Properties

• The key problem is the choice of sampling frequency (or number of observations per day).

— Bandi and Russell (2003) propose a data-based method for choosing frequency that minimizes the MSE of the measurement error.
— Simulations and empirical examples suggest optimal sampling is around 1-3 minutes for equity returns.

RV Models - Variation

• Another method: AR model for volatility:

\[ |\varepsilon_t| = \alpha + \gamma |\varepsilon_{t-1}| + \nu_t \]

The \( \varepsilon_t \) are estimated from a first step procedure - i.e., a regression. Asymmetric/Leverage effects can also be introduced.

OLS estimation possible. Make sure that the variance estimates are positive.
Other Models - Parkinson’s (1980) estimator

• The Parkinson’s (1980) estimator:

\[ s_t^2 = \left( \sum_t (\ln(H_t) - \ln(L_t))^2 / (4\ln(2)T) \right), \]

where \( H_t \) is the highest price and \( L_t \) is the lowest price.

• There is an RV counterpart, using HF data: Realized Range (RR):

\[ RR_t = \left( \sum_j [100 * (\ln(H_{t,j}) - \ln(L_{t,j}))^2] / (4\ln(2)) \right), \]

where \( H_{t,j} \) and \( L_{t,j} \) are the highest and lowest price in the \( j^{th} \) interval.

• These “range” estimators are very good and very efficient.


Stochastic volatility (SV/SVOL) models

• Now, instead of a known volatility at time \( t \), like ARCH models, we allow for a stochastic shock to \( \sigma_t, \eta_t \) or \( \nu_t \):

\[ \sigma_t = \omega + \beta_1 \sigma_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \]

Or using logs:

\[ \log \sigma_t = \omega + \beta_1 \log \sigma_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma_\nu^2) \]

• The difference with ARCH models: The shocks that govern the volatility are not necessarily the shocks to the mean process, \( \varepsilon_t \)’s.

• Usually, the standard model centers log volatility around \( \omega \):

\[ \log \sigma_t = \omega + \beta_1 (\log \sigma_{t-1} - \omega) + \nu_t, \]

Then,

\[ \mathbb{E}[\log(\sigma_t)] = \omega \]

\[ \text{Var}[\log(\sigma_t)] = \sigma^2 = \sigma_\nu^2 / (1 + \beta_1^2) \]

\[ \Rightarrow \text{Unconditional distribution: } \log(\sigma_t) \sim N(\omega, \sigma^2) \]
Stochastic volatility (SV/SVOL) models

- We have 3 SVOL parameters to estimate: $\varphi = (\omega, \beta, \sigma^2_v)$.

- Like ARCH models, SV models produce returns with kurtosis $> 3$ (and, also, positive autocorrelations between squared excess returns):
  \[
  \text{Var}[r_t] = \mathbb{E}[(r_t - \mathbb{E}[r_t])^2] = \mathbb{E}[\sigma_v^2 z_t^2] = \mathbb{E}[\sigma_v^2] \mathbb{E}[z_t^2] \\
  = \mathbb{E}[\sigma_v^2] = \exp(2\omega + 2\kappa^2) \quad \text{(property of log normal)}
  \]
  \[
  \text{kurt}[r_t] = \mathbb{E}[(r_t - \mathbb{E}[r_t])^4] / \left( \left( \mathbb{E}[(r_t - \mathbb{E}[r_t])^2] \right)^2 \right) \\
  = \mathbb{E}[\sigma_v^4] \mathbb{E}[r_t^4] / \left( \left( \mathbb{E}[\sigma_v^2] \right)^2 (\mathbb{E}[z_t^2])^2 \right) \\
  = 3 \exp(4\omega + 8\kappa^2) / \exp(4\omega + 4\kappa^2) = 3 \exp(4\kappa^2) > 3!
  \]

- Estimation:
  - Bayesian: Using MCMC methods (mainly, Gibbs sampling). Modern approach.

Stochastic volatility (SV/SVOL) models

- The Bayesian approach takes advantage of the idea of hierarchical structure:
  - $f(y \mid h_t)$ (distribution of the data given the volatilities)
  - $f(h_t \mid \varphi)$ (distribution of the volatilities given the parameters)
  - $f(\varphi)$ (distribution of the parameters)

Algorithm: MCMC (JPR (1994).)
Augment the parameter space to include $h_t$.
Using a proper prior for $f(h_t, \varphi)$ MCMC methods provides inference about the joint posterior $f(h_t, \varphi \mid y)$. We'll cover this topic in Lecture 17.