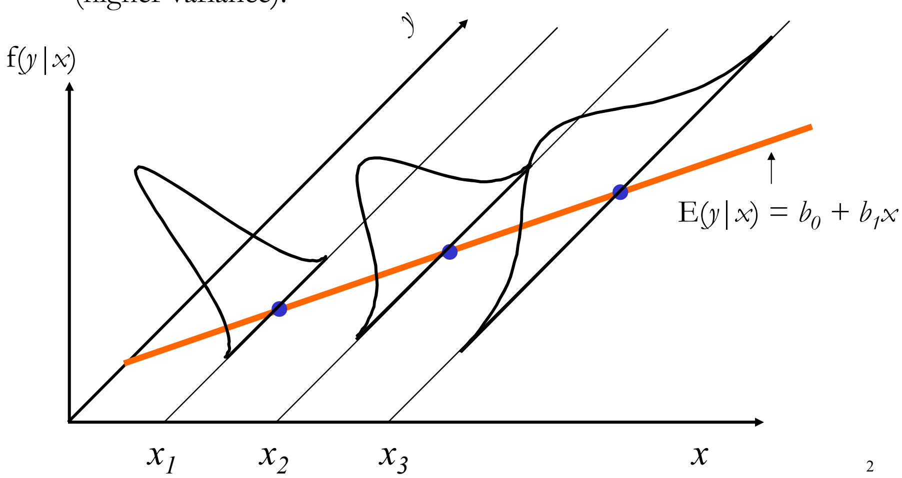
Lecture 12
Heteroscedasticity

(for private use, not to be posted/shared online)

1

Heteroscedasticity

• Assumption (**A3**) is violated in a particular way: **ε** has unequal variances, but ε_i and ε_j are still not correlated with each other. Some observations (lower variance) are more informative than others (higher variance).



2

Heteroscedasticity

• Now, we have the CLM regression with hetero-(different) scedastic (variance) disturbances.

(**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3'**) $\mathrm{Var}[\varepsilon_i] = \sigma^2\,\omega_i$, $\omega_i > 0$. (CLM $\Rightarrow \omega_i = 1$, for all i.)

(**A4**) $\mathbf{X}$ has full column rank – rank($\mathbf{X}$)=k–, where $T \geq k$.

• Popular normalization: $\sum_{i=1}^{T} \omega_i = 1$. (A scaling, absorbed into σ^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

$\omega_i / \sum_{i=1}^{T} \omega_i \rightarrow 0$ as $T \rightarrow \infty$. -i.e., no single observation becomes dominant.

$(1/T) \sum_{i=1}^{T} \omega_i \rightarrow$ some stable constant. (Not a plim!)

GR Model and Testing

• Implications for conventional OLS and hypothesis testing:

1. $\mathbf{b}$ is still unbiased.
2. Consistent? We need the more general proof. Not difficult.
3. If plim $\mathbf{b} = \boldsymbol{\beta}$, then 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 $s^2(\mathbf{X'X})^{-1}$?
A: It depends on $\mathbf{X'\Omega X} - \mathbf{X'X}$. If they are nearly the same, the OLS covariance matrix will give OK inferences.

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

$$\frac{1}{T}\sum_{i=1}^{T} \omega_i\, x_i^2 \xrightarrow{p} E[x^2] \qquad \text{–just like } \frac{1}{T}\sum_{i=1}^{T} x_i^2.$$

• Remark: For the heteroscedasticity to be a significant issue for estimation and inference by OLS, the weights must be correlated with $\boldsymbol{x}$ 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 σ_i^2 may be related to some explanatory variables $z_1, \ldots, z_j$ and/or $z_1^2, \ldots, z_j^2$.

Examples:

1. Following the *error-learning models*, as people learn, their errors of behavior become smaller over time. Then, σ_i^2 is expected to decrease.6

2. As data collecting techniques improve, σ_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, σ_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 the following specifications:
– H1 : σ_t^2 is a function of past ε_t^2 and past σ_t^2 (ARCH models).

– H2 : σ_t^2 increases monotonically with one (or several) exogenous variable(s) ($z_1, \ldots, z_j$).

– H3 : σ_t^2 decreases monotonically with *Size* (Market Cap).

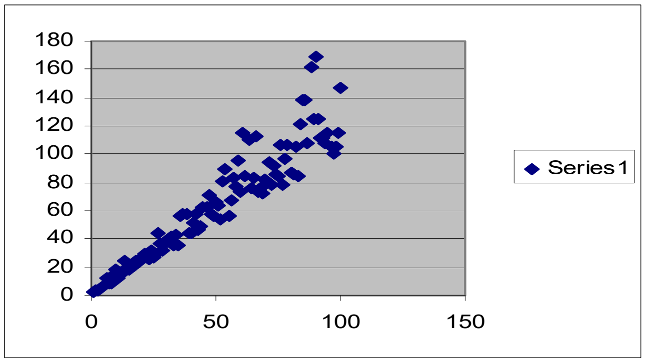
– H4 : σ_t^2 is the same within p subsets of the data but differs across the subsets (*grouped heteroscedasticity*). This specification allows for structural breaks.

• 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.



9

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: $E(\varepsilon_i^2 \mid x_1, x_2, \ldots, x_k) = \sigma^2$

• H_1 and the structure of the test depend on what we consider the drivers of σ_i^2 – i.e., in the previous examples: H1, H2, H3, H4, etc

• The key is whether $E[\varepsilon_i^2] = \sigma_i^2$ is related to $\boldsymbol{X}$ and/or $\boldsymbol{X}^2$. Suppose we suspect a particular independent variable, say $\boldsymbol{x}_j$, is driving σ_i^2:

$$\sigma_i^2 = f(\boldsymbol{x}_j)$$

• Then, a simple test: Check the RSS for large values of $\boldsymbol{x}_j$, and the RSS for small values of $\boldsymbol{x}_j$. This is the Goldfeld-Quandt (GQ) test.

Testing for Heteroscedasticity

• **The Goldfeld-Quandt test**

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

H_1: $\sigma_i^2 = f(\boldsymbol{x}_j)$

• Easy to compute:

– **Step 1**. Arrange the data from small to large values of the independent variable suspected of causing heteroscedasticity, $\boldsymbol{x}_j$.

– **Step 2**. Run two separate regressions, one for small values of $\boldsymbol{x}_j$ and one for large values of $\boldsymbol{x}_j$, omitting ***d*** middle observations (***d*** ≈ 20%). Get the RSS for each regression: RSS_1 for small values of $\boldsymbol{x}_j$ and RSS_2 for large $\boldsymbol{x}_j$'s.

– **Step 3**. Calculate the F ratio

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

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 ω_i's, this test is not very useful.

• But, the GQ test is a popular to 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

```
> 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
```

⇒ cannot reject H_0 at 5% level.

• GE returns

```
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
```

⇒ reject H_0 at 5% level.

23

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:

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

We have two models:

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

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

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

$$\ln L_U = -\frac{T}{2}[\ln(2\pi)+1] - \sum_{i=1}^{g}\frac{T_i}{2}\ln\hat{\sigma}_i^2;\quad \hat{\sigma}_i^2 = \tfrac{1}{T_i}(y_i - X_i b)'(y_i - X_i b)$$

Testing for Heteroscedasticity: LR Test

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

$$LR = 2(\ln L_U - \ln L_R) = T \ln \hat{\sigma}^2 - \sum_{i=1}^{g} T_i \ln \hat{\sigma}_i^{\,2} \xrightarrow{a} \chi^2_{g-1}$$

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

• Using specific functions for σ_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: $E(\varepsilon_i^2 \,|\, x_1, x_2, \ldots, x_k) = \sigma^2$ against an H_1 with a general functional form.

• Recall the central issue is whether $E[\varepsilon_i^2] = \sigma_i^2$ is related to $\boldsymbol{X}$ and/or $\boldsymbol{X}^2$. Then, a simple strategy is to use OLS residuals to estimate disturbances and look for relationships between e_i^2 and x_i and/or x_i^2.

• Suppose that the relationship between ε^2 and $\mathbf{X}$ is linear:

$$\varepsilon_i^2 = \boldsymbol{x}_i{}'\boldsymbol{\alpha} + v_i$$

Then, we test: H_0: $\boldsymbol{\alpha} = 0$ against H_1: $\boldsymbol{\alpha} \neq 0$.

• We base the test on how the squared OLS residuals, e, correlate with $\mathbf{X}$.

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 H_0: 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 tests 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:

$$h_i(\alpha_0 + z_{i,1}\,\alpha_1 + z_{i,2}\,\alpha_2 + ... + z_{i,m}\,\alpha_m) = \sigma_i^2$$

• We want to test: H_0: $E(\varepsilon_i^2 \,|\, z_1, z_2, \ldots, z_m) = h_i(\mathbf{z}_i'\boldsymbol{\alpha}) = \sigma^2$

or H_0: $\alpha_1 = \alpha_2 = ... = \alpha_m = 0$ (m restrictions)

• Assume normality. That is, the log likelihood function is:

$$\log L = \textit{constant} + ½\, \Sigma \log \sigma_i^2 - ½\, \Sigma\, \varepsilon_i^2/\sigma_i^2$$

Then, construct an LM test:

$$LM = \mathbf{S}(\boldsymbol{\theta}_R)'\, \mathbf{I}(\boldsymbol{\theta}_R)^{-1}\, \mathbf{S}(\boldsymbol{\theta}_R) \qquad \boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$$

$\mathbf{S}(\boldsymbol{\theta}) = \partial \log L/\partial\theta' = [-\Sigma\sigma_i^{-2}\mathbf{X'}\varepsilon_i\, ; -½\Sigma(\partial h/\partial\boldsymbol{\alpha})\mathbf{z}_i\sigma_i^{-2} + ½\Sigma\sigma_i^{-4}\,\varepsilon_i^2(\partial h/\partial\boldsymbol{\alpha})\mathbf{z}_i]$

$\mathbf{I}(\boldsymbol{\theta}) = E[-\,\partial^2 \log L/\partial\theta\partial\theta']$

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

$$LM = \mathbf{S}(\alpha_0, \mathbf{0})'\, [\mathbf{I}_{22} - \mathbf{I}_{21}\, \mathbf{I}_{11}\, \mathbf{I}_{21}]^{-1}\, \mathbf{S}(\alpha_0, \mathbf{0})$$

Testing for Heteroscedasticity: BP Test

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

$$LM = \mathbf{S}(\alpha_0,\mathbf{0})' [\mathbf{I}_{22} - \mathbf{I}_{21}\ \mathbf{I}_{11}\ \mathbf{I}_{21}]^{-1}\ \mathbf{S}(\alpha_0,\mathbf{0})$$

$$\mathbf{S}(\alpha_0,\mathbf{0}) = -½\ \Sigma_i\ (\partial h/\partial \boldsymbol{\alpha} | \alpha_{0,R},\mathbf{0})\mathbf{z}'\ \sigma_R^{-2} + ½\ \Sigma_i\ \sigma_R^{-4} e_i^2\ (\partial h/\partial \boldsymbol{\alpha}\ | \alpha_{0,R},\mathbf{0})\mathbf{z}'$$

$$= ½\ e_i^2/\sigma_R^{-2}\ (\partial h/\partial \boldsymbol{\alpha}\ | \alpha_{0,R},\mathbf{0})\ \textstyle\sum_{i=1}^{T} \mathbf{z}_i(e_i^2/\sigma_R^2 - 1)$$

$$= ½\ e_i^2/\sigma_R^{-2}\ (\partial h/\partial \boldsymbol{\alpha}\ | \alpha_{0,R},\mathbf{0})\ \textstyle\sum_{i=1}^{T} \mathbf{z}_i w_i$$

$$w_i = e_i^2/\sigma_R^2 - 1$$

$$\mathbf{I}_{22}(\alpha_0,\mathbf{0}) = E[-\ \partial^2 \log L/\partial\ \boldsymbol{\alpha}\partial\ \boldsymbol{\alpha}'] = ½\ [\sigma_R^{-2}\ (\partial h/\partial \boldsymbol{\alpha}\ | \alpha_{0,R},\mathbf{0})]^2 \textstyle\sum_{i=1}^{T} \mathbf{z}_i \mathbf{z}_i'$$

$$\mathbf{I}_{21}(\alpha_0,\mathbf{0}) = \mathbf{0}$$

$$\sigma_R^2 = (1/T)\textstyle\sum_{i=1}^{T} e_i^2 \qquad \text{(MLE of } \sigma \text{ under } H_0\text{)}.$$

Then,

$$LM = ½\ (\textstyle\sum_{i=1}^{T} \mathbf{z}_i w_i)'\ [\sum_{i=1}^{T} \mathbf{z}_i \mathbf{z}_i']^{-1}\ (\sum_{i=1}^{T} \mathbf{z}_i w_i)$$

$$= ½\ \mathbf{W}'\mathbf{Z}\ (\mathbf{Z}'\mathbf{Z})^{-1}\ \mathbf{Z}'\mathbf{W}\ \sim \chi_m^2$$

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

Testing for Heteroscedasticity: BP Test

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

$$LM = ½\ \mathbf{W}'\mathbf{Z}\ (\mathbf{Z}'\mathbf{Z})^{-1}\ \mathbf{Z}'\mathbf{W} = ½\ ESS$$

ESS = Explained SS in regression of w_i ($= e_i^2/\sigma_R^2 - 1$) against $\mathbf{z}_i$.

• Under the usual regularity conditions, and under H_0,

$$\sqrt{T}\ (\boldsymbol{\alpha}_{ML} - \boldsymbol{\alpha}) \xrightarrow{d} N(0, 2\ \sigma^4\ (\mathbf{Z}'\mathbf{Z}/T)^{-1})$$

Then,

$$\text{LM-BP} = (2\ \sigma_R^4)^{-1}\ ESS_e \xrightarrow{d} \chi_m^2.$$

ESS_e = ESS in regression of e_i^2 against $\mathbf{z}_i$.

Since $\sigma_R^4 \xrightarrow{p} \sigma^4 \quad \Rightarrow \text{LM-BP} \xrightarrow{d} \chi_m^2$

Testing for Heteroscedasticity: BP Test

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

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

- Variations:

(1) Glesjer (1969) test. Use absolute values instead of e_i^2 to estimate the varying second moment. Following our previous example,

$$|e_i| = \alpha_0 + \alpha_1\, z_{i,1} + \alpha_2\, z_{i,2} + ... + \alpha_m\, z_{i,m} + v_i$$

(2) Harvey-Godfrey (1978) test. Use $\ln(e_i^2)$. Then, the implied model for σ_i^2 is an exponential model.

$$\ln(e_i^2) = \alpha_0 + \alpha_1\, z_{i,1} + \alpha_2\, z_{i,2} + ... + \alpha_m\, z_{i,m} + v_i$$

Note: Implied model for $\sigma_i^2 = \exp\{\alpha_0 + \alpha_1 z_{i,1} + ... + \alpha_m z_{i,m} + v_i\}$.

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 ε is normal. Koenker (1981) proposed studentizing the statistic LM-BP by

$$\text{LM-S} = (2\,\sigma_R^4)\ \text{LM-BP}/[(\textstyle\sum_{i=1}^{T}(\varepsilon_i^2 - \sigma_R^2))^2/T] \xrightarrow{d} \chi_m^2$$

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/σ_R^2 on the variables $\mathbf{Z}$. (Omitting σ_R^2 from the denominator is OK.)

Testing for Heteroscedasticity: BP Test

- We have the following steps:

- **Step 1**. Run OLS on DGP:

$\mathbf{y} = \mathbf{X}\,\beta + \varepsilon$. – Keep e_i and compute $\sigma_R^2 = \text{RSS}/T$

- **Step 2**. (Auxiliary Regression). Run the regression of e_i^2 on the *m* explanatory variables, $\mathbf{z}$. In our example,

$$e_i^2/\sigma_R^2 = \alpha_0 + \alpha_1\, z_{1,i} + ... + \alpha_m\, z_{m,i} + v_i \quad \text{– Keep } R^2.$$

- **Step 3**. Use the R^2 from Step 2. Let's call it R_{e2}^2. Calculate

$$\text{LM} = T * R_{e2}^2 \xrightarrow{d} \chi_m^2.$$

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:

```
fit_ibm_ff3 <- lm (ibm_x ~ Mkt_RF + SMB + HML)   # Step 1 – OLS in DGP (3-factor FF model)
e_ibm <- fit_ibm_ff3$residuals                    # Step 1 – keep residuals
e_ibm2 <- e_ibm^2                                 # Step 1 – squared residuals
Mkt_RF2 <- Mkt_RF^2
fit <- lm (e_ibm2 ~ Mkt_RF2)                      # Step 2 – Auxiliary regression
Re_2 <- summary(fit_BP)$r.squared                 # Step 2 – keep R^2
LM_BP_test <- Re2 * T
> LM_BP_test                                      # 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
> p_val
[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-value*= **.6168.**

27

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:

```
> bptest(ibm_x ~ Mkt_RF + SMB + HML)     #bptest only allows to test H1:σi²=f(xi=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 ($\chi^2_{[3],.05}$≈**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.

28

Testing for Heteroscedasticity: Example – DIS

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

```
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$residuals                     # Step 1 – keep residuals
e_dis2 <- e_dis^2                                  # Step 2 – squared residuals
fit <- lm (e_dis2 ~ Mkt_RF2)                       # Step 2 – Auxiliary regression
Re_e2 <- summary(fit_BP)$r.squared                 # Step 2 – Keep R^2 from Auxiliary reg
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 ($\chi^2_{[1],.05}$≈**3.84**); with a *p-value* = **.0001.**

26

Testing for Heteroscedasticity: Example – DIS

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

```
SMB2 <- SMB^2
fit <- lm (e_dis2 ~ SMB2)
Re_e2 <- summary(fit_BP)$r.squared
LM_BP_test <- Re_e2 * T
> LM_BP_test
[1] 7.564692
>  p_val <- 1 - pchisq(LM_BP_test, df = 1)  # p-value of LM_test
>  p_val
[1] 0.005952284
```

LM-BP Test: **7.56** ⇒ reject H_0 at 5% level ($\chi^2_{[1],.05}$≈**3.84**); with a *p-value*= **.006.**

27

Testing for Heteroscedasticity: White Test

- Based on the difference between OLS and true OLS variances:

$$\sigma^2 (\mathbf{X'\Omega X} - \mathbf{X'X}) = \mathbf{X'\Sigma X} - \sigma^2\mathbf{X'X} = \sum_{i=1}^{T} (\text{E}[\varepsilon_i^2] - \sigma^2)\boldsymbol{x}_i{}'\boldsymbol{x}_i$$

- Empirical counterpart: $(1/T) \sum_{i=1}^{T} (e_i^2 - s^2)\, \boldsymbol{x}_i{}'\boldsymbol{x}_i$

- We can express each element of the $k(k+1)$ matrix as:

$(1/T) \sum_{i=1}^{T} (e_i^2 - s^2)\, \psi_i{}'$ $\qquad \psi_i$: Kolmogorov-Gabor polynomial

where

$\psi_i = (\psi_{1i}\ \psi_{2i}, \ldots, \psi_{mi})'$ $\qquad \psi_{li} = \psi_{qi}\, \psi_{pi},\ \ p \geq q,\ \ p, q = 1, 2, \ldots, k$

$l = 1, 2, \ldots, m\ \ \&\ \ m = k^*(k-1)/2$

- Let $\mathbf{D}_T = \text{Var}\, [\frac{1}{T}\sum_{i=1}^{T} (e_i^2 - s^2)\, \psi_i{}']$. Then, White LM test:

$$W = [\tfrac{1}{T}\textstyle\sum_{i=1}^{T} (e_i^2 - s^2)\, \psi_i{}']'\, \mathbf{D}_T^{-1}\, [\tfrac{1}{T}\textstyle\sum_{i=1}^{T} (e_i^2 - s^2)\, \psi_i{}'] \xrightarrow{d} \chi^2_m$$

Testing for Heteroscedasticity: White Test

• White heteroscedasticity test:

$$W = [\frac{1}{T}\sum_{i=1}^{T}(e_i^2 - s^2)\,\psi_i']'\,\mathbf{D}_T^{-1}\,[\frac{1}{T}\sum_{i=1}^{T}(e_i^2 - s^2)\,\psi_i'] \xrightarrow{d} \chi^2_m$$

where

$$\mathbf{D}_T = \text{Var}\,[\frac{1}{T}\sum_{i=1}^{T}(e_i^2 - s^2)\,\psi_i']$$

Note: W is asymptotically equivalent to a $T\,R^2$ test, where R^2 is calculated from a regression of e_i^2/σ_R^2 on the ψ_i's.

Testing for Heteroscedasticity: White Test

• Usual calculation of the White test

– **Step 1**. Run OLS on DGP:

$\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$. –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 $(x_j * x_i)$.

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_{e2}^2.

– **Step 3**. Compute the LM statistic: $LM = T * R_{e2}^2 \xrightarrow{d} \chi^2_m$.

Testing for Heteroscedasticity: White Test

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

```
HML2 <- HML^2;
Mkt_HML <- Mkt_RF*HML
Mkt_SMB <- Mkt_RF*SMB
SMB_HML <- SMB*HML
xx2 <- cbind(Mkt_RF2, SMB2, HML2, Mkt_HML, Mkt_SMB, SMB_HML)
fit_ibm_W <- lm(e_ibm2 ~ xx2)                  # Not including original variables OK
r2_e2 <- summary(fit_ibm_W)$r.squared          # Keep R^2 from Auxiliary regression
> r2_e2
[1] 0.0166492
lm_t <- T * r2_e2                              # Compute LM test: R^2 * sample size (T)
> lm_t
[1] 10.93483
df_lm <- ncol(xx2)
qchisq(.95, df = df_lm)
```

LM-White Test: **10.93** ⇒ cannot reject H_0 at 5% level ($\chi^2_{[6],.05}$≈**12.59**).

Testing for Heteroscedasticity: White Test

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

• For **DIS**, we get:

```
fit_dis_W <- lm (e_dis2 ~ xx2)
Re_2W <- summary(fit_dis_W)$r.squared
LM_W_test <- Re_2W * T
> LM_W_test
[1] 25.00148                         ⇒ reject H0 at 5% level (χ2[6],05 ≈ 12.59).
>qchisq(.95, df = df_lm)
[1] 12.59159
>  p_val <- 1 - pchisq(LM_W_test, df = 6)  # p-value of LM_test
>  p_val
[1] 0.0003412389
```

• For **GE**, we get:

LM-White Test: **20.15** (*p-value*=**0.0026**) ⇒ reject H_0 at 5% level.

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):

```
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 H0 at 5% level
```

33

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 BP 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 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 heterosekdasticity-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:

$$\text{Var}[\varepsilon \,|\, \boldsymbol{x}] = \sigma^2\, h(\boldsymbol{x})$$

• The key is to figure out what $h(\boldsymbol{x})$looks like. Suppose that we know h_i. For example, $h_i(\boldsymbol{x}) = x_i^2$. (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 $h_i(\boldsymbol{x}) = x_i^2$. Then, use $1/\sqrt{x_i^2}$ to transform the model:

$$\text{Var}[\frac{\varepsilon_i}{\sqrt{x_i^2}} \,|\, \boldsymbol{x}] = \sigma^2$$

• Thus, if we divide our whole equation by $\sqrt{h_i(\boldsymbol{x})}$ 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 $h_i(\boldsymbol{x})$.

• Typically, we start by assuming a fairly flexible model, such as

$\text{Var}[\varepsilon \mid \boldsymbol{x}] = \sigma^2 \exp(\mathbf{X}\boldsymbol{\delta})$ –make sure $\text{Var}[\varepsilon_i \mid \boldsymbol{x}] > 0$.

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

$\boldsymbol{\varepsilon}^2 = \sigma^2 \exp(\mathbf{X}\boldsymbol{\delta})\, \boldsymbol{v}$ with $E(\boldsymbol{v} \mid \mathbf{X}) = 1$.

Then, if $E(\boldsymbol{v}) = 1$

$$\ln(\boldsymbol{\varepsilon}^2) = \mathbf{X}\boldsymbol{\delta} + \mathbf{u} \qquad (*)$$

where $E(\mathbf{u}) = 0$ and $\mathbf{u}$ is independent of $\mathbf{X}$.

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

Estimation: FGLS

• Now, an estimate of h is obtained as $\hat{h} = \exp(\hat{g})$, 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, $\boldsymbol{e}$. Get $\ln(\boldsymbol{e}^2)$.

(2) Regress $\ln(\boldsymbol{e}^2)$ on all of the independent variables. Get fitted values, $\hat{g}$.

(3) Do WLS using $1/\sqrt{\exp(\hat{g})}$ 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 $\boldsymbol{\varepsilon}$. Under normality (**A5**):

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

• Suppose $\sigma_i^2 = \exp(\alpha_0 + \alpha_1\, z_{1,i} + ... + \alpha_m\, z_{m,i}) = \exp(\boldsymbol{z}_i{}'\boldsymbol{\alpha})$

• Then, the first derivatives of the log likelihood wrt $\boldsymbol{\theta}=(\boldsymbol{\beta}, \boldsymbol{\alpha})$ are:

$$\frac{\partial \ln L}{\partial \beta} = -\sum_{i=1}^{T}\mathbf{x}_i\,\varepsilon_i / \sigma_i^2 = X'\Sigma^{-1}\varepsilon$$

$$\frac{\partial \ln L}{\partial \alpha_i} = -\frac{1}{2}\sum_{i=1}^{T} 1/\sigma_i^2 \exp(z_i{}'\alpha) z_i - (-\frac{1}{2})\sum_{i=1}^{T}\varepsilon_i^2/\sigma_i^4 \exp(z_i{}'\alpha) z_i = \frac{1}{2}\sum_{i=1}^{T} z_i(\varepsilon_i^2/\sigma_i^2 - 1)$$

• 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 :

$$\frac{\partial \ln L^2}{\partial\beta\partial\beta'} = -\sum_{i=1}^{T}\mathbf{x}_i\,\mathbf{x}_i{}'/\sigma_i^2 = X'\Sigma^{-1}X$$

$$\frac{\partial \ln L}{\partial\beta\partial\alpha_i{}'} = -\frac{1}{2}\sum_{i=1}^{T} x_i z_i{}'\varepsilon_i/\sigma_i^2$$

$$\frac{\partial \ln L}{\partial\alpha_i\partial\alpha_i{}'} = -\frac{1}{2}\sum_{i=1}^{T} z_i z_i{}'\varepsilon_i^2/\sigma_i^2$$

• Then,

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

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

$$\theta_{j+1} = \theta_j - H_j^{-1}\, g_j \qquad\qquad g_j = \partial\log L_j/\partial\theta'$$

Estimation: MLE

• We estimate the model using Newton's method:

$$\theta_{j+1} = \theta_j - H_j^{-1}\, g_j \qquad\qquad g_j = \partial\log L_j/\partial\theta'$$

where subscript j stands for values computed under iteration j.

Since H is block diagonal, then

$\beta_{j+1} = \beta_j - (\mathbf{X}'\, \boldsymbol{\Sigma}_j^{-1}\, \mathbf{X})^{-1}\, \mathbf{X}'\, \boldsymbol{\Sigma}_j^{-1}\, \boldsymbol{\varepsilon}_j$

$\alpha_{j+1} = \alpha_j - (½\mathbf{Z}'\mathbf{Z})^{-1}\, [½ \sum_{i=1}^{T} \boldsymbol{z}_i(\varepsilon_{i,j}^2/\sigma_{i.j}^2 - 1)] = \alpha_j - (\mathbf{Z}'\mathbf{Z})^{-1}\, \mathbf{Z}'\boldsymbol{v}$,

where

$$\boldsymbol{v} = (\varepsilon_{i,j}^2/\sigma_{i,j}^2 - 1)$$

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

• We have an iterative algorithm $\quad\Rightarrow$ Iterative FGLS = MLE!

Heteroscedasticity: Log Transformations

• A log transformation of the data, can eliminte (or reduce) a certain type of heteroskedasticity.

- Assume
  - $\mu_t = \mathrm{E}[Z_t]$
  - $\mathrm{Var}[Z_t] = \delta\, \mu_t^2$ (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: Var[*f(X)*] using delta method:

$$Var[f(X)] \approx f'(\theta)^2\, Var[X]$$

• Then, the variance of $\log(Z_t)$ is roughly constant:

$$Var[\log(Z_t)] \approx (1/\mu_t)^2\, Var[Z_t] = \delta$$

ARCH Models

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

$$y_t = E[y_t \mid I_t] + \varepsilon_t, \qquad \varepsilon_t \sim D(0, \sigma^2)$$

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:

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

• 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:

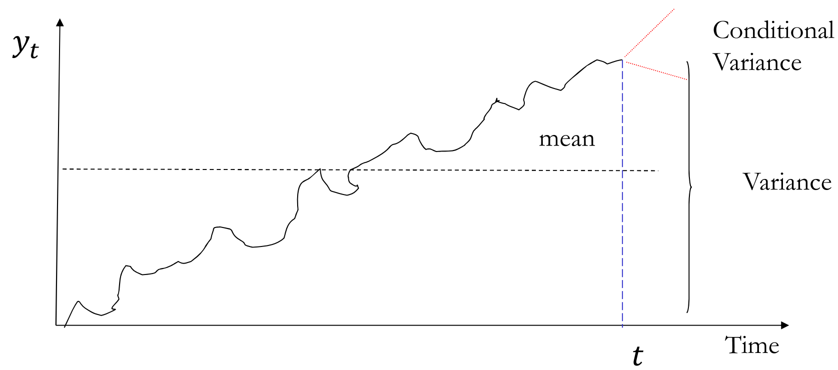
$$\text{Var}_{t-1}[y_t \mid I_{t-1}] = E_{t-1}[y_t - E[y_t]])^2] = E_{t-1}[\varepsilon_t^2]$$

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: 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.

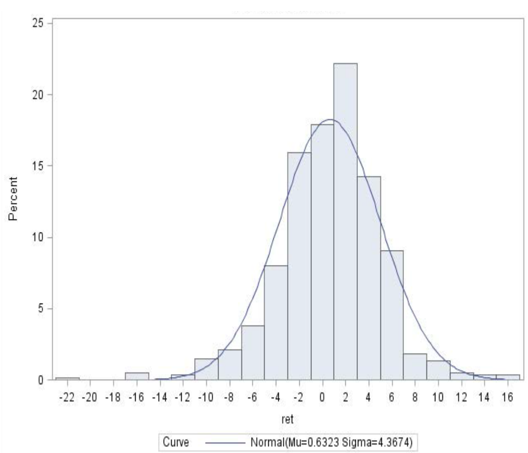
- *Co-movements in volatility* – Ramchand and Susmel (1998): Volatility is positively correlated across markets/assets.

• We need a model that accommodates all these facts.

ARCH Models: Stylized Facts of Asset Returns

• Easy to check leptokurtosis (Stylized Fact #1)

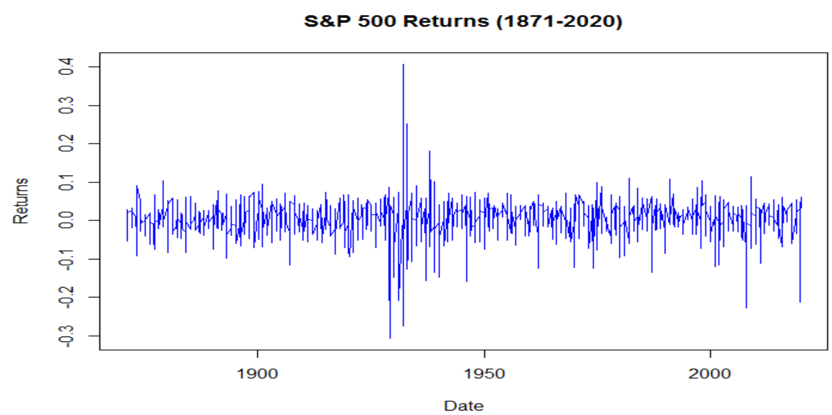
Figure: Descriptive Statistics and Distribution for Monthly S&P500 Returns



	Statistic
Mean (%)	0.626332 (p-value: 0.0004)
Standard Dev (%)	4.37721
Skewness	-0.43764
Excess Kurtosis	2.29395
Jarque-Bera	145.72 (p-value: <0.0001)
AR(1)	0.0258 (p-value: 0.5249)

ARCH Models: Stylized Facts of Asset Returns

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



ARCH Models: Engle (1982)

• We start with assumptions (**A1**) to (**A5**), but with a specific (**A3'**):

$$y_t = \gamma X_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, {\sigma_t}^2)$$

(**A3'**) ${\sigma_t}^2 = Var_{t-1}[\varepsilon_t] = E_{t-1}[\varepsilon_t^2] = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2,$

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 ε_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 ε_t^2's: High (low) past ε_t^2's produce a high (low) ε_t^2 today.

52

ARCH Models: Engle (1982)

• 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, ε_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

$$\omega > 0 \text{ and } \alpha_i > 0 \text{ for all } i.$$

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

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

Robert F. Engle (1942, USA)

53

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 σ^2, we impose another restriction: $(1 - \sum_{i=1}^{q} \alpha_i) > 0$

Example: ARCH(1)

$$Y_t = \beta X_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_t{}^2)$$
$$\sigma_t{}^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 \qquad \Rightarrow \sigma^2 = \frac{\omega}{1 - \alpha_1}$$

• We need to impose restrictions: $\omega > 0$, $\alpha_1 > 0$, & $(1 - \alpha_1) > 0$.

54

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:

$$\kappa(\varepsilon_t) = E[\varepsilon_t^4]/E[\varepsilon_t^2]^2$$

• 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 ~ 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[ε_t]=0) moment:

$$E[\varepsilon_t^4] = E[z_t^4]E[\sigma_t^4]$$

55

ARCH Models: Leptokurtosis

• Then, using Jensen's inequality:

$$E[\varepsilon_t^4] = E[z_t^4]E[\sigma_t^4] > E[z_t^4]E[\sigma_t^2]^2 = E[z_t^4]E[\varepsilon_t^2]^2$$
$$= 3\ E[\varepsilon_t^2]^2$$
$$\kappa(\varepsilon_t) = E[\varepsilon_t^4]/E[\varepsilon_t^2]^2 > 3.$$

where we have used the fact that since E[ε_t] = 0, then E[ε_t^2] = E[σ_t^2]

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

$$\kappa(\varepsilon_t) = \frac{3(1-\alpha^2)}{1-3\alpha^2} \qquad \text{if } 3\alpha^2 < 1.$$

56

ARCH Models: Alternative Representation

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

$$\varepsilon_t = \sqrt{\sigma_t^2}\, z_t$$

where υ_t is *i.i.d.* with mean 0, and 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 σ_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):

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$
$$= \omega + \alpha(L)\varepsilon^2 + \beta(L)\sigma^2$$

GARCH: Bollerslev (1986)

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

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

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: Var[ε_t^2] = $\sigma^2 = \frac{\omega}{1-\alpha_1-\beta_1}$

⇒ Restrictions: ω > 0, α_1> 0, β_1> 0; ($1 - \alpha_1 - \beta_1$) > 0.

• <u>Technical details</u>: This is *covariance stationary* if all the roots of

$$\alpha(\text{L}) + \beta(\text{L}) = 1$$

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 ε_t exist:

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

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

GARCH-X

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

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

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

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.

Examples: Glosten et al. (1993) and Engle and Patton (2001) use 3-mo T-bill rates for modeling stock return volatility. Hagiwara and Herce (1999) use interest rate differentials between countries to model FX return volatility. The US congressional budget office uses inflation in an ARCH(1) model for interest rate spreads.

IGARCH

• Recall the technical detail: The standard GARCH model:

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon^2 + \beta(L)\sigma^2$$

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

$\Rightarrow$ Today's variance remains important for all future forecasts

IGARCH

• 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: $y_t = X_t\gamma + \delta\,\sigma_t^2 + \varepsilon_t$, $\varepsilon_t \sim$ N(0, σ_t^2)

Variance equation: $\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$

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

• Finance intuition says that δ has to be positive and significant. However, in empirical work, it does not work well: δ is not significant or negative.

63

GARCH: Variations – Asymmetric GJR

• GJR-GARCH model – Glosten, Jagannathan & Runkle (JF, 1993):

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{q}\gamma_i\varepsilon_{t-i}^2 * I_{t-i} + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2$$

where I_{t-i} = 1 if $\varepsilon_{t-i} < 0$;
= 0 otherwise.

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

• The GARCH(1,1) version:

$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2 + \gamma_1\varepsilon_{t-1}^2\, I_{t-1} + \beta_1\sigma_{t-1}^2$$

where I_{t-1} = 1 if $\varepsilon_{t-1} < 0$;
= 0 otherwise.

64

GARCH: Variations – Asymmetric GJR

• The GARCH(1,1) version:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 I_{t-1} + \beta_1 \sigma_{t-1}^2$$

When $\varepsilon_{t-1} < 0 \Rightarrow \sigma_t^2 = \omega + (\alpha_1 + \gamma_1)\, \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

$\varepsilon_{t-1} > 0 \Rightarrow \sigma_t^2 = \omega + \alpha_1\, \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

• 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.

65

GARCH: Variations – EGARCH

• EGARCH model – Nelson (Econometrica, 1991).
It models an exponential function for the time-varying variance:

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i (z_{t-i} + \gamma(|z_{t-i}| - E|z_{t-i}|)) + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2)$$

where z is a standardized *i.i.d.* D(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 ($z_{t-i} < 0$) increase σ_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

• Non-linear ARCH model NARCH – Higgins and Bera (1992) and Hentschel (1995).

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.

67

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," $(\varepsilon_{t-i} < \kappa)$:

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

There are two variances:

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

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

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

68

GARCH: Variations – SWARCH

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

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

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

The key is to select a parsimonious representation:

$$\frac{\sigma^2{}_t}{\gamma_{s_t}} = \omega + \sum_{i=1}^{q} \alpha_i \frac{\varepsilon_{t-i}^2}{\gamma_{s_{t-i}}}$$

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

$$\sigma_t^2 = \omega\gamma_1 + \alpha_1 \varepsilon_{t-1}^2 \gamma_1 / \gamma_1, \qquad s_t = 1, s_{t-1} = 1$$
$$\sigma_t^2 = \omega\gamma_1 + \alpha_1 \varepsilon_{t-1}^2 \gamma_1 / \gamma_2, \qquad s_t = 1, s_{t-1} = 2$$
$$\sigma_t^2 = \omega\gamma_2 + \alpha_1 \varepsilon_{t-1}^2 \gamma_2 / \gamma_1, \qquad s_t = 2, s_{t-1} = 1$$
$$\sigma_t^2 = \omega\gamma_2 + \alpha_1 \varepsilon_{t-1}^2 \gamma_2 / \gamma_2, \qquad s_t = 2, s_{t-1} = 2$$

GARCH: Forecasting and Persistence

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

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

Taking expectation at time t

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

Then, by repeated substitutions:

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

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

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}$$

GARCH: Forecasting and Persistence

• Today's forecast in a GARCH(1,1) model for horizon j :

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

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 σ_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:

$$f(y_1, y_2, \dots, y_T; \boldsymbol{\theta}) = f(y_1|\mathbf{x}_1; \boldsymbol{\theta}) f(y_2|y_1, \mathbf{x}_2, \mathbf{x}_1; \boldsymbol{\theta}) f(y_3|y_2, y_1, \mathbf{x}_3, \mathbf{x}_2, \mathbf{x}_1; \boldsymbol{\theta}) \dots$$
$$\dots f(y_T|y_{T-1}, \dots, y_1, \mathbf{x}_{T-1}, \dots, \mathbf{x}_1; \boldsymbol{\theta}).$$

Taking logs:

$$L = \log(f(y_1, y_2, \dots, y_T; \boldsymbol{\theta})))) = \log(f(y_1|\mathbf{x}_1; \boldsymbol{\theta})) + \log(f(y_2|y_1, \mathbf{x}_2, \mathbf{x}_1; \boldsymbol{\theta}))$$
$$+ \dots + \log(f(y_T|y_{T-1}, \dots, y_1, \mathbf{x}_{T-1}, \dots, \mathbf{x}_1; \boldsymbol{\theta}))$$
$$= \sum_{t=1}^{T} \log(f(y_t|Y_{t-1}, X_t; \boldsymbol{\theta}))$$

• We maximize this function with respect to the k mean parameters (γ) and the m variance parameters $(\omega, \alpha_1, \beta_1)$. Let $\boldsymbol{\theta}$ =(γ, $\omega, \alpha_1, \beta_1$). 72

ARCH Estimation: MLE

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

- We have a (k+m x 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, $\boldsymbol{\theta}$)) can be easily implemented.

- Given the AR structure, we will need to make assumptions about σ_0 (and ε_0, ε_1, ..., ε_p if we assume an AR(p) process for the mean).

- Alternatively, we take σ_0 (& ε_0, ε_1, ..., ε_p) as parameters to estimate.

Note: The $\frac{\partial L}{\partial \boldsymbol{\gamma}} = \mathbf{0}$ (k x1 vector of f.o.c.'s) will give us GLS (or WLS).

$$\frac{\partial L}{\partial \boldsymbol{\gamma}} = -\sum_{t=1}^{T} \boldsymbol{X'}_t \varepsilon_t / \sigma^2_{t,MLE} = \sum_{t=1}^{T} \boldsymbol{X'}_t (y_t - \boldsymbol{X}_t \boldsymbol{\gamma}_{MLE}) / \sigma^2_{t,MLE} = 0$$

$$= \sum_{t=1}^{T} \frac{\boldsymbol{X'}_t}{\sigma_{t,MLE}} \left(\frac{y_t}{\sigma_{t,MLE}} - \frac{\boldsymbol{X}_t}{\sigma_{t,MLE}} \boldsymbol{\gamma}_{MLE}\right) = 0$$

ARCH Estimation: MLE

• MLE = GLS for mean parameters; y_t & $\boldsymbol{X}_t$ weighted by 1/ $\sigma_{t,MLE}$.

• If the conditional density is well specified and θ_0 belongs to Ω, then

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

• 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 θ_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: $y_t = \boldsymbol{X}_t\boldsymbol{\gamma} + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma_t^2)$

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 $\mathcal{L}$ (the joint pdf):

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

We take logs to form the log likelihood, $L = \log \mathcal{L}$:

$$L = \sum_{t=1}^{T} \log(f_t) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T} \varepsilon^2{}_t/\sigma_t^2$$

Then, we maximize L with respect to $\boldsymbol{\theta} = (\gamma, \omega, \alpha_1)$ the function L.

75

ARCH Estimation: MLE – ARCH(1)

Example (continuation): ARCH(1) model.

$$L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \log(\omega + \alpha_1\varepsilon_{t-1}^2) - \frac{1}{2}\sum_{t=1}^{T} \frac{\varepsilon_t^2}{(\omega + \alpha_1\varepsilon_{t-1}^2)}$$

Taking derivatives with respect to $\boldsymbol{\theta} = (\omega, \alpha_1, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a vector of k mean parameters:

$$\frac{\partial L}{\partial \omega} = -\frac{1}{2}\sum_{t=1}^{T} \frac{1}{\omega + \alpha_1\varepsilon_{t-1}^2} - (-\frac{1}{2}) \sum_{t=1}^{T} \frac{\varepsilon_t^2}{(\omega + \alpha_1\varepsilon_{t-1}^2)^2}$$

$$\frac{\partial L}{\partial \alpha_1} = -\frac{1}{2}\sum_{t=1}^{T} \frac{\varepsilon_{t-1}^2}{\omega + \alpha_1\varepsilon_{t-1}^2} - (-\frac{1}{2}) \sum_{t=1}^{T} \frac{\varepsilon_t^2\varepsilon_{t-1}^2}{(\omega + \alpha_1\varepsilon_{t-1}^2)^2}$$

$$\frac{\partial L}{\partial \boldsymbol{\gamma}} = -\sum_{t=1}^{T} \boldsymbol{X}_t\varepsilon_t/\sigma_t^2$$ (kx1 vector of derivatives)

ARCH Estimation: MLE – Example (in R)

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

```
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 – rho*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
}
```

ARCH Estimation: MLE – Example (in R)

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

```
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
logL_g11
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))
plot.ts(chf_usd)                                     # time series plot of data
```

78

ARCH Estimation: MLE – Example (in R)

Example 1 (continuation):

```
> 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.
eigen() decomposition
$values                             # Eigenvalues: if positives => Hessian is pd
[1] 1.687400e+08 6.954454e+05 7.200084e+03 5.120984e+02 2.537958e+02

$vectors
              [,1]          [,2]          [,3]          [,4]          [,5]
[1,]  4.265907e-05  9.999960e-01 -0.0011397586  0.0018331957 -0.0018541203
[2,] -3.333961e-06 -2.188159e-03 -0.0010048203  0.9769058449 -0.2136566699
[3,]  9.999998e-01 -4.223001e-05 -0.0003544245  0.0001291633  0.0005770707
[4,] -3.599974e-06 -1.702277e-03 -0.8603563865 -0.1097470278 -0.4977344477
[5,] -6.893837e-04  6.416141e-04 -0.5096905472  0.1833226197  0.8405994743
```

79

ARCH Estimation: MLE – Example (in R)

Example 1 (continuation):

```
> 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
```

80

ARCH Estimation: MLE – Example (in R)

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

$e_{f,t} = [\log(S_t) - \log(S_{t-1})] = a_0 + a_1\, e_{f,t-1} + \varepsilon_t, \qquad \varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2)$

$\sigma_t^2 = \omega + \alpha_1\, \varepsilon_{t-1}^2 + \beta_1\, \sigma_{t-1}^2$

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

The estimated model for $e_{f,t}$ is given by:

$e_{f,t} = \underset{(.0012)}{-0.00211} + \underset{(0.044)}{0.02600}\, e_{f,t-1},$

$\sigma_t^2 = \underset{(0.00096)^*}{0.00012} + \underset{(0.050)^*}{0.19003}\, \varepsilon_{t-1}^2 + \underset{(0.040)^*}{0.71007}\, \sigma_{t-1}^2.$

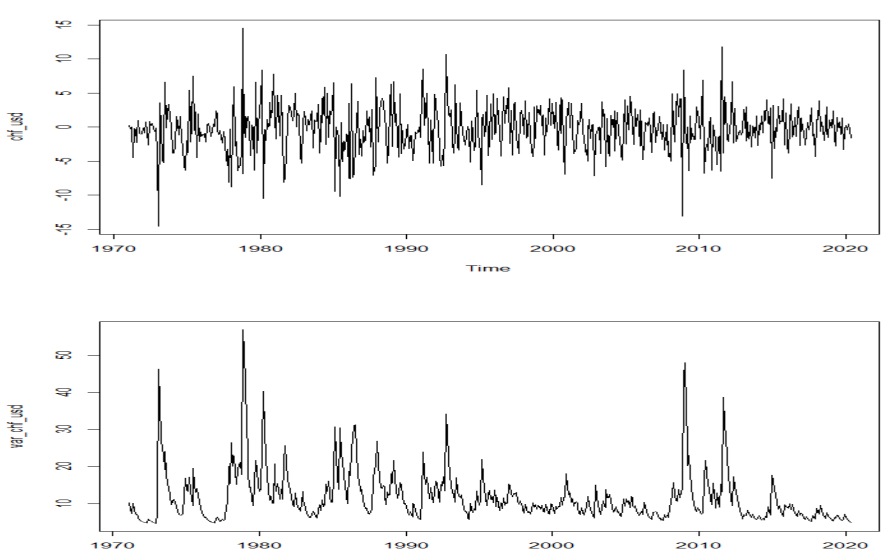
Unconditional $\sigma^2 = 0.00012 / (1 - 0.19003 - 0.71007) = 0.001201201$

Log likelihood: 1745.197

Note: $\alpha_1 + \beta_1 = .90 < 1$. (Persistent.)

81

ARCH Estimation: MLE – Example (in R)



82

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})] = $a_0 + a_1 r_{t-1} + \varepsilon_t$, $\quad \varepsilon_t \mid I_{t-1} \sim N(0, \sigma_t^2)$
$\sigma_t^2 = \omega + \alpha_1\, \varepsilon_{t-1}^2 + \beta_1\, \sigma_{t-1}^2$

The estimated model for s_t is given by:

$$r_t = \underset{(.08)^*}{0.338} + \underset{(0.025)^*}{0.278}\, r_{t-1},$$

$$\sigma_t^2 = \underset{(0.151)^*}{0.756} + \underset{(0.017)^*}{0.126}\, \varepsilon_{t-1}^2 + \underset{(0.021)^*}{0.826}\, \sigma_{t-1}^2.$$

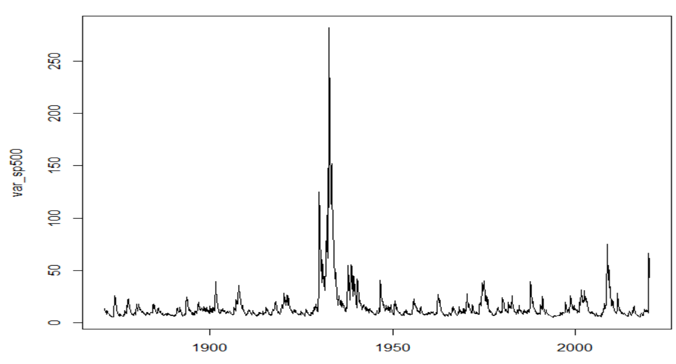
Unconditional σ^2 = 0.756 /(1 - 0.126 - 0.826) = 15.4630
Log likelihood: 4795.08

Note: $\alpha_1 + \beta_1$ = .952 < 1. (Very persistent.)

83

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 $E[\ln(\alpha_1 z_t^2 + \beta_1)] < 0$, 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, θ_1, and variance parameters, θ_2.

Then, $\delta\mu_t(\theta)/\delta\theta_2 = 0$ and, although, $\delta\sigma_t(\theta)/\delta\theta_1 \neq 0$, 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

$$l_t = \ln(\Gamma(0.5(v+1))\Gamma(0.5v)^{-1}(v-2)^{-1/2}(1+z_t(v-2)^{-1})^{-(v+1)/2}) - 0.5\ln(\sigma^2{}_t)$$

where Γ is the gamma function and v is the degrees of freedom.
As υ→∞, this tends to the normal distribution.

• GED Distribution - Nelson (1991)

ARCH Estimation: GMM

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

$$(1) - E[m_1] \equiv E[\mathbf{x_t}'(y_t - \mathbf{x_t}\boldsymbol{\gamma})] = 0$$
$$(2) - E[m_2] \equiv E[\boldsymbol{\varepsilon}_\mathbf{t}^\mathbf{2}{}'(\varepsilon_t^2 - \sigma_t^2)] = 0$$
$$(3) - E[m_3] \equiv E[\varepsilon_t^2 - \omega/(1 - \alpha_1 - ... - \alpha_q)] = 0$$

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

• GMM objective function:

$$Q(\mathbf{X},\ \mathbf{y};\ \boldsymbol{\theta}) = \hat{E}[\mathbf{m}(\boldsymbol{\theta};\mathbf{X},\ \mathbf{y})]'\ \mathbf{W}\ \hat{E}[\mathbf{m}(\boldsymbol{\theta};\mathbf{X},\ \mathbf{y})]$$

where

$$\hat{E}[\mathbf{m}(\boldsymbol{\theta};\mathbf{X},\ \mathbf{y})] \equiv [\hat{E}[\mathbf{m}_1]'\ \hat{E}[\mathbf{m}_2]'\ \hat{E}[\mathbf{m}_3]'\ ]'$$

ARCH Estimation: GMM

• $\boldsymbol{\gamma}$ has k free parameters; α has q free parameters. Then, we have $J = k + q + 1$ parameters. Note that:

$\mathbf{m}(\boldsymbol{\theta}; \mathbf{X}, \boldsymbol{y})$ has $J = k + q + 1$ equations.

Dimensions: Q is 1x1; E[$\mathbf{m}(\boldsymbol{\theta}; \mathbf{X}, \boldsymbol{y})$] is Jx1; $\mathbf{W}$ is JxJ.

• Problem is *over-identified*: more equations than parameters so cannot solve E[$\mathbf{m}(\boldsymbol{\theta}; \mathbf{X}, \boldsymbol{y})$]=0, exactly.

• Choose a weighting matrix $\mathbf{W}$ for objective function and minimize using numerical optimization.

• Optimal weighting matrix: $\mathbf{W} = \{E[\mathbf{m}(\boldsymbol{\theta}; \mathbf{X}, \boldsymbol{y})]\ E[\mathbf{m}(\boldsymbol{\theta}; \mathbf{X}, \boldsymbol{y})]'\}^{-1}$.

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

where $\mathbf{D} = \delta E[\mathbf{m}(\boldsymbol{\theta}; \mathbf{X}, \boldsymbol{y})]/\delta\theta'$ –expressions evaluated at θ_{GMM}.

ARCH Estimation: Testing

• Standard BP test, with auxiliary regression given by:

$$e_t^2 = \alpha_0 + \alpha_1\, e_{t-1}^2 + ... + \alpha_q\, e_{t-q}^2 + v_t \qquad \text{-Keep } R^2.$$

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

Then, the LM test is $(T - q)^* R^2 \xrightarrow{d} \chi_q^2$ – 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. θ_0 is often on the boundary of Ω. (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), ω and β_1 are not jointly identified. See Davies (1977).

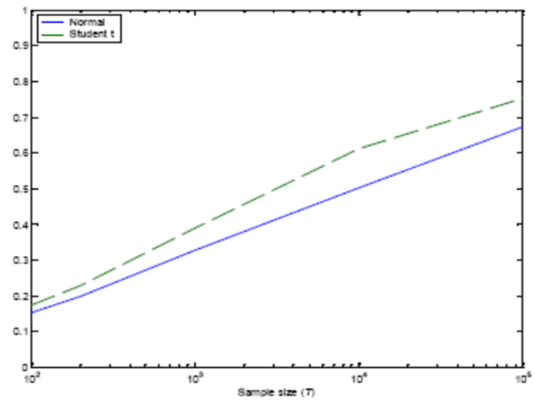
• Ignoring ARCH

- You suspect y_t has an AR structure: $y_t = \gamma_0 + \gamma_1\, y_{t-1} + \varepsilon_t$

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%.



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** ⇒ reject H_0 at 5% level ($\chi^2_{[1],05} \approx 3.84$), the usual result for financial time series.

GARCH: Forecasting and Persistence (Again)

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

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

Taking expectation at time t

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

Then, by repeated substitutions:

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

• When $\alpha_1 + \beta_1 = 1$, today's volatility affect 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 σ_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** (=$\sigma_{2020:9}$ = sqrt(**0.00367**) = 6.1%)

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

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

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

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

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

$$E_{2020:9}[\sigma_{2020:12}^2] = 0.00012 * \{1 + (0.9) + (0.9)^2\} + 0.00367 * (0.9)^3$$
$$= 0.00300063$$

96

GARCH: Forecasting and Persistence

Example 1 (continuation):
We forecast volatility for March 2021:

$$E_{2020:6}[\sigma^2_{2021:03}] = \mathbf{0.00012} * \{1 + (\mathbf{0.9}) + (\mathbf{0.9})^2 + \ \ldots\ + (\mathbf{0.9})^5\} +$$
$$+\ \mathbf{0.00367} * (\mathbf{0.9})^6 = 0.002512659$$

<u>Remark</u>: We observe that as the forecast horizon increases ($j \rightarrow \infty$), the forecast reverts to the unconditional variance:

$$\omega/(1 - \alpha_1 - \beta_1) = \mathbf{0.00012}/(1 - \mathbf{0.9}) = 0.0012$$

$\Rightarrow \sigma$ = sqrt(0.0012) = 0.0346 (3.46% ≈ close to sample SD = **3.36%**)

97

GARCH: Forecasting and Persistence

Example 2: On August 2020, we forecast the December's variance for the S&P500 changes. Recall we estimated σ^2_t:

$$\sigma^2_t = \mathbf{0.756} + \mathbf{0.125}\ \varepsilon^2_{t-1} + \mathbf{0.826}\ \sigma^2_{t-1}.$$

getting $\sigma^2_{2020:8}$ = **43.037841**

We based the $\sigma^2_{2020:12}$ forecast on:

$$E_t[\sigma^2_{t+j}] = \omega * [\textstyle\sum_{i=0}^{j-1} (\alpha_1 + \beta_1)^i] + \sigma^2_t\ (\alpha_1 + \beta_1)^j$$

Then, since $(\boldsymbol{\alpha_1} + \boldsymbol{\beta_1})$ = **0.952**

$$E_{2020:8}[\sigma^2_{2020:12}] = \mathbf{0.756} * \{1 + (\mathbf{0.952}) + (\mathbf{0.952})^2 + (\mathbf{0.952})^3\} +$$
$$+\ \mathbf{43.037841} * (\mathbf{0.952})^4 = 38.02797$$

Lower variance forecasted for the end of the year, but still far from the unconditional variance of **15.4**.

98

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:

$$\boldsymbol{s^2} = \frac{1}{(T-1)} \sum_{i=1}^{\boldsymbol{T}} (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2$$

• 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 $\overline{\boldsymbol{x}}$ 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:

$$Variance_t = \sum_{i=1}^{k} r_{t,i}^2$$

100

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 ($i \rightarrow 0$),

$$RV_t \rightarrow \text{Return Variation } [t-1, t].$$

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

101

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.

102

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 ∞. 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.

105

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:

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

$r_{t,j}$ 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, $RVol$, is:

$$RVol_{t=30-min} = \sqrt{RV_{t=30-min}}$$

106

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.

SYMBOL	DATE	TIME	PRICE	SIZE
SPY	20140102	9:30:00	183.98	500
SPY	20140102	9:30:00	183.98	500
SPY	20140102	9:30:00	183.98	200
SPY	20140102	9:30:00	183.98	500
SPY	20140102	9:30:00	183.98	1000
SPY	20140102	9:30:00	183.98	1000
SPY	20140102	9:30:00	183.98	800
SPY	20140102	9:30:00	183.98	100
SPY	20140102	9:30:00	183.98	100
SPY	20140102	9:30:00	183.97	200
SPY	20140102	9:30:00	183.98	100
SPY	20140102	9:30:00	183.97	200
SPY	20140102	9:30:00	183.98	1000
SPY	20140102	9:30:00	183.97	100
SPY	20140102	9:30:00	183.98	1000
SPY	20140102	9:30:00	183.98	2600
SPY	20140102	9:30:00	183.98	1000
SPY	20140102	9:30:00	183.97	400

107

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

SYMBOL	DATE	TIME	BID	OFR	BIDSIZ	OFRSIZ	MODE	EX
AAPL	20140102	4:00:00	455.39	0	1	0	12	T
AAPL	20140102	4:00:00	553.5	558	2	2	12	P
AAPL	20140102	4:00:01	455.39	561.02	1	2	12	T
AAPL	20140102	4:00:45	552.1	558	1	2	12	P
AAPL	20140102	4:00:51	552.1	558.4	1	2	12	P
AAPL	20140102	4:00:51	552.1	558.8	1	2	12	P
AAPL	20140102	4:00:51	552.1	559	1	1	12	P
AAPL	20140102	4:01:14	553	559	1	1	12	P
AAPL	20140102	4:01:30	553.01	561.02	1	2	12	T
AAPL	20140102	4:01:43	553.01	559	1	1	12	T
AAPL	20140102	4:01:44	553.05	559	1	1	12	P
AAPL	20140102	4:01:49	455.39	559	1	1	12	T
AAPL	20140102	4:01:49	553.61	559	1	1	12	T
AAPL	20140102	4:02:02	553.05	559	1	2	12	P
AAPL	20140102	4:02:04	455.39	559	1	1	12	T
AAPL	20140102	4:02:04	548.28	559	1	1	12	T
AAPL	20140102	4:02:33	553.05	558.83	1	2	12	P
AAPL	20140102	4:02:33	555.17	558.83	2	2	12	P
AAPL	20140102	4:03:50	555.2	558.83	5	2	12	P

108

RV Models: High Frequency – TAQ

Example (continuation): We read SPY trade data for 2014:Jan.

```
> HF_da <- read.csv("c:/Financial Econometrics/SPY_2014.csv", head=TRUE, sep=",")
> summary(HF_da)
SYMBOL         DATE             TIME          PRICE         SIZE           G127
SPY:6800865  Min.   :20140102  9:30:00 :  21436  Min.   :176.6  Min.   :      1  Min.   :0
             1st Qu.:20140110  16:00:00: 11352   1st Qu.:178.9  1st Qu.:    100  1st Qu.:0
             Median :20140121  9:30:01 :  5922   Median :182.6  Median :    100  Median :0
             Mean   :20140119  15:59:59: 4090    Mean   :181.4  Mean   :    337  Mean   :0
             3rd Qu.:20140128  15:59:55: 3198    3rd Qu.:183.5  3rd Qu.:    300  3rd Qu.:0
             Max.   :20140131  15:50:00: 2916    Max.   :189.2  Max.   :4715350  Max.   :0
                               (Other) :6751951
    CORR              COND              EX
 Min.   :0.0e+00   @      :3351783   T      :1649158
 1st Qu.:0.0e+00   F      :2888182   P      :1335135
 Median :0.0e+00          : 524409   Z      :1182126
 Mean   :1.9e-04   O      :  18057   D      :1062382
 3rd Qu.:0.0e+00   4      :   9098   K      : 437900
 Max.   :1.2e+01   6      :   8142   J      : 356539
                   (Other):   1194   (Other): 777625
```

10
9

RV Models: High Frequency – TAQ

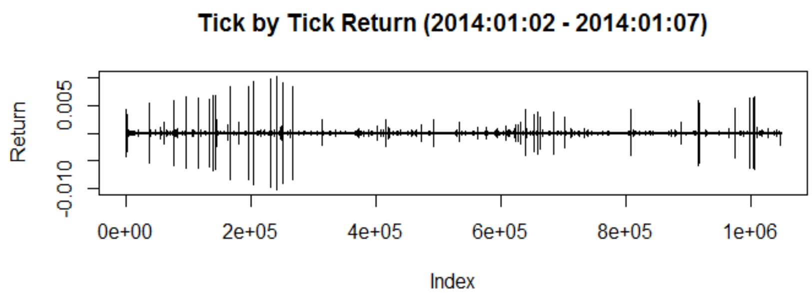
Example (continuation): Using the SPY trade data, we calculate using 5'-returns a daily realized volatility for the first 4 days in 2014 (2014:01:02 - 2014:01:07). Originally, we have T = 1,048,570.

```
HF_da <- read.csv("http://www.bauer.uh.edu//rsusmel//4397//SPY_2014.csv",
head=TRUE, sep=",")
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
                                             # pt pastes together DATE and Time.
spy_p <- as.numeric(hf_1$PRICE)              # Read price data as numeric

T <- length(spy_5_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)
```

11
0

RV Models: High Frequency – TAQ

Example (continuation): We plot the tick-by-tick data.

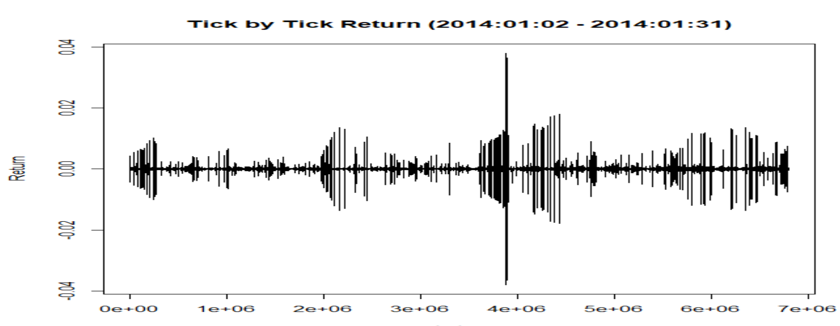


Very noisy data, with lots of "jumps":
Mean tick by tick return: -3.7365e-09
Tick-by-tick SD: 6.3163e-05

111

RV Models: High Frequency – TAQ

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

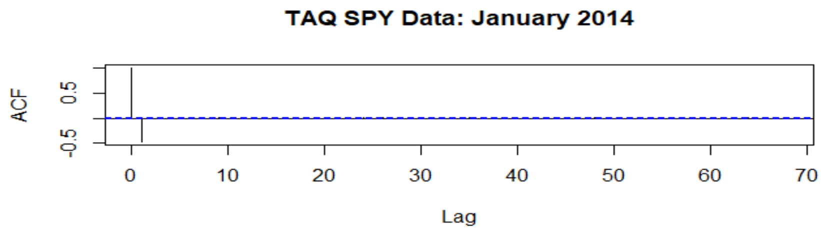



```
> mean(spy_ret)
[1] -4.796933e-09
> sd(spy_ret)
[1] 7.804991e-05
```

112

RV Models: High Frequency – TAQ

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



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.**

113

RV Models: High Frequency – TAQ

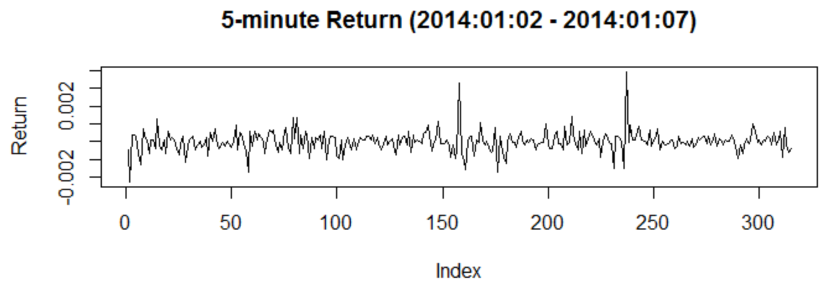
Example (continuation): We aggregate the tick-by-tick data in 5' intervals using the function *aggregateTrades* in the R package *highfrequency*. It needs as an input an xts object (hf_1, for us).

```
library(highfrequency)
spy_5 <- aggregateTrades(
hf_1,
on = "minutes",                    # you can use also seconds, days, weeks, etc.
k = 5,                             # number of units in for "on"
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)")
```

114

RV Models: High Frequency – TAQ

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

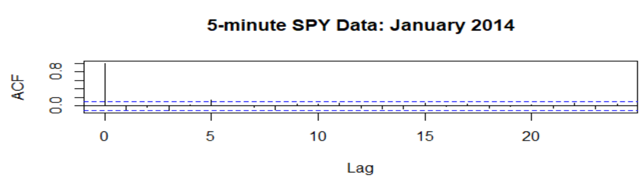


$RVol_{t=2014:01:02} = 0.0053344$
$RVol_{t=2014:01:03} = 0.0043888$
$RVol_{t=2014:01:04} = 0.0059836$
$RVol_{t=2014:01:05} = 0.0052772$

115

RV Models: High Frequency – TAQ

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
```

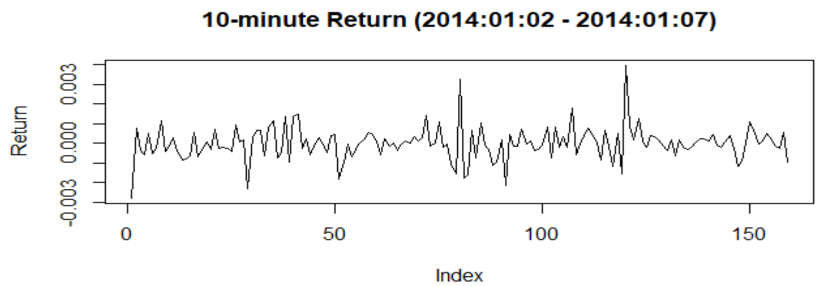
0	1	2	3	4	5	6	7	8	9	10
1.000	-0.105	-0.024	-0.104	0.018	0.147	0.016	-0.024	-0.088	0.048	0.037

Note: We have a negative 1st-order autocorrelation: **-0.105**, thought not significant. However, the autocorrelation of order 5 is significant.

116

RV Models: High Frequency – TAQ

Example (continuation): We plot the 10-minute return data. Smoothing increases.

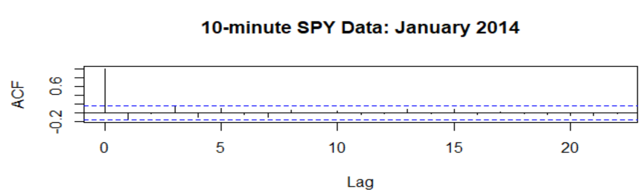


$RVol_{t=2014:01:02} = 0.005478294$
$RVol_{t=2014:01:03} = 0.004256046$
$RVol_{t=2014:01:04} = 0.006190508$
$RVol_{t=2014:01:05} = 0.005145601$

117

RV Models: High Frequency – TAQ

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



Note: Now, none of the autocorrelations is significant. The 10-minute returns look independent.

118

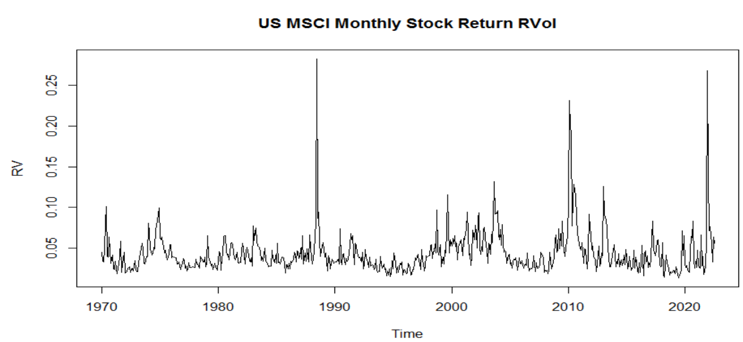
RV Models: R Script

Example: R script to compute realized volatility

```
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).




```
> mean(rvol)          # average monthly Rvol in the sample
[1] 0.04326531        ⇒ very close to monthly S&P Volatility: 4.49%
> sd(rvol)            # 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:

$$RV_{t=J} = J * RV_{t=j}$$

The $RVOL_{t=j}$ is the square root of $RV_{t=j}$.

RV Models: Log Rules

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

$$RV_{t=annual} = 260 * RV_{t=daily}$$

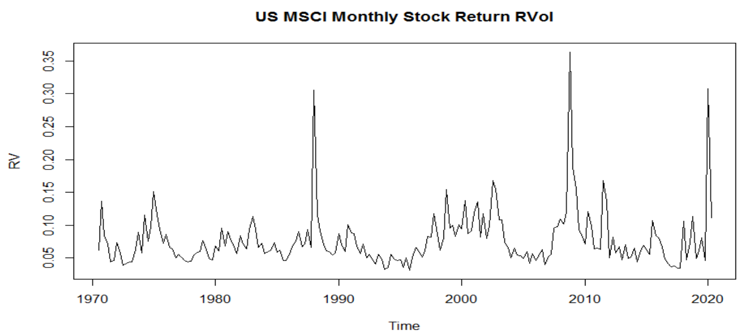
where 260 is the number of trading days in the year. The annualized RVOL is the squared root of RV_{annual}:

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

We can use time series models –say, an ARIMA model- for RV_t to forecast daily volatility.

RV Models: Quarterly RV From Daily Data

Example: Using daily data we calculate 3-mo Realized Volatility (k = 66 days) for log returns for the MSCI (1970: March – 2020: Oct).




```
> mean(rvol)        # average monthly Rvol in the sample
[1] 0.07725361       ⇒ log approximation: sqrt(3) * 0.04326 = 0.07493 (close!)
> sd(rvol)          # standard deviation of monthly Rvol in the sample
[1] 0.02592653
```

RV Models: Properties

• Under some conditions (bounded kurtosis and autocorrelation of squared returns less than 1), RV_t 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 ε_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:

$$s_t^2 = \{ \textstyle\sum_t (\ln(H_t) - \ln(L_t))^2 / (4\ln(2)T) \},$$

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):

$$\text{RR}_t = \{ \textstyle\sum_j [100 * (\ln(H_{t,j}) - \ln(L_{t,j}))^2]/(4\ln(2)) \},$$

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.

Reference: Christensen and Podolskij (2005).

Stochastic volatility (SV/SVOL) models

• Now, instead of a known volatility at time t, like ARCH models, we allow for a stochastic shock to σ_t, η_t or υ_t:

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

Or using logs:

$$\log \sigma_t = \omega + \beta_1 \log \sigma_{t-1} + \upsilon_t, \qquad \upsilon_t \sim N(0, \sigma_\upsilon^2)$$

• The difference with ARCH models: The shocks that govern the volatility are not necessarily the shocks to the mean process, ε_t's.

• Usually, the standard model centers log volatility around ω:

$$\log \sigma_t = \omega + \beta_1 (\log \sigma_{t-1} - \omega) + \upsilon_t,$$

Then,

$$E[\log(\sigma_t)] = \omega$$
$$Var[\log(\sigma_t)] = \varkappa^2 = \sigma_\upsilon^2/(1 - \beta_1{}^2).$$

$\Rightarrow$ Unconditional distribution: $\log(\sigma_t) \sim N(\omega, \varkappa^2)$ 12

Stochastic volatility (SV/SVOL) models

• We have 3 SVOL parameters to estimate: φ = (ω, β_1, σ_υ^2).

• Like ARCH models, SV models produce returns with kurtosis > 3 (and, also, positive autocorrelations between squared excess returns):

$$\text{Var}[r_t] = \text{E}[(r_t - \text{E}[r_t])^2] = \text{E}[\sigma_t^2\, z_t^2\,] = \text{E}[\sigma_t^2\,]\, \text{E}[z_t^2\,]$$
$$= \text{E}[\sigma_t^2\,] = \exp(2\omega + 2\,\varkappa^2) \qquad \text{(property of log normal)}$$

$$\text{kurt}[r_t] = \text{E}[(r_t - \text{E}[r_t])^4] \;/\; \{(\text{E}[(r_t - \text{E}[r_t])^2]\,)^2\,\}$$
$$= \text{E}[\sigma_t^4\,]\, \text{E}[r_t{}^4] \;/\; \{(\text{E}[\sigma_t^2\,])^2\;\; (\text{E}[z_t^2\,])^2\,\}$$
$$= 3\exp(4\omega + 8\varkappa^2) \;/\; \exp(4\omega + 4\varkappa^2) = 3\exp(4\varkappa^2) > 3!$$

• Estimation:
- GMM: Using moments, like the sample variance and kurtosis of returns. Complicated -see Anderson and Sorensen (1996).
- Bayesian: Using MCMC methods (mainly, Gibbs sampling). Modern approach.

12
9

Stochastic volatility (SV/SVOL) models

• The Bayesain approach takes advantage of the idea of hierarchical structure:
- f($\boldsymbol{y}$ | h_t) (distribution of the data given the volatilities)
- f(h_t | φ) (distribution of the volatilities given the parameters)
- f(φ) (distribution of the parameters)

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

Classic references: Jacquier, E., Poulson, N., Rossi, P. (1994), "Bayesian analysis of stochastic volatility models," *Journal of Business and Economic Statistics*. (Estimation). Heston, S.L. (1993), "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*. (Theory)