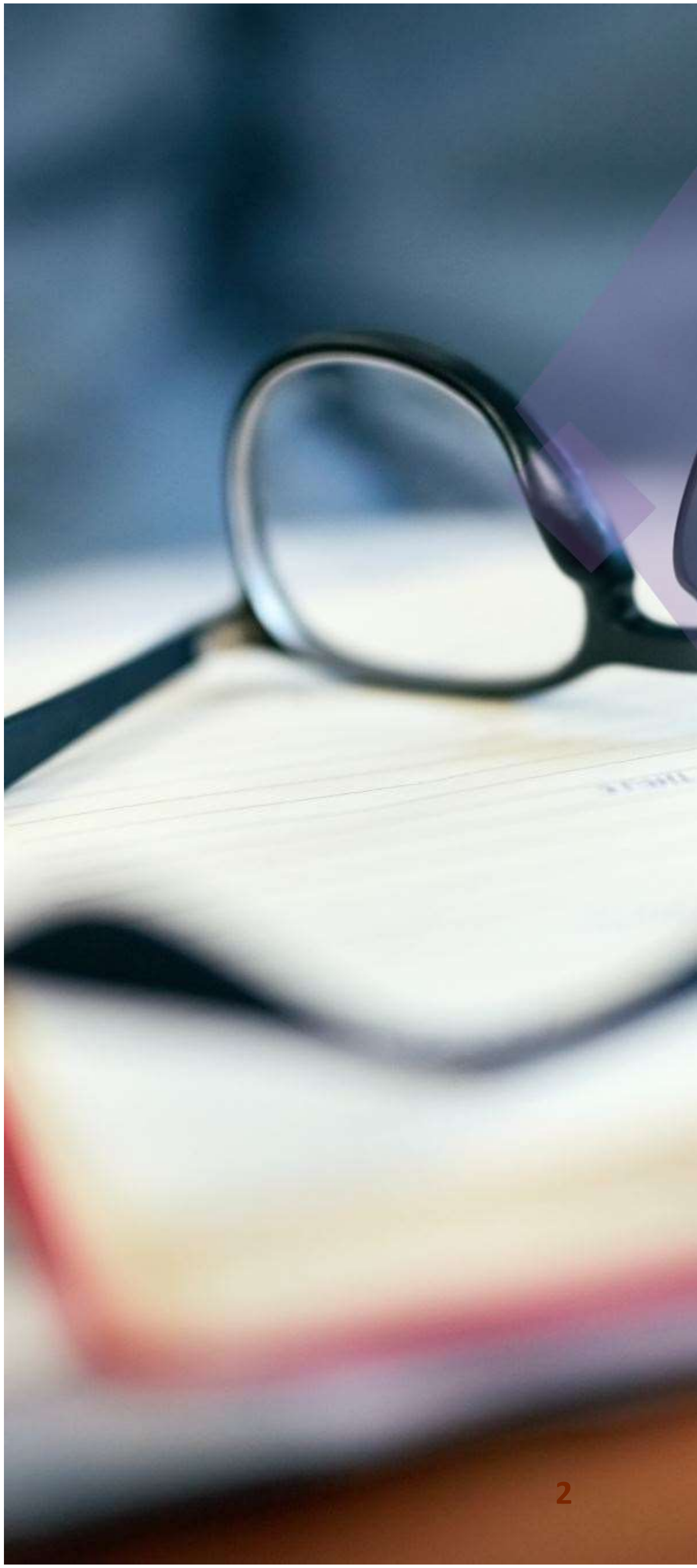
Quantitative Analysis of Financial Markets
Non-Stationary Time Series

Benjamin Ee
November, 2020

Administrative matters

- Student presentations will occupy the entirety of class time next week
- For students who are participating remotely, I will co-host the presenters
- For students who are presenting in person, you can just do so from your computer, or (if you wish) my laptop
- Around 15 minutes per group, with several minutes of Q&A
- Presentation order will be predetermined prior to class
- If you have a preference, email it to me by this coming Friday. It will be assigned at that point and emailed to the class.



Today

1. ARIMA
2. ECM
3. GARCH

Non stationary processes

Trend-stationary processes:

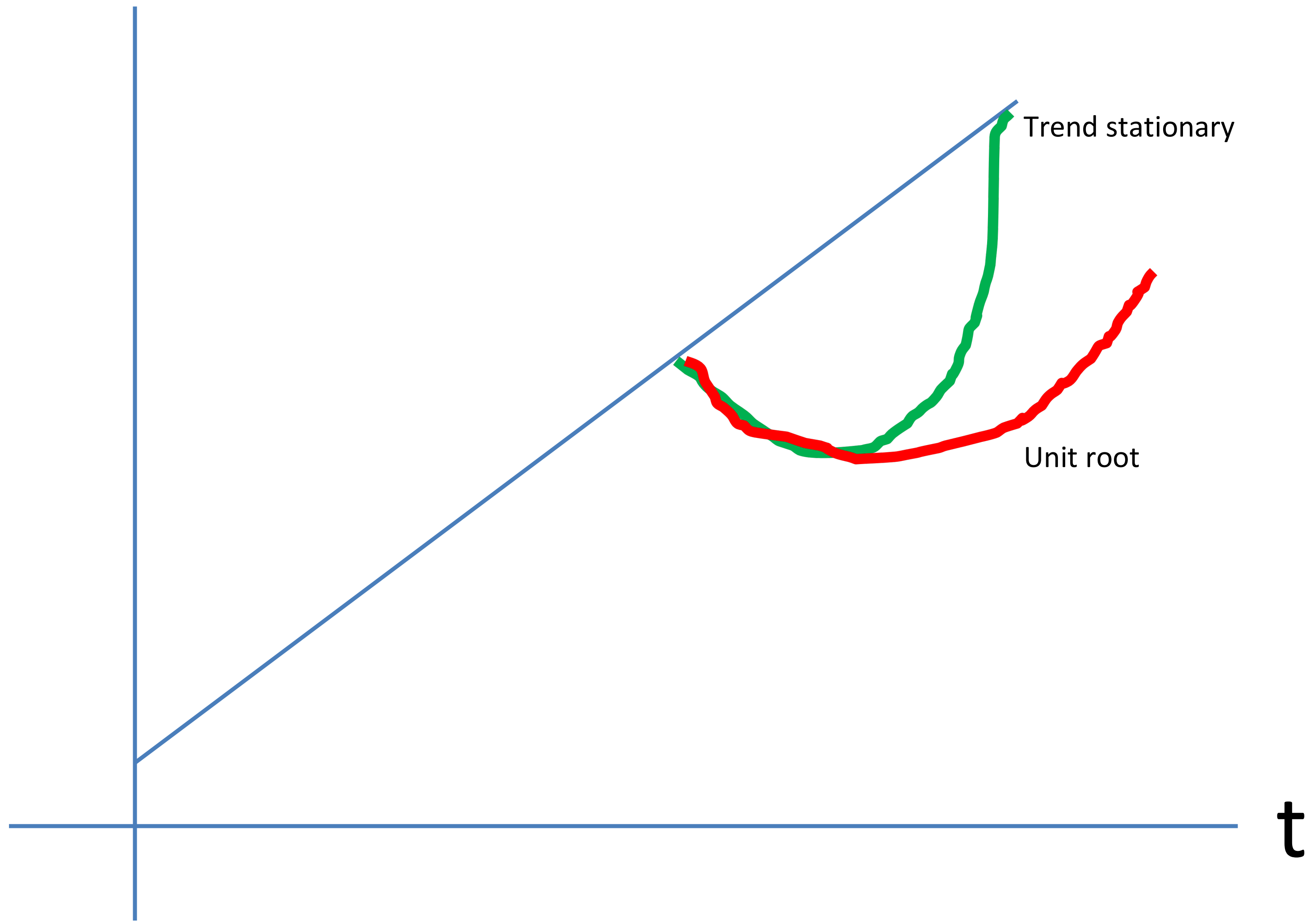
- Time series has a deterministic trend
- Stochastic shocks only have a temporary effect. After shocks, variable tends toward a deterministically determined mean

Cyclo-stationary processes:

- No trend like behavior
- Varies cyclically with time

Unit roots

- A trend in the mean of the variable can be due to unit roots
- Stochastic shocks have permanent effects
- Process is not mean-reverting
- Consider $y_t = py_{t-1} + u_t$., where t is time, p is a coefficient and u_t is the error term.
  - A unit root is present if p = 1


Trend stationary
Unit root
t

ARIMA(p,d,q)

- Parameters of the ARIMA(p,d,q) model are:
  - P: Number of lagged terms included in the model
  - D: Number of times that the observations are differenced to result in a stationary distribution
  - Q: Length of the moving average window

- The additional "D" term involves subtracting observations from observations at the previous time step, to obtain a stationary time series. This can serve two purposes:
  - Remove trend terms
  - Remove seasonality

- Overall model is estimated with linear regression

- When D = 0, model reverts to a typical ARMA(p,q) situation

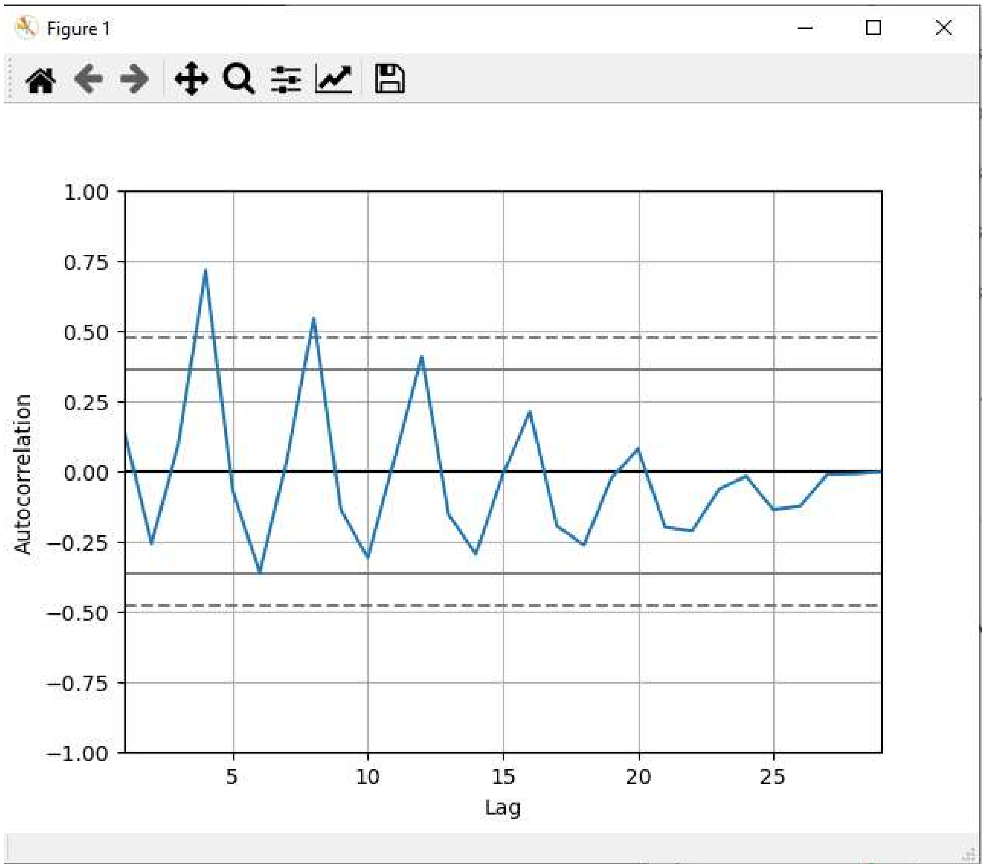
An example of ARIMA differencing

- Given: $y_1, y_2, y_3, \ldots. y_{t-1}, y_t, y_{t+1}, \ldots.$
- Define:
- First difference: $\Delta\, y_t = y_t - y_{t-1}$
- Second difference: $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$
- Lag between differenced terms does not have to be 1. We can allow it to be an arbitrary lag s, e.g. for seasonal differencing
- First difference: $\Delta\, y_t = y_t - y_{t-s}$ where s = seasonal difference
- A seasonal ARIMA (SARIMA) model can also be written as "ARIMA$(P,D,Q)_m$" where m is the seasonal lag (this is not the only notation used)
- Overall, differenced data is subsequently used for estimation of an ARMA model

Examples of ARIMA models

- ARIMA(0,0,0): white noise
- ARIMA(0,1,0): Random walk
    - $y_t = y_{t-1} + e_t$
    - $y_t = c + y_{t-1} + e_t$ (random walk with constant drift c)

Autocorrelation plot of Apple’s revenues

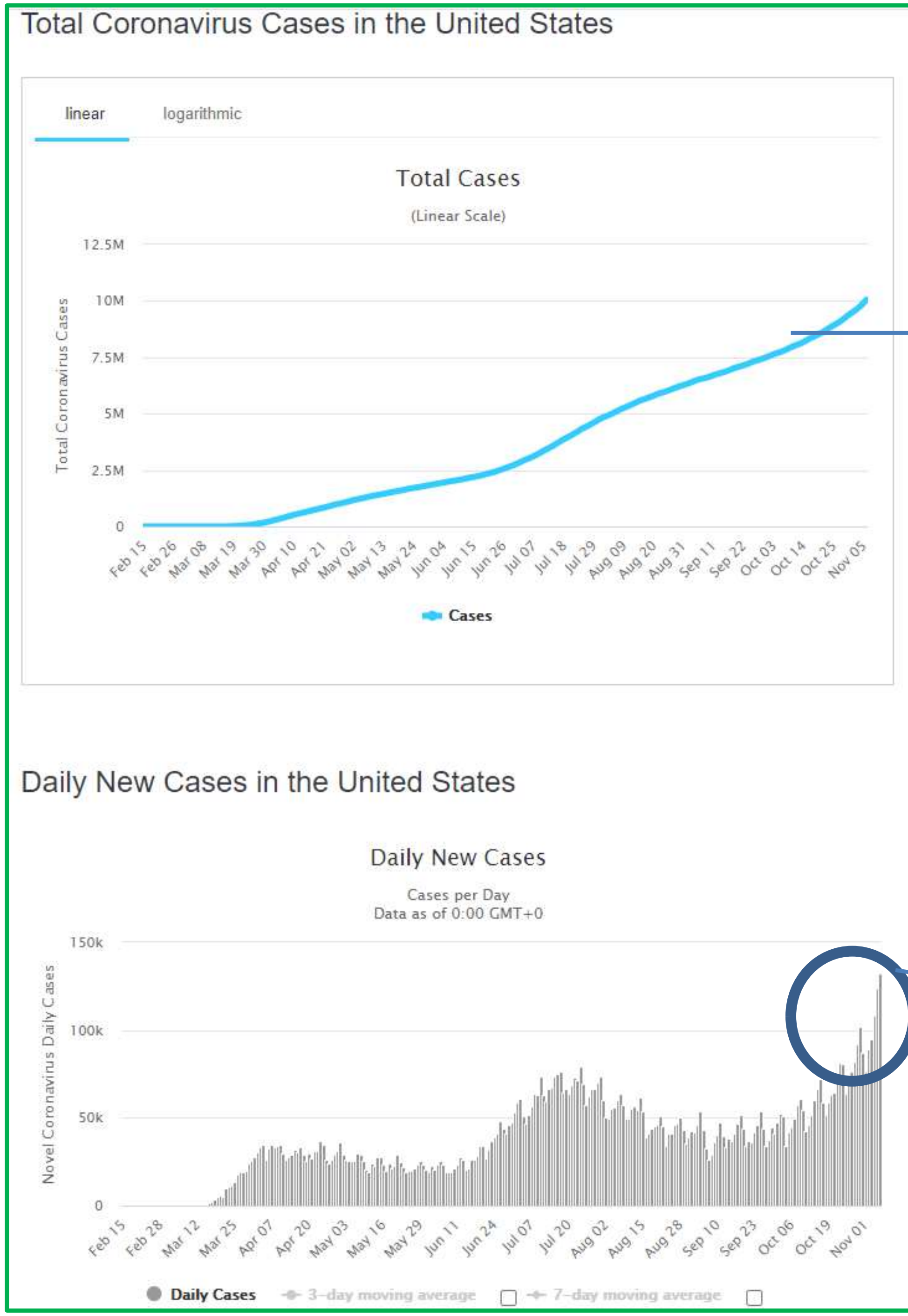



```
In [2]: from pandas.plotting import autocorrelation_plot

In [3]: autocorrelation_plot(appleresampled['revenue']
   ...: )
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x26898b67520>

In [4]: plt.show()
```

```python
import statsmodels.api as sm
mod_sarimax = sm.tsa.SARIMAX(appleresampled['revenue'], order = (0,1,0))
res_sarimax = mod_sarimax.fit()
res_sarimax.summary()
```

```
In [19]: res_sarimax.summary()
Out[19]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                               SARIMAX Results
==============================================================================
Dep. Variable:                revenue   No. Observations:                   29
Model:               SARIMAX(0, 1, 0)   Log Likelihood                -702.903
Date:                Sun, 08 Nov 2020   AIC                           1407.806
Time:                        11:11:39   BIC                           1409.138
Sample:                    09-30-2013   HQIC                          1408.213
                         - 09-30-2020
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      3.604e+20   1.24e+20      2.911      0.004    1.18e+20    6.03e+20
===================================================================================
Ljung-Box (Q):                      117.92   Jarque-Bera (JB):                 1.00
Prob(Q):                              0.00   Prob(JB):                         0.61
Heteroskedasticity (H):               1.62   Skew:                             0.11
Prob(H) (two-sided):                  0.48   Kurtosis:                         2.10
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
"""
```

Example of ARIMA model application



Time trend in total # cases

Seasonality in rate of increase?

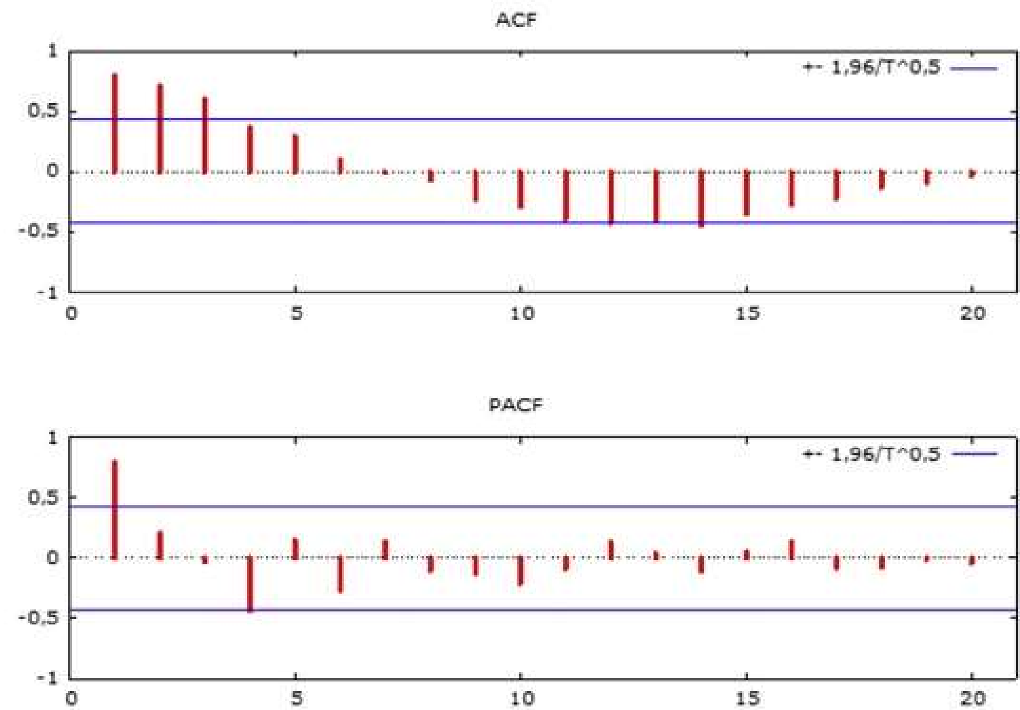
Reference: "Application of the ARIMA model on the Covid-2019 epidemic dataset", Data in Brief, April 2020 V29

https://www.sciencedirect.com/science/article/pii/S2352340920302341

ARIMA modelling of epidemic spread

Using the John Hopkins epidemiological data, the authors build an ARIMA model to forecast incidence level and growth

ARIMA(1,0,3) model was selected as the most appropriate fit



Vector ARIMA

SO FAR, WE ARE ONLY MODELLING A SINGLE VARIABLE

VECTOR ARIMA (VAR) MODELS CAN DESCRIBE A SET OF K VARIABLES, WHICH ARE MODELLED AS A LINEAR FUNCTION OF ALL OF THEIR PAST VALUES ONLY

EACH OF THE K VARIABLES CAN DEPEND ON LAGGED VALUES OF ALL OF THE OTHER K VARIABLES

Vector ARIMA example

$$Price_{wheat(t)} = B_0 Price_{wheat(t-1)} + B_1 Price_{corn(t-1)} + B_2 Price_{rice(t-1)}$$
$$Price_{corn(t)} = B_3 Price_{wheat(t-1)} + B_4 Price_{corn(t-1)} + B_5 Price_{rice(t-1)}$$
$$Price_{rice(t)} = B_6 Price_{wheat(t-1)} + B_7 Price_{corn(t-1)} + B_8 Price_{rice(t-1)}$$

- Note all variables have to be of the same order of integration (slides 15 to 21)
- If all the variables are cointegrated, then the error correction term should be included in the VAR
- What is the error correction term?

2. Error Correction Models

Cointegration

We have been studying the relationship of variables with their own history

i.e. single variable situations

In estimations with multiple time series variables, it is possible for all the variables to contain time trends

For a set of n time series variables, the set of variables is cointegrated if both of the following are true:

- Each individual variable have order of integration d
- Order of integration "I(d)" of a time series variable is the minimum number of differences needed to transform the variable into a covariance stationary time series
- A linear combination of the variables is integrated of order < d

Order of integration

1. Order of integration, I(d) of a time series is the minimum number of differences needed to obtain a covariance-stationary series

2. Covariance-stationarity requires that **only** that:

    - First moment (mean) of the time series does not vary with time. i.e. constant mean
    - Second moment (autocovariance) of the time series does not vary with time.
    - Autocovariance is finite (for all times)

3. Covariance-stationarity is also referred to as weak stationarity

4. This is in contrast to strict-sense stationarity, which requires that the entire cumulative distribution function is not a function of time

Spurious relationships

- Regressing ***independent*** non-stationary variables (e.g. with a unit root) against each other will frequently lead to Type 1 errors

- An example from economics involves ***nominal*** economic variables
  - Because both independent and dependent variables may involve the growth of prices over time (e.g. inflation), we may reject the null hypotheses even without a casual effect
  - This is an example of a confounding variable (e.g. T -> X and T-> Y)
  - Monte carlo simulations demonstrate that simple OLS in this situation will generate very high R squared, high t-statistic and low DW statistic, despite the variables being independent
  - We note that such spurious results ***may*** still be interesting because we could be interested in the confounding stochastic variable that is causing the spurious relationship. However, in this case, it is better to identify the variable directly

Testing for cointegration

For cointegrated I(1) processes, detrending does not eliminate spurious correlation. However, two I(1) trends can be co-integrated only if there is a true relationship between the two

The Engle-Granger two-step method is a commonly used test for cointegration

If y_t and x_t are non-stationary and I(1), a linear combination of them will be stationary if they are cointegrated. i.e.:

$Y_t - Bx_t = u_t$ where u_t is stationary

Step 1: We estimate B first using OLS
Step 2: Test the residuals u_t for stationarity

Engle-Granger two step-method (DF test)

Procedure:

1. Test individual time series to confirm they are non-stationary.

2. This can be done via the Dickey-Fuller test
   - Dickey-Fuller tests the null hypothesis that a unit root is present in an AR model, versus the alternative that the data is not
   - Recall: $y_t = py_{t-1} + u_t$, where a unit root is present so $p = 1$
   - Let $\Delta\, y_t = (p-1)\, y_{t-1} + u_t$
   - We estimate the model, and test that $(p-1) = 0$
   - Critical regions (depending on sample size) are read from the Dickey-Fuller table (we cannot use a standard t-distribution to provide critical values)

Engle-Granger two step-method (Cointegration test)

Procedure (continued):

3. If DF testing indicates both series are I(0), we run OLS as per class 3 & 4

4. If they have different orders of cointegration, we have to transform the variables so that they have the same order of cointegration (recall that differencing reduces I(..))

5. If both variables are integrated to the same order, we can estimate an **error correction model** of the form:
    - $\varepsilon_t = y_t - B_0 - B_1 x_t$, and save predicted residuals
    - Test residuals for a unit root to verify that the variables are cointegrated. There **will not** be a unit root in residuals if variables are cointegrated
    - Note this statistical test is weak because we are testing that the null hypothesis is **not rejected**

Engle-Granger two step-method (ECM modelling)

Procedure (continued):

6. Assuming that variables are indeed cointegrated, we can model the relationship via an error correction model of the form:

    - $\Delta y_t = \gamma + B_1 * \Delta x_t + \alpha\varepsilon_{t-1} + v_t$

7. Second term describes short run impact of a change in x_t on y_t

8. Third term (with α) describes long run correction towards equilibrium relationship between the variables

9. Last term reflects random shocks

3. GARCH

INTRODUCTION

VAR

ESTIMATING VOLATILTIY

EWMA

GARCH(1,1)

MLE

VOLATILITY TERMS STRUCTURE

TAKEAWAYS

Introduction

- The objective is to use historical data to produce estimates of the current and future levels of volatilities and correlations.

- You need to recognize that volatilities and correlations are not constant.

- Applications that motivate this module:

  VaR
  Valuation of derivatives

- Three important models:

  EWMA: Exponentially Weighted Moving Average

  ARCH: Autoregressive Conditional Heteroscedasticity

  GARCH: Generalized Autoregressive Conditional Heteroscedasticity

Risks and Risk Management

✓ Major Risks

- § Market risks
- § Credit risks
- § Operational risks
- § Liquidity risks
- § Legal risks
- § Political risks
- § Model risks

✓ Industry Practices

- § Regulatory capital adequacy
- § Bank's internal risk control
- § Corporations' investments
- § Firm's hedging of transactions
- § Exchanges' margining rules and practices

Introduction: Risk Measure

- Risk Management is a procedure for shaping a loss distribution.

- Despite some serious shortcomings, Value-at-Risk, or VaR, is the most popular portfolio risk measure used by risk management practitioners.

- VaR is a number constructed on day t such that the portfolio losses on day $t + 1$ will only be larger than the VaR forecast with probability p, e.g. 5%.

- The main objective of this lesson is to see how GARCH model is applied in forecasting VaR.
  ✓ Question: What risk does VaR address?

Value at Risk

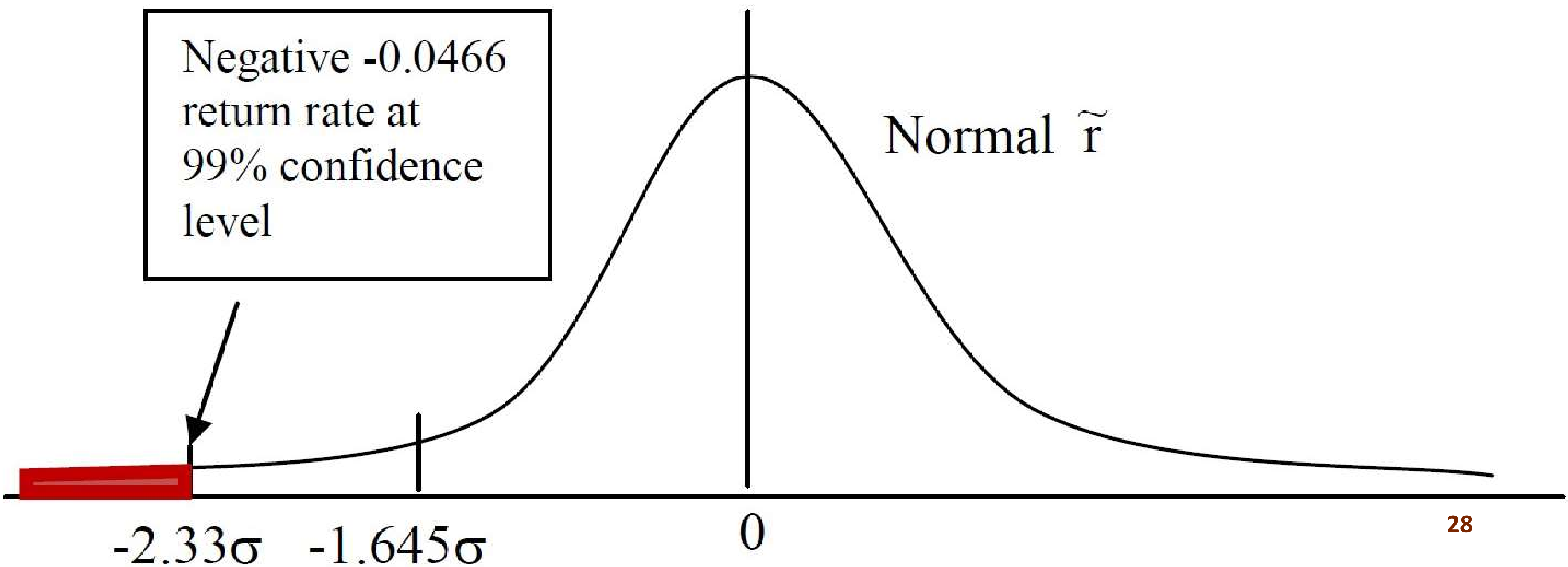
- VaR was popularized by J.P. Morgan in the 1990s. The executives at J.P. Morgan wanted their risk managers to generate one number at the end of each day to summarize the risk of the firm's entire portfolio.
- What they came up with was VaR.
- If the 95% VaR of a portfolio is $400, then we expect the portfolio will lose $400 or less in 95% of the scenarios, and lose more than $400 in 5% of the scenarios.
- We can define VaR for any confidence level, but 95% has become an extremely popular choice.
- VaR is a one-tailed confidence interval.

Definition of Value at Risk and Example

Definition of VaR

VaR is the maximum loss over a specified horizon at a given confidence level (e.g. 95%).

Example: Suppose log return $\tilde{r} \overset{d}{\sim} N(\mu, \sigma^2)$ Suppose daily volatility σ = 2% and daily mean μ = 0.



Example

- Suppose the portfolio value is $P_0 = \$100$ million.
- The daily log return is

$$\tilde{r} = \ln \frac{\tilde{P}_1}{P_0}$$

- Since $r = -2.33 \times 0.02 = -0.0466$, we have

$$-0.0466 = \ln \frac{\tilde{P}_1}{\$100\text{m}} \quad \text{or} \quad \tilde{P}_1 = \$100\text{m} \times e^{-0.0466}$$

- The VaR is, at the 99% confidence level

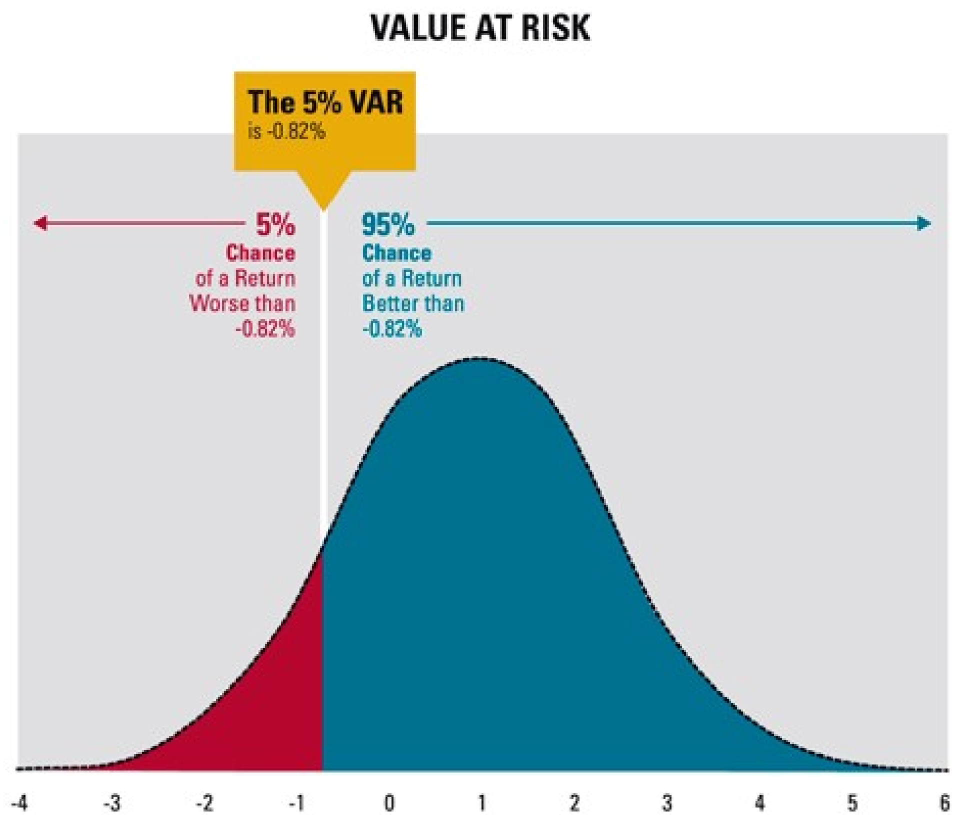
$$\$100\text{m} \times (1 - e^{-0.0466}) = \$4.553\text{m}.$$

Value at Risk Time Horizon

The time horizon needs to be specified for VaR.

On trading desks, with liquid portfolios, it is common to measure the one-day 95% VaR.

In other settings, in which less liquid assets may be involved, time frames of up to one year are not uncommon.



VaR Exceedance

- If an actual loss equals or exceeds the predicted VaR, that event is known as an **exceedance**.
- For a one-day 95% VaR, the probability of an exceedance event on any given day is 5%.
- Let the random variable L represent the loss to your portfolio. L is simply the negative of the return to your portfolio. If the return of your portfolio is -$600, then the loss, L, is +$600.
- For a given confidence level, α , then, we can define value at risk as

$$P\left(L \geq \mathrm{VaR}_{\alpha}\right) = 1 - \alpha.$$

- If a risk manager says that the one-day 95% VaR of a portfolio is $400, it means that there is a 5% probability that the portfolio will lose $400 or more on any given day (that L will be more than $400).

Remarks

- We can also define VaR directly in terms of returns. If we multiply both sides of the inequality in by -1, and replace -*L* with *R*, we come up with

$$P(R \leq -\mathrm{VaR}_{\alpha}) = 1 - \alpha.$$

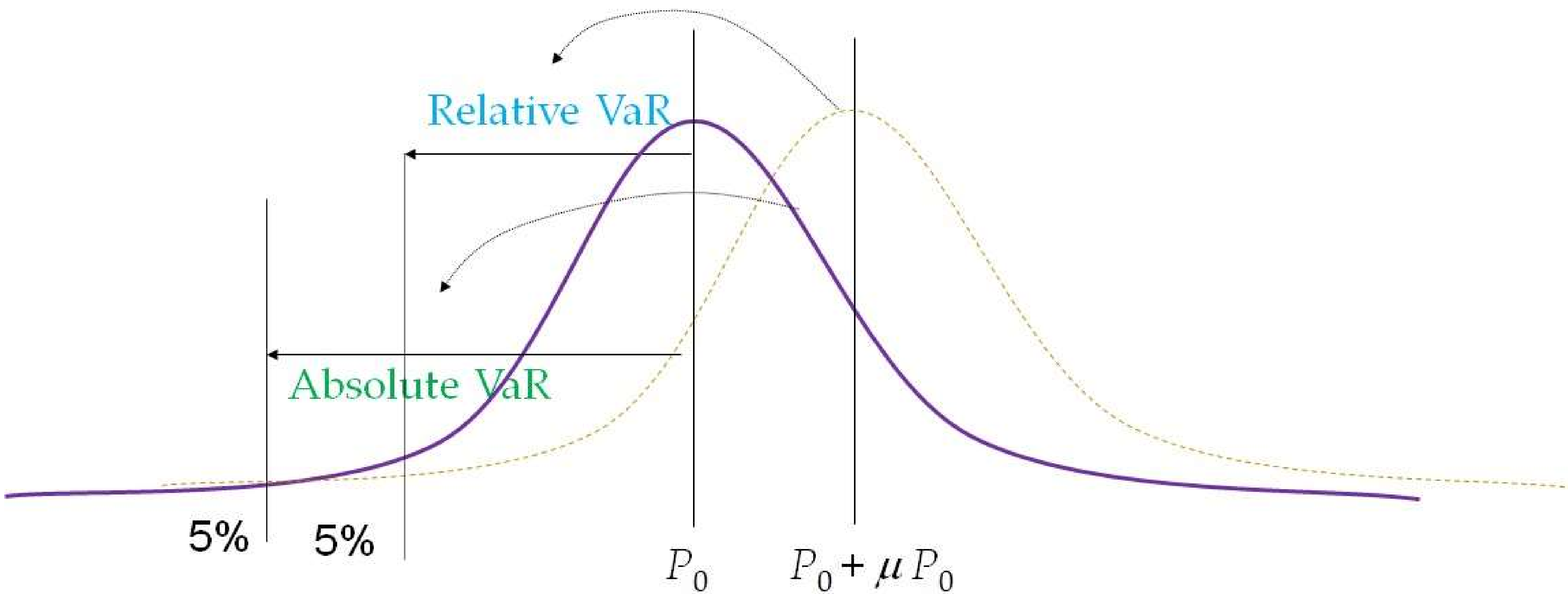
- A loss of $400 or more and a return of -$400 or less are exactly the same.

- Notice that the definition does not invoke the assumption that the distribution is normal.

Absolute Versus Relative VaR

- Absolute VaR is measured with respect to the current marked-to-market portfolio value regardless of μ
- Relative VaR is computed taking into account the loss also of the expected profit μP_0.
- Suppose $\mu =$ 1%. At the 99% confidence level, the critical value is-2.33.

Illustration



Definition of Volatility and Variance Rate

- Define σ_n as the volatility of a market variable on day n, as estimated at the end of day $n-1$.

The square of the volatility σ_n^2 on day n is the **variance rate**.

Denote S_i as the value of a variable at the end of day

- Define the log return as $u_i := \ln \frac{S_i}{S_{i-1}}$.
  - An unbiased estimate of the mean of u_i, using the most recent m observations with respect to "today" n is

$$\overline{u} = \frac{1}{m} \sum_{i=1}^{m} u_{n-i}.$$

An unbiased estimate of the variance rate per day, σ_n^2 ,is

$$\sigma_n^2 = \frac{1}{m-1} \sum_{i=1}^{m} (u_{n-i} - \overline{u})^2.$$

Alternative definitions of volatility?

- Intuitively, volatility is a measure of a financial asset's readiness to move from one price to another price, in such a way that leaves the market participants unsure about the next quantum of price change.

- The more ready the asset's price is to move, the more uncertain its price in the future will be. The resulting price uncertainty makes the asset risky to the investors, as the price may move in the direction contrary to what they expect, and they become exposed to the consequent losses.

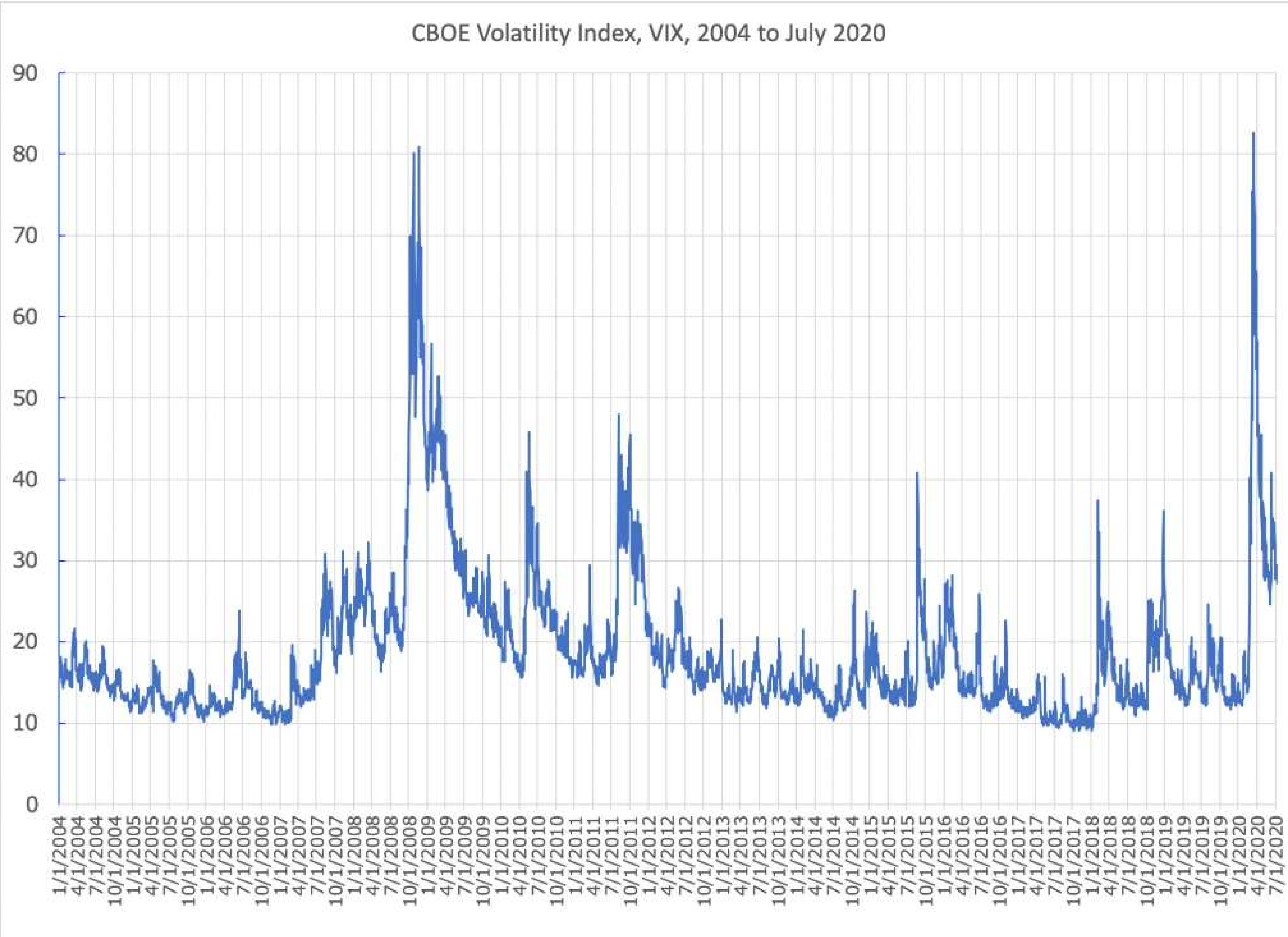
- Just like there are many different definitions of returns, volatility is defined differently for different purposes and from different sources.

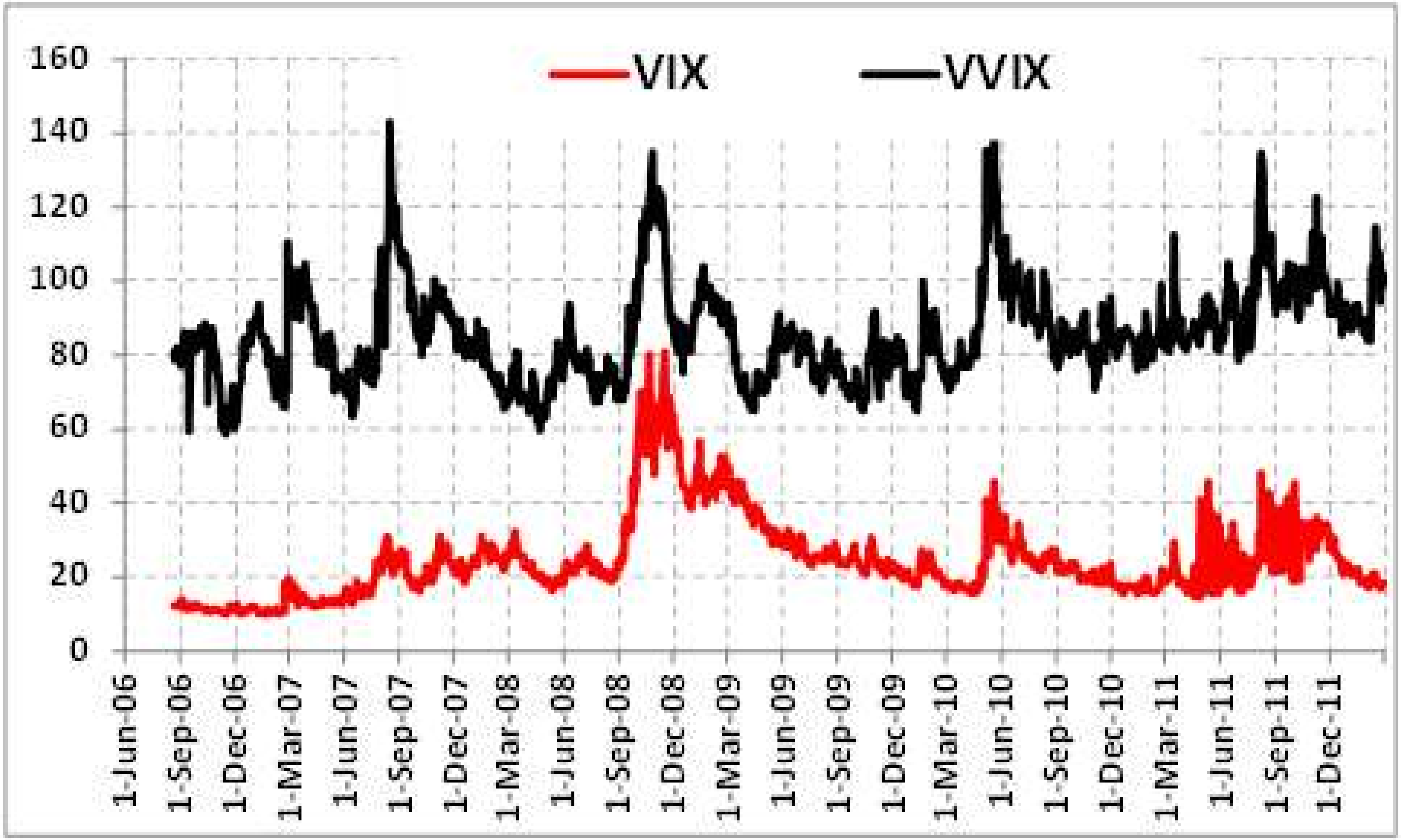
Implied Volatilities

Definition

The implied volatility of an option is the volatility that gives the market price of the option when it is substituted into the pricing model.

- An example of implied volatility: **VIX Index**


CBOE Volatility Index, VIX, 2004 to July 2020
90
80
70
60
50
40
30
20
10
0
1/1/2004
4/1/2004
7/1/2004
10/1/2004
1/1/2005
4/1/2005
7/1/2005
10/1/2005
1/1/2006
4/1/2006
7/1/2006
10/1/2006
1/1/2007
4/1/2007
7/1/2007
10/1/2007
1/1/2008
4/1/2008
7/1/2008
10/1/2008
1/1/2009
4/1/2009
7/1/2009
10/1/2009
1/1/2010
4/1/2010
7/1/2010
10/1/2010
1/1/2011
4/1/2011
7/1/2011
10/1/2011
1/1/2012
4/1/2012
7/1/2012
10/1/2012
1/1/2013
4/1/2013
7/1/2013
10/1/2013
1/1/2014
4/1/2014
7/1/2014
10/1/2014
1/1/2015
4/1/2015
7/1/2015
10/1/2015
1/1/2016
4/1/2016
7/1/2016
10/1/2016
1/1/2017
4/1/2017
7/1/2017
10/1/2017
1/1/2018
4/1/2018
7/1/2018
10/1/2018
1/1/2019
4/1/2019
7/1/2019
10/1/2019
1/1/2020
4/1/2020
7/1/2020


VIX
VVIX
160
140
120
100
80
60
40
20
0
1-Jun-06
1-Sep-06
1-Dec-06
1-Mar-07
1-Jun-07
1-Sep-07
1-Dec-07
1-Mar-08
1-Jun-08
1-Sep-08
1-Dec-08
1-Mar-09
1-Jun-09
1-Sep-09
1-Dec-09
1-Mar-10
1-Jun-10
1-Sep-10
1-Dec-10
1-Mar-11
1-Jun-11
1-Sep-11
1-Dec-11

Weighting Scheme

- So far, we give an equal weight to $u^2_{n-1}, u^2_{n-2}, \ldots, u^2_{n-m}$.

To estimate the *current* level of volatility, give more weight α_i to recent data.

$$\sigma_n^2 = \sum_{i=1}^{m} \alpha_i u_{n-i}^2. \qquad (1)$$

- The weights $\alpha_i, i = 1, 2, \ldots, m$ must satisfy the following:
  - positive: $\alpha_i > 0$.
  - lesser weight for older observations: $\alpha_i < \alpha_j$ when $i > j$.
  - summed to unity: $\sum_{i=1}^{m} \alpha_i = 1.$

Model for Estimating the Variance Rate: ARCH

- Let V_L be the long-run variance rate and γ be the weight assigned to V_L. The model for estimating the variance rate is

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2.$$

- The weights must sum to one.

$$\gamma + \sum_{i=1}^{m} \alpha_i = 1.$$

- Define $\omega := \gamma V_L$. The model is rewritten as

$$\sigma_n^2 = \omega + \sum_{i=1}^{m} \alpha_i u_{n-i}^2.$$

Model for Estimating the Variance Rate: ARCH

- The EMWA model is a special case of (1) where the weights α_i decrease exponentially as we move back through time.

- Specifically, with $0 < \lambda < 1$,

$$\alpha_{i+1} = \lambda \alpha_i,$$

the EMWA model is

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2.$$

  - Only the current estimate of the variance rate and the most recent observation on the value of the market variable are needed.

Exponential Decline

- We know

$$\begin{aligned}\sigma_n^2 &= \lambda\sigma_n^2 + (1-\lambda)u_{n-1}^2 \\ &= \lambda\left(\lambda\sigma_{n-2}^2 + (1-\lambda)u_{n-2}^2\right) + (1-\lambda)u_{n-1}^2 \\ &= (1-\lambda)u_{n-1}^2 + \lambda(1-\lambda)u_{n-2}^2 + \lambda^2\sigma_{n-2}^2\end{aligned}$$

- Substituting for σ^2_{n-2}, then for σ^2_{n-3}, then for σ^2_{n-4}, and so on:

$$\begin{aligned}\sigma_n^2 =& (1-\lambda)u_{n-1}^2 + \lambda(1-\lambda)u_{n-2}^2 + \lambda^2(1-\lambda)u_{n-3}^2 + \cdots \\ & \cdots + \lambda^{m-1}(1-\lambda)u_{n-m}^2 + \lambda^m\sigma_{n-m}^2.\end{aligned}$$

- Weights start at $1-\lambda$ and decline at rate λ.

Attractions of EWMA

Relatively little data needs to be stored

- We need only remember the current estimate of the variance rate and the most recent observation on the market variable
- Tracks volatility changes
- 0.94 is a popular choice for λ

ARCH

- A model for a process on the variance of u_t

$$\mathbb{V}(u_t) = \alpha_0 + \alpha_1 u_{t-1}^2$$

- More generally, with $e_t \overset{d}{\sim} N(0, 1)$,

$$u_t = e_t \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$$

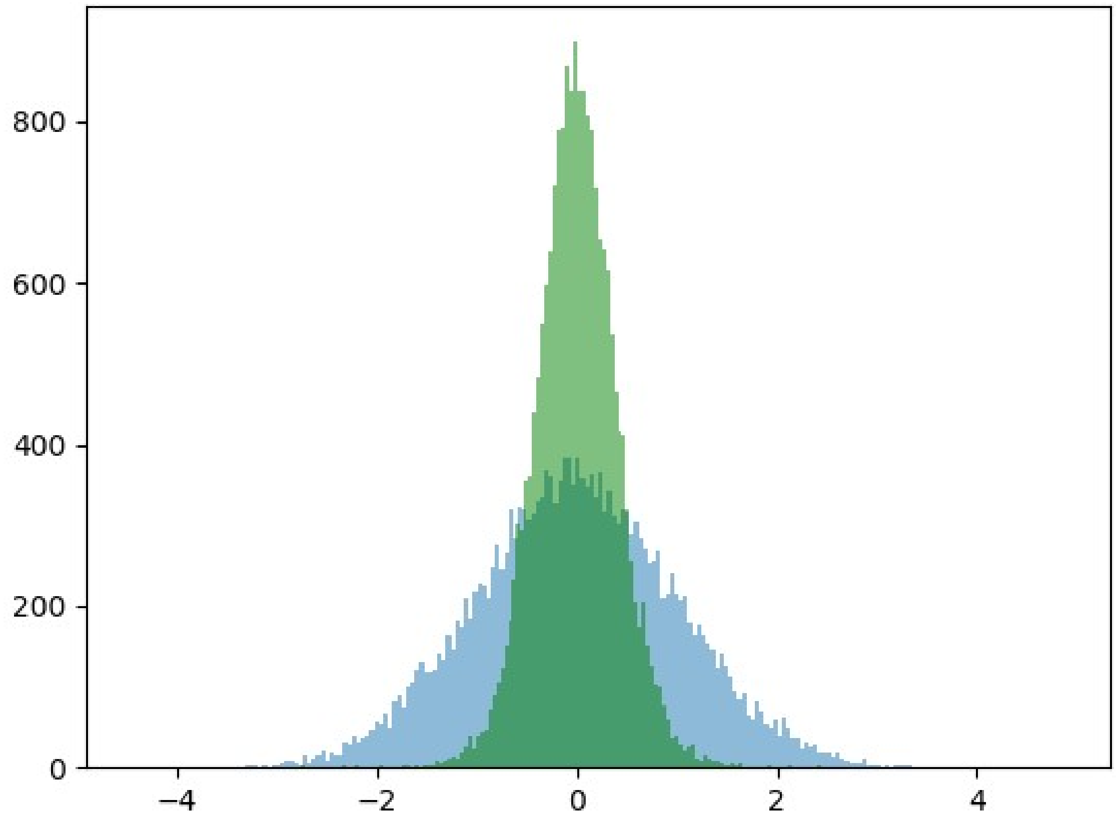
- The process u_t is unconditionally not a normal distribution. However, conditional on u_{t-1}, u_t is normally distributed.

ARCH Simulation Code

```
from_future import division, print_function
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import jarque_bera

alpha0, alpha1, n = 0.1, 0.4, 20000
np.random.seed(137)
e = np.random.normal(size=n)

plt.plot(e)
plt.show()

u = np.zeros(n)
u[0] = e[0]
for t in range(1,n):
   u[t] = e[t]*np.sqrt(alpha0 + alpha1 * u[t-1]**2)

plt.plot(u)
plt.show()

JBstat_e, JBstat_u = jarque_bera(e), jarque_bera(u)
print(JBstat_e)
print(JBstat_u)

he, hu = plt.hist(e, 200, alpha=0.5), plt.hist(u, 200, facecolor='g', alpha=0.5)
plt.savefig('ARCH_Simulate.png')
plt.show()
```

Histogram of Simulated ARCH(1) Process

- The Jarque-Bera test statistic of 1.74 does not reject the null
- But at 7,732, the Jarque-Bera test strongly rejects the null hypothesis for the ARCH(1) process u_t in green.



GARCH (1,1)

- Bollerslev in 1986 proposed the GARCH(1, 1) model.
- In GARCH(1,1), is calculated from a **long-run average variance rate**, V_L, as well as from σ^2 and u^2 .
- The GARCH(1,1) model is

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 .$$

- Since the weights must sum to unity, it follows that

$$\gamma + \alpha + \beta = 1.$$

- The EWMA model is a particular case of GARCH(1,1) where $\gamma = 0$, $\alpha = 1 - \lambda$ and $\beta = \lambda$.

Parameters of GARCH (1,1)

- Let $\omega := \gamma V_L$. The GARCH (1,1) is rewritten as

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 .$$

- Once the parameters ω, a, and β have been estimated, we can calculate γ as $1 - \alpha - \beta$.

- The **long-term variance** V_L can then be calculated as

$$V_L = \frac{\omega}{\gamma} = \frac{\omega}{1 - a - \beta}.$$

- For a stable GARCH(1,1) process, we require $\alpha + \beta < 1$, and obviously $\omega > 0$.

Unconditional Stationarity

- While GARCH processes are conditionally nonstationary with changing variances, they are still unconditionally stationary processes.

$$
\begin{aligned}
\sigma_n^2 &= \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \\
&= \omega + \alpha u_{n-1}^2 + \beta \Big(\omega + \alpha u_{n-2}^2 + \beta \sigma_{n-2}^2 \Big) \\
&= \omega \big(1 + \beta \big) + \alpha \big(u_{n-2}^2 + \beta u_{n-2} \big) + \beta^2 \Big(\omega + \alpha u_{n-3}^2 + \beta \sigma_{n-3}^2 \Big) \\
&= \omega \big(1 + \beta + \beta^2 + \cdots \big) + \alpha \big(u_{n-1}^2 + \beta u_{n-2}^2 + \beta^2 u_{n-3}^2 + \cdots \big)
\end{aligned}
$$

- Taking unconditional expectation on both sides, so that

$$
\sigma^2 := \mathbb{E}\left(u_n^2\right) = \mathbb{E}\left(u_{n-1}^2\right) = \mathbb{E}\left(u_{n-2}^2\right) = \cdots .
$$

Unconditional Stationarity (cont'd)

- Then

$$\begin{aligned}\sigma^2 &= \frac{\omega}{1-\beta} + \alpha(\sigma^2 + \beta\sigma^2 + \beta^2\sigma^2 + \cdots) \\ &= \frac{\omega}{1-\beta} + \frac{\alpha\sigma^2}{1-\beta}\end{aligned}$$

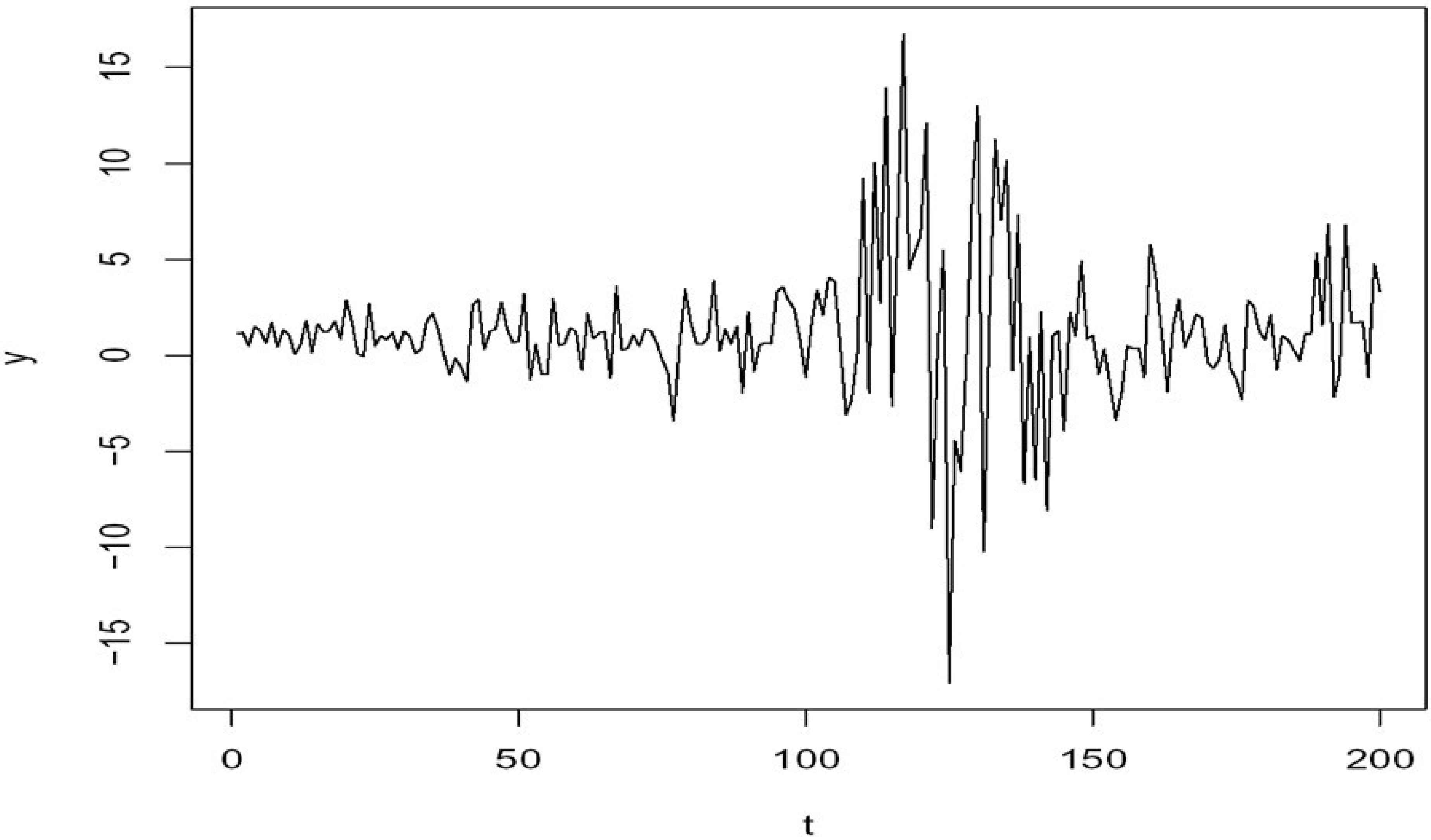
- The unconditional variance σ^2 is constant!

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta} = V_L,$$

provided $\omega > 0$, and $|\alpha + \beta| < 1$.

Simulated GARCH Process

$$\sigma_n^2 = 0.50 + 0.25 u_{n-1}^2 + 0.70 \sigma_{n-1}^2$$



Example

- Suppose $\sigma_n^2 = 0.000002 + 0.13u_{n-1}^2 + 0.86\sigma_{n-1}^2$.
- The long-run variance rate is 0.0002
- Suppose that the current estimate of the volatility is 1.6% per day and the most recent percentage change in the market variable is 1%.
- The new variance rate is

$$0.000002 + 0.13 \times 0.01^2 + 0.86 \times 0.016^2.$$

Consequently, the new volatility is 1.53% per day.

Mean Reversion

- The GARCH (1,1) model recognizes that over time the variance tends to get pulled back to a long-run average level of V_L.

- Stochastic differential equation for the variance V:

$$dV = a(V_L - V)dt + \xi V\,dz,$$

where time is measured in days, $a = 1 - \alpha - \beta$, and $\xi = \alpha\sqrt{2}$.

- Mean reversion: The variance has a drift that pulls it back to V_L at rate a. When $V > V_L$, the variance has a negative drift; when $V < V_L$, it has a positive drift.

Estimating a Constant Variance

- Maximum likelihood method involves choosing values for the parameters that maximize the chance (or likelihood) of observations occurring.

- Example 1: We observe that a certain event happens one time in 10 trials. What is our estimate of the proportion of the time, p, that it happens?

- The probability of the event happening on one particular trial and not on the others is $p(1-p)^9$.

- We maximize this probability to obtain a maximum likelihood estimate. Result: $p = 0.1$ (as expected)

Example 2

- Estimate the variance of observations from a normal distribution with mean zero

$$\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{u_i^2}{2v}\right)$$

- Taking logarithm of the function is equivalent to maximizing

$$\sum_{i=1}^{m} \left(-\ln(v) - \frac{u_i^2}{v}\right)$$

- Result:

$$v = \frac{1}{m}\sum_{i=1}^{m} u_i^2.$$

Application to GARCH

- We choose parameters that maximize

$$\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v_i}} \exp\left(-\frac{u_i^2}{2v_i}\right)$$

or

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

How Good Is the Model?

- If a GARCH model is working well, it should remove the autocorrelation in u_i^2/σ_i^2.

- We can test whether it has done so by considering the autocorrelation structure for the variables u_i^2 If these show very little autocorrelation, our model for σ_i^2 has succeeded in explaining the autocorrelations in u_i^2.

- Ljung-Box statistic for the u_i^2/σ_i^2 time series.
    - Chi-square

Variance Targeting

- One way of implementing GARCH(1,1) that increases stability is by using variance targeting.

- We set the long-run average volatility equal to the sample variance.

- Only two other parameters then have to be estimated.

Forecasting Future Volatility

- The variance rate estimated at the end of day $n-1$ for day n, when GARCH(1,1) is used, is

$$\sigma_n^2 = (1-\alpha-\beta)V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2.$$

- Iffollows that

$$\sigma_n^2 - V_L = \alpha\left(u_{n-1}^2 - V_L\right) + \beta\left(\sigma_{n-1}^2 - V_L\right)$$

- On day $n+t$ in the future and with σ^2,

$$\sigma_{n+t}^2 - V_L = \alpha\left(u_{n+t-1}^2 - V_L\right) + \beta\left(\sigma_{n+t-1}^2 - V_L\right)$$

- The expected value of u_{n+t-1}^2 is σ_{n+t-1}^2. Hence

$$\mathbb{E}\left(\sigma_{n+t}^2 - V_L\right) = (\alpha+\beta)\,\mathbb{E}\left(\sigma_{n+t-1}^2 - V_L\right).$$

Forecasting Future Volatility (cont'd)

- Applying the above equation repeatedly yields

$$\mathbb{E}\left(\sigma_{n+t}^2 - V_L\right) = (\alpha+\beta)^t(\sigma_n^2 - V_L),$$

  which is

$$\mathbb{E}\left(\sigma_{n+t}^2\right) = V_L + (\alpha+\beta)^t(\sigma_n^2 - V_L).$$

- This equation allows you to forecast the volatility on day $n+t$ using the information available at the end of day n.

- **Real-World Application**
  The maintenance margin x is set as, given today's settlement price of F_n,

$$x \geqq 1.645\sqrt{\mathbb{E}\left(\sigma_{n+1}^2\right)}\,F_n$$

Maximum Likelihood Estimators

- With $e_t \overset{d}{\sim} N(0,1)$,consider a model $Y_t = f(\boldsymbol{X};\boldsymbol{\theta}) + e_t$. Then $Y_t - f(\boldsymbol{X};\boldsymbol{\theta})$ is distributed as i.i.d. N 0, σ . Its probability density function is

$$\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2}\left(\frac{Y_t - f(\boldsymbol{X};\boldsymbol{\theta})}{\sigma_e}\right)^2\right)$$

- Given a sample size N of $\{Y_1, Y_2, \ldots, Y_N$, and $\boldsymbol{X}\}$, the likelihood function is

$$L = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}}\right)^N \exp\left(-\frac{1}{2}\sum_{t=1}^{N}\left(\frac{Y_t - f(\boldsymbol{X};\boldsymbol{\theta})}{\sigma_e}\right)^2\right)$$

- Suppose we find estimate values $\boldsymbol{\theta}$ and σ_e that maximizes this log-likelihood function log L. Then the estimates are called Maximum Likelihood estimates.

Cramer-Rao Inequality

- The parameters to be estimated and data are assembled into vectors:

$$\boldsymbol{\Lambda} := \begin{pmatrix} \boldsymbol{\theta} \\ \sigma_e \end{pmatrix} \quad \text{and} \quad \boldsymbol{Z} := (\boldsymbol{Y} \ \ \boldsymbol{X})$$

- Probabilities add up to 1:

$$\int_{-\infty}^{\infty} L(\boldsymbol{Z}\,;\,\boldsymbol{\Lambda})\,d\boldsymbol{Z} = 1; \qquad \int_{-\infty}^{\infty} \frac{\partial L(\boldsymbol{Z}\,;\,\boldsymbol{\Lambda})}{\partial \boldsymbol{\Lambda}}\,d\boldsymbol{Z} = 0$$

- **Fisher's information matrix**

$$\boldsymbol{R}(\boldsymbol{\Lambda}) := -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial \boldsymbol{\Lambda} \partial \boldsymbol{\Lambda}^\top}\right)$$

- Let $h(\boldsymbol{Z})$ be an unbiased estimator of $\boldsymbol{\Lambda}$. Then

$$\mathbb{C}\big(h(\boldsymbol{Z})\big) \geq \boldsymbol{R}^{-1}$$

- Question: What's the interpretation of Cramer-Rao Inequality?

Volatility Term Structures

- Suppose it is day n. Define

$$V(t) := \mathbb{E}\left(\sigma_{n+t}^2\right), \qquad \text{and} \qquad a := \ln \frac{1}{\alpha + \beta}.$$

- The predictive equation becomes

$$V(t) = V_L + e^{-at}\big(V(0) - V_L\big).$$

Here, $V(t)$ is an estimate of the instantaneous variance rate in t days.

Volatility Term Structures (cont'd)

- The average variance rate per day between today and time T is given by

$$\frac{1}{T}\int_0^T V(t)\,dt = V_L + \frac{1-e^{-aT}}{aT}\big(V(0)-V_L\big).$$

- Then the volatility per annum for an option lasting T days is

$$\sigma(T) = \sqrt{252\left(V_L + \frac{1-e^{-aT}}{aT}(V(0)-V_L)\right)}.$$

- So from GARCH(1,1), we have a volatility term structure, which is the relationship between the forward-looking volatilities and the maturities.

S&P Volatility Term Structure Predicted from GARCH(1,1)

- Note that a is positive since $\alpha + \beta < 1$.
- ω= 0.0000013465, α = 0.083394, and β = 0.910116.

$$a = \ln\left(\frac{1}{0.083394 + 0.910116}\right) = 0.006511$$

Option Life (days)	10	30	50	100	500
Volatility (% per annum)	27.4	27.1	26.9	26.4	24.3

Impact of Volatility Changes

- We note that $V(0) = \sigma(0)^2/252$.
- When instantaneous volatility $\sigma(0)$ changes by $\Delta\sigma(0)$, volatility for T-day maturity changes by approximately

$$\Delta\sigma(T) \approx \frac{1 - e^{-aT}}{aT} \frac{\sigma(0)}{\sigma(T)} \Delta\sigma(0).$$

- Impact of 1% change in the instantaneous volatility predicted from GARCH (1,1)

Option Life (days)	10	30	50	100	500
Volatility increase (%)	0.97	0.92	0.87	0.77	0.33

Summary

- In the EWMA and the GARCH(1,1) models, the weights assigned to observations decrease exponentially as the observations become older.
- The GARCH(1,1) model differs from the EWMA model in that some weight is also assigned to the long-run average variance rate. It has a structure that enables forecasts of the future level of variance rate to be produced relatively easily.
- Maximum likelihood methods are usually used to estimate parameters from historical data in the EWMA, GARCH(1,1), and similar models.
- Once its parameters have been determined, a GARCH(1,1) model can be judged by how well it removes autocorrelation from the u^2.