Volatility of Asset Returns

We can almost directly observe the return (simple or log) of an asset over any given period.

All that it requires is the observed price at the beginning of the period and the observed price at the end of the period.

We are interested in how much variability there is in these returns from period to period.

The variability in the returns is called the “volatility”.

Can we measure it?

Not as easy as we might think.

First, we need to define it more precisely.
Volatility of Asset Returns

As a first stab, we will tie it to a well-understood abstraction in probability models.

We will associate the volatility of an asset with the standard deviation of the asset return.

We notice some incompleteness in this statement.

We haven’t specified the period, and we haven’t specified the type of return (simple or log).
Volatility of Asset Returns

We will choose 1 year as the time interval.

To emphasize this we sometimes say the “annualized return” and the “annualized volatility”.

Now we have an idealized definition of volatility of an asset; it is a standard deviation of a random variable representing asset returns.

We will continue also to use “volatility” in a non-technical sense.
**Volatility of Asset Returns**

While the definition of volatility as a standard deviation of annualized returns fixes in our mind the general idea, it is far from being a workable definition, or even a realistic one.

The first issue is whether the distribution of the return varies over time.

Recall the plot of the *daily* log returns of the S&P 500 Index of two weeks ago.
Stochastic Volatility

Without really addressing the issue of how the volatility changes over time, let’s just agree to think of the variance as a time-varying measure, so we will denote it by $\sigma_t^2$.

We can think of it as a random variable that we can express as a conditional expectation.

We condition on what has happened before. Let $\mathcal{F}_{t-1}$ represent “all that has happened up to time $t$”. ($\mathcal{F}_{t-1}$ is the $\sigma$-field generated by all previous random variables in the stochastic process.)

We now define volatility as a conditional variance, conditioned on what has happened before:

$$\sigma_t^2 = \mathbb{V}(R_t|\mathcal{F}_{t-1}).$$

In the finance literature, we call this “stochastic volatility”.
Length of Period of the Returns

The real problem, of course, is to estimate $\sigma_t^2$.

This is related to the problem of defining a sample statistic that corresponds to this concept of volatility.

(Note that these are actually two separate problems; we may want to make inferences about a probability model, or we may just want to describe observed data.)

First, we note that a one-year period for measuring returns seems somewhat excessive.

How could we realistically expect to have data from enough years to be able to say anything about the volatility over several years.

(Who’s even interested in the volatility over several years?)
Length of Period of the Returns

Beauty of the log returns...

Suppose $m$ counts time in months, and we have the log returns

$$r_m, r_{m+1}, \ldots, r_{m+11}$$

and so the log return in one year

$$r_y = r_m + r_{m+1} + \cdots + r_{m+11},$$

where the time represented by "$y$" is the same time as represented by "$m + 11$".

So how does the variance of the data-generating process giving rise to $$\ldots, r_{m-1}, r_m, r_{m+1} \ldots$$ compare with the variance of the data-generating process giving rise to $$\ldots, r_{y-1}, r_y, r_{y+1} \ldots$$?
Length of Period of the Returns

Let’s wave our hands (and assume 0 correlations).

We have

\[ V(\text{yearly returns}) \approx 12 V(\text{monthly returns}) \]

or

\[ \sigma_y \approx \sqrt{12} \sigma_m, \]

where \( \sigma_y \) is the volatility of the yearly log returns, and \( \sigma_m \) is the volatility of the monthly log returns.

We will generally use the yearly log returns, however, for exploratory analyses and plotting, we will often use the daily returns.
Stationarity of Returns of Financial Assets

Let’s return to a more pressing issue; modeling returns.

Are the returns stationary?

Most glaring departure: nonconstant variance.
Models of Returns of Financial Assets

The concept of volatility can be well-described as a parameter of a probability distribution, but of course we cannot directly observe it, and it’s not clear how we could relate any observable quantities to this measure.

Volatility is one of the most important characteristics of financial data. It is one of the basic considerations in
• asset allocation,
• managing financial portfolios,
• pricing derivative assets, and
• measuring value at risk.

The fact that it is time dependent complicates any practical application.

Measuring volatility is one of the fundamental problems in finance.
Measuring and/or Estimating Volatility

We can consider some quantity that is computable from observations and that corresponds in a meaningful way to the concept of volatility.

One way of approaching the problem is to assume that over short intervals of time the volatility is relatively constant. If we also assume the the returns have zero serial correlations, we could use the sample variance.

```r
IBM_mc <- get.stock.price("IBM", 
    start.date=c(1,1,2011),stop.date=c(12,31,2013),freq="m")
IBM_mr <- log.ratio(IBM_mc)
sqrt(12)*stdev(IBM_mr)
[1] 0.1411888

IBM_mc <- get.stock.price("IBM", 
    start.date=c(1,1,2009),stop.date=c(12,31,2011),freq="m")
IBM_mr <- log.ratio(IBM_mc)
sqrt(12)*stdev(IBM_mr)
[1] 0.1418932
```
Historical or Statistical Volatility

The standard deviation of a sequence of observed returns, as computed on the previous slide, is called the “historical volatility” or “statistical volatility”.

It is rarely as constant over different time periods, as in the case of IBM that we considered.

For Dollar General, for example, over the same two periods it was 0.195 and 0.248.
Implied Volatility

Another way of estimating the volatility would be to use an invertible model that relates the volatility to observable quantities:

\[ y_t = f(x_t, \sigma_t), \]

where \( y_t \) and \( x_t \) are observable. (Remember \( y_t \) and \( x_t \) are vectors!)

Plugging in observed values for \( y_t \) and \( x_t \) and solving for \( \sigma_t \) yields an “implied volatility”.

The standard model for this is the model relationship between the fair value of a specific stock option of a certain type and the current price of the stock.
Options

There are various types of options.

Some types of options give the owner the right to buy or sell an asset ("exercise" the option) at a specific price ("strike" price) at a specific time ("expiry") or at specific times, or anytime before a specific time frame.

The asset that can be bought or sold is called the "underlying".

If the option gives the right to buy, it is called a "call option".

If the option gives the right to sell, it is called a "put option".

A "European" option can only be exercised at a specific time.

An "American" option can be exercised at anytime there is trading in the underlying before a specific time.

Almost all stock options for which there is a market are American options.
Options

An option is created by a “sell to open” or a “buy to open” order.

An option is terminated by a “sell to close” or a “buy to close” order.

The price of the transaction is called the “option premium”.

As with any negotiable instrument, there are two sides to a trade.
Options

Another class of options is based on an “underlying” that cannot be bought or sold (such as a stock index or some measure of the weather); hence, the option is settled for cash, rather than being exercised.

Most options of this class for which there is a market are European options.

The only options of this class that require settlement are those that are open at expiry (European options). Some options of this class also allow settlement at other specific dates.

Such options behave in other respects as options on negotiable underlyings.
Option Values

The value of an option is obviously a decreasing function of the strike price, call it $K$, minus the current price of the underlying, call it $S_t$, and an increasing function of the expiry, call it $T$, minus the current time, call it $t$.

The relationship between $K$ and $S_t$ is called the “moneyness”.

For call options,
If $K > S_t$ the option is “out-of-the-money”
If $K = S_t$ the option is “at-the-money”
If $K < S_t$ the option is “in-the-money”, and the difference $S_t - K$ is called the “intrinsic value”; otherwise, the intrinsic value defined as 0.

For put options, the same terms are used for the opposite relationships between $K$ and $S_t$. 
Option Values

The value of an option minus its intrinsic value is called the “time value” of the option.

For some types of options, under certain circumstances, the time value can be negative.

The value of an option also depends on the volatility of the underlying.
Implied Volatility Using Option Values and Prices

If we have a formula that relates the value of an option to observable quantities, we can observe the market price of the option, assume that price is the same as the value of the option, and solve for the volatility.
Implied Volatility Using the Black-Scholes Formula

At time $t$ for a stock with price $S$ that has a dividend yield of $q$, given a risk-free interest rate of $r$, the Black-Scholes pricing formula for a European call option with strike price $K$ and expiry date $T$ is

$$C_{BS}(t, S) = S e^{-q(T-t)} \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2),$$

where

$$d_1 = \frac{\log(S/K) + (r - q + \frac{1}{2}\sigma^2)(T - t)}{\sigma \sqrt{T - t}},$$

and

$$d_2 = d_1 - \sigma \sqrt{T - t}.$$
Pricing Options

The Black-Scholes pricing formula for a European call option is based on a fairly simple model of geometric Brownian motion.

The factors $e^{-q(T-t)}$ and $e^{-r(T-t)}$ come from the “carrying cost” of owning the stock.

There is a similar pricing formula for a European put option.

There would be an arbitrage opportunity would exist if there was not a fixed relationship between the call and put prices. It is called the put-call parity:

$$P = C - S e^{-q(T-t)} + Ke^{-r(T-t)}.$$
Implied Volatility Using the Black-Scholes Formula

The Black-Scholes formula is the simplest equation of the general type of function $y_t = f(x_t, \sigma_t)$ alluded to above.

$y_t$ is the observed price of the option and $x_t$ is all the other stuff.

We do not have a closed form for $\sigma_t$ from the Black-Scholes formula, but we can solve it numerically.
Implied Volatility Using Prices of European Options

The function `EuropeanOptionImpliedVolatility` in RQuantLib numerically solves for the volatility in the Black-Scholes formula for pricing European options.

It is part of a software suite called QuantLib, developed by StatPro, which is a UK risk-management provider.

One of the most popular underlyings for European options is the CBOE S&P 100 index (OEX).
Implied Volatility Using Prices of OEX European Options

Using the 14 March 810 options on the OEX we get

```r
library(RQuantLib)
Tminust <- as.numeric(difftime("2014-03-22","2014-02-26"))/365.24

EuropeanOptionImpliedVolatility(type='call', value=9.55, underlying=812.26, strike=810, dividendYield=0, riskFreeRate=0.0003, maturity=Tminust, volatility=0.14)

$impliedVol
[1] 0.1000934

EuropeanOptionImpliedVolatility(type='put', value=9.25, underlying=812.26, strike=810, dividendYield=0, riskFreeRate=0.0003, maturity=Tminust, volatility=0.14)

$impliedVol
[1] 0.1238616
```
Implied Volatility

The two values are different. (Surprise!) And if we’d do some more, they’d be different too.

This ain’t no exact science.
Prices of OEX European Options

Although the bid and ask prices for deep in-the-money options on OEX are set artificially, the midpoint price may have negative time value.

On February 26, 2014, for example, when the value of the OEX was 812.26, the bid and ask on the March 700 call were 110.70 and 112.60. Using the midpoint, 111.65, we have a negative time value of 0.61.

The function `EuropeanOptionImpliedVolatility` in RQuantLib will not work when the time value is negative.

```r
EuropeanOptionImpliedVolatility(type='call', value=111.65, underlying=812.26, strike=700, dividendYield=0, riskFreeRate=0.0003, maturity=Tminust, volatility=0.14)
```

Pricing American Options

European options can be exercised only at expiry, whereas American options can be exercised any time the market is open prior to expiry.

This means that the SDE model of the geometric Brownian motion has free boundary conditions, and so a closed-form solution is not available (although there are some fairly good closed-form approximations).

It is never optimal to exercise an American call option on a non-dividend paying stock prior to expiry; therefore, for all practical purposes the price of an American call is the same as that of a European call with the same characteristics.
Implied Volatility

We use current prices of call options, which we can get at Yahoo Finance, together with the current price of the underlying to solve for $\sigma$ in the formula on the previous slide. (Yahoo Finance gives only one chain and it’s a weekly for months that have open weeklys.)

For example, on February 14, 2014, we had
IBM last price 183.69
dividend yield 2.10%
consider April call and put at a strike of 185 (expiry Apr 19)
Apr 185c last price 4.50
Apr 185p last price 5.40
$r = 0.0003$ (ZIRP!)
Implied Volatility

The function `AmericanOptionImpliedVolatility` in RQuantLib uses finite differences to solve the SDE for pricing American options. (As the default, it uses 150 time steps at 151 gridpoints.)

```r
library(RQuantLib)

AmericanOptionImpliedVolatility(type='call', value=4.5,
                   underlying=183.69, strike=185, dividendYield=0.021,
                   riskFreeRate=0.0003, maturity=Tminust, volatility=0.14,
                   timeSteps=150, gridPoints=151)

$impliedVol
[1] 0.1756564

AmericanOptionImpliedVolatility(type='put', value=5.4,
                   underlying=183.69, strike=185, dividendYield=0.021,
                   riskFreeRate=0.0003, maturity=Tminust, volatility=0.14,
                   timeSteps=150, gridPoints=151)

$impliedVol
[1] 0.141408
```
Implied Volatility

Again, the two values are different. (Surprise!)

We need an authority to tell us what the volatility is.

(And an authority who will do things for money.)
Measuring the Volatility of the Market

A standard measure of the overall volatility of the market is the CBOE Volatility Index, VIX, which CBOE introduced in 1993 as a weighted average of the Black-Scholes-implied volatilities of the S&P 100 Index (OEX — see a previous slide) from at-the-money near-term call and put options.

(“At-the-money” is defined as the strike price with the smallest difference between the call price and the put price.)

In 2004, futures on the VIX began trading on the CBOE Futures Exchange (CFE), and in 2006, CBOE listed European-style calls and puts on the VIX.

Another measure is the CBOE Nasdaq Volatility Index, VXN, which CBOE computes from the Nasdaq-100 Index, NDX, similarly to the VIX. (Note that the more widely-watched Nasdaq Index is the Composite, IXIC.)
The VIX

In 2006, CBOE changed the way the VIX is computed. It is now based on the volatilities of the S&P 500 Index implied by several call and put options, not just those at the money, and it uses near-term and next-term options (where “near-term” is the earliest expiry more than 8 days away).

It is no longer computed from the Black-Scholes formula.

It uses the prices of calls with strikes above the current price of the underlying, starting with the first out-of-the-money call and sequentially including all with higher strikes until two consecutive such calls have no bids. It uses the prices of puts with strikes below the current price of the underlying in a similar manner.

The price of an option is the “mid-quote” price, i.e. the average of the bid and ask prices.
Let $K_1 = K_2 < K_3 < \cdots < K_{n-1} < K_n = K_{n+1}$ be the strike prices of the options that are to be used.

The VIX is defined as $100 \times \sigma$, where

$$\sigma^2 = \frac{2e^{rT}}{T} \left( \sum_{i=2; i \neq j}^{n} \frac{\Delta K_i}{K_i^2} Q(K_i) + \frac{\Delta K_j}{K_j^2} \left( Q(K_j \text{put}) + Q(K_j \text{call}) \right) / 2 \right)$$

$$- \frac{1}{T} \left( \frac{F}{K_j} - 1 \right)^2,$$

- $T$ is the time to expiry (in our usual notation, we would use $T - t$, but we can let $t = 0$),
- $F$, called the “forward index level”, is the at-the-money strike plus $e^{rT}$ times the difference in the call and put prices for that strike,
- $K_i$ is the strike price of the $i^{th}$ out-of-the-money strike price (that is, of a put if $K_i < F$ and of a call if $F < K_i$),
- $\Delta K_i = (K_{i+1} - K_{i-1})/2$,
- $Q(K_i)$ is the mid-quote price of the option,
- $r$ is the risk-free interest rate, and
- $K_j$ is the largest strike price less than $F$. 

33
Computing the VIX

Time is measured in minutes, and converted to years.

Months are considered to have 30 days and years are considered to have 365 days.

There are \( N_1 = 1,440 \) minutes in a day.
There are \( N_{30} = 43,200 \) minutes in a month.
There are \( N_{365} = 525,600 \) minutes in a year.

A value \( \sigma_1^2 \) is computed for the near-term options with expiry \( T_1 \), and a value \( \sigma_2^2 \) is computed for the next-term options with expiry \( T_2 \), and then \( \sigma \) is computed as

\[
\sigma = \sqrt{ \left( T_1 \sigma_1^2 \frac{N_{T_2} - N_{30}}{N_{T_2} - N_{T_1}} + T_2 \sigma_1^2 \frac{N_{30} - N_{T_1}}{N_{T_2} - N_{T_1}} \right) \frac{N_{365}}{N_{30}}}.
\]
The VIX Daily Closes

VIX Daily Prices

Jan 1, 2008 to Dec 31, 2013
The VIX

The VIX is an index, "^VIX".

It is not traded; rather, futures are traded on it.

On 2011-07-08 VIX was 15.95 and 11 Nov 20c was $3.40; on 2011-08-04 VIX was 31.66 and 11 Nov 20c was $5.50.

On 2011-09-02 VIX was 33.50 and 12 Jan 30p was $4.40; on 2011-12-21 VIX was 21.43 and 12 Jan 30p was $6.00.

Those were not bad profits.

The VIX has been no fun since 12 Jun.
Modeling Returns

Now, let’s return to the problem of modeling the asset returns. We see that because of the nonconstant variance, an ARMA model by itself is not going to make it.

Let’s start with the conditional mean and variance of the returns:

\[ \mu_t = \mathbb{E}(R_t|\mathcal{F}_{t-1}) \]

and

\[ \sigma_t^2 = \mathbb{V}(R_t|\mathcal{F}_{t-1}). \]

The first, and simplest question regards the conditional mean function.

Does it exhibit serial correlations?

If so, we should try to take care of them first, maybe with an ARMA model.
Characteristics of the Returns

Recall Tsay’s analysis of monthly log returns for IBM for 1926-1-1 to 2008-12-31 on pages 20 and 33.

He found no significant autocorrelations.

How about for 2009-1-1 to 2013-12-31?
Serial Characteristics of the Returns

What about the autocorrelations?

ACF for IBM Monthly Returns
Serial Characteristics of the Returns

> Box.test(IBM_mr,lag=5,type="Ljung")

      Box-Ljung test

data:  IBM_mr
X-squared = 5.6456, df = 5, p-value = 0.3422

This is consistent with Tsay's results on page 32 for the earlier and longer time period.
Serial Characteristics of the Returns

What do we expect about the serial properties of the log returns?

If we were to look at several different stocks such as IBM, we would find tremendous variation.

Both Tsay’s analysis of log returns for IBM for 1926-1-1 to 2008-12-31 and ours for 2009-1-1 to 2013-12-31 showed no significant autocorrelations.

But how about the S&P 500?

While the S&P 500 is not necessarily representative or an “average”, if is usually more instructive to study it.
Serial Characteristics of the Returns

Recall the plot of the S&P 500 log returns:

S&P 500 Daily Returns

Jan 1, 2008 to Dec 31, 2013
Serial Characteristics of the Returns

What about the autocorrelations? Look at the ACF and the portmanteau test.
Serial Characteristics of the Returns

> Box.test(GSPC_dr, lag=5, type="Ljung")

Box-Ljung test

data:  GSPC_dr
X-squared = 34.2285, df = 5, p-value = 2.144e-06
Serial Characteristics of the Monthly Index

Returns

The serial characteristics of a time series often depends on the frequency of the series.

This is especially true for financial time series, and it is especially true for time series of rates of return (because, among other reasons, the return depends on the interval).

Tsay says (p. 113) that the “daily returns of a market index often show minor serial correlations, but monthly returns of the index may not contain any significant serial correlation.”

Well, let’s try the monthly returns of the most watched index of all.

Look at the ACF and the portmanteau test.
Serial Characteristics of the Returns

ACF for S&P 500 Monthly Returns
Serial Characteristics of the Returns

> Box.test(GSPC_dr,lag=5,type="Ljung")

Box-Ljung test

data:  GSPC_mr
X-squared = 12.7043, df = 5, p-value = 0.02631

Unfortunately, there is a lot of variability in the characteristics of returns. Some have stronger serial correlations than others.
Portmanteau Tests of Serial Correlations of the Monthly Returns of the Dow Stocks

P-values of Box-Ljung test.

<table>
<thead>
<tr>
<th>Stock</th>
<th>p-value</th>
<th>Stock</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXP</td>
<td>0.04273</td>
<td>MCD</td>
<td>0.7316</td>
</tr>
<tr>
<td>T</td>
<td>0.1233</td>
<td>MRK</td>
<td>0.5474</td>
</tr>
<tr>
<td>BA</td>
<td>0.5637</td>
<td>MSFT</td>
<td>0.02527</td>
</tr>
<tr>
<td>CAT</td>
<td>0.01994</td>
<td>NKE</td>
<td>0.09713</td>
</tr>
<tr>
<td>CVX</td>
<td>0.1215</td>
<td>PFE</td>
<td>0.8731</td>
</tr>
<tr>
<td>CSCO</td>
<td>0.2347</td>
<td>PG</td>
<td>0.2813</td>
</tr>
<tr>
<td>DD</td>
<td>0.5078</td>
<td>KO</td>
<td>0.07811</td>
</tr>
<tr>
<td>XOM</td>
<td>0.4612</td>
<td>MMM</td>
<td>0.5517</td>
</tr>
<tr>
<td>GE</td>
<td>0.06802</td>
<td>TRV</td>
<td>0.1157</td>
</tr>
<tr>
<td>GS</td>
<td>0.7577</td>
<td>UTX</td>
<td>0.7398</td>
</tr>
<tr>
<td>HD</td>
<td>0.7477</td>
<td>UNH</td>
<td>0.5675</td>
</tr>
<tr>
<td>INTC</td>
<td>0.188</td>
<td>VZ</td>
<td>0.1786</td>
</tr>
<tr>
<td>IBM</td>
<td>0.1562</td>
<td>V</td>
<td>0.08173</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.2867</td>
<td>WMT</td>
<td>0.02941</td>
</tr>
<tr>
<td>JPM</td>
<td>0.07467</td>
<td>DIS</td>
<td>0.0731</td>
</tr>
</tbody>
</table>
How About the Dow Itself?

> Box.test(DJI_mr,lag=5,type="Ljung")

Box-Ljung test

data:  x_mr
X-squared = 10.0765, df = 5, p-value = 0.0731

Based on the foregoing examples, an ARMA model for the conditional mean may be called for (rather than just going for an ARCH or GARCH model immediately, as Tsay does in some examples).
Models of Returns of Financial Assets

Now, let’s again return to the problem of modeling the asset returns.

Given that the conditional mean and variance of the returns are time dependent, we consider models that capture that dependence.

1. We first model the conditional mean.

\[ \mu_t = \mathbb{E}(R_t | \mathcal{F}_{t-1}) \]

2. After fitting the mean, we model the return as the mean plus a shock,

\[ R_t = \mu_t + A_t, \]

and then model the conditional variance.

\[ \sigma_t^2 = \mathbb{V}(R_t | \mathcal{F}_{t-1}). \]

3. We evaluate our model from steps 1. and 2., and iterate as the data indicate.
Step 1: Modeling the Conditional Mean

The general expression for the conditional mean is

$$\mu_t = \mathbb{E}(R_t|\mathcal{F}_{t-1}).$$

Is there a trend? This may be answered by visual inspection.

If there is a trend, can we model it:

$$\mu_t = f(t, \theta)?$$

If we can, we fit the model and look at residuals ("detrend" the data).

If the mean is a polynomial in time, that is, if

$$\mu_t = \beta_0 + \beta_1 t + \cdots + \beta_d t^d,$$

there is a simple way to detrend. (What is it?)
Step 1: Modeling the Conditional Mean

After detrending, we check for serial correlations. (How? What are the tools?)

If there are serial correlations, we hope we can account for them with something simple, such as a low-order ARMA. (When we have a hammer, everything looks like a nail.)

The ARMA model might also include other variables, as in equation (3.3) in Tsay. (We will not explore the multitude of possibilities here.)

There are several possibilities beyond ARMA models.

Seasonality is a big issue for economic data.

There are several things that can be done about seasonality: seasonal ARIMA models, seasonal adjustment, etc.

These steps are still part of modeling the conditional mean.
In this course, we are not so concerned about seasonality as about something else that obviously occurs in asset returns: constancy of the variance of the residuals.

This is always a big issue in any modeling. (Recall in regression modeling, often the variance of the errors is proportional to the magnitude of the dependent variable.)

In modeling asset returns, the variation in the variance (the “stochastic volatility”) often has a serial correlation, and that is our main concern now.
2. After fitting the mean, we model the return as the mean plus a shock,

\[ R_t = \mu_t + A_t, \]

and then model the conditional variance of the shocks, or residuals,

\[ \sigma_t^2 = \text{V}(A_t | \mathcal{F}_{t-1}). \]

We first decide whether the residuals indicate that the mean model errors were iid.

If the residuals are iid, our model of the mean is adequate, and so we stop.

There are several alternative possibilities, but our main concern will be whether they have serial correlations.
Step 2: Testing for the ARCH Effect and Modeling Variance

The question is whether the errors have serial correlations.

We apply the Ljung-Box portmanteau test (Box.test in R) up to about 12 lags to the squared residuals (see Tsay, page 114).

There are other tests for this also, but we won’t use them.

The presence of serial correlations in the residuals from an ARMA fits is sometimes called an “ARCH effect” because of the type of model that can be used to account for them.

If there are serial correlations, we try a simple AR model for \{A_t\}, but with a difference; we allow Conditional Heteroscedasticity.
Autoregressive with Conditional Heteroscedasticity (ARCH)

We will assume that there is an underlying strong white noise process (that is, an iid process) \( \{\epsilon_t\} \) with mean 0 and variance 1, such that

\[
A_t = \sigma_t \epsilon_t
\]

and

\[
\sigma_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2,
\]

with \( \alpha_0, \alpha_1 > 0 \).

This is called an ARCH(1) model. (There are other formulations of the ARCH model.)

The key characteristic is that while the unconditional variance is constant (in this formulation), the **conditional variance**, 

\[
V(A_t|\mathcal{F}_{t-1})
\]

is not.
ARCH(1); Another Formulation

Again, assuming a strong white noise process \( \{ \epsilon_t \} \), we define the process \( \{ A_t \} \) by the conditions

\[
E(A_t|\mathcal{F}_{t-1}) = 0,
\]

and

\[
A_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2 + \epsilon_t,
\]

with \( \alpha_1 > 0 \), \( \alpha_0 \geq \epsilon_t \) for any allowable value of \( \epsilon_t \).

Aside from this last condition (which is satisfied by a bound on the probability), this formulation is easier to work with.
ARCH(1)

Using the alternate formulation, we easily get the properties in Section 3.4.1 of Tsay in terms of conditional expectations given the state at a previous point.

At any lag \( h \), we have

\[
E(A_t|\mathcal{F}_{t-h}) = 0,
\]

\[
\text{Cov}(A_t, A_{t+k}|\mathcal{F}_{t-h}) = 0
\]

and

\[
\text{V}(A_t|\mathcal{F}_{t-h}) = \alpha_0 \frac{1 - \alpha_1^h}{1 - \alpha_1} + \alpha_1^h E_{t-h}^2.
\]

We get this latter expression by taking the recursion

\[
A_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2 + \epsilon_t
\]

back \( h \) steps.
ARCH(1); The Normality Assumption and Stationarity of Higher-Order Moments

For higher moments we need to make additional assumptions about the white noise or the distribution of $A_t$.

If we assume that $A_t|F_{t-1} \sim N(0, \alpha_0 + \alpha_1 A_{t-1}^2)$, we can work out the conditional even-degree moments.

These moments are stationary only if $3\alpha_1^2 < 1$.

For the fourth moment, we get

$$E(A_t^4) = \frac{3\alpha_0^2}{(1 - \alpha_1)^2} \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}$$

and the kurtosis $E(A_t^4)/(E(A_t^2))^2$ is greater than 3.
ARCH\((m)\)

Engle (1982) introduced the ARCH(1) model and the generalization ARCH\((m)\).

For ARCH\((m)\), in the same formulation of the ARCH(1), we assume that there is an underlying strong white noise process (that is, an iid process) \(\{\epsilon_t\}\) with mean 0 and variance 1, such that

\[
A_t = \sigma_t \epsilon_t
\]

and

\[
\sigma_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2 + \cdots + \alpha_m A_{t-m}^2,
\]

with \(\alpha_0, \alpha_m > 0\) and \(\alpha_i \geq 0\) for \(i = 1, \ldots, m - 1\).
ARCH\((m)\); Another Formulation

If we let \( \eta_t = A_t - \sigma_t^2 \), and plug this into the general ARCH model equation, we get the kind of expression we showed before:

\[ A_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2 + \cdots + \alpha_m A_{t-m}^2 + \eta_t, \]

which really looks like an AR model. … Except

what about the distribution of \( \{\eta_t\} \)?

Remember, \( \{\epsilon_t\} \) is an iid process with mean 0.

This means that \( \{\eta_t\} \) has mean 0.

The independence of \( \{\epsilon_t\} \) means that \( \{\eta_t\} \) has 0 correlation. What property of the model assures this?

However, \( \{\eta_t\} \) is neither i nor i.
Fitting an ARCH($m$) Model

The first problem in working with an ARCH($m$) model is the choice of $m$.

This choice shares all the caveats I’ve given about any process of building models. The autoregressive nature of the model leads us to the same tool we used for selecting the order of an AR($p$) model. The partial autocorrelation at lag greater than the order of the model is 0.

Hence, we use the PACF of the things we’re going to model, and choose the ARCH order as the last lag that appears to be different from 0. (I say “the things we’re going to model” because they’re not the returns themselves.)

Once the order is fixed the maximum likelihood is the best way to estimate the parameters.
Fitting an ARCH($m$) Model

An ARCH($m$) model is most commonly fit by maximum likelihood (ML).

To use ML, we do **four** things:
- assume a family of probability models for the observable random variable(s),
- write the probability model for the sample of observable random variables,
  (This is a PDF with unknown parameters or unknown form.)
- change arguments of the probability model and assume the sample of random variables have been replaced by given constants,
  (This new expression is the “likelihood”.)
- determine the value of all variable terms that maximize the likelihood.
Terminology

Tsay on pages 120 through 122 develops the conditional likelihood functions for the normal and the Student’s t distributions and two variations of these called “skew Student’s t” distribution and “generalized error” distribution (GED).

I was puzzled by his reference to the likelihoods as “prior” models and likelihoods in four places on pages 120 and 121.

One interpretation of “prior” is as a general family before specific values for some parameters are assigned (by ML or some other means). I finally decided that he just meant “previous” by “prior”.

There are some terms that have specific common meanings in statistical applications.

Terms, such as “consistent”, “efficient”, “prior”, etc., that have specific technical meanings should not be used by statisticians without qualifiers unless the terms are to carry that technical meaning
Maximum Likelihood in Time Series Models

To use ML in fitting an ARCH($m$) model, we first must assume a family of probability models for the observable random variable.

Recall in the development of the ARCH model, we first introduced it as a model of something that was not observable.

We then applied it to a “fitted observable”.

Now, in applications of the model

$$A_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2 + \cdots + \alpha_m A_{t-m}^2,$$

or

$$A_t^2 = \alpha_0 + \alpha_1 A_{t-1}^2 + \cdots + \alpha_m A_{t-m}^2 + \eta_t,$$

$A_t$ is observable.
Maximum Likelihood in Time Series Models

So now we assume a family of probability models for the observable random variable.

$A_t$ is the observable random variable, but we should model its distribution in terms of either $\epsilon_t$ or $\eta_t$.

Which can you do?

Here is the basic problem in use of ML in time series models.

The distribution of the observables is conditional.

Let’s choose a probability model for the iid random variables, $\{\epsilon_t\}$. 
So for step 1, we choose a family of probability models based on an assumed distribution of \( \{\epsilon_t\} \).

Remember, the model specifies that \( \{\epsilon_t\} \) has constant mean of 0 and constant variance of 1.

In a very general modeling problem, we could let the form of the distribution of \( \{\epsilon_t\} \) be part of the maximization process.

That’s hard. Let’s not do it.

Let’s choose \( N(0,1) \) as the distribution of \( \{\epsilon_t\} \).

We’ll use this for the next few slides.

OK. The next step is to write the probability model for the sample of observable random variables. This is a joint PDF.

Let’s assume we will have \( a_1, \ldots, a_n \). (Tsay uses “\( T \)” for “\( n \)”.)
Conditional Distributions

In time series applications, we usually have to write the joint PDF in terms of the conditional PDFs.

I'll say that again.
Maximum Likelihood in Time Series Models

In time series applications, we usually have to write the joint PDF in terms of the conditional PDFs.

For the parameter $\alpha = (\alpha_1, \ldots, \alpha_m)$, the joint PDF associated with $a_1, \ldots, a_n$ is

$$
f(a_1, \ldots, a_n; \alpha) = f(a_n; \mathcal{F}_{n-1}) f(a_{n-1}; \mathcal{F}_{n-2}) \cdots f(a_{m+1}; \mathcal{F}_m) f(a_1, \ldots, a_m; \alpha)
$$

$$
= \prod_{t=m+1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-a_t^2/2\sigma_t^2} f(a_1, \ldots, a_m; \alpha).
$$

Now, the next thing we usually do in using ML in time series applications is figure out what to do with $f(a_1, \ldots, a_m; \alpha)$; that is, with the joint PDF of the initial state of the model.
Maximum Likelihood in Time Series Models

The joint PDF of the initial state of the model, \( f(a_1, \ldots, a_m; \alpha) \) is likely to be quite complicated, and it is even in this simple case.

What can we do?

The standard thing is just to consider \( a_{m+1}, \ldots, a_n \) and to condition on \( a_1, \ldots, a_m \).

This doesn’t have much effect if \( n \) is large relative to \( m \), which it should be to estimate \( m \) parameters.

So we just need the conditional PDF \( f(a_{m+1}, \ldots, a_n; \alpha, a_1, \ldots, a_m) \).

We get it in the usual way:

\[
f(a_{m+1}, \ldots, a_n; \alpha, a_1, \ldots, a_m) = \prod_{t=m+1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-a_t^2/2\sigma_t^2}.
\]
Maximum Likelihood in Time Series Models

Now, the next step is to write the conditional likelihood:

\[ L(\alpha; a_1, \ldots, a_m, a_{m+1}, \ldots, a_n) = \prod_{t=m+1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-a_t^2/2\sigma_t^2}. \]

That looks a lot like the joint PDF. **It is not!**

For any likelihood in this form, the log likelihood minus any constants is easier to work with:

\[ l_L(\alpha; a_1, \ldots, a_m, a_{m+1}, \ldots, a_n) = - \sum_{t=m+1}^{n} \left( \log(\sigma_t^2) - \frac{a_t^2}{\sigma_t^2} \right). \]

This we can maximize (by minimizing the negative).
What about other distributions of $\{\epsilon_t\}$? Maybe ones with heavier tails.

One of the easiest families of distributions with heavier tails is the Student’s t. This family has one parameter, $\nu$, called the “degrees of freedom”.

As $\nu$ varies from 1 to $\infty$, the t family goes from the Cauchy to the standard normal.

For $\nu \leq 2$, the variance of the t distribution is infinite or does not exist. (Technically, those are different, but for purposes of modeling $\{\epsilon_t\}$, they are effectively the same.)

For $\nu > 2$, the variance of the t is $\nu/(\nu - 2)$, and the mean is 0. Since the model specifies that $\{\epsilon_t\}$ has constant mean of 0 and constant variance of 1, we can take

$$\epsilon_t = X_\nu / \sqrt{\nu/(\nu - 2)},$$

where $X_\nu$ is a random variable with a t distribution with $\nu$ DF.
Flexible Families of Probability Distributions for Modeling Applications

There are several families of probability distributions that have one or more parameters whose values lead to a variety of useful shapes of densities (or probabilities).

One that most statistics students are very familiar with is the beta family, which can have several shapes over any finite range, \([a, b]\).

A useful family over any semi-infinite range \([a, \infty[\) or \(]\infty, a]\) is the gamma family.

We have already mentioned that the Student’s t family provides a range of tail weights for symmetric distributions over \(]\infty, \infty[\).
Location-Scale Families of Probability Distributions

Given the distribution of any random variable $X$, we can define a location-scale family of distributions as the family of distributions of random variables of the form

$$Y = a + bX,$$

where $b \neq 0$.

Some familiar families of distributions are location-scale families, such as the normal and the logistic.

Other familiar families can serve as the basis of a larger family by a location-scale transformation.

That is essentially what we did with the Student’s t distribution; we made a scale transformation.
Transformations of Random Variables to Get More Interesting Probability Distributions

Linear transformations only affect the first two moments.

If we want to change the skewness (third moment) or tail weight (fourth moment), we much consider other transformations.

A power transformation can do that.

Consider a linear transformation in the kernel of a $N(0, 1)$ PDF $e^{-x^2}$ followed by a power transformation of the absolute value. We get the kernel

$$k(x) = e^{-|x-\gamma|/\beta|^{\alpha}},$$

where $\alpha, \beta > 0$.

The **exponential power family** is the set of distributions whose PDFs have a kernel of this form.

This family is also sometimes called the generalized error family.
Other Flexible Families of Probability Distributions for Modeling Applications

A specific member of the exponential power family of distributions is often used in finance, and is called the generalized error family of distributions (GED), as in Tsay on p. 122.

This family is symmetric about 0, and the other two parameters are reduced to a single one (although written as two). The PDF, as written by Tsay and by most economists, is

$$f(x) = \frac{v}{2^{1+\frac{1}{v}}\lambda\Gamma(1/v)} e^{-\frac{1}{2}|x/\lambda|^v},$$

where \(\lambda = \sqrt{2^{-2/v}\Gamma(1/v)/\Gamma(3/v)}\).

This distribution is much simpler than the Student’s t, which we mentioned earlier for modeling heavy tails.
Other Flexible Families of Probability Distributions for Modeling Applications

The other variation we need is to find families of distributions whose skewness can be controlled.

The familiar families of skewed distributions, such as log-normal and gamma, have support that is semi-finite.

We need families with support on $]\infty, \infty[$.

The “skew-Student-t” distribution of Fernández and Steel (which I admit, I don’t think I had heard of) is one such family.

It is described on p 122.

It has 2 parameters, one of which, $\nu$, is the t DF and so controls tail weight and the other, $\xi$, is for the proportion of the probability in the positive and negative regions, and so controls skewness.

(In the way the PDF is commonly written, and in Tsay’s notation, there are two other “parameters”, $\varpi$ and $\varrho$.)
Other Flexible Families of Probability Distributions for Modeling Applications

A general method of forming a skew distribution from a symmetric one is by skewing with a CDF.

The kernel of the probability density of a CDF-skewed distribution is

\[ k(x) = p(x)P(\lambda x), \]

where \( p(\cdot) \) is the density of the underlying symmetric distribution, and \( P(\cdot) \) is a CDF.

It is not necessary that the CDF be for the same distribution.

The idea also extends to multivariate distributions.
The Skew-Normal Distribution

The most common CDF-skewed distribution is the “skew-normal”.

The (standard) skew-normal distribution has density

\[ g(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \Phi(\lambda x) \quad \text{for} \quad -\infty \leq x \leq \infty, \]

where \( \Phi(\cdot) \) is the standard normal CDF, and \( \lambda \) is a constant such that \( -\infty < \lambda < \infty \).

For \( \lambda = 0 \), the skew-normal distribution is the normal distribution, and in general, if \( |\lambda| \) is relatively small, the distribution is close to the normal.

For larger \( |\lambda| \), the distribution is more skewed, either positively or negatively.
The purpose of the foregoing discussion was to introduce alternative probability models for the error distribution in financial models.

Probability models are needed for use of maximum likelihood in fitting models to data.

The key feature of these distributions is that they had parameters ("tuning parameters") that controlled features of interest, such as skewness or tail weight.

Nonstandard values of these features probably yield better fits of real data than the standard ones (skewness of 0, and kurtosis of 3).

What values of these tuning parameters should we choose?

A beauty of ML fitting is that we could include those parameters in the optimization procedure if we choose to ("let the data speak").
Understanding a Data-Generating Process

The main objective in statistical applications is to gain understanding of a data-generating process, whether it’s the rate of return of ordinary stocks, the rate structure of US bonds, patterns of crop rotation by farmers in Iowa, or the spread of an infectious disease.

All statistical methods rely on observational data to improve this understanding.

The better the understanding, the better the data-generating process can be simulated.

Since simulation models can be adjusted easily, simulation often provides a quicker way to understand a data-generating process.

We try a model, say a skew t, and if it matches the observational data, then we know a lot about the data-generating process.

(To paraphrase George Box, the model is not “right”; but it is useful.)
Performance of Statistical Methods

Statistical methods are developed and analytic properties of their performance are derived in the context of some set of assumptions about the underlying distribution.

The question is How do these methods perform when some of the assumptions are violated?

For instance, if the statistical method is developed under an assumed symmetric distribution, how does the method perform if the underlying distribution is asymmetric?

Flexible families of probability distributions are often useful in Monte Carlo studies of the performance of various statistical methods.
Order Determination in the ARCH Model

If an ARCH\((m)\) model is to be used, the first question is the order \(m\).

The model is fundamentally of the AR type, and we can use the PACF as we did with an AR model.

We use PACF on the squared residuals.
Fitting ARCH Models

As we have noted, ARCH models are usually fit by maximum likelihood.

This requires the specification of a distribution of course.

The normal is one choice, and often provides a good approximation.

The Student’s t family and the exponential power family provide a choice of tail weight.

The skewed Student’s t family and and CDF-skewed family provide a choices of skewness.
Fitting Time Series Models

Given a time series, we first look at it, and then we compute the ACF and the PACF. We might also do a Box-Ljung test.

```
plot(x)
acf(x)
pacf(x)
```

Consider the following example.
Time Series
Partial ACF
Fitting Time Series Models

The PACF suggests an AR effect of lag 2, and maybe lag 3. (Notice that while $|\phi_{3,3}|$ is large, it is considerably smaller than $|\phi_{2,2}|$.)

Also, we note that the ACF dies off at lag 1 or 2. We may model an MA effect of order 1 or 2.

So we do a preliminary ARMA fit and look at the residuals.

```r
fit <- arima(x, order=c(2,0,2))
res <- fit$residuals
acf(res)
pacf(res)
Box.test(res^2, lag=5,type="Ljung")
qqnorm(res)
```

We might also do a PACF on the residuals and their squares and do a Box-Ljung test on them.
Residuals ACF

Series res
Residuals PACF
Residuals Q-Q Plot

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles
Serial Correlation in Residuals

> Box.test(res^2, lag=5, type="Ljung")

Box-Ljung test

data: res^2
X-squared = 1.6065, df = 5, p-value = 0.9005
Fitting Time Series Models

The foregoing seemed to indicate that the time series in our example could be fit reasonably well with an ARMA model, possibly of order (2,2), as we used.

Suppose it doesn’t work out that way.

Consider a different time series.
Time Series

Time garch

2012-10-01 2013-04-01 2013-10-01
-0.015 -0.010 -0.005 0.000 0.005 0.010 0.015

Graph
Fitting Time Series Models

It looks a little different from the previous one, but let’s go ahead and do a preliminary ARMA fit and look at the residuals, as before.

```r
fitxx <- arima(xx, order=c(2,0,2))
resxx <- fitxx$residuals
Box.test(resxx^2, lag=5,type="Ljung")
```

This time we get

```
Box-Ljung test

data:  resxx^2
X-squared = 112.97, df = 5, p-value < 2.2e-16
```

An ARCH effect is present.
Let's fit a GARCH(1,1) to the residuals.

garchFit(data=resxx, formula=\sim garch(1,1), trace=FALSE)

We get

Coefficient(s):

\begin{tabular}{cccc}
mu & omega & alpha1 & beta1 \\
-1.0533e-05 & 5.1126e-07 & 1.5189e-01 & 8.0261e-01 \\
\end{tabular}
Fitting Time Series Models

Fitting a GARCH(1,1) to the residuals is the same as fitting a ARMA(2,2) + GARCH(1,1) model to the original data.

garchFit(data=xx, formula=~arma(2,2)+garch(1,1), trace=FALSE)

We get

Coefficient(s):

<table>
<thead>
<tr>
<th></th>
<th>mu</th>
<th>ar1</th>
<th>ar2</th>
<th>ma1</th>
<th>ma2</th>
<th>alpha1</th>
<th>beta1</th>
<th>alpha2</th>
<th>beta2</th>
<th>alpha3</th>
<th>beta3</th>
<th>alpha4</th>
<th>beta4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-5.8643e-05</td>
<td>3.4314e-01</td>
<td>-1.6114e-01</td>
<td>6.8433e-01</td>
<td>2.1825e-01</td>
<td>5.0707e-07</td>
<td>1.5263e-01</td>
<td>8.0223e-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As a general rule, the PACF of the residuals from an ARMA fit may be expected to suggest the order $m$ in a GARCH model.

It sometimes does, but it often does not.

The Box-Ljung test on the residuals from a GARCH fit often is highly significant.

```r
gfitxx <- garchFit(data=xx, formula=~arma(2,1)+garch(1,1), trace=FALSE)
gresxx <- residuals(gfitxx)
Box.test(gresxx^2, lag=5,type="Ljung")
```

yielded a p-value of essentially 0.

There are remaining serial correlations; the variance is changing!

The variances of the statistics in time series analyses are large!
The GARCH Model

Bollersev (1986) introduced a generalization of the ARCH\( (m) \) model, called the GARCH\( (m, s) \) model.

\[
A_t = \sigma_t \epsilon_t
\]

as before, and

\[
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i A_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2.
\]

This model has a good balance of complexity, and is probably the most widely-used of the simple discrete-time models for financial asset returns.
Variations on GARCH: The IGARCH Model

If the previous values are persistent, as in unit-root nonstationarity, we apply differences before fitting the GARCH model.

This leads to an IARCH($m, s, d$) model.

We usually take only $d = 1$. 
Variations on GARCH: EGARCH and TGARCH Models

In another variation, which will account for asymmetry in the residuals (that is, differences in the distribution of positive and negative returns), we consider an asymmetric function of the $\epsilon_t$:

$$g(\epsilon_t) = \theta \epsilon_t + \gamma (|\epsilon_t| - E(|\epsilon_t|)).$$

This ultimately leads a model for the log of $\sigma_t^2$, equation (3.25) in Tsay.

Because of the log, this is called an exponential GARCH model.

Another way of handling asymmetry is to use a “threshold” for negative residuals in the model for the returns.

This TGARCH model is given on page 149 of Tsay.
Variations on GARCH: The APARCH Model

Ding, Grange, and Engle (1993):

\[ A_t = \sigma_t \epsilon_t \]

as before, and

\[ \sigma_t^\delta = \alpha_0 + \sum_{i=1}^{m} \alpha_i (|A_{t-i} - \gamma_i A_{t-i}|^\delta + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^\delta. \]

This model includes several of the other variations on the GARCH model.
Fitting GARCH Models

GARCH models are usually fit by maximum likelihood.

This requires the specification of a distribution of course.

The normal is one choice, and often provides a good approximation.

The Student’s t family of distributions provides a choice of tail weight.
R Functions for Fitting GARCH Models

There are (at least) three packages in R that has GARCH modeling capabilities: \texttt{fGarch}, \texttt{tseries}, and \texttt{GARCH4.2}. I don't know anything about the last of these, but Tsay gives an example using it.

Probably the best one is \texttt{fGarch}, which was developed by Rmetrics, which is a software development project started in 2001 by Diethelm Wuertz at the Swiss Federal Institute of Technology (ETH) in Zurich.
R Functions for Fitting GARCH Models

Most of the R functions for fitting GARCH models allow the ARMA fitting to be done at the same time.

garchFit in fGarch fits GARCH models.

It is illustrated on pages 128, 129 of Tsay.

The key argument in garchFit is formula.

You can specify something like arma(2,1)+garch(1,1).
Simulation of Time Series

Data from ARIMA processes can be simulated using `arima.sim` in the `stats` package of R. (The result is stored in a `ts` object.)

The function `garchSim` function in the `fGarch` package simulates data from ARMA+GARCH processes. (The result is stored in a `timeSeries` object.)

garchSim also allows errors to follow the skewed-t and GED distributions described in Tsay.
Simulation of MA Data-Generating Processes

We need to understand some of the issues in generating data from time series models.

An MA($q$) process follows the simple model

$$X_t = \theta_0 - \sum_{i=1}^{q} \theta_i A_{t-i} + A_t,$$

where $\{A_t\}$ is white noise (i.e., iid, finite variance, and 0 mean).

This is straightforward to simulate; all we must do is simulate $\{A_t\}$ and take the appropriate weighted sum.
Simulation of AR Data-Generating Processes

An AR($p$) process follows the simple model

$$X_t = \phi_0 - \sum_{i=1}^{p} \phi_i X_{t-i} + A_t,$$

where $\{A_t\}$ is white noise.

This is not so straightforward to simulate; the problem is how to get started.

Recall that an AR($p$) process is effectively an MA($\infty$) process.

This is the problem.
Simulation of AR Data-Generating Processes

One way of simulating an AR series is to begin with a single variable from a white noise and then generate the AR series one term at a time.

For an AR(1) process for example, given the realization $a_1$, we form $a_2 + \phi_1 a_1$, then $a_3 + \phi_1 (a_2 + \phi_1 a_1)$, and so on, so that the $t^{th}$ term is

$$(1 + \phi_1 B + \cdots + \phi_1^{t-1} B^{t-1})a_t,$$

where $B$ is the backshift operator or lag operator.

Notice that if we let $X_t$ be the expression above, we have

$$X_t = \phi_1 X_{t-1} + a_t,$$

so it looks like an AR(1) process — but it is not!

Compare the variance of $X_t$ with that of $X_{t-1}$. 

110
Simulation of AR Data-Generating Processes

When do these terms become the same as terms in an AR process?

Recall that for an AR(1) to be invertible (non-explosive), we must have $|\phi_1| < 1$. In that case, depending on $t$, the terms farther back in the series have less effect.

Notice that if $t$ is “large enough” and $|\phi_1|$ is “small enough”, the variance of $X_t$ is approximately the same as that of $X_{t-1}$, and we have “almost” an AR(1) process.

The point of this is that we must have a “burn-in” period. We discard the first several realizations.

This is a characteristic of many time series, for example that in an MCMC application.
Simulation of AR(2) Data-Generating Processes

Just to fix the points, let’s consider simulation of an AR(2) process defined by the recursion

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + A_t, \]

using the same approach as before.

We have \( a_1, \ a_2 + \phi_1 a_1, \ a_3 + \phi_1(a_2 + \phi_1 a_1) + \phi_2 a_1, \ a_4 + \phi_1(a_3 + \phi_1(a_2 + \phi_1 a_1) + \phi_2 a_1) + \phi_2(a_2 + \phi_1 a_1), \ldots \)

We can write an expression for the general \( t^{\text{th}} \) term, but it is very complicated.
Following Dunne (1992), we begin with the reciprocals $b_1$ and $b_2$ of the roots of the characteristic equation

$$\phi_2 z^2 + \phi_1 z + 1 = 0,$$

(We have let $1/b_1$ and $1/b_2$ be the two roots.) so these must satisfy the relations

$$b_1 + b_2 = \phi_1 \quad \text{and} \quad b_1 b_2 = -\phi_2.$$

Recall that invertibility requires that the roots lie outside the unit circle, so in that case we must have $|b_1| < 1$ and $|b_2| < 1$; that is, we are going to write an expression whose terms will become progressively smaller.
Simulation of AR(2) Data-Generating Processes

Now we can write the general $t^{th}$ term as

$$X_t = \frac{-b_1}{b_2 - b_1} (1 + b_1 B + \cdots + b_1^{t-1} B^{t-1}) a_t + \frac{b_2}{b_2 - b_1} (1 + b_2 B + \cdots + b_2^{t-1} B^{t-1}) a_t,$$

and we note that

$$X_t = (b_1 + b_2) X_{t-1} - b_1 b_2 X_{t-2} + a_t,$$

or

$$X_t = \phi_1 X_{t-1} \phi_2 X_{t-2} + a_t,$$

which is in the correct form for an AR(2) process.

As before, however, $\{X_t\}$ is not second-order stationary.

Again, as before, if $|b_1|$ and $|b_2|$ are small enough and $t$ is large enough, the process is approximately an AR process.

We use a burn-in period.