The AR(2) model. Assume $\mathrm{E}(y_t) = 0$, and ignore constant

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t$$

where ϵ_t iid $(0,\sigma_\epsilon^2)$. We can, but we won't, compute the variance and autocovariances and autocorrelations by hand. But we will look at the forecasting formula

$$\begin{aligned}
E\left(y_{t+1}|I_t\right) &= E\left(\rho_1 y_t + \rho_2 y_{t-1} + \epsilon_{t+1}|I_t\right) \\
&= \rho_1 y_t + \rho_2 y_{t-1} \\
E\left(y_{t+2}|I_t\right) &= E\left(\rho_1 y_{t+1} + \rho_2 y_t + \epsilon_{t+2}|I_t\right) \\
&= \rho_2 E\left(y_{t+1}|I_t\right) + \rho_2 y_t \\
&= \rho_2\left(\rho_1 y_t + \rho_2 y_{t-1}\right) + \rho_2 y_t \\
&= \rho_2(1+\rho_1) y_t + \rho_2 y_{t-1}
\end{aligned}$$

So we build up forecasts at longer horizons recursively.

The AR(2) will be stationary if $|\rho_1 + \rho_2| < 1$.

We going to look at an AR(2) with $\rho_1 = 0.8, \rho_2 = -0.8$.The IR oscillates.

Let's look at $\rho_1 = 0.8, \rho_2 = 0.1$

How do we know which model to estimate? ARMA(p,q). How do we choose p and q? Remember a couple lectures ago, we met the log likelihood function? And it looked a little like

$$\frac{LL}{T} = -\ln\left(\hat{\sigma}_\epsilon^2\right)^{\frac{1}{2}}$$

(I think it was labeled as eq.(12)), where $\hat{\sigma}_\epsilon^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t^2\left(y_t, y_{t-1}, ..\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2}, ...\right)$where $\hat{\epsilon}_t^2\left(y_t, y_{t-1}, ..\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2}, ...\right)$ means the form is model dependent. ML estimation asks computer to choose parameters to maximize the LL. We cannot just compute the log likelihood for each ARMA(p,q) and take the one with the highest LL. Solution is to penalize yourself for adding parameters. Akaike was the first to do this. It is called the AIC. IC stands for information criterion (basically the LL), the A he says is the first alphabet. But we really know it stands for Akaike.

$$AIC = \ln\left(\hat{\sigma}_\epsilon^2\right) + \frac{2k}{T}$$

where k is the number of parameters (the p and q in ARMA). Search over the various models and choose the one that minimizes AIC.

A bunch of people studied the properties of AIC and concluded that it often overparameterizes the model. So Swartz came up with a different penalty.

$$BIC = \ln\left(\hat{\sigma}_\epsilon^2\right) + \frac{k}{T}\ln\left(T\right)$$

But sometimes BIC leads to an underparameterized model. So Hannan and Quinn came up with

$$HQIC = \ln\left(\hat{\sigma}_\epsilon^2\right) + \frac{2k}{T}\ln\left(\ln\left(T\right)\right)$$

which usually splits the difference between AIC and BIC.

A couple of words about the driftless random walk. This is when $\rho = 1$ in the AR(1) model

$$y_t = y_{t-1} + \epsilon_t$$

with ϵ_t iid $(0,\sigma_\epsilon^2)$.

$$E\left(y_{t+1}|I_t\right) = E\left(y_t + \epsilon_{t+1}|I_t\right) = y_t$$

Think of y_t as the (with dividend) log price of a security or portfolio.

$$r_t = y_t - y_{t-1}$$

Stock returns are hard to predict, so as an approximation the log price may follow a random walk.

$$\begin{aligned} E\left(y_{t+2}|I_t\right) &= E\left(y_{t+1} + \epsilon_{t+2}|I_t\right) \\ &= E\left(y_t + \epsilon_{t+1}|I_t\right) \\ &= y_t \end{aligned}$$

What is

$$\begin{aligned} E\left(r_{t+1}|I_t\right) &= E\left(y_{t+1} - y_t|I_t\right) \\ &= 0 \end{aligned}$$

You might want to see if your ARMA(p,q) model forecasts better than the random walk. Another feature of the random walk is that shocks are permanent. To see this, get the MA representation

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t \\ &= y_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &= y_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\ \\ &= y_0 + \epsilon_1 + \epsilon_2 + \cdots \epsilon_t \end{aligned}$$

Random walk with drift.

$$\begin{aligned} y_t &= d + y_{t-1} + \epsilon_t \\ E\left(y_t - y_{t-1}\right) &= d \end{aligned}$$

Movements in returns unpredictable although over long periods of time the return is d.Backward substitution

$$\begin{aligned}
y_t &= d + y_{t-1} + \epsilon_t \\
&= d + d + y_{t-2} + \epsilon_{t-1} + \epsilon_t \\
&= 2d + d + y_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\
&= 3d + y_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\
\\
&= td + y_0 + \underbrace{(\epsilon_1 + \epsilon_2 + \cdots \epsilon_t)}_{\text{driftless random walk}}
\end{aligned}$$

If the truth is a random walk with drift, you might be tempted to "detrend" it by regressig y_t on a constant and trend. But don't! Because the error term is a driftless random walk. You won't get anything sensible. You have to work in first differences of y_t.