ARMA Properties and Forecasting

This Lecture

- Properties of stationary and nonstationary ARMA processes
- Lag polynomials
- Stochastic cycles/business cycles
- Forecasting behavior and forecast error variance using ARMA

Ignore this warning message

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0.  You will still be able to use
## 'loadSymbols' to automatically load data.  getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.

## [1] "DCOILWTICO"
## [1] "GOLDPMGBD228NLBM"
```

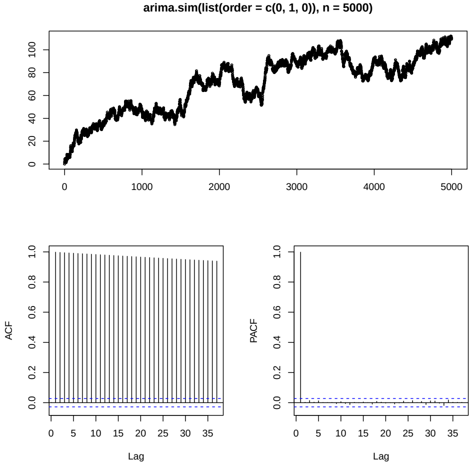
ACF/PACF Behavior

- Recall that ACFs and PACFs either cut off or exponentially decay/oscillate for stationary series
- For unit root nonstationary series, ACF stays close to 1 and PACF first lag or lags close to 1 while others cut off.
- Why?

ACF/PACF Behavior

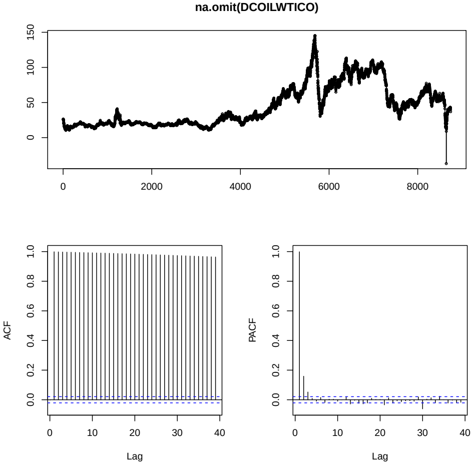
Simulated

$$y_t = y_{t-1} + a_t$$



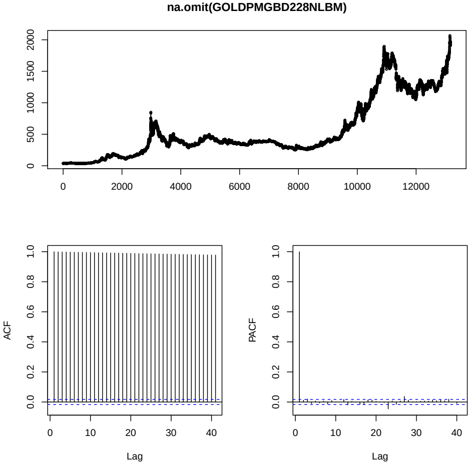
ACF/PACF Behavior

Oil price (not returns)



ACF/PACF Behavior

Gold price (not returns)



AR(1)

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \;\; a_t \sim white\ noise(0, \sigma_a^2)$$

$$E(r_t | r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \;\; Var(r_t | r_{t-1}) = \sigma_a^2$$

- weak stationarity requires
  $E(r_t) = \mu, \; Var(r_t) = \gamma_0, \; Cov(r_t, r_{t-s}) = \gamma_s$ are all constants not dependent on t.

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}) \Longrightarrow \mu = \frac{\phi_0}{1-\phi_1}, \; or \; \phi_0 = (1-\phi_1)\mu$$

- This implies $\phi_1 \neq 1$, AND $\mu = 0$ if and only if $\phi_0 = 0$

AR(1)

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \quad a_t \sim \textit{white noise}(0, \sigma_a^2)$$

$$\begin{aligned}
Var(r_t) &= \tfrac{\sigma_a^2}{1-\phi_1^2} \\
&= E\left[(r_t - \mu)^2\right] \\
&= E\left[(\phi_1(r_{t-1} - \mu) + a_t)^2\right] \\
&= E\left[\phi_1^2(r_{t-1} - \mu)^2 + a_t^2\right] \\
&\text{(independent } a_t \text{ means cross products are zero)} \\
&= \phi_1^2 E\left[(r_t - \mu)^2\right] + \sigma_a^2
\end{aligned}$$

- We need $\phi_1^2 < 1$ for finite, non-negative variance. Stationarity requires $|\phi_1| < 1$.
- Similarly, $\gamma_s = E\left[(r_{t-s} - \mu)(r_t - \mu)\right] = \phi_1 \gamma_{s-1}$
- Dividing by γ_0 we have $\rho_s = \phi_1 \rho_{s-1} = \phi_1^s$ because $\rho_0 = 1$

$AR(1)$ Permanent Shocks

Recall we can write

$$\begin{aligned} r_t - \mu &= \phi_1(r_{t-1} - \mu) + a_t \\ &= \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \end{aligned}$$

If $\phi_1 = 1$ all past shocks are permanent.

AR(p)

$$
\begin{aligned}
r_t &= \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t \\
E(r_t) &= \mu = \frac{\phi_0}{1-\phi_1-\cdots-\phi_p} \\
r_t - \mu &= \phi_1(r_{t-1} - \mu) + \cdots + \phi_p(r_{t-p} - \mu) + a_t
\end{aligned}
$$

AR(p) requires us to think more about stationarity condition and ACF behavior

$$
\begin{aligned}
\gamma_s &= \phi_1\gamma_{s-1} + \cdots + \phi_p\gamma_{s-p} \\
\implies \rho_s &= \phi_1\rho_{s-1} + \cdots + \phi_p\rho_{s-p} \\
(1 - \phi_1 L - \cdots - \phi_p L^p)\rho_s &= 0
\end{aligned}
$$

Lag polynomial or backshift polynomial. Roots of this polynomial determine stationarity and patterns of decay in autocorrelation.

MA(1)

$$r_t = \mu + a_t + \theta a_{t-1}$$

$$E(r_t) = \mu$$

$$E(r_t - \mu)^2 = (1 + \theta^2)\sigma_a^2$$

First autocovariance/autocorrelation:

$$\gamma_1 = E(r_t - \mu)(r_{t-1} - \mu) = E(a_t + \theta a_{t-1})(a_{t-1} + \theta a_{t-2}) = \theta\sigma_a^2$$

$$\rho_1 = \gamma_1/\gamma_0 = \frac{\theta}{1 + \theta^2}$$

Notice that $|\rho_1| < \frac{1}{2}$. If you found otherwise, series could not be MA(1).
Further autocovariances/autocorrelations:

$$E(r_t - \mu)(r_{t-j} - \mu) = E(a_t + \theta a_{t-1})(a_{t-j} + \theta a_{t-j-1}) = 0 \; \forall \; j = 2, 3, \ldots$$

MA(q)

- $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}$
- $E(y_t) = \mu$
- $E(y_t - \mu)^2 = (1 + \theta_1^2 + ... + \theta_q^2)\sigma^2$
- $E(y_t - \mu)(y_{t-j} - \mu) =$
  $E(\epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q})(\epsilon_{t-j} + \theta_1 \epsilon_{t-j-1} + ... + \theta_q \epsilon_{t-j-q})$
  - $= (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+1}\theta_2 + ... + \theta_q \theta_{q-j})\sigma^2$ for $j = 1, 2, ..., q$
  - $= 0$ for $j > q$
- An MA(q) process is covariance stationary for **any** values of $\theta_1 ... \theta_q$

ARMA(1,1)

$$r_t = \phi_1 r_{t-1} + a_t + \theta_1 a_{t-1}$$

$$\begin{aligned}
\mu = E(r_t) &= \tfrac{\phi_0}{1-\phi_1} \\
Var(r_t) &= E(\phi_1 r_{t-1} + a_t + \theta_1 a_{t-1})^2 \\
&= \phi_1^2 Var(r_{t-1}) + \sigma_a^2 + \theta_1^2 \sigma_a^2 + 2\phi_1\theta_1 E(r_{t-1} a_{t-1}) \\
&= \tfrac{(1+2\phi_1\theta_1+\theta_1^2)\sigma_a^2}{1-\phi_1^2}
\end{aligned}$$

- Just as in AR case, need $\phi_1^2 < 1$.
- For MA(1), $Var(r_t) = (1+\theta_1^2)\sigma_a^2$
- For AR(1), $Var(r_t) = \frac{\sigma_a^2}{1-\phi_1^2}$
- Behavior is a mix of AR and MA parts

Lag polynomials

We can rewrite

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-1} + a_t + \theta_1 a_{t-1} + ... + \theta_q a_{t-q}$$

as

$$(1 - \phi_1 L - ... - \phi_p L^p) r_t = \phi_0 + (1 + \theta_1 L + ... + \theta_q L^q) a_t$$

$$\phi(L) r_t = \phi_0 + \theta(L) a_t$$

- Recall an MA(q) is stationary for any finite q and finite θ's.
- Intuition: just a weighted average of white noise.
- Properties of AR polynomial $\phi(L)$ matter a lot.

AR(2) Lag polynomial

$$(1-\phi_1 L-\phi_2 L^2)r_t = a_t$$

Consider

$$(1-\phi_1 x-\phi_2 x^2) = 0$$

- Roots: $x=\frac{\phi_1 \pm \sqrt{\phi_1^2+4\phi_2}}{-2\phi_2}$
- inverse of the solutions are called characteristic roots ω_1, ω_2.

AR(2) Lag polynomial

- If both roots are real, can factor $(1-\omega_1 L)(1-\omega_2 L)$ - double exponential decay/oscillation in ACF.
- If not real, ω_1, ω_2 are complex conjugate pairs.
  - Time series has a cyclical component (e.g., business cycle) with random period
  - dependence/ACF will decay sinusoidally.
- Average period/stochastic cycle $k = \frac{2\pi}{cos^{-1}\left[\phi_1/\left(2\sqrt{-\phi_2}\right)\right]}$
- For stationarity, roots of lag polynomial should be greater than one, characteristic roots should be less than one in modulus/absolute value $(||\omega|| < 1)$

Companion Form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t$$

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}$$

$$\mathbf{y_t} = \mathbf{F}\mathbf{y_{t-1}} + \mathbf{w_t}$$

- The characteristic roots of the lag polynomial ω_1, ω_2 are the eigenvalues of the **F** matrix.
- All $||\omega|| < 1$ is a stationary system. If any $||\omega|| = 1$, system has a unit root. If any $||\omega|| > 1$ system is explosive.
- Consider behavior of $\frac{\partial y_{t+s}}{\partial w_t}$ as s grows. "Impulse Response Function".

AR(2) Lag polynomial

- More generally for AR(p): $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) r_t = \mu + a_t$
  - lag polynomial has p roots. Complex will be of form $a \pm bi$, with
    - $k = \frac{2\pi}{cos^{-1}\left[a/\left(\sqrt{a^2+b^2}\right)\right]}$
    - $\sqrt{a^2 + b^2}$ is the absolute value, or modulus of $a \pm bi$
  - there may be multiple stochastic cycles with different periodicity layered on top of each other.

Forecasting from AR(p)

- Suppose we are at time h (forecast origin) and want to forecast l periods ahead (forecast horizon)
- $\hat{r}_h(l)$ is the forecast of r_{h+l} that minimizes squared error loss function

$$min_{\hat{r}_h(l)} E\left[\left(r_{h+l} - \hat{r}_h(l)\right)^2 | F_h\right]$$

- forecast minimizes expected squared distance from true value, conditional on all we know at time h

1 Step Ahead

- From AR(p),
  $\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \phi_1 r_h + \phi_2 r_{h-1} + ... + \phi_p r_{h-p+1}$
  - Forecast error: $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$
  - $Var(e_h(1)) = \sigma_a^2$
  - So the variance of the forecast error is just the variance of the residual.
  - 95% confidence intervals for 1-step ahead: $\hat{r}_h(1) \pm 1.96 \cdot \sigma_a$

2 Step Ahead

- $r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \phi_2 r_h + ... + \phi_p r_{h-p+2} + a_{h+2}$
- $\hat{r}_h(2) = E(r_{h+2}|F_h) = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + ... + \phi_p r_{h-p+2}$
- Forecast error:
  $e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1(r_{h+1} - \hat{r}_h(1)) + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}$
- $Var(e_h(2)) = (1 + \phi_1^2)\sigma_a^2 > \sigma_a^2$

Multistep ahead

- $\hat{r}_h(l) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(l-i)$
- $\hat{r}_h(l) \rightarrow E(r_t)$ *as* $l \rightarrow \infty$ "mean reversion"
- $Var(e_h(l)) \rightarrow Var(r_t)$

Forecasting from MA(q)

Forecasts return to the mean more quickly than AR models.
One-step ahead for an MA(1) at forecast origin h,

$$r_{h+1} = \mu + a_{h+1} + \theta_1 a_h$$

$$E(r_{h+1}|F_h) = \hat{r}_h(1) = \mu + \theta_1 a_h$$
$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$$
$$Var(e_h(1)) = Var(a_{h+1}) = \sigma_a^2$$

This is the variance of each independent shock.

Two-step ahead

$$r_{h+2} = \mu + a_{h+2} + \theta_1 a_{h+1}$$

$$E(r_{h+2}|F_h) = \hat{r}_h(2) = \mu$$

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} + \theta_1 a_{h+1}$$

$$Var(e_h(2)) = Var(a_{h+2} + \theta_1 a_{h+1}) = (1 + \theta_1^2)\sigma_a^2$$

This is the variance of r_t when MA(1)
For an MA(q), forecast reverts to the mean and forecast error variance reverts to the unconditional variance after q steps.

Forecasting from ARMA

MA component disappears after q periods, then behaves like an AR forecast. Notice an ARMA can be represented in three ways:

1. ARMA(p, q):
$$\phi(B)r_t = \phi_0 + \theta(B)a_t$$
2. MA(∞):
$$r_t = \frac{\phi_0}{\phi(B)} + \frac{\theta(B)}{\phi(B)}a_t = \mu + \psi(B)a_t = \mu + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots$$
3. AR(∞):
$$\frac{\phi(B)}{\theta(B)}r_t = \pi(B)r_t = \frac{\phi_0}{\theta(B)} + a_t = \frac{\phi_0}{1 + \theta_1 + \ldots + \theta_q} - \pi_1 r_{t-1} - \pi_2 r_{t-2} - \ldots + a_t$$

Forecasting from ARMA

One-step ahead for ARMA(p, q) at forecast origin h,

$$E(r_{h+1}|F_h) = \hat{r}_h(1) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} + \sum_{i=1}^{q} \theta_i a_{h+1-i}$$

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$$

$$Var(e_h(1)) = Var(a_{h+1}) = \sigma_a^2$$

This is the variance of each independent shock.

Forecasting from ARMA

l-step ahead:

$$E(r_{h+l}|F_h) = \hat{r}_h(l) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(l-i) + \sum_{i=1}^{q} \theta_i a_h(l-i)$$

where $\hat{r}_h(l-i) = r_{h+l-i}$ if $l - i \leq 0$ (if the lags of the forecast horizon are in the sample).
Similarly, $a_h(l-i) = 0$ if $l - i > 0$ and $a_h(l-i) = a_{h+l-i}$ otherwise.

Forecasting from ARMA

How to describe the forecast error:

$$e_h(l) = r_{h+l} - \hat{r}_h(l)$$

Using the MA(∞) representation, we have

$$\hat{r}_h(l) = \mu + \psi_l a_h + \psi_{l+1} a_{h-1} + \ldots$$

because we have not yet observed the shocks a_h, $a_{h+1}, \ldots,$ a_{h+l}. (Notice also that as l gets large and ψ_l gets small, the forecast approaches the mean.) The forecast error accumulates those as yet unobserved shocks, so

$$e_h(l) = a_{h+l} + \psi_1 a_{h+l-1} + \ldots + \psi_{l-1} a_{h+1}$$

$$Var(e_h(l)) = (1 + \psi_1^2 + \ldots + \psi_{l-1}^2)\sigma_a^2$$

Finally, from the MA(∞) representation, we have that $Var(r_t) = (1 + \psi_1^2 + \psi_2^2 + \ldots)\sigma_a^2$. We can see that $Var(e_h(l))$ converges to this as l gets large.