Statistics 5350/7110
Forecasting

Lecture 11
ARMA Models

Professor Robert Stine

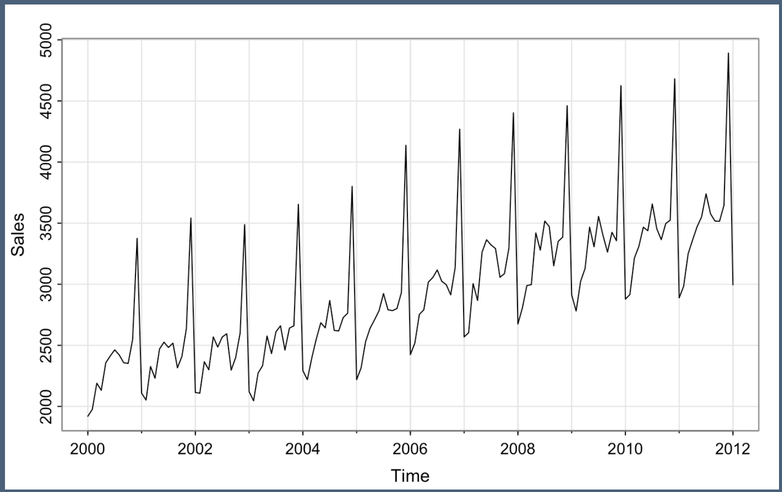
Admin Issues

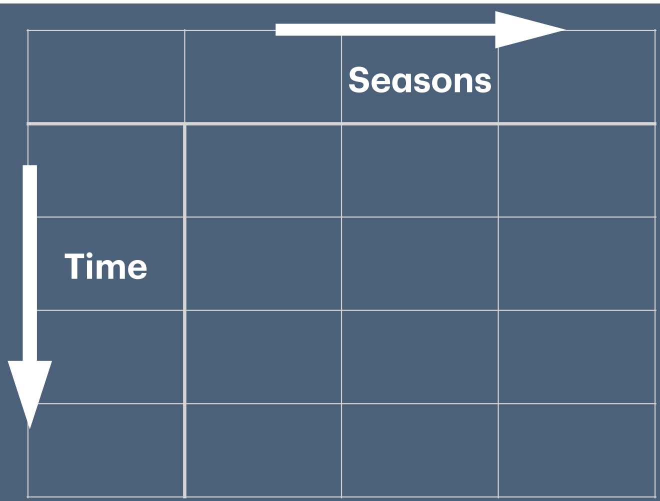
- Questions, comments
  - Fall Break, so no class on Thursday this week.
- Assignments
  - Next one handed out next week.
  - There will be questions regarding STL.
- Quick review
  - Decomposition of time series
  - Seasonally adjusted data
  - Methods

    Median polish

    Generalized additive model (a little)

    STL (iterative loess fits for trend and seasonal)





Today's Topics

Text, §4.1-4.2

- Have seen simplest cases
  - First-order autoregression

$$X_t = \phi_1 X_{t-1} + w_t, \quad |\phi_1| < 1$$

  - Moving average, centered in time

$$X_t = \frac{w_{t+1} + w_t + w_{t-1}}{3}$$

- Generalize
  - The notation varies from source to source (particularly choice of signs)

Definition 4.8

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

  - Covariances are key: If Gaussian, mean plus autocovariances are sufficient statistics
- Issues
  - Stationarity, identifiability, constraints (causality and invertibility)
  - Backshift notation and polynomials
  - Covariance and correlation functions

- Properties of the AR(1) model
  - Restriction on the coefficient needed for stationarity (assume mean is zero, or $X_t = (x_t - \mu)$)

$$X_t = \phi_1 X_{t-1} + w_t, \quad |\phi_1| < 1$$

  - Backshift polynomial form

$$(1 - \phi_1 B)\, X_t = w_t$$

  - Back-substitution or polynomial trick yields the infinite moving average representation

Definition 4.4

$$X_t = w_t + \phi_1\, w_{t-1} + \phi_1^2\, w_{t-1} + \cdots + \phi_1^s\, w_{t-s} + \cdots = \sum_{s=0}^{\infty} \phi_1^s\, w_{t-s}$$

Causal

- Variance
  - Derive from assumption of stationarity or from moving average representation
  - Assuming stationarity, take variance of expression for X_t

$$\text{Var}(X_t) = \phi_1^2 \text{Var}(X_{t-1}) + \sigma_w^2 = \frac{\sigma_w^2}{1 - \phi_1^2}$$

Autocorrelations of AR(1)

- Autocovariances
  - Assuming stationarity, multiply through equation by lags and take expectations
  - We'll find $\gamma(h)$ for h = 0,1,2,..., and find the others by symmetry $\gamma(h) = \gamma(-h)$
  - Equation from multiplying by X_t is a little different

$$\gamma_x(0) = \phi_1 \gamma_x(1) + \sigma_w^2$$

  - Multiplying by lags X_{t-1} and so forth are simpler since uncorrelated with w_t

$$\gamma_x(1) = \phi_1 \gamma_x(0)$$
$$\gamma_x(j) = \phi_1 \gamma_x(j-1)\ , \quad j = 2,3,\ldots$$

- Autocorrelations
  - Since we know γ(0), we have

$$\gamma(j) = \phi_1^j \gamma(0)$$

  - Hence autocorrelations decay geometrically

$$\rho(j) = \phi_1^j$$

Text derives from the moving average expression

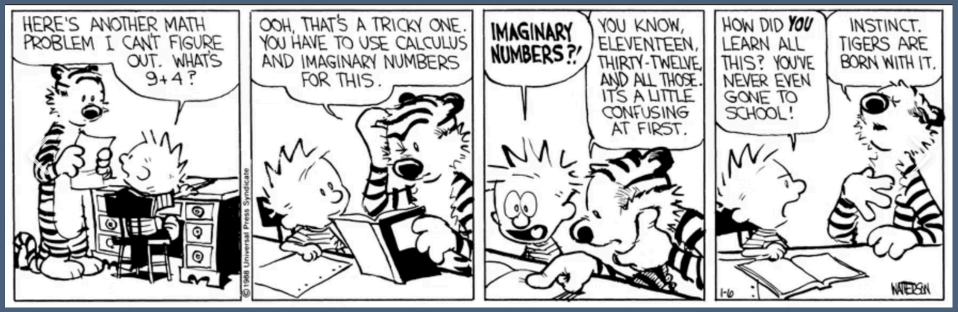
Background

- Before considering general AR(p) case…
- Fundamental Theorem of Algebra
  - Every polynomial equation of degree k with complex number coefficients has k roots $r_1,\ldots,r_k$ in the complex numbers (roots a.k.a. zeros or solutions).
  - Hence every polynomial factors

$$A(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_k z^k = c(z - r_1)(z - r_2)\cdots(z - r_k) = 0$$

  - Proved by Gauss and others, ~ 1800

Calvin & Hobbs



- Special case
  - Coefficients are real numbers (no imaginary component)
  - Roots must either be real numbers or complex conjugate pairs
  - Complex conjugates: complex numbers of the form z = x ± i y so that $(x + i\,y)(x - iy) = x^2 + y^2$ is a real number.

Appendix C
Text reviews basic facts about complex numbers

Autoregression, General Case

- Properties of the AR(p) model
  - Restrictions on the coefficients that assure stationarity are "complex"
$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t, \quad |\phi_1| < 1$$
  - Backshift polynomial form
$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\, X_t = \phi(B) X_t = w_t$$
  - Stationarity requires zeros of the polynomial $\phi(z)$ to lie outside unit circle
$$\phi(r_j) = 0 \iff |r_j| > 1$$
  - Back-substitution or polynomial trick yields the causal moving average representation
$$X_t = w_t + \psi_1\, w_{t-1} + \psi_2\, w_{t-2} + \cdots + \psi_s\, w_{t-s} + \cdots = \sum_{s=0}^{\infty} \psi_s\, w_{t-s}$$
- AR(1) case
  - Zero of polynomial easy to find: it's the reciprocal of the coefficient
$$\phi(z) = 1 - \phi_1 z = 0 \iff r_1 = 1/\phi_1$$

$E(X_t) = 0$ throughout

Regression model puts no such constraint on the coefficients

Example in RMD

Autocovariance Function

- Yule-Walker equations
  - Multiply expression for X_t by lags X_{t-1}, X_{t-2}, ... X_{t-p}, then take expectations (assume mean is zero)
  - Obtain a system of equations to solve for autocovariances

$$\begin{array}{rcl} \gamma(1) & = & \phi_1\gamma(0) + \phi_2\gamma(1) + \cdots + \phi_p\gamma(p-1) \\ \gamma(2) & = & \phi_1\gamma(1) + \phi_2\gamma(2) + \cdots + \phi_p\gamma(p-2) \\ \vdots & = & \vdots \\ \gamma(p) & = & \phi_1\gamma(p-1) + \phi_2\gamma(2) + \cdots + \phi_p\gamma(0) \end{array}$$

$$\gamma = \Gamma\,\phi$$

- Solving for autocovariances
  - Further autocovariances follow recursion

$$\gamma(k) = \phi_1\,\gamma(k-1) + \phi_2\,\gamma(k-2) + \phi_p\,\gamma(k-p), \quad p \le k$$

  - Add equation for γ(0) to get p+1 equations in p+1 unknowns

$$\gamma(0) = \phi_1\gamma(1) + \cdots + \phi_p\gamma(p) + \sigma_w^2$$

  - AR(2) special case

$$\gamma(0) = \frac{\sigma^2(1-\phi_2)}{((\phi_2-1)^2-\phi_1^2)(\phi_2+1)}$$

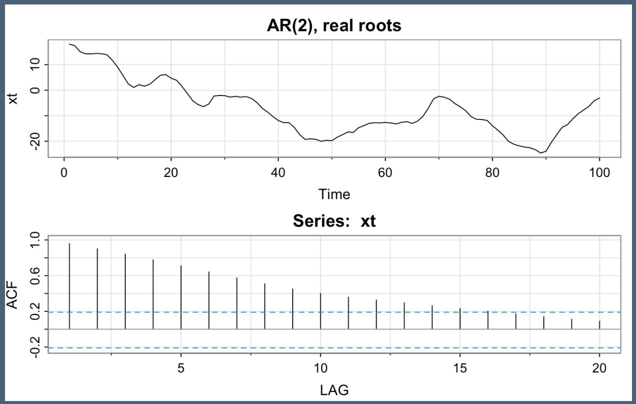
```
Solve[
 {γ0 == ϕ1 γ1 + ϕ2 γ2 + σ^2 &&
   γ1 == ϕ1 γ0 + ϕ2 γ1 &&
   γ2 == ϕ1 γ1 + ϕ2 γ0}, {γ0, γ1, γ2}]
```

Mathematica

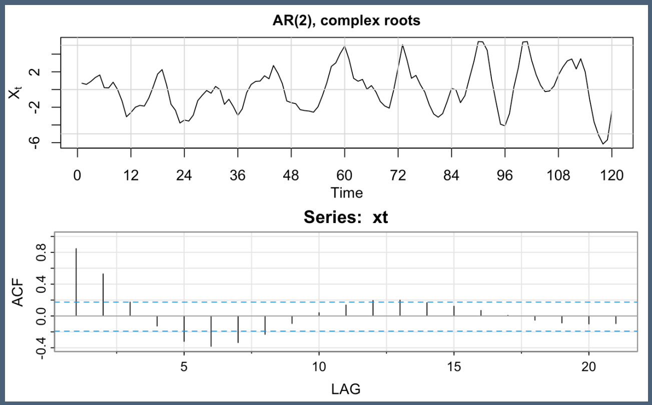
$$\left\{\left\{\gamma_0 \to -\frac{\sigma^2\,(-1+\phi_2)}{(1+\phi_2)\,(1-\phi_1^2-2\,\phi_2+\phi_2^2)},\ \gamma_1 \to \frac{\sigma^2\,\phi_1}{(1+\phi_2)\,(1-\phi_1^2-2\,\phi_2+\phi_2^2)},\ \gamma_2 \to \frac{\sigma^2\,(\phi_1^2+\phi_2-\phi_2^2)}{(1+\phi_2)\,(1-\phi_1^2-2\,\phi_2+\phi_2^2)}\right\}\right\}$$

Examples of AR(2) Covariances

- $\phi(B)$ has real roots
  - $X_t = 1.7\, X_{t-1} - 0.72\, X_{t-2} + w_t$
  - MA weights ψ_j decay slowly



- $\phi(B)$ has complex pair of roots
  - Textbook (p 70)
  - $X_t = 1.5\, X_{t-1} - 0.75\, X_{t-2} + w_t$
  - MA weights oscillate
  - "Quasi-periodic" with period ≈ 12



Moving Average Processes

- MA(1) process

$$X_t = w_t + \theta_1 w_{t-1}$$

$E(X_t) = 0$ throughout

- Always stationary with covariances (symmetric around 0)

$$\gamma(0) = \sigma_w^2(1 + \theta_1^2)$$
$$\gamma(1) = \sigma_w^2\,\theta_1$$
$$\gamma(h) = 0, \quad h = 2,3,\ldots$$

- Hence autocorrelation is

$$\rho(1) = \frac{\theta_1}{1 + \theta_1^2}, \qquad \rho(h) = 0,\ h = 2,3,\ldots$$

  - Interesting property: $|\rho(1)| \leq 0.5$
  - In comparison to AR models, ACF "cuts off" after 1 term

Identifiability

- Model as described cannot be identified from autocovariances
  - Processes with different coefficients and variances have the same autocovariances
  - Ambiguity between white noise variance and MA coefficient
  - Example

$$\sigma_w^2 = 25 \quad \theta_1 = 1/5 \qquad \text{and} \qquad \sigma_w^2 = 1 \quad \theta_1 = 5$$

  In both cases

$$\gamma(0) = 25(1 + 1/25) = 1(1 + 25)$$
$$\gamma(1) = 25(1/5) = 1(5)$$

- Invertibility condition
  - Require that any moving average process be expressible as a stationary, causal autoregression
  - Implies that the polynomial $\theta(z) = 1 + \theta_1 z$ has zeros outside the unit circle
  - In the case of the prior MA(1), we choose the representation with the larger noise variance.

Moving Average Processes

- General form, MA(q)

$$X_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} = \theta(B)\, w_t$$

More like a moving sum rather than a moving average

where invertibility implies that zeros $r_1, r_2, \ldots, r_q$ of $\theta(z)$ lie outside the unit circle

$$\theta(r_k) = 0 \iff |r_k| > 1$$

- Autocovariances are more easily found than for AR(p) model
  - Variance is

$$\text{Var}(X_t) = \sigma_w^2(1 + \theta_1^2 + \cdots + \theta_q^2)$$

  - First autocovariances

$$\gamma(1) = \sigma_w^2\,(\theta_1 + \theta_1\theta_2 + \cdots + \theta_{q-1}\theta_q)$$
$$\gamma(2) = \sigma_w^2\,(\theta_2 + \theta_1\theta_3 + \cdots + \theta_{q-2}\theta_q)$$

  - Zero after q lags

$$\gamma(h) = 0, \quad h = q+1, q+2, \ldots$$

Autoregressive Moving Average

- Combine AR(p) with MA(q), ARMA(p,q)

α is not the mean!

$$X_t = \alpha + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

where w_t is a white-noise process

- Compact expression

$$\phi(B)\, X_t = \theta(B)\, w_t$$

- Conditions on the coefficients
  - For stationarity, the zeros of ϕ(B) lie outside the unit circle (causal)
  - For identifiability, the zeros of θ(B) lie outside the unit circle (invertible)
  - For identifiability, ϕ(B) and θ(B) have no common zeros (see examples 4.9-4.11)
- Add non-zero mean
  - Taking expectations of the definition shows that

$$E(X_t) = \mu = \alpha + \phi_1 \mu + \cdots + \phi_p \mu \qquad \text{or} \qquad \alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$$

Why these models?

- Substantive motivation
  - Hard to come by outside special physical systems
- An answer
  - They work!
  - Analogous to use of polynomials in regression
  - ARMA models can predict anything (within reason, especially once paired with differencing!)
- Deeper answer
  - Parsimonious way to achieve promise of the Wold Representation Theorem
    Every second-order stationary time series $\{X_t\}$ has a causal moving average representation

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \, w_{t-j}$$

  - The few AR and MA coefficients determine the infinite series $\{\psi_j\}$

ARMA(1,1) Autocovariances

Example 4.16, p 78

- Illustrates the general properties of ARMA processes
  - Process is (zero mean)

$$X_t = \phi_1 X_{t-1} + w_t + \theta_1 w_{t-1}$$

  - Conditions require that $\phi \neq -\theta_1$ else would just be complicated way to define white noise.
- Find the variance
  - Approach used previously, multiplying and taking expectations, is harder

$$\gamma(0) = \phi_1\gamma(1) + \sigma_w^2 + \theta_1 \text{Cov}(X_t, w_{t-1}) = \phi_1\gamma(1) + \sigma_w^2 + \sigma_w^2\theta_1(\phi_1 + \theta_1)$$

$$\gamma(1) = \phi_1\gamma(0) + \theta_1\sigma_w^2$$

Leading terms are different

$$\gamma(2) = \phi_1\gamma(1)$$

$$\gamma(3) = \phi_1\gamma(2)$$

It's the eventual pattern that's imporant:
Recursion as found for an AR(1) process

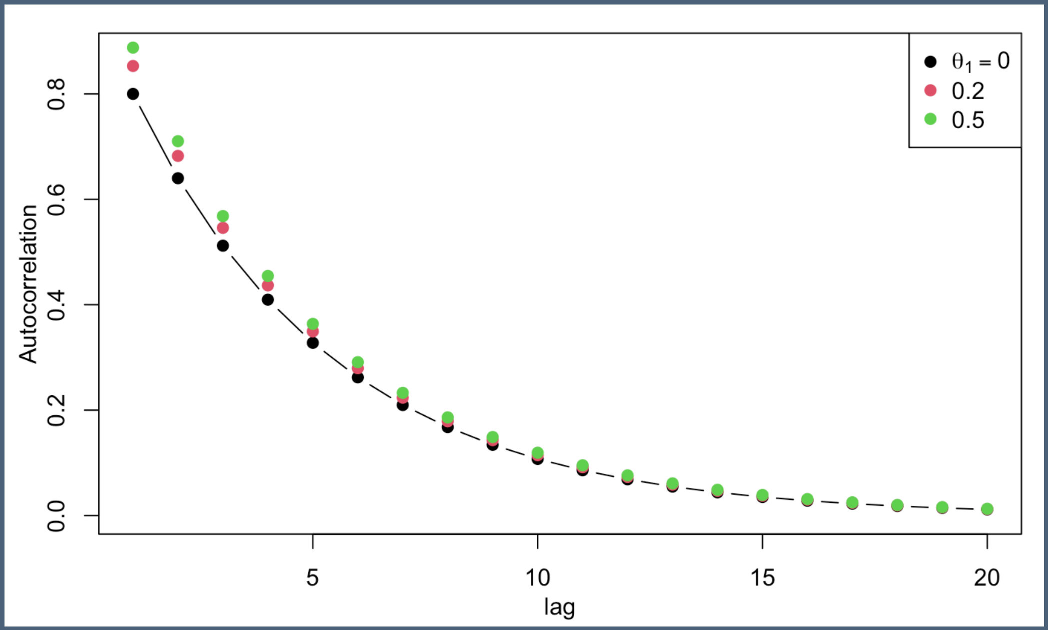
  - Solve as would any system of equations. e.g., plugging second equation into first gives

$$\gamma(0) = \sigma_w^2 \frac{1 + \theta_1^2 + 2\phi_1\,\theta_1}{1 - \phi^2}$$

The denominator resembles divisor in the expression for the variance of an AR(1) process.

ARMA Autocorrelation

- Similar to that of the associated AR(1) process
  - So similar that it is hard to distinguish the processes
- Example
  - Process is $X_t = 0.8\,X_{t-1} + w_t + \theta_1 w_{t-1},$



ACF falls off geometrically in every case, with the starting correlations determined by θ_1

One Motivation for ARMA Processes

- ARMA(1,1) processes appear frequently in analysis
- One explanation: add noise to an AR(1) process
  - AR(1) process $Y_t = \phi Y_{t-1} + w_t$
  - Independent noise u_t
  - Observe sum $X_t = Y_t + u_t$
- Autocovariances
  - Generalized difference of data $X_t - \phi X_{t-1} = w_t + u_t - \phi\, u_{t-1}$
  - Define $\epsilon_t = w_t + u_t$. Then have two process with the same autocovariances

$$w_t + u_t - \phi\, u_{t-1} \stackrel{D}{=} \epsilon_t + \xi\, \epsilon_{t-1}, \quad \xi = -\phi \sqrt{\frac{\sigma_w^2}{\sigma_w^2 + \sigma_u^2}}$$

  - Hence, adding independent noise to an AR(1) process produces a process whose autocovariances are those of an ARMA(1,1) process

What's next?

- The identification problem
  - Choosing p and q from an estimated ACF
  - Model selection criteria (AIC, BIC)

- Partial autocorrelations
  - Autocorrelations of the ARMA(1,1) process fall off roughly geometrically, but for the first
  - Hence, hard to distinguish from an AR(1) process
  - Partial autocorrelations offer big help to identify the model.

Enjoy Fall Break!