Statistics 5350/7110
Forecasting

Lecture 12
Identifying ARMA Models

Professor Robert Stine

Preliminaries

- Questions?

- Assignments
  - Assignment #3, due next Tuesday.

- Quick review
  - ARMA(p,q) process

$$X_t = \alpha + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} \quad \text{and} \quad \phi_p, \theta_q \neq 0$$

  - Causality and invertibility
  - Backshift polynomials and zeros/roots

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

  - Behavior of autocorrelations for AR, MA and ARMA: Decay vs cut-off.

Today's Topics

Text, §4.1-4.2

- Partial autocorrelation function (PACF)

- Role
  - Help identifying the model
  - Decide values for p and q (AR and MA orders)

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

  - Hard from ACF only: ACF decays geometrically for both AR and ARMA models

- Approaches
  - Without much computational resources: Identify from shape of ACF, PACF
  - With unbounded computational resources: Penalized likelihood, within limits
  - Estimating the parameters by maximum likelihood

Partial Autocorrelation Function

- Partial correlation
  - Correlation between two random variables holding other variables fixed
  - Use regression to remove the effects of other variables
  - Construction of the added variable plot in regression

- Partial autocorrelation function (PACF)
  - Partial correlation conditional on intervening variables

$$\phi_{11} = \text{Corr}(X_{t+1}, X_t)$$

Definition 4.17

$$\phi_{hh} = \text{Corr}(X_{t+h}, X_t \mid X_{t+h-1}, \ldots, X_{t+1}), \quad h = 2,3,\ldots$$

  - Remove the influence of the variables between X_{t+h} and X_t
  - Correlation between residuals from two regressions:

$$\phi_{hh} = \text{Corr}(X_{t+h} - \hat{X}_{t+h}, X_t - \hat{X}_t), \quad h = 2,3,\ldots$$

Regular correlation

where $\hat{X}_{t+h}$ and $\hat{X}_t$ are the predictions from regressing X_{t+h} and X_t on $X_{t+h-1}, \ldots, X_{t+1}$.

Partial Autocorrelation Examples

- Example for AR(1)
  - Initial value ϕ_{11} is the usual autocorrelation
  - Second value ϕ_{22} is the correlation between X_t and X_{t-2} after conditioning on X_{t-1}
  - Residual from regressing X_t on X_{t-1} is w_t which is uncorrelated with prior terms
  - Hence, $\phi_{22} = 0$. Similarly for ϕ_{hh}, $h = 3,4,\ldots$
- Example for AR(2)
  - ϕ_{11} is the usual autocorrelation, which happens to be $\phi_1/(1 - \phi_2)$
  - ϕ_{22} is the correlation between X_t and X_{t-2} given X_{t-1}, which is ϕ_2
  - In general, for an AR(p) process, $\phi_{pp} = \phi_p$
  - $\phi_{33} = 0$ since regression of X_t on X_{t-1} and X_{t-2} is w_t, uncorrelated with prior variables
- Example for MA(q)
  - Recall that invertible MA(q) is an infinite order AR, so ϕ_{hh} decay toward zero rather than cut off

Summary of Patterns in ACF

- AR(p)
  - Autocorrelation geometrically decays
  - Partial autocorrelation cuts off after p nonzero values
- MA(q)
  - Autocorrelation cuts off after q nonzero values
  - Partial autocorrelation geometrically decays
- ARMA(p,q)
  - Both autocorrelation and partial autocorrelation decay geometrically
- Identification procedure
  - Plot estimated autocorrelation and partial autocorrelation functions
  - Look for cut-off of autocorrelation (signals MA) or partial autocorrelation (signals AR)
  - If neither cut off, result is ARMA
  - Great conceptually, but we observe estimated correlations, not $\rho(h)$ and ϕ_{hh}

Table 4.1 Summarizes

Properties of Estimated ACF, PACF

- Recall properties of estimated autocorrelation
- Sampling variance
  - White noise standard error is approximately 1/√n
    $$\text{Var}(\hat{\rho}(h)) \approx \frac{1}{n}$$
    Similar asymptotic estimate holds for partial correlations
  - For an MA process of order q (i.e., as if ρ(h) = 0 for h > q)
    $$\text{Var}(\hat{\rho}(h)) \approx \frac{1}{n}\left(1 + 2(\rho(1)^2 + \rho(2)^2 + \cdots + \rho(q)^2)\right), \quad \text{for} \quad h > q$$
- Dependence
  - Estimated autocorrelations are typically more autocorrelated than the process itself.
  - Consequence: Once the ACF deviates from ρ(h), the deviations persist for several lags.

Examples of Identification

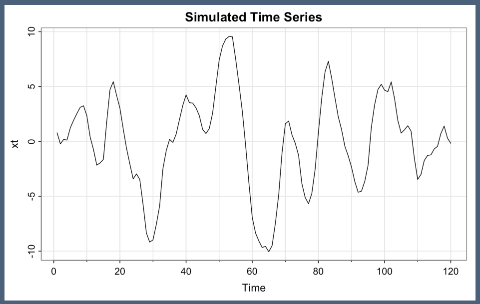
- Simulated Gaussian time series from several ARMA models
- Ideal cases
  - All are stationary
  - Long series with n = 120 values
  - None of the time series has outliers
  - None of the time series has measurement error
  - There's no change in any of these processes over time

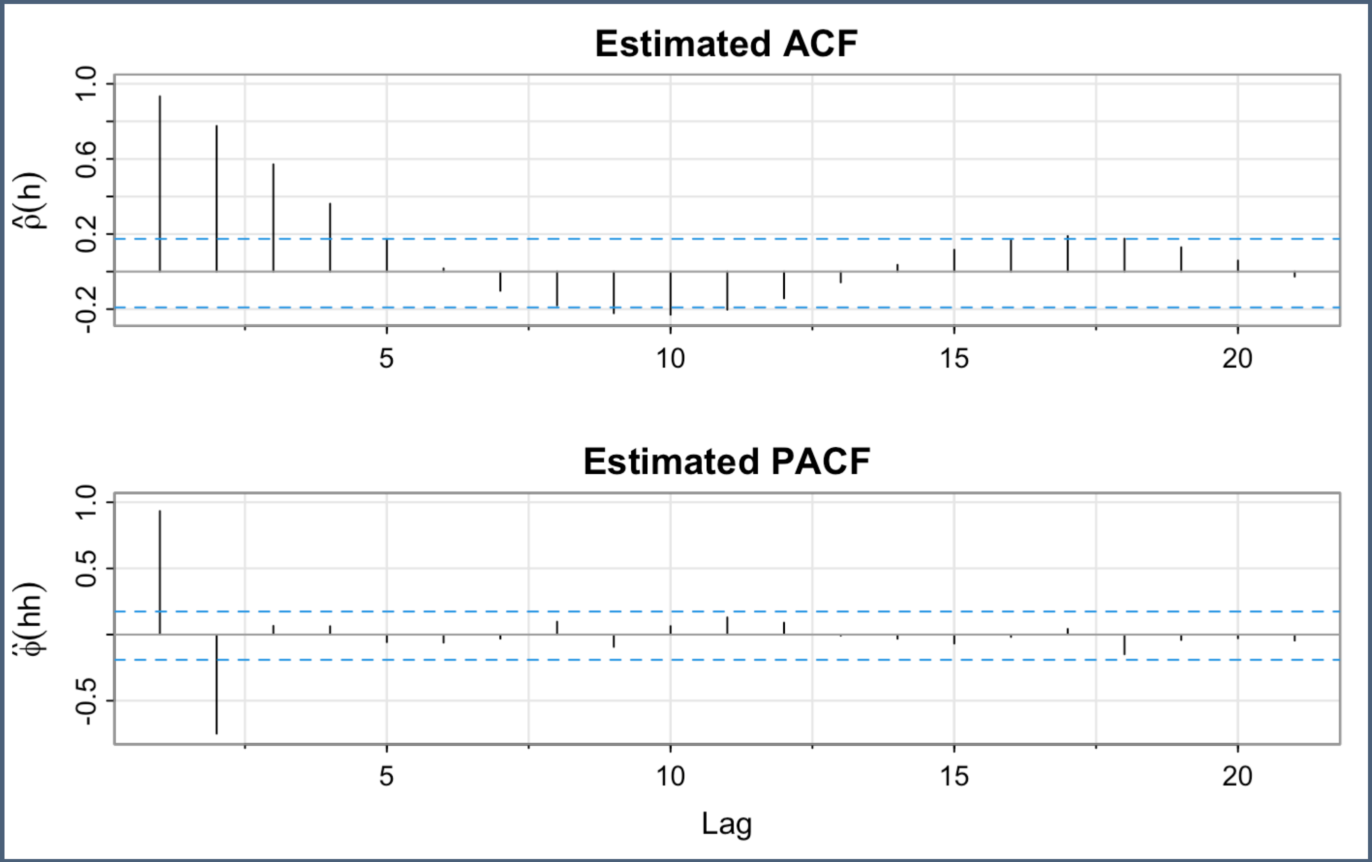
as if 10 years of clean monthly economic data with no recessions or pandemics

- Code generating these examples in the Rmd file for this class

Examples

- Identify orders p and q of an ARMA process from these graphs
  - Gaussian process
  - n = 120

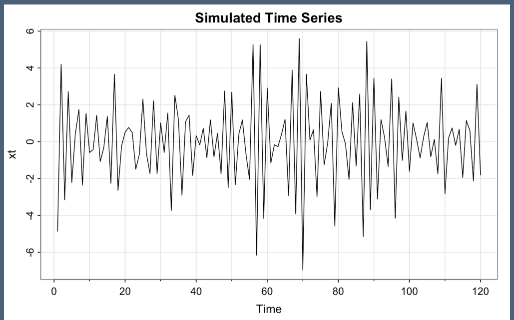


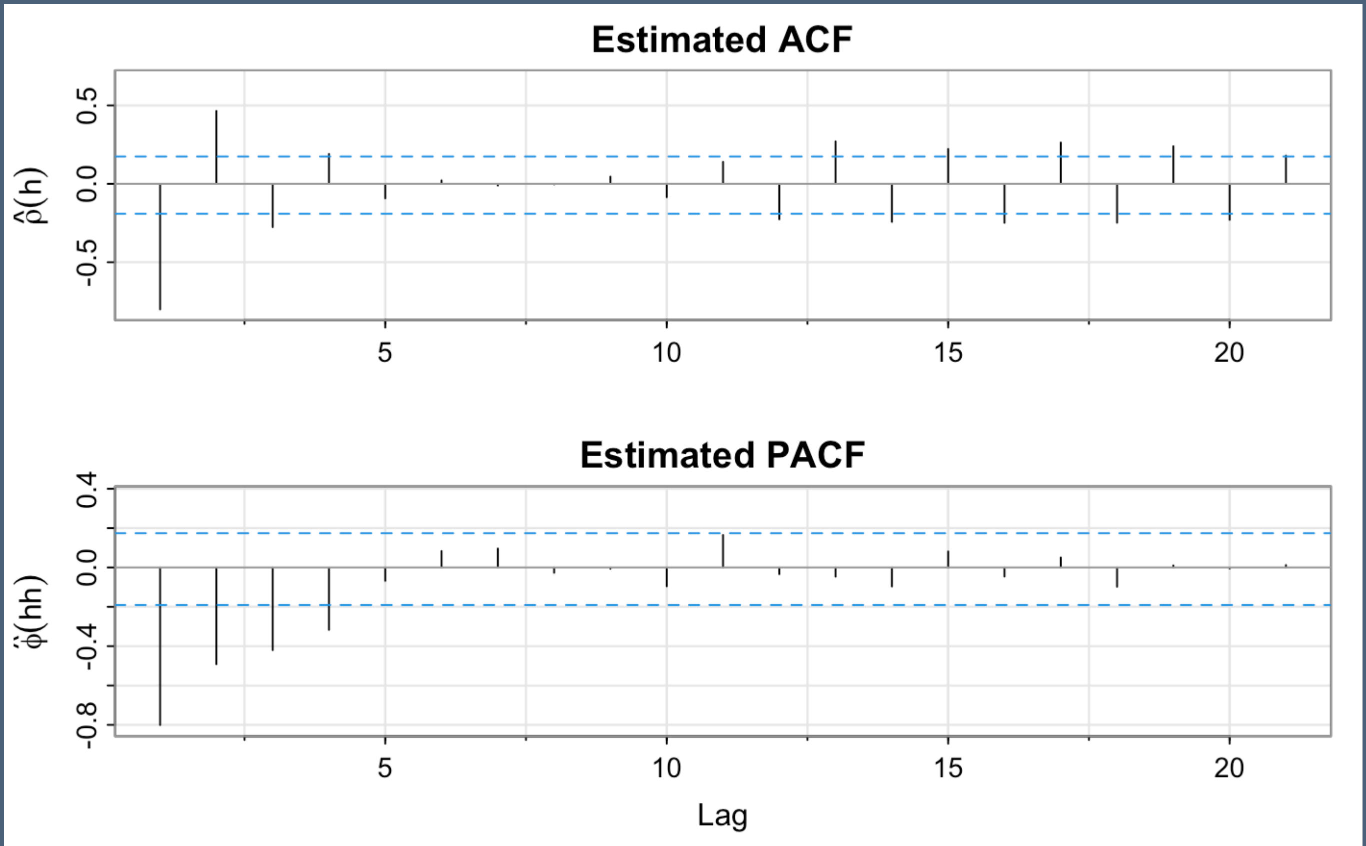


AR(2)

Examples

- Identify orders p and q of an ARMA process from these graphs
  - Gaussian process
  - n = 120

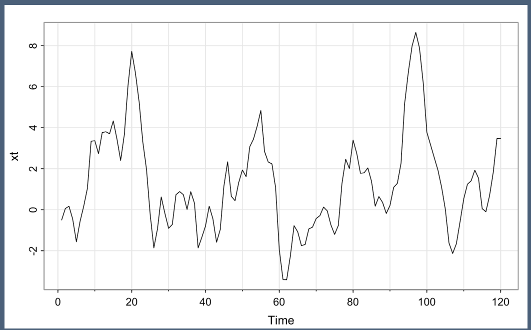


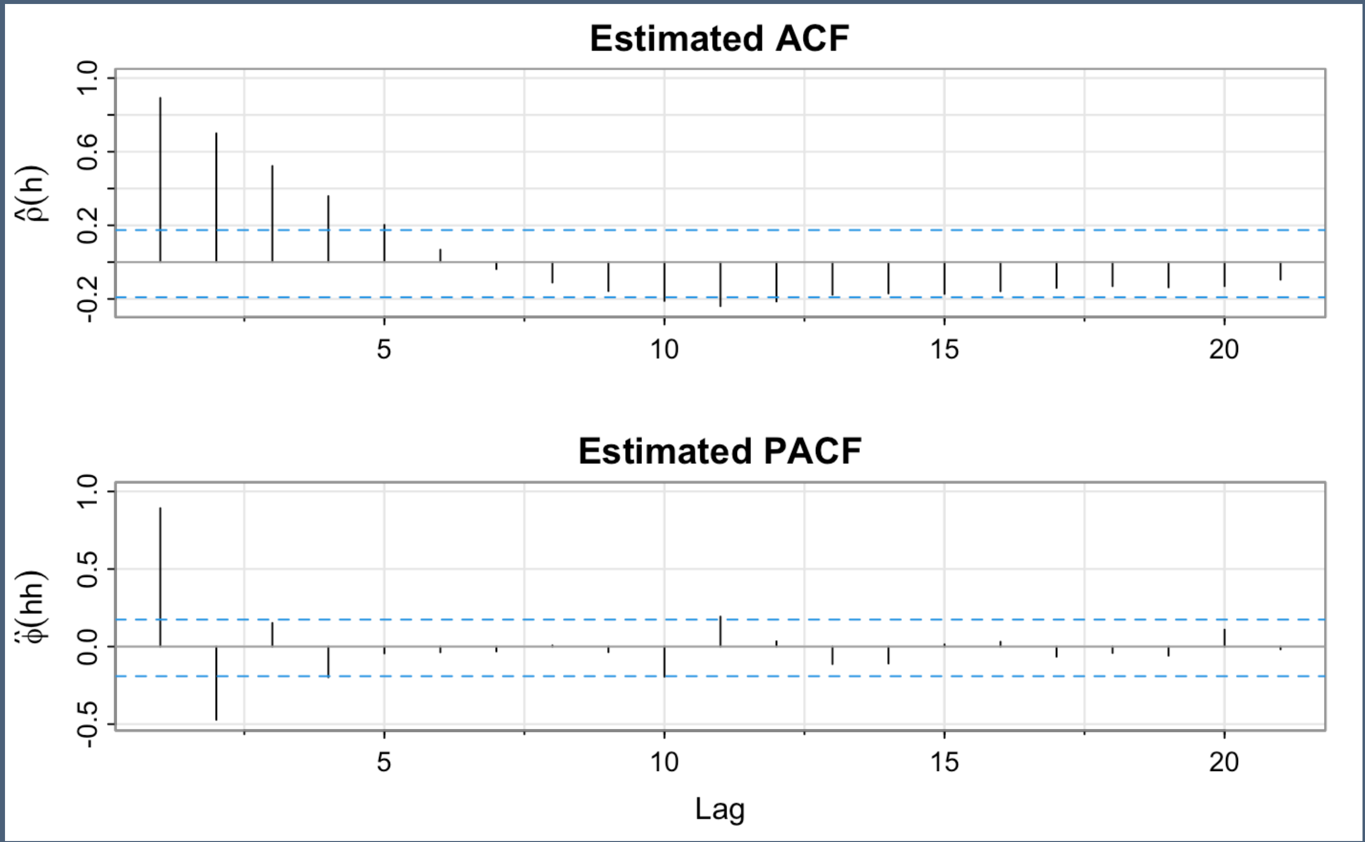


MA(3)

Examples

- Identify orders p and q of an ARMA process from these graphs
  - Gaussian process
  - n = 120

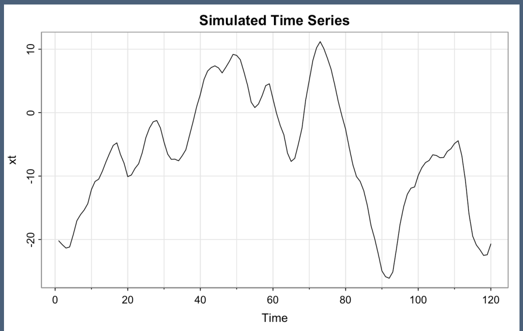


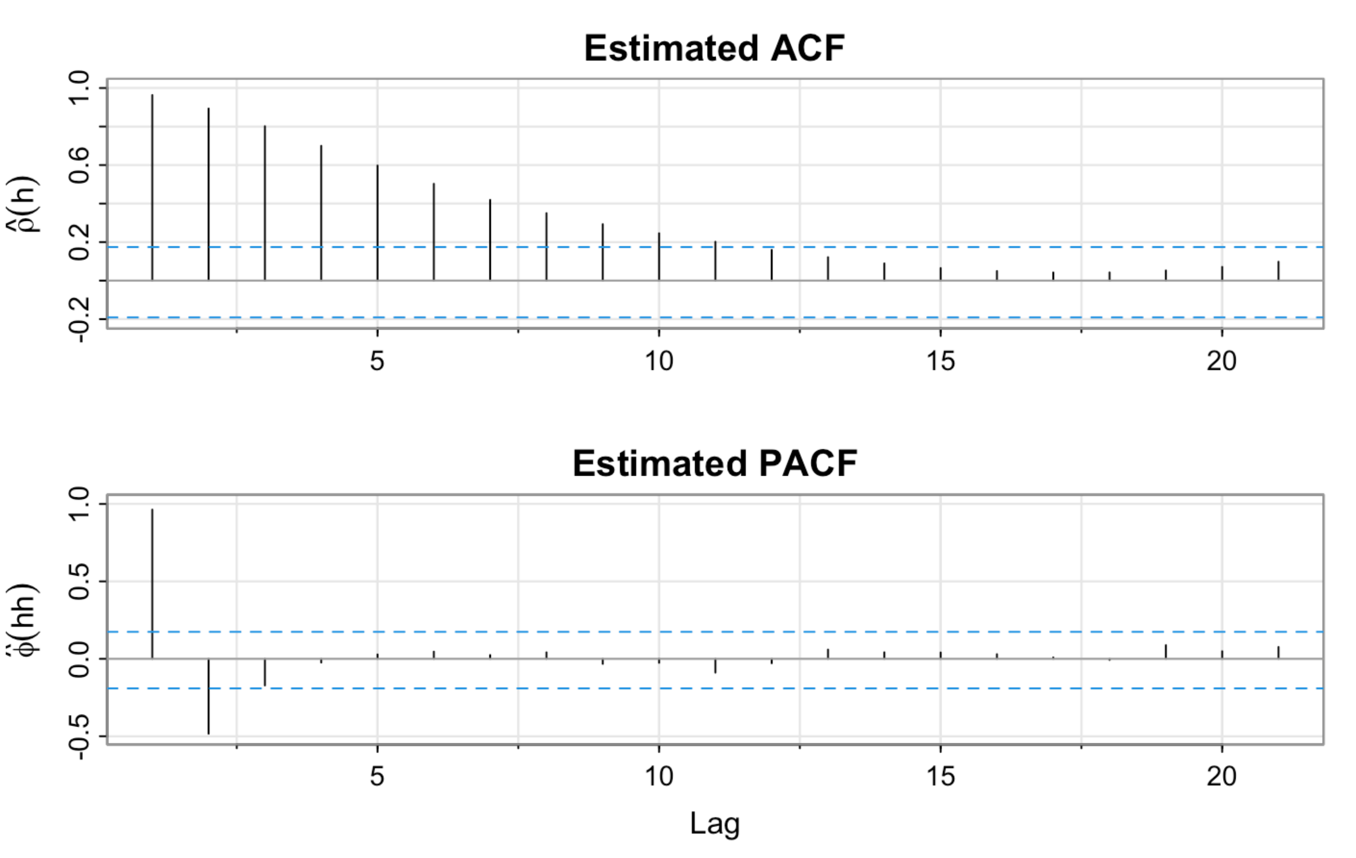


ARMA(1,1)

Examples

- Identify orders p and q of an ARMA process from these graphs
  - Gaussian process
  - n = 120





AR(3)

Discussion of Identification

- Hard to identify p and q from ACF and PACF unless you have a long series
  - Large standard errors
  - When an estimate is not far from zero, has it cut off or is it declining geometrically?
  - Estimates often resemble random walk, looks like decay but just random variation
  - Complication: Longer the series, larger the chance that the model has changed!

- Alternative: Model selection criteria
  - Assume we know how to fit a probability model (next lecture)
  - Assume time series is Gaussian (or at least close to Gaussian)
  - Maximize Gaussian likelihood for a given specification of orders p and q
  - Essentially minimizes sum of squared error as in least squares regression
  - Defer to model selection criterion:

    Penalize likelihood using AIC or BIC

    Choose p, q that maximize penalized likelihood

Model Selection

- Compute the log-likelihood for range of values of p and q
  - Like R^2 and the residual sum-of-squares in regression, the model with the most parameters will always have the highest log-likelihood.

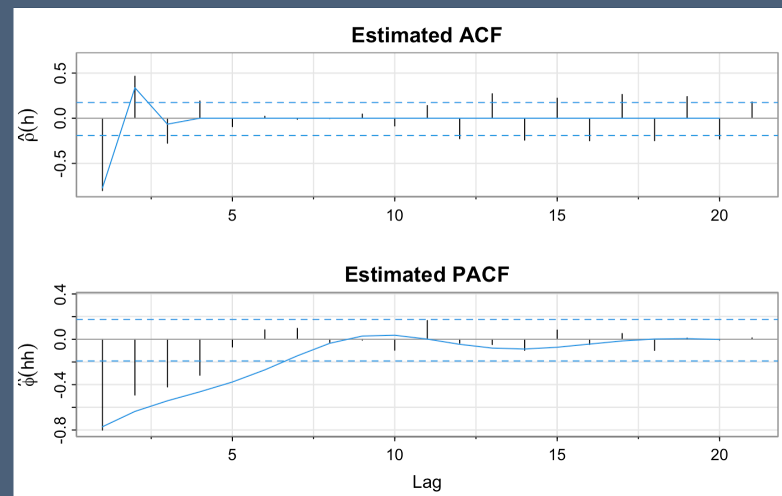
	q=0	q=1	q=2	q=3	q=4	q=5
p=0	-276.69	-211.01	-180.48	-168.65	-164.48	-159.59
p=1	-179.32	-159.66	-159.27	-159.08	-159.06	-157.87
p=2	-162.47	-159.16	-156.76	-158.78	-158.78	-157.61
p=3	-160.80	-158.80	-158.72	-156.04	-156.37	-155.50
p=4	-158.18	-158.09	-156.68	-156.45	-157.18	-155.01
p=5	-158.13	-157.82	-156.85	-156.38	-154.18	-153.89

- Penalize for the size of the model
  - Example: "Best" model has minimum AIC
    $AIC(k) = (-2 \log \text{likelihood} + 2k)/n$
  - Table shown to right subtracts the minimum AIC statistic from all the values
  - The best model (minimum AIC) identified by the entry with zero. BIC will often differ.

	q=0	q=1	q=2	q=3	q=4	q=5
p=0	231.86	102.49	43.45	21.78	15.45	7.65
p=1	39.12	1.80	3.02	4.64	6.61	6.23
p=2	7.43	2.79	0.00	6.04	8.04	7.70
p=3	6.09	4.09	5.91	2.56	5.22	5.49
p=4	2.84	4.66	3.84	5.38	8.85	6.50
p=5	4.75	6.11	6.18	7.24	4.84	6.26

Comments Before R Code

- Estimated ACF vs process ACF
  - Since simulated, we know the true ACF
  - As in this example, deviations persist.



- AIC vs BIC
  - AIC often picks higher (p,q) than process
  - BIC eventually gets this right, assuming there is a right answer (BIC is "consistent")

- Parameter estimates
  - Explains why AIC chooses higher (p,q)
  - MA(3), but AIC prefers ARMA(3,4)
  - Estimates are significant

	q=0	q=1	q=2	q=3	q=4	q=5
p=0	256.23	116.84	23.73	0.80	2.70	3.07
p=1	127.63	34.88	3.07	2.64	3.17	3.44
p=2	91.38	26.24	4.56	6.98	0.15	5.30
p=3	57.24	13.59	1.35	3.28	0.00	4.14
p=4	24.89	5.81	3.23	3.06	1.99	0.14
p=5	9.08	16.04	4.84	6.83	2.31	5.14

Coefficients:

	Estimate	SE	t.value	p.value
ar1	-0.0074	0.1050	-0.0706	0.9438
ar2	-0.7257	0.0523	-13.8659	0.0000
ar3	-0.4474	0.0990	-4.5181	0.0000
ma1	-2.1349	0.0854	-25.0076	0.0000
ma2	2.6807	0.1694	15.8243	0.0000
ma3	-1.9932	0.1708	-11.6719	0.0000
ma4	0.7627	0.1036	7.3639	0.0000
xmean	-0.0038	0.0101	-0.3755	0.7080

What's next?

- Estimating models
  - We picked a model using an elaborate procedure that penalizes the log of the likelihood function
  - What does the maximum likelihood estimation procedure do?
  - Turns out that it's very similar to fitting a least-squares regression, as we can see from the similarity of maximum likelihood estimates in the examples to regression estimates

  - Regression models require a careful diagnostic analysis.
    Where's that for fitting an ARMA model?