Statistics 5350/7110
Forecasting

Lecture 17
Forecasting ARIMA Models

Professor Robert Stine

Preliminaries

- Questions?
- Grading
- Assignments
  - Assignment 4 is posted
- No office hours today... Halloween!
- Quick review
  - Forecasting ARMA models
  - Role of difference equation form
  - Role of moving average form

Today's Topics

Text, §5.1

- Definition of an ARIMA model
  - ARMA model fit to the differences

- Examples of ARIMA models

- Forecasting special cases of ARIMA models… the details
  - Random walk versus a stationary AR(1) model
  - Integrated autoregression, ARIMA(1,1,0) or ARI(1,1)
  - Integrated moving average, ARIMA(0,1,1) or IMA(1,1)

ARIMA Model

- Incorporate differencing into specification of model
  - Three choices specify model

    p Order of the autoregression

    d Degree of differencing (p, d, q) Definition 5.1

    q Order of the moving average

$$\nabla^d X_t = \alpha + \phi_1(\nabla^d X_{t-1}) + \cdots + \phi_p(\nabla^d X_{t-p}) + w_t + \theta_1 w_{t-1} + \cdots \theta_q w_{t-q}$$

$$\phi(B)\ \nabla^d\, X_t = \alpha + \theta(B)\ w_t \qquad \nabla^d = (1-B)^d$$

  - Intercept in the model is

$$\alpha = \delta\ (1 - \phi_1 - \phi_2 - \cdots - \phi_p), \quad \delta = E(\nabla^d X_t)$$

  - Equivalent to saying that the differenced data $(1\text{-}B)^d X_t$ is ARMA(p,q)
  - Practice: Identify if non-stationary, then model differences as needed as ARMA
- Expanded scope
  - Random walk is a trivial ARIMA(0,1,0) model
  - Exponential smoothing is an ARIMA(0,1,1) model (next class)

Estimating an ARIMA Model

- Same as for ARMA models
  - Difference the data
  - Fit parameters to differences as if starting with an ARMA process

- Advantage of ARIMA style
  - Why bother since "essentially" an ARMA model: Computing forecasts
  - Software sums differences as needed
  - Computes prediction
  - Computes standard error of prediction (MSPE, mean squared prediction error)

- R details
  - `arima` and `sarima` functions incorporate choice of differencing parameter
  - `ARMAtoMA` works for non-stationary models to get MA weights

Examples of Nonstationary Macro Time Series

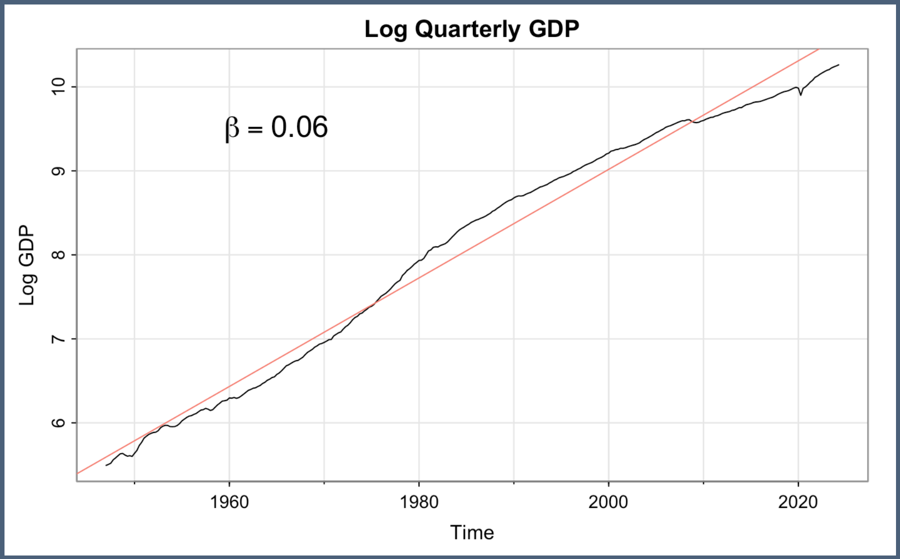
- Gross domestic product (GDP)
- Returns on the stock market (SP500)
- Consumer price index

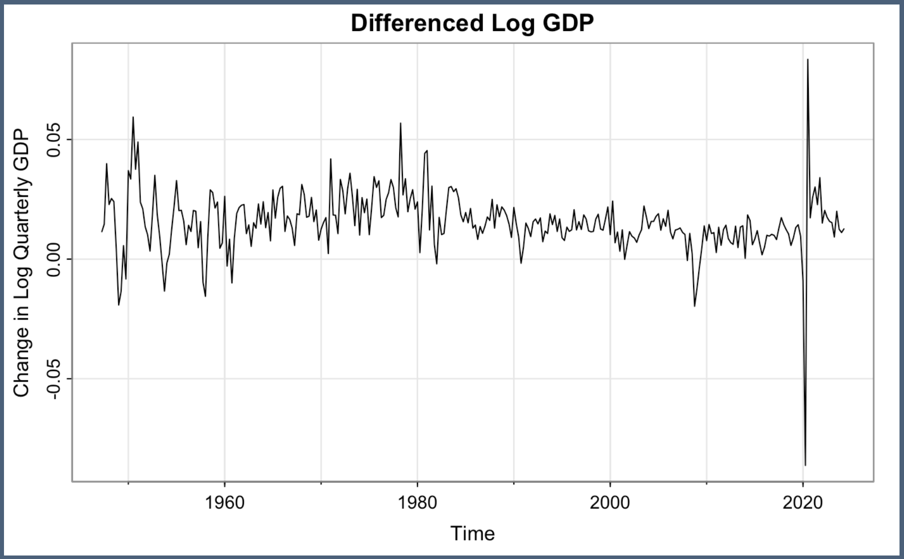
Examples of Nonstationary Time Series

- Nominal US Gross Domestic Product
  - Seasonally adjusted
  - Log scale shows long-term trend
  - Differences look stationary, but for huge Covid outlier
  - Change in volatility associated with (inflation)

Change (log) ≈ Percent Change

$$\begin{aligned}\nabla \log X_t &= \log \frac{X_t}{X_{t-1}} \\ &= \log \frac{X_t - X_{t-1} + X_{t-1}}{X_{t-1}} \\ &= \log \left(1 + \frac{X_t - X_{t-1}}{X_{t-1}}\right) \\ &\approx \frac{X_t - X_{t-1}}{X_{t-1}}\end{aligned}$$



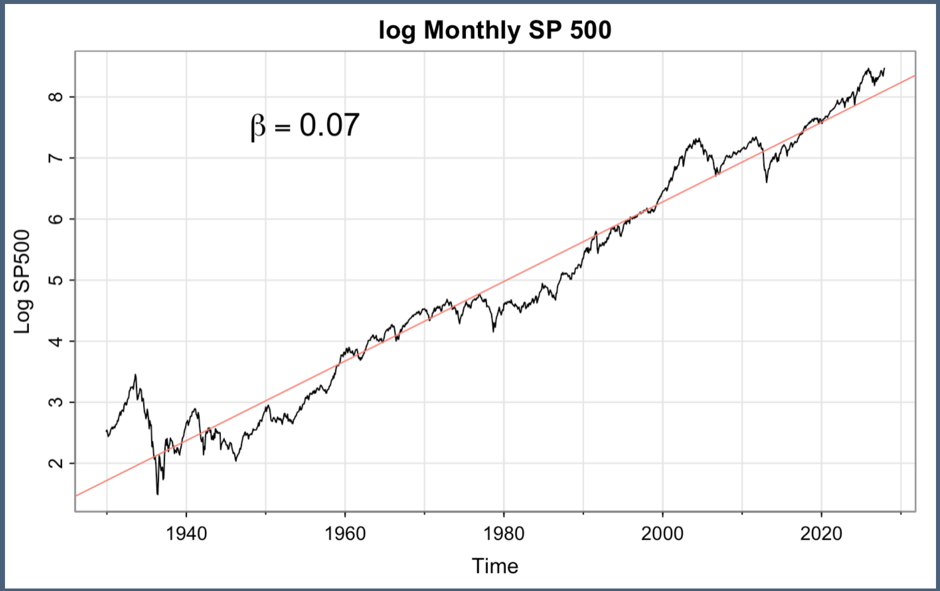


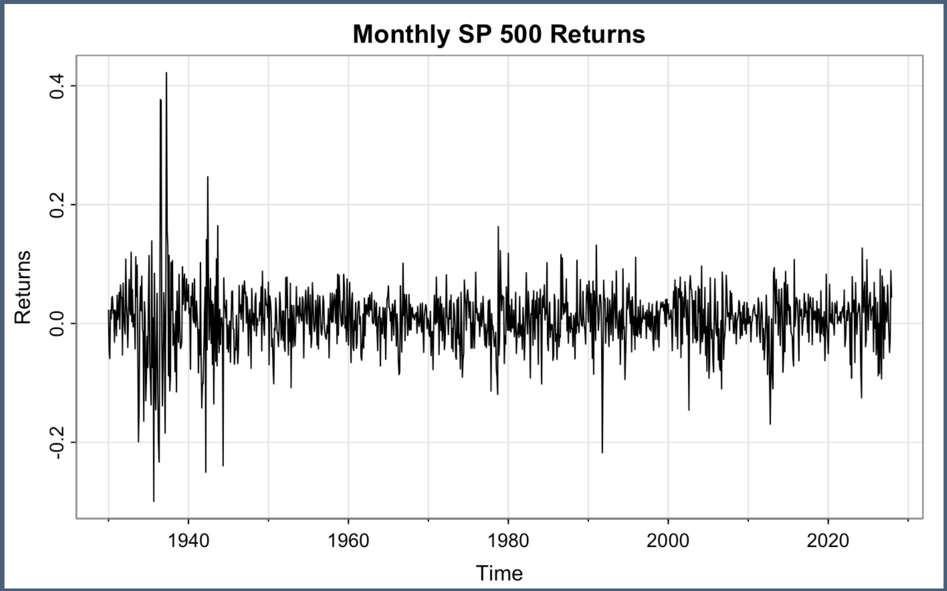
FRED data

Text, Example 5.6 goes though 2002

Examples of Nonstationary Time Series

- SP 500 Stock Market Index
  - Log scale shows long-term trend
  - Return defined as $(P_t - P_{t-1})/P_{t-1}$
  - Periods of persistent high volatility

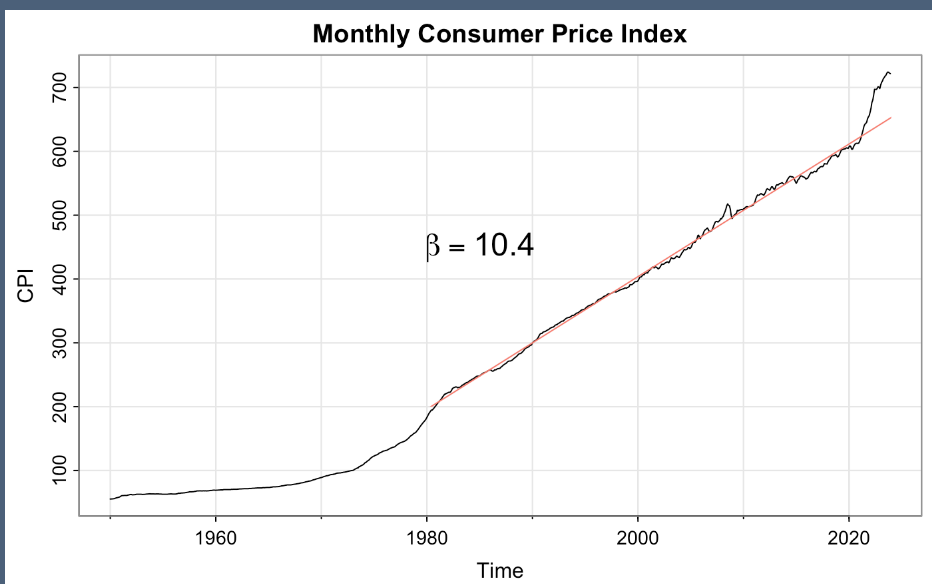


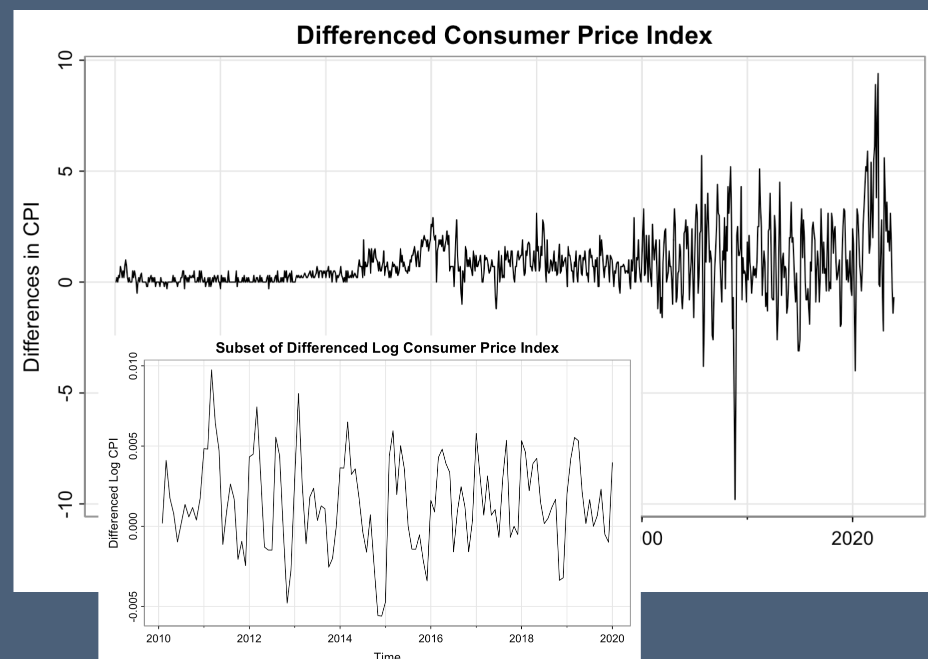


CRSP data, via WRDS

Examples of Nonstationary Time Series

- Consumer price index (CPI)
  - Index growth, with different regimes (linear from 1980 through 2019)
  - Differences appear more stationary
  - But differencing reveals changes in volatility





CRSP data, via WRDS

Three Models for Nonstationary Time Series

• Random walk versus AR(1)	Example 5.3
• Integrated autoregression	Example 5.4
• Integrated moving average	Example 5.5

Random Walk with Drift

Example 5.3

- Model

$$X_t = \delta + X_{t-1} + w_t$$

- Forecast
  - Assume we know the model and have data $X_1, X_2, \ldots, X_n$.
  - First forecast is obvious

$$\hat{X}_{n+1} = \delta + X_n$$

  - Take conditional expectation to find the next

$$X_{n+2} = \delta + X_{n+1} + w_{n+2} \quad \Rightarrow \quad \hat{X}_{n+2} = \delta + (\delta + X_n) + 0 = 2\,\delta + X_n$$

  - Predictions are linear trend from last value

$$\hat{X}_{n+m} = m\,\delta + X_n$$

Differences are $\delta + w_t$, so just add them up as extrapolate

- Mean squared prediction error (MSPE)
  - Back-substitution expression shows the accumulation of errors (MA weights $\psi_j = 1$)

$$X_{n+m} = m\,\delta + \underbrace{w_{n+m} + w_{n+m-1} + \cdots + w_{n+1}}_{m \text{ terms}} \quad \Rightarrow \quad E(X_{n+m} - \hat{X}_{n+m})^2 = m\,\sigma_w^2$$

Grows and grows with no limit

Random Walk with Drift

Example 5.3

- Random walk
  - Difference equation sets $\phi = 1$

$$X_t = \delta + X_{t-1} + w_t$$

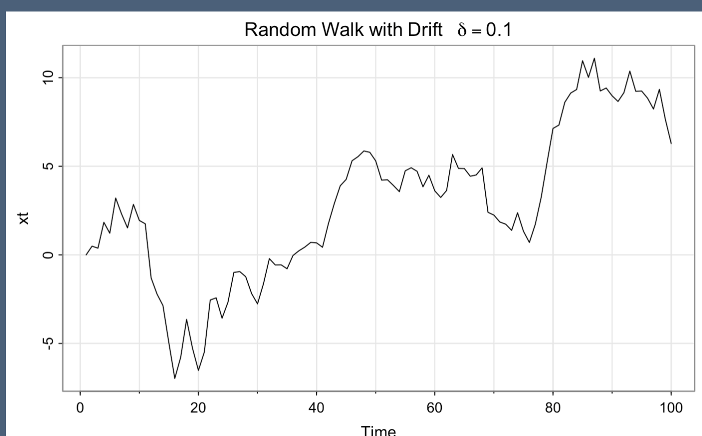
  - Forecast does not revert to mean (no constant mean to revert towards)
  - Forecast expected squared error MSPE grows with extrapolation
- AR(1) model
  - Constrains $|\phi| < 1$

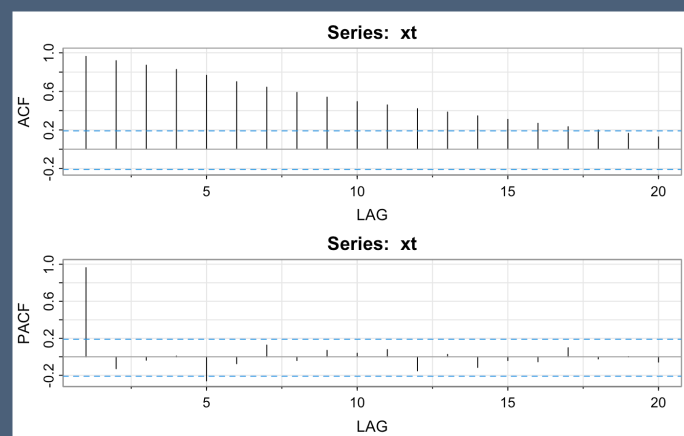
$$X_t = \alpha + \phi X_{t-1} + w_t$$

  - Stationary, though ϕ might be close to 1.
  - Forecast reverts to mean $\mu = \alpha + \phi\mu \quad \Rightarrow \quad \mu = \alpha/(1-\phi)$
  - MSPE grows to limit at $\text{Var}(X_t) = \sigma_w^2/(1-\phi^2)$
- Distinguishing these models
  - Known as the "unit root" problem in statistics/econometrics (Text §8.2)

Competing Risks

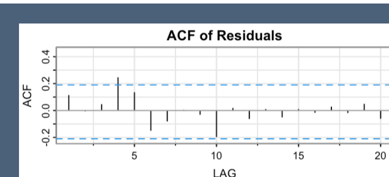
- Which is the worse mistake?
  - The process is a random walk, but you model it as an AR(1)
  - The process is an AR(1), but you model it as a random walk
- Short term
  - AR(1) forecasts revert to the mean, whereas random walk forecasts drift
  - AR(1) MSPE approaches variance, whereas random walk MSPE grows without bound
- Example






```
Coefficients:
      Estimate     SE t.value p.value
ar1     0.9650 0.0223 43.1967  0.0000
xmean   2.9453 2.4602  1.1972  0.2341

sigma^2 estimated as 1.150809 on 98 de
```



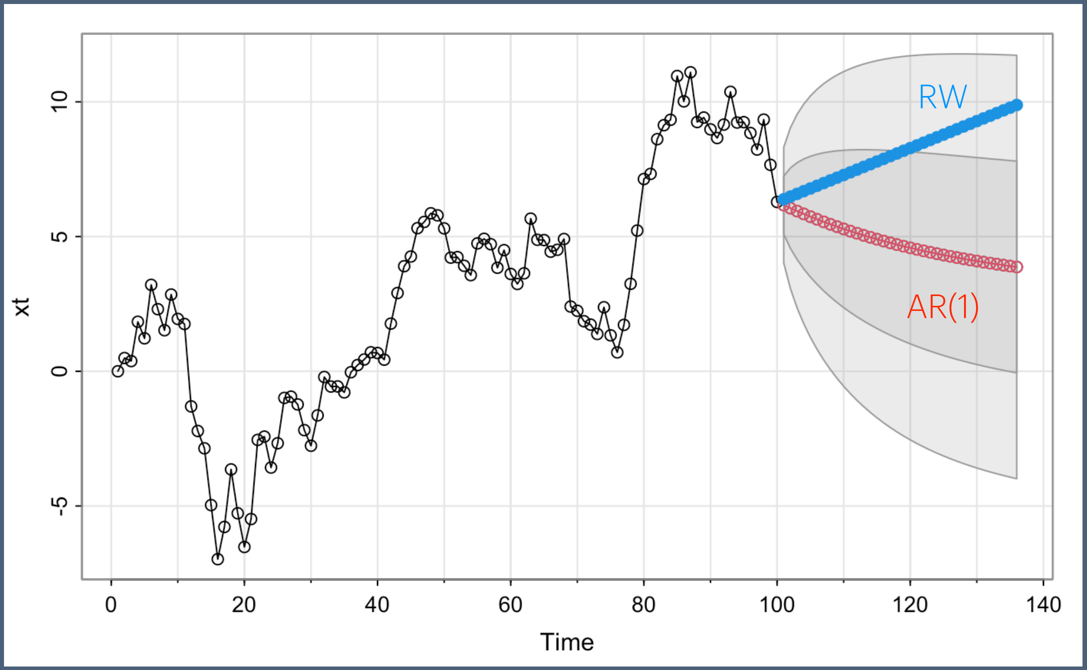
Not easy to distinguish

Competing Risks

- Situation
  - Process is a random walk with small positive drift δ = 0.1
  - Fitted model is AR(1), leaving small amount of residual autocorrelation

- Forecasts
  - AR(1) forecasts revert to process mean
  - Actual mean trends up (δ = 0.1)
- Econometric preference
  - Common to assume RW unless convinced that it is not
  - As if null hypothesis is H_0: X_t is RW rather than other way around.
  - Not the right standard error!

```
Coefficients:
      Estimate     SE t.value p.value
ar1     0.9650 0.0223 43.1967  0.0000
xmean   2.9453 2.4602  1.1972  0.2341
```

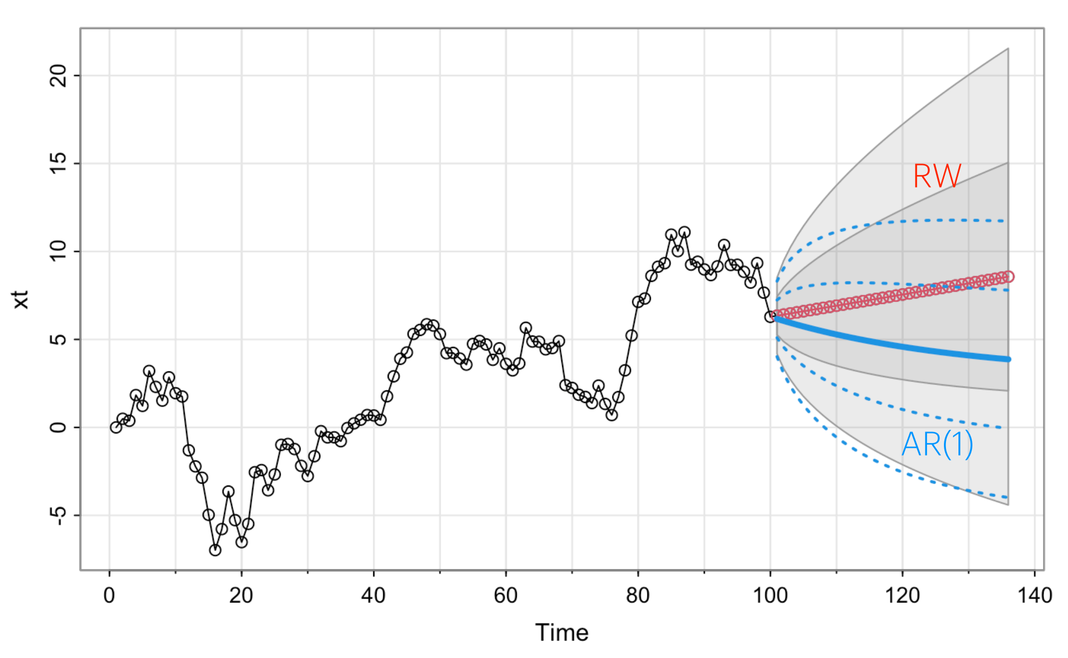


Competing Risks

- Situation
  - Process is a random walk (RW) with small positive drift δ = 0.1
  - Suppose model it instead as RW with possible drift

- Forecasts
  - RW forecasts trend upward
  - RW standard errors grow

- False sense of accuracy?
  - AR(1) will overstate accuracy if time series is a random walk.
  - RW is more conservative:
    Less confident of what the future.



Integrated Autoregression

Example 5.4

- ARIMA(1,1,0)
  - Differences are AR(1) rather than uncorrelated as in a random walk
  - Resembles an AR(2)

a.k.a. ARI(1,1)

$$\nabla X_t = \delta + \phi\,(\nabla X_{t-1}) + w_t \quad \Rightarrow \quad X_t = \delta + (1+\phi)\,X_{t-1} + \phi\,X_{t-2} + w_t$$

- Forecasting
  - 1 step ahead

$$\hat{X}_{n+1} = [X_{n+1}] = [\delta + (1+\phi)X_n - \phi X_{n-1} + w_{n+1}] = \delta + (1+\phi)X_n - \phi X_{n-1} + 0$$

error $\quad X_{n+1} - [X_{n+1}] = w_{n+1}$

  - 2 steps ahead

$$[X_{n+2}] = [\delta + (1+\phi)X_{n+1} - \phi X_n + w_{n+2}] = \delta + (1+\phi)[X_{n+1}] - \phi X_n + 0$$

error $\quad X_{n+2} - [X_{n+2}] = w_{n+2} + (1+\phi)(X_{n+1} - [X_{n+1}]) = w_{n+2} + (1+\phi)w_{n+1}$

weight on w_{n+1} increases

  - 3 steps ahead

$$[X_{n+3}] = [\delta + (1+\phi)X_{n+2} - \phi X_{n+1} + w_{n+3}] = \delta + (1+\phi)[X_{n+2}] - \phi[X_{n+1}] + 0$$

error $\quad X_{n+3} - [X_{n+3}] = w_{n+3} + (1+\phi)w_{n+2} + (1+\phi+\phi^2)w_{n+1}$

Do you see a pattern in the MA weights?

Integrated Autoregression

- Another way to see what's happening to prediction error
- Moving average representation
  - Not a stationary process, but can still compute the MA weights.
  - Very simple r.h.s. since $\theta(B) = 1$

$$\phi(B)\,\psi(B) = \theta(B) = 1$$

  - Solve by equating coefficients recursively (as before)

$$\left(1 - (1+\phi)B + \phi B^2\right)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

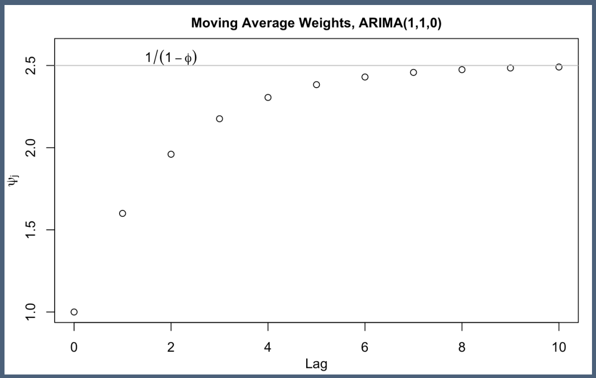
  - Method is more "mechanical" than the direct manipulation on prior slide.
- General pattern
  - Geometric series

$$\psi_j = 1 + \phi + \phi^2 + \cdots + \phi^j, \quad j = 1,2,\ldots$$

  - Limiting value is

$$\psi_j \to 1/(1-\phi)$$

Do you see a pattern in these MA weights?

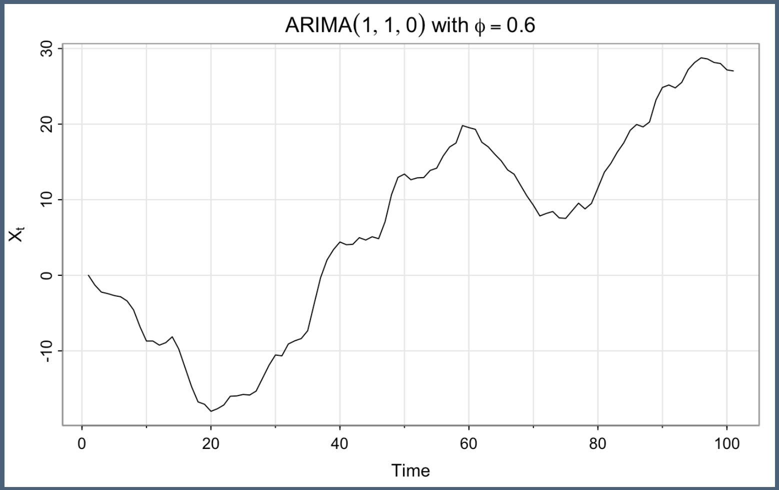


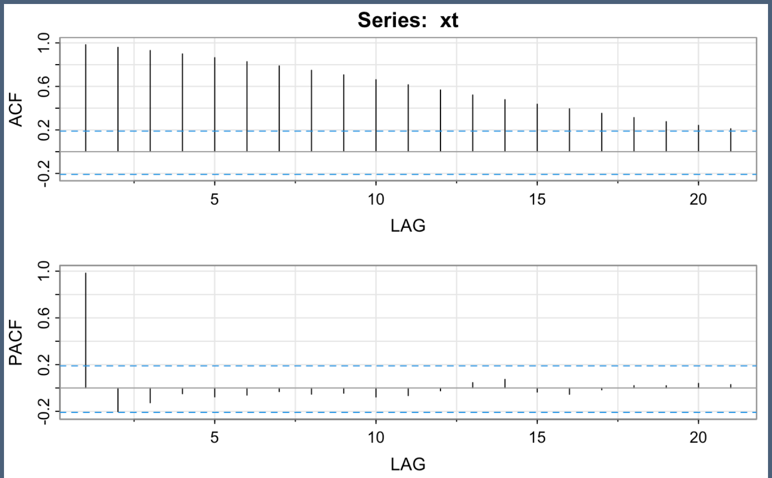
Example ARIMA(1,1,0)

- ARIMA(1,1,0) process
  - Model is

$$\nabla X_t = 0.6\,\nabla X_{t-1} + w_t$$

- Observed time series



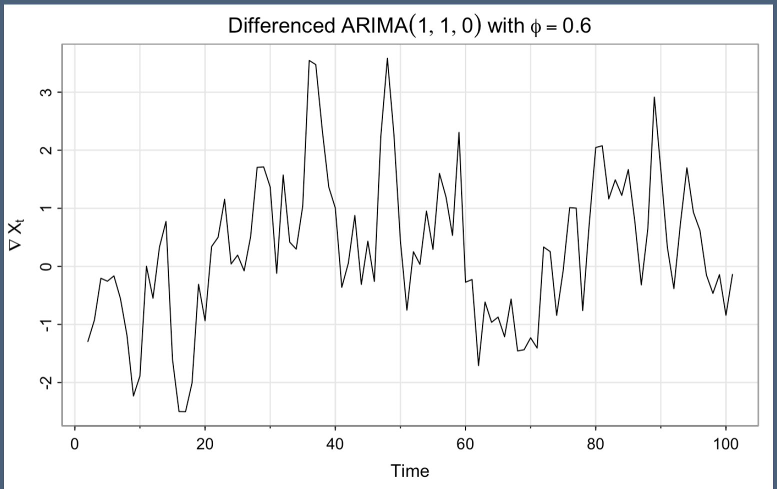


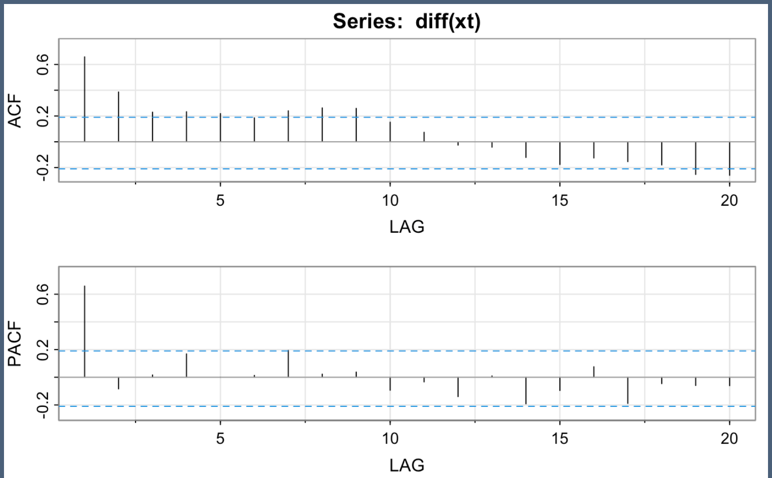
Example ARIMA(1,1,0)

- ARIMA(1,1,0) process
  - Model is

$$\nabla X_t = 0.6\,\nabla X_{t-1} + w_t$$

- Observed differences





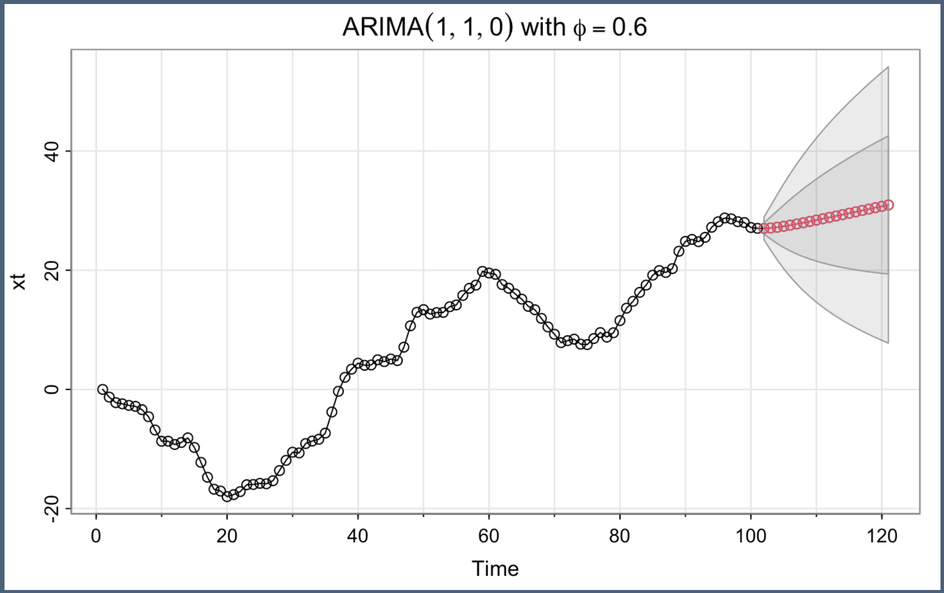
Estimates and Forecasts

- Summary of estimates

```
Coefficients:
          Estimate     SE t.value p.value
ar1         0.6629 0.0744  8.9134  0.0000
constant    0.2329 0.2774  0.8397  0.4031

sigma^2 estimated as 0.9068293 on 98 degrees
```

- Forecasts



Integrated Moving Average

Example 5.5

a.k.a. IMA(1,1)

Text uses different notation
$X_t = X_{t-1} + w_t - \lambda\, w_{t-1}$
because of connection to exponential smoothing.

- ARIMA(0,1,1)
  - Differences are moving average k
  - Resembles an AR(2)

$$\nabla X_t = \delta + \theta w_{t-1} + w_t \quad \Rightarrow \quad X_t = \delta + X_{t-1} + \theta w_{t-1} + w_t$$

- Forecasting
  - 1 step ahead

$$\hat{X}_{n+1} = [X_{n+1}] = [\delta + X_n + \theta w_n + w_{n+1}] = \delta + X_n + \theta w_n + 0$$

error $\quad X_{n+1} - [X_{n+1}] = w_{n+1}$

  - 2 steps ahead

$$[X_{n+2}] = [\delta + X_{n+1} + \theta w_{n+1} + w_{n+2}] = \delta + [X_{n+1}] + 0 + 0 \quad = 2\,\delta + X_n + \theta w_n$$

error $\quad X_{n+2} - [X_{n+2}] = w_{n+2} + \theta\, w_{n+1} + w_{n+1} = w_{n+2} + (1+\theta)\, w_{n+1}$

  - 3 steps ahead

$$[X_{n+3}] = [\delta + X_{n+2} + \theta w_{n+2} + w_{n+3}] = \delta + [X_{n+2}] + 0 + 0 \quad = 3\,\delta + X_n + \theta w_n$$

error $\quad X_{n+3} - [X_{n+3}] = w_{n+3} + (1+\theta)\, w_{n+2} + (1+\theta)\, w_{n+1}$

Integrated Autoregression

- Another way to see what's happening to prediction error
  - Polynomial algebra replaces expressions for prediction error

- Moving average representation
  - Backshift polynomial form is

$$\phi(B)\,\psi(B) = \theta(B) \quad \Rightarrow \quad (1-B)\,\psi(B) = 1 + \theta\,B$$

  - Solve by equating coefficients recursively (as before)

$$(1-B)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1 + \theta\,B$$

  - General pattern emerges quickly

$$\psi_0 = 1,\ \psi_1 = 1+\theta,\ \psi_2 = \psi_1,\ \psi_3 = \psi_2,\ \ldots \qquad \text{or} \qquad \psi_j = 1+\theta, \quad j = 1,2,\ldots$$

- Mean squared prediction error
  - Squared error grows linearly with increasing extrapolation

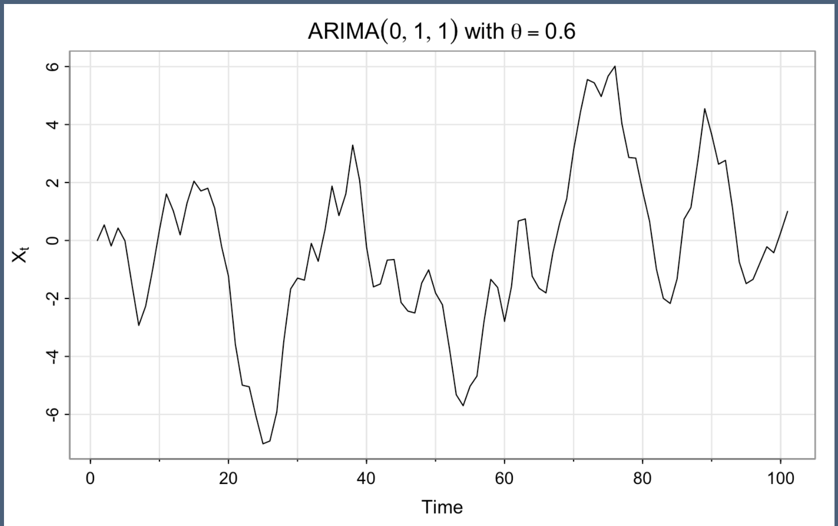
$$E(X_{n+m} - \hat{X}_{n+m})^2 = \sigma_w^2\,\left(1 + (m-1)(1+\theta)^2\right)$$

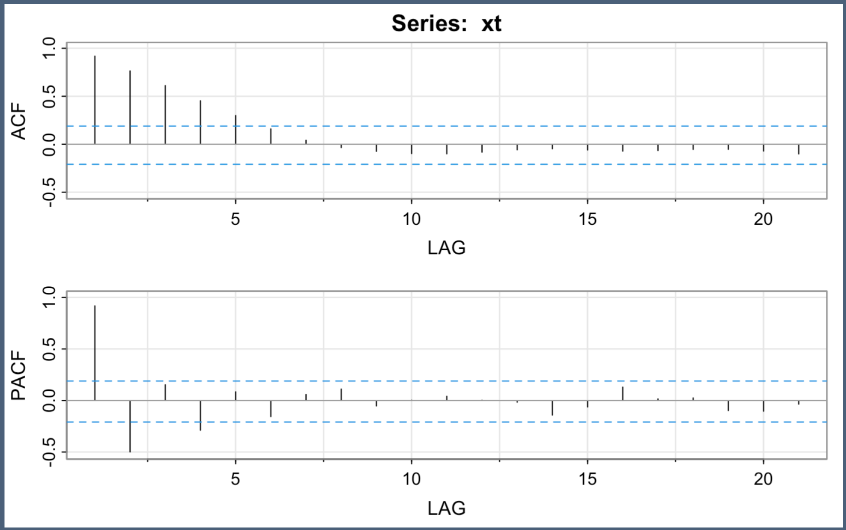
Example ARIMA(0,1,1)

- IMA(1,1) process
  - Model is

$$\nabla X_t = 0.6\, w_{t-1} + w_t$$

- Observed time series
  - Sequence plot looks stationary
  - ACF/PACF both appear to decay geometrically



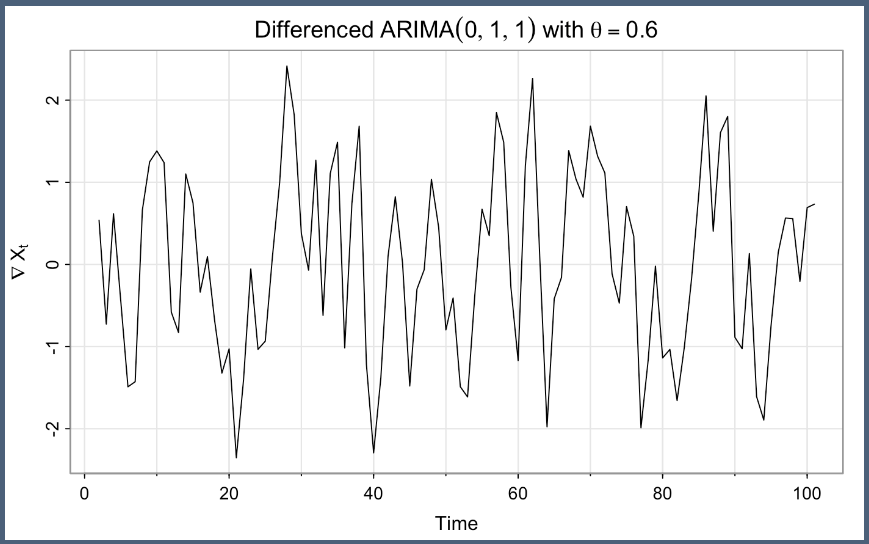


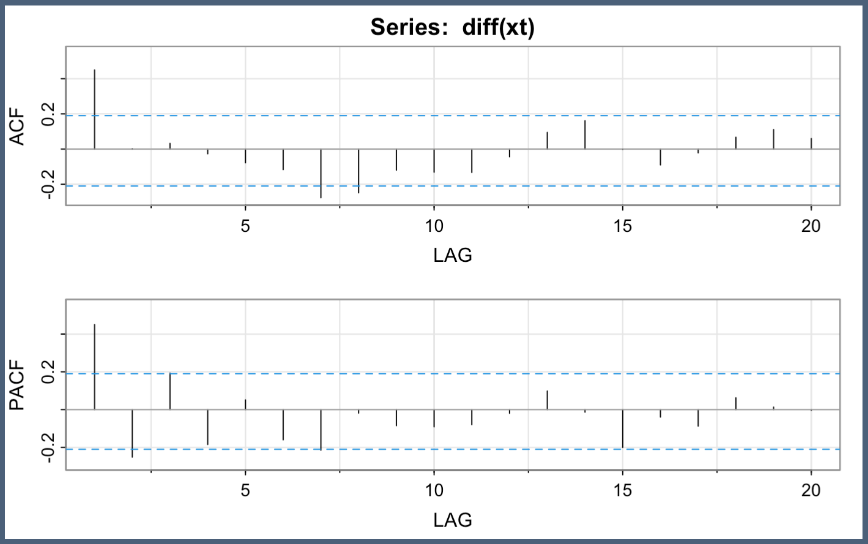
Example ARIMA(0,1,1)

- IMA(1,1) process
  - Model is

$$\nabla X_t = 0.6\, \nabla X_{t-1} + w_t$$

- Observed differences
  - ACF/PACF suggest MA(1) if ignore the ACF after lag 6





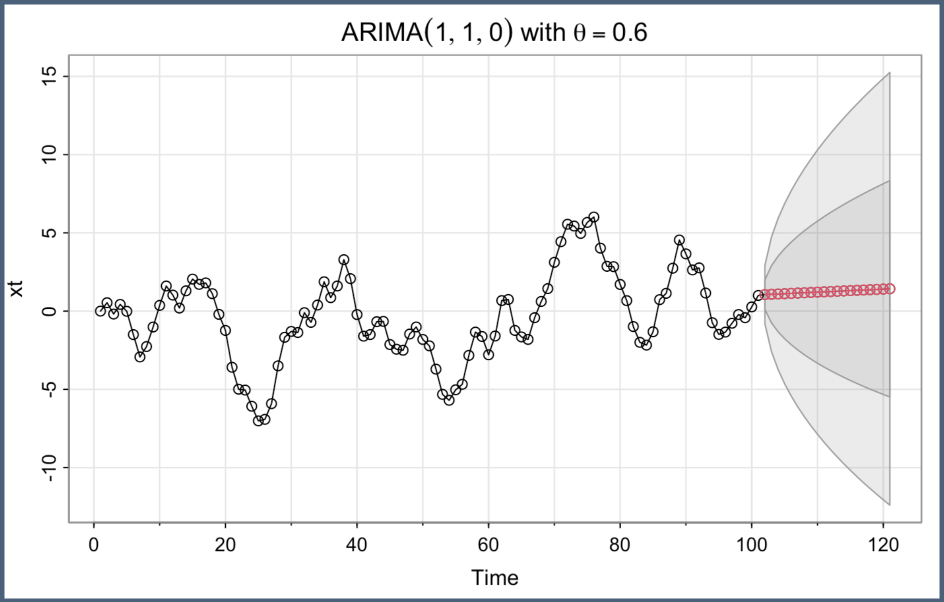
Estimates and Forecasts

- Summary of estimates
  - Small estimate of δ

```
Coefficients:
         Estimate     SE t.value p.value
ma1        0.6810 0.0710  9.5890  0.0000
constant   0.0195 0.1565  0.1248  0.9009

sigma^2 estimated as 0.8741228 on 98 degrees
```

- Forecasts
  - Small estimate of δ leads to very gradual increase
  - Very wide prediction intervals



What's next?

- Modeling examples
  - More model diagnostics (we've seen these in the output)
  - Applying these ideas to real data
- Model testing
  - How do we decide when a model is good enough?
  - How much does it matter