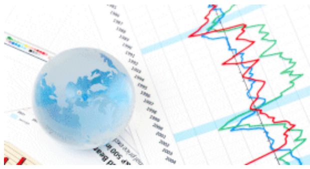


Lecture notes on forecasting

Robert Nau
Fuqua School of Business
Duke University

http://people.duke.edu/~rnau/forecasting.htm

Introduction to ARIMA models

– Nonseasonal

– Seasonal

ARIMA models

- **A**uto-**R**egressive **I**ntegrated **M**oving **A**verage
- Are an adaptation of discrete-time filtering methods developed in 1930's-1940's by electrical engineers (Norbert Wiener et al.)
- Statisticians George Box and Gwilym Jenkins developed systematic methods for applying them to business & economic data in the 1970's (hence the name "Box-Jenkins models")

What ARIMA stands for

- A series which needs to be differenced to be made stationary is an "integrated" (**I**) series
- Lags of the stationarized series are called "auto-regressive" (**AR**) terms
- Lags of the forecast errors are called "moving average" (**MA**) terms
- *We've already studied these time series tools separately*: differencing, moving averages, lagged values of the dependent variable in regression

ARIMA models put it all together

- *Generalized random walk models* fine-tuned to eliminate all residual autocorrelation
- *Generalized exponential smoothing models* that can incorporate long-term trends and seasonality
- *Stationarized regression models* that use lags of the dependent variables and/or lags of the forecast errors as regressors
- The most general class of forecasting models for time series that can be *stationarized** by transformations such as differencing, logging, and or deflating

* A time series is "stationary" if all of its statistical properties—mean, variance, autocorrelations, etc.—are constant in time. Thus, it has no trend, no heteroscedasticity, and a constant degree of "wiggliness."

Construction of an ARIMA model

1. *Stationarize* the series, if necessary, by differencing (& perhaps also logging, deflating, etc.)
2. Study the pattern of *autocorrelations* and *partial autocorrelations* to determine if lags of the stationarized series and/or lags of the forecast errors should be included in the forecasting equation
3. Fit the model that is suggested and check its residual diagnostics, particularly the residual ACF and PACF plots, to see if all coefficients are significant and all of the pattern has been explained.
4. Patterns that remain in the ACF and PACF may suggest the need for additional AR or MA terms

ARIMA terminology

- A non-seasonal ARIMA model can be (almost) completely summarized by three numbers:

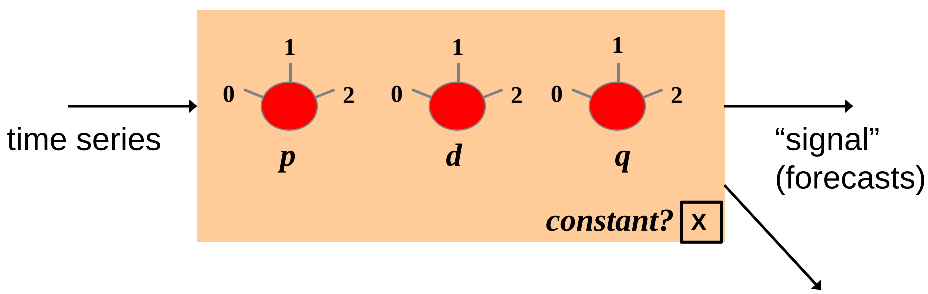
  p **= the number of** ***autoregressive*** **terms**

  d **= the number of** ***nonseasonal differences***

  q **= the number of** ***moving-average*** **terms**

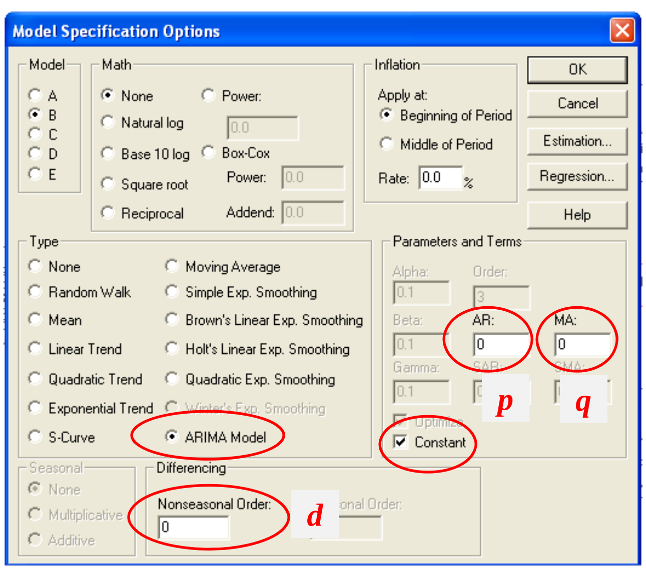
- This is called an "ARIMA(*p,d,q*)" model
- The model may also include a *constant* term (or not)

The ARIMA "filtering box"



Objective: adjust the knobs until the residuals are "white noise" (uncorrelated)

In Statgraphics:



ARIMA options are available when model type = ARIMA

ARIMA models we've already met

- ARIMA(0,0,0)+c = mean (constant) model
- ARIMA(0,1,0) = RW model
- ARIMA(0,1,0)+c = RW with drift model
- ARIMA(1,0,0)+c = regress Y on Y_LAG1
- ARIMA(1,1,0)+c = regr. Y_DIFF1 on Y_DIFF1_LAG1
- ARIMA(2,1,0)+c = " " plus Y_DIFF_LAG2 as well
- ARIMA(0,1,1) = SES model
- ARIMA(0,1,1)+c = SES + constant linear trend
- ARIMA(1,1,2) = LES w/ damped trend (leveling off)
- ARIMA(0,2,2) = generalized LES (including Holt's)

ARIMA forecasting equation

- Let Y denote the *original* series
- Let y denote the *differenced* (stationarized) series

No difference (d=0): $y_t = Y_t$

First difference (d=1): $y_t = Y_t - Y_{t-1}$

Second difference (d=2): $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

$= Y_t - 2Y_{t-1} + Y_{t-2}$

Note that the second difference is not just the change relative to two periods ago, i.e., it is *not* $Y_t - Y_{t-2}$. Rather, it is the change-in-the-change, which is a measure of local "acceleration" rather than trend.

Forecasting equation for *y*

constant AR terms (lagged values of *y*)

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}$$

$$- \theta_1 e_{t-1} \ldots - \theta_q e_{t-q}$$

MA terms (lagged errors)

By convention, the AR terms are + and the MA terms are –

Not as bad as it looks! Usually $p+q \leq 2$ and either $p=0$ or $q=0$ (pure AR or pure MA model)

Undifferencing the forecast

The differencing (if any) must be *reversed* to obtain a forecast for the original series:

$$\text{If } d = 0: \quad \hat{Y}_t = \hat{y}_t$$

$$\text{If } d = 1: \quad \hat{Y}_t = \hat{y}_t + Y_{t-1}$$

$$\text{If } d = 2: \quad \hat{Y}_t = \hat{y}_t + 2Y_{t-1} - Y_{t-2}$$

Fortunately, your software will do all of this automatically!

Do you need both AR and MA terms?

- In general, you *don't*: usually it suffices to use only one type or the other.
- Some series are better fitted by AR terms, others are better fitted by MA terms (at a given level of differencing).
- Rough rules of thumb:
  - If the stationarized series has positive autocorrelation at lag 1, AR terms often work best. If it has negative autocorrelation at lag 1, MA terms often work best.
  - An MA(1) term often works well to *fine-tune* the effect of a nonseasonal difference, while an AR(1) term often works well to compensate for the *lack* of a nonseasonal difference, so the choice between them may depend on whether a difference has been used.

Interpretation of AR terms

- A series displays autoregressive (AR) behavior if it apparently feels a "restoring force" that tends to pull it back toward its mean.
  - In an AR(1) model, the AR(1) coefficient determines how fast the series tends to return to its mean. If the coefficient is near zero, the series returns to its mean quickly; if the coefficient is near 1, the series returns to its mean slowly.
  - In a model with 2 or more AR coefficients, the sum of the coefficients determines the speed of mean reversion, and the series may also show an oscillatory pattern.

Interpretation of MA terms

- A series displays moving-average (MA) behavior if it apparently undergoes random "shocks" whose effects are felt in two or more consecutive periods.
  - The MA(1) coefficient is (minus) the fraction of last period's shock that is still felt in the current period.
  - The MA(2) coefficient, if any, is (minus) the fraction of the shock two periods ago that is still felt in the current period, and so on.

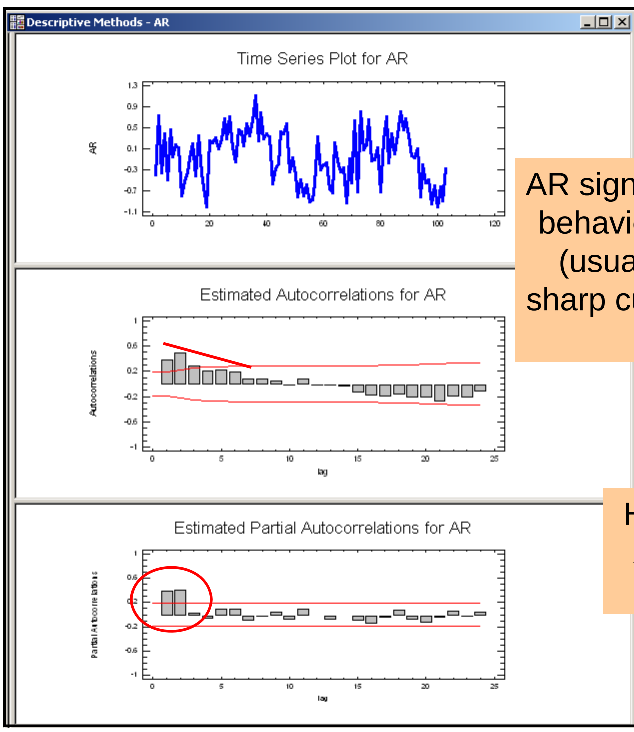
Tools for identifying ARIMA models: ACF and PACF plots

- The autocorrelation function (ACF) plot shows the *correlation* of the series with *itself* at different lags
  - The autocorrelation of *Y* at lag *k* is the correlation between *Y* and LAG(*Y*,*k*)
- The *partial* autocorrelation function (PACF) plot shows the amount of autocorrelation at lag *k* that is *not explained by lower-order autocorrelations*
  - The partial autocorrelation at lag *k* is the coefficient of LAG(*Y*,*k*) in an AR(*k*) model, i.e., in a regression of *Y* on LAG(*Y*, 1), LAG(*Y*,2), … up to LAG(*Y*,*k*)

AR and MA "signatures"

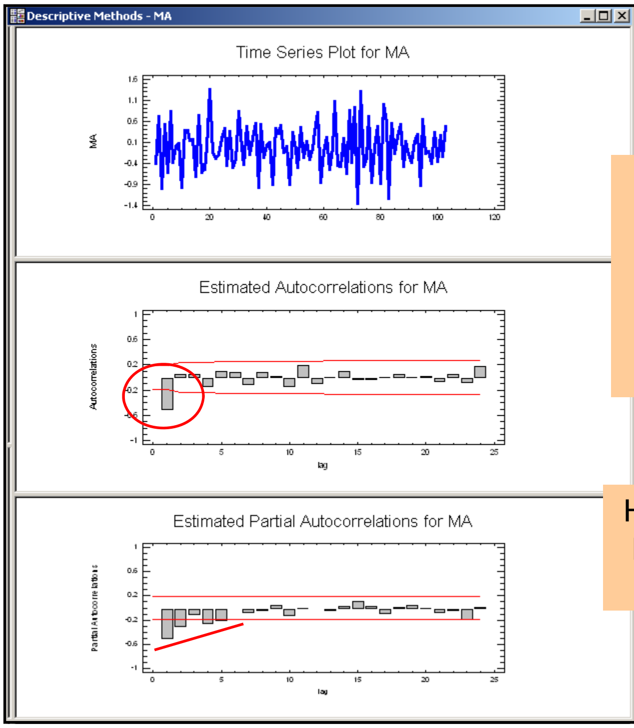
- ACF that dies out gradually and PACF that cuts off sharply after a few lags ⇒ **AR signature**
  - An AR series is usually *positively autocorrelated at lag 1* (or even borderline nonstationary)

- ACF that cuts off sharply after a few lags and PACF that dies out more gradually ⇒ **MA signature**
  - An MA series is usually *negatively autcorrelated at lag 1* (or even mildly overdifferenced)



AR signature: mean-reverting behavior, slow decay in ACF (usually positive at lag 1), sharp cutoff after a few lags in PACF.

Here the signature is AR(2) because of 2 spikes in PACF.



MA signature: noisy pattern, sharp cutoff in ACF (usually negative at lag 1), gradual decay in PACF.

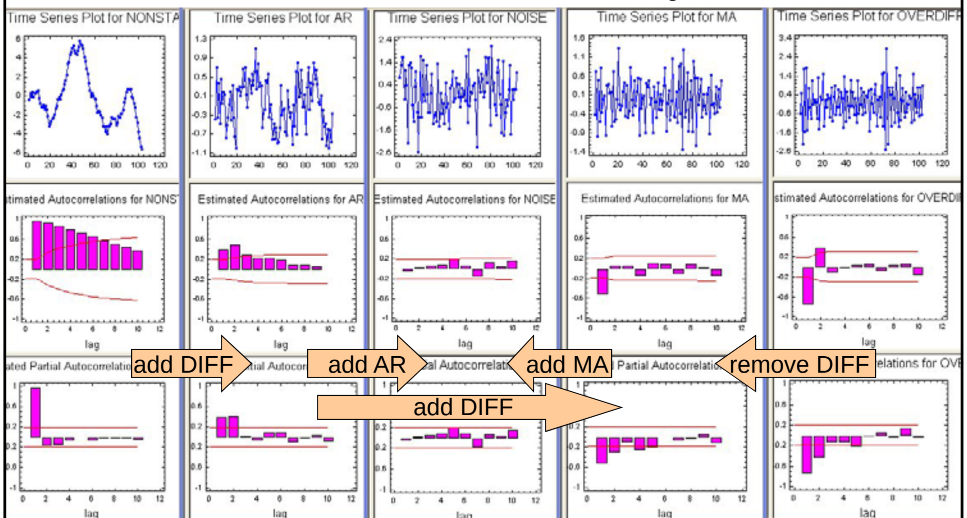
Here the signature is MA(1) because of 1 spike in ACF.

AR or MA? It depends!

- Whether a series displays AR or MA behavior often *depends on the extent to which it has been differenced*.
- An “underdifferenced” series has an AR signature (positive autocorrelation)
- After one or more orders of differencing, the autocorrelation will become more negative and an MA signature will emerge
- Don’t go too far: if series already has zero or negative autocorrelation at lag 1, don’t difference again

The autocorrelation spectrum

←Positive autocorrelation | No autocorrelation | Negative autocorrelation→



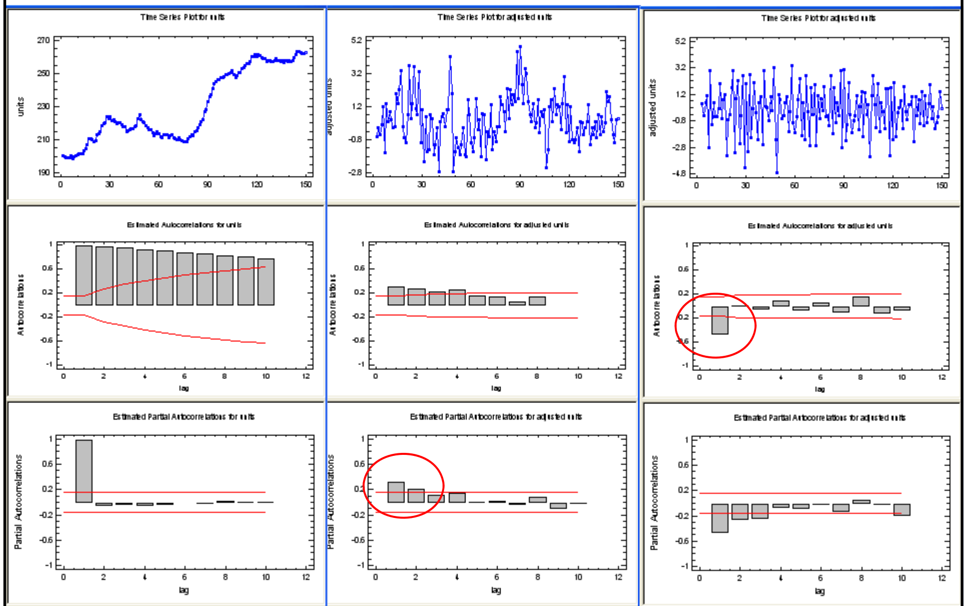
Nonstationary | Auto-Regressive | White Noise | Moving-Average | Overdifferenced

Model-fitting steps

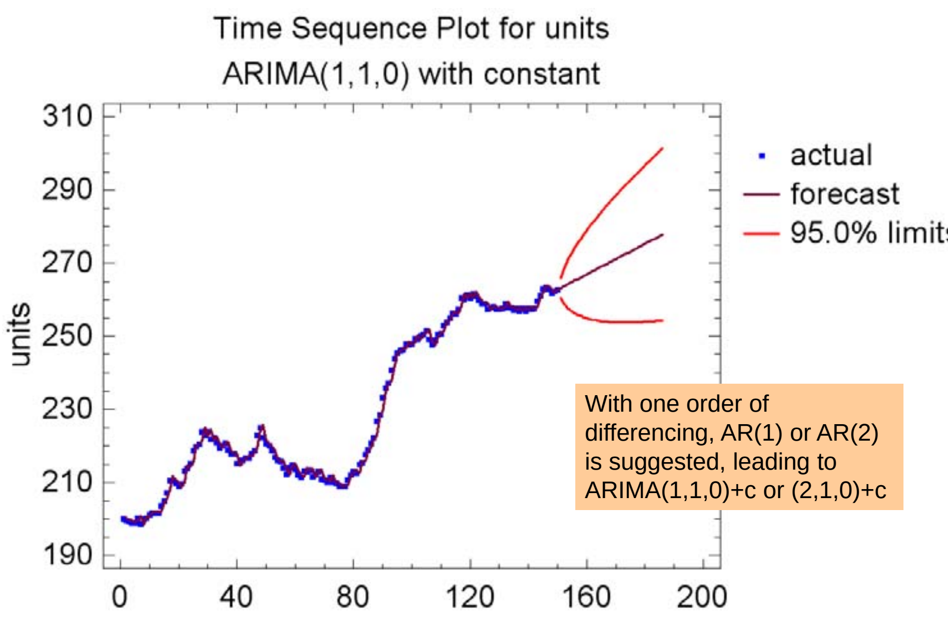
1. Determine the order of differencing
2. Determine the numbers of AR & MA terms
3. Fit the model—check to see if residuals are "white noise," highest-order coefficients are significant (w/ no "unit "roots"), and forecasts look reasonable. If not, return to step 1 or 2.

In other words, move right or left in the "autocorrelation spectrum" by appropriate choices of differencing and AR/MA terms, until you reach the center (white noise)

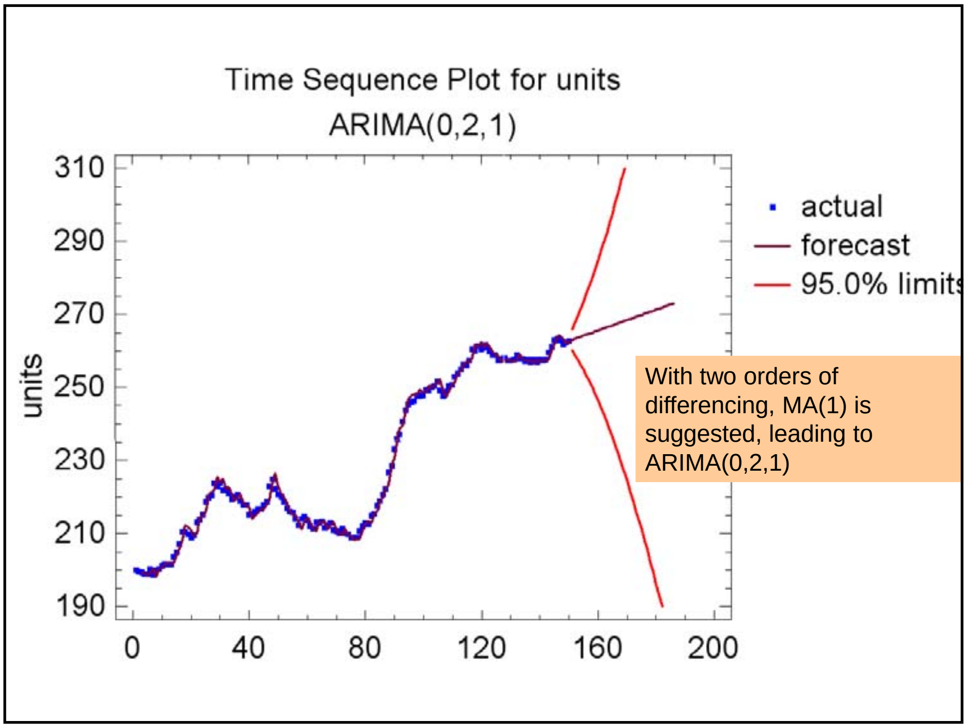
“Units” example



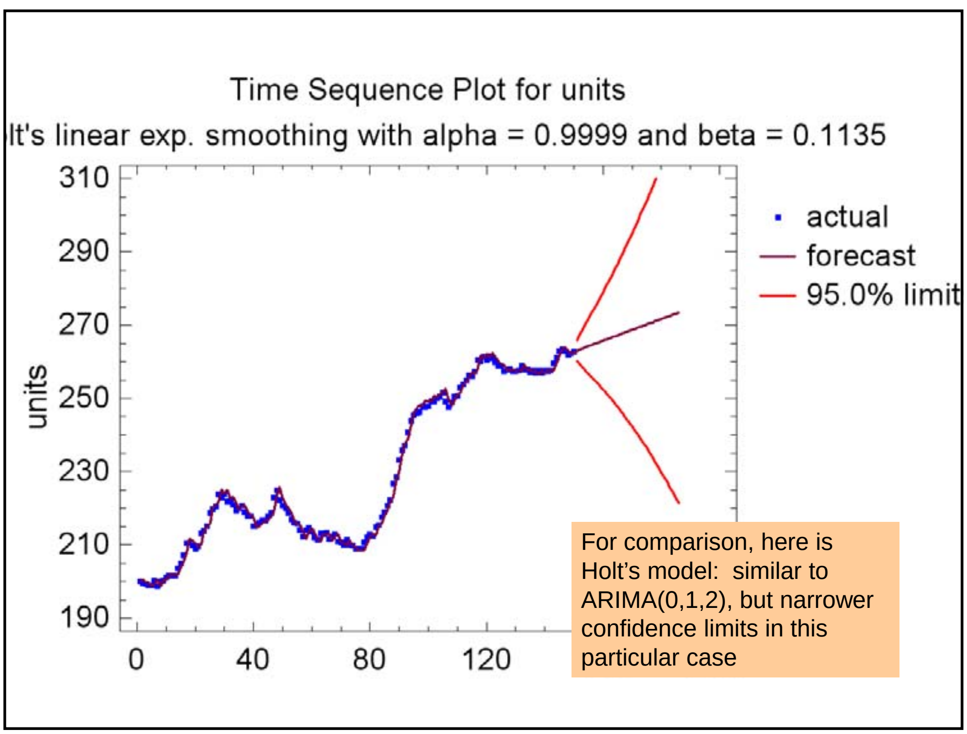
Original series: nonstationary 1st difference: AR signature 2nd difference: MA signature



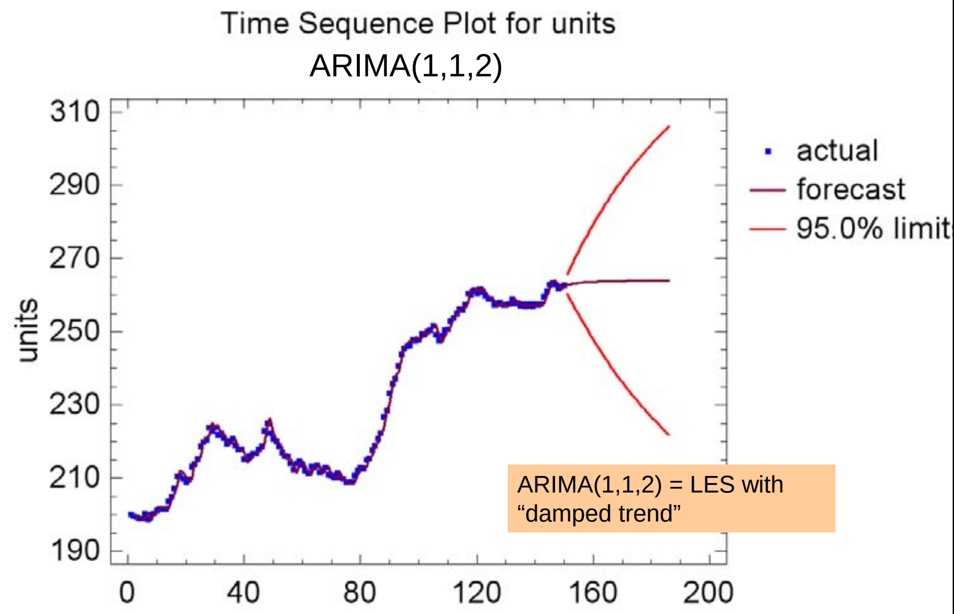
With one order of differencing, AR(1) or AR(2) is suggested, leading to ARIMA(1,1,0)+c or (2,1,0)+c



With two orders of differencing, MA(1) is suggested, leading to ARIMA(0,2,1)



For comparison, here is Holt’s model: similar to ARIMA(0,1,2), but narrower confidence limits in this particular case



Model Comparison
Data variable: units
Number of observations = 150
Start index = 1.0
Sampling interval = 1.0

Models
(A) ARIMA(1,1,0) with constant
(B) ARIMA(0,2,1)
(C) ARIMA(1,1,2)
(D) Simple exponential smoothing with alpha = 0.9999
(E) Holt's linear exp. smoothing with alpha = 0.9999 and beta = 0.1135

Estimation Period

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*
(A)	1.37619	1.05058	0.462858	0.00208321	-0.0011386
(B)	1.36987	1.07665	0.473588	0.0133783	0.0105393
(C)	1.34551	1.04936	0.462074	0.143321	0.0639647
(D)	1.49927	1.15338	0.507076	0.417375	0.17929
(E)	1.39	1.07169	0.471833	0.000867136	0.00544249

Model	*RMSE*	*RUNS*	*RUNM*	*AUTO*	*MEAN*	*VAR*
(A)	1.37619	*	OK	OK	OK	OK
(B)	1.36987	OK	OK	OK	OK	*
(C)	1.34551	OK	OK	OK	OK	*
(D)	1.49927	OK	*	***	**	OK
(E)	1.39	OK	*	OK	OK	OK

All models that involve at least one order of differencing (a trend factor of some kind) are better than SES (which assumes no trend). ARIMA(1,1,2) is the winner over the others by a small margin.

Technical issues

- Backforecasting
  - Estimation algorithm begins by forecasting backward into the past to get start-up values
- Unit roots
  - Look at sum of AR coefficients and sum of MA coefficients—if they are too close to 1 you may want to consider higher or lower of differencing
- Overdifferencing
  - A series that has been differenced one too many times will show *very* strong negative autocorrelation and a strong MA signature, probably with a unit root in MA coefficients

Seasonal ARIMA models

- We've previously studied three methods for modeling seasonality:
  - Seasonal adjustment
  - Seasonal dummy variables
  - Seasonally lagged dependent variable in regression
- A 4th approach is to use a *seasonal ARIMA model*
  - Seasonal ARIMA models rely on *seasonal lags and differences* to fit the seasonal pattern
  - Generalizes the regression approach

Seasonal ARIMA terminology

- The seasonal part of an ARIMA model is summarized by three *additional* numbers:

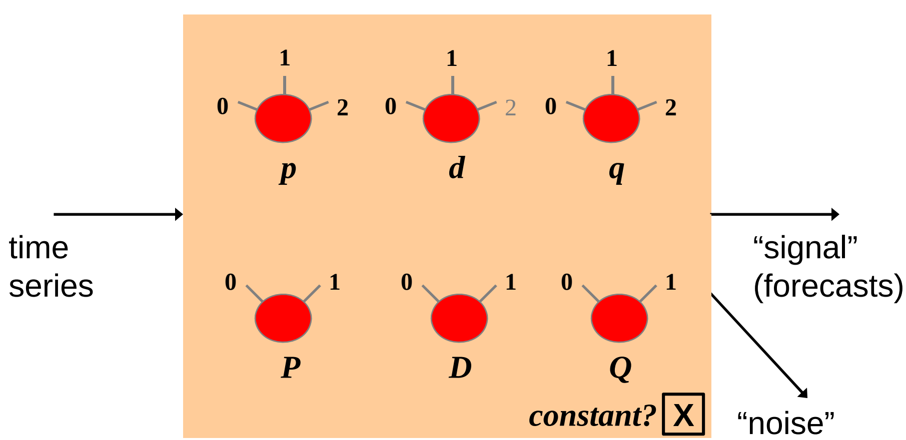
  P **= # of *seasonal autoregressive* terms**

  D **= # of *seasonal differences***

  Q **= # of *seasonal moving-average* terms**

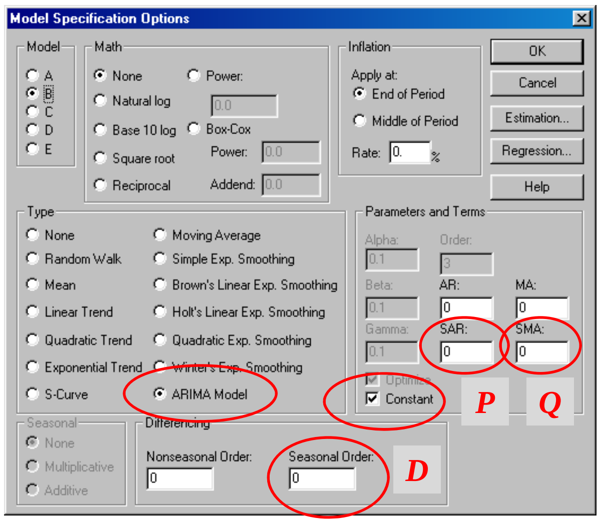
- The complete model is called an "ARIMA(*p,d,q*)×(*P,D,Q*)" model

The "filtering box" now has 6 knobs:



Note that *P*, *D*, and *Q* should never be larger than 1 !!

In Statgraphics:



Seasonal ARIMA options are available when model type = ARIMA and a number has been specified for "seasonality" on the data input panel.

Seasonal differences

How non-seasonal & seasonal differences are combined to stationarize the series:

If d=0, D=1: $y_t = Y_t - Y_{t-s}$

s is the seasonal period, e.g., *s*=12 for monthly data

If d=1, D=1: $y_t = (Y_t - Y_{t-1}) - (Y_{t-s} - Y_{t-s-1})$

$$= Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$$

D should never be more than 1, and *d*+*D* should never be more than 2. Also, if *d*+*D* =2, the constant term should be suppressed.

SAR and SMA terms

How SAR and SMA terms add coefficients to the model:

- Setting *P* =1 (i.e., SAR=1) adds a multiple of $y_{t\text{-}s}$ to the forecast for y_t
- Setting *Q* =1 (i.e., SMA=1) adds a multiple of $e_{t\text{-}s}$ to the forecast for y_t

Total number of SAR and SMA factors usually should not be more than 1 (i.e., either SAR=1 or SMA=1, not both)

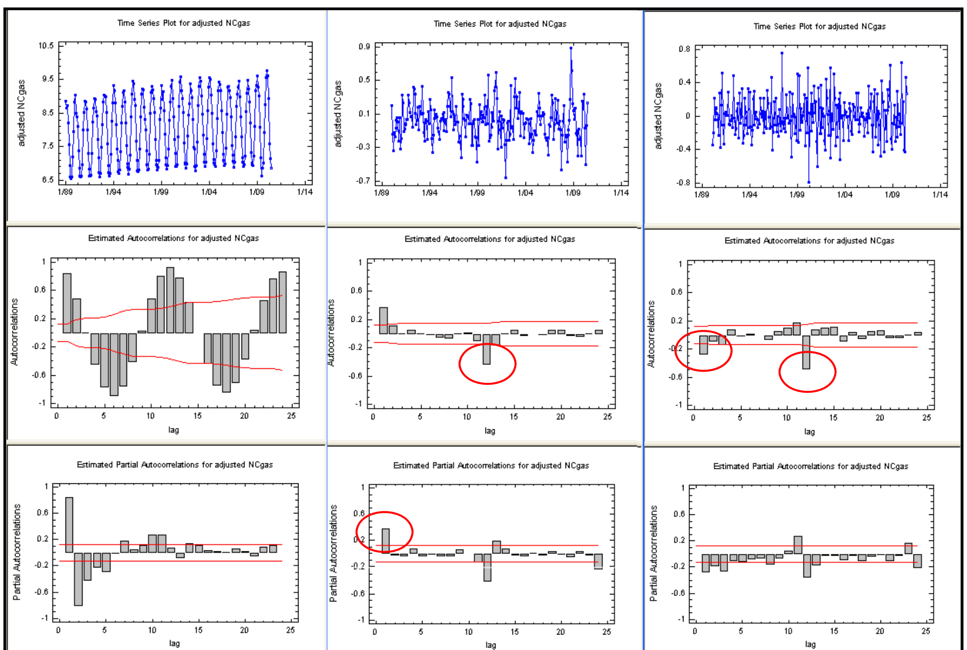
Model-fitting steps

- Start by trying various combinations of one seasonal difference and/or one non-seasonal difference to stationarize the series and remove gross features of seasonal pattern.
- If the seasonal pattern is strong and stable, you *MUST* use a seasonal difference (otherwise it will “die out” in long-term forecasts)

Model-fitting steps, continued

- After differencing, inspect the ACF and PACF at multiples of the seasonal period (*s*):
  - *Positive* spikes in ACF at lag s, 2s, 3s…, single positive spike in PACF at lag s ⇒ SAR=1
  - *Negative* spike in ACF at lag s, negative spikes in PACF at lags s, 2s, 3s,… ⇒ SMA=1
  - SMA=1 often works well in conjunction with a seasonal difference.

Same principles as for non-seasonal models, except focused on what happens *at multiples of lag s* in ACF and PACF.



Original series: nonstationary | Seas. diff: need AR(1) & SMA(1) | Both diff: need MA(1) & SMA(1)

A common seasonal ARIMA model

- Often you find that the “correct” order of differencing is *d*=1 and *D*=1.
- With one difference of each type, the autocorr. is often *negative* at *both* lag 1 and lag *s.*
- This suggests an ARIMA(0,1,1)×(0,1,1) model, a common seasonal ARIMA model.
- Similar to Winters’ model in estimating *time-varying trend and time-varying seasonal pattern*

Another common seasonal ARIMA model

- Often with *D*=1 (only) you see a borderline nonstationary pattern with *A*R(*p*) signature, where *p*=1 or 2, sometimes 3
- After adding AR=1, 2, or 3, you may find negative autocorrelation at lag *s* (⇒ SMA=1)
- This suggests ARIMA(*p,*0,0)×(0,1,1)+c, another common seasonal ARIMA model.
- Key difference from previous model: assumes a *constant* annual trend

Bottom-line suggestion

- When fitting a time series with a strong seasonal pattern, you generally should try

  ARIMA(0,1,q)×(0,1,1) **model (*q*=1 or 2)**

  ARIMA(p,0,0)×(0,1,1)+c **model (*p*=1, 2 or 3)**

  … in addition to other models (e.g., RW, SES or LES with seasonal adjustment; or Winters)

- If there is a significant trend and/or the seasonal pattern is multiplicative, you should also try a *natural log* transformation.

Take-aways

- Seasonal ARIMA models (especially the (0,1,*q*)x(0,1,1) and (*p*,0,0)x(0,1,1)+c models) compare favorably with other seasonal models and often yield better short-term forecasts.

- Advantages: solid underlying theory, stable estimation of time-varying trends and seasonal patterns, relatively few parameters.

- Drawbacks: no explicit seasonal indices, hard to interpret coefficients or explain "how the model works", danger of overfitting or mis-identification if not used with care.