RegARIMA Modeling: Transformations & ARIMA Models

Seasonal Adjustment With X-13ARIMA-SEATS

2019

Economic Statistical Methods Division

U.S. Census Bureau

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Objectives

- **At the end of this unit, you should understand –**
  - The basic idea and the notation of regARIMA modeling
  - The types of transformations, regressors, and ARIMA models available from X-13ARIMA-SEATS

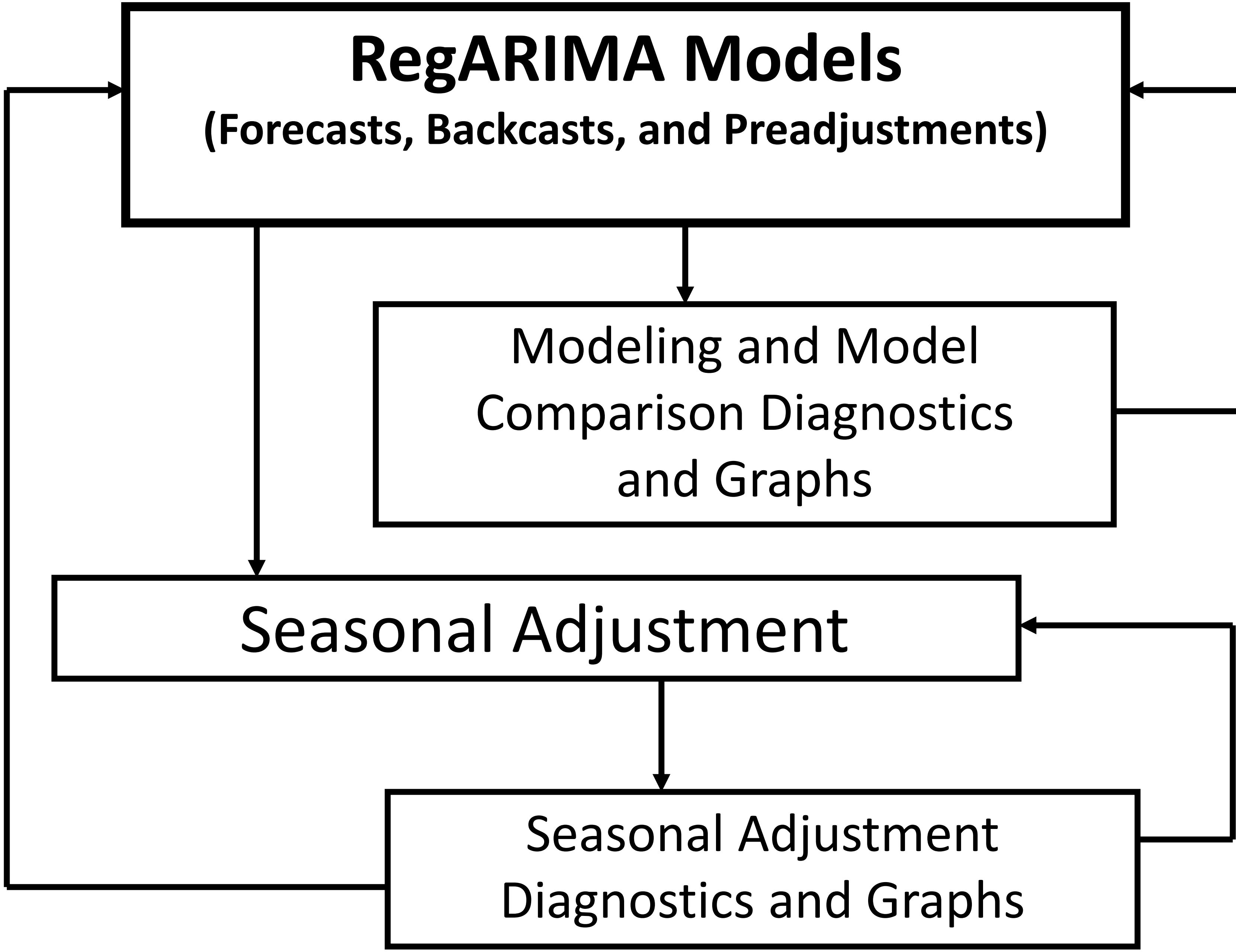
Outline

- Overview
- Transformations & Adjustments
- ARIMA Processes
- Regression Effects

Learning RegARIMA Modeling

- *Basic* overview here

- Bill Bell (ADRM) – Chapter 4 of the X-13A-S manual and previous course notes (1992, 1998 updates)


RegARIMA Models
(Forecasts, Backcasts, and Preadjustments)
Modeling and Model Comparison Diagnostics and Graphs
Seasonal Adjustment
Seasonal Adjustment Diagnostics and Graphs

Regression Model Review

$$\log(\ Y\ ) = \beta' X + \varepsilon$$

Transformation(s)

Regression

error

X = Regressors

The traditional regression model assumes that data are independent, identically distributed; but time series data are autocorrelated.

Autocorrelation
(Statistical Dependence)

- Cannot rely on statistical theory that assumes independence and identical distribution
- Can take advantage of the autocorrelation to fit models, identify patterns, **forecast future values**

Modeling Time Series or How to Use Autocorrelation

RegARIMA =

Regression + ARIMA

ARIMA = Autoregressive Integrated Moving Average

RegARIMA Model

$$\log\left(Y_t / D_t \right) = \beta' X_t + Z_t$$

Transformation(s)

Regression

ARIMA Process

$X_t =$ Regressor for trading day and holiday or calendar effects, additive outliers, temporary changes, level shifts, ramps, user-defined effects

$D_t =$ Leap-year adjustment, or "subjective" strike adjustment, etc.

RegARIMA Model (No log transform)

$$Y_t - D_t = \beta' X_t + Z_t$$

Regression ARIMA Process

$X_t =$ Regressor for trading day and holiday or calendar effects, additive outliers, temporary changes, level shifts, ramps, user-defined effects

$D_t =$ "Subjective" strike adjustment, etc.

RegARIMA Model Uses

- Extend the series with forecasts
  - Beneficial for seasonal adjustment
- Detect and directly estimate trading day effects and other effects (e.g. moving holiday effects, user-defined effects)
- Detect and adjust for outliers to improve the forecasts and seasonal adjustments
- (Estimate missing values)

Some important time series vocabulary/concepts

Sample Mean/Variance/ Autocovariance

For a discrete time series $y_1, y_2, \ldots, y_n$

Mean: $\hat{\mu}_t = \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$

Variance: $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$

Autocovariance at lag k: $\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$

Note that $\hat{\gamma}_0 = \hat{\sigma}^2$

Sample Autocorrelation

The autocorrelation at lag *k* is

$$\hat{\rho}_k = {}^{\hat{\gamma}_k}\!/_{\hat{\gamma}_0}$$

The sequence $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \ldots$ is the sample autocorrelation function, or ACF.

Partial Autocorrelation Function (PACF)

- The partial autocorrelation at lag k, φ_{kk}, is the correlation between y_t and y_{t+k} with the effects of the in-between observations (y_{t+1}, …, y_{t+k-1}) removed.

- These can be found by solving the Yule-Walker equations

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

Stationarity
This Is a Key Concept

- Many time series statistical procedures rely on ***stationarity*** of the series
- A time series y_t is stationary if the following are constant over time t
  - Mean: $E(y_t) \equiv \mu_t = \mu$
  - Variance: $Var(y_t) \equiv \sigma^2_t = \sigma^2$
  - Covariance of any two points the same distance apart – in other words, autocovariance for any lag k:

    $Cov(y_t, y_{t+k}) = \gamma_k$ for all t

    note: $\gamma_0 = Cov(y_t, y_t) = \sigma^2$

Results of Stationarity

- Stationarity implies that the autocorrelation at any lag k is constant over time t

  $Corr(y_t, y_{t+k}) = \rho_k$ for all t

  note: $\rho_0 = 1$

- Stationarity implies that the autocorrelation decays to 0 with increasing lag

  $\rho_k \to 0$ as $k \to \infty$

Transformations and Adjustments

$$\log(\, Y_t \,/\, D_t \,) = \beta' X_t + Z_t$$

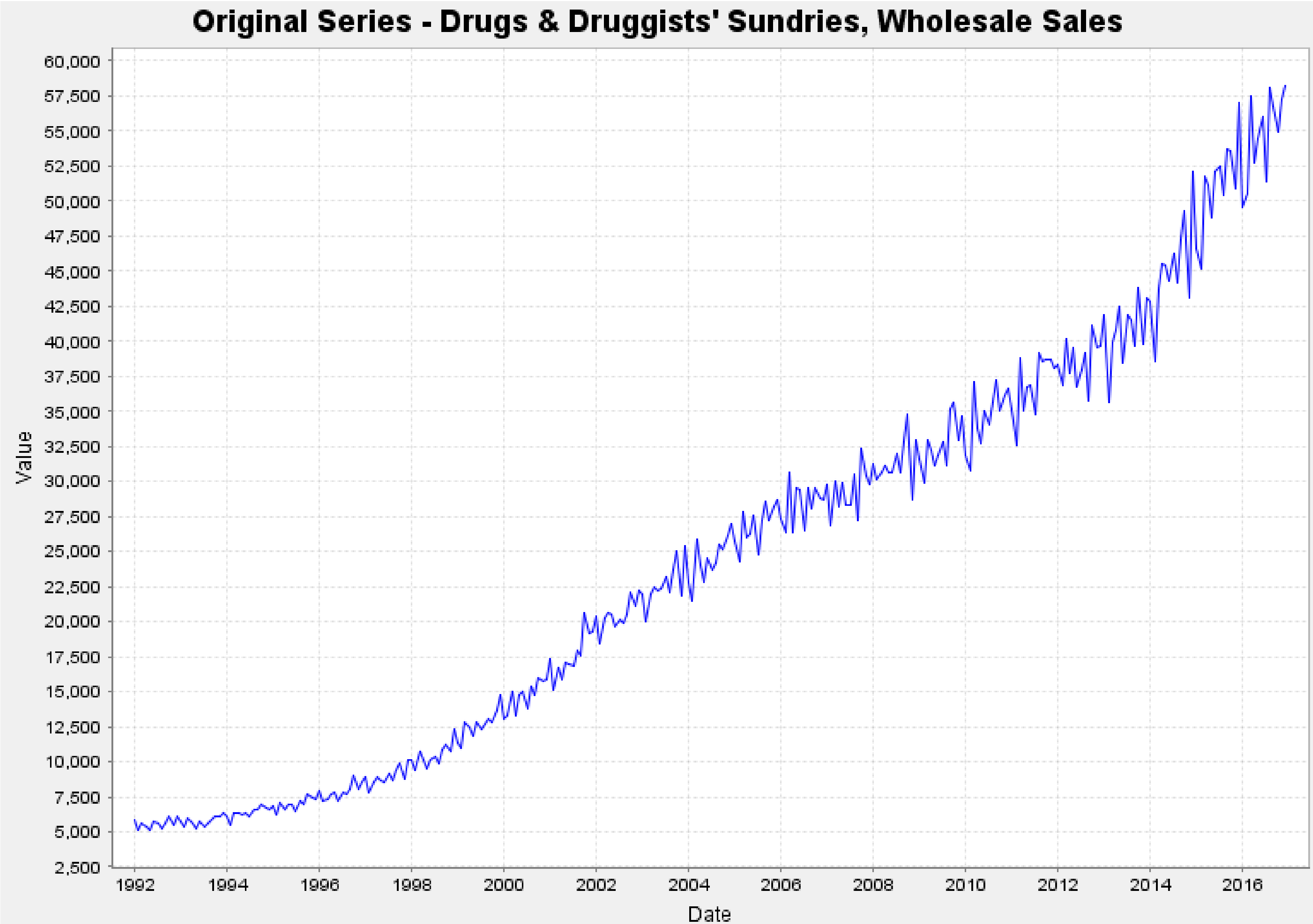
Transformation(s)

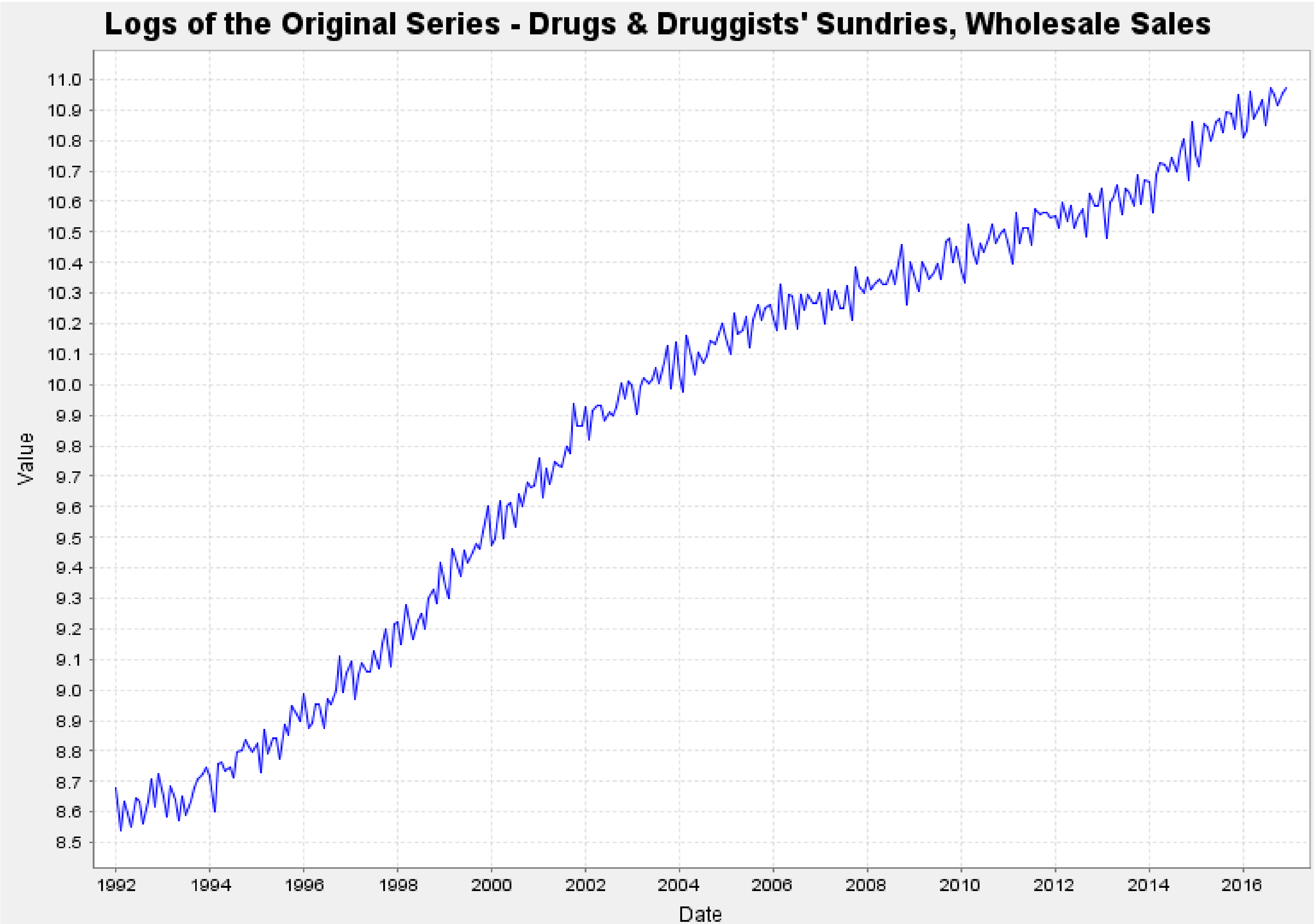
ARIMA Process

Regression

Transformations & Stationarity

- When the variance of a series increases as the level increases, the log transformed series might have constant variance


Original Series - Drugs & Druggists' Sundries, Wholesale Sales
Value
60,000
57,500
55,000
52,500
50,000
47,500
45,000
42,500
40,000
37,500
35,000
32,500
30,000
27,500
25,000
22,500
20,000
17,500
15,000
12,500
10,000
7,500
5,000
2,500
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Date


Logs of the Original Series - Drugs & Druggists' Sundries, Wholesale Sales
Value
11.0
10.9
10.8
10.7
10.6
10.5
10.4
10.3
10.2
10.1
10.0
9.9
9.8
9.7
9.6
9.5
9.4
9.3
9.2
9.1
9.0
8.9
8.8
8.7
8.6
8.5
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Date

`Transform` Spec

- Specifies Box-Cox power transformations for time series modeling
- If performing seasonal adjustment, only log transformation or no transformation is allowed
  - Automatic selection option using likelihood test

`transform{} function`

- **`function = auto  savelog = atr`**
  - Tests for log transform vs. no transform
- **`function = log`**
  - Multiplicative adjustment: A = Y / S
- **`function = none`**
  - Additive adjustment: A = Y – S

- Test first then set **`function`** to **`log`** or **`none`**

AIC Test for Log Transformation

- **`function = auto`** calculates the AICC of the series with a log transformation and without
- By default, there is a slight preference for log transformation built in to the test
  - If AICC(no log) – AICC(log) < -2 select no transformation; else select log transformation
  - This is controlled with the **`aicdiff`** argument; by default, **`aicdiff = -2.0`**

AIC

- Standard model comparison diagnostic
  - Based on Akaike's Information Criterion
- **Prefer minimum value**
  - "MAIC" = minimum AIC

"Smaller" value?

Not magnitude, just minimum

–400 < –300, so prefer AIC of –400 over –300

AIC Formula

Akaike's Information Criterion

$$AIC_N = -2\hat{L}_N + 2\mathrm{p}$$

where

$\hat{L}_N$ is the maximized value of the log-likelihood function from *N* observations, and

p is the number of estimated parameters

AICC Formula

AIC Corrected (for sample size)

$$AICC_N = -2\hat{L}_N + \frac{2p}{1 - \frac{p+1}{N}}$$

Note: As N (number of observations) gets larger, AICC approaches the AIC

AIC Test for Transformation

From the Drugs & Druggists' Sundries output file:

Likelihood statistics for model fit to untransformed series.

Likelihood Statistics

AICC (F-corrected-AIC)	4954.4937

Likelihood statistics for model fit to log transformed series.

Likelihood Statistics

AICC (F-corrected-AIC)	4755.2045

***** AICC (with aicdiff=-2.00) prefers **log transformation** *****

***** Multiplicative seasonal adjustment will be performed. ****

Transformation Changes and Revisions

- A change in the transformation leads to a change in the seasonal adjustment decomposition method
- This will result in changes to the seasonal adjustment throughout the series

Alternate transformations

- Other transformations can be used but ONLY for modeling, not for seasonal adjustment:
  - **`function = sqrt | inverse | logistic`**
  - **`power =`** λ specifies a Box-Cox power transformation

Other Uses of the `transform` Spec

- The series can be adjusted by prior factors.
  - Some are built in to the program (leap year and length of month/quarter) but ONLY for a log-transformed series.
  - Users can specify prior factors to be divided from (in the case of a log-transformed series) or subtracted from (in the case of an untransformed series) the series.

Transform Spec Adjustments

- Length-of-month/quarter adjustments: series is divided by (# days in month/quarter)/(average # days in all months/quarters)

```
adjust = lom | loq
```

- Leap-year adjustments: series Februaries/Q1 are divided by (# days in February)/(average # days in February/Q1)

```
adjust = lpyear
```

A 2 Prior-adjustment factors

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	AVGE
1992	100.0	102.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.2
1993	100.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
1994	100.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
1995	100.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
1996	100.0	102.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.2
1997	100.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
1998	100.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9

User Prior-Adjustment Factors

- Prior adjustments not built into X-13A-S can be specified with an external file of factors
  - Survey of Construction (Housing Starts, etc.) prior adjustment to remove boost discontinued in 1999
  - Building Permits prior adjustment for universe size (no longer used)

- Use `file` or `data` argument in the `transform` spec to input the data, along with any necessary `format` or `start` arguments.

Transformation Factor Types

- **`type = permanent`**
  - Excluded from final seasonally adjusted series
    - Possible example, deflator
- **`type = temporary`**
  - Included in final seasonally adjusted series
    - Possible example, boost factor that accounts for universe change

- Can specify one of each type of factor

Transformation Factor Modes

- **`mode = ratio`**
  - Y_t / D_t $D_t = 1.50$
- **`mode = percent`**
  - Default if log transformation
  - $Y_t / (D_t / 100)$ $D_t = 150.00$
- **`mode = diff`**
  - For additive series (no log transformation)
  - $Y_t – D_t$

Transform Spec Syntax

```
transform {
 function = auto | log | none
 adjust = lom | loq | lpyear
 file = "MyTransformData.dat"
 format = "datevalue"|"x12save"|"free“
 # with data argument or free format:
 # start = yyyy.mm
```

`Transform` Spec Syntax (continued)

```
…
  type = temporary | permanent
  #   [ default: permanent ]
  mode = percent | ratio | diff
  #   [ default: percent ]

  print = See Manual/Quick Reference
  save = See Manual/Quick Reference
  savelog = autotransform
}
```

Differencing

- First difference – subtracting the previous value from the current value

$$z_t - z_{t\text{-}1}$$

- Seasonal difference – subtracting the previous year's value from the current value

$$z_t - z_{t\text{-}12} \qquad \text{(monthly series)}$$

$$z_t - z_{t\text{-}4} \qquad \text{(quarterly series)}$$

- Differencing reduces the number of values available for estimating the model coefficients

Date	Value	1^{st} Diff	4^{th} Diff	1^{st} & 4^{th}
2003.4	242139			
2004.1	228132	-14007		
2004.2	237880	9748		
2004.3	237709	-171		
2004.4	251362	13653	9223	
2005.1	240957	-10405	12825	3602
2005.2	249386	8429	11506	-1319
2005.3	250421	1035	12712	1206

Differencing Helps Attain Stationarity

- Difference a series to achieve a constant mean
- Typically an economic series needs both a first difference and a seasonal difference

$$z_t - z_{t-1} - (z_{t-4} - z_{t-5}) \qquad \text{(quarterly series)}$$
$$= z_t - z_{t-1} - z_{t-4} + z_{t-5}$$

Differencing Notation

- B is the Backshift operator

$$B z_t = z_{t-1}$$

$$B^4 z_t = z_{t-4}$$

- So a first difference can be written

$$z_t - z_{t-1} = (1 - B)z_t$$

- One first and one quarterly difference:

$$(1 - B)(1 - B^4)z_t = (1 - B - B^4 + B^5)z_t$$

$$= z_t - z_{t-1} - z_{t-4} + z_{t-5}$$

Differencing and the ACF/PACF

- In a nonstationary series, $\hat{\rho}_k$ decreases slowly as k increases
- In a stationary series, $\hat{\rho}_k$ rapidly approaches zero as k increases
- Plots of the ACF can help identify a proper level of differencing

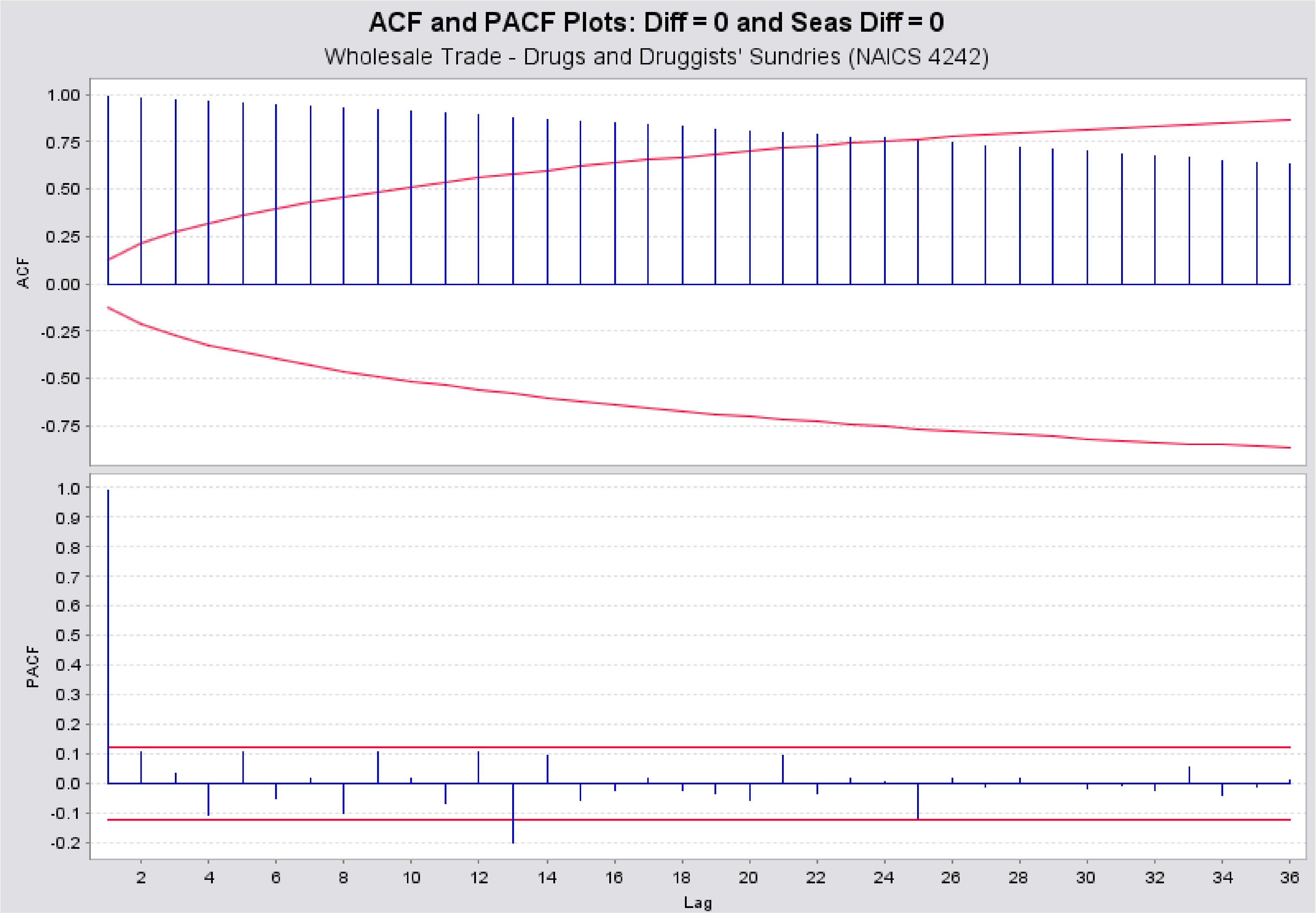
Identify Spec

Used to find the ACF and PACF of a series at the specified levels of differencing.

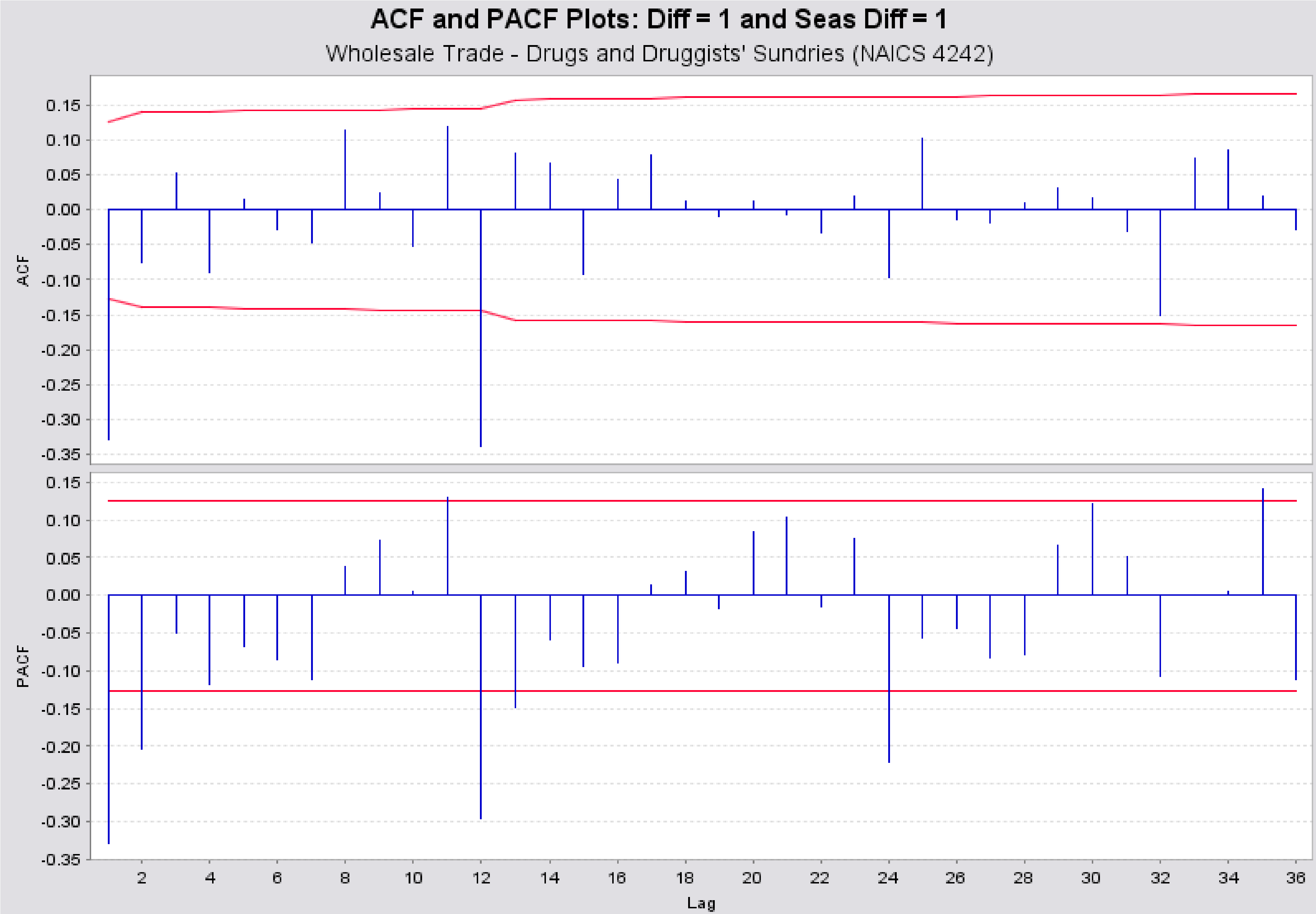
`identify{diff=(0 1) sdiff=(0 1)}`

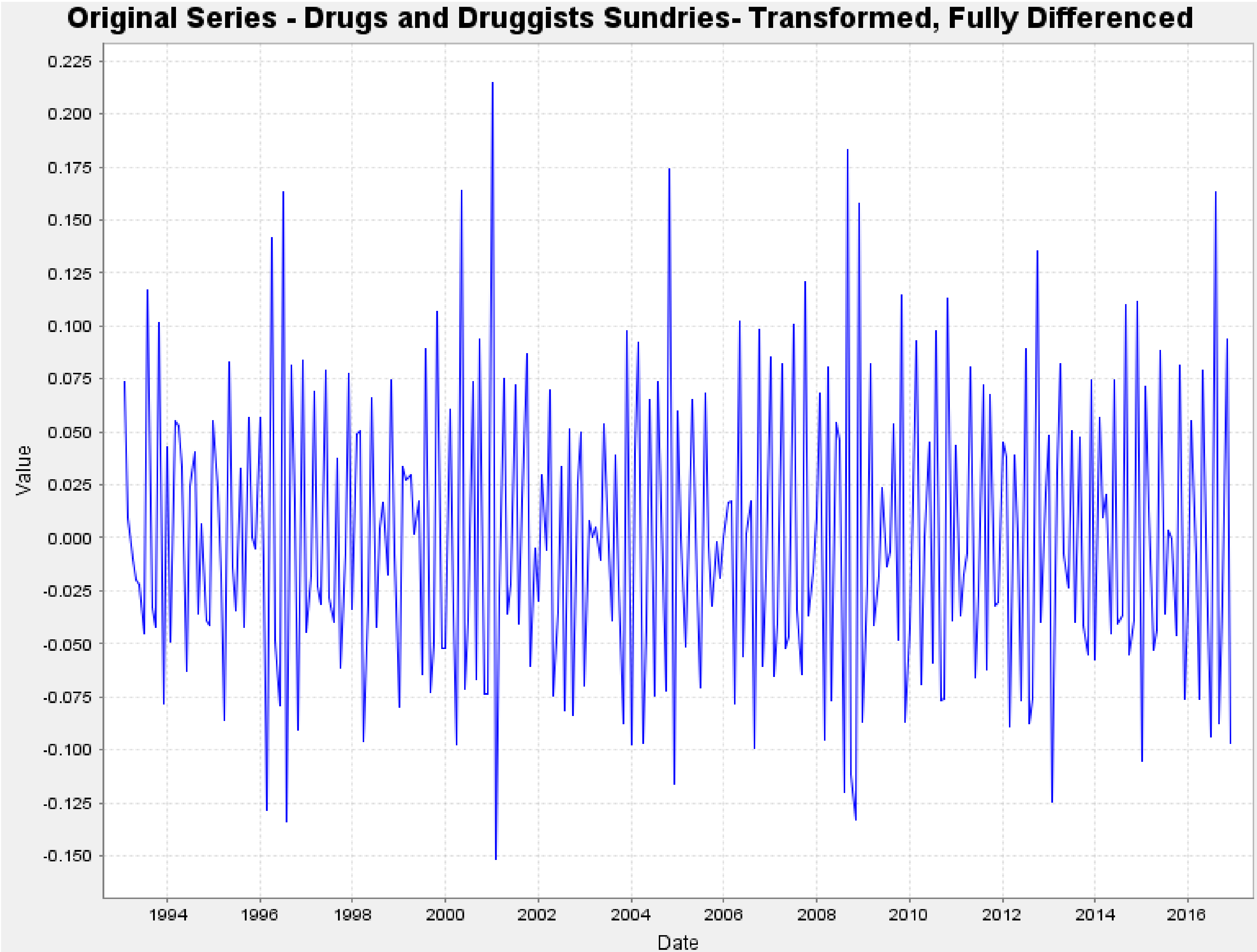
calculates the ACF and PACF with

- no differencing
- one first difference
- one seasonal difference
- one first and one seasonal difference



ACF and PACF Plots: Diff = 0 and Seas Diff = 0
Wholesale Trade - Drugs and Druggists' Sundries (NAICS 4242)
ACF
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
PACF
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
Lag


ACF and PACF Plots: Diff = 1 and Seas Diff = 1
Wholesale Trade - Drugs and Druggists' Sundries (NAICS 4242)
ACF
PACF
Lag
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
-0.35
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36



Original Series - Drugs and Druggists Sundries- Transformed, Fully Differenced
Value
0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
-0.025
-0.050
-0.075
-0.100
-0.125
-0.150
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Date

ARIMA Process

$$\log(Y_t / D_t) = \beta' X_t + \boldsymbol{Z}_t$$

Transformation(s)

Regression

ARIMA Process

- Start with simplest processes and build on those

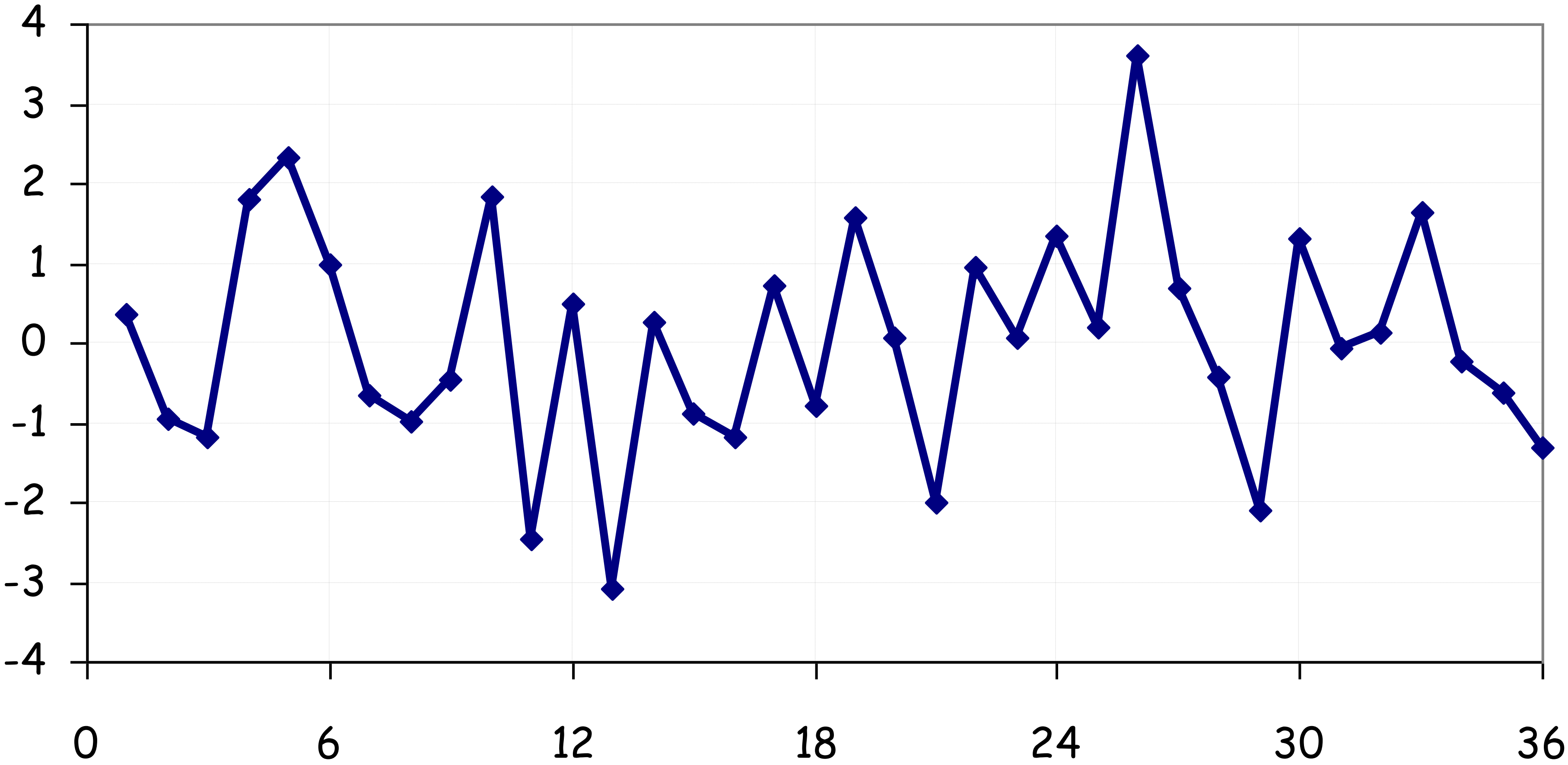
White Noise Process

- Random drawings from a fixed distribution, usually assumed to be Normal with mean 0 and variance σ_a^2.
- Notation: $z_t = a_t$

- What is the best one-step ahead forecast for a white noise series?

(What is the expected value of a_{t+1}?)

White Noise (Mean 0, Variance 2)



Simple Stochastic Process – Random Walk

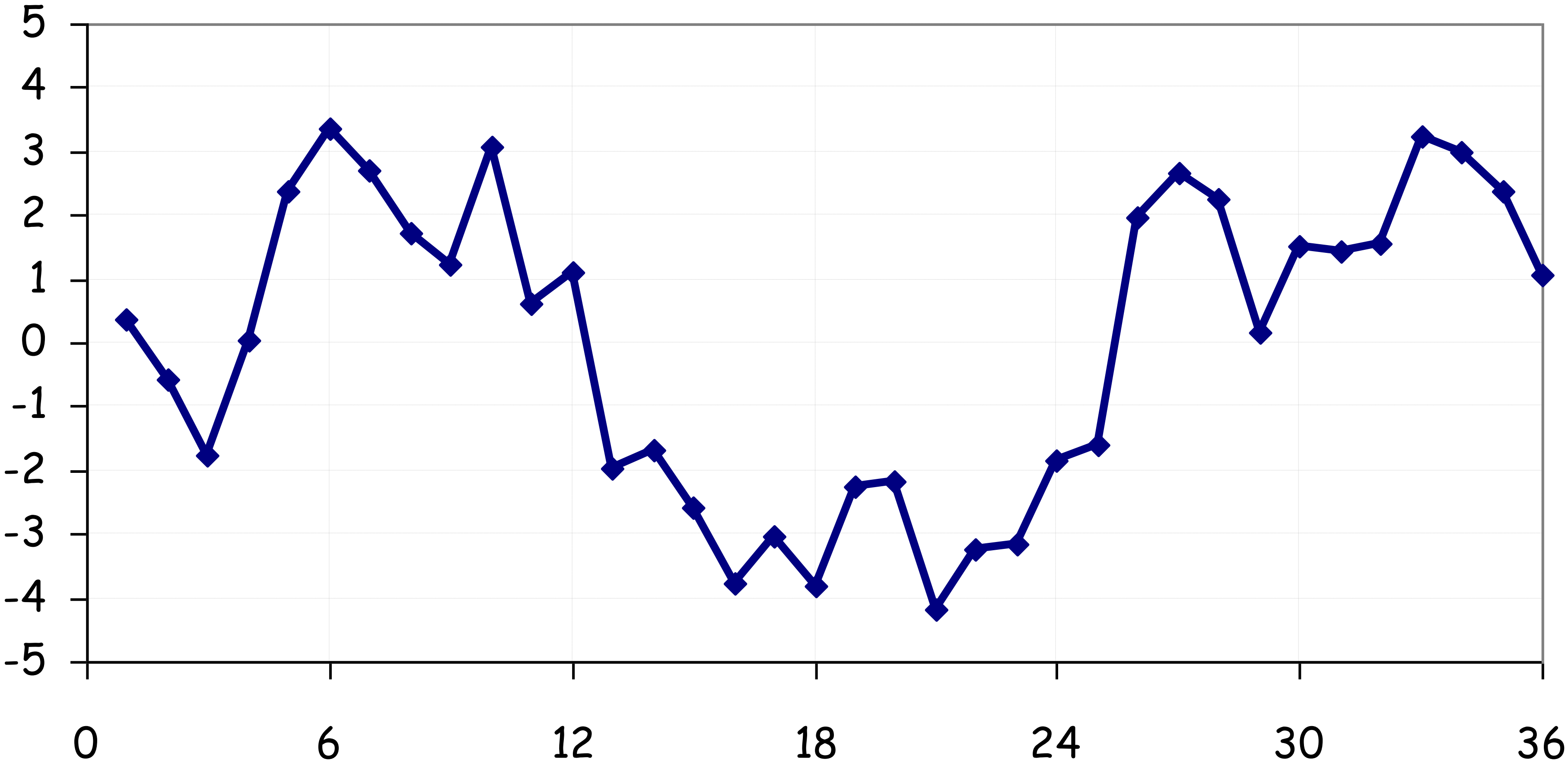
- Start at the previous value, and add a random component

$$z_t = z_{t-1} + a_t$$

where a_t is white noise

- What is the best one-step ahead forecast (expected value of z_{t+1}) for a random walk?

Random Walk



Autoregressive (AR) Process

- The current value depends on the previous value times a constant

$$z_t = \varphi z_{t-1} + a_t$$

where a_t is white noise and φ is a constant

- Autoregressive – the series is regressed on past values of itself
  - Because there is one term, and the *y* values are one step apart, it's also called an AR(1)

AR(1) Models

- AR(1) models are stationary if $|\varphi| < 1$, and nonstationary if $|\varphi| \geq 1$
- The process can be written as

$$(1 - \varphi B)z_t = a_t$$

so

$$\begin{aligned} z_t &= (1 - \varphi B)^{-1} a_t \\ &= (1 - \varphi B - \varphi^2 B^2 - \varphi^3 B^3 - \ldots) a_t \\ &= a_t - \varphi a_{t-1} - \varphi^2 a_{t-2} - \cdots \end{aligned}$$

AR(2) Process

- Current value depends on the previous two values times constants

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + a_t$$

where a_t is white noise and φ_1 and φ_2 are constants

- Also written

$$(1 - \varphi_1 B - \varphi_2 B^2) z_t = a_t$$

Seasonal Process

- Series values may be related to past values at lags of a year
  - For example, a monthly time series could be related at lag 12

$$z_t = \Phi z_{t-12} + a_t$$

Note: if the series is related at lag 12, then lags 24 and 36 might have significant autocorrelation also

More Complicated Process

- Series values may be related to the previous value and the value a year ago

$$z_t = \varphi z_{t-1} + \Phi z_{t-s} + a_t$$

where a_t is white noise,

φ (lowercase) and Φ (capital) are constants, and

s is the period (12 for monthly series

and 4 for quarterly series)

Moving Average Process

- The current value depends on lags of the white noise instead of lags of itself

$$z_t = a_t - \theta a_{t-1} \quad \text{or } z_t = (1 - \theta B) a_t$$

where a_t is white noise and $|\theta| \leq 1$ is a constant

- This is an MA(1) process.

MA and AR Models

- Just as an AR(1) can be written as an infinite MA model, an MA(1) is also an infinite AR:

$$(1 - \theta B)^{-1} z_t = a_t$$
$$(1 - \theta B - \theta^2 B^2 - \theta^3 B^3 - \ldots) z_t = a_t$$

- In practice, this may mean a high-ordered MA model can be replaced with an AR(1), or vice versa

ARIMA

- AR -Autoregressive
- I -Integrated (differenced)
- MA -Moving Average

Standard ARIMA Notation

(p d q) (P D Q)

p is the nonseasonal AR order

d is the nonseasonal order of differencing

q is the nonseasonal MA order

P is the seasonal AR order

D is the seasonal order of differencing

Q is the seasonal MA order

ARIMA (p d q)(P D Q)

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D z_t = \theta(B)\Theta(B^s)a_t$$

where

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \ldots - \Phi_P B^{Ps}$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \ldots - \Theta_Q B^{Qs}$$

Example Equation

$$z_t - z_{t-1} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

$$= (1 - B)z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$$

ARIMA (0 1 2)

1 difference

2nd-order MA process

Can you figure out the (p d q)(P D Q) form of this model?

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)^2(1 - B^{12})Y_t$$
$$= (1 - \Theta_1 B^{12})a_t$$

Estimating Coefficients

- Specify the model (p d q)(P D Q), and the program estimates the coefficients ϕ, Φ, θ, Θ and the innovation variance
- Usually ARIMA model coefficients are between -1 and 1

ARIMA Model From Output

```
ARIMA Model:  (0 1 1)(0 1 1)
  Nonseasonal differences: 1
  Seasonal differences:    1
                                       Standard
Parameter         Estimate               Errors
-----------------------------------------
Nonseasonal MA
  Lag  1            0.1484              0.19661
Seasonal MA
  Lag  4            0.9830              0.30551
Variance            0.37857E-04
SE of Var           0.15455E-04
```

Well-Known ARIMA Model

Airline Model (0 1 1) (0 1 1)

Box and Jenkins used this model to fit airline passenger series

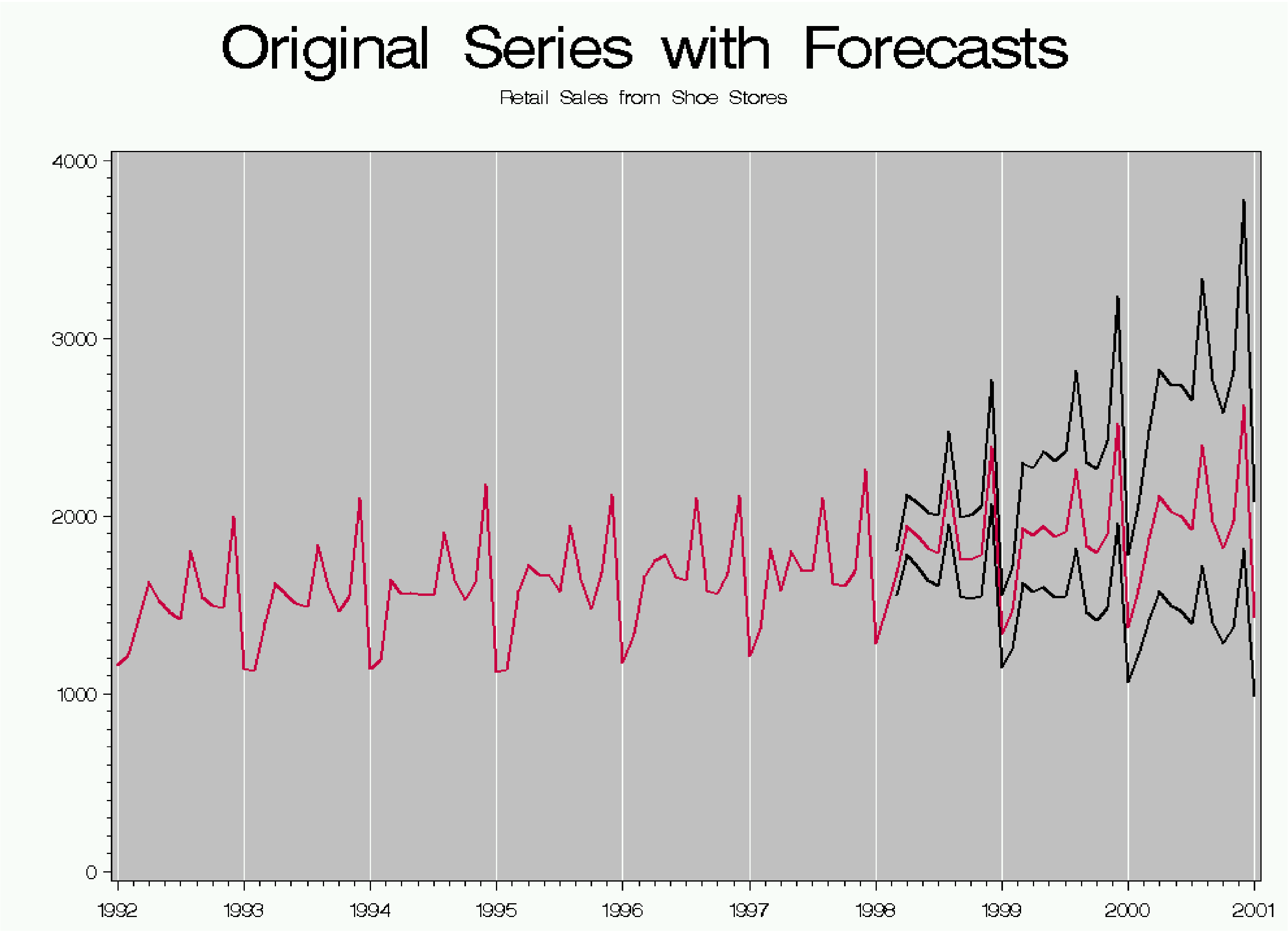
$(1 - B)(1 - B^s)Z_t = (1 - \theta B)(1 - \Theta B^s)a_t$

s is usually 12 (monthly) or 4 (quarterly)

Airline-Model Forecasts

- Forecasts from the airline model are exponentially weighted averages of past values and exponentially weighted averages of past seasonal values

Forecasts from an example airline series



ARMA Models and Autocorrelation Functions

- For each ARMA model the theoretical autocorrelation and partial autocorrelation function can be derived.
- Traditionally, model selection was done by calculating the sample ACF and PACF and matching their form to the theoretical ACF and PACF of an ARMA model.

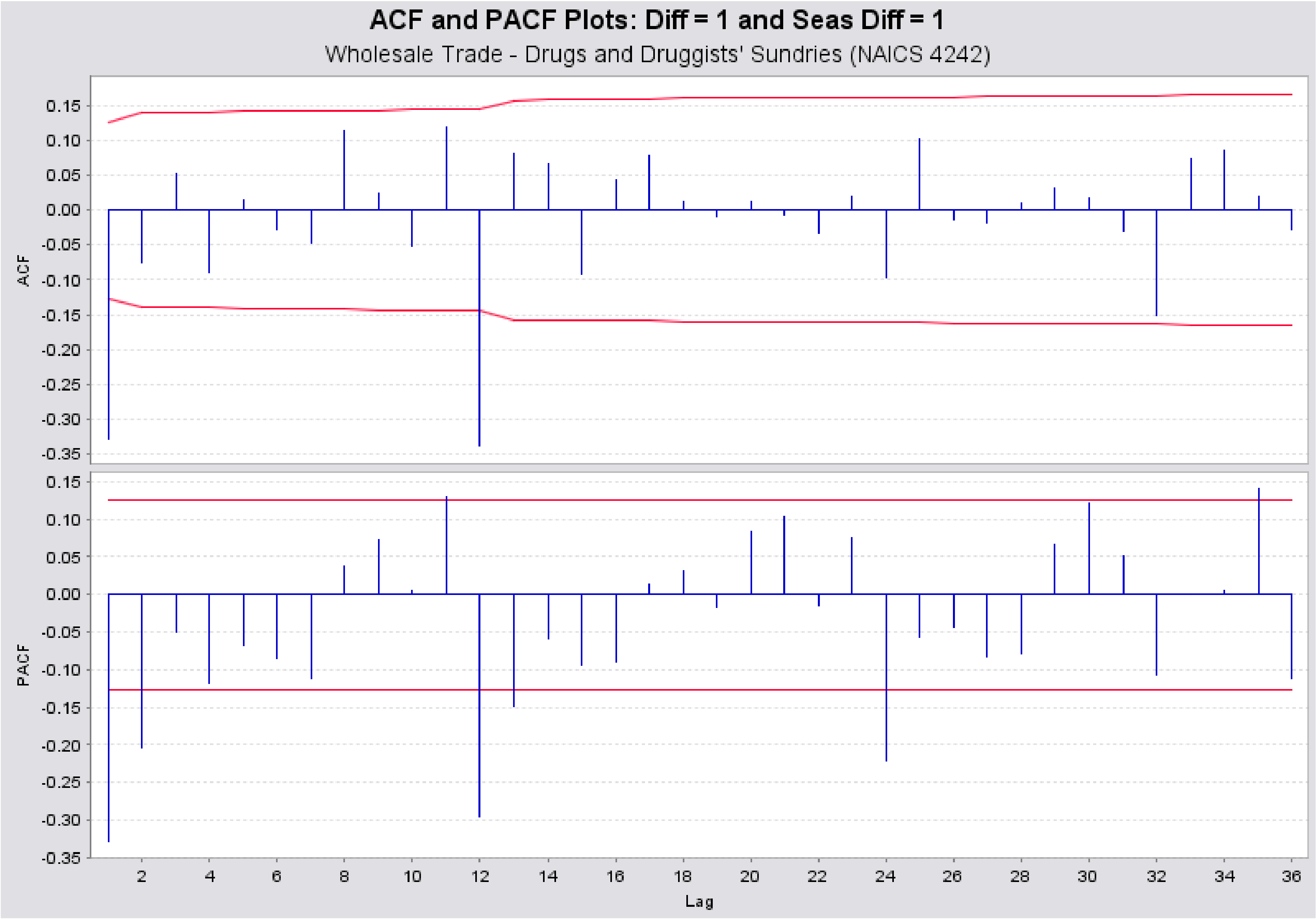
ACF and PACF of AR Models

Model	Lag	ACF	PACF
AR(1)	Lag 1	ϕ_1	ϕ_1
	Lag 2	ϕ_1^2	0
	Lag k	ϕ_1^k	0
AR(p)		Exponential decay	Spikes at lag 1 through p; zero after lag p

ACF and PACF of MA Models

Model	Lag	ACF	PACF
MA(1)	Lag 1	$-\theta_1/(1 + \theta_1^2)$	$-\theta_1/(1+\theta_1^2)$
	Lag 2	0	…
	Lag k	0	$-\theta_1^k(1-\theta_1^2)$ $/(1-\theta_1^{2(k+1)})$
MA(q)		Spikes at lags 1 through q; zero after lag q	Exponential decay

ACF/PACF From an Airline-Model Series



Model Selection
X-13ARIMA-SEATS Specs

- **`Automdl`**
  - Automatic model selection
- **`Pickmdl`**
  - Model selection from a list of possible models
- **`Identify`**
  - Autocorrelation information to choose a model “from scratch”
- **`ARIMA`**
  - Known model

Model Selection
Identify an ARIMA Model

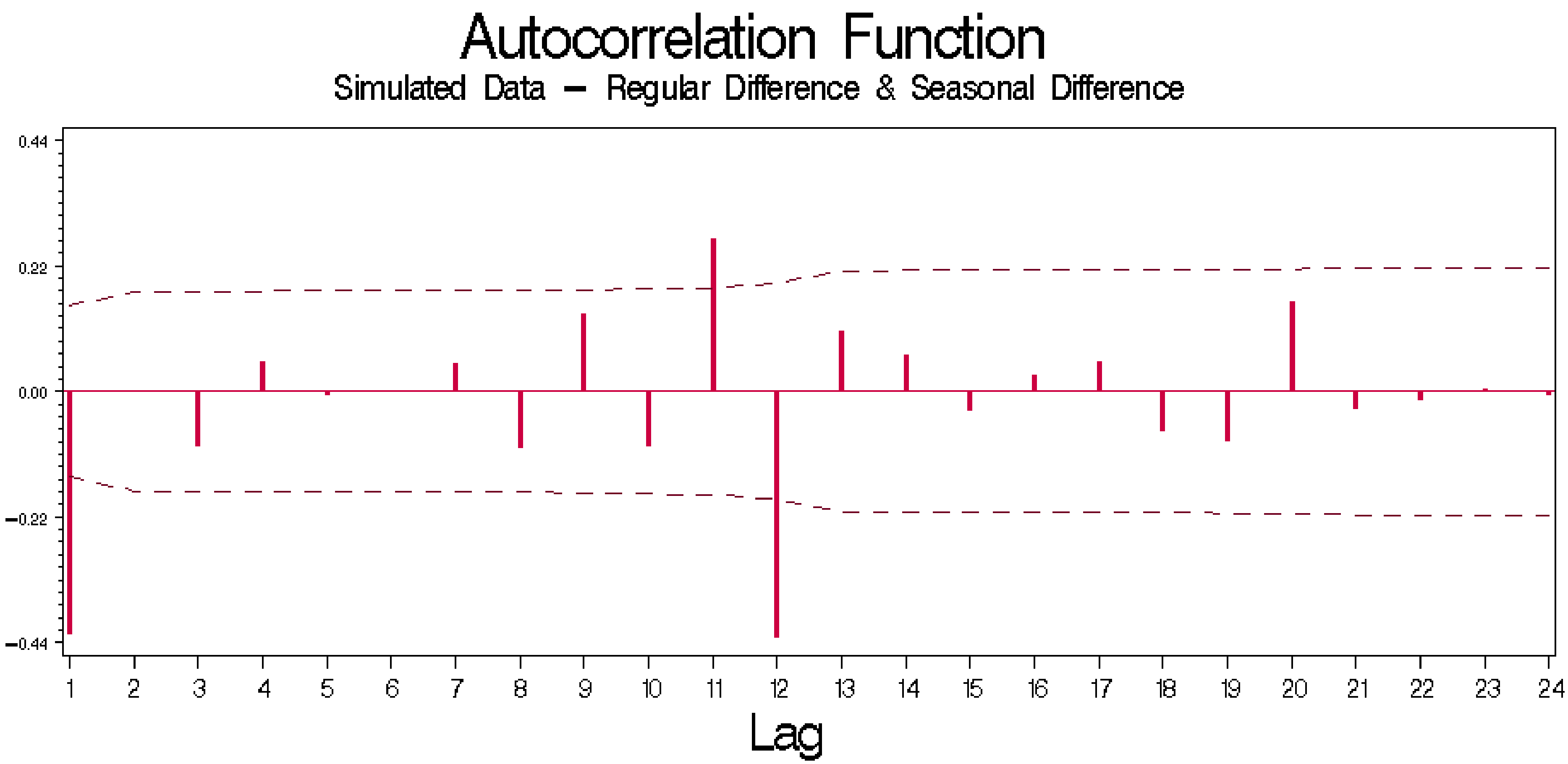
- Use ACF and PACF plots to identify the ARIMA model
  - Different plots for different orders of differencing
- Look for significance at meaningful lags (low lags and seasonal frequencies)

Model Selection
Identify Spec

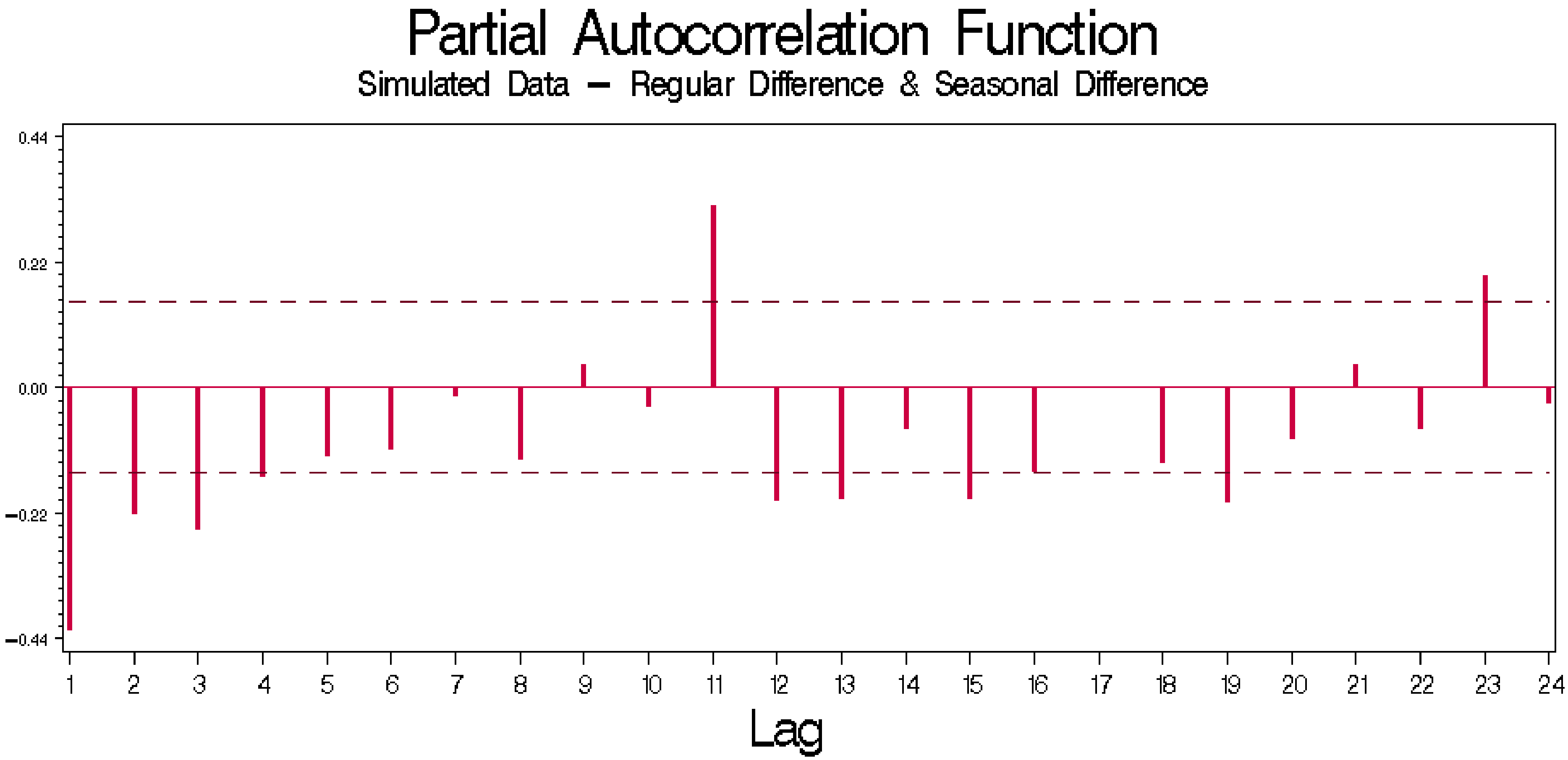
```
identify { diff = 1 sdiff = 1 }
```

OR

```
identify { diff = (0 1)
           sdiff = (0 1)
}
```



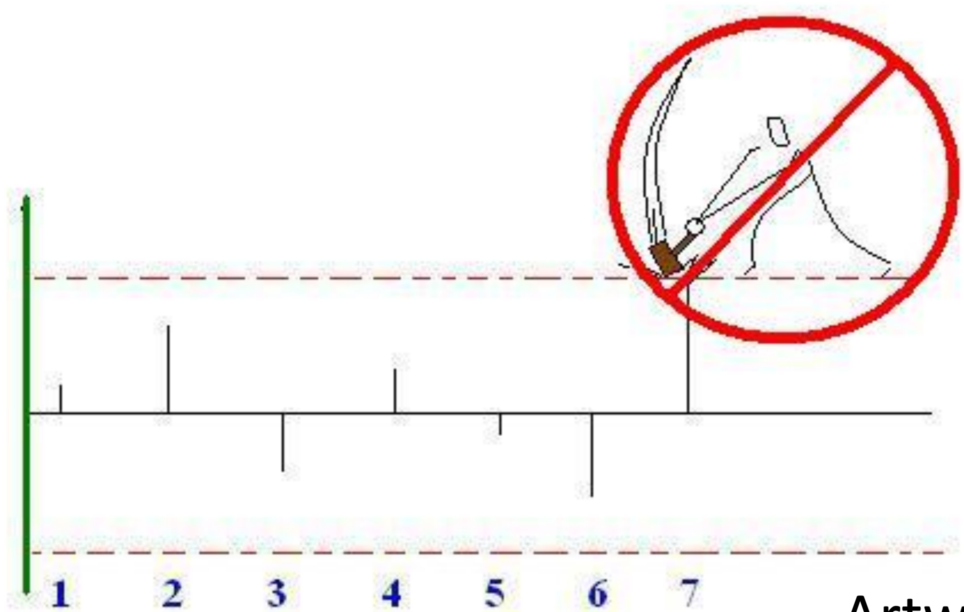
Autocorrelation Function
Simulated Data – Regular Difference & Seasonal Difference
0.44
0.22
0.00
-0.22
-0.44
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Lag



Partial Autocorrelation Function
Simulated Data – Regular Difference & Seasonal Difference
0.44
0.22
0.00
-0.22
-0.44
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
Lag

Interpreting the ACF

Disregard significance at lags that have no reason to be significant

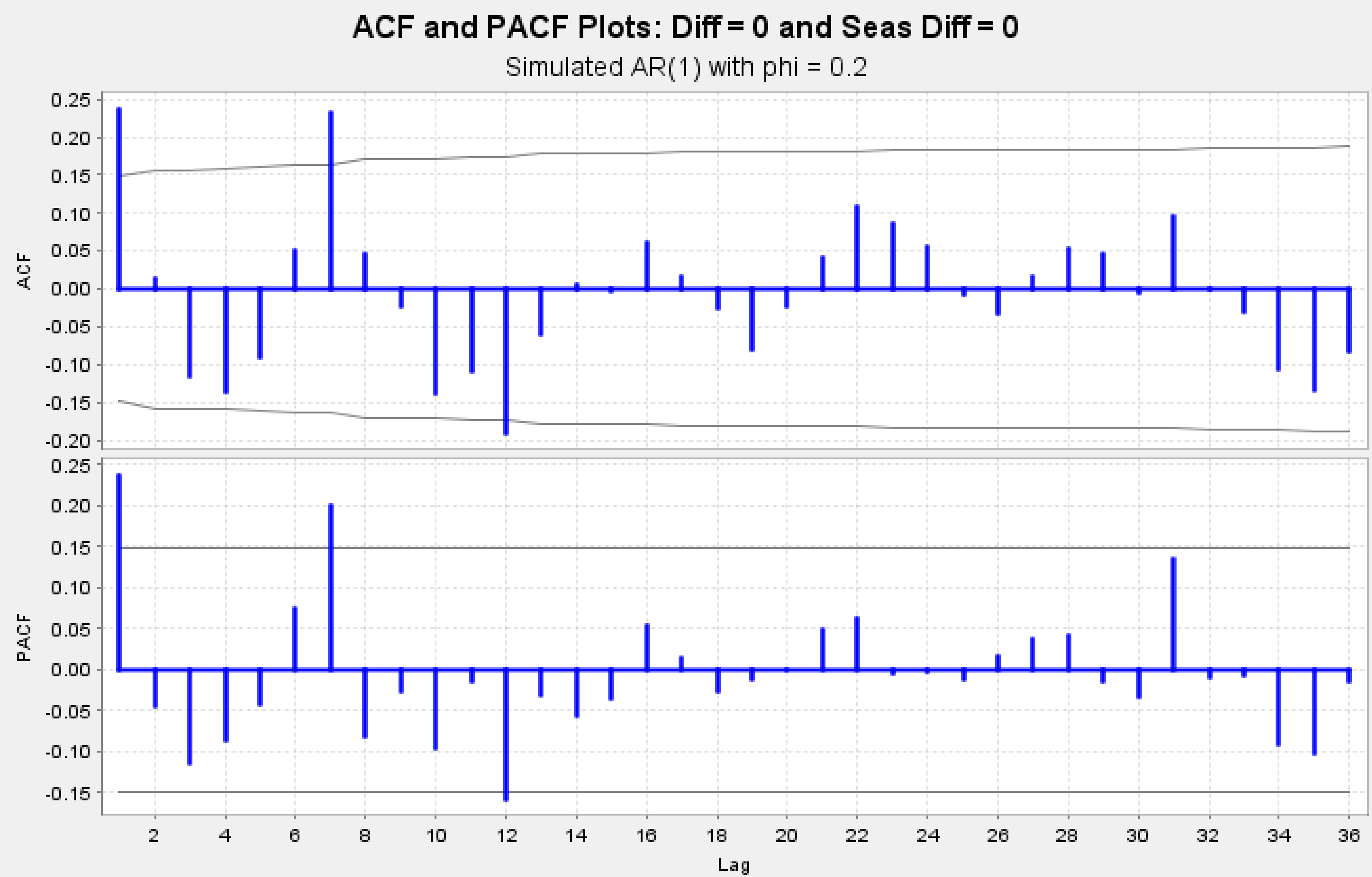


Artwork courtesy of Kevin Tolliver, 2010

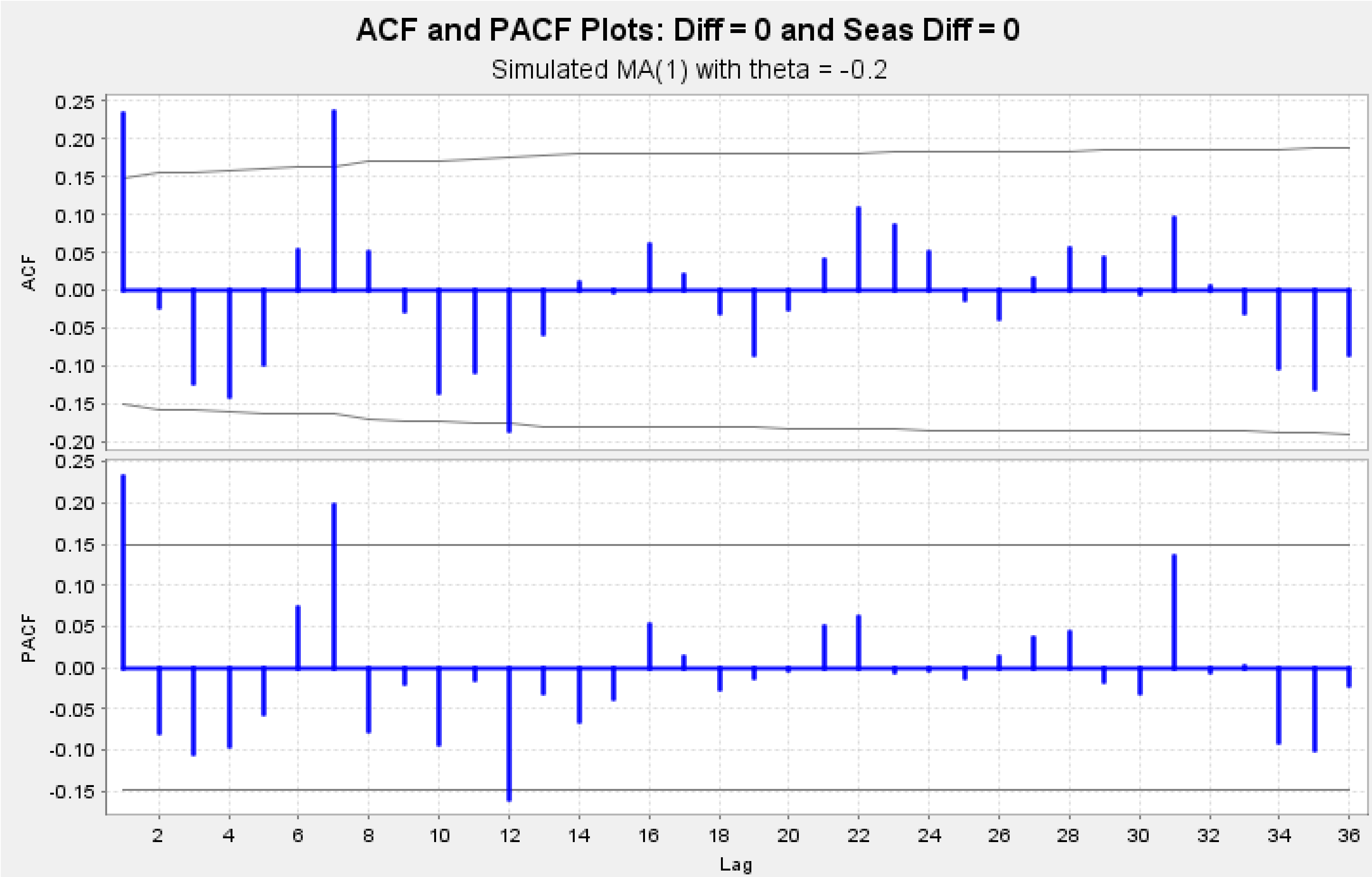
Model Identification Considerations

- The ACF and PACF may lead to multiple models to consider.

- Using the same white noise sequence, here are the ACF and PACF of
  - A simulated AR(1) with phi = 0.2
  - A simulated MA(1) with theta = -0.2



ACF and PACF Plots: Diff = 0 and Seas Diff = 0
Simulated AR(1) with phi = 0.2
ACF
PACF
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
Lag

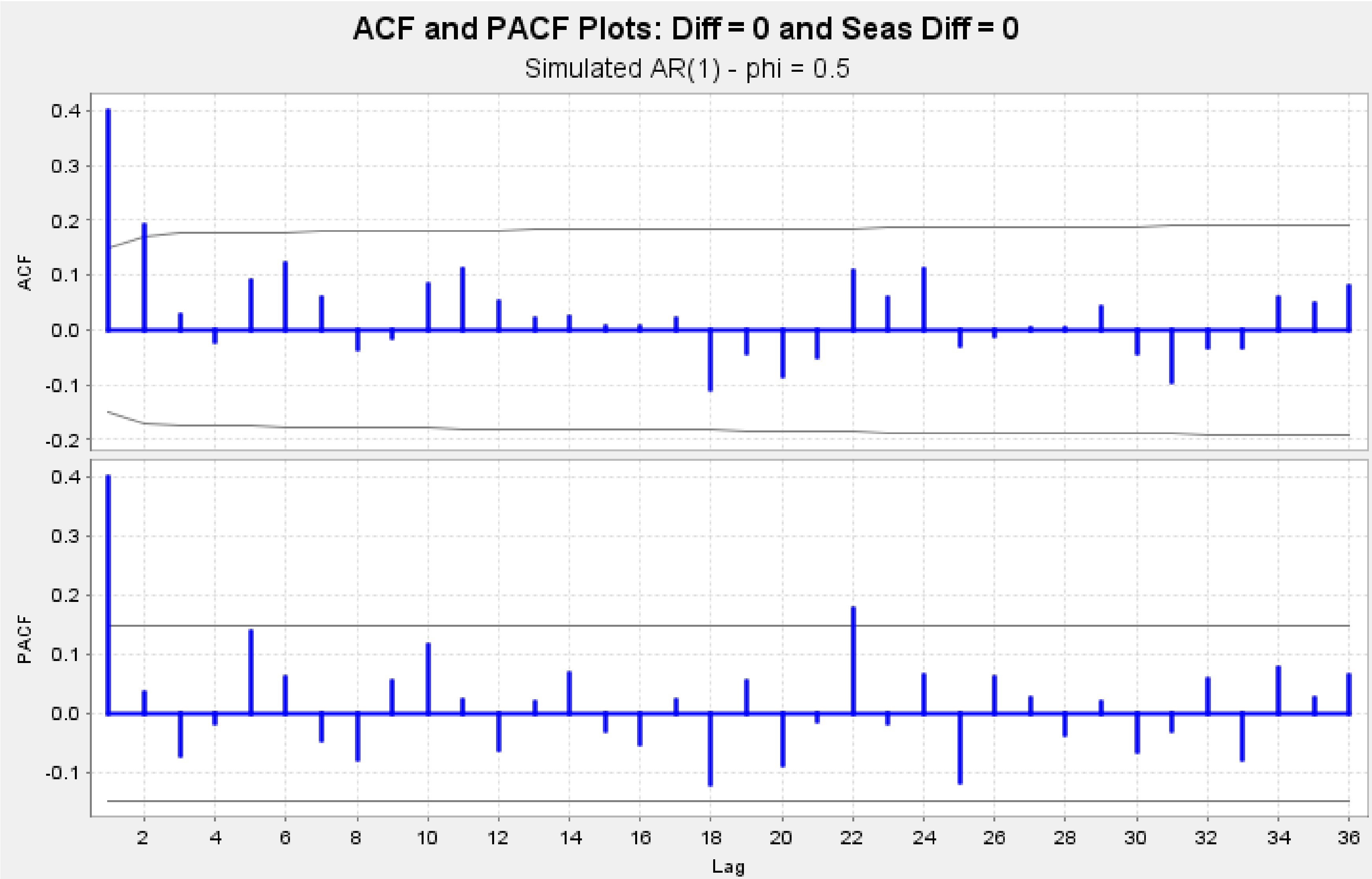

ACF and PACF Plots: Diff = 0 and Seas Diff = 0
Simulated MA(1) with theta = -0.2
ACF
PACF
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
Lag

Model Identification Considerations (Example 1)

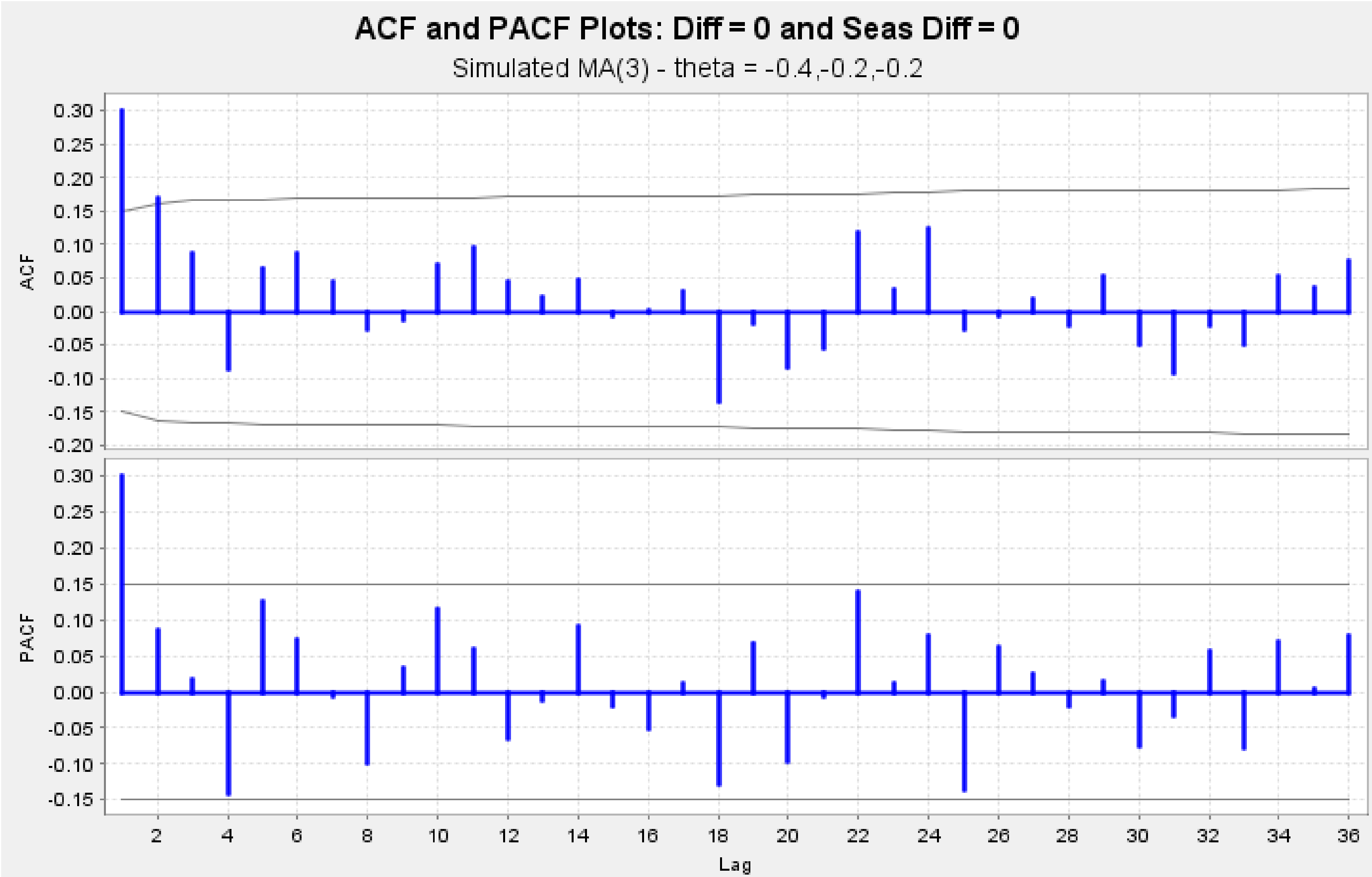
- The ACF and PACF of the two models looks similar. For these two series, both the MA(1) model and the AR(1) model may be an adequate model choice.

Model Identification Considerations (Example 1 cont)

- Note also:
  - Lag 7 is significant, but because these are simulated series we know they're not related at lag 7. Spurious spikes in the ACF and PACF happen, which is why trying to select a model to account for lags which don't make sense isn't recommended.
  - These series were not modeled to be seasonal, but lag 12 is significant. Be careful deciding that a series is seasonal due to one diagnostic.



ACF and PACF Plots: Diff = 0 and Seas Diff = 0
Simulated AR(1) - phi = 0.5
ACF
PACF
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
Lag


ACF and PACF Plots: Diff = 0 and Seas Diff = 0
Simulated MA(3) - theta = -0.4,-0.2,-0.2
ACF
PACF
Lag
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36

Model Identification Considerations (Example 2)

- The preceding ACF and PACF plots use the same white noise sequence to make:
  - AR(1) with phi = 0.5
  - MA(3) with theta1 = -0.4, theta2 = theta3 = -0.2
- The similar patterns indicate that the AR(1) series may be modeled with an MA(3), and vice versa.

Model Estimation, Simulated AR(1) Series

Series modeled as:		Estimate	Standard Error
AR(1)	Nonseasonal AR		
	Lag 1	0.40502	0.06831
MA(3)	Nonseasonal MA		
	Lag 1	-0.40282	0.07450
	Lag 2	-0.25744	0.07797
	Lag 3	-0.08623	0.07354

Model Estimation, Simulated MA(3) Series

Series modeled as:		Estimate	Standard Error
AR(1)	Nonseasonal AR		
	Lag 1	0.30568	0.07118
MA(3)	Nonseasonal MA		
	Lag 1	-0.32133	0.07369
	Lag 2	-0.21877	0.07576
	Lag 3	-0.16578	0.07287

Model Selection

- When multiple models are candidates, use model diagnostics to determine if one is preferred

Model Selection
Automatic Modeling

- **`automdl { savelog = amd }`**
  - **`automdl`** identifies models through step-by-step procedure
  - Based on TRAMO (Gómez and Maravall)

- **`pickmdl { savelog = amd }`**
  - **`pickmdl`** chooses from models in a file (5 models by default)
  - From Statistics Canada (Dagum)
  - In X-12-ARIMA, Version 0.2.10 and earlier, this was called **`automdl`**

An overview of the `automdl` procedure

1. A default model is estimated. Initial outliers and regressors are selected.
2. The program selects the order of differencing using unit root tests
3. Models with varying AR and MA orders are fit and ranked using a Bayesian Information Criterion, and the AR and MA orders are selected
4. The selected model is compared to the default model
5. The final model is checked for adequacy

Section 7.2 of the X-13A-S manual describes the procedure in detail

`automdl` Spec Options

- **`maxdiff = (d D)`**
  - Sets the maximum nonseasonal and seasonal difference to search for
  - Maxdiff = (2 1) is the default.
  - Maximum d is 2; maximum D is 1.
- **`diff = (d D)`**
  - Sets the nonseasonal and seasonal differences
  - Cannot be used in same spec as maxdiff

`automdl` Spec Options (2)

- **`maxorder = (pq PQ)`**
  - Gives the maximum nonseasonal and seasonal order of the AR and MA polynomials
  - Given an ARIMA model (p d q)(P D Q), **`maxorder = (3 1)`** means that p and q are at most 3, and P and Q are at most 1
  - Both pq and PQ must be greater than zero. Maximum pq is 4; maximum PQ is 2.
  - Default is **`maxorder = (2 1)`**

`automdl` Spec Options (3)

- **`Checkmu`**
  - Controls whether X-13A-S checks for significance of a constant term
  - If **`checkmu = yes`** (the default), then if significant, a constant term will be part of the regression
  - If **`checkmu = no`**, the regression includes a constant only if the user specifies it in the **`regression`** spec

`automdl` Spec Options (4)

- Other **`automdl`** options control how a model is chosen – the acceptance criteria for certain diagnostics checked during the **`automdl`** procedure
  - We rarely change these from the default

ShoesAutomdl.spc

```
series{title =
 "Retail Sales: Shoe Stores (Automdl)"
    file = "shoes.dat"  format = "datevalue"
    span = (2001.1, )
}
transform { function = log }
automdl{ maxorder = (3 1) }
check { print = all }
estimate {print=all} forecast {print=none}
x11 { … }
```

ARIMA Spec

- Choose the ARIMA model by **identify**, **automdl**, or **pickmdl**, and use the **arima** spec to hard-code it
  - Can run **identify** spec in addition to an automatic modeling spec (not required)
  - Only one of **automdl**, **pickmdl**, and **arima** specs can be in a spec at any time
- Always hard-code the ARIMA model for production runs

ARIMA Spec Syntax

```
arima{model=(p d q)(P D Q)

#  ar = (initial coefficients for AR, or
#  fixed values with suffix f, e.g. -.6f)
#  ma = (initial coefficients for MA, or
#  fixed values with suffix f, e.g. -.6f)
# sometimes use initial coefficients
# if X-13A-S has trouble with estimation
}
```

ShoesARIMA.spc

```
series{title =
 "Retail Sales: shoe stores (Airline Model)"
    file = "shoes.dat"  format = "datevalue"
    span = (2001.1, )
}
transform { function = log }
#automdl{  maxorder = (3 1)  }
arima { model = (0 1 1)(0 1 1)  }
check { print = all }
estimate {print=all} forecast {print=none}
x11 { … }
```

Skipped-Lag Model

```
arima{model=(0 1 [1 6])(0 1 1)}
```

$$(1 - B)\ (1 - B^{12})\ Y_t =$$
$$(1 - \vartheta_1 B - \vartheta_6 B^6)\ (1 - \Theta_{12} B^{12}) a_t$$

Skipped-Lag Model Interpretation

(0 1 [1 6])(0 1 1)

=

(0 1 6)(0 1 1)

With the 2nd-, 3rd-, 4th-, and 5th-order MA coefficients constrained to be 0

Skipping Lags

- Never start with a skipped-lag model
  - Do not base decision to skip a lag on the acf or pacf plot(s)
- Always fit the full model
  - If intermediate lag coefficients are not significantly different from 0 then those lags may be skipped

Skipped-Lag Models, Yes/No?

- Skipping lags is not required even if the coefficient is 0
  - Bill Bell recommends against skipping lags
- Skipping a lag reduces the number of parameters that X-13ARIMA-SEATS has to estimate
  - May lead to better likelihood diagnostics, but this improvement is rather artificial
- X-13A-S will not perform a SEATS adjustment for a skipped-lag model

Skipping Lags, AR Models

- Generally avoid skipping lags with AR models
  - More complicated than skipping lags with MA models
    - AR process operates on differenced series
    - MA process operates on the white noise
- Not a hard-and-fast rule
  - Always look at the full model
  - Never skip a lag if it is significant in the full model

Example: Series Run with (0 1 4)(0 1 1)

	Estimate	Standard Error
Nonseasonal MA		
Lag 1	0.52548	0.05569
Lag 2	0.04732	0.06352
Lag 3	-0.02171	0.06371
Lag 4	0.15671	0.05606
Seasonal MA		
Lag 12	0.45587	0.04998

We might try the model (0 1 [1 4])(0 1 1), since lag 2 and lag 3 are both <1 standard error from zero

Mixed Models

- A model is mixed if there are positive (nonzero) AR and MA orders in the same model component

  p and q both > 0

  or

  P and Q both > 0

- Generally avoid mixed models

Mixed Model Examples (Avoid)

(1 1 1) (0 0 0): $p = 1$, $q = 1$

(3 0 1) (0 1 0): $p = 3$, $q = 1$

(0 1 1) (1 1 1): $P = 1$, $Q = 1$

(1 1 2) (1 0 1): $p = 1$, $q = 2$; $P = 1$, $Q = 1$

Nonmixed Model Examples (Okay)

(1 1 0) (0 1 1): $p = 1, q = 0; P = 0, Q = 1$

(0 1 6) (1 1 0): $p = 0, q = 6; P = 1, Q = 0$

(1 0 0) (1 0 0): $p = 1, q = 0; P = 1, Q = 0$

(0 1 1) (0 1 1): $p = 0, q = 1; P = 0, Q = 1$

Mixed models and `automdl`

- By default, **`automdl`** searches for mixed models
- Use **`mixed = no`** to avoid mixed models
- Note: you (rarely) may still get a mixed model if no acceptable model was found

Maximum Model Order

- Nonseasonal
  - Monthly series: Prefer to see order <= 3 but depends on knowledge and experience
  - Quarterly series: Expect to see order <=1 for most series (possibly 2?)

- Seasonal
  - Use order <= 1

`automdl{}` examples

Example 1: Drugs and Druggists' Sundries Wholesale Sales automdl results:

Results of Unit Root Test for identifying orders of differencing:

Regular difference order : 1 Seasonal difference order : 1

Mean is not significant.

Example 1: Drugs and Druggists' Sundries Wholesale Sales automdl results (2):

Best Five ARIMA Models

Model # 1 : (0 1 1)(0 1 1) (BIC2 = -4.725)
Model # 2 : (1 1 1)(0 1 1) (BIC2 = -4.706)
Model # 3 : (0 1 2)(0 1 1) (BIC2 = -4.706)
Model # 4 : (2 1 0)(0 1 1) (BIC2 = -4.697)
Model # 5 : (1 1 0)(0 1 1) (BIC2 = -4.690)

Preliminary model choice : (0 1 1)(0 1 1)

Example 1: Drugs and Druggists’ Sundries Wholesale Sales automdl results (3):

Final Checks for Identified Model

Checking for Unit Roots.

No unit root found.

Checking for nonseasonal overdifferencing.

Nonseasonal MA not within 0.001 of 1.0 - model passes test

Checking for insignificant ARMA coefficients.

Final automatic model choice : (0 1 1)(0 1 1)

Example 1: Drugs and Druggists' Sundries Wholesale Sales automdl results (4):

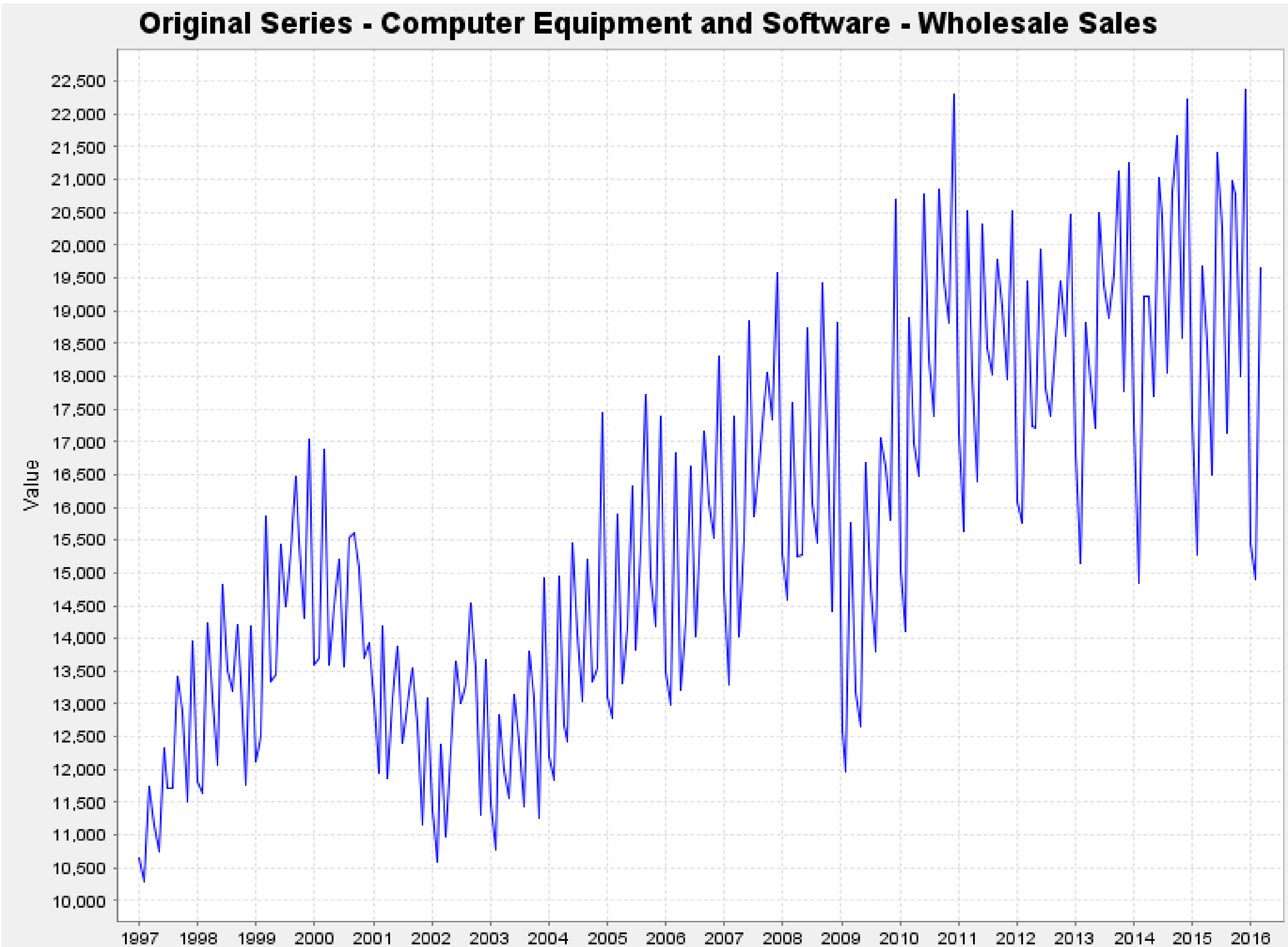
ARIMA Model

	Estimate	Standard Error
Nonseasonal MA		
Lag 1	0.44009	0.05265
Seasonal MA		
Lag 12	0.62228	0.04627

Model Innovation Variance

Variance	0.47919E-03
Standard Error of Variance	0.40002E-04

Example 2 – Computer Equipment and Software Wholesale Sales



Example 2 – Computer Equipment and Software Wholesale Sales (2)

With automdl{ print = all }, a (2 1 2)(0 1 1) model is selected:

Best Five ARIMA Models

Model # 1 : (2 1 2)(0 1 1) (BIC2 = -3.530)

Model # 2 : (2 1 1)(0 1 1) (BIC2 = -3.514)

Model # 3 : (2 1 0)(0 1 1) (BIC2 = -3.510)

Model # 4 : (0 1 1)(0 1 1) (BIC2 = -3.507)

Model # 5 : (1 1 0)(0 1 1) (BIC2 = -3.490)

Example 2 – Computer Equipment and Software Wholesale Sales (3)

	Estimate	Standard Error
Nonseasonal AR		
Lag 1	-0.84933	0.07866
Lag 2	-0.84101	0.071
Nonseasonal MA		
Lag 1	-0.63556	0.10941
Lag 2	-0.61448	0.10179
Seasonal MA		
Lag 12	0.60573	0.05836

Example 2 – Computer Equipment and Software Wholesale Sales (4)

Roots of ARIMA Model

	Real	Imaginary	Modulus	Frequency
	Nonseasonal AR			
Root 1	-0.50494	0.966475	1.09043	0.326625
Root 2	-0.50494	-0.96648	1.09043	-0.32663
	Nonseasonal MA			
Root 1	-0.51716	1.16617	1.27569	0.316432
Root 2	-0.51716	-1.16617	1.27569	-0.31643
	Seasonal MA			
Root 1	1.65091	0	1.65091	0

Example 2 – Computer Equipment and Software Wholesale Sales (5)

Series run with automdl{ mixed = no print = all }

Best Five ARIMA Models

Model # 1 : (2 1 0)(0 1 1) (BIC2 = -3.510)

Model # 2 : (0 1 1)(0 1 1) (BIC2 = -3.507)

Model # 3 : (1 1 0)(0 1 1) (BIC2 = -3.490)

Model # 4 : (0 1 2)(0 1 1) (BIC2 = -3.483)

Model # 5 : (0 1 0)(0 1 1) (BIC2 = -3.453)

Preliminary model choice : (2 1 0)(0 1 1)

Example 2 – Computer Equipment and Software Wholesale Sales (6)

(2 1 0)(0 1 1) parameter estimates:

	Estimate	Standard Error
Nonseasonal AR		
Lag 1	-0.304	0.0661
Lag 2	-0.21622	0.06585
Seasonal MA		
Lag 12	0.54432	0.05817

Example 2 – Computer Equipment and Software Wholesale Sales (7)

Series run with
automdl{ mixed = no maxorder = (3 1) }

	Estimate	Standard Error
Nonseasonal AR		
Lag 1	-0.26487	0.06586
Lag 2	-0.1622	0.06733
Lag 3	0.19869	0.06682
Seasonal MA		
Lag 12	0.56451	0.05737

Example 2 – Computer Equipment and Software Wholesale Sales (8)

Diagnostics of the various models:

	AICC	LBQ Failures	Sig ACF	Sig PACF	V.S. Resid Peaks
(2 1 2)(0 1 1)	3421.3	none			S3 T2
(2 1 0)(0 1 1)	3432.4	4 to 24	3 6	3 6	
(3 1 0)(0 1 1)	3426.0	none			T1