Seasonality in Data and How to Adjust

Brian C. Monsell
U. S. Census Bureau
Brian.C.Monsell@census.gov

September 23, 2009
Workshop for Transportation Forcasters

USCENSUSBUREAU

X-12-ARIMA Collaborators

TS Research Staff

- Brian C. Monsell
- Tucker McElroy
- Natalya Titova
- Christopher Blakely (student)
- Ekaterina Sotiris (student)

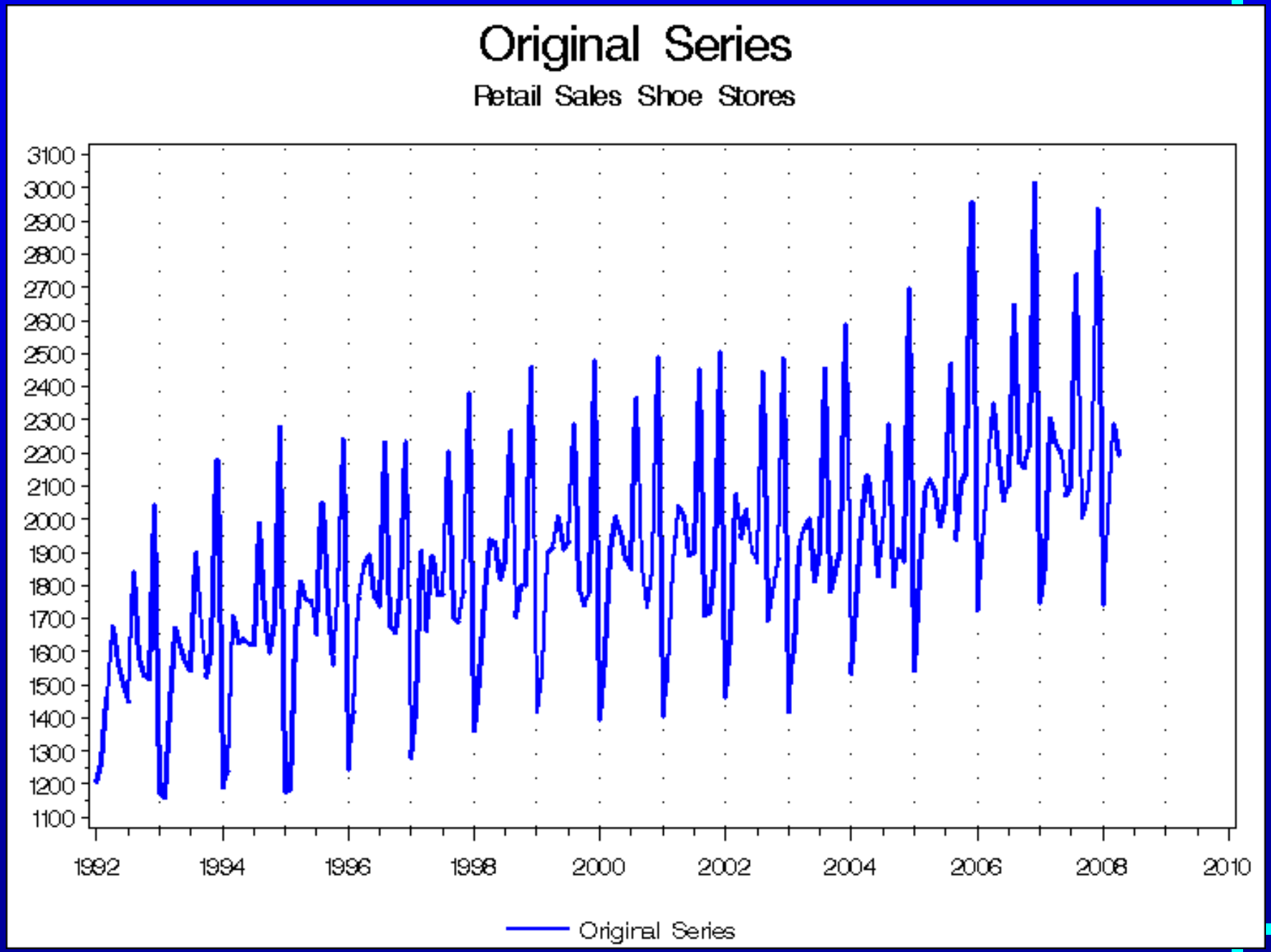
TS Methods Staff

- Kathleen McDonald-Johnson
- Ayonda Dent
- Isaac Dompreh
- Demetra Lytras
- Monica Wroblewski

Senior Math Stats: David Findley (Emeritus), Bill Bell

Outline

- Review some basic concepts
  - » Definitions of components
- Current methods and software
  - » regARIMA modeling in Seasonal Adjustment
  - » X-11 Seasonal Adjustment Overview
- What's Next?


Original Series
Retail Sales Shoe Stores
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
Original Series

Purpose of seasonal adjustment

- Bell and Hillmer (1984)
  - “Seasonal adjustment is done to simplify the data so that they may be more easily interpreted … without a significant loss of information”
- Large seasonal movements can obscure other movements of importance.
- Easier to see related movements in different series

Background: Components of a Time Series

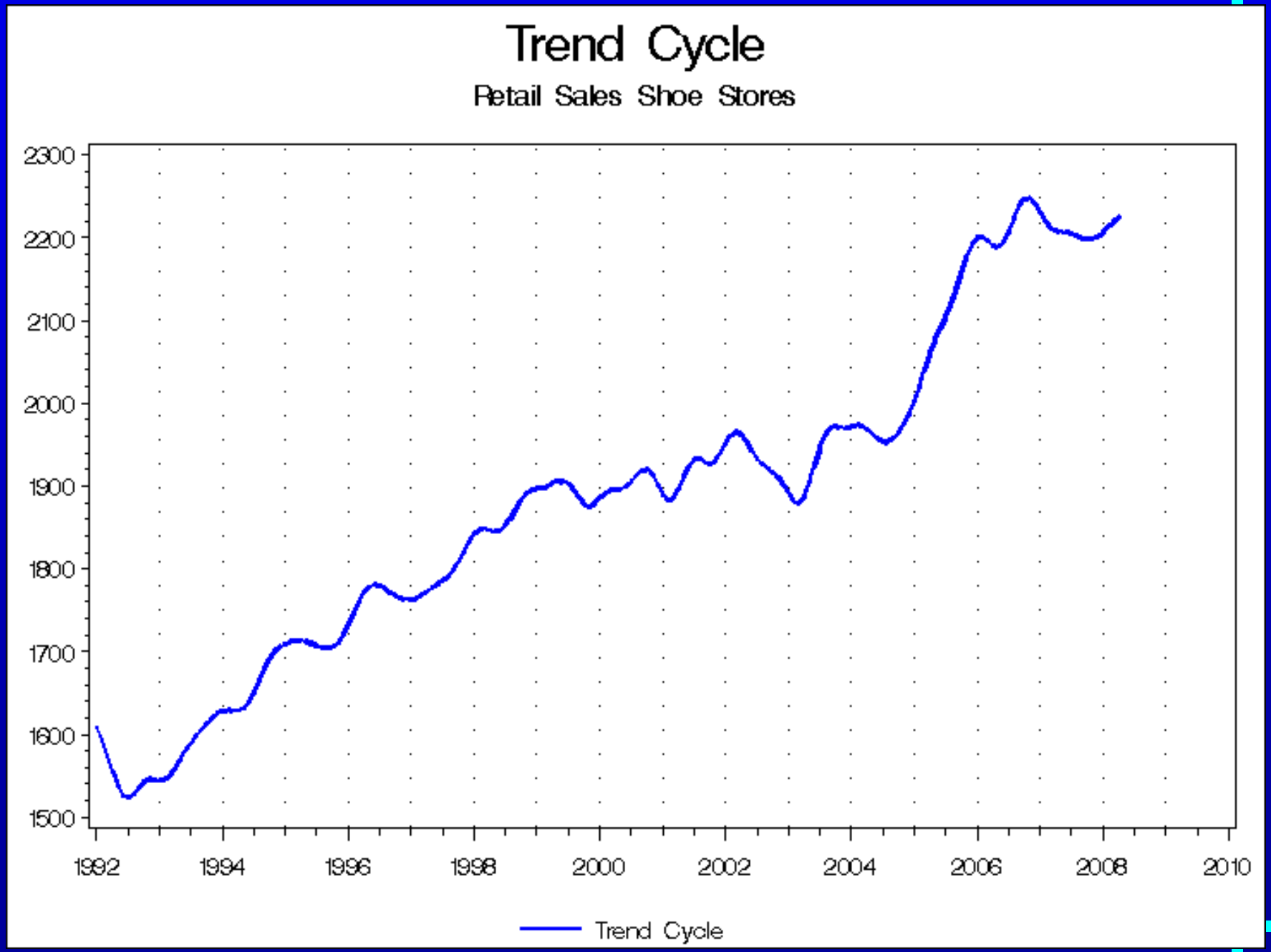
- Y = C x S x I (or C + S + I)

where

- Y = Original series
- C = Trend-cycle
- S = Seasonal effects (+ other effects)
- I = Irregular

Trend-Cycle

- Level of the series
  - » Local level estimates for the purpose of estimating seasonal factors
- Reasonably smooth, includes movements and cycles that last longer than a year
  - » Find turning points in the trend


Trend Cycle
Retail Sales Shoe Stores
2300
2200
2100
2000
1900
1800
1700
1600
1500
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
Trend Cycle

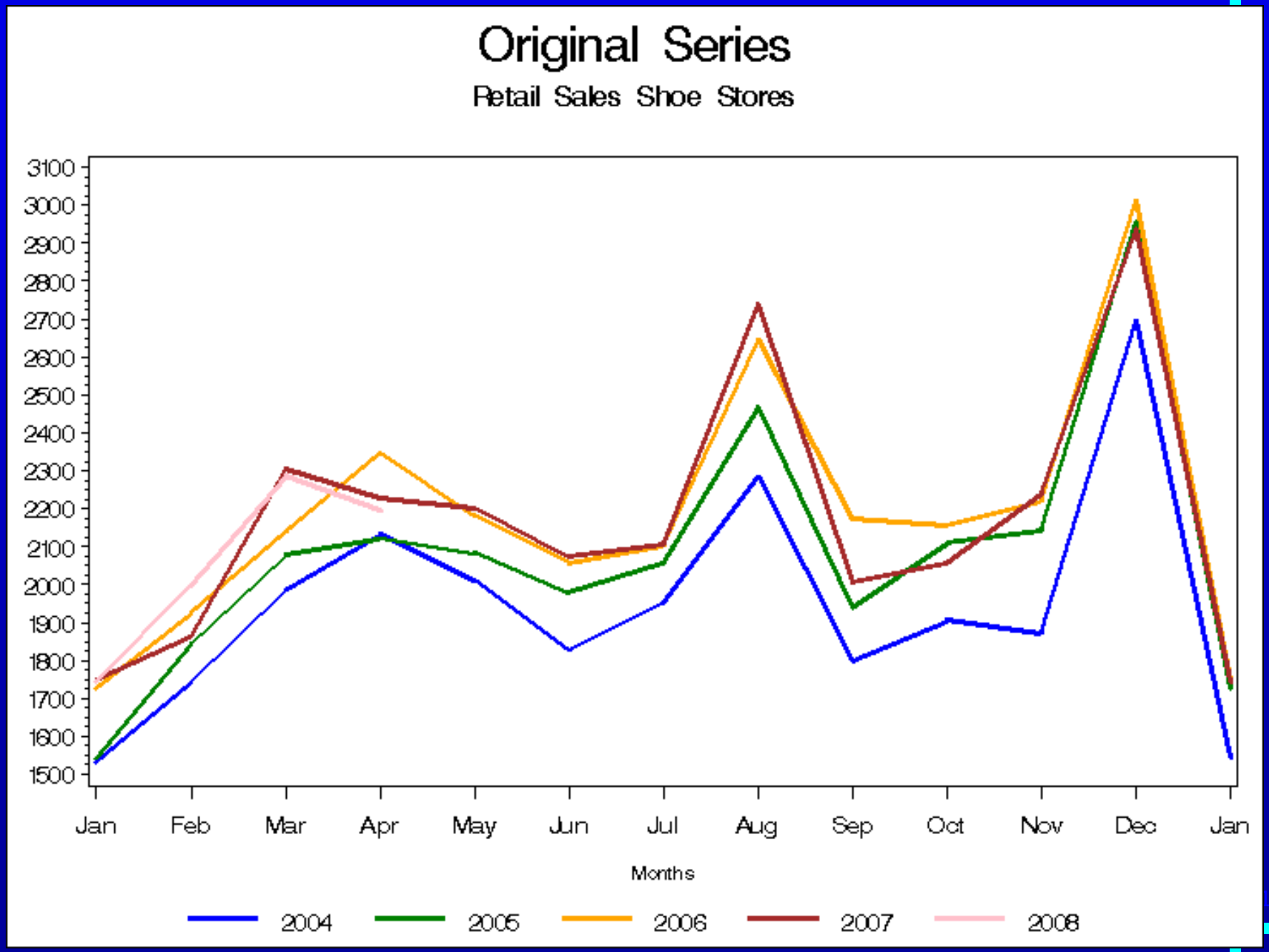
Seasonal Effects

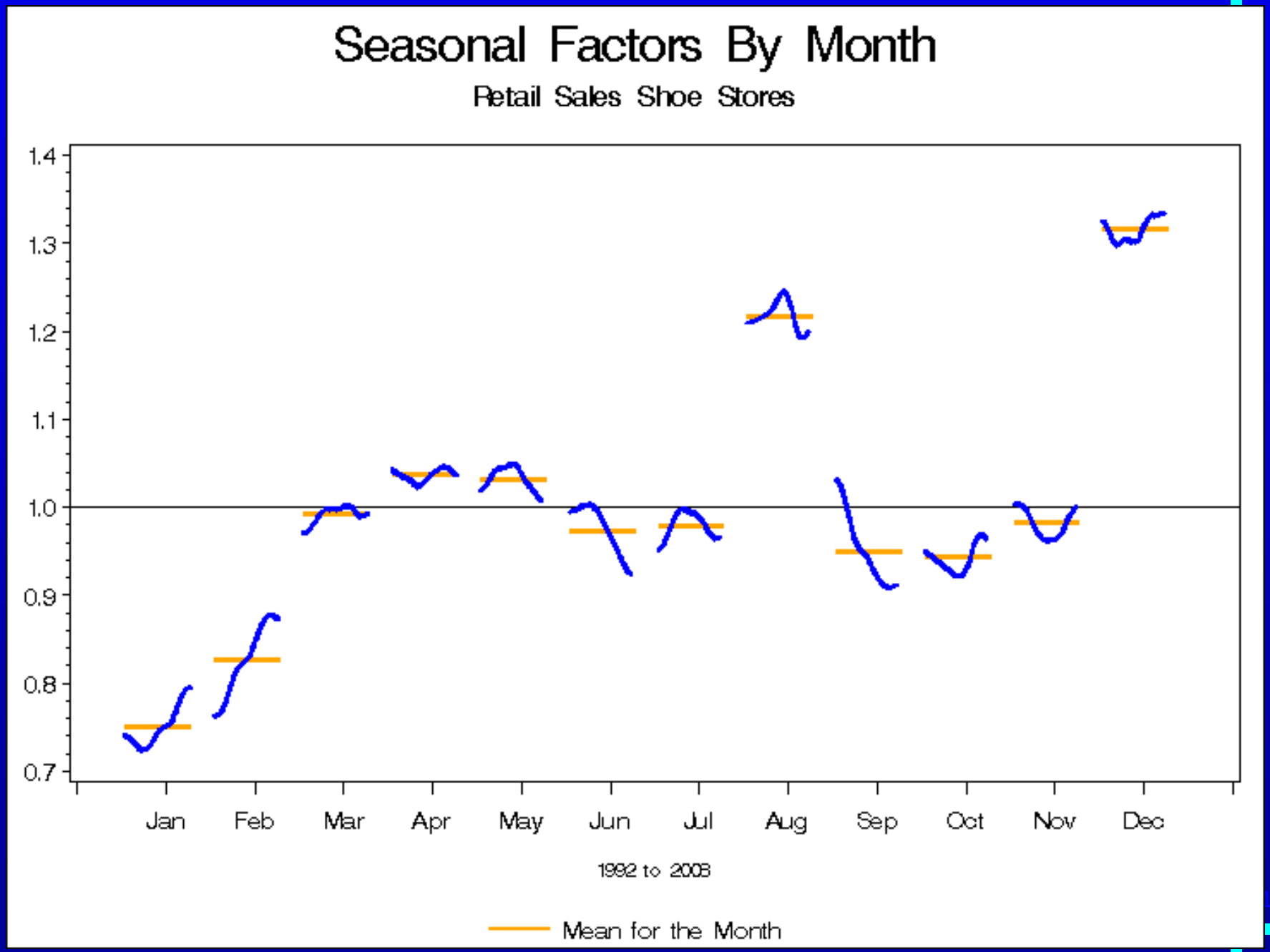
- Reasonably stable in terms of
  - » Annual timing
    - Within same month or quarter
  - » Direction
  - » Magnitude

Determining If a Series Has Seasonality

- F statistics *(X-12-ARIMA)*
- "M7“ *(X-12-ARIMA)*
- Graphs
- Spectrum diagnostics

- In addition, other diagnostics help determine if the observed seasonality is stable enough to adjust – series must have past values that reliably describe what is happening now

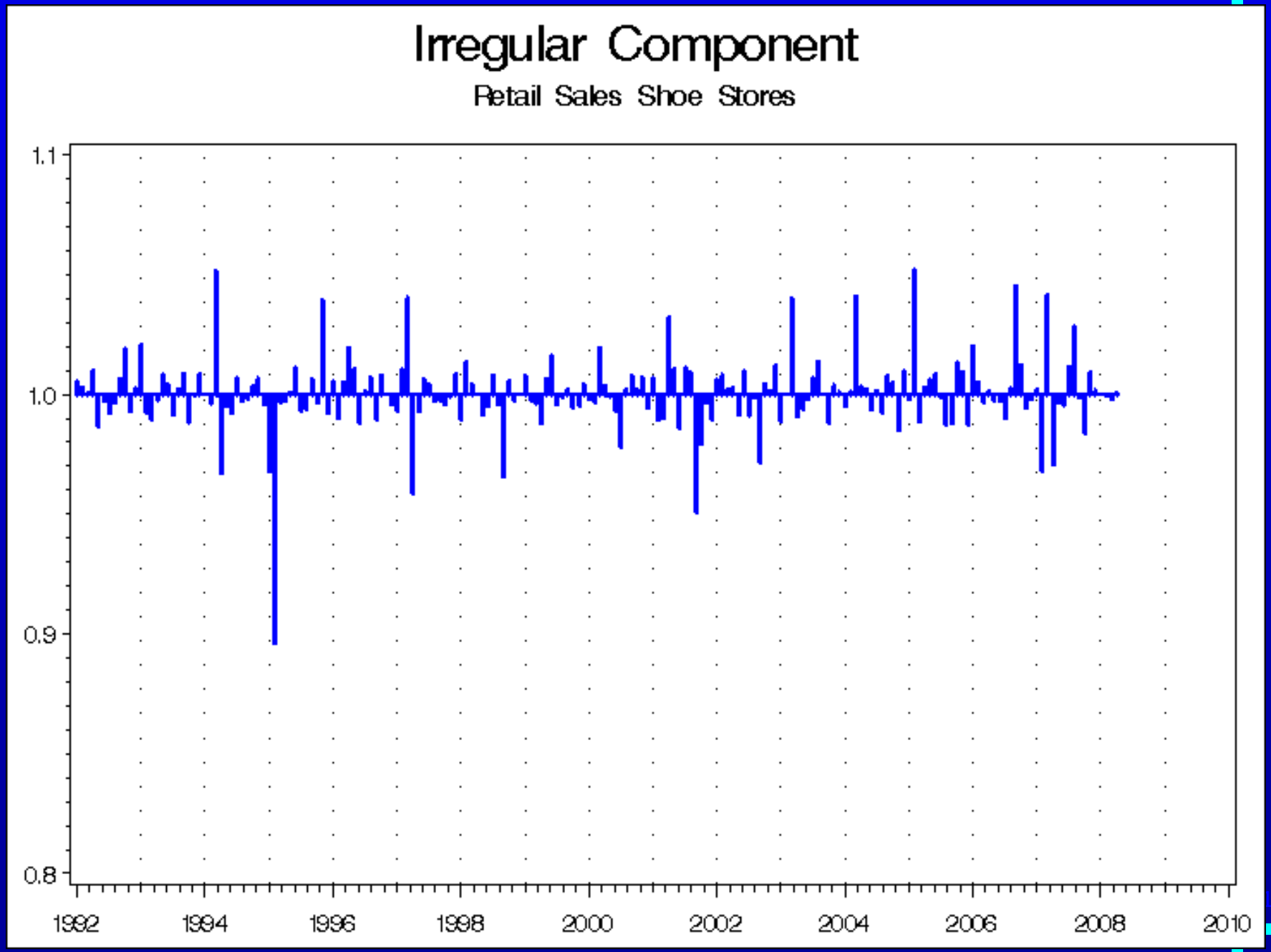

Original Series
Retail Sales Shoe Stores
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Months
2004
2005
2006
2007
2008


Seasonal Factors By Month
Retail Sales Shoe Stores
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
1992 to 2008
Mean for the Month

Irregular Effects

- Unpredictable in terms of
  - » Timing
  - » Impact
  - » Duration

- Residual after removing seasonal and trend


Irregular Component
Retail Sales Shoe Stores
1.1
1.0
0.9
0.8
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010

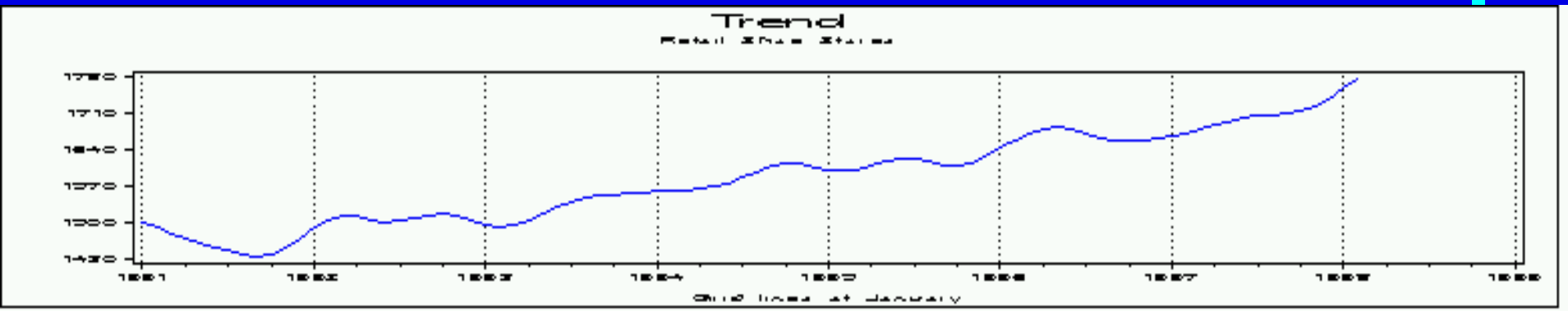
Seasonal Adjustment

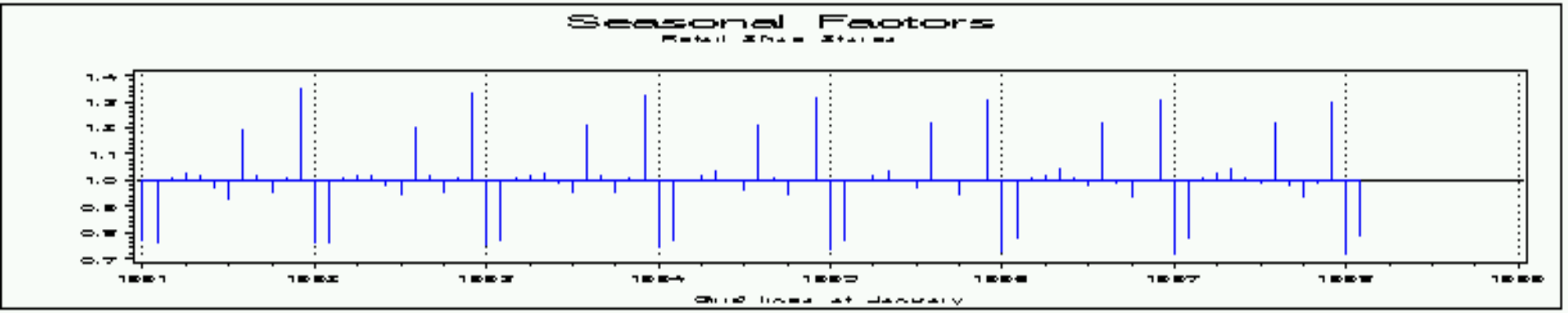
- Seasonal effects are stable in annual timing, direction, and magnitude
- Removing the typical seasonal effects is ***seasonal adjustment*** or ***seasonal decomposition***

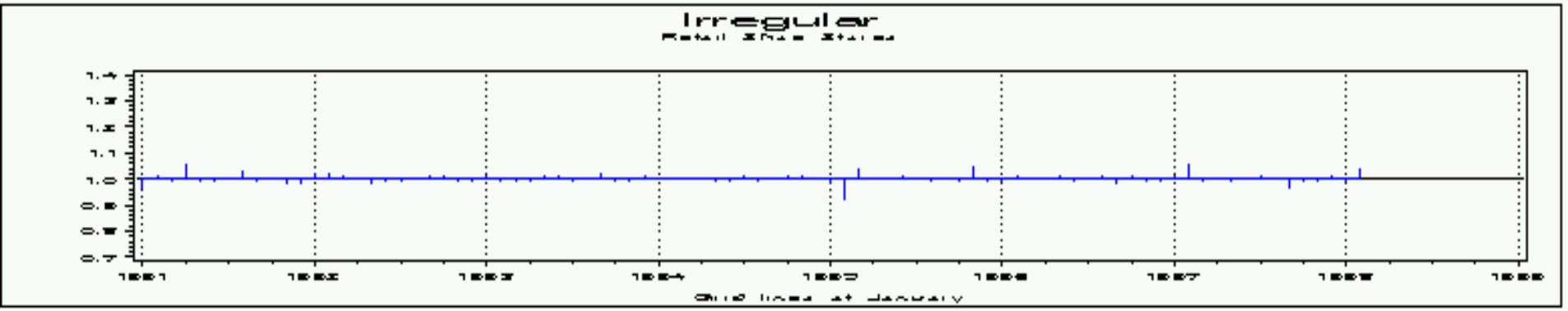
Types of Decompositions

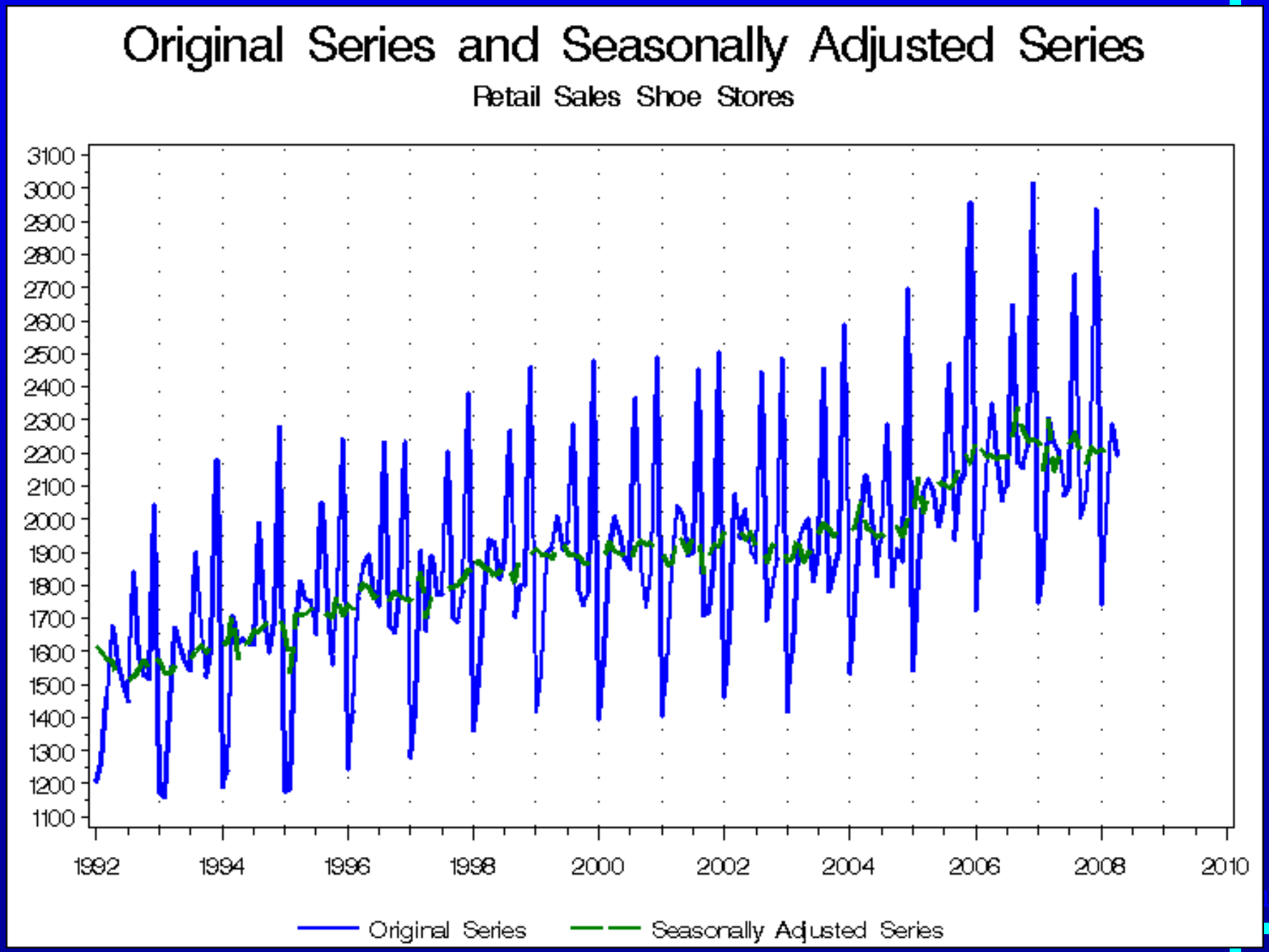
Multiplicative	$Y_t = S_t C_t I_t$	$A_t = C_t I_t$
Additive	$Y_t = S_t + C_t + I_t$	$A_t = C_t + I_t$
Log-Additive	$Ln(Y_t) = C_t + S_t + I_t$	$A_t = \exp(C_t + I_t)$

Y_t Original series
C_t Trend-Cycle component
S_t Seasonal component
I_t Irregular component


Trend


Seasonal Factors


Irregular


Original Series and Seasonally Adjusted Series
Retail Sales Shoe Stores
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
Original Series
Seasonally Adjusted Series

Current methods (and software) for seasonal adjustment

- Non-parametric methods
  - » The X-11 family (U. S. Census Bureau, Statistics Canada)
  - » SABL (Bell Labs)
  - » STL (Seasonal-Trend Loess – Bell Labs)
- Parametric (model-based) methods
  - » TRAMO/SEATS (Bank of Spain)
  - » STAMP (Andrew Harvey)

Currently

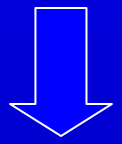
- Most statistical agencies and central banks are using
  - » X-12-ARIMA
  - » TRAMO-SEATS

X-12-ARIMA

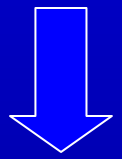
- Developed at the Census Bureau – Findley, Monsell, Bell and Otto (1990)
- Current method for statistical agencies in United States, UK, Canada, New Zealand, Japan, and other countries

“X-11” Family

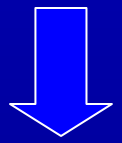
X-11



X-11-ARIMA



X-12-ARIMA



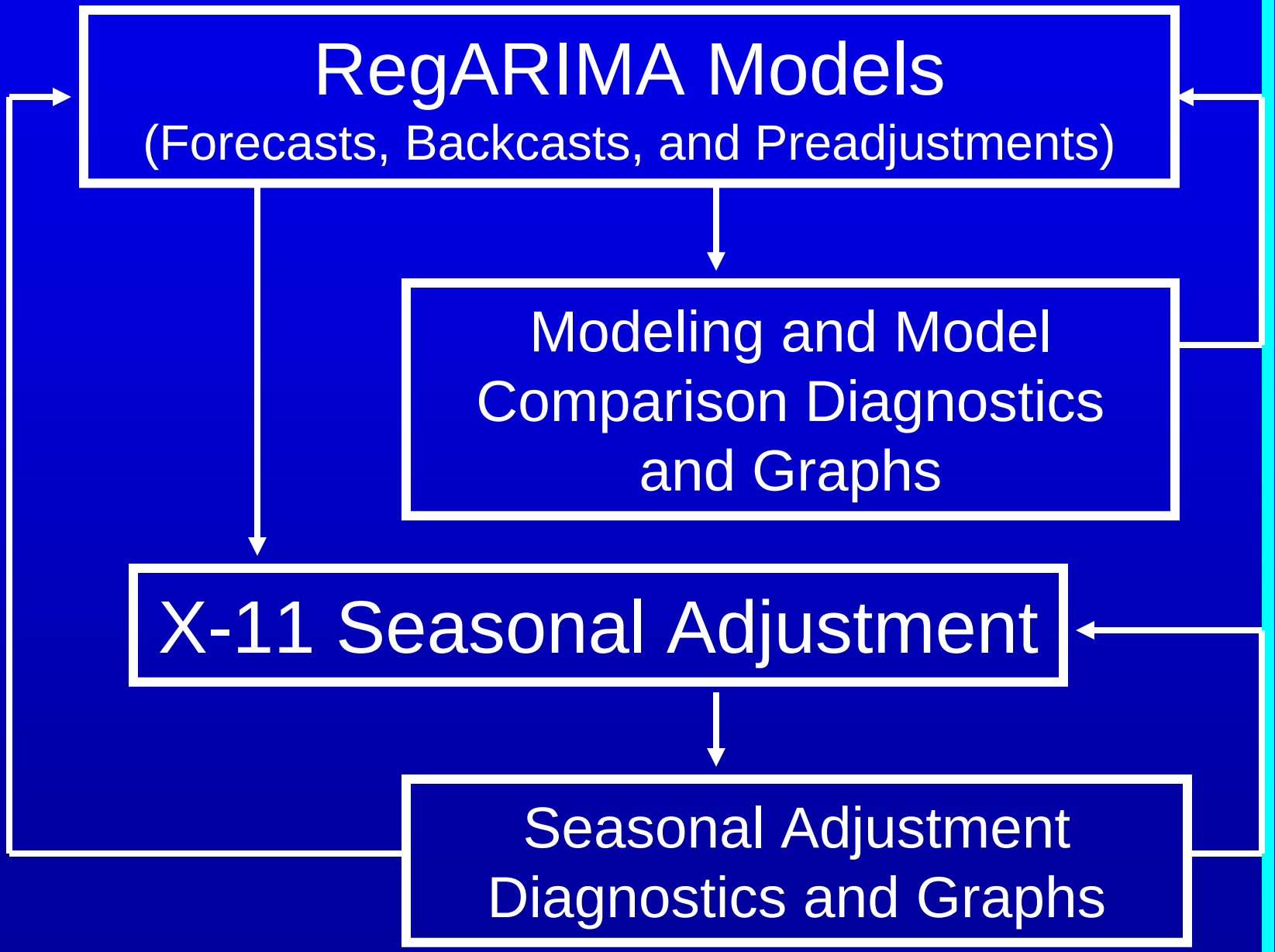
X-13-ARIMA-SEATS

Features of X-12-ARIMA

- Wide variety of seasonal and trend filter options;
- Suite of modeling and seasonal adjustment diagnostics, including
  - » Spectral diagnostics;
  - » Diagnostics of the quality and stability of the seasonal adjustments;
  - » Out of sample forecast error model selection diagnostics.

Features of X-12-ARIMA

- Extensive time series modeling and model selection capabilities
  - » linear regression models with ARIMA errors (regARIMA models);
  - » Automatic model selection options;
  - » User-defined regression variables.


RegARIMA Models
(Forecasts, Backcasts, and Preadjustments)
Modeling and Model Comparison Diagnostics and Graphs
X-11 Seasonal Adjustment
Seasonal Adjustment Diagnostics and Graphs

TRAMO/SEATS (Bank of Spain)

- Consists of two linked programs
  - » **TRAMO** is a complete regARIMA modeling package, with automatic identification of ARIMA models, outliers and other components
  - » **SEATS** takes modeling results from TRAMO and performs a model-based signal extraction
- Used by European statistical agencies

SEATS in a nutshell

- SEATS uses filters derived from an ARIMA model to describe the different components.

- The model-based approach allows
  - » **Interpretation:** explicit model expression for each component of the time series.
  - » **Diagnostics:** the joint distribution of the estimators can be derived, and hence standard tests can be performed.
  - » **Inference:** for example, optimal forecasts of the rate of growth of the seasonally adjusted series, with the standard errors.

Airline Model – ARIMA(0 1 1)(0 1 1)

$$(1 - B)(1 - B^{12})\, z_t = (1 - \theta B)(1 - \Theta_{12} B^{12})\, a_t$$

- Can infer
  - » A model for the seasonality from seasonal MA term
  - » A model for the trend from the nonseasonal MA term
- **Hillmer and Tiao (1978)**

Time Series Modeling

RegARIMA =

Regression + ARIMA

ARIMA = Autoregressive Integrated Moving Average

RegARIMA Model

$$\log\left(\frac{Y_t}{D_t}\right) = \beta' X_t + Z_t$$

transformation

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.

Uses of RegARIMA Models in Seasonal Adjustment

- Detect and adjust for outliers and other distorting effects to improve the forecasts and seasonal adjustments (automatic option)

- Detect and estimate additional components (e.g. calendar effects)

Uses of RegARIMA Models in Seasonal Adjustment (cont.)

- X-12-ARIMA
  - » Forecast (and Backcast) extension of series before applying X-11 filters
- TRAMO/SEATS
  - » Use ARIMA portion of model to generate a model-based signal extraction

Regression Effects

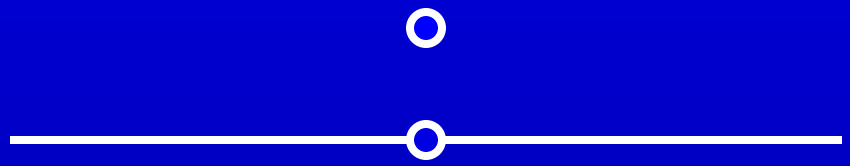
- X-12-ARIMA and TRAMO has many built-in regressors and options for users to define their own regressors
- Regressors we will discuss
  - » Outlier effects
  - » Calendar effects
    - Trading day
    - Easter

Outlier Effects

- Additive outliers (AO)
  - » Also called point outliers
- Level shifts (LS)
- Temporary changes (TC)
- Ramps (RP)

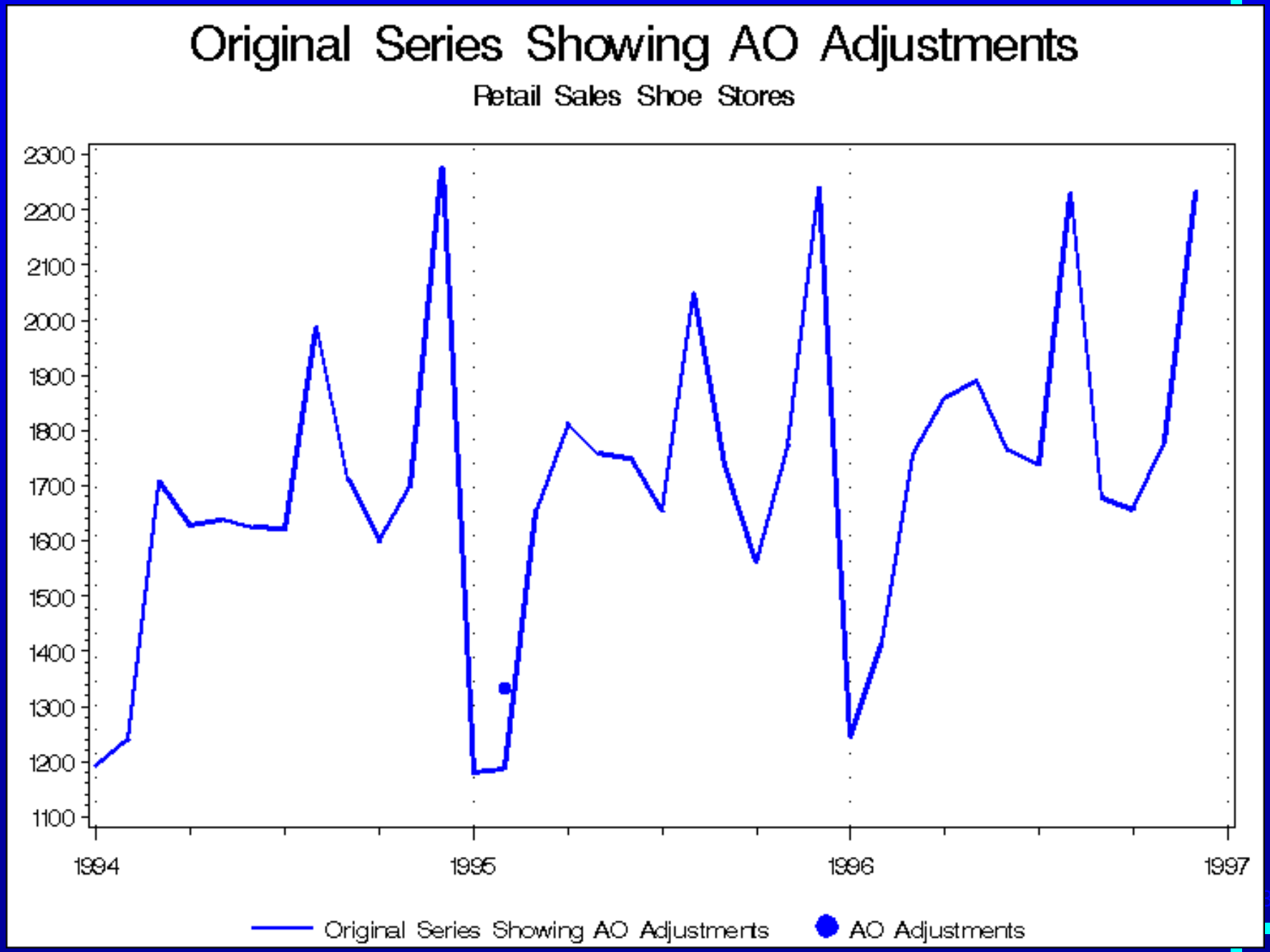
- Automatic detection available for AO, LS, TC

Additive Outlier (AO)

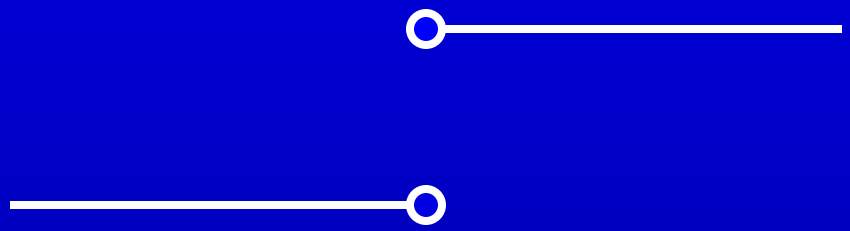


(point outlier)

AO*yyyy.mm*
(ao1989.9 or ao1989.09 or ao1989.Sep)


Original Series Showing AO Adjustments
Retail Sales Shoe Stores
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1994
1995
1996
1997
Original Series Showing AO Adjustments
AO Adjustments

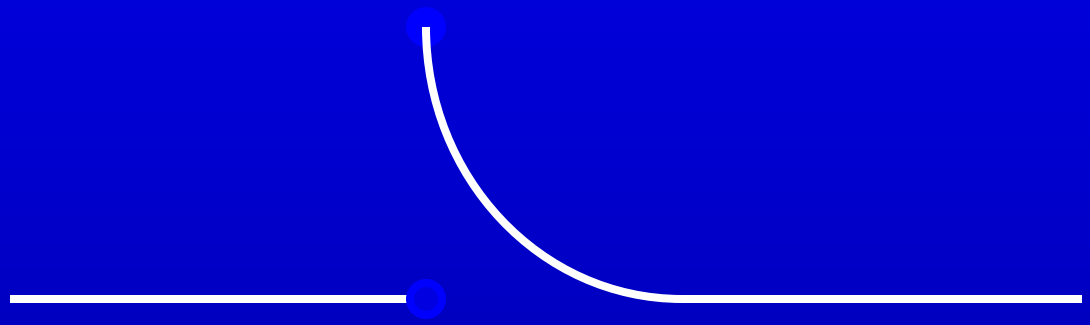
Level Shift (LS)



LS*yyyy.mm*

(ls1989.4 or ls1989.04 or ls1989.Apr)

Temporary Change (TC)



TC*yyyy.mm*

(tc1989.4 or tc1989.04 or tc1989.Apr)

Temporary Change Regressor

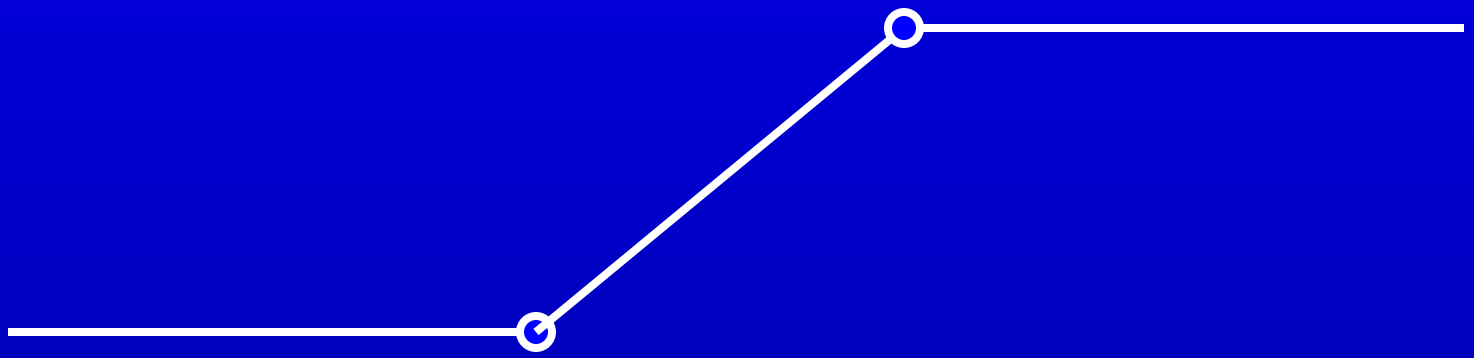
Temporary change at t_0

TC regressor

$$\begin{cases} 0 & \text{for } t < t_0 \\ \alpha^{t - t_0} & \text{for } t \geq t_0 \end{cases}$$

where α is the rate of decay back to the previous level, $0 < \alpha < 1$ (default: 0.7 for monthly and 0.343 for quarterly series)

Ramp



Rpyyyy.mm

(rp1999.09-2000.01 or rp1999.9-2000.1 or rp1999.Sep-2000.Jan)

Ramp Regressor

Ramp at t_0 through t_1

Ramp regressor

$$\begin{cases} -1 & \text{for } t \le t_0 \\ (t - t_0) / (t_1 - t_0) - 1 & \text{for } t_0 < t < t_1 \\ 0 & \text{for } t \ge t_1 \end{cases}$$

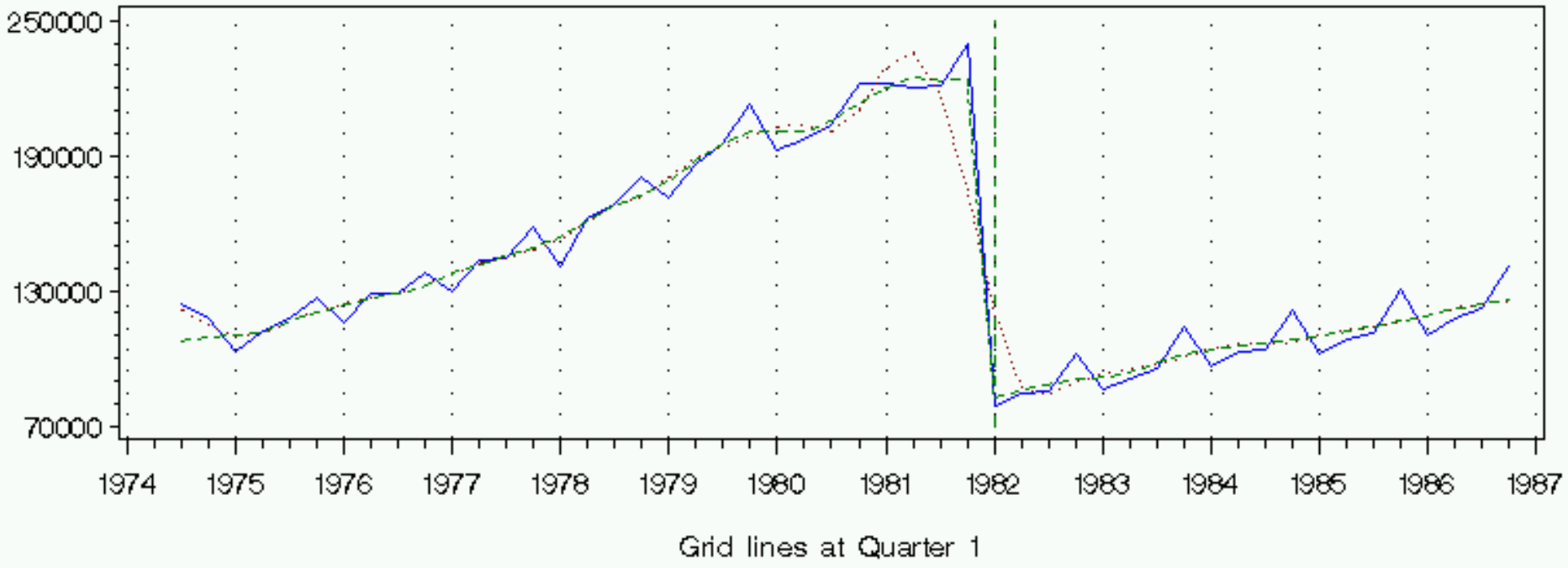
Outliers in Seasonal Adjustment

- Outlier effects removed from series before seasonal adjustment procedure – but they come back after the seasonal component is calculated

- AOs, TCs assigned to Irregular
- Level shifts and ramps assigned to Trend

Original Series and Trend

S30000 – Total Net Sales and S30000 – Total Net Sales



Grid lines at Quarter 1

S30000 – Total Net Sales: Original Series Trend

S30000 – Total Net Sales: Trend

Vertical Lines mark dates of Level Shifts, Temporary Changes, and the beginnings and endings of Ramps.

Calendar Effects

- Trading or Working Day:
  - » Effects related to:
    - Which weekdays (Mondays,…, Sundays) occur five times in a month
    - Effects of variable length of February.

August 2007

Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	***29***	***30***	***31***	

August 2008

Sun	Mon	Tue	Wed	Thu	Fri	Sat
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	***29***	***30***
31						

Flow Trading Day – Six Coefficients

TD_{1t} = (# of Mondays) – (# of Sundays)

TD_{2t} = (# of Tuesdays) – (# of Sundays)

. . .

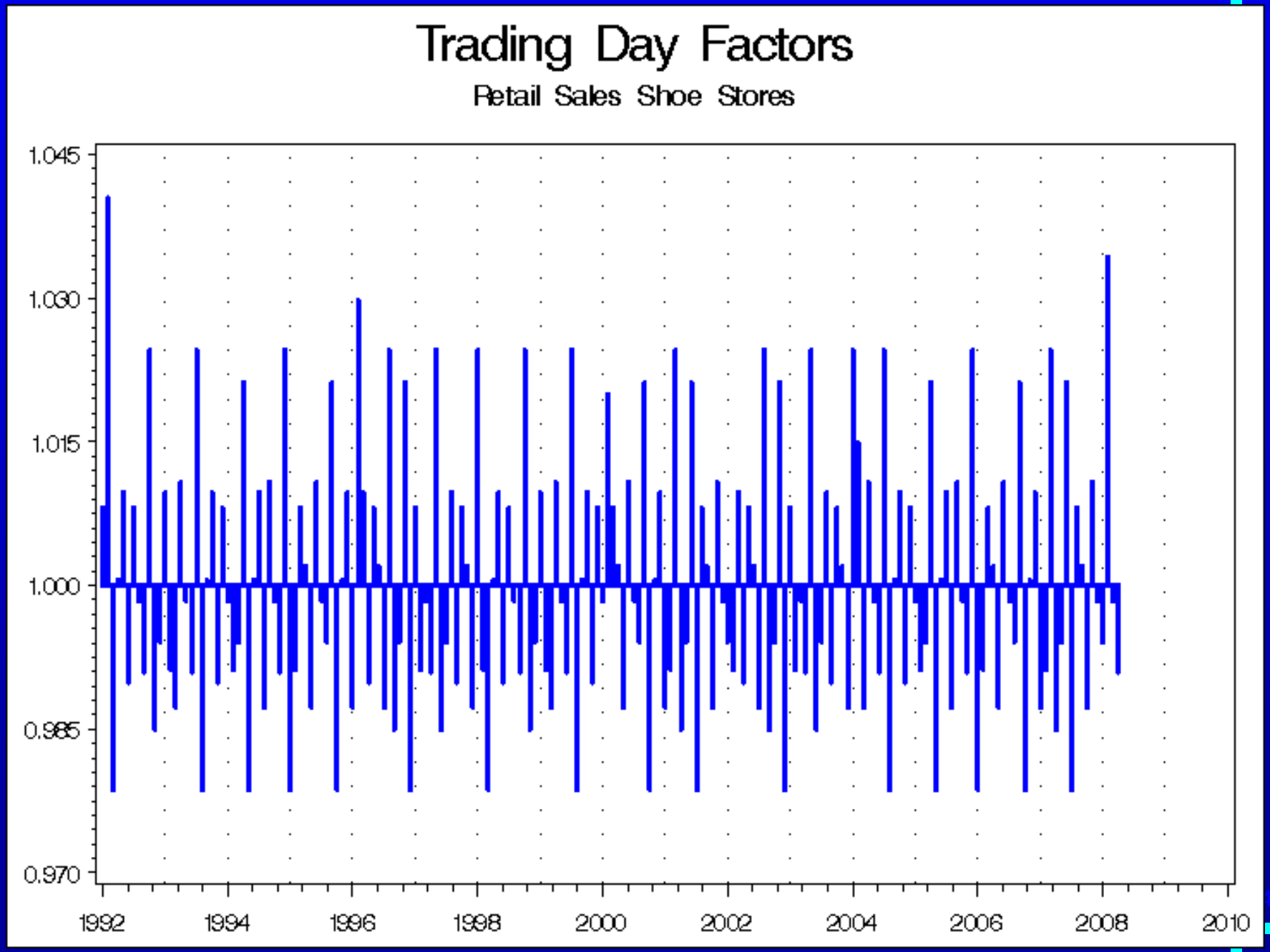
TD_{6t} = (# of Saturdays) – (# of Sundays)

(Regressors are –1, 0, 1)

Six regressors because seven regressors are constrained to sum to zero to avoid long-term adjustments to the series level – see Bell and Hillmer (1983)

Additional Flow Trading Day Effects

- Leap-year adjustment
  - » Nonseasonal component of the length-of-month effect
  - » Prior adjustment factor (multiplicative)
  - » Regression effect (additive)


Trading Day Factors
Retail Sales Shoe Stores
1.045
1.030
1.015
1.000
0.985
0.970
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010

Calendar Effects (Continued)

- Moving Holidays:
  - » Effects of holidays
    - With changing dates
    - Which can impact more that one month in a way that depends on the date.
  - » Examples : Easter, Chinese New Year, Ramadan, etc

Easter Effect (Moving Holiday)

- Effect (often an elevated level) occurring for w days, ending on the day before Easter
  - » $1 \leq w \leq 25$

Example

- If Easter is April 3 and the effect is 8 days long
  - » 2/8 of effect in April (2nd Quarter)
  - » 6/8 of effect in March (1st Quarter)

Easter [w]

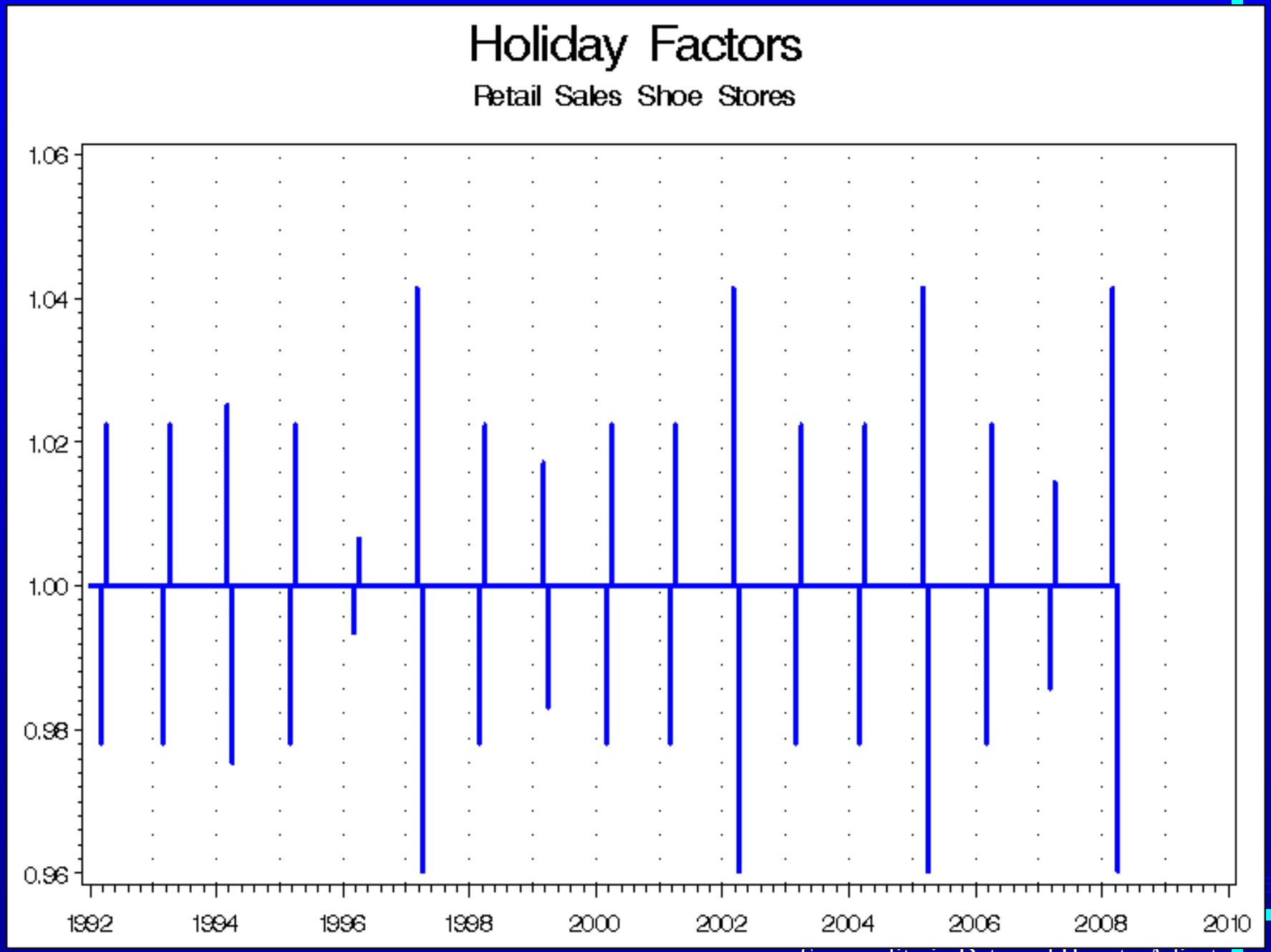
$$E_{wt} = \frac{W_t}{w} - \overline{E_{wt}}$$

w = length of the Easter effect (how many days before Easter)

W_t = number of the w days in month/quarter t

$\overline{E_{wt}}$ = long-term means of W_t / w

$E_{wt} > 0$ only in Mar and April (maybe Feb) (1st and 2nd quarter)


Holiday Factors
Retail Sales Shoe Stores
1.06
1.04
1.02
1.00
0.98
0.96
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010

Deciding to Model Trading Day or Easter Effects

- Likelihood tests
- Significance tests
- Forecast comparisons
- Spectrum diagnostics (trading day only)

X-11 Approach to Seasonal Adjustment

- Use moving average filters to identify and remove the seasonal pattern
- X-11 method has been in use since the 1950s

Step-by-Step Through X-11

Seasonal Adjustment With the X-11 Method

by Dominique Ladiray and Benoît Quenneville (2001)

Springer-Verlag Lecture Notes in Statistics 158

X-12-ARIMA Tables/Iterations

A. Prior adjustments before the core X-11 procedures
B. Preliminary estimation of Seasonal, Trend, and extreme values
C. Intermediate estimation of Seasonal and Trend, final estimation of extreme values
D. Final estimation of Seasonal, Trend, Irregular

Iterative Refinement in X-11

1. Estimate simple trend
2. Remove the trend from the series
3. Estimate seasonal factors
4. Calculate seasonally adjusted series

Repeat Steps 1 — 4

Re-estimate trend and irregular

Complete the procedure three times (B, C, D)

Step 1 – Estimate Simple Trend

- "Centered 12-term moving average"
  = 2x12 moving average
  Estimate for February 2004:

$$\frac{2003.8 + 2003.9 + \ldots + 2004.2 + \ldots + 2004.7}{+ 2003.9 + \ldots + 2004.2 + \ldots + 2004.7 + 2004.8}{24}$$

Step 2 – Remove Trend Estimate From Prior-adjusted Series

- SI-Ratio
- Prior-adjusted series with estimated trend removed
  - » "Detrended" series

$$SI_t = S_t \times I_t = \frac{Y_t^{adj}}{C_t}$$

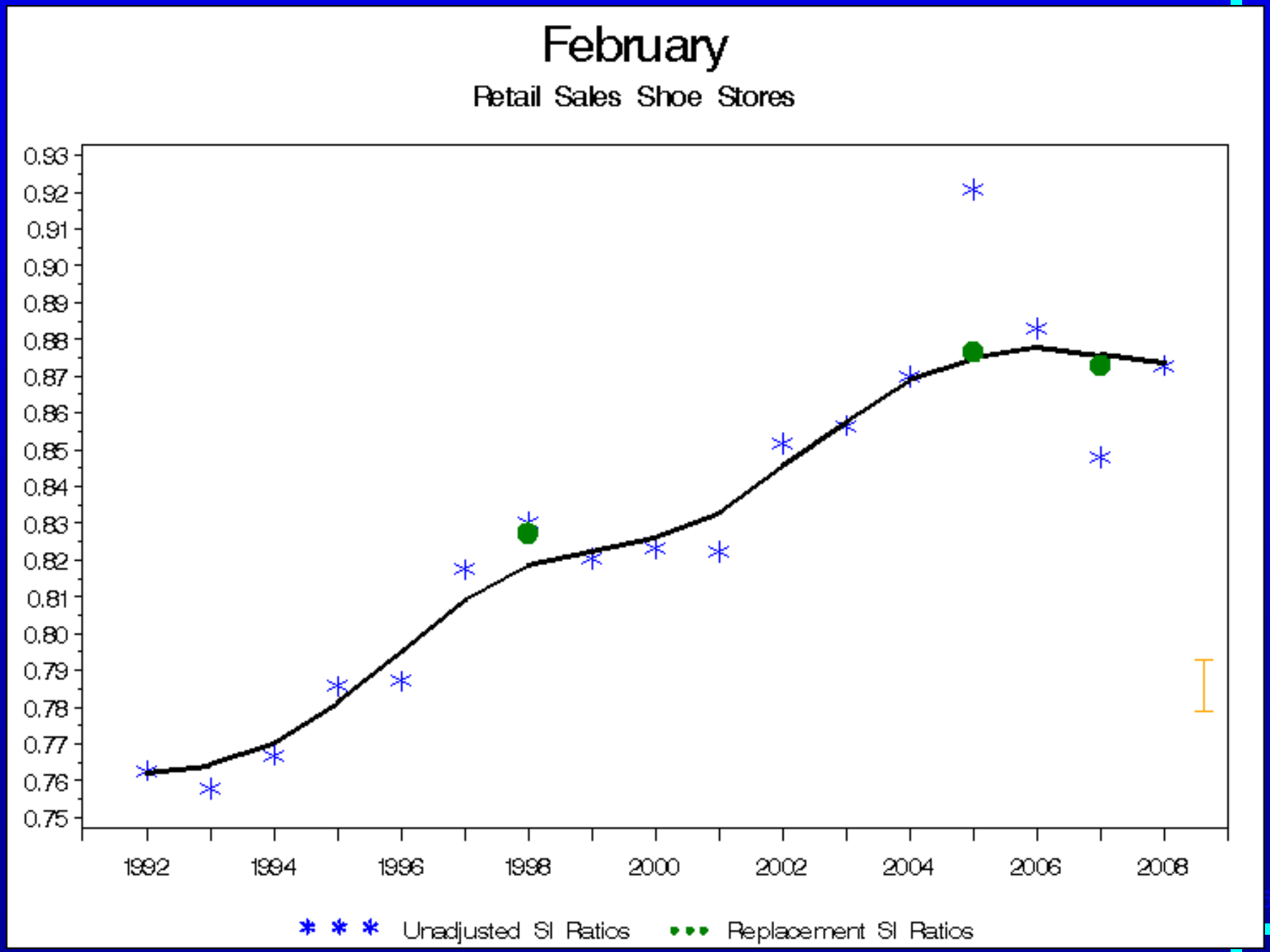
Multiplicative

$$SI_t = S_t + I_t = Y_t^{adj} - C_t$$

Additive

Extreme Values

- In practice, the detrended series needs robustifying
- Smooth the robustified detrended series to calculate seasonal factors


February
Retail Sales Shoe Stores
0.93
0.92
0.91
0.90
0.89
0.88
0.87
0.86
0.85
0.84
0.83
0.82
0.81
0.80
0.79
0.78
0.77
0.76
0.75
1992
1994
1996
1998
2000
2002
2004
2006
2008
Unadjusted SI Ratios
Replacement SI Ratios

Step 3 – Estimate Seasonal Factors

- Average of robust SI ratios
  - » Average Months or Quarters
- Monthly series, 3x3 moving average filter
  - » Other filters possible

$$S_t^{3\times3} =$$

$$\frac{1}{9} SI_{t-24} + \frac{2}{9} SI_{t-12} + \frac{1}{3} SI_t + \frac{2}{9} SI_{t+12} + \frac{1}{9} SI_{t+24}$$

Step 4 – Estimate Seasonally Adjusted Series

Multiplicative $$A_t = \frac{Y^*_t}{S_t}$$

Y* includes outlier effects but not trading day, holiday effects

Additive $$A_t = Y^*_t - S_t$$

Final Trend From Henderson Filter

- Average consecutive values
- Chosen Henderson filter depends on the ratio of Irregular to changes in Trend (I/C)
  - » More irregular, more smoothing
  - » More change in the trend, less smoothing

Final Irregular

- Calculate final irregular component by removing the final trend from the final seasonally adjusted series

Multiplicative $$I_t = \frac{A_t}{C_t}$$

Additive $$I_t = A_t - C_t$$

Seasonal Adjustment Diagnostics Overview

- Spectrum diagnostics
- Revisions history
- Sliding spans

What's Next?

X-13ARIMA-SEATS =

X-13A-S =

X-12-ARIMA + SEATS

What is X-13A-S?

- Users can choose between model-based seasonal adjustments from SEATS and non-parametric adjustments with X-11.

- Collaboration between the U. S. Census Bureau and the current developers of SEATS, **Agustin Maravall** of the Bank of Spain and **Gianluca Caporello**.

Why X-13A-S?

- Allows users to
  - » generate X-11 and SEATS seasonal adjustments using the same interface
  - » compare X-11 and SEATS seasonal adjustments using a common set of diagnostics

On the Web:

TRAMO/SEATS website

www.bde.es/servicio/software/econome.htm

X-12-ARIMA website

www.census.gov/srd/www/x12a

Contact Information

brian.c.monsell@census.gov

Brian Monsell

U.S. Census Bureau

SRD, Room 5K018

Washington DC 20233

Disclaimer

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Graphics appearing in this talk were generated using X-12-Graph (Hood 2002, Lytras 2006) using Base SAS® software, SAS/AF® software, and SAS/GRAPH® software, Versions 8 and 9 of the SAS System for Windows. Copyright © 1999-2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Spectral graphs

Spectrum graphs are useful to assess the cycles in the time series.

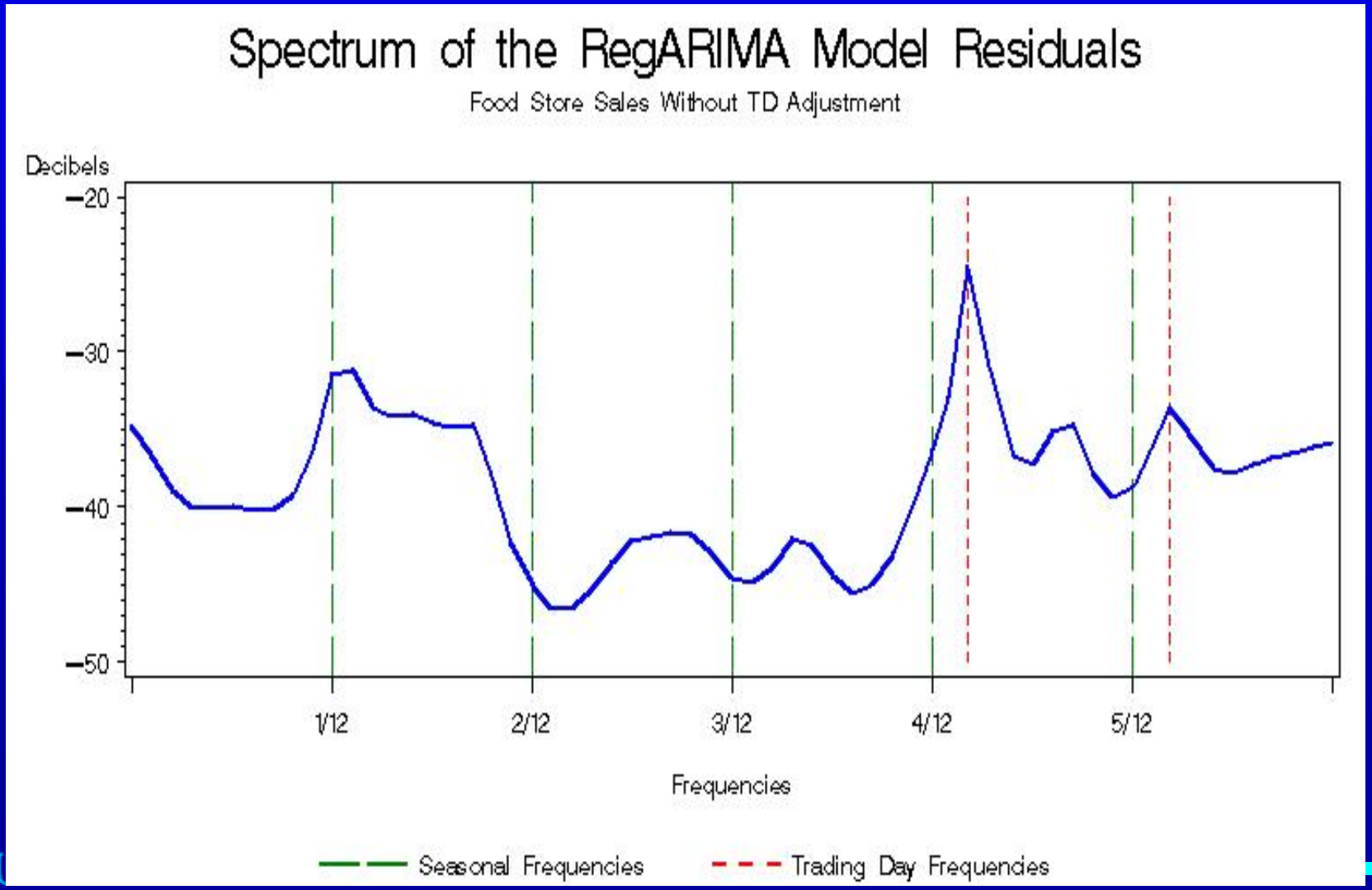
Seasonal frequencies

- Occur at 1/12, 2/12, 3/12, 4/12, 5/12, 6/12
- Prominent spectrum peak at these frequencies usually indicates that a seasonal effect is present.

Trading day frequencies

- Occur at 0.348 and 0.432
- Prominent peaks at this frequency suggest that a trading day effect is present.

Food Sales: RegARIMA (no TD)
TD Peaks at .348 and .412



History diagnostics

- How to measure stability of the seasonal adjustment?
  - » General approach is to do many seasonal adjustments on a sequence of increasing data spans, at a new time point each time.
  - » Allows an assessment of how the parameters can change over time and the magnitude of revisions

Types of History Analysis

- Revision Histories
  - » evaluation and comparison of competing adjustments
- Out-of-Sample Forecast Errors
  - » comparison of competing models (e.g. different TD decisions)
- AIC History
  - » evaluate the stability of AIC's choice
  - » compare competing data adjustments to past data

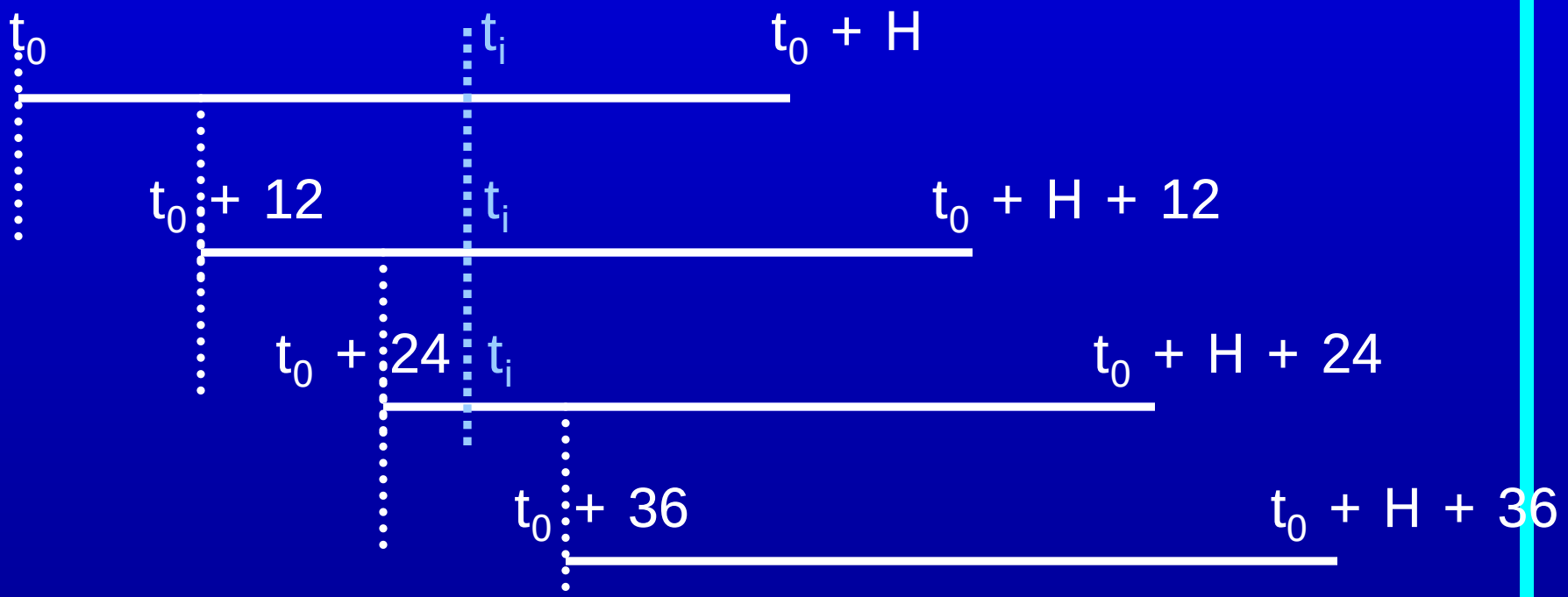
Sliding Spans Diagnostics

- Can also be used to look at the stability of the seasonal adjustment.
- Compares seasonal adjustments from overlapping sub spans of the time series.
- Reference:
  - » Findley, Monsell, Shulman and Pugh (1990) "Sliding Spans Diagnostics for Seasonal and Related Adjustments" Journal of the American Statistical Association 85, 345-355.

Sliding Spans (continued)

- Construct four spans of length H where H depends on the seasonal moving average
- Each span starts one year after the previous span
- Compare any adjustments that occur in more than one span

Sliding Spans



Sample Sliding Spans - Seasonal Factors

Month	Span 1	Span 2	Span 3	Max % Diff.
1990-11	109.7	108.8	108.2	1.4
1990-12	113.0	111.5	110.3	2.0
1991-01	74.6	73.9	72.9	2.3
1991-02	84.0	80.7	76.1	10.3