RegARIMA Modeling: Regression Effects

Seasonal Adjustment With X-13ARIMA-SEATS

2019

Economic Statistical Methods Division

U.S. Census Bureau

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

Outline

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

Regression Effects

$$\log\left(Y_t \,/\, D_t \right) = \boldsymbol{\beta}' \mathbf{X}_t + Z_t$$

Transformation(s)

ARIMA Process

Regression

$X_t =$ 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.

Regression Effects Available in X-13ARIMA-SEATS

- Constant Term
- Outlier Effects
  - Additive (or Point) Outliers
  - Level Shifts
  - Temporary Level Shifts
  - Temporary Changes
  - Ramps
  - Seasonal Outliers

Regression Effects Available in X-13ARIMA-SEATS (2)

- Seasonal Effects
  - Calendar month indicators*
  - Trigonometric Seasonal (Sines-Cosines)*
- Calendar Effects
  - Trading Day (Flow or Stock)*
  - Leap-year February*, Length of Month*
  - Moving Holidays (e.g. Easter)

* Two-regime option available (to test for pattern changes)

Regression Effects Available in X-13ARIMA-SEATS (3)

- User-Defined Effects
  - For effects that are not built-in
  - Often used for holidays not celebrated widely in the United States, such as Ramadan, Chinese New Year, Easter Monday, etc.
  - For complicated or new situations, such as the Super Bowl or moving holidays with before and after effects, etc.

Note: Regression coefficients can be fixed, but generally we choose to re-estimate them

Regression Specs in X-13ARIMA-SEATS

- **`Regression`**
  - Specify regressors with **`variables`** argument
  - Specify user regressors
  - Test for trading day/holiday
- **`Outlier`**
  - Automatic identification of some outlier types

The Regression Matrix and Differencing

- If the series is differenced, the regression is fit with the differenced data. The regression matrix is also differenced:

$$\Delta Y_t = \beta'(\Delta X_t) + \Delta Z_t$$

- So regressors can't be used if they're annihilated by differencing.

Calculating Regression Factors: Log Transformed Series

- A RegARIMA model is $\log(y_t) = \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \log(Z_t)$ where Z_t is an ARIMA process.
- So $y_t = \exp(\beta_1 X_{1,t}) \ldots \exp(\beta_k X_{k,t}) z_t$
- If $\beta_i X_{i,t}$ is small (<0.3 or so), then 100*$\beta_i X_{i,t}$ is the average percentage increase or decrease in y_t due to the *i*th regression effect.
- X-13A-S calculates and prints out outlier factors, trading day factors, holiday factors, etc, which are calculated by multiplying together the relevant $\exp(\beta_i X_{i,t})$.
- To see $X_1, \ldots, X_k$, the regression matrix, for your adjustment use **`save=rmx`** in the **regression{}** spec.

Constant Term

- Trend Constant

$$\log(Y_t/D_t) = \beta_0 + Z_t$$

- Usually needed only if no nonseasonal differencing in the ARIMA model
  - **`Automdl`** may add one when there is differencing – it's good practice to compare model without it

Regression Spec With Constant Term

```
regression{
  variables = const
}
```

Outlier Effects

- Additive (or Point) outliers (AO)
- Level shifts (LS)
- Temporary level shifts (TL)
- Temporary changes (TC)
- Ramps (RP, QI, QD)
- Seasonal outliers (SO)

Automatic Detection Available

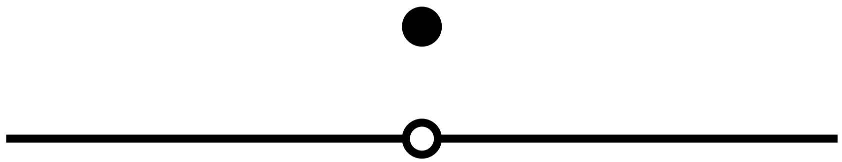
Default types

- Additive (or Point) outliers (AO)
- Level shifts (LS)

Also possible to detect, if specified

- Temporary changes (TC)

Additive Outlier (AO)



(point outlier)

AO*yyyy.mm*

(ao1989.9 or ao1989.09 or ao1989.Sep)

Additive Outlier Regressor

Additive outlier (point outlier) at t_0

AO regressor

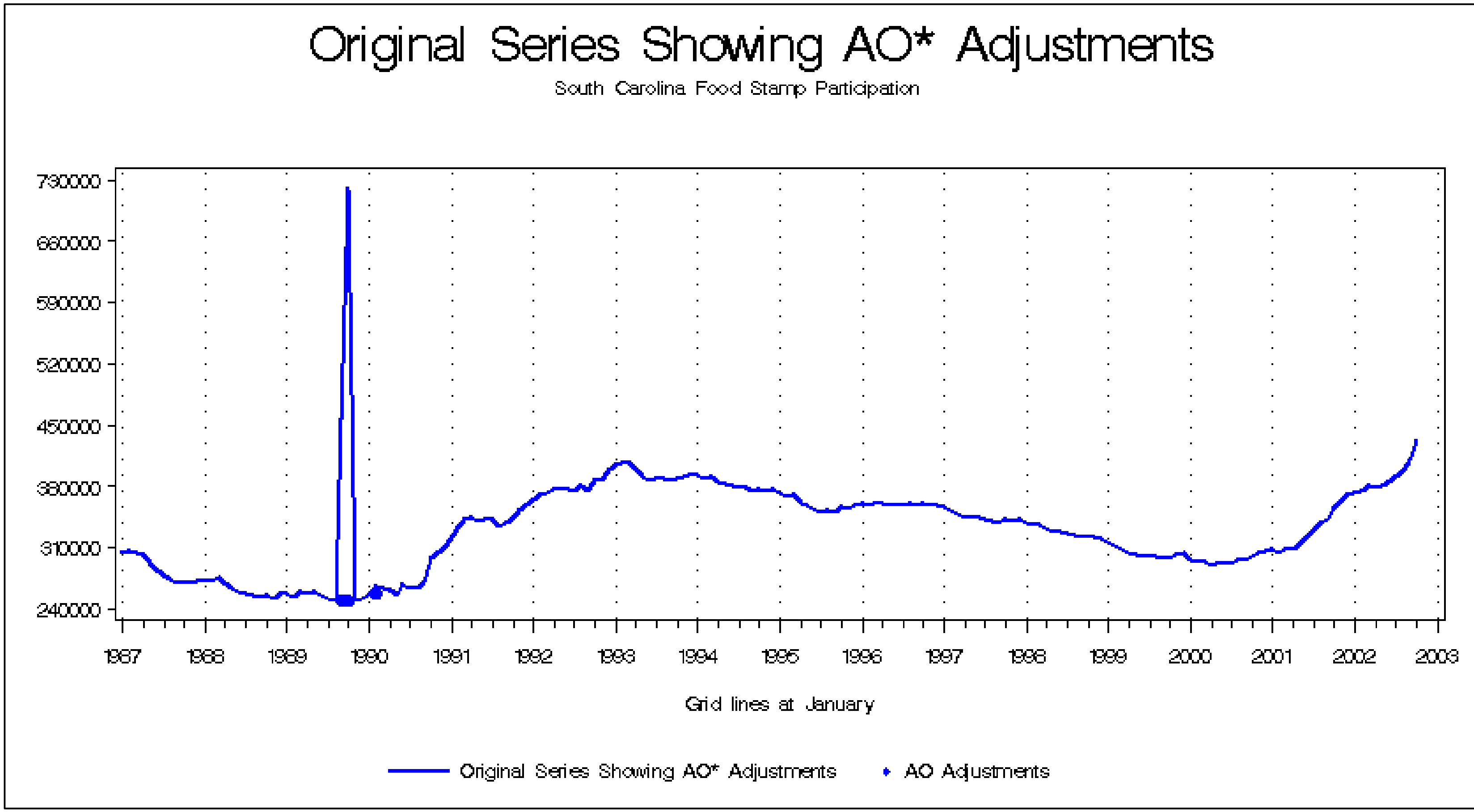
$$\begin{cases} 1 \text{ for } t = t_0 \\ 0 \text{ for } t \neq t_0 \end{cases}$$

South Carolina SNAP Recipients

Hurricane Hugo

September 1989

After natural disasters, emergency Supplemental Nutrition Assistance Program cards (food stamps) are often issued with different eligibility requirements – victims of the disaster are more likely to be eligible


Original Series Showing AO* Adjustments
South Carolina Food Stamp Participation
730000
660000
590000
520000
450000
380000
310000
240000
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
Grid lines at January
Original Series Showing AO* Adjustments
AO Adjustments

South Carolina – 3 AOs

September 1989

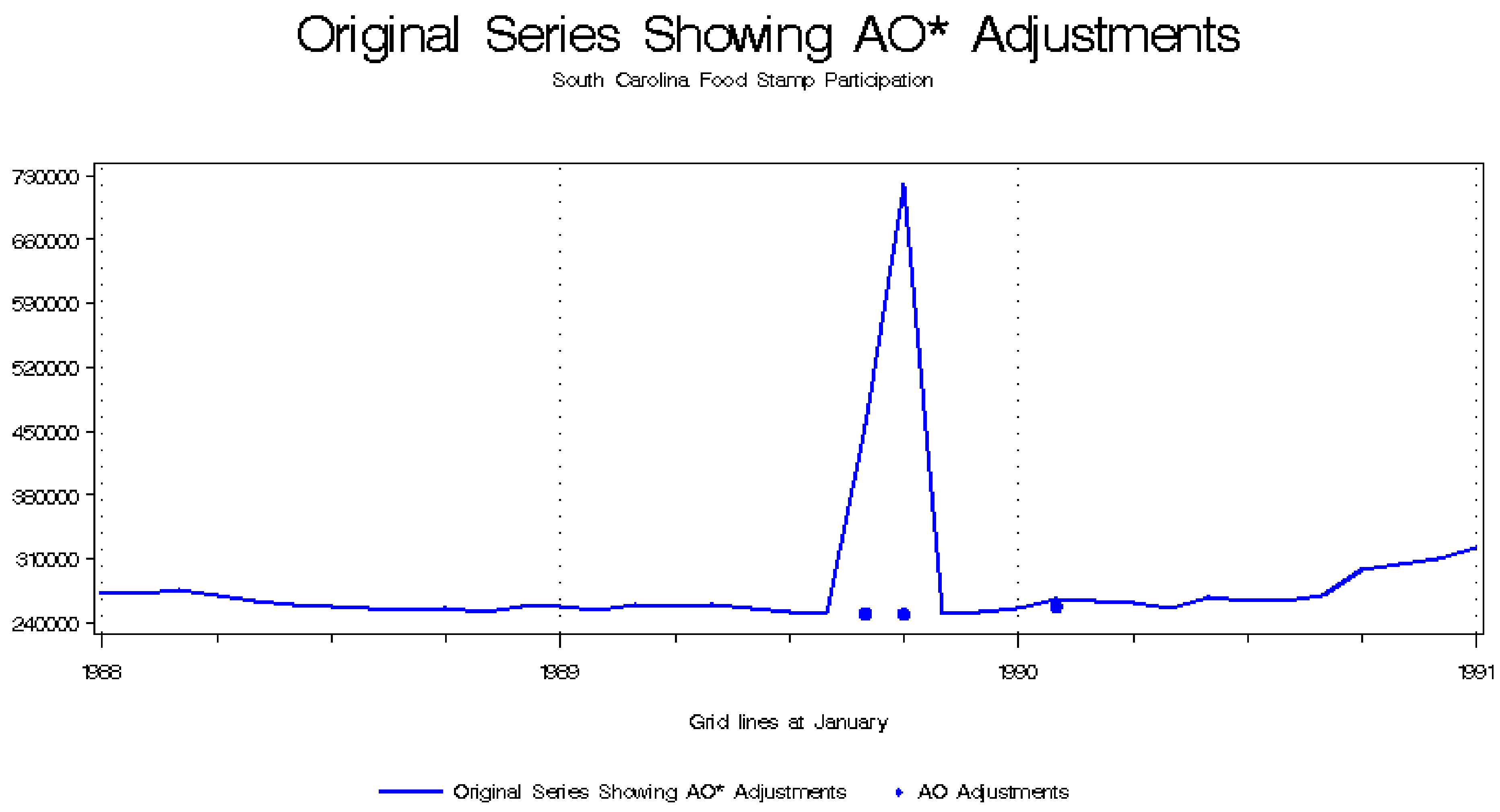
October 1989

February 1990

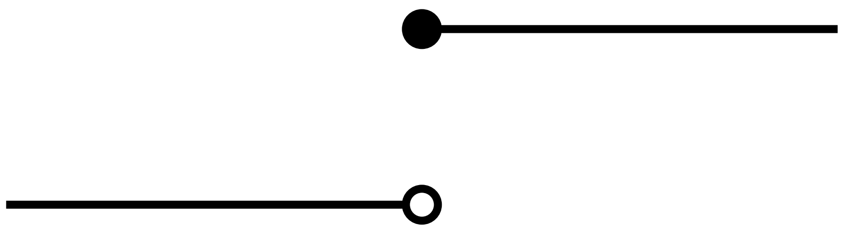
AO Parameters

Variable	Parameter Estimate	Standard Error	t-value
AO1989.Sep	0.6007	0.00589	101.92
AO1989.Oct	1.0614	0.00595	178.25
AO1990.Feb	0.0334	0.00520	6.43

. . .


Original Series Showing AO* Adjustments
South Carolina Food Stamp Participation
730000
660000
590000
520000
450000
380000
310000
240000
1988
1989
1990
1991
Grid lines at January
Original Series Showing AO* Adjustments
AO Adjustments

Level Shift (LS)



LS*yyyy.mm*

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

Level Shift Regressor

Level shift at t_0

LS regressor

$$\begin{cases} -1 & \text{for } t < t_0 \\ 0 & \text{for } t \geq t_0 \end{cases}$$

Level shift adjustment changes the past level(s) to match the current level

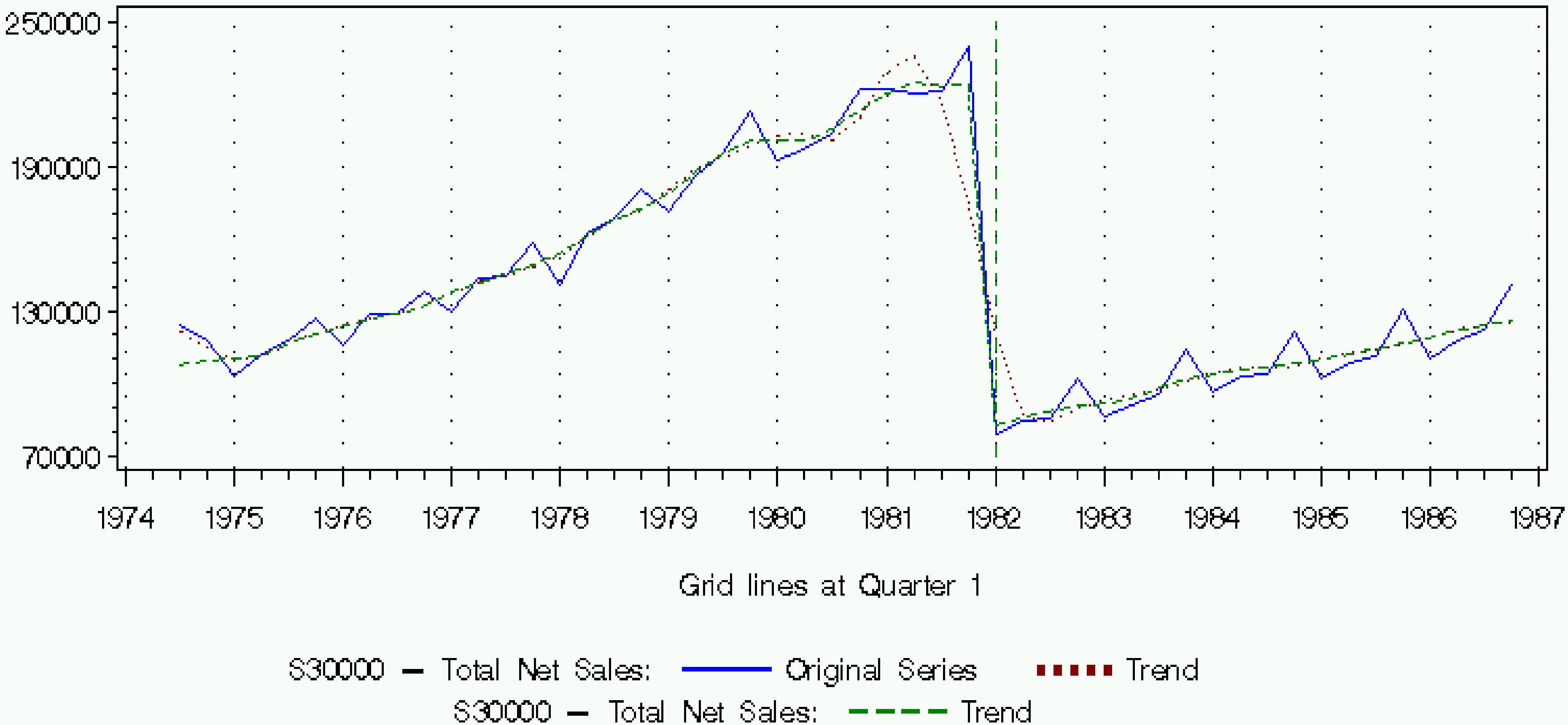
Quarterly Financial Report

Paperwork Reduction Act

Implemented 1st quarter 1982

Original Series and Trend

S30000 – Total Net Sales and S30000 – Total Net Sales



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

Canceling Level Shifts

- Also called temporary level shifts
  - NOT temporary changes!
- 2 (or more) level shifts whose effects cancel
- Can replace them with additive outliers (over short spans) or use a temporary level shift regressor
- More details in last set of class slides

Temporary Level Shift

TL*yyyy.mm-yyyy.mm*

(TL1989.4-1989.6

or

TL1989.Apr-1989.Jun)

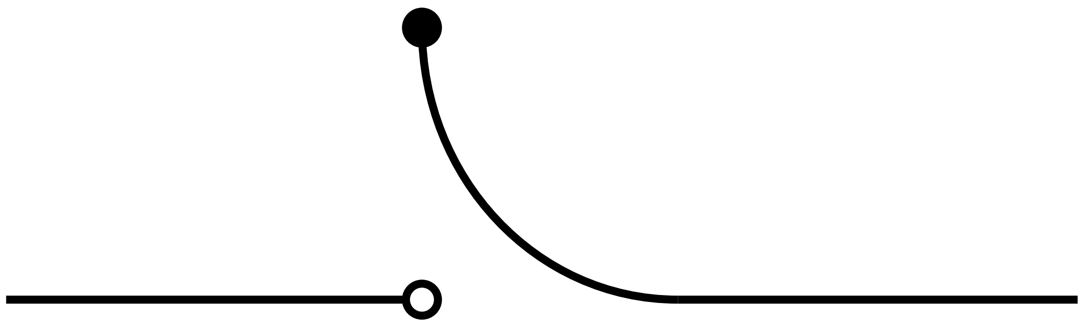
Temporary Level Shift Regressor

TL at t_0 through t_1

Temporary Level Shift regressor

$$\begin{cases} 0 & \text{for } t \leq t_0 \\ 1 & \text{for } t_0 < t < t_1 \\ 0 & \text{for } t \geq t_1 \end{cases}$$

Temporary Change (TC)



TC*yyyy.mm*

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

Temporary Change Regressor

Temporary change at t_0

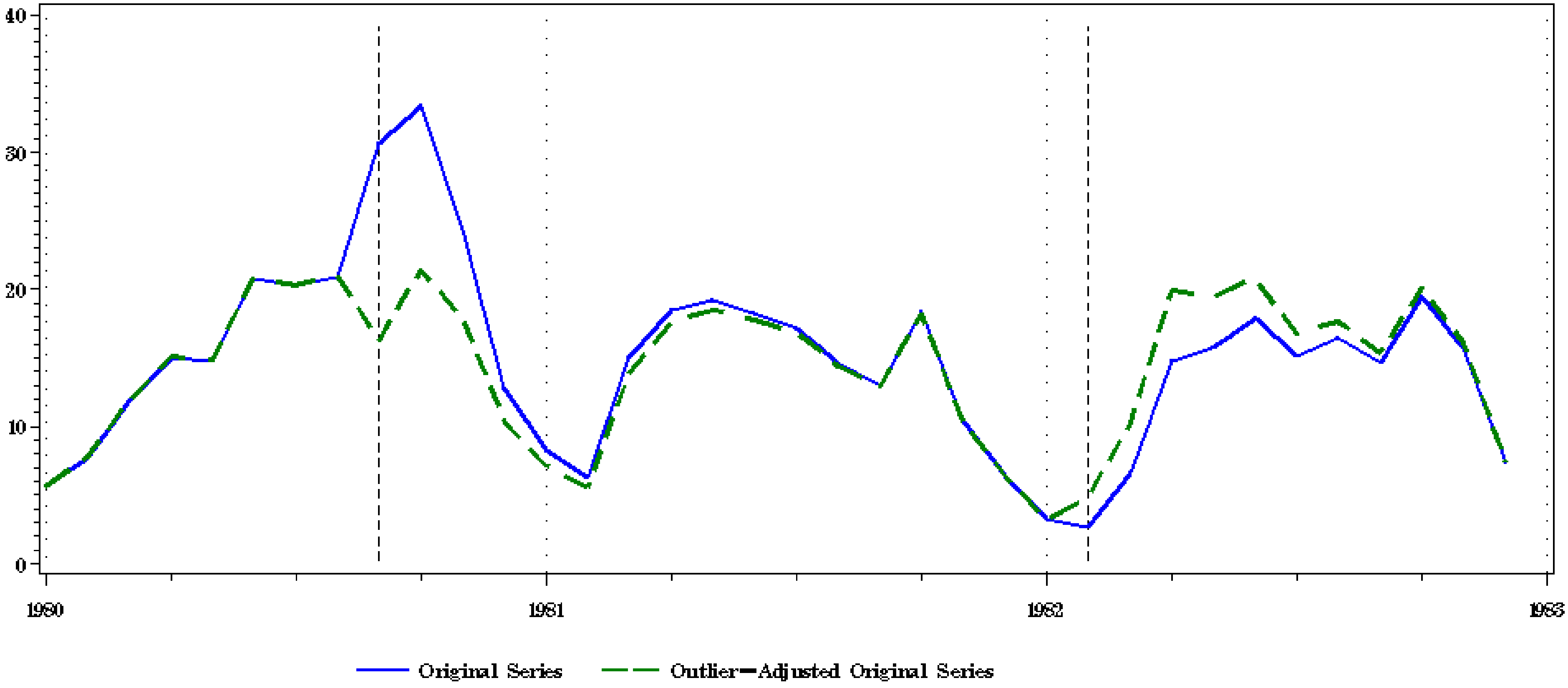
TC regressor

$$\begin{cases} 0 & \text{for } t < t_0 \\ \alpha^{t - t0} & \text{for } t \geq t_0 \end{cases}$$

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)

Original Series and Outlier–Adjusted Original Series

Starts MW Total



Seasonal Outlier (SO)

- SO*yyyy.mm* – can be used when the seasonal pattern shifts in one month.

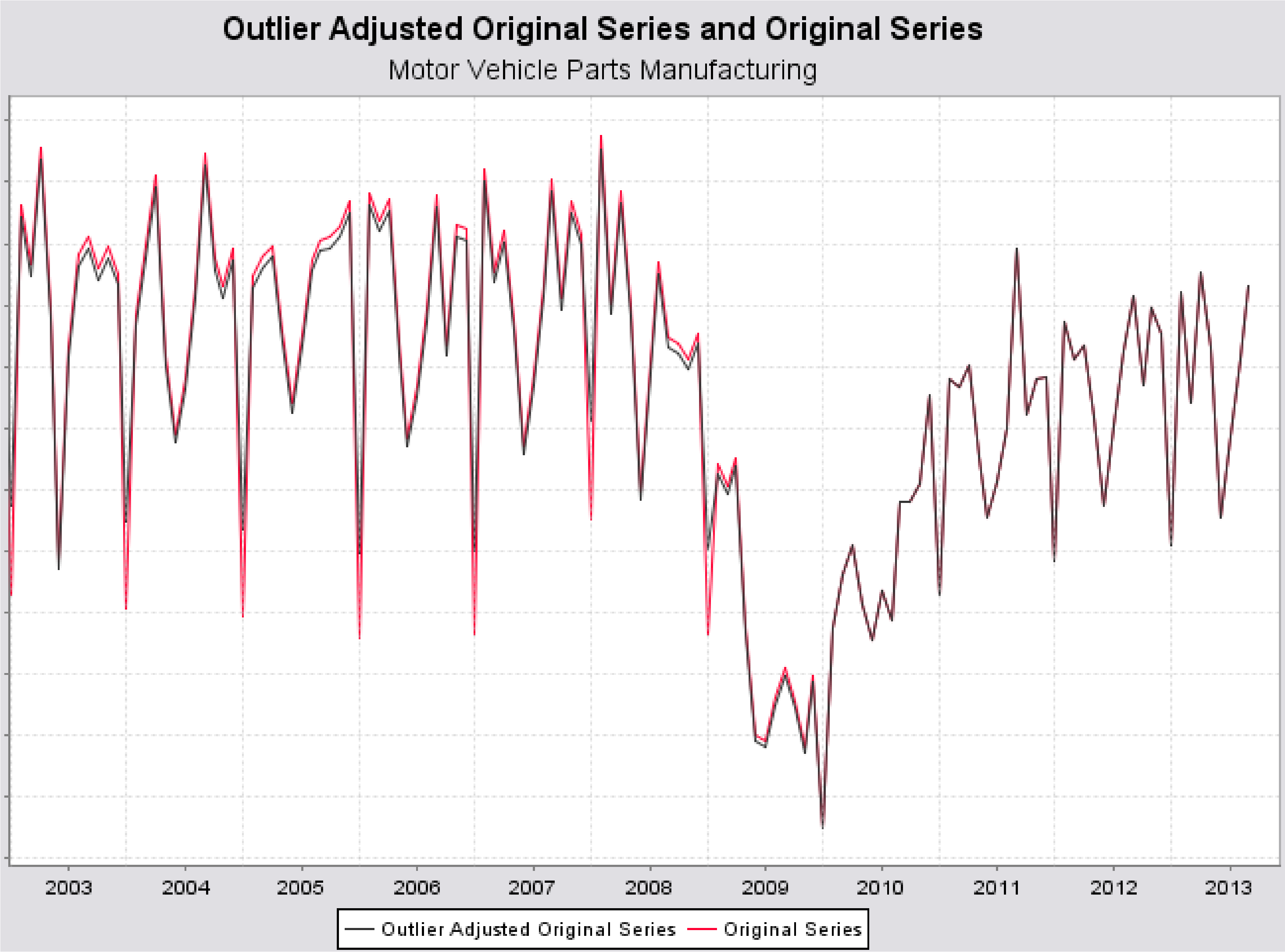
- The built-in regressor shifts the level of the changed month, and also all other months to compensate. This may not be the best choice for your series.

Seasonal Outlier Regressor

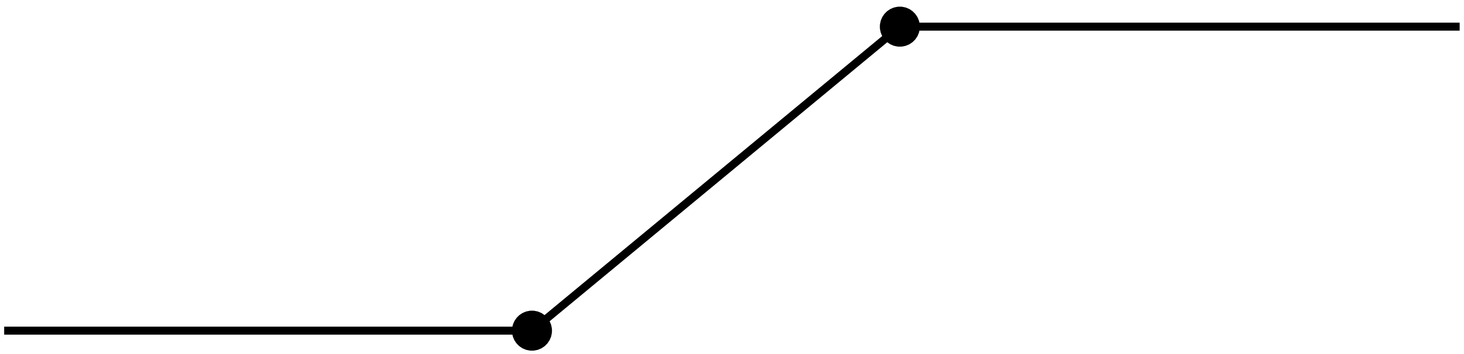
SO at t_0:

Seasonal Outlier regressor

$$
\begin{cases}
0 & \text{for } t \geq t_0 \\
1 & \text{for } t < t_0 \text{, } t \text{ same month as } t_0 \\
-1/(s-1) & \text{otherwise}
\end{cases}
$$


Outlier Adjusted Original Series and Original Series
Motor Vehicle Parts Manufacturing
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Outlier Adjusted Original Series
Original Series

Ramp



RP*yyyy.mm-yyyy.mm*

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

Ramp Regressor

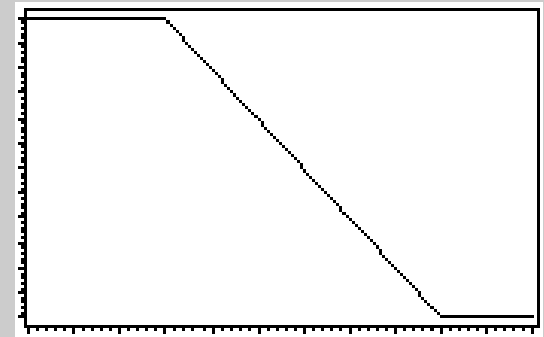
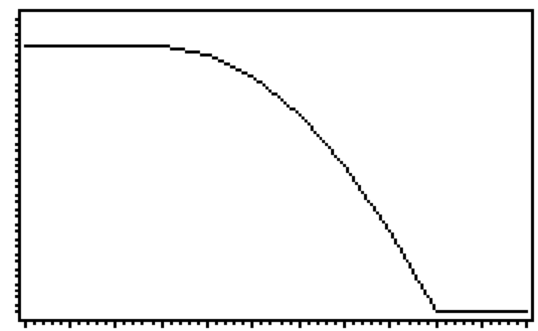
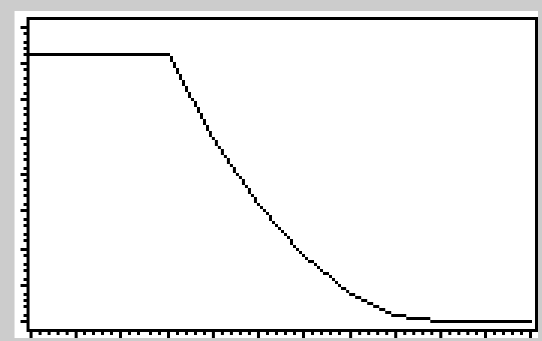
Ramp at t_0 (start date) through t_1 (end date)

Ramp regressor

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

Quadratic Ramps

- Similar to linear ramps, but fit a quadratic pattern
- QI*yyyy.mm-yyyy.mm*: effect increases in slope
- QD*yyyy.mm-yyyy.mm*: effect decreases in slope

Ramp Effect	**Regression Variable**	**Graph of 6 Month Decline**
Linear Ramp	$RP_t^{(t0,t1)} = \begin{cases} t_0 - t_1 & for\ t \leq t_0 \\ t - t_1 & for\ t_0 < t < t_1 \\ 0 & for\ t \geq t_1 \end{cases}$	
Increasing Quadratic Ramp	$QI_t^{(t0,t1)} = \begin{cases} -(t_1 - t_0)^2 & for\ t \leq t_0 \\ (t - t_0)^2 - (t_1 - t_0)^2 & for\ t_0 < t < t_1 \\ 0 & for\ t \geq t_1 \end{cases}$	
Decreasing Quadratic Ramp	$QD_t^{(t0,t1)} = \begin{cases} -(t_1 - t_0)^2 & for\ t \leq t_0 \\ -(t_1 - t)^2 & for\ t_0 < t < t_1 \\ 0 & for\ t \geq t_1 \end{cases}$	

Using Ramps

- At Census Bureau, we use ramps as interventions, to handle effects where we know there is a reason for a decrease/increase – the recent recession
- For some guidance on using ramps, “Modeling Recession Effects and the Consequences on Seasonal Adjustment” by Lytras & Bell, 2013: www.census.gov/ts/papers/jsm2013lytrasfinal.pdf

Outlier Sequences

- AOS*yyyy.mm-yyyy.mm* and LSS*yyyy.mm-yyyy.mm* place an AO or an LS at every point from the start date to the end date
- Can specify a critical value with
  `tlimit = ##` in the **`regression`** spec;
  if the absolute value of the t-statistic of an outlier in the sequence is less than the given critical value, X-13A-S removes the outlier from the sequence

`Outlier` Spec Use

- Automatically identify outliers
  - Additive (point) outliers (AO)
  - Level shifts (LS)
  - Temporary changes (TC)
- By default AO and LS are identified, but can request to identify TC
  - **`types = (ao ls)`** is the default
  - **`types=all`** *or* **`types = (ao ls tc)`** identifies AO, LS, TC

Automatic Identification

- X-13ARIMA-SEATS calculates t statistics for every possible outlier type for every point in the outlier span

  *.fts file

- If the outlier absolute t value is greater than the outlier critical value, that outlier is selected
- By default, outliers added one at a time (**`method = addone`**) but there is an option for all outliers with a large enough t value to be added (**`method = addall`**)
- Backward elimination – if the t value of the added outlier regressor is lower than the critical value, the outlier is removed from the model
- Continue adding/eliminating until no new outliers to add

- For more details see the Reference Manual

Variable Outlier Critical Value

- Default outlier identification threshold depends on the length of the outlier span (Greta Ljung 1993)
- User can specify an outlier threshold with the **`critical`** argument
  - Higher values make it less likely to identify a value as an outlier
  - Each outlier type *can* have a different **`critical`** value (typically we use the same for all types)

Outlier Span Length	Default Critical Value
1	1.96
2	2.24
48	3.63
96	3.80
120	3.85
240	3.99
360	4.07

Outlier Spec Syntax

```
outlier{
  types = none | ao | ls | tc | all
#[ default: ( ao ls ) ]
  critical = value for outlier testing
  | ( < AO >, < LS >, < TC > )
#[ default: depends on length of span ]
  span = ( yyyy.mm, yyyy.mm )
#[ default values from series spec]
```

`Outlier` Spec Syntax (continued)

```
    lsrun = [ 0 up to 5 ]
#        number of successive
#        level shifts to test as
#        temporary [ default: 0 ]

    method = addone | addall
#          [ default: addone ]

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

`Outlier` Spec Example

```
outlier{
  types = ( ao ls tc )
  critical = 4.5
  span = ( 2007.Oct, )
}
```

If No Outliers Are Found

- Output gives the outlier regressor with the largest absolute t value
- Output also indicates “almost” outliers

OUTLIER DETECTION

Critical |t| for AO outliers : 3.94
Critical |t| for LS outliers : 3.94

"Almost" Outliers

	t(AO)	t(LS)
LS2005.Sep	-2.42	-3.80
LS2011.Jan	1.86	3.50

No AO and LS outliers identified
Largest outlier t-value : -3.80023 (LS2005.Sep)

If Outliers Are Found

ALWAYS, ALWAYS, ALWAYS

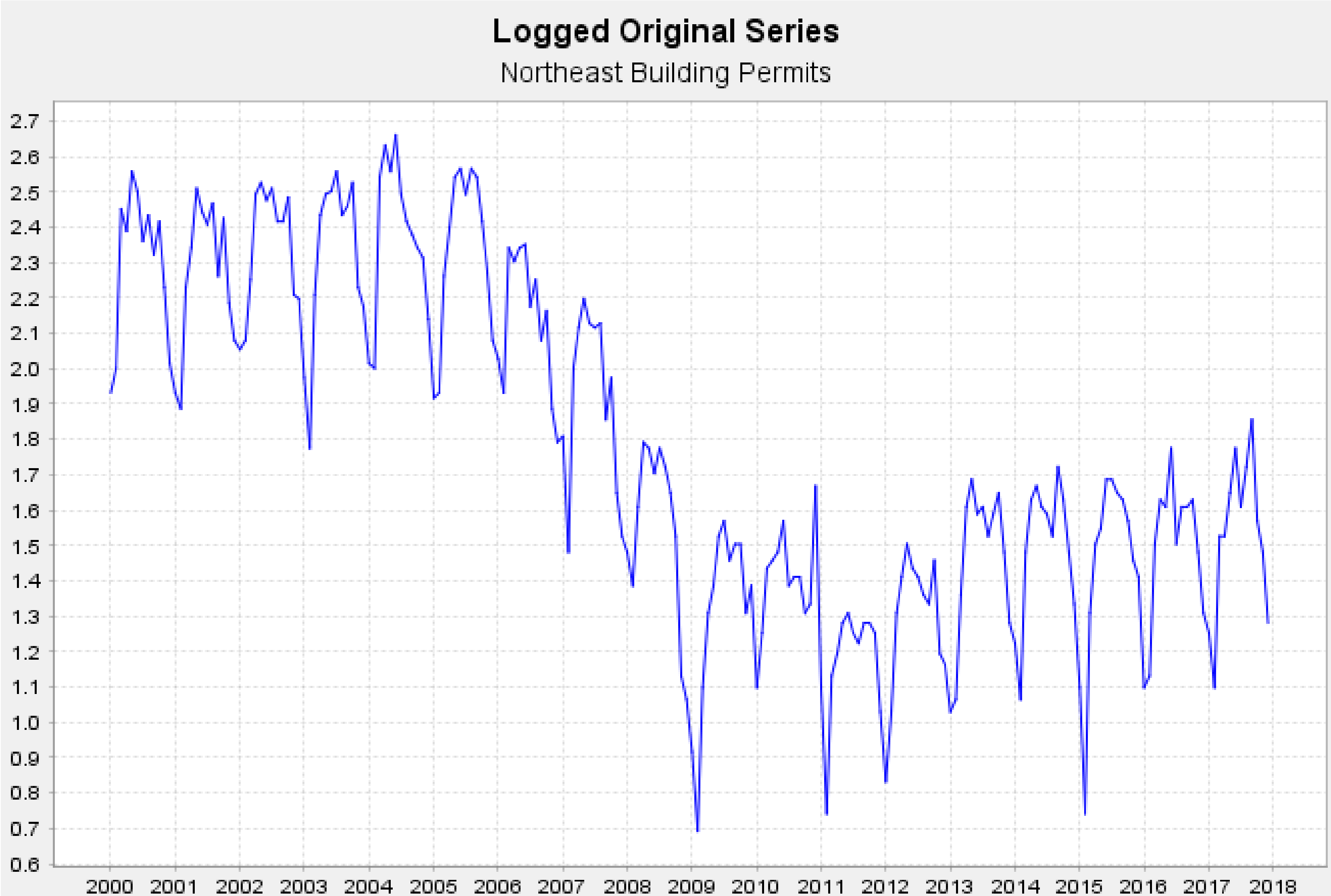
- Check if outliers are reasonable
- Hard-code known outliers
- Do not hard-code outliers for estimates that may be revised (if the estimates may change, they can't be known outliers)
  - Usually a concern only for the most recent part of the series

Outliers in Production

- Hard-code all known outliers in the historical part of the series (not going to be revised)
- Set outlier span for more recent/new values only
  - Remember shorter span lowers the default critical value
- Set a high enough critical value so newly identified outliers truly are significant
- Hard-code new outliers when verified

Example: Northeast Building Permits

```
series{ file = "bpne1.dat" format = datevalue
   title = "Northeast Building Permits" }
transform{ function = log }
regression{ variables = (td1coef) }
outlier{ types = (AO LS TC) }
arima{ model = (1 1 0)(0 1 1) }
forecast{ maxlead = 36 }
x11{ seasonalma = s3x5 }
```



Where do you think the outliers are?

Example: Northeast Building Permits Outlier Identification Results (1)

OUTLIER DETECTION

From 2000.Jan to 2017.Dec
Observations 216

Types : All types
Method : add one
Critical |t| for AO outliers : 3.97
Critical |t| for LS outliers : 3.97
Critical |t| for TC outliers : 3.97

Example: Northeast Building Permits Outlier Identification Results (2)

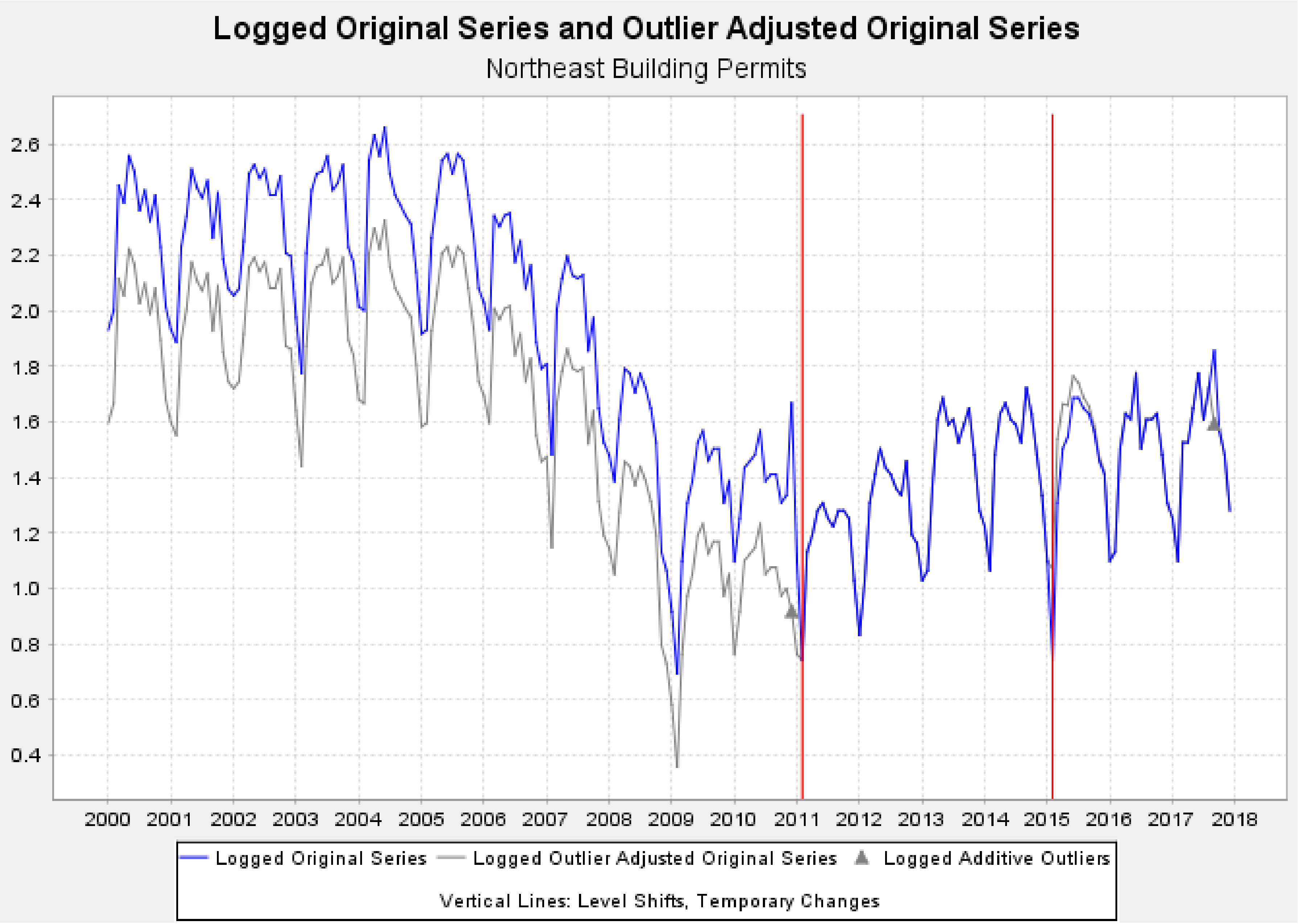
	Parameter Estimate	Standard Error	t-value
1-Coefficient Trading Day			
Weekday	0.0125	0.00121	10.31
Sat/Sun (derived)	-0.0312	0.00303	-10.31
Automatically Identified Outliers			
AO2010.Dec	0.4135	0.06015	6.87
LS2011.Feb	-0.3347	0.07184	-4.66
TC2015.Feb	-0.3278	0.06710	-4.89
AO2017.Sep	0.2624	0.05807	4.52

Example: Northeast Building Permits Outlier Identification Results (3)

The following time series values might later be identified as outliers when data are added or revised. They were not identified as outliers in this run either because their test t-statistics were slightly below the critical value or because they were eliminated during the backward deletion step of the identification procedure, when a non-robust t-statistic is used.

"Almost" Outliers

	t(AO)	t(LS)	t(TC)
AO2010.Feb	3.77	1.24	2.58



Logged Original Series and Outlier Adjusted Original Series
Northeast Building Permits
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Logged Original Series
Logged Outlier Adjusted Original Series
Logged Additive Outliers
Vertical Lines: Level Shifts, Temporary Changes

Example: Northeast Building Permits Production Spec File

```
series{ file = "bpne1.dat" format = datevalue
   title = "Northeast Building Permits" }
transform{ function = log }
regression{ variables = (td1coef AO2010.Dec
   LS2011.Feb TC2015.Feb AO2017.Sep) }
outlier{ types = (AO LS TC) span = (2018.1, )
   critical = 4.0 }
arima{ model = (1 1 0)(0 1 1) }
forecast{ maxlead = 36 }
x11{ seasonalma = s3x5 }
```

Outlier Summary

- Outlier types
  - Additive outlier
  - Level shift
  - Temporary level shift
  - Temporary change
  - Ramp (multiple types)
  - Seasonal outlier
- Identify outliers using the `outlier` spec
- Set known outliers using the `regression` spec

Seasonal Regressors

Sometimes called

- Seasonal dummies
- Fixed seasonal effects

Monthly Seasonal Regressors

$SR_{1,t}$ = 1 in Jan., −1 in Dec., 0 otherwise

$SR_{2,t}$ = 1 in Feb., −1 in Dec., 0 otherwise

. . .

$SR_{11,t}$ = 1 in Nov., −1 in Dec., 0 otherwise

Constrained to 11 regressors

Quarterly Seasonal Regressors

$SR_{1,t}$ = 1 in 1st Q, −1 in 4th Q, 0 otherwise

$SR_{2,t}$ = 1 in 2nd Q, −1 in 4th Q, 0 otherwise

$SR_{3,t}$ = 1 in 3rd Q, −1 in 4th Q, 0 otherwise

Constrained to three regressors

Seasonal Regression Effects

- Currently de-emphasized, less common than in the past
- May be useful if series has very stable seasonality
  - Not for series without stable seasonality
- ***Cannot*** be used in combination with a seasonal difference, which makes the regression variables all zero
- May need to combine with seasonal

  (P 0 0) or (0 0 Q)

When to Use Seasonal Regression

- Fit ARIMA model (p d q) (0 1 1)
- Check if the seasonal MA coefficient Θ is not significantly different from 1
  - Check for difference of two standard errors
  - $\Theta = 1$ implies very stable seasonality*
- Also check if Θ is close to 1 in a practical sense (greater than 0.9)
- ***May*** want to use seasonal regressors

- *Alternately, it could mean there's a seasonal overdifference, or a very poorly fitting model.

```
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
```

Stable Seasonality

- Seasonal MA coefficient is 0.9830
- Standard error is 0.30551
  - Two standard errors = 0.61102
- Coefficient is not significantly different from 1 but the standard error is very large
- Coefficient is 0.98, close to 1
- Try seasonal regression

Constant Effect?

- Because a seasonal difference implies an additional regular difference, check for significance of constant term when removing a seasonal difference to include seasonal regressors

$$(1 - B^{12}) = \mathbf{(1 - B)}(1 + B + B^2 + \ldots + B^{11})$$

How to Use Seasonal Regressors

- Determine if seasonal regressors are an option
- Remove the seasonal part of the ARIMA model
- Add seasonal and constant variables in the regression spec
- Check whether the regressors are significant
- (If needed for model fit, add a seasonal

    AR: $(1\ 0\ 0)_s$ or seasonal MA: $(0\ 0\ 1)_s$)

Regression Spec With Seasonal Effects

```
ARIMA{model=(0 1 3)
# or model=(0 1 3)(0 0 0)
}
regression{
  variables = (const seasonal)
## note const and not constant
}
```

Regression Model

Variable	Parameter Estimate	Standard Error	t-value
Constant	0.0129	0.00096	13.47
Seasonal			
1st	-0.0235	0.00113	-20.89
2nd	0.0016	0.00106	1.55
3rd	-0.0095	0.00106	-8.98
*4th(derived)	0.0314	0.00113	27.87
Automatically Identified Outliers			
AO2003.4	0.0175	0.00432	4.04

Chi-squared Tests for Groups of Regressors

Regression Effect	df	Chi-Square	P-Value
Seasonal	3	1287.09	0.00

F Tests for Seasonal Regressors

Regression Effect	df	F-statistic	P-Value
Seasonal	3,56	407.18	0.00

Which Do We Choose?

- Soon we will talk about diagnostics to help us choose between competing regARIMA models

Trading Day Regression

- Capture the effect of the weekday composition of each month
- Every month has at least four of every type of day (Sundays, Mondays, etc.)
- Months may have five of some types of days

August 2011

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 2014

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						

Limitations on Trading Day

- Quarterly trading day effects are subtle, usually not significant enough to model
- Generally need 7 or more years of a monthly series to estimate a trading day effect
  - Trading day is a weekly effect we are trying to measure with monthly or quarterly observations

Types of Trading Day (IMPORTANT)

- Flow
  - Activity summed over the month
- Stock
  - Inventory, snapshot

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)

Constrained to six regressors

One-Coefficient Trading Day – Weekday vs. Weekend (Flow)

D_{jt} = number of days of type *j* in month *t*

j = 1 (Mon), 2 (Tues), . . ., 7 (Sun)

$$TD_t = \sum_{j=1}^{5} D_{jt} - \frac{5}{2}\sum_{j=6}^{7} D_{jt}$$

Estimates an effect for Mon-Fri vs. Sat-Sun

Flow Trading Day Regressors

- The six-coefficient td regressor shown was **`tdnolpyear`**
- The one-coefficient td regressor shown was **`td1coefnolpyear`**

- The two most common flow td regressors, **`td`** and **`td1coef`**, include leap year effects along with the trading day regressors
  - Multiplicative => leap year prior adjustment factor
  - Additive => regression effect
- If no seasonal adjustment performed, adjustment for length of month included in the trading-day effect
  - Less common

Additive Adjustment TD

```
transform { function = none}
regression { variables = td }
```

is equivalent to

```
transform { function = none }
regression { variables =  (tdnolpyear lpyear) }
```

Multiplicative Adjustment TD

```
transform { function = log}
regression { variables = td }
```

is equivalent to

```
transform { function = log
    adjust = lpyear}
regression { variables = tdnolpyear }
```

*If you think the leap year adjustment over/underadjusts the series, you can use **`variables=(tdnolpyear lpyear)`** for multiplicative adjustments as well

A2 Prior-adjustment factors

Leap year (from trading day regression) adjustments.

From 1991.Jan to 2007.Nov

Observations 203

	Jan	Feb	Mar	Apr	May...	Dec	AVGE
1991	100.0	99.1	100.0	100.0	100.0...	100.0	99.9
1992	100.0	102.7	100.0	100.0	100.0...	100.0	100.2
1993	100.0	99.1	100.0	100.0	100.0...	100.0	99.9
1994	100.0	99.1	100.0	100.0	100.0...	100.0	99.9
1995	100.0	99.1	100.0	100.0	100.0...	100.0	99.9
1996	100.0	102.7	100.0	100.0	100.0...	100.0	100.2

A 2 Prior-adjustment factors
Leap year (from trading day regression)
adjustments.
From 2003.4 to 2007.4
Observations 17

	1st	2nd	3rd	4th	AVGE
2003				100.00	100.00
2004	100.83	100.00	100.00	100.00	100.21
2005	99.72	100.00	100.00	100.00	99.93
2006	99.72	100.00	100.00	100.00	99.93
2007	99.72	100.00	100.00	100.00	99.93

Stock (Inventory) Trading Day

- **`tdstock[ω]`**
  - ω is the day of inventory
- Let ϖ be the smaller of ω and the length of month (equal to ω for $\omega \leq 28$)
  - D_{1t} = 1 when ϖth day is Monday, -1 when ϖth day is Sunday, and 0 otherwise
  - D_{2t} = 1 when ϖth day is Tuesday, -1 when ϖth day is Sunday, and 0 otherwise
  - …
  - D_{6t}

Setting Stock Trading Day Regressor

- Day of inventory is ω
- Regressor form for end-of-month inventories is **`tdstock[31]`**
- Default value is 31 for **`AICtest`** (defined soon)
  - Recall ϖ is the smaller of the length of month and ω, so ϖ is 30 for April
- Not for inventories taken on a given day of the week (for example, not for inventories taken on the first Wednesday)

One-Coefficient Trading Day for Stock Series

- How it's calculated:
  $I(\omega)_t = -0.6D(\omega)_{1,t} - 0.2D(\omega)_{2,t} + 0.2D(\omega)_{3,t} + 0.6D(\omega)_{4,t} + D(\omega)_{5,t}$

  where $D(\omega)_{i,t}$ is the i^{th} stock trading day regressor defined as above for stock day ω
- Regressor is called **`tdstock1coef[w]`**

See "Modeling Stock Trading Day Effects Under Flow Day-of-Week Effect Constraints" by Findley and Monsell (2009)

Example: Grocery Store Retail Sales

```
series{ file = "grocery stores.dat"
    format = datevalue }
transform{ function = log }
regression{ variables = (td easter[8]
    AO1999.Dec AO2000.Jan ) }
outlier{ types = all }
arima{ model = (1 1 0)(0 1 1) }
forecast{ maxlead = 12 }
x11{ seasonalma = s3x5 }
```

Example: Grocery Store Retail Sales Regression Model Table

	Parameter Estimate	Standard Error	t-value
Trading Day			
Mon	-0.0049	0.00073	-6.68
Tue	-0.0049	0.00074	-6.61
Wed	0.0024	0.00073	3.26
Thu	0.0025	0.00072	3.50
Fri	0.0032	0.00074	4.25
Sat	0.0078	0.00074	10.48
*** Sun (derived)**	-0.0061	0.00074	-8.32
Easter[8]	0.0206	0.00147	14.05
AO1999.Dec	0.0417	0.00543	7.68
AO2000.Jan	-0.0258	0.00543	-4.76

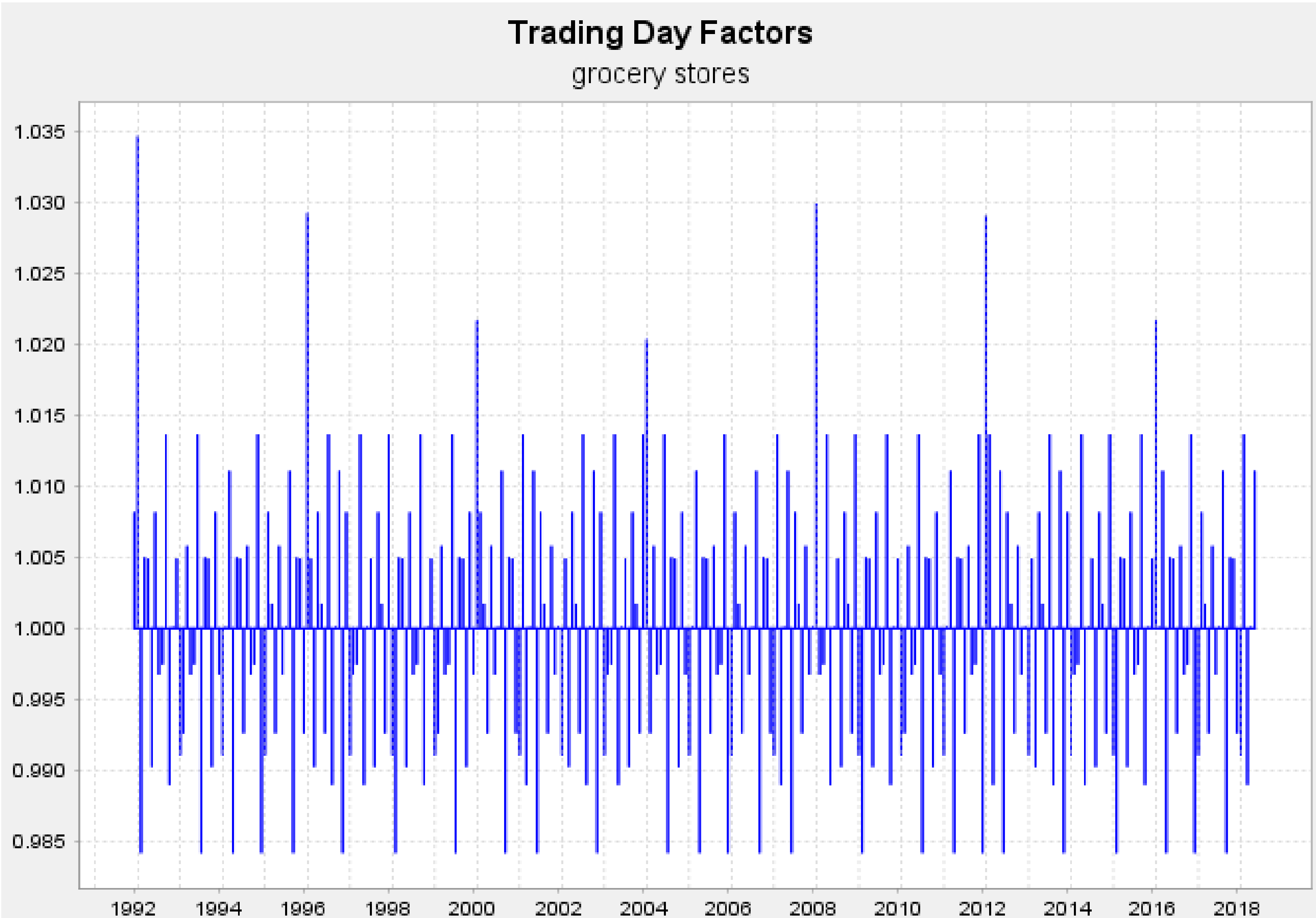
Example: Grocery Store Retail Sales Trading Day F-Test

- X-13A-S tests the significance of the group of trading day regressors

F Tests for Trading Day Regressors

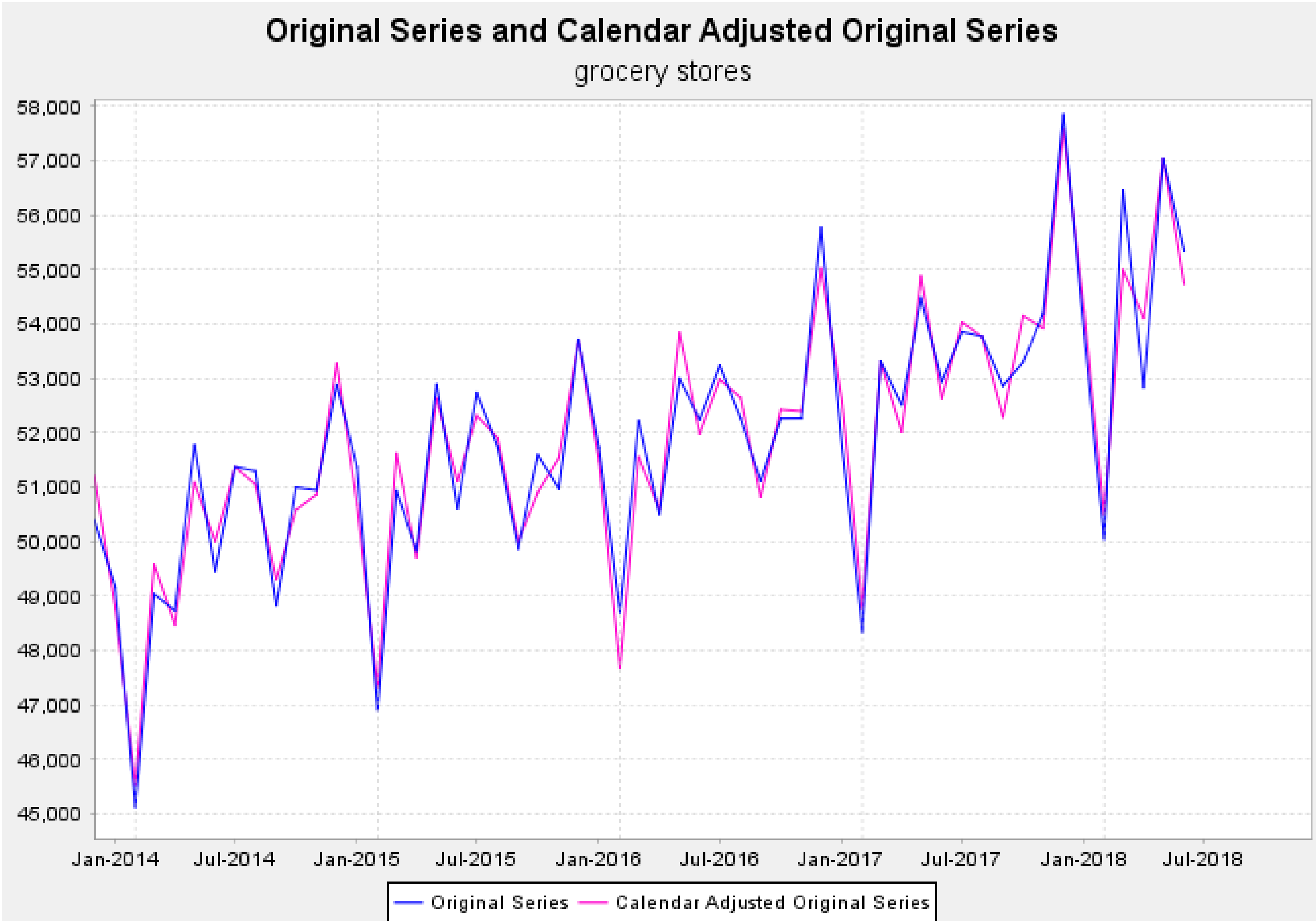
	df	F-statistic	P-Value
Trading Day	6, 296	172.51	0.00

Example: Grocery Stores Trading Day Factors



Example: Grocery Store Sales Calendar Adjusted Series

Original series / (trading day factors * Easter factors)



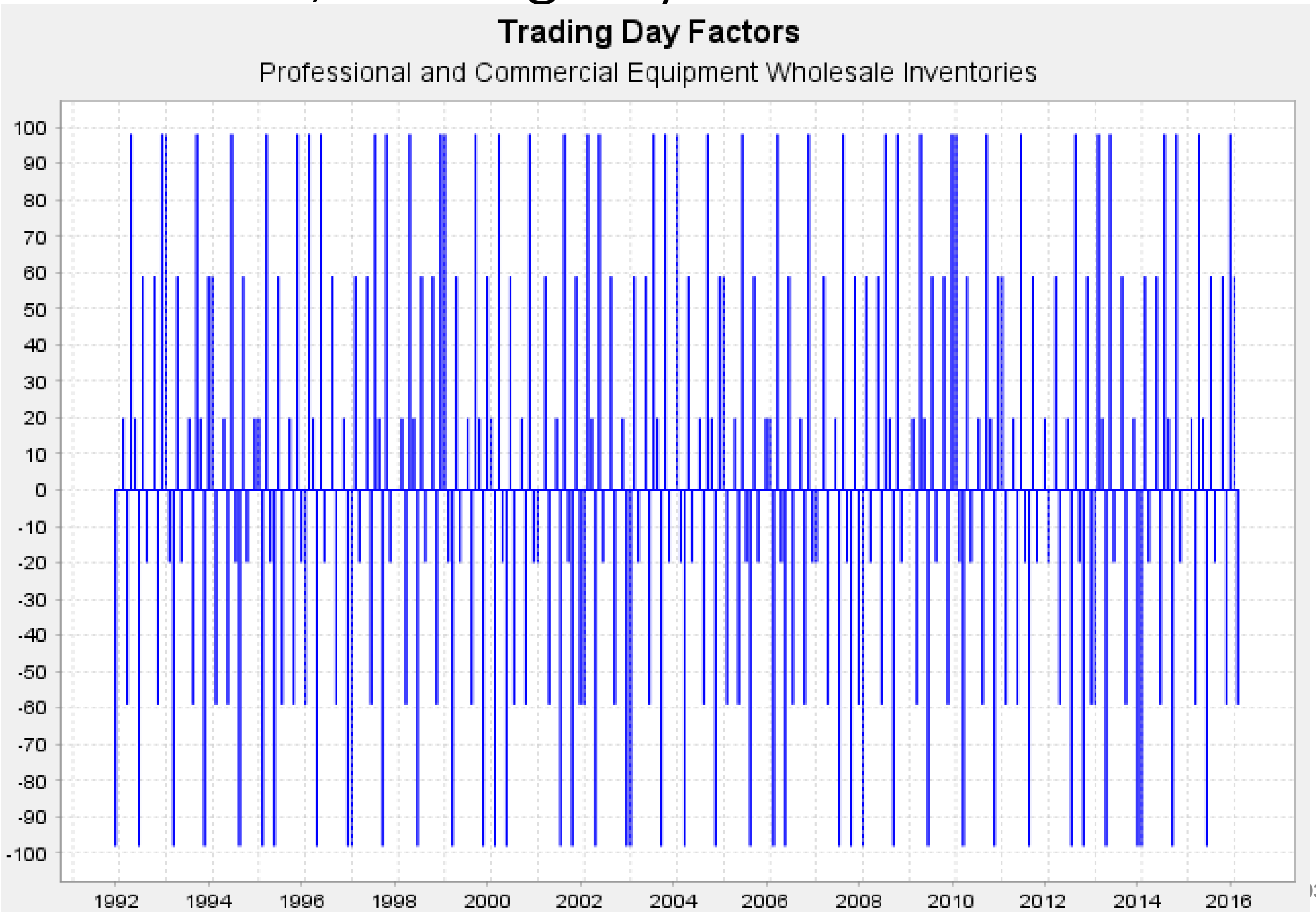
Example: Professional and Commercial Equipment and Supplies Wholesale Inventories

```
series{ file="professional equipment.dat"
    format = datevalue }
transform{ function = none }
regression{ variables = (tdstock1coef[31]
     easterstock[1] ) }
outlier{ types = all }
arima{ model = (0 1 1)(0 1 1) }
forecast{ maxlead =  60 }
x11{ seasonalma = s3x5 }
```

Example: Professional Equipment Wholesale Inventories, Regression Model Table

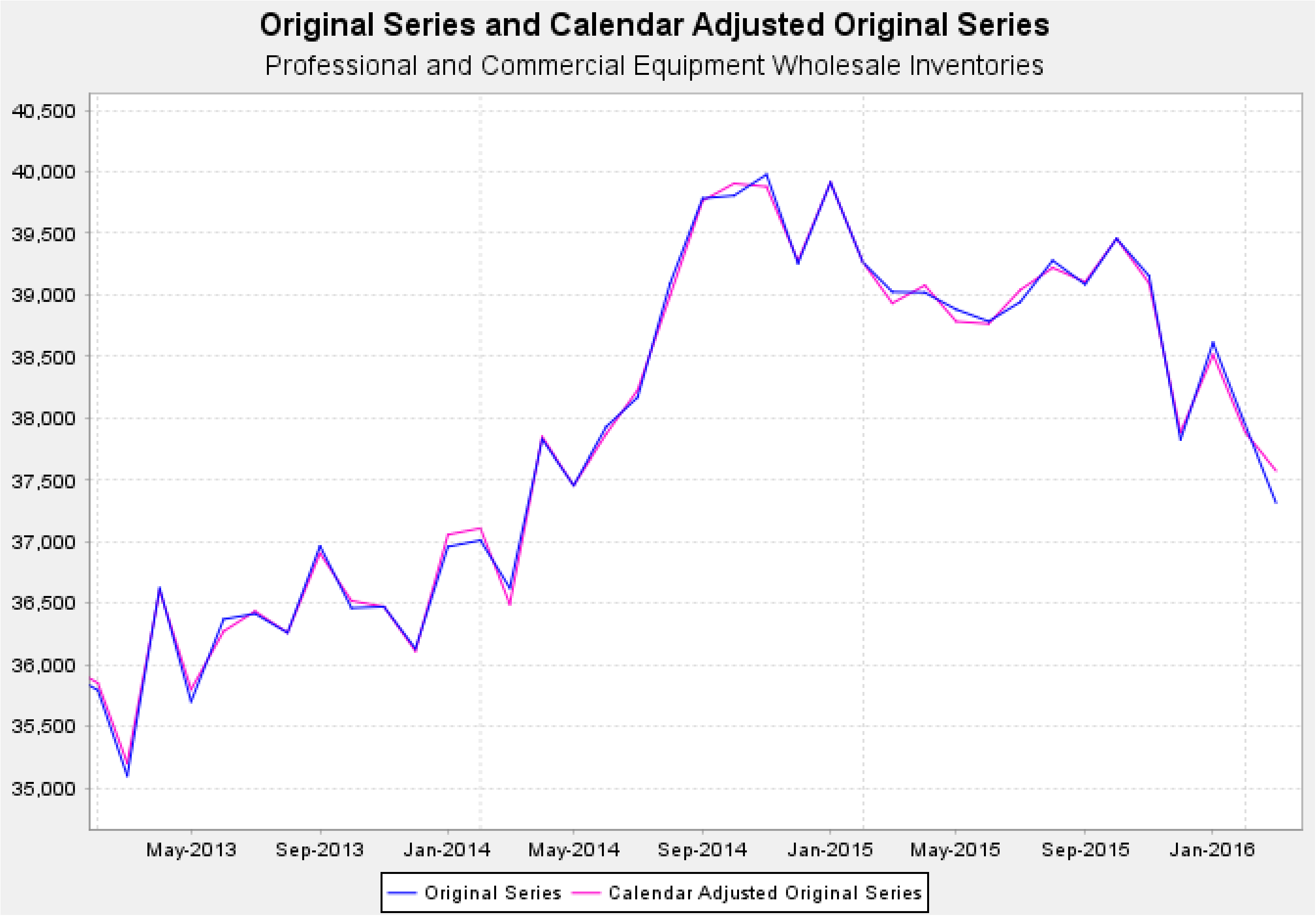
	Parameter Estimate	Standard Error	t-value
1-Coefficient Stock Trading Day[31]			
Weekday	-97.8555	21.94870	-4.46
**** Sat/Sun (derived)**	244.6386	54.87174	4.46
StockEaster[1]	-274.1645	136.48426	-2.01

Example: Professional Equipment Wholesale Inventories, Trading Day Factors



Example: Professional Equipment Wholesale Inventories, Calendar Adjusted Series

Original series – trading day factors – Easter factors



Easter Effect

- Primarily for Retail Sales but we've seen evidence of the effect in other series
- Regressor measures an effect (often an elevated level for Retail series) seen in the series for w days, ending on the day before Easter, $1 \leq w \leq 25$
  - The built-in regressor assumes the effect is the same over those w days

Easter[w]

$$E_{wt} = \frac{W_t}{w} - \bar{E}_{wt}$$

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

W_t = number of the w days in month t

$\bar{E}_{wt}$= the long-term monthly means of W_t / w

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

10-Day Easter Effect: `Easter [10]`

If Easter is on April 3 and we use `Easter[10]` (w=10)

- W_{March} / w = 8/10
- W_{April} / w = 2/10
- W / w = 0 for all other months

Easter[0]

- New regressor available
- Based on research from “Big Data” pilot
  - Daily retail electronic sales transactions indicated that Easter activity levels were lower than other Sundays
- Measures an effect on the day itself, not before and not after

- *Indistinguishable from Easter[1] unless Easter is on April 1

End-of-the-Month Stock Easter Regressor

$$E(w)_{m,y}^{stock} = \begin{cases} 0, & \text{for } m = 1 \\ E(w)_{2,y}^{flow}, & \text{for } m = 2 \\ E(w)_{2,y}^{flow} + E(w)_{3,y}^{flow}, & \text{for } m = 3 \\ 0, & \text{for } 4 \le m \le 12 \end{cases}$$

Stock Easter Regressor Syntax

```
regression{
 variables = (tdstock[31]
                easterstock[9]
            ao1992.Nov ls1989.Jul ) }
```

- For more information see Findley (2009)
- Titova and Monsell (2009) applied stock calendar regressors to Census Bureau stock series

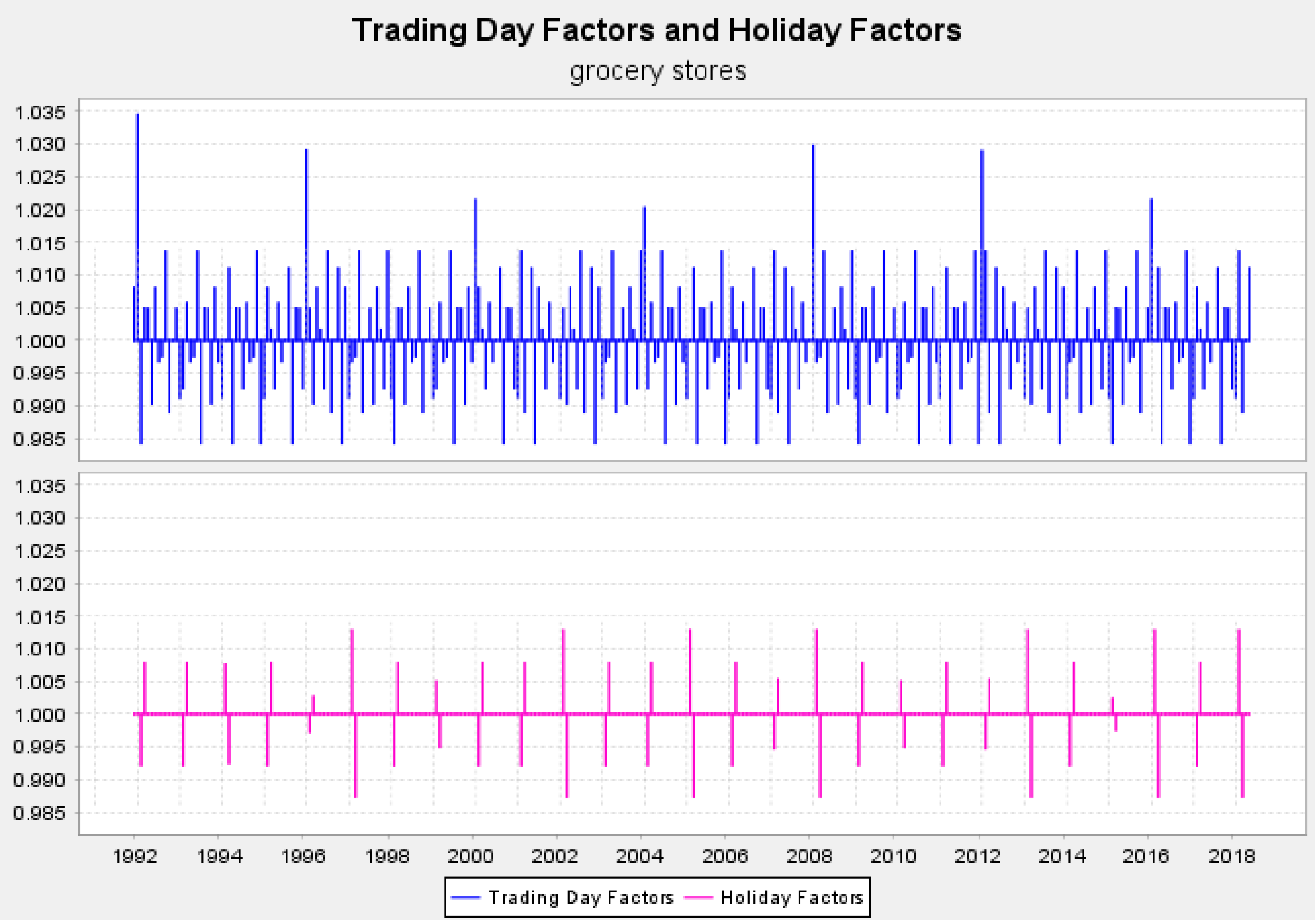
Example: Grocery Store Retail Sales Spec File

```
series{ file = "grocery stores.dat"
    format = datevalue }
transform{ function = log }
regression{ variables = (td easter[8]
    AO1999.Dec AO2000.Jan ) }
outlier{ types = all }
arima{ model = (1 1 0)(0 1 1) }
forecast{ maxlead = 12 }
x11{ seasonalma = s3x5 }
```

Example: Grocery Store Retail Sales Regression Model Table Showing Easter

	Parameter Estimate	Standard Error	t-value
Trading Day			
Mon	-0.0049	0.00073	-6.68
Tue	-0.0049	0.00074	-6.61
Wed	0.0024	0.00073	3.26
Thu	0.0025	0.00072	3.50
Fri	0.0032	0.00074	4.25
Sat	0.0078	0.00074	10.48
*** Sun (derived)**	-0.0061	0.00074	-8.32
Easter[8]	0.0206	0.00147	14.05
AO1999.Dec	0.0417	0.00543	7.68
AO2000.Jan	-0.0258	0.00543	-4.76

Example: Grocery Store Retail Sales Trading Day and Holiday Factors



Other Moving Holidays

Labor Day

```
Labor[ 1 to 25 ]
```

Thanksgiving (Christmas effect)

```
Thank[ -8 to 17 ]
```

Primarily for retail series

Not appropriate for quarterly series

RegARIMA Model Note

- Stock (inventory) series
  - Stock trading day, end-of-month stock Easter
  - Currently no built-in stock Labor Day or Thanksgiving
- Don’t use regressors meant for flow series to model stock series (and vice versa)

Regression Variables Argument

```
variables = ( const
   seasonal
   td | td1coef | tdstock[1 to 31]
   | tdnolpyear | td1nolpyear
   | tdstock1coef[1 to 31]
   lpyear | lom | loq
   easter[1 to 25] | easterstock[1 to 25]
   labor[1 to 25]   thank[-8 to 17]
   aoyyyy.mm  lsyyyy.mm  tcyyyy.mm
   rpyyyy.mm-yyyy.mm
   soyyyy.mm
)
```

Regression Effects and the Final Seasonally Adjusted Series

- Trading day, Easter, and other moving holiday effects (a.k.a. calendar effects) are removed permanently from the seasonally adjusted series

- Outliers are removed for estimating seasonal factors but put back into the final seasonally adjusted series

- Seasonal regressors and constant are for modeling and forecasting only
  - Under special circumstances, we might use seasonal regressors to adjust part of the series

User-Defined Regressors

- Used when the predefined regressors are not sufficient
  - Commonly used to define regressors for holidays not widely celebrated in the United States (`Id al Fitr/Ramadan in Islamic countries, Chinese New Year, Easter Monday, etc.)

`Regression` Spec for User-Defined Regressors

```
regression {
. . .
  user = ( ChNY
#  name(s) of user-defined
#  regression variable(s)
)
. . .
```

`Regression` Spec for User-Defined Regressors (2)

```
. . .
#  data = ( )

#  start = yyyy.mm
## Do not use data argument and
## file argument in the same
## regression spec
. . .
```

`Regression` Spec for User-Defined Regressors (3)

```
. . .
 file = "MyRegressors.txt"
 format = "datevalue" |  "x12save" | "free"
# or some other accepted format
# [ default: free ]

. . .
```

`Regression` Spec for User-Defined Regressors (4)

```
. . .
  usertype = ( constant  seasonal
    td  tdstock
    lpyear  lom  loq
    easter  thanks  labor  holiday
     holiday2-holiday5
    ao  ls  rp  tc  so
    user transitory )
 }
```

Example Spec File: User-Defined Regressors

```
series { file="sales.dat" format="free"
  start = 1989.jan }
transform { function = log }
regression { variables = ( td ao1990.1 )
     user = ( cny1 cny2 cny3 )
     file = "cnyregrs.dat"
     format = "datevalue"
     usertype = holiday
}
arima { model = (0 1 1)(0 1 1) }
check { … } estimate { … }  forecast { … }
x11{ … }
```

`Regression` Spec AIC Tests

- X-13A-S can perform AIC tests for
  - Flow and stock trading day
  - Flow and stock Easter
  - User defined regressors

AIC Test Example

```
aictest = (
   td | tdstock |
   easter | easterstock |
   user
)
```

`aictest = td`

- X-13ARIMA-SEATS estimates AICCs of models with
  - **`td`** (Six-coefficient)
  - **`td1coef`** (One-coefficient)
  - No td

  and then selects the model with minimum AICC

Testing for Trading Day Effects

```
regression {
  variables =(td | td1coef
    | tdstock[w] | tdstock1coef)
  aictest = td
}
```

These settings will test ***only*** the specified trading day effect vs. no effect

Testing for Stock Trading Day

- **`aictest = tdstock`** will estimate the AICC of
  - **`tdstock[31]`**
  - **`tdstock1coef[31]`**
  - no trading day
- To test other inventory days

```
regression{ variables = tdstock[1]
  aictest = td }
```

aictest = easter

- If **variables = (easter[w] …)** is in spec file, compare AICCs of models with and without **easter[w]** regressors

- If no specific Easter variable, compare AICCs
  - without Easter
  - with **easter[1]**
  - with **easter[8]**
  - with **easter[15]**

- Choose model with minimum AICC

Testing for Easter Effects

```
regression {
  variables =
 ( easter[w] | easterstock[w] )
  aictest = easter
}
```

These settings will test ***only*** the specified Easter effect vs. no effect

Purely AICC?

- When running the **`automdl`** spec, X-13ARIMA-SEATS does a significance test in addition to the AICC test
    - 5% level
- With a set ARIMA model, X-13ARIMA-SEATS does a pure AICC test without additional requirements

AIC Tests in Output

- If running **`automdl`**, the AIC tests for trading day and Easter are not printed in the output file
- If the ARIMA model is hard-coded, then the tables are printed.

Example: Grocery Store Retail Sales AIC Test for TD and Easter

...

```
regression{ aictest = (td easter) }
arima{ model = (1 1 0)(0 1 1) }
```

...

Likelihood statistics for model without td

Likelihood Statistics

AICC (F-corrected-AIC)	4813.8426

Likelihood statistics for model with td

AICC (F-corrected-AIC)	4330.1283

Likelihood statistics for model with td1coef

AICC (F-corrected-AIC)	4771.1070

***** AICC (with aicdiff = 0.0000) prefers model **with td** *****

Spurious AIC Results

- Research has shown that TD and Easter may be chosen by the AIC tests when they aren't present
  - TD1 in particular very likely to be chosen
  - Test for the effect only if it makes sense
    - Ask if the selected regressor makes sense
  - Look at the associated t statistics and p values
  - Look at the difference in AICC: Is AICC with TD much smaller than AICC without?

Model Span

- By default the same span of data that is being adjusted is used for modeling
- Can use a subspan of the full series
  - Applies to all modeling specs (**`regression, arima, outlier`***, etc.)

  - Can shorten the outlier span even more
    - **`Outlier{types = ao span = (2014.1,)}`**

`modelspan` Argument

```
series{
   span=(1991.1, 2007.12)
   modelspan=(1999.1, )
  . . .
}
```

Model Spans and Regressors

- The model span limits the data used to calculate the parameter estimates for the regression effects
  - Regressors (like td, easter) are still applied to the full span of data, but the regressor is estimated using only the data in the model span

Using a Model Span

- If all series in a group must start at the same date, but some individual series show evidence of change over time, shortening the model span might improve forecasts and model fit for these series
  - Model spans do not have to agree for groups

When to Avoid a Model Span

- If the series is very different from start to end, try changing the full series span
  - Abruptly or rapidly changing seasonal patterns will make it harder to seasonally adjust the full series, may need to shorten the series for better estimation

- Avoid using a model span when there are outliers prior to the model span start

Change of Regime

- For seasonal and trading day regressors
- Useful to model or detect changes in the seasonal or td pattern

Change of Regime Types

- **`variables = seasonal/2000.Jan/`**
  - Estimates seasonal effects for the full series and a change in the seasonal effects through Dec 1999
- **`variables = td//2004.Jan/`**
  - Estimates trading day effects for the series starting only at Jan 2004
- **`variables = tdstock[31]/2002.Jan//`**
  - Estimates stock trading day from start of model span through Dec 2001 only

Change-of-regime F-test Results

	Df	F Statistic	P Value
Trading Day (after 2000.Jan)	6, 226	21.59	0.00
Trading Day (change for before 2000.Jan)	6, 226	2.90	0.01*
Combined Trading Day Regressors	12, 227	27.96	0.00

*Significant Change at the 95% level ($\alpha = 0.05$)

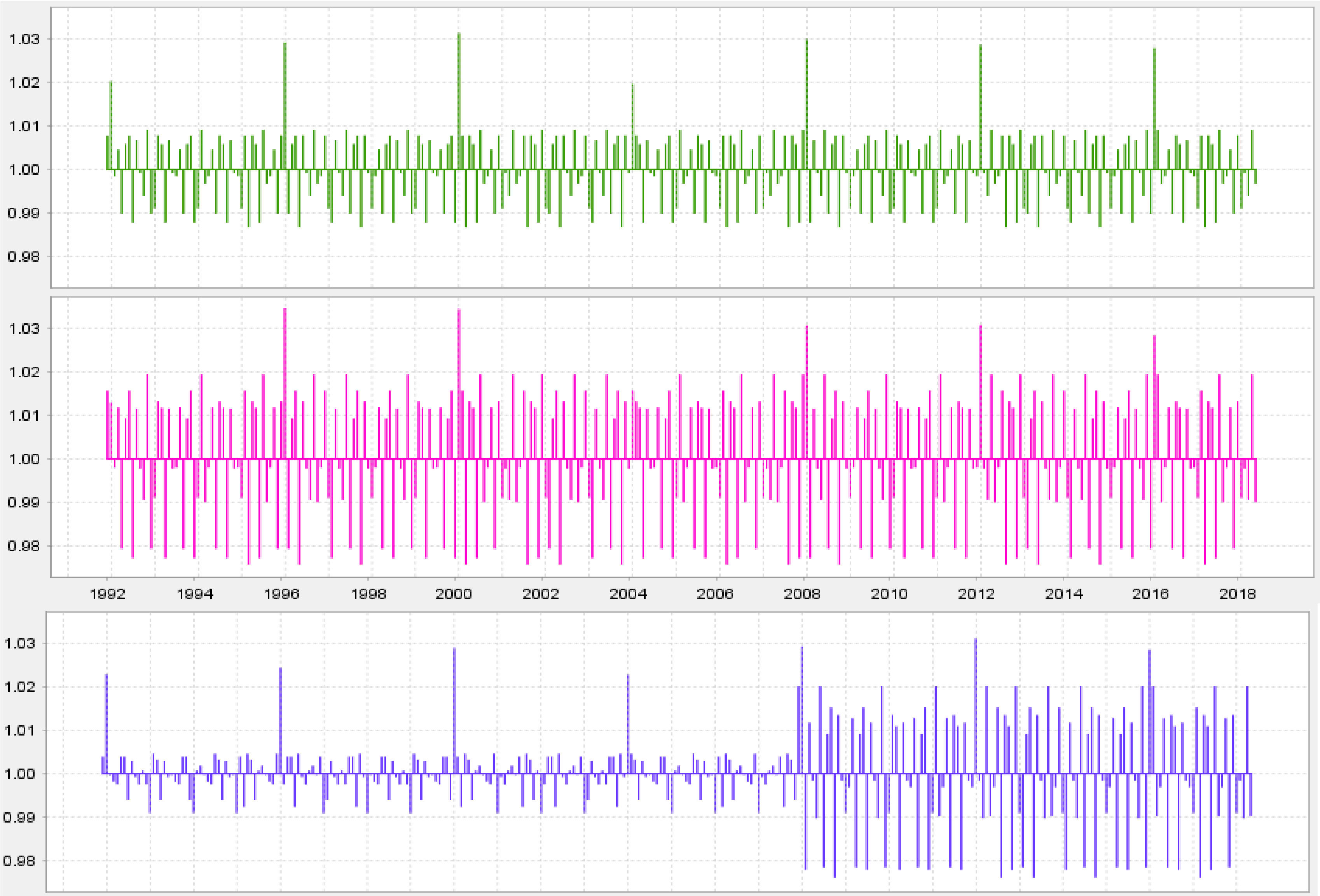
Example: Printing & Related Support Activities: Manufacturing Value of Shipments

Run with:

```
1. series{…} regression{ variables = td }
2. series{ modelspan = (2008.1, ) }
   regression{ variables = td }
3. series{…} regression{ variables = td/2008.1/ }
```

Trading day is significant with both spans, and the change is significant.

Example: Trading Day Factors of the Three Runs



Regression Spec, Other Arguments

```
regression { . . .
  print=See Manual/Quick Reference
  save=See Manual/Quick Reference
  savelog = aictest
}
```

Purposes of RegARIMA Models in X-13ARIMA-SEATS

- Directly estimate trading day and holiday effects
- Detect and adjust for outliers and other distorting effects to improve the forecasts and seasonal adjustments (automatic option)
- Estimate missing values
- Choose between competing prior adjustments, calendar effect models, etc.
- Forecast the series