Accommodating Weather Effects in Seasonal Adjustment

Osbert Pang
U.S. Census Bureau

November 4, 2016

Acknowledgements

Joint work with Brian Monsell, William Bell, and James Livsey.

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau.

Outline

1. Introduction
2. Taking weather data and forming regressors
3. Applying computed weather regressors to Midwest regional construction series
4. Conclusions and other thoughts

Introduction

- Seasonal adjustment attempts to account for regular seasonal patterns in a time series.
- Weather effects can contribute to seasonal patterns, but may vary greatly from year to year.
- Unaccounted for weather effects may be viewed as outliers.
- This is inspired in part by the work of Boldin and Wright (2015), who found weather effects in national employment data.

Weather Data (1)

Monthly summary data for weather stations from the National Climatic Data Center (NCDC) for 16 weather variables

Each variable can be categorized into 1 of the following 4 groups (NCDC variable name in parentheses):

Number of days in a month where ...	Monthly extreme values	Monthly totals	Monthly means
Precipitation $\geq$ 0.1 in (DP01)	Max daily precipitation (EMXP)	Precipitation (TPCP)	Mean max temperature (MMXT)
Precipitation $\geq$ 0.5 in (DP05)	Max daily temperature (EMXT)	Snowfall (TSNW)	Mean min temperature (MMNT)
Precipitation $\geq$ 1.0 in (DP10)	Min daily temperature (EMNT)		Mean temperature (MNTM)
Min temperature $\leq$ 0 F (DT00)	Max snow depth (MXSD)		
Min temperature $\leq$ 32 F (DT32)			
Max temperature $\leq$ 32 F (DX32)			
Max temperature $\geq$ 90 F (DT90)			

Weather Data (2)

- Data obtained stretches from Jan 1944 through Dec 2014.
- Stations get mapped to cities, with a total of 92 cities.
- Not all stations have monthly summary numbers for all variables previously described.
- We attempt to create monthly city value with no missing values by splicing associated stations.

Weather Data (3)

To illustrate, suppose that for variable x we have data from 3 stations associated with a city, and the beginning of the series looks as follows:

Station A	x_{A1}	x_{A2}				x_{A6}	x_{A7},
Station B		x_{B2}	x_{B3}		x_{B5}		x_{B7},
Station C				x_{C4}	x_{C5}	x_{C6}	x_{C7}.
City	x_{A1}	$f(x_{A2}, x_{B2})$	x_{B3}	x_{C4}	$f(x_{B5}, x_{C5})$	$f(x_{A6}, x_{C6})$	$f(x_{A7}, x_{B7}, x_{C7})$

- f in this discussion is the median, but other options are available (e.g., mean, min/max).

- Additional step: for variables that are counts of days, divide by the number of days in the corresponding month to convert to proportion.

Weather Data (4)

- Mean-center using values for same month from previous 20 years; that is, if $x_{k,t}$ is the monthly value for variable x at time t of some city k, then the mean-centered value is
$$\tilde{x}_{k,t} = x_{k,t} - \frac{1}{N}\sum_{\delta=1}^{20} x_{k,t-12\delta}.$$

- Obtain regional regressor based on weather variable by summing population-weighted city values; suppose $w_{k,t}$ is the weight for city k at time t, then for a region K, the regional regressor $v_{K,t}$ at time t would take the value
$$v_{K,t} = \sum_{k \in K} w_{k,t}\tilde{x}_{k,t}.$$

- 2 ideas for population weighting: (1) using only city population figures, (2) using both city and state population figures

Weather Data (5)

Weighting by city population only: cities with a missing value are assigned weight 0; otherwise

$$\text{weight for city } j = \frac{\text{population for city } j}{\text{sum of population for eligible cities}}.$$

Weighting by city and state population: cities with a missing value are assigned weight 0; otherwise

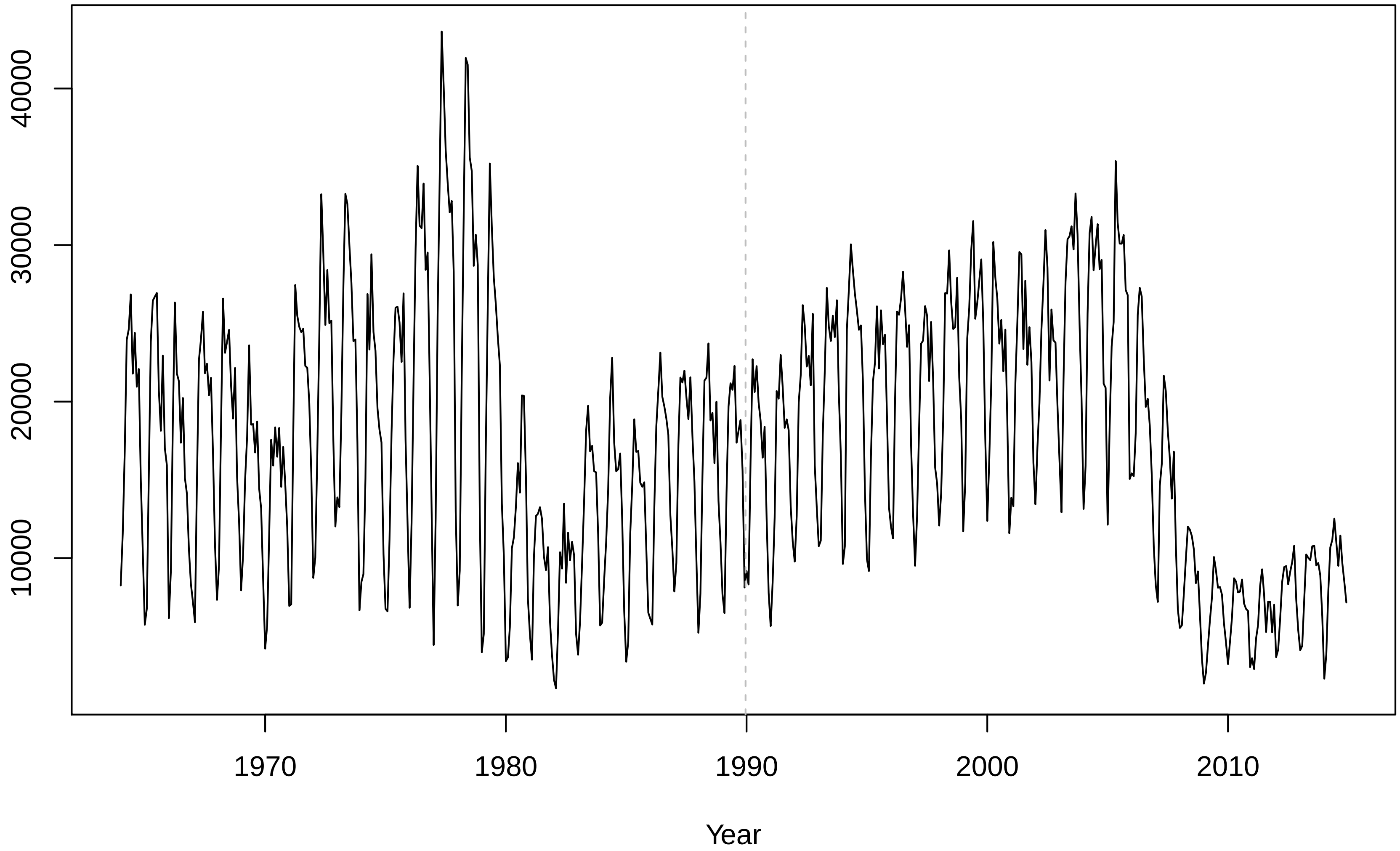
$$\text{weight for city } j = \frac{\text{population for city } j}{\text{sum of population for cities in } j\text{'s state}} \times \frac{\text{population for } j\text{'s state}}{\text{sum of population for states in } j\text{'s region}}.$$

Weather Data (6)

Quick Recap

1. Monthly summary data for weather variable from NCDC
2. Make monthly series for a city by splicing together monthly series from associated stations
3. Convert count of days in month into proportion of month
4. Mean-centered city values
5. Apply weights to obtain single regional value from city values.

Application: Midwest Regional Housing Starts



Application: MW Starts (2)

The data come from the Survey of Construction; time span used is Jan 1964 through Dec 2014.

Default pass with a $(0\ 1\ 1)(0\ 1\ 1)_{12}$ airline model (no regressors) through X-13ARIMA-SEATS suggests the following regARIMA model on the log-transformed starts:

- 1-coefficient trading-day regressor
- 6 point outliers (AO): Dec 1973, Jan 1977, Feb 1981, Feb 1982, Dec 2011, Jan 2014
- 3 level shifts (LS): Jan 1979, Mar 1979, Jan 1980

Re-examine using time span above as well as two subspans (Jan 1964 through Dec 1989, Jan 1990 through Dec 2014), with outliers fixed at those identified previously.

Application: MW Starts (3)

	1964.1–2014.12		1964.1–1989.12		1990.1–2014.12	
Regressor	AICC	AICC change	AICC	AICC change	AICC	AICC change
None	10 935.7		5510.2		5191.1	
DT90	10 937.4	1.7	5511.2	1.0	5193.2	2.1
DX32	10 818.5	−117.2	5439.4	−70.8	5154.1	−37.0
DT00	10 800.3	−135.4	5439.6	−70.6	5141.2	−50.0
DT32	10 924.6	−11.1	5505.2	−5.0	5189.2	−1.9
DP01	10 916.1	−19.6	5498.9	−11.3	5182.1	−9.1
DP05	10 922.4	−13.3	5504.3	−5.9	5183.9	−7.3
DP10	10 931.0	−4.7	5510.1	−0.1	5187.5	−3.6
EMXP	10 934.0	−1.7	5510.5	0.2	5190.1	−1.0
MXSD	10 882.9	−52.8	5479.6	−30.6	5171.2	−20.0
TPCP	10 921.8	−13.9	5504.3	−6.0	5183.4	−7.8
TSNW	10 877.8	−57.9	5481.3	−29.0	5162.8	−28.3
EMXT	10 910.6	−25.1	5491.2	−19.0	5188.4	−2.8
EMNT	10 885.5	−50.2	5484.9	−25.3	5174.3	−16.9
MMXT	10 865.5	−70.3	5475.3	−34.9	5165.1	−26.0
MMNT	10 872.7	−63.0	5480.7	−29.5	5167.8	−23.3
MNTM	10 865.2	−70.5	5475.9	−34.3	5165.0	−26.1

Application: MW Starts (4)

- Days where max temperature exceeds 90 F does not have a significant effect.
- Days where max temp is below 32 F and Days where min temp is below 0 see largest decreases in AICC.
- The 0.1 in threshold for precipitation counts of days has a stronger effect than the other 2 thresholds.
- Snowfall has more effect on AICC than precipitation, though.
- Average monthly temperature regressors matter more than than extreme temperature regressors.
- Best single regressors from an AICC perspective deal with temperature and snowfall.

Application: MW Starts (5)

	Additive Outliers						Level Shifts				
Reg	1973 Dec	1977 Jan	1981 Feb	1982 Feb	2011 Dec	2014 Jan	1979 Jan	1979 Mar	1980 Jan	1980 Sep	2008 Dec
None	X/X	X/X	X/	X/X	X/X	X/X	X/X	X/X	X/X		/X
DX32	X/	X/	X/X	X/X	/X	X/X	X/X	X/	X/X		
DT00	X/X			X/X	X/X				X/X		
DP01	X/X	X/X		X/X	X/X	X/X	X/X	X/X	X/X		/X
EMXP	X/X	X/X		X/X	X/X	X/X	X/X	X/X	X/X		/X
MXSD		X/X	LS/	X/	/X	X/X	X/		X/X		/X
TPCP	X/X	X/X		X/X	X/X	X/X	X/X	X/X	X/X		/X
TSNW		X/X		X/X	/X	/X	X/X	X/	X/X		
EMNT	X/X	X/X	LS/LS	X/X	X/X	X/X	X/X	X/X	X/X	X/X	/X
MMXT	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X		/X
MMNT	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X		/X
MNTM	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X	X/X		/X

Application: MW Starts (6)

- Using just the first half does not result in Feb 1981 being viewed as an outlier; using just the second half results in Dec 2008 being identified as an outlier.
- The regressors based on average monthly temperatures do not account for any previously identified outliers.
- The inclusion of a low temperature regressor (DT00 specifically) can account for multiple outliers that were previously identified.
- Total snow can account for a few of the previously identified outliers.

United States Census Bureau

Application: MW Starts (7)

	1964.1–2014.12		1964.1–1989.12		1990.1–2014.12	
Regressor(s)	AICC	AICC change	AICC	AICC change	AICC	AICC change
None	10 935.7		5521.8		5174.8	
DT00	10 800.3	−135.4	5452.6	−69.3	5130.5	−44.3
DX32	10 818.5	−117.2	5463.7	−58.2	5142.1	−32.6
DT00	10 800.3		5452.6		5130.5	
{DT00 DX32}	10 787.1	−13.1	5448.8	−3.8	5128.4	−2.1
{DT00 MMXT}	10 791.9	−8.4	5452.1	−0.4	5126.0	−4.5
{DT00 TSNW}	10 792.4	−7.8	5445.6	−7.0	5128.6	−1.8
{DT00 MXSD}	10 795.5	−4.7	5451.1	−1.5	5129.8	−0.7
{DT00 DP01}	10 773.8	−26.4	5433.7	−18.8	5121.2	−9.3
{DT00 TPCP}	10 779.8	−20.4	5440.3	−12.3	5122.2	−8.3
{DT00 DP01}	10 773.8		5433.7		5121.2	
{DT00 DP01 DX32}	10 756.7	−17.2	5425.2	−8.6	5118.7	−2.5
{DT00 DP01 MMXT}	10 768.1	−5.8	5433.7	−0.0	5118.9	−2.3
{DT00 DP01 TSNW}	10 771.9	−1.9	5431.4	−2.4	5122.0	0.8
{DT00 DP01 MXSD}	10 771.9	−1.9	5434.2	0.5	5121.6	0.4
{DT00 DP01 MNTM}	10 769.2	−4.6	5433.8	0.0	5120.0	−1.2
{DT00 DP01 DX32}	10 756.7		5425.2		5118.7	
{DT00 DP01 DX32 MMXT}	10 758.5	1.8	5427.3	2.2	5119.7	0.9
{DT00 DP01 DX32 TSNW}	10 758.7	2.1	5426.6	1.4	5120.9	2.2
{DT00 DP01 DX32 MXSD}	10 758.5	1.9	5427.1	2.0	5120.6	1.9
{DT00 DP01 DX32 EMNT}	10 758.4	1.7	5426.9	1.7	5120.4	1.7

Application: MW Starts (8)

- Best results seem to occur when using 3 weather variables: **Days where min temp is below 0 F** (DT00), **Days where prec exceeds 0.1 in** (DP01), and **Days where max temp is below 32 F** (DX32).

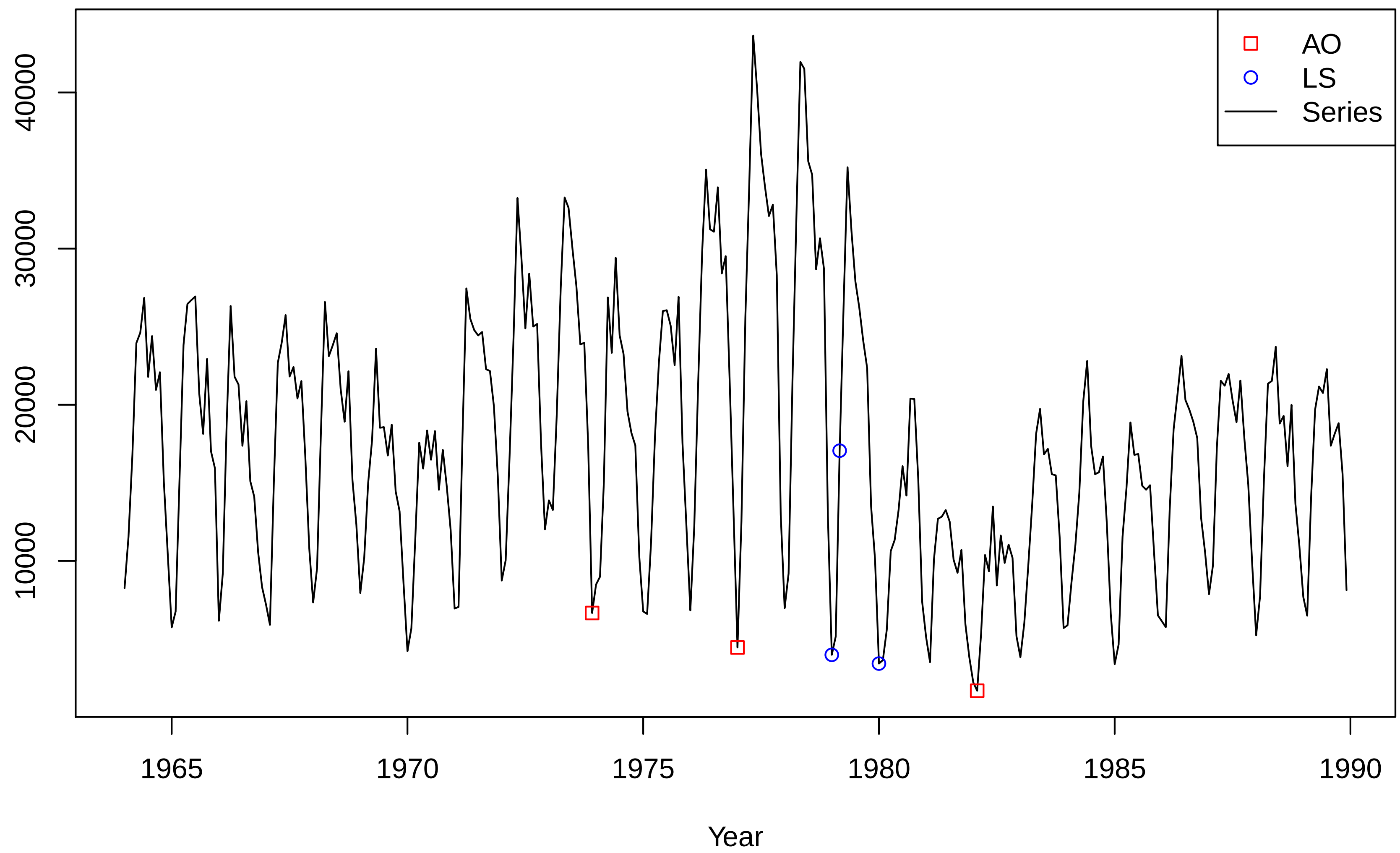
- DT00 and DX32 are (strongly) correlated (0.74).

- DP01 is close to uncorrelated with the other two.

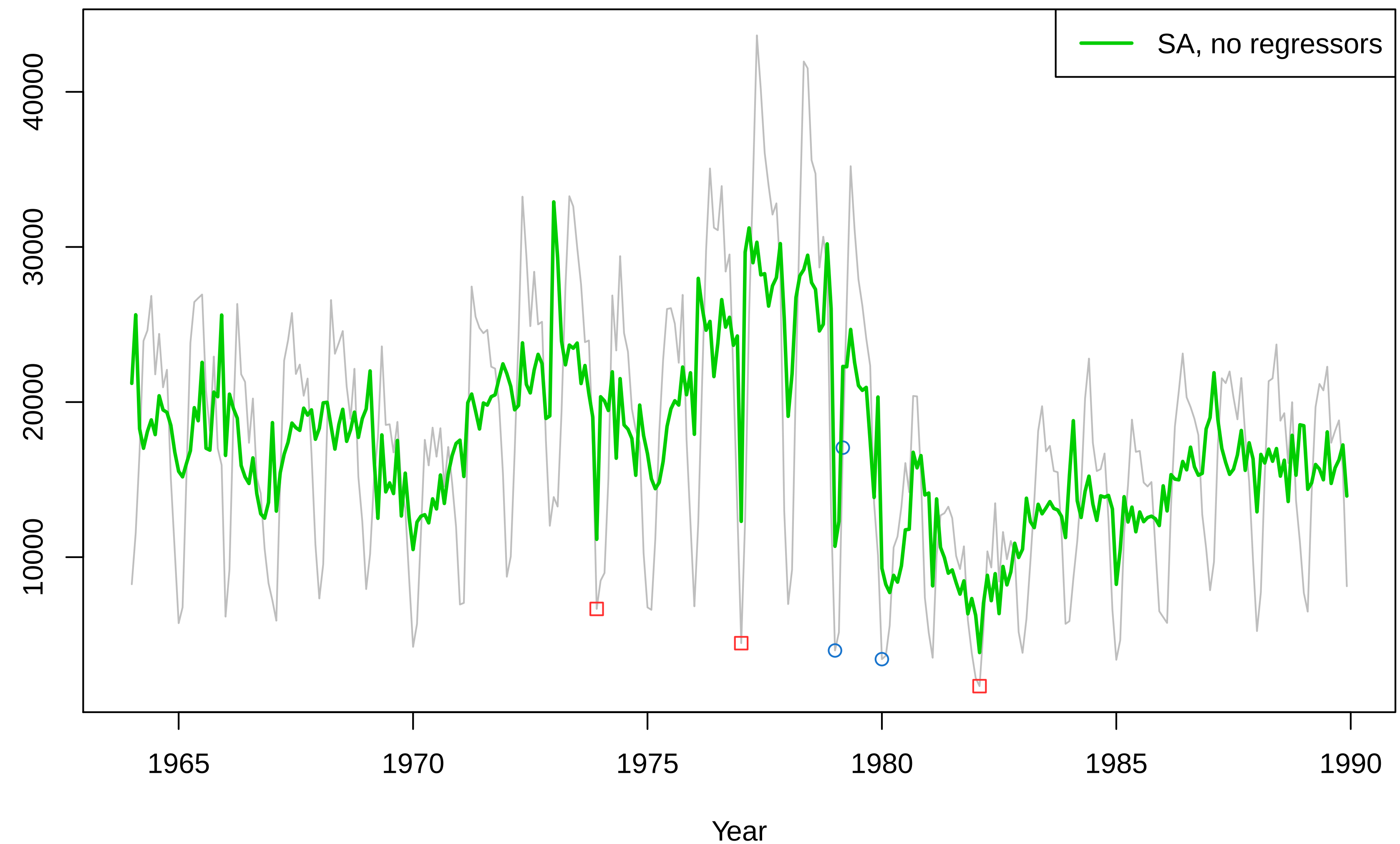
- Outliers identified using the set of 3: point outliers (Dec 1973, Feb 1982, Dec 2011), level shift (Jan 1980) (+ subspan-only point outlier Jan 2014).

- All 3 regressors are significant in each span.

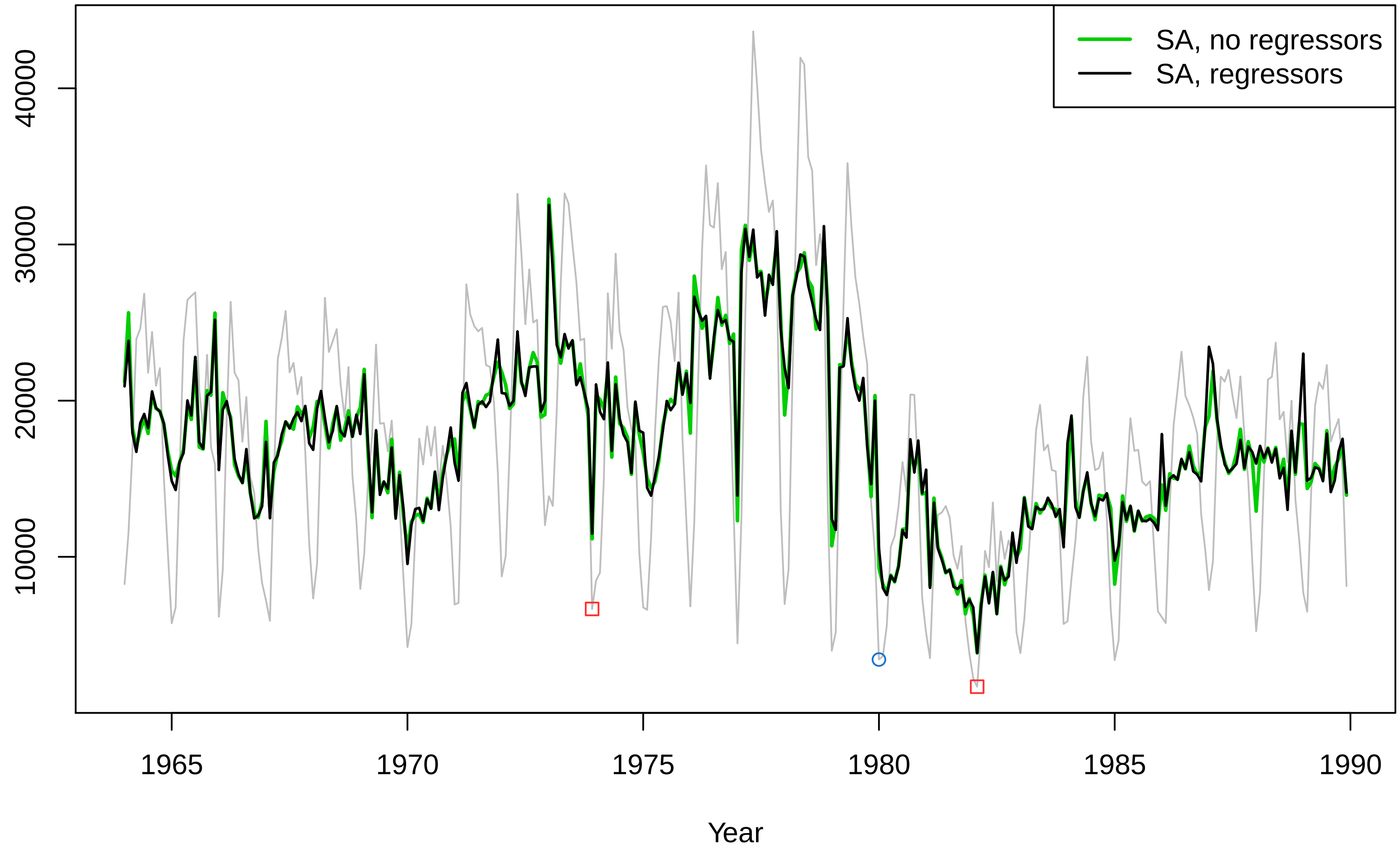
Application: MW Starts (9)



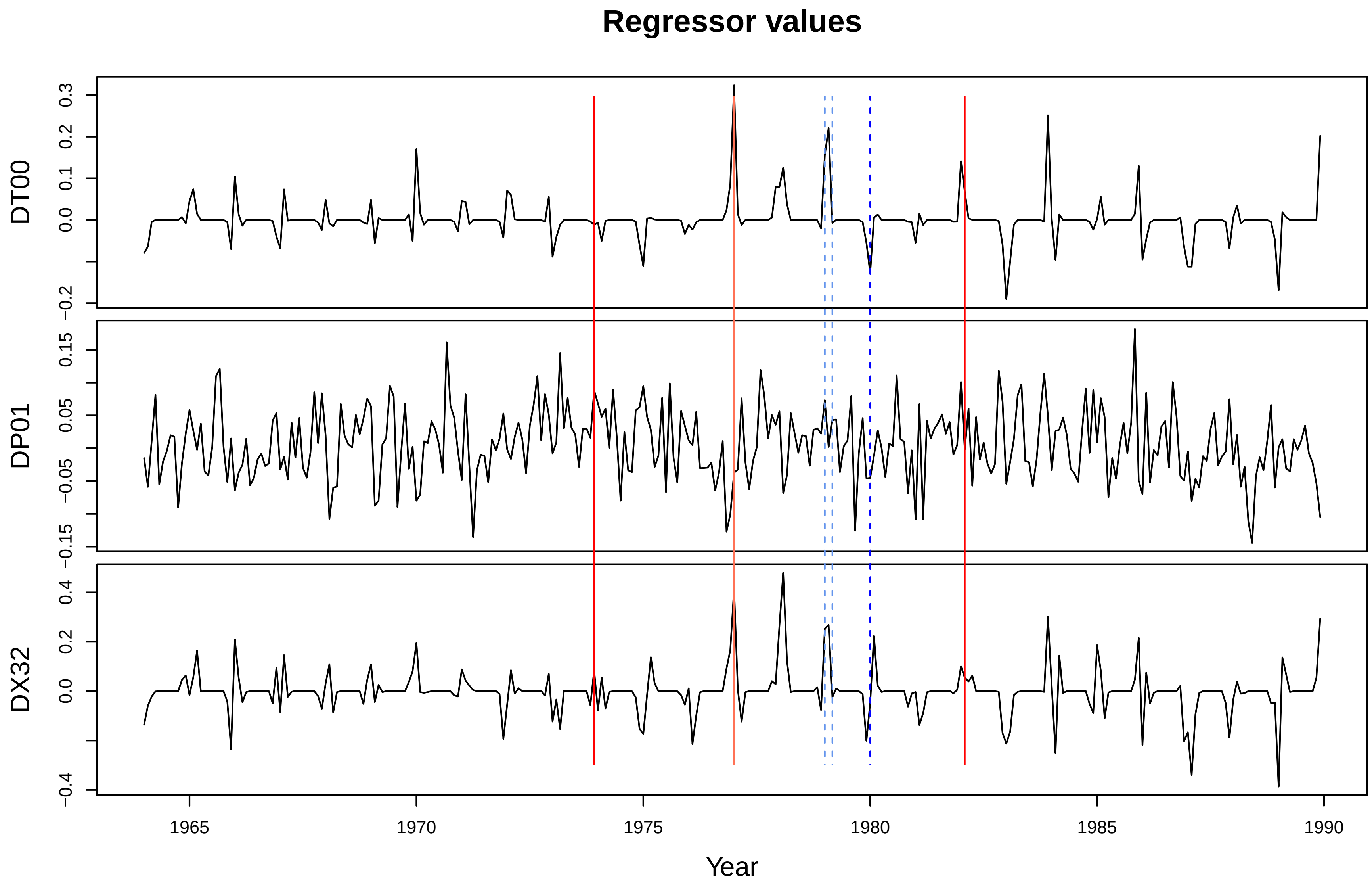
Application: MW Starts (10)



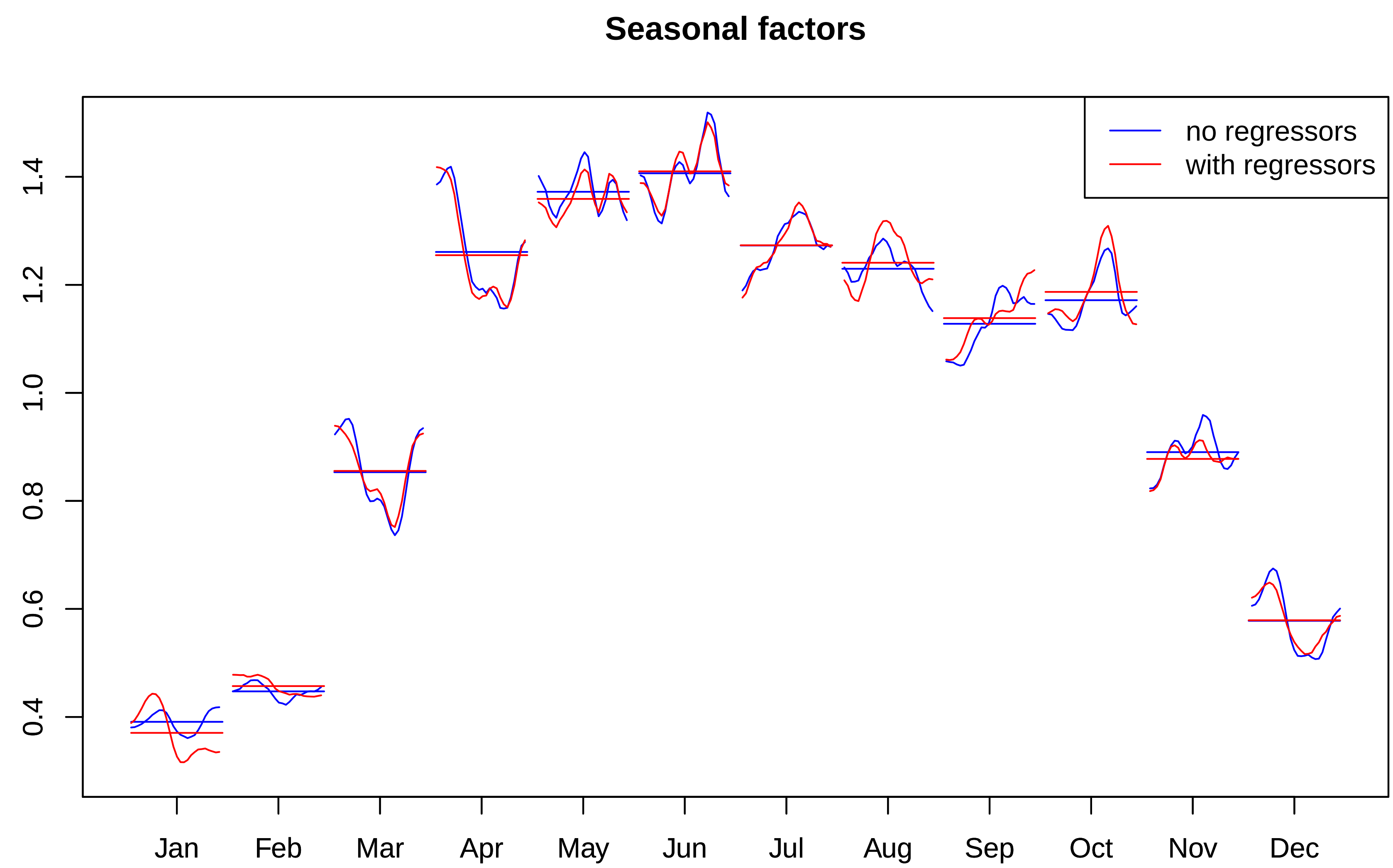
Application: MW Starts (11)



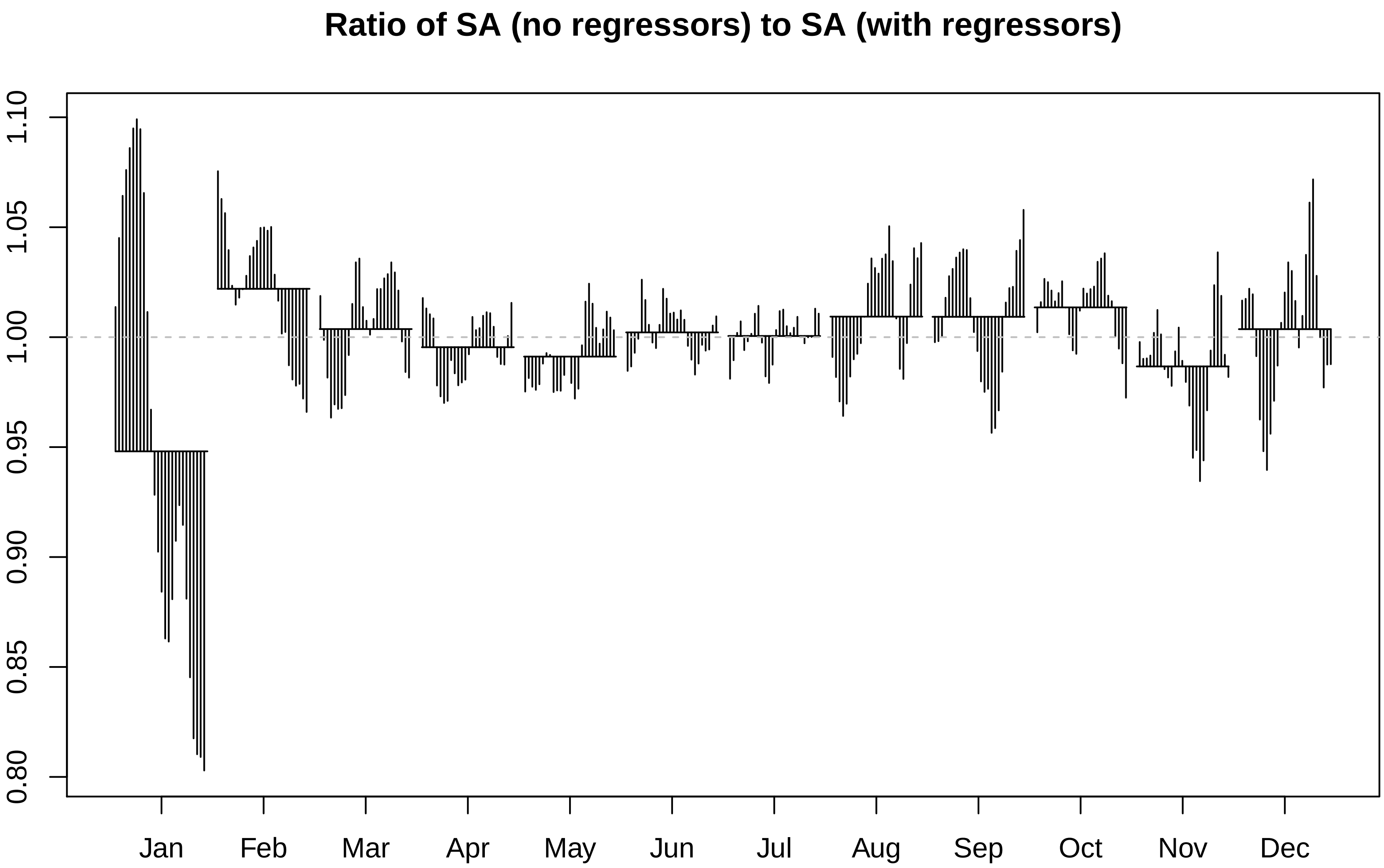
Application: MW Starts (12)



Application: MW Starts (13)



Application: MW Starts (14)



Application: MW Starts (15)

- Looking at the first subspan (1964 through 1989), half of outliers identified by model with no weather regressors are accounted for by a model with weather regressors.
- The regressor set used has a small effect on the seasonal factors for some months, but a much more noticeable effect in other months.
- In particular, minimal difference between the two seasonal adjustments in June and July.
- In addition, noticeable differences between seasonal adjustments in January (and to a lesser extent, February).
- Variability of the ratio not constant across seasons.

Conclusions and other thoughts

- Weather adjustment does help explain some unusual observations in regional construction series, but not all.
- Failure to account for weather effects can lead to sizable deviations in the seasonal adjustment.
- Thresholds different from the default ones used by NCDC may have more value, but require the use of daily data instead of monthly data.
- This approach may be better suited to smaller geographies; large geographies can experience large disparity in weather conditions across the region.
- Look at using sets of 12 seasonal (monthly) dummies in place of single regressors.

Contact: osbert.c.pang@census.gov