Paper 233-25

Cinema and Seasonality: A SAS® System PROC X11 Application

Joseph Earley, Loyola Marymount University, Los Angeles
Jia Lin Wang, Loyola Marymount University, Los Angeles

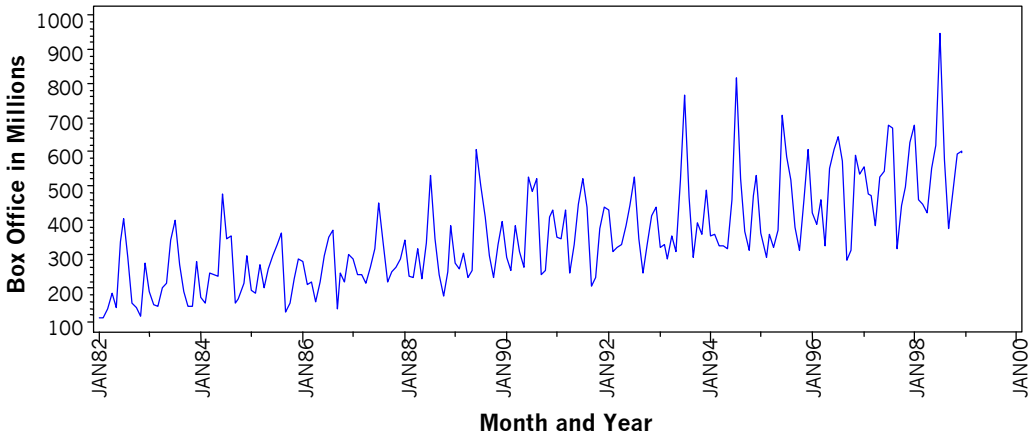
Abstract

Cinema in the United States is a multi-billion dollar annual business which exhibits a strong seasonal pattern. In order to insure that their films have a high probability of success, film companies must have knowledge of competitive film openings as well as the size of their potential audience. An important aspect of deriving this information is the ability to model monthly seasonal components of the time series. The SAS procedure, PROC X11 provides the software tool to achieve this goal. This paper illustrates how PROC X11 may be used to detect seasonal patterns in the time series for United States Domestic Box Office Revenue. The study uses the ARIMA option of PROC X11 as well as the new technique of sliding spans analysis.

Introduction

The SAS System offers a procedure, PROC X11, which may be used to seasonally adjust either monthly or quarterly data. The procedure was originally developed by the U.S. Bureau of the Census in 1967, and has achieved widespread use for the seasonal adjustment of government data. The SAS implementation of the X11 procedure now incorporates an ARIMA option, which was developed bv researchers at Statistics Canada.

Original Data
Box Office Gross Monthly for 1982 - 1998



Seasonality Detection and PROC X11

Seasonal adjustment of a time series presupposes that a series can be partitioned into the following components[1]:

$$O_t = S_t\ C_t\ D_t\ I_t$$

where: O_t Original series
S_t Seasonal component
C_t Trend cycle component
D_t Trading day component
I_t Irregular component

The basic idea behind the procedure is that the series components can be filtered from the original time series using a series of symmetrical moving average filters. Following the original work of Arthur Burns and Wesley Clair Mitchell, the seasonal component of the series is treated as noise which should be filtered from the series in order to get at the more important trend-cycle component. Using monthly data, a general outline of the X11 procedure would be[2]:

1. Use a 12 month centered moving average as an

[1] SAS OnlineDoc™: Version 7-1 Chapter 21, The X11 Procedure

[2] Jaditz: Seasonality: Economic data and model estimation.

estimate of the trend-cycle component.

2. Difference the original series and the centered moving average for an estimate of the sum of the seasonal and irregular components.
3. Apply a 5 term moving average separately to each month in order to extract seasonal factors.
4. Further smooth the seasonal factors and subtract them to yield an estimate of the irregular components.
5. Further smooth the initial estimate of the trend and re-estimate the seasonals and irregulars.
6. Finally, the seasonal adjusted series is obtained by subtracting the final estimate of the seasonal from the raw series.

Sliding Spans Analysis

Version 7.0 of the SAS System for Windows incorporates a new feature - sliding spans analysis.[3] The purpose of sliding spans analysis is to determine whether or not the time series under investigation is suitable for seasonal adjustment. Sliding spans analysis is based on the idea that the components of a stable time series should be reasonably invariant to small changes in the amount of data used for estimation. If there are large changes in the components due to small changes in the amount of data, the series may not be suitable for seasonal adjustment. Sliding spans analysis works by sequentially applying seasonal adjustment to moving spans of data. Usually the initial span is eight years. The following notation is used in sliding spans analysis[4]

X_t	Component X in month(or quarter) t, computed from data in span k.
$S_t(k)$	Seasonal factors

N_t = {k: span k contains month t}

Thus since we are interested in how much the seasonal factors vary from month to month, the following is calculated for the additive model:

$$\text{Max}_{k \in Nt}\ S_t(k) - \text{min}_{k \in Nt}\ S_t(k)$$

And for the multiplicative model, we define the MPD_t - the maximum percent difference at time period t, as:

$$MPD_t = \frac{100\left(\text{Max}_{k \in Nt}\ S_t(k) - \text{min}_{k \in Nt}\ S_t(k)\right)}{\text{min}_{k \in Nt}\ S_t(k)}$$

A large MPD_t indicates that the series seasonality factor at time period t may be unreliable. Findley and others have concluded that a series which has more than 25 % or more of the MPD values exceeding 3.0 % is almost always unstable.

Empirical Results

The SAS X11 procedure output is divided into seven major parts, as follows:

A. Prior adjustments, if any
B. Preliminary estimates of irregular component weights and regression trading day factors
C. Final estimates of above
D. Final estimates of seasonal, trend-cycle and irregular components
E. Analytical tables
F. Summary measures
G. Charts

The original series, stable seasonality, moving seasonality and combined test for the presence of identifiable seasonality, final seasonal factors and final seasonally adjusted series derived by using PROC X11 are shown below. The stable seasonality test is a one-way analysis of variance using the seasons - months- as the factor of Table D8, the final unmodified SI ratios. Large F values indicate that a significant amount of variation in the SI ratios is due to months i.e. seasonality exists. Technically, the null hypothesis is that there is no effect due to months. An F value of 47.31 with a p-value of less that 0.0001 indicate that months do have an influence on the time series. The moving seasonality test is a two-way analysis of variance, using both months and years. Using the same unmodified SI ratios, this test allows for determining whether or not there is slowly evolving seasonality in the time series. The null hypothesis in this case is that years have no significant effect after accounting for variation due to months. An F value of 1.59 with the associated p-value of 0.0761 indicates that the null hypothesis cannot be rejected at either the 1.0% or the 5% levels of significance.

PROC X11 now presents a new test for seasonality which combines the previous two F tests with the Kruskal-Wallis chi-squared test for stable seasonality to determine what is called identifiable seasonality. The null hypothesis is that identifiable seasonality is not present. The result of this test for the boxinmillions variable is that identifiable seasonality was shown to be present in the series.

X11 Procedure Output

Following is a subset of output derived from using PROC X11. The output includes the original series, the final seasonal factors, and the final seasonally adjusted series.

[3] See Chapter 21: The X11 Procedure for a discussion of the SAS implementation of this analysis.

[4] See Findley, Monsell, Shulman and Pugh, 1990.

B1 Original Series

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1982	111.905	114.594	138.411	184.208	141.667	335.024	405.152	288.503	155.066	142.546	116.517	274.932
1983	187.595	150.246	149.105	203.240	213.477	342.190	400.655	269.151	189.898	145.146	148.048	276.397
1984	171.211	154.184	241.809	237.633	237.315	476.368	344.969	354.336	156.725	166.281	215.285	296.157
1985	191.428	183.131	267.909	203.392	255.924	294.994	326.847	361.837	131.812	156.598	228.421	284.448
1986	277.173	209.483	216.439	157.667	217.890	294.330	347.153	370.629	140.489	242.466	219.741	299.718
1987	286.600	241.367	238.604	212.786	260.291	315.997	451.792	330.402	218.522	247.900	262.185	287.146
1988	341.016	233.484	232.871	313.336	228.641	330.287	531.895	345.401	240.753	175.121	248.699	380.593
1989	273.712	255.401	304.581	231.852	252.272	606.326	500.401	407.493	294.281	232.081	327.966	395.878
1990	288.581	251.641	382.812	308.733	259.558	524.107	483.938	522.775	239.265	252.809	407.056	428.330
1991	347.345	346.097	429.090	243.224	327.070	445.815	522.160	435.281	206.174	229.820	372.549	438.600
1992	430.342	306.238	320.022	329.365	384.645	440.642	524.565	343.266	243.811	329.910	413.499	435.621
1993	319.122	327.564	287.494	351.744	305.765	516.372	765.369	465.199	291.653	390.129	355.358	488.265
1994	353.630	357.803	321.734	325.706	313.373	457.280	814.949	522.953	366.645	309.585	472.607	530.343
1995	362.623	290.551	355.299	320.433	369.643	707.530	584.437	515.816	378.957	313.166	460.351	607.326
1996	421.228	386.402	456.606	323.090	549.251	606.684	643.399	572.685	282.260	310.308	590.403	535.612
1997	553.072	475.622	470.474	381.406	526.033	541.541	675.239	668.290	314.325	441.533	494.328	625.501
1998	678.467	459.072	446.095	419.604	551.151	619.477	945.625	575.947	372.992	485.890	592.440	602.761

Stable Seasonality Test

Source	Sum of Squares	DF	Mean Square	F Value
Between Months	135495	11	12318	47.31
Error	49992	192	260.4	
Total	185487	203		

Probability of a Larger F is < 0.0001

Moving Seasonality Test

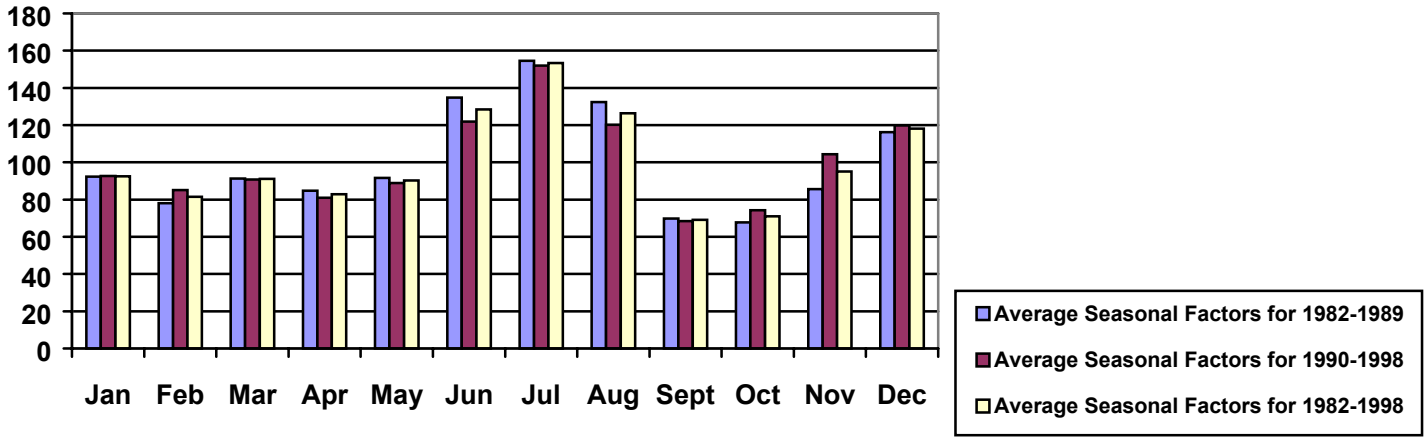
Source	Sum of Squares	DF	Mean Square	F Value
Between Months	49305465	16	3081592	1.59
Error	341571835	176	1940749	
Total	390877300	192		

Probability of a Larger F is 0.0761

Combined Test for the Presence of Identifiable Seasonality:
Result: Identifiable seasonality present

D10 Final Seasonal Factors

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1982	76.983	68.911	90.267	94.999	92.729	151.048	159.965	134.348	64.351	64.787	79.300	120.439
1983	79.498	70.619	91.179	92.925	93.478	147.940	157.608	136.005	63.456	65.030	80.280	119.778
1984	84.221	73.553	92.593	89.172	95.155	141.098	154.419	136.692	63.146	66.396	82.450	118.186
1985	91.318	77.194	93.218	85.180	95.056	133.435	151.313	136.781	64.658	67.541	85.081	117.143
1986	97.349	80.896	93.778	81.212	94.000	126.188	150.113	134.773	68.061	68.551	88.261	116.308
1987	101.306	82.988	93.986	78.990	89.612	123.963	149.544	133.323	71.476	68.657	91.155	116.100
1988	101.290	84.148	95.119	77.929	86.033	124.181	150.504	129.839	73.655	68.669	94.023	115.974
1989	99.306	83.936	96.672	79.657	83.116	126.526	149.826	126.406	73.798	68.002	97.899	117.539
1990	94.876	84.158	97.347	81.769	82.058	128.142	150.447	122.818	72.533	67.776	101.300	119.256
1991	91.836	84.623	95.729	83.887	80.895	128.320	152.380	120.993	71.021	68.056	104.191	120.855
1992	89.823	85.433	92.517	84.372	81.549	126.259	155.523	118.990	71.029	69.114	105.063	122.357
1993	89.881	85.422	89.230	83.883	83.345	123.992	157.159	117.550	71.841	69.493	105.917	123.187
1994	91.220	85.535	87.264	81.643	86.472	121.702	156.048	118.313	72.351	70.801	104.514	123.530
1995	93.216	85.524	87.355	79.203	89.851	119.578	153.597	120.077	71.265	72.714	104.170	121.046
1996	95.801	85.770	88.795	76.860	94.166	117.698	149.602	122.078	69.225	75.133	103.228	119.340
1997	97.801	86.051	90.197	75.441	97.758	116.741	147.206	122.132	66.907	76.897	103.793	116.618
1998	99.169	86.468	91.117	74.886	99.957	117.156	145.444	122.348	65.081	77.748	103.586	115.362



Next is the output derived from the sliding spans option - SSPAN.

S 0.A Sliding Spans Analysis for variable boxinmills: Number, Length, and Dates of Spans

Number of Spans 4

Length of Spans (Months) 96

Span	Begin	End
1	JAN1988	DEC1995
2	JAN1989	DEC1996
3	JAN1990	DEC1997
4	JAN1991	DEC1998

Seasonality Type	Span 1	Span 2	Span 3	Span 4
Stable Seasonality	25.94	25.64	23.96	23.96
Moving Seasonality	0.416	0.430	0.726	0.967
Identifiable Seasonality?	Yes	Yes	Yes	Yes

S 1.A Range Analysis of Seasonal Factors: Means of Seasonal Factors for Each Month (Movements Within a Month Should be Small)

Month	Span 1	Span 2	Span 3	Span 4	Difference	Maximum Percentage Average
January	89.9	88.982	91.393	92.424	3.87	90.675
February	82.835	83.543	84.587	85.455	3.16	84.105
March	91.841	92.262	90.972	85.817	7.51	90.223
April	83.767	82.735	81.791	80.395	4.19	82.172
May	79.717	81.309	86.464	91.12	14.30	84.653
June	125.7	127.65	123.14	119.27	7.03	123.94
July	159.85 MAX	154.65 MAX	149.64 MAX	154.84 MAX	6.82	154.74 MAX
August	122.11	121.22	120.97	117.95	3.53	120.56
September	74.565	71.953	70.303 MIN	69.339 MIN	7.54	71.54 MIN
October	67.958 MIN	69.289 MIN	72.788	77.873	14.59	71.977
November	102.01	106.07	106.26	103.8	4.17	104.54
December	121.32	122.27	122.26	120.27	1.66	121.53

Maximum Percent Difference = 100 * (max - min) / min

Average = average taken over all 4 spans

S 1.B Summary of Range Measures

	Range Means	R-R Means	Min SF	Max SF	Range SF	R-R SF
Span 1	91.89	2.352	91.89	178.2	112.7	2.721
Span 2	85.37	2.232	85.37	160.1	94.42	2.438
Span 3	79.34	2.129	79.34	154.6	90.01	2.392
Span 4	85.50	2.233	85.50	162.2	97.00	2.488
Ave. S	83.20	2.163	83.20	178.2	113.5	2.757

R-R = Range Ratio = Max / Min, SF = Seasonal Factors

S 2.A.1 Breakdown of 3.00% or Greater Differences in the Seasonal Factors by Month

Month	Exceed Count	Total Count	Average Maximum Percentage Difference
January	4	9	3.780
February	4	9	2.653
March	7	9	7.088
April	1	9	1.656
May	8	9	7.873
June	6	9	4.008
July	7	9	7.768
August	5	9	4.046
September	9	9	8.695
October	8	9	10.46
November	6	9	3.568
December	3	9	2.722

S 2.A.2 Breakdown of 3.00% or Greater Differences in the Seasonal Factors by Year

Year	Exceed Count	Total Count	Average Maximum Percentage Difference
1989	5	12	3.674
1990	7	12	4.203
1991	8	12	7.504
1992	7	12	5.220
1993	6	12	4.354
1994	7	12	5.306
1995	12	12	7.775
1996	10	12	6.642
1997	6	12	3.563

S 2.A.3 Breakdown Summary of Flagged Observations Based on 3.00% Cutoff for Seasonal Factors

Type	Number Found	Total Tested	Percentage
Flagged MPD	68	108	62.96

For Flagged MPD, usually, 15% is too high, and 25% is much too high.

S 2.A.3 Breakdown Summary of Flagged Observation Based on 3.00% Cutoff for Seasonal Factors

Type	Number Found	Total Tested	Percentage
Sign Change	2	108	1.85
Turning Point	31	84	36.90
Flagged Sign Change	2	2	100.00
Flagged Turning Point	22	31	70.97

For Flagged MPD, usually, 15% is too high, and 25% is much too high.

S 2.B Histogram of Flagged Observations Using 3.00% Base Cutoff for Seasonal Factors (See Table S 7.A)

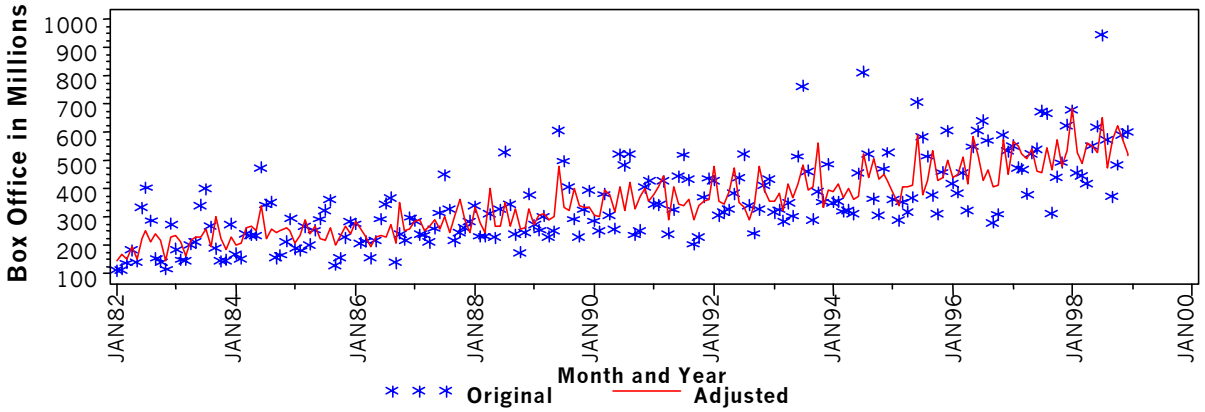
Symbol Used	Range	Number Found
%	[3.0%-4.0%)	14.00
%%	[4.0%-5.0%)	10.00
%%%	[5.0%-6.0%)	9.000
%%%%	>=6.0%	35.00

S 2.C Statistics for Maximum Percentage Difference of the Seasonal Factors

	Minimum	0.00	
	25th Percentile	2.07	
	median	4.04	
->	65th percentile	5.30	<-
	75th percentile	7.46	
->	85th percentile	9.56	<-
	maximum	19.41	

Adjusted Data

Box Office Gross Monthly for 1982 - 1998



Conclusion

Results from the X11 procedure indicate that the time series United States Domestic Box Office Revenue does exhibit seasonality. The stable seasonality test indicates that months have a significant effect on the series. The moving seasonality test did not indicate a significant year effect on the series. The high mean percentage difference (MPD) from sliding spans analysis however casts doubt on whether or not the series is suitable for seasonal adjustment.

Contact Information

Joseph Earley
Department of Economics
Loyola Marymount University
7900 Loyola Blvd.
Los Angeles, CA 90045-8410
Phone: (310) 338-4572
Fax: (310) 338-1950
e-mail: jearley@lmumail.lmu.edu

References

Bernstein, Jake Seasonality: Systems, Strategies, and Signals New York: John Wiley & Sons, 1998.

Findley, David F., Monsell, Brian C., Shulman, Holly B., and Pugh, Marian "Sliding-Spans Diagnostics for Seasonal and Related Adjustments", Journal of the American Statistical Association, June 1990, Vol. 85, No. 410, pp. 345-355.

Jaditz, Ted, "Seasonality: Economic Data and Model Estimation", Monthly Labor Review, December 1994, pp. 17-22.

McIntire, Robert J. "Revision of Seasonally Adjusted Labor Force Series", Bureau of Labor Statistics. Http://stats.bls.gov/cpsrs.htm.

Miron, Jeffrey A., The Economics of Seasonal Cycles Cambridge, Massachusetts: The MIT Press, 1996.

Nerlove, Marc, Grether, David M., and Carvalho, Jose L. Analysis of Economic Time Series: A Synthesis, New York: Academic Press, 1979.

SAS Institute Inc., SAS/ETS User's Guide. Version 6. Second Edition, Cary, NC:SAS Institute Inc., 1993.

SAS Institute Inc., SAS/ETS Software: Applications Guides 1 and 2, Version 6, First Edition, 1993.

SAS Institute Inc., SAS/STAT User's Guide, Vols I and II, 1990 Version 6, Fourth Edition. Cary, N.C.: SAS Institute Inc.