BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-84/01

SEASONAL DECOMPOSITION OF DETERMINISTIC EFFECTS

by

William Bell

U.S. Bureau of the Census
Washington, D.C. 20233

Recommended by: Myron J. Katzoff

Report completed: November 17, 1983

Report issued: January 10, 1984

SEASONAL DECOMPOSITION OF

DETERMINISTIC EFFECTS

William Bell

U.S. Bureau of the Census

November 17, 1983

Seasonal Decomposition of Deterministic Effects

Let f_t be a deterministic function of time. Suppose we have modeled a time series as

$$z_t = f_t + (z_t - f_t).$$

To seasonally adjust z_t we must assign part of f_t to the seasonal component. The remaining part of f_t can be broken down into an indigenous f-effect (such as trading day and holiday) and part of the trend. We first show how to remove the seasonal part for a general f_t and then show how to decompose trading day and holiday effects.

Removing Seasonality from f_t

Suppose $f_t = fs_t + fn_t$ where fs_t is the seasonal part of f_t and fn_t the nonseasonal part. Then we require that:

(1) $fs_t = fs_{t+12j} \quad j = \pm 1, \pm 2, \ldots$

(2) $U(B)fs_t = 0$ where $U(B) = 1+B+\ldots+B^{11}$, so fs_t sums to zero over any twelve consecutive months

(3) fs_t and fn_t should be orthogonal over the long run, that is

$$\frac{1}{n} \sum_{1}^{n} (fs_t \times fn_t) \to 0 \qquad \text{as } n \to \infty$$

To achieve this we proceed as follows:

(i) compute $\quad \bar{f}_i = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} f_{i+12j} \qquad i=1,\ldots,12,$

the long run monthly means of f_t (we assume these exist)

(ii) compute $\bar{\bar{f}} = \frac{1}{12} \sum_{i=1}^{12} \bar{f}_i$ the overall long run mean of f_t

(iii) if (j, i) corresponds to t (year j , month i) set

$$fs_t = \bar{f}_i - \bar{\bar{f}} \qquad fn_t = f_t - \bar{f}_i + \bar{\bar{f}}$$

Thus, we remove the monthly means and add back the overall mean to get fn_t.

The above procedure is equivalent to regressing f_t on the variables SM_{it} $i = 1,\dots,11$ over an infinite time horizon and taking the predicted values for fs_t and the residuals for fn_t where the SM_{it} are as follows:

month of t	SM_{1t}	SM_{2t}	SM_{3t}	...	SM_{10t}	SM_{11t}
Jan.	1	0	0		0	0
Feb.	0	1	0		0	0
Mar.	0	0	1		0	0
Apr.	0	0	0		0	0
May	0	0	0		0	0
June	0	0	0		0	0
July	0	0	0	...	0	0
Aug.	0	0	0		0	0
Sept.	0	0	0		0	0
Oct.	0	0	0		1	0
Nov.	0	0	0		0	1
Dec.	-1	-1	-1		-1	-1
Jan.	1	0	0		0	0
Feb.	0	1	0		0	0
	⋮			⋮		

Notice that SM_{it} $i = 1,\dots,11$ and 1_t (=1 for all t) span the same space as the monthly mean variables M_{1t} = 1 in Jan. and 0 otherwise, M_{2t} = 1 in Feb. and 0 otherwise, etc. All deterministic seasonal components are in the space spanned by the SM_{it} $i=1,\dots 11$, which is orthogonal to the space spanned by 1_t. If f_t has been decomposed as $fs_t + fn_t + x_t$ where x_t is another effect (for example trading day or holiday) we can check that the decomposition is correct by checking that in the long run

(i) fn_t and x_t (or the independent variables used in them) are orthogonal to SM_{it} i=1,...,11

(ii) fs_t can be perfectly predicted from the SM_{it} i=1,...,11

(iii) the independent variables used in x_t are orthogonal to all the trend independent variables in fn_t (often just 1_t, but perhaps including t, t^2, or something else).

An alternative approach to removing long term monthly means and adding back the long term overall mean would be to just regress f_t on $SM_{1t},...,SM_{11,t}$ over the observed stretch of data (t=1,...,m say) and take the residuals. This makes $(fn_1,...,fn_m,)'$ orthogonal to $(SM_{i1},...,SM_{im})'$ i=1,...,11 but has the disadvantages that

(i) the definition of seasonality and trend change depending on the time frame of the data set

(ii) the decomposition may have to be recomputed for each new data set instead of just substituting in some parameter

(iii) the orthogonality properties need not be preserved in the long run, or even as a few new data points are added

Seasonal Decomposition of Trading Day Effects

I. Flow Series

In this case the trading day plus level effect is $\sum_{i=1}^{7} \beta_i X_{it}$ which may be written as (X_{it} = # of i days in month t i=1,...,7)

$$\sum_{i=1}^{7} \beta_i X_{it} = \sum_{i=1}^{6} \hat{\beta}_i \; (X_{it} - X_{7t}) + \bar{\beta} \; m_t$$

$$\bar{\beta} = \frac{1}{7}\sum_{i=1}^{7} \beta_i , \ \hat{\beta}_i = \beta_i - \bar{\beta}, \ m_t = \sum_{i=1}^{7} X_{it} = \text{length of month } t$$

Notice

$$\frac{1}{n}\sum_{t=1}^{n} (X_{it} - X_{7t}) \to 0 \qquad i = 1,\ldots 6$$

$$\frac{1}{n}\sum_{t=1}^{n} (X_{it} - X_{7t}) SM_{jt} = \frac{1}{n}\sum_{t=1}^{n} (X_{it} - X_{7t}) M_{jt} - \frac{1}{n}\sum_{t=1}^{n} (X_{it} - X_{7t}) M_{12,t}$$

$$\to 0 + 0 \quad \text{as } n \to \infty \quad i=1,\ldots,6 \ j=1,\ldots,11$$

Thus $\sum_{i=1}^{6} \hat{\beta}_i (X_{it} - X_{7t})$ is orthogonal to the seasonal variables SM_{it} and the constant variable 1_t, and so is part of the pure trading day effect.

We now need to decompose $\bar{\beta}\ m_t$, or just m_t. The long term monthly means of m_t are (ignoring the loss of a leap year every 400 years)

31 Jan., March, May, July, Aug., Oct., Dec.

30 April, June, Sept., Nov.

28.25 Feb.

The mean of these is $\frac{365.25}{12} = 30.4375$. Thus, the seasonal component of m_t is

$$\xi_t = \begin{cases} .5625 & \text{Jan., March, May, July, Aug., Oct., Dec.} \\ -.4375 & \text{April, June, Sept., Nov.} \\ -2.1875 & \text{Feb.} \end{cases}$$

The deviations of m_t from its monthly means are zero for every month but February, and for February are

-.25 -.25 -.25 .75 -.25 -.25 -.25 .75 ...

We now write m_t as

$$m_t = (m_t - LF_t - 30.4375) + LF_t + 30.4375 = \xi_t + LF_t + 30.4375$$

where

$$LF_t = \begin{cases} -.25 \text{ in a non-leap year Feb.} \\ .75 \text{ in a leap Feb.} \\ 0 \text{ otherwise} \end{cases}$$

Here LF_t sums to zero over any 48 months, so it is a leap year effect which we assign to the trading day component. Its long term monthly means are zero so it is orthogonal to both the seasonal and constant trend components. Also notice $m_t - LF_t - 30.4375$ sums to zero over any twelve consecutive months. Notice that $(1-B^{12})\, m_t = 0 + (1-B^{12})\, LF_t + 0$ which is nonzero only in leap year Februaries and the immediately following Februaries.

Finally, we write the trading day plus level effect as

$$\sum_{i=1}^{7} \beta_i X_{it} = \sum_{i=1}^{6} \tilde{\beta}_i (X_{it} - X_{7t}) + \bar{\beta}\, LF_t \qquad \text{trading day effect}$$

$$+ \bar{\beta}\, \xi_t \qquad \text{seasonal effect}$$

$$+ \bar{\beta}(30.4375) \qquad \text{nonseasonal effect}$$

II. Stock Series

Let F_t be a monthly flow series and let

$$I_t = I_0 + \sum_{j=1}^{t} F_j$$

be the end of the month stock series, starting at I_0 at t=0. Assume the trading day effect in I_0 is zero. Then, if the trading day effect in F_j is $\sum_{i=1}^{7} \beta_i X_{ij}$, that in I_t is

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \beta_i X_{ij} = \sum_{j=1}^{t} \sum_{i=1}^{7} [(\beta_i - \bar{\beta}) X_{ij} + \bar{\beta} X_{ij}]$$

$$= \sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_{ij} + \bar{\beta} \sum_{j=1}^{t} m_j .$$

One way to look at $\sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_{ij}$ is to consider what happens on a daily basis- the appropriate $\tilde{\beta}_i$ is added in for each day in months 1, ...,t. If k_t is the type of day month t ends on and k_0 is the type of day just before the start of month 1, then if $k_0 < k_t$

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_{ij} = (\tilde{\beta}_{k_0+1}+\ldots+\tilde{\beta}_7+\tilde{\beta}_1+\ldots+\tilde{\beta}_{k_0}) + (\tilde{\beta}_{k_0+1}+\ldots+\tilde{\beta}_{k_0})+\ldots+$$

$$(\tilde{\beta}_{k_0+1}+\ldots+\tilde{\beta}_{k_0}) + \tilde{\beta}_{k_0+1}+\ldots+\tilde{\beta}_{k_t}$$

$$= \tilde{\beta}_{k_0+1}+\ldots+\tilde{\beta}_{k_t}$$

since each of the sums in parentheses is zero. Also, for $k_0 = k_t$ we see $\sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_{ij} = 0$, and for $k_0 > k_t$

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_{ij} = \tilde{\beta}_{k_0+1}+\ldots+\tilde{\beta}_7+\tilde{\beta}_1+\ldots+\tilde{\beta}_{k_t}$$
$$= -(\tilde{\beta}_{k_t+1}+\ldots+\tilde{\beta}_{k_0}).$$

Now define

$$\gamma_1 = \tilde{\beta}_1 + \gamma_7$$
$$\gamma_2 = \tilde{\beta}_1 + \tilde{\beta}_2 + \gamma_7$$
$$\vdots$$
$$\gamma_6 = \tilde{\beta}_1 +\ldots+ \tilde{\beta}_6 + \gamma_7$$
γ_7 arbitrary

Then

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_{ij} = \begin{cases} \gamma_{k_t} - \gamma_{k_0} & k_0 \le k_t \\ -(\gamma_{k_0} - \gamma_{k_t}) & k_0 > k_t \end{cases}$$

$$= \gamma_{k_t} - \gamma_{k_0}$$

$$= \sum_{k=1}^{7} \gamma_k \cdot I_t(k) - \gamma_{k_0}$$

where $I_t(k)$ is 1 if month t ends on a k day and is zero otherwise. By picking $\gamma_7 = -(\tilde{\beta}_1 + \ldots + \tilde{\beta}_{k_0})$ we make $\gamma_{k_0} = 0$. Then the trading day effect in I_t

is ($\bar{\gamma} = \frac{1}{7} \sum_{i=1}^{7} \gamma_i$)

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \beta_i X_{ij} = \sum_{i=1}^{7} \gamma_i I_t(i) + \bar{\beta} \sum_{j=1}^{t} m_j$$

$$= \sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) I_t(i) + \bar{\gamma} + \bar{\beta} \sum_{j=1}^{t} m_j .$$

Notice that $\bar{\gamma}$ is a level effect (part of trend) and

$$\frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) I_{j+12t}(i) = \sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) \frac{1}{n} \sum_{t=1}^{n} I_{j+12t}(i)$$

$$\to \sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) \frac{1}{7} = 0$$

so the long term monthly means (and also the overall mean) of $\sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) I_t(i)$ are

zero, and it is a trading day effect. It remains to decompose $\sum_{j=1}^{t} m_j$

From the previous results we can write

$$\sum_{j=1}^{t} m_j = \sum_{j=1}^{t} \xi_j + \sum_{j=1}^{t} LF_j + (30.4375)t$$

Now $\sum_{j=1}^{t} \xi_j = \sum_{j=1}^{t+12} \xi_j$, since ξ_j sums to zero over any twelve consecutive months; thus, this is a series of monthly means. The overall mean is

$$\frac{1}{12} \sum_{t=1}^{12} \sum_{j=1}^{t} \xi_j$$

which depends on the month the series starts in (t = 1). The overall means for each starting month are

$\bar{\xi}_k$

J	F	M	A	M	J
-.6979	-1.2604	.9271	.3646	.8021	.2396
J	**A**	**S**	**O**	**N**	**D**
.6771	.1146	-.4479	-.0104	-.5729	-.1354 .

Denote these by $\bar{\xi}_k$ k=1,...,12. If the series starts in month k then $\sum_{j=1}^{t} \xi_j - \bar{\xi}_k$ will be a seasonal effect and $\bar{\xi}_k$ will be part of the level of the series.

Notice $\sum_{j=1}^{t} \xi_j = \sum_{j=1}^{r} \xi_j$ (define $\sum_{j=1}^{0} \xi_j = 0$) where r is the remainder on taking t/12, that is, the number of months in 1,...,t in excess of complete years.

Next we must decompose $\sum_{j=1}^{t} LF_j$. Since LF_j sums to zero over any 48 consecutive months $\sum_{j=1}^{t} LF_j = \sum_{j=1}^{t+48} LF_j$, and since LF_j is nonzero only in February, $\sum_{j=1}^{t} LF_j$ is constant from any February to the following January, and it changes again in the next February. For example, if t = 1 corresponds to the January following a leap year, then $\sum_{j=1}^{t} LF_j$ behaves as follows:

t=1

J	F	M	...	J	F	M	...	J	F
0	-.25	-.25	...	-.25	-.5	-.5	...	-.5	-.75
M	...	J	F	M	...	J	F	M	...
-.75	...	-.75	0	0	...	0	-.25	-.25	...

Other patterns will arise depending on what month we start with. Since $\sum_{j=1}^{t} LF_j$ is constant from any February to the following January it is easy to see that $\sum_{j=1}^{t} LF_j$ is orthogonal to SM_{it} $i=1,\ldots,12$ over each such period, and hence $\sum_{j=1}^{t} LF_j$ is orthogonal to the seasonal component in the long run. All we need now do is remove the long term mean of $\sum_{j=1}^{t} LF_j$, which is the same as the mean over the first 48 months since $\sum_{j=1}^{t} LF_j = \sum_{j=1}^{t+48} LF_j$, or equivalently, over 48 months starting with the first February. This mean is the same as the mean of the four February values, and depends on whether the first, second, third, or fourth February is the leap year February, according to the following table

Feb. that is a leap year Feb.	Long run mean of $\sum_{j=1}^{t} LF_j$		
1st	$\frac{.75 + .5 + .25 + 0}{4}$	=	.375
2nd	$\frac{-.25 + .5 + .25 + 0}{4}$	=	.125
3rd	$\frac{-.25 - .5 + .25 + 0}{4}$	=	.125
4th	$\frac{-.25 - .5 - .75 + 0}{4}$	=	.375

Let $\delta_{\ell t}$ take on the ℓ^{th} value (ℓ =1,2,3,or4) in the above table for all t depending on whether the first, second, third, or fourth February is a leap year February. Then $\sum_{j=1}^{t} LF_j - \delta_{\ell t}$ is a leap year effect (part of the trading day effect) which is orthogonal in the long run to the seasonal and trend components.

We must also decompose the linear function t since it contains a seasonal component (if the series starts in January, t increases each year from January to December). Let

$$YR_t = 6.5 + 12[(t-1)/12] \quad [\ ] \text{ denotes the greatest integer function}$$

and consider

$$t = (t - YR_t) + YR_t.$$

YR_t makes a step increase of 12 every 12 months (6.5,18.5,30.5,...) and thus follows a linear trend when considered from year to year. We show that $t - YR_t$ sums to zero over any twelve months (let $t=12k+h$ $0<k, 1<h\leq 12$):

$$\sum_{j=t}^{t+11} (j-YR_j) = \frac{(t+11)(t+12)-(t-1)t}{2} - 12(6.5) - 12\sum_{j=t}^{t+11}\left[\frac{j-1}{12}\right]$$

$$= \frac{t^2+23t+132-t^2+t}{2} - 78 - 12\sum_{12k+h}^{12k+12}\left[\frac{j-1}{12}\right]$$

$$- 12\sum_{12(k+1)+1}^{12(k+1)+h-1}\left[\frac{j-1}{12}\right]$$

$$= 12t+66-78 - 12\sum_{12k+h}^{12k+12} k - 12\sum_{12(k+1)+1}^{12(k+1)+h-1}(k+1)$$

$$=12(12k+h)-12-12k(12-h+1)-12(k+1)(h-1)$$

$$= 144k +12h-12-144k + 12kh-12k-12kh-12h+12k+12$$

$$= 0.$$

Thus, $t - YR_t$ is a seasonal effect.

Combining all the above results, we write the trading day terms used in modeling a stock series as

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \hat{\beta}_i X_{ij} + \bar{\beta} \sum_{j=1}^{t} m_j = \sum_{j=1}^{7} (\gamma_i - \bar{\gamma}) I_t(i) + \bar{\gamma}$$

$$+ \bar{\beta}\{(\sum_{i=1}^{t} \xi_j - \bar{\xi}_k) + \bar{\xi}_k + (\sum_{j=1}^{t} LF_j - \delta_{\ell t})$$

$$+ \delta_{\ell t} + 30.4375((t-YR_t) + YR_t)\}$$

$$= \sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) I_t(i) + \bar{\beta}(\sum_{j=1}^{t} LF_j - \delta_{\ell t}) \quad \text{trading day effect}$$

$$+ \bar{\beta}\{(\sum_{j=1}^{t} \xi_j - \bar{\xi}_k) + 30.4375(t-YR_t)\} \quad \text{seasonal effect}$$

$$+ \bar{\gamma} + \bar{\beta} \{\bar{\xi}_k + \delta_{\ell t} + (30.4375)YR_t\} \quad \text{trend effect}$$

$$= \sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) I_t(i) + \bar{\beta}(\sum_{j=1}^{t} LF_j - \delta_{\ell t})$$

$$+ \bar{\beta}\{\sum_{j=1}^{t} (m_j - LF_j) - \bar{\xi}_k - (30.4375)YR_t\}$$

$$+ \bar{\gamma} + \bar{\beta} \{\xi_k + \delta_{\ell t} + (30.4375) YR_t\} .$$

Seasonal Decomposition of Holiday Effects

The same basic principles as used before can be used to break down holiday effects. We investigate the Easter effect modeled as $E_t = \alpha H(\tau,t)$ where $H(\tau,t)$ is the proportion of the time period τ days before Easter that falls in month t. If $\tau \leq 21$, then $H(\tau,t)$ is zero except in March and April. Let $\eta_{3\tau}$ denote the long term monthly mean of $H(\tau,t)$ for March, and $\eta_{4\tau}$ that for April. Notice $\eta_{3\tau} + \eta_{4\tau} = 1$. Let MA_t be $\eta_{3\tau}$ in March, $\eta_{4\tau}$ in April, and zero otherwise. Then we write

$$H(\tau,t) = [H(\tau,t) - MA_t] + [MA_t - \frac{1}{12}] + \frac{1}{12}$$

and notice that the long run monthly means of $H(\tau,t) - MA_t$ are all zero, and $MA_t - \frac{1}{12}$ sums to zero over any twelve consecutive months. Thus, $\alpha [H(\tau,t) - MA_t]$ is the (Easter) holiday effect, $\alpha [MA_t - \frac{1}{12}]$ is a seasonal effect, and $\frac{\alpha}{12}$ is part of the trend. To apply this we use the values for $\eta_{4\tau}$ given in the following table and use $\eta_{3\tau} = 1 - \eta_{4\tau}$.

τ	1	2	3	4	5	6	7	8	9	10
$\eta_{4\tau}$	.740	.730	.715	.698	.680	.663	.646	.630	.614	.599
τ	11	12	13	14	15	16	17	18	19	20
$\eta_{4\tau}$	.582	.564	.548	.531	.515	.498	.480	.463	.446	.429
τ	21	22	23	24	25					
$\eta_{4\tau}$	.412	.395	.379	.363	.349					

These values were obtained by averaging $H(\tau,t)$ values for Aprils from 1901 through

2100. We could expect slightly different values to be obtained if a different long stretch of data was used. The values for $\tau > 21$ are correct for $\eta_{4\tau}$, but when $\tau > 21$ $H(\tau, t)$ can be nonzero in February so that $\eta_{3\tau}$ is not $1 - \eta_{4t}$, and we could calculate a nonzero average, $\eta_{2\tau}$, of $H(\tau, t)$ for February. Then $\eta_{2\tau} + \eta_{3\tau} + \eta_{4\tau} = 1$. For $\tau = 22, 23,$ 24, or 25 the $\eta_{2\tau}$'s will be small, and $\eta_{3\tau}$ will not differ much from $1 - \eta_{4\tau}$.

Final Note

If one is concerned only with the adjusted data then it is not necessary to distinguish effects such as trading day or holiday effects from the seasonal effect. We need only decompose f_t into trend and other effects. The more detailed decomposition is useful if we wish to examine the relative magnitudes of the seasonal and trading day or holiday effects.