BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-95/01

CORRECTION TO "SEASONAL DECOMPOSITION OF DETERMINISTIC EFFECTS"

by

William R. Bell
Statistical Research Division
U. S. Bureau of the Census
Washington, DC 20233-9100

Report issued: January 5, 1995

1. Introduction

In Bell (1984) I discussed decomposition of a deterministic function of time f_t into seasonal and nonseasonal parts as $f_t = fs_t + fn_t$. This is important for decomposing regression effects, such as trading–day and holiday effects, in seasonal adjustment. The general approach presented in Bell (1984) for a monthly time series is as follows:

1. Compute the long–run monthly means of f_t, $\bar{f}_k$ for $k = 1, \dots, 12$.

2. Compute the long–run overall mean of f_t, $\bar{f} = (\bar{f}_1 + \cdots + \bar{f}_{12})/12$.

3. Let $fs_t = \bar{f}_{k(t)} - \bar{f}$, where $k(t)$ is the calendar month of time t (i.e., in any January $k(t) = 1$, in any February $k(t) = 2$, etc.). Let $fn_t = f_t - fs_t$.

It is assumed that the long–run monthly and overall means exist. With easy modifications, this approach applies to data with a seasonal period other than monthly.

Bell (1984) applied this procedure to flow and stock trading–day effects, and to Easter holiday effects. This note corrects the results given there for stock trading–day effects in regard to the treatment of the stock length–of–month effect. This effect arises as the cumulative summation over time of $\bar{\beta}m_t$, the flow length–of–month effect, where m_t denotes the length of month t. This can be broken into three terms, the last of which is a time trend (constant times t). It is the latter which I treat differently here. The treatment of a time trend in seasonal adjustment is discussed in Section 2. Section 3 then redoes the seasonal decomposition of stock trading–day effects in light of this result. Section 3 also improves on the notation of Bell (1984) to clarify the presentation.

2. Time Trends and Seasonal Adjustment

Consider the function $f_t = t$. Suppose the time frame of the series under consideration begins in January of some year. If the series contains the time trend effect βt, this will tend to make Decembers higher than Januaries (if $\beta > 0$). This could be interpreted as a seasonal effect, suggesting that $f_t = t$ be decomposed into seasonal and nonseasonal parts. This was the reasoning employed in Bell (1984). The result of this is shown in Figure 1 for a time series of eight years ($t = 1, \dots, 96$). The "nonseasonal" part of $f_t = t$ is a step function constant over calendar years that starts at 6.5 and increases by 12 each year. The graph of this is superimposed on the graph of $f_t = t$ in Figure 1. The seasonal part of $f_t = t$ increases linearly within calendar years from –5.5 in January to 5.5 in December, repeating this pattern each year. This is the dotted line graphed about the horizontal axis in Figure 1, and also in Figure 2.

A major problem with this approach is that the nature of the "seasonal component" of $f_t = t$ depends on the starting month of the series. For example, consider the seasonal component in $f_t = t$ for a series that starts in July. In this case the computed seasonal increases linearly from –5.5 in July to 5.5 in June. When this pattern is plotted over calendar years we get the solid line in Figure 2. Note the shift from the dotted line, the seasonal pattern when the series starts in January. Thus, we see the seasonal component in $f_t = t$ defined this way shifts depending on the starting month.

(Actually, a sign of problems with seasonal decomposition of $f_t = t$ is that the general scheme given in the introduction cannot be applied directly, since the long–run monthly and overall means of $f_t = t$ do not exist. Essentially, this difficulty was sidestepped in Bell (1984) by defining the seasonal fs_t as the difference between $f_t = t$ and the mean of $f_t = t$ for the calendar year that includes t, this difference being the same for all calendar years.)

Although an argument can be made for the seasonal decomposition of $f_t = t$ as shown in the graphs, it seems undesirable to use a seasonal decomposition that changes if the

Fig. 1. Time trend and seasonal decomposition

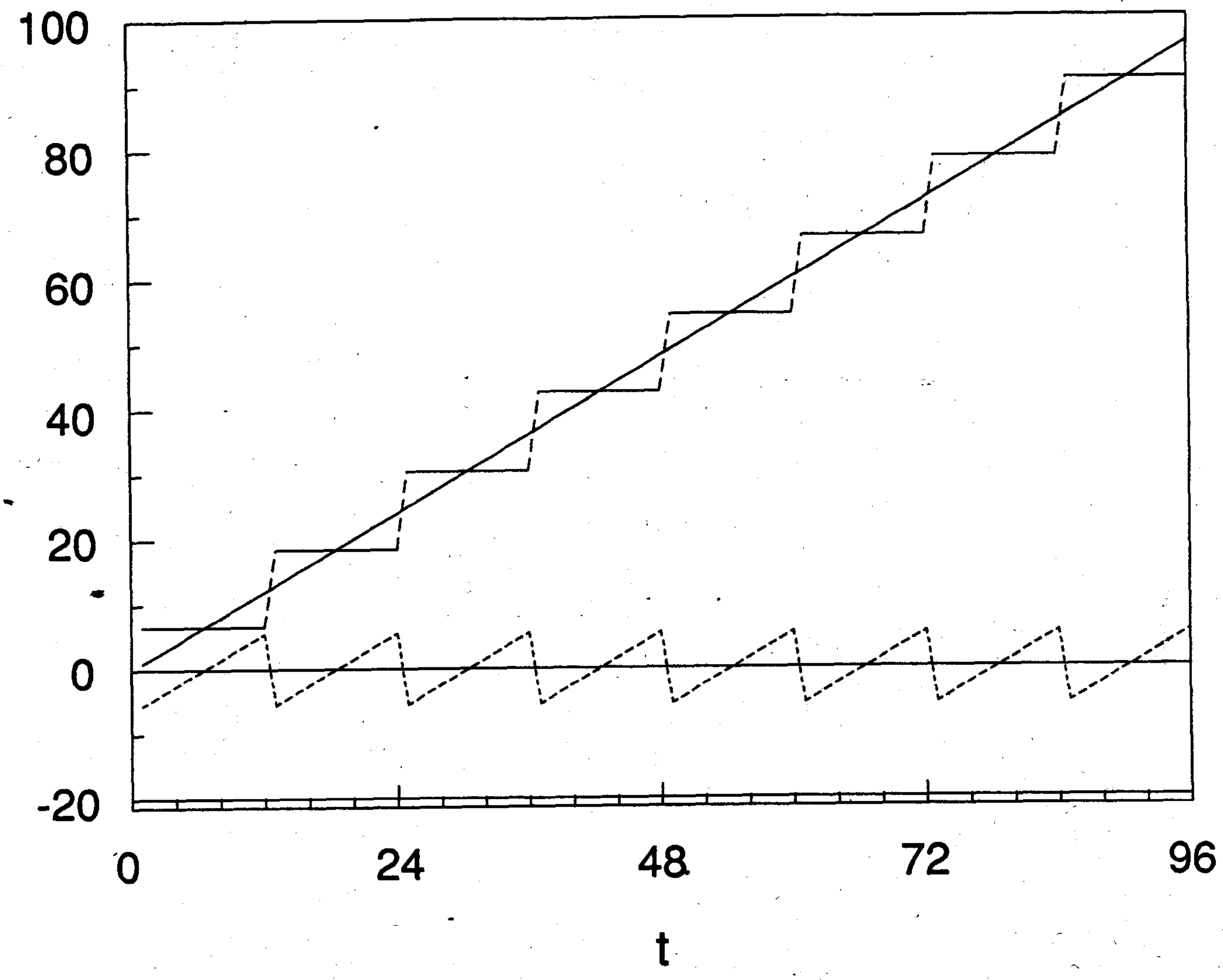
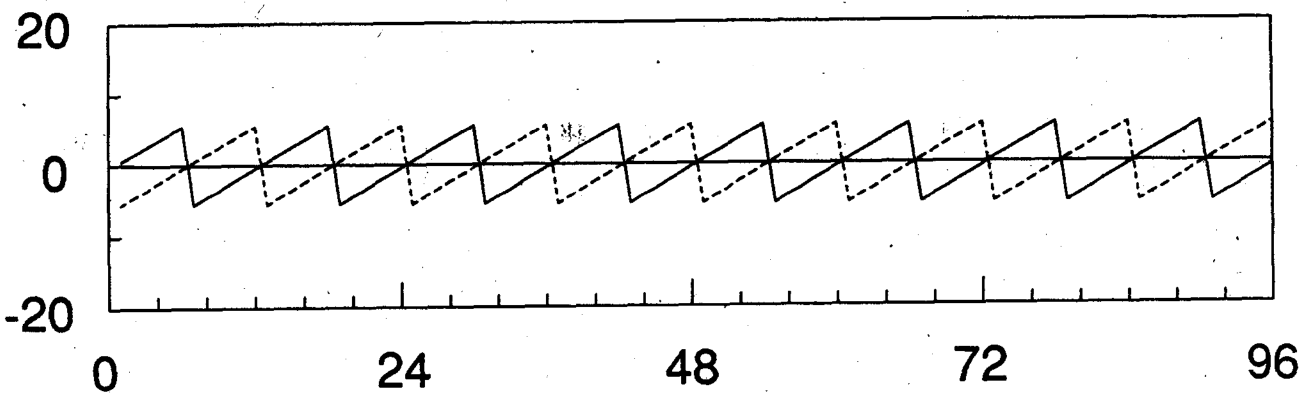


Fig. 2. Two possible seasonals for a time trend

— series starts in Jan.   --- series starts in July



starting month of the series changes (e.g., if we drop some of the first data points). In fact, a related point made in Bell (1984) is that the general decomposition approach given in Section 1 should not compute the seasonal and overall means defined only over the finite time frame of a given data set primarily because such decompositions would generally change if the time frame of the data set changed. Thus, the decomposition of $f_t = t$ in Bell (1984) violates a generalization of this principle, which can be stated as follows:

Principle. The seasonal component of a deterministic effect f_t should be the same over calendar years regardless of the starting or ending date of the time series.

Except for the time trend arising in stock length–of–month effects, the other decompositions of deterministic effects in Bell (1984) satisfy this principle. This is reviewed in the next section for stock trading–day effects.

I now feel that the preferred treatment of $f_t = t$ in seasonal adjustment is to assign it entirely to the nonseasonal as part of the trend. Thus, it has no seasonal component, and $f_t = t$ will not differentially affect the seasonal component of the series depending on the starting month. The function $f_t = t$ will differentially affect the trend level depending on the starting month, but this is not of concern since the series will presumably contain other level effects that will confound this anyway.

3. Seasonal Decomposition of Trading–Day Effects in Stock Series — New Approach

From Bell (1984) the trading–day regression effect for a stock series can be written

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

where X_{ij} is the number of times day i occurs in month j, $I_t(i)$ is 1 if month t ends on an i day and is 0 otherwise, and m_j is the length of month j. The β_i's and γ_i's are parameters with $\bar{\beta} = (\beta_1 + \cdots + \beta_7)/7$ and $\bar{\gamma} = (\gamma_1 + \cdots + \gamma_7)/7$. The first term on the right hand side of (1) can be written as

$$\sum_{i=1}^{7} (\gamma_i - \bar{\gamma}) I_t(i) = \sum_{i=1}^{7} (\gamma_i - \bar{\gamma})[I_t(i) - I_t(7)] + I_t(7) \sum_{i=1}^{7} (\gamma_i - \bar{\gamma})$$

$$= \sum_{i=1}^{6} \tilde{\gamma}_i D_{it} \tag{2}$$

where $\tilde{\gamma}_i = \gamma_i - \bar{\gamma}$ and $D_{it} = I_t(i) - I_t(7)$. Since the long–term overall means of the D_{it}'s are all zero, this contains no seasonality or trend — it is a pure stock trading–day effect. The second term on the right hand side of (1), $\bar{\gamma}$, is part of the trend. Bell (1984, p. 7) shows that $\sum_{j=1}^{t} m_j$ can be written as follows:

$$\sum_{j=1}^{t} m_j = \sum_{j=1}^{t} \xi_j + \sum_{j=1}^{t} LF_j + (30.4375)t \tag{3}$$

where $\xi_j \equiv \xi_{k(j)}$ is the seasonal component of m_j (.5625 in 31–day months, –.4375 in 30–day months, and –2.1875 in February), and LF_j is the leap February variable (–.25 in a non–leap February, .75 in a leap February, and zero otherwise). Bell (1984) broke the last term in (3) into seasonal and nonseasonal parts. We now see that the effect of this term, $\bar{\beta}(30.4375)t$, should merely be assigned to the trend.

For the first term on the right hand side of (3), Bell (1984, p. 8) notes that $\sum_{j=1}^{t} \xi_j$ is a series of monthly means since any 12 consecutive ξ_j's sum to zero, and thus

$$\sum_{j=1}^{t} \xi_j = [\sum_{j=1}^{t} \xi_j - \bar{\xi}_*] + \bar{\xi}_* \qquad (4)$$
$$= \text{seasonal} + \text{level}$$

where $\bar{\xi}_*$ is the long–term overall mean, $\bar{\xi}_* = (1/12) \sum_{t=1}^{12} \sum_{j=1}^{t} \xi_j = (1/12)[12\xi_{k(1)} + 11\xi_{k(2)} + \cdots + \xi_{k(12)}]$, which depends on the month k(1) that the series starts in. This raises a question as to whether this decomposition of $\sum_{j=1}^{t} \xi_j$ satisfies the invariance principle put forward in Section 2. However, computation of $\sum_{j=1}^{t} \xi_j - \bar{\xi}_*$ for t = 1, ... , 12 starting from each possible month k(1) = Jan., ... , k(1) = Dec. yields the same seasonal pattern over calendar months. This pattern is as follows.

$$\sum_{j=1}^{t} \xi_j - \bar{\xi}_*$$

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1.2604	–.9271	–.3646	–.8021	–.2396	–.6771	–.1146	.4479	.0104	.5729	.1354	.6979

The sum of these 12 numbers is 0 (within round–off error). Thus, this is the seasonal part of $\sum_{j=1}^{t} \xi_j$ regardless of the starting month. The level part of $\sum_{j=1}^{t} \xi_j$, $\bar{\xi}_*$, depends on the starting month k(1). But this is not of concern, since the level will be absorbed into the overall mean (if the series is not differenced), or annihilated by the differencing.

A similar analysis can be performed for the second term on the right hand side of (3), $\sum_{j=1}^{t} LF_j$. Note $\sum_{j=1}^{t} LF_j$ has period 48 since the sum of LF_j over 4 years is zero. Bell (1984, p. 9) notes that $\sum_{j=1}^{t} LF_j$ is orthogonal to a pure seasonal effect in the long run (actually, over each 4 year period). Thus, $\sum_{j=1}^{t} LF_j$ consists of a pure leap–year effect, plus a level effect δ_* which is constant over time but depends on whether the first, second, third, or fourth February in the series is a leap–year February. The level values δ_* are .375, .125, –.125, and –.375, depending on whether the first leap–year February occurs in the first, second, third, or fourth year of the series (Bell 1984, p. 9). We can thus write

$$\sum_{j=1}^{t} LF_j = [\sum_{j=1}^{t} LF_j - \delta_*] + \delta_* \tag{5}$$

$$= \text{leap–year effect} + \text{level}.$$

Computation of $\sum_{j=1}^{t} LF_j - \delta_*$ shows that it follows a 48 month pattern that does not depend on when the first leap–year February occurs. This pattern is as follows (Feb* = leap–year February, Feb = non–leap–year February).

$$\sum_{j=1}^{t} LF_j - \delta_*$$

	Jan	Feb*	Mar		Jan	Feb	Mar		Jan	Feb	Mar
...	–.375	.375	.375	...	.375	.125	.125	...	.125	–.125	–.125

	Jan	Feb	Mar		Jan	Feb*	Mar	
...	–.125	–.375	–.375	...	–.375	.375	.375	...

The long–term monthly means of this sequence are all zero, so it is a pure leap–year effect.

Using (2) – (5), the stock trading–day regression effect (1) can be decomposed as follows:

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \beta_i X_{ij} = \sum_{i=1}^{6} \tilde{\gamma}_i D_{it} + \bar{\gamma} + \bar{\beta}[\sum_{j=1}^{t} \xi_j + \sum_{j=1}^{t} LF_j + (30.4375)t]$$

$$= \sum_{i=1}^{6} \tilde{\gamma}_i D_{it} + \bar{\gamma} + \bar{\beta}(30.4375)t + \bar{\beta}[(\sum_{j=1}^{t} \xi_j - \bar{\xi}_*) + \bar{\xi}_* + (\sum_{j=1}^{t} LF_j - \delta_*) + \delta_*]$$

$$
\begin{aligned}
&= \sum_{i=1}^{6} \tilde{\gamma}_i D_{it} && \text{(pure stock trading–day)} \\
&+ \bar{\beta}\Big(\sum_{j=1}^{t} \xi_j - \bar{\xi}_*\Big) && \text{(pure seasonal)} \\
&+ \bar{\beta}\Big(\sum_{j=1}^{t} LF_j - \delta_*\Big) && \text{(pure leap–year)} \\
&+ \bar{\gamma} + \bar{\beta}(\bar{\xi}_* + \delta_*) + \bar{\beta}(30.4375)t && \text{(level and trend).}
\end{aligned}
\tag{6}
$$

Only $\bar{\xi}_*$ and δ_* in the last line depend on the month the series starts in. But since these will be confounded with the overall level, there is no need to break them out separately. Thus, if $\sum_{j=1}^{t} \sum_{i=1}^{7} \beta_i X_{ij} = \sum_{i=1}^{7} \beta_i \sum_{j=1}^{t} X_{ij}$ has been estimated, it can be broken down into the first three lines in (6), and $\bar{\beta}(30.4375)t$ can be added to the level estimate for the series with these regression effects removed (however the level estimate is obtained).

When estimating stock trading–day models, we need to include the variables $D_{1t}, \ldots, D_{6t}$. If we also want to allow for the cumulative length–of–month effect $(\beta \sum_{j=1}^{t} m_j)$, we have several choices about which variable to include:

(i) $\sum_{j=1}^{t} m_j$, in which case, after estimation, $\bar{\beta} \sum_{j=1}^{t} m_j$ should be decomposed as in (6);

(ii) $(\sum_{j=1}^{t} LF_j - \delta_*) + (30.4375)t$, if the model includes fixed seasonal effects or seasonal differencing;

(iii) $\sum_{j=1}^{t} LF_j - \delta_*$, if the model includes two differences (to wipe out t) one of which is seasonal (to wipe out fixed seasonal effects), or only a seasonal difference but with a trend constant, or only a nonseasonal difference but with a trend constant and fixed seasonal effects, or no differences but with fixed seasonal effects and a linear time trend.

When the conditions are met for (iii), using it gives a direct estimate of $\beta(\sum_{j=1}^{t} LF_j - \delta_*)$, the pure stock leap–year effect. This may have an advantage of convenience. Using (iii) all the time, though, forces us to meet the conditions of (iii) by either including sufficient differencing in the model, or by including fixed seasonal effects and/or a time trend (or trend constant) in the model. Similarly, always using (ii) forces us to meet its condition by including fixed seasonal effects or seasonal differencing in the model. Thus, (1) is more general, but it requires the decomposition in (6) for interpretation. Another alternative would be to use the variable

$$\text{(iv)} \quad (\sum_{j=1}^{t} \xi_j - \bar{\xi}_*) + (\sum_{j=1}^{t} LF_j - \delta_*) = \sum_{j=1}^{t} m_j - (\bar{\xi}_* + \delta_*) - (30.4375)t \,.$$

This is like (iii), and requires the same conditions, except that it does not separate out the seasonal part of the cumulated length–of–month variable. Thus, β times the above can be regarded as the stock length–of–month effect.

<u>Final note</u>: This presentation assumes, as in Bell (1984), that length–of–month effects in the flow series which aggregates to the stock series of interest are to be modelled using a length–of–month regression variable (m_t, possibly with its fixed seasonal and leap–year effects removed). There are two reasons why this may not be the case. First, if the flow series would be transformed before modelling (or if the flow series was the difference of two flow series that would be transformed), then this condition is violated since length–of–month effects in the transformed flow scale would not aggregate to anything simple in the stock series. Second, research done since my earlier work (Bell 1992, Findley and Chen 1994) suggests that length–of–month effects in flow series are better handled by dividing the series by length–of–month, i.e., if y_t is the original (untransformed) series

being modelled we take y_t/m_t. This at least seems preferable for series y_t that are transformed by taking logs or otherwise. If length–of–month effects in the flow series are handled by division, there is again no simple expression for the length–of–month effect in the corresponding stock series. Therefore, the decomposition discussed here is strictly relevant only for the case where the flow series is best modelled without transformation (additive seasonal decomposition of the flow series). In other cases use of $\sum_{j=1}^{t} m_j$ as a regression variable is not strictly appropriate, though it may sometimes be used as an approximation. However, a better alternative in such cases is probably to simply ignore possible length–of–month effects in a stock series. If one thinks of directly addressing the modelling of calendar effects in a stock series, there is no obvious reason why the series should depend on some function of the length–of–month time series m_t.

REFERENCES

Bell, William R. (1984) "Seasonal Decomposition of Deterministic Effects," Research Report 84/01, Statistical Research Division, Bureau of the Census.

Bell, William R. (1992) "Alternative Approaches to Length of Month Adjustment," Research Report 92/17, Statistical Research Division, Bureau of the Census.

Findley, David F. and Chen, Bor–Chung (1994) "Trading–Day Effect Models for Flow Series and Stock Series," paper in progress, Statistical Research Division, Bureau of the Census.