

RESEARCH REPORT SERIES
(Statistics #1992-17)

**Alternative Approaches to Length
of Month Adjustment**

William R. Bell

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Report Issued: May 7, 1992

Disclaimer: This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233

**BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES**

TECHNICAL NOTE SERIES, No. TN-92/01

**ALTERNATIVE APPROACHES TO LENGTH
OF MONTH ADJUSTMENT**

by

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

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Report issued: May 7, 1992

ALTERNATIVE APPROACHES TO LENGTH OF MONTH ADJUSTMENT

**William R. Bell
Statistical Research Division
U. S. Bureau of the Census**

May 7, 1992

ALTERNATIVE APPROACHES TO LENGTH OF MONTH ADJUSTMENT

William R. Bell
Statistical Research Division
U. S. Bureau of the Census

ABSTRACT

We consider two approaches to adjustment for length of month variation in nonnegative flow time series observed monthly. One approach is to divide the observed series value in each month by the length of that month and then multiply all series values by the average length of month (30.4375). The other approach is to include length of month as an explanatory variable in a regression model with ARIMA time series errors (REGARIMA model), and then estimate and remove the length of month effect. For additive models we observe that the two approaches will be different, and that arguments can be made for either approach so that the choice between them may be a matter of personal preference. For multiplicative models (additive models for the logged series), we observe that the two approaches are approximately equivalent if and only if the estimated length of month coefficient is approximately .035. Since this is also the value that would be expected for the length of month coefficient in a model for the logged series, we argue that in multiplicative models one should adjust for length of month by division and then rescaling, rather than by using an estimated term from a REGARIMA model.

Key Words: calendar effects, REGARIMA model, time series

Let Y_t be a nonnegative flow time series observed monthly. Then Y_t could be expected to vary according to the length of month t , which we denote as m_t . Y_t may also be subject to trading day and other variations (e.g. holiday variation). Two general approaches to accounting and adjusting for length of month variation are (i) division by m_t , yielding the daily rate Y_t/m_t , (which is then typically rescaled to the average month length of 30.4375 days) and (ii) inclusion of m_t as a regression variable in a REGARIMA model, and subsequent removal of the estimated effect. In the following, we compare these two approaches and examine when their results would be approximately the same.

1. Additive Model

A general additive REGARIMA model is

$$Y_t = \beta_0 m_t + \sum_{i=1}^6 \beta_i T_{it} + \underline{x}'_t \underline{\alpha} + Z_t \quad (1.1)$$

where T_{1t} through T_{6t} are trading-day variables as in Bell and Hillmer (1983), the \underline{x}_t vector contains other regression variables and $\underline{\alpha}$ the corresponding parameters (e.g. for holiday or outlier effects), and Z_t has mean zero and follows some ARIMA model. In some cases, the $\underline{x}'_t \underline{\alpha}$ term will be absent from (1.1). Also, there may be no trading-day variation apart from length of month, in which case $\beta_1 = \dots = \beta_6 = 0$ or we drop the $\sum_{i=1}^6 \beta_i T_{it}$ term from (1.1).

If we divide (1.1) by m_t we get

$$Y_t/m_t = \beta_0 + \sum_{i=1}^6 \beta_i (T_{it}/m_t) + (\underline{x}'_t/m_t) \underline{\alpha} + Z_t/m_t. \quad (1.2)$$

If the trading-day and $\underline{x}'_t \underline{\alpha}$ terms are absent from (1.1), then β_0 is clearly the average daily rate. This interpretation still holds with other regression variables in (1.1) as long as the

long-run averages of these variables divided by m_t are zero, so that these new variables are asymptotically orthogonal to the constant term. The long-run averages of the T_{it}/m_t variables, which are the proportion of days in month t that are Mondays minus the proportion that are Sundays, etc., are in fact zero.

Comparing (1.1) and (1.2), we see two basic differences. The first is the change in the regression variables from (1.1) to (1.2). One concern about (1.2) may be interpretation. The interpretation of the T_{it}/m_t is straightforward, but some x'_t/m_t variables may not have a natural interpretation. (One regression variable for which interpretation after dividing by m_t is not a problem is the additive outlier indicator variable, $AO_t^{(t_0)}$, which is 1 for $t = t_0$ and is 0 otherwise, since $AO_t^{(t_0)}/m_t$ is just a rescaling of $AO_t^{(t_0)}$.) Also, notice that if we difference (1.2), we will annihilate β_0 and hence will not be able to explicitly estimate it, nor will we need to. If we difference (1.1), we do not annihilate $\beta_0 m_t$ and still need to explicitly estimate β_0 . (There is no natural value for β_0 in the additive model; this will vary from series to series.) Notice if the model for Z_t/m_t involves differencing, then (1.2) accounts for length of month effects with one less parameter than does (1.1).

The second basic difference between (1.1) and (1.2) is the ARIMA modeling of Z_t in (1.1) versus the ARIMA modeling of Z_t/m_t in (1.2). Let $Z_t = \sum_{i=1}^{m_t} Z_{it}$ where Z_{it} is the daily flow series with means removed. If Z_{it} is i.i.d. then $\text{Var}(Z_t) = m_t \text{Var}(Z_{it})$ and $\text{Var}(Z_t/m_t) = m_t^{-1} \text{Var}(Z_{it})$. In more general settings, it may also be reasonable to assume autocovariances of Z_t are proportional to m_t , and so those of Z_t/m_t are proportional to m_t^{-1} . In both cases there is a time dependence in variances and autocovariances which could be accounted for by modeling $Z_t/(m_t)^{1/2}$ instead. Since the variation in m_t is small, this may not make much difference. Also, except for leap year Februaries, the variation in m_t is perfectly seasonal, and so could be confounded with seasonal fluctuations in variance that might arise from other sources. For the remainder of this note, we shall assume that the

effects of dividing by m_t or not on the autocovariance structure of the series are negligibly different from a constant rescaling.

Though (1.1) and (1.2) should be at least approximately consistent with each other as models for Y_t , they suggest different length of month adjustments. (1.1) naturally suggests the additive length of month adjustment, $Y_t - \beta_0(m_t - \bar{m})$, where $\bar{m} = 30.4375$ is the average month length. (1.2) naturally suggests taking $\bar{m}Y_t/m_t$ as the length of month adjusted series. (The rescaling by \bar{m} is to maintain the average level of the series. We take up this adjustment again in the next section. Note we can use (1.2) multiplied by \bar{m} as a model for the adjusted series; the relevant statistical properties are unchanged by the rescaling by \bar{m} .) One could argue that the additive adjustment, $Y_t - \beta_0(m_t - \bar{m})$, is the more natural when using models with additive effects as in (1.1) and (1.2). On the other hand, this adjustment requires estimation of the parameter β_0 , whereas taking $\bar{m}Y_t/m_t$ does not. In the end, the choice between these two approaches to length of month adjustment might depend on personal preference.

2. Multiplicative (log-additive) Model

A multiplicative version of (1.1) is

$$Y_t = \exp[\beta_0 m_t + \sum_{i=1}^6 \beta_i T_{it} + \underline{x}'_t \alpha] Z_t. \quad (2.1)$$

On taking logarithms of this we have

$$\log(Y_t) = \beta_0 m_t + \sum_{i=1}^6 \beta_i T_{it} + \underline{x}'_t \alpha + \log(Z_t). \quad (2.2)$$

Dividing (2.2) by m_t gives

$$\log(Y_t)/m_t = \beta_0 + \sum_{i=1}^6 \beta_i (T_{it}/m_t) + (\underline{x}'_t/m_t)\underline{\alpha} + \log(Z_t)/m_t \quad (2.3)$$

Notice (2.2) and (2.3) are analogous to (1.1) and (1.2), so our previous remarks about (1.1) and (1.2) apply to (2.2) and (2.3). The remark about the effect on autocovariance structure of dividing by m_t applies if Z_{it} is a multiplicative daily effect so $Z_t = Z_{1t} \times \dots \times Z_{m_t, t}$ and $\log(Z_t) = \sum_{i=1}^{m_t} \log(Z_{it})$. We shall again assume dividing by m_t has effects on autocovariance structure that are negligibly different from a constant rescaling.

The individual regression terms in (2.1) can be converted into multiplicative form. For length of month, this is $e^{\beta_0 m_t} = [e^{\beta_0}]^{m_t}$. If β_0 is not large (and it shouldn't be, for reasons seen subsequently) then $e^{\beta_0} \approx 1 + \beta_0$ and $e^{\beta_0 m_t} \approx (1 + \beta_0)^{m_t}$. We see $(1 + \beta_0)^{m_t}$ functions as a compounding factor, with $1 + \beta_0$ giving an "average" proportionate increase in the series each day. Thus, $100\beta_0$ is the average percentage increase each day.

An alternative to (2.1) – (2.3) is to start with regression effects as in (1.1) for $\bar{m}Y_t/m_t$, but in a multiplicative structure, i.e. .

$$\bar{m}Y_t/m_t = \exp[\beta_0 + \sum_{i=1}^6 \beta_i T_{it} + \underline{x}'_t \underline{\alpha}] Z_t \quad (2.4)$$

Here $\bar{m} = 30.4375$, the average month length, is incorporated as a rescaling factor to avoid changing the overall level of the series. Taking logarithms in (2.4) gives

$$\log(\bar{m}Y_t/m_t) = \beta_0 + \sum_{i=1}^6 \beta_i T_{it} + \underline{x}'_t \underline{\alpha} + \log(Z_t) . \quad (2.5)$$

We now examine under what circumstances (2.4) – (2.5) may be regarded as approximately consistent with (2.1) – (2.3). We examine this question two ways.

2.1 Approximation 1

Notice that

$$\begin{aligned}
 \log(\bar{m}Y_t/m_t) &= \log(Y_t) + \log(\bar{m}/m_t) \\
 &= \log(Y_t) - \log(m_t/\bar{m}) \\
 &= \log(Y_t) - \log[1 + (m_t - \bar{m})/\bar{m}] \\
 &\approx \log(Y_t) - (m_t - \bar{m})/\bar{m} \\
 &= \log(Y_t) - m_t/\bar{m} + 1
 \end{aligned} \tag{2.6}$$

using a one term Taylor series to approximate $\log[1 + (m_t - \bar{m})/\bar{m}]$. The accuracy of this approximation can be seen in the following table.

m_t	28	29	30	31
$\log(\bar{m}/m_t)$.0835	.0484	.0145	-.0183
$1 - m_t/\bar{m}$.0801	.0472	.0144	-.0185

The approximation is excellent for months other than February, and appears adequate for most purposes in February as well.

The additive constant 1 in (2.6) is immaterial for modeling purposes (but not for scaling) since it will be annihilated by differencing. Apart from this, (2.6) shows taking $\log(\bar{m}Y_t/m_t)$ is approximately the same as taking $\log(Y_t) + 1 - \beta_0 m_t$ as implied by (2.2) if and only if β_0 is approximately $1/\bar{m} = .0329$. Notice that, to avoid level differences, we take $\log(Y_t) + 1 - .0329m_t$, and not just $\log(Y_t) - .0329m_t$, because the long term average of $1 - .0329m_t$ is zero. In the original scale we take $Y_t \exp[1 - .0329m_t]$ and the average effect on the level is the long-term average of $\exp[1 - m_t/\bar{m}]$, which is about 1.0004, or approximately 1. Hence, the level in the original scale is not altered, on average, by taking $Y_t \exp[1 - m_t/\bar{m}]$. If instead we wish to estimate β_0 , and then use $\hat{\beta}_0$ to adjust for length of

month, we should take $\log(Y_t) + 1 - \hat{\beta}_0 m_t$ (not just $\log(Y_t) - \hat{\beta}_0 m_t$) and, in the original scale, $Y_t \exp[1 - \hat{\beta}_0 m_t]$, to keep the scaling consistent with that of $\bar{m} Y_t / m_t$.

2.2 Approximation 2

Suppose our model involves seasonal differencing. Taking Δ_{12} of (2.5) annihilates β_0 , leaving no trace there of the length of month effect. Taking Δ_{12} of (2.2) suggests taking $\Delta_{12} \log(Y_t) - \beta_0 \Delta_{12} m_t$ to remove length of month effects from the seasonally differenced data. Notice $\Delta_{12} m_t$ is nonzero only in leap year Februaries and the following Februaries, and $\Delta_{12} m_t = \Delta_{12} LF_t$ where LF_t is the "leap February variable" used in REGARIMA.

Now notice that

$$\begin{aligned}
 \Delta_{12} \log(\bar{m} Y_t / m_t) &= \Delta_{12} \log(Y_t) + \log(m_{t-12} / m_t) + \Delta_{12} \log(\bar{m}) \\
 &= \Delta_{12} \log(Y_t) + \log[1 + (m_{t-12} - m_t) / m_t] + 0 \\
 &\approx \Delta_{12} \log(Y_t) + (m_{t-12} - m_t) / m_t \\
 &= \Delta_{12} \log(Y_t) + \begin{cases} -1/29 & t - \text{leap Feb.} \\ 1/28 & t-12 - \text{leap Feb.} \\ 0 & \text{otherwise} \end{cases} \\
 &\approx \Delta_{12} \log(Y_t) - .035 \Delta_{12} m_t
 \end{aligned} \tag{2.7}$$

where a one term Taylor series is used to approximate $\log[1 + (m_{t-12} - m_t) / m_t]$ and we use $1/29 \approx .03448 \approx .035$ and $1/28 \approx .03571 \approx .035$. The Taylor series approximation is exact in all months other than February (since then $\Delta_{12} m_t = 0$) and gives $-1/29$ and $1/28$ instead of $\log(28/29) = -.0351$ and $\log(29/28) = .0351$ in leap year and the following Februaries, respectively. We see (2.7) implies approximately the same length of month adjustment as (2.2) if and only if $\beta_0 = .035$. This is a very slightly different value from .0329, the required value of β_0 for the approximation in section 2.1.

The approach here could also be used with a nonseasonal difference Δ in place of Δ_{12} , but the resulting Taylor series approximation is less accurate than that of section 2.1.

2.3 What value would we expect for β_0 ?

Consider (2.1). If there is no trading day variation present, and no other regression effects, then $Y_t \approx (1 + \beta_0)^{m_t} Z_t$, since $\exp(\beta_0) \approx 1 + \beta_0$ if β_0 is not large. If t corresponds to a leap-year February then

$$Y_t/Y_{t-12} \approx (1 + \beta_0)Z_t/Z_{t-12}$$

If all other things are equal and $\log(Z_t/Z_{t-12})$ is stationary, we would expect Z_t/Z_{t-12} to be around 1 "on average", and would expect Y_t to proportionally exceed Y_{t-12} according to the lengths of months t and $t-12$; that is, we would expect $1 + \beta_0 \approx m_t/m_{t-12} = 29/28$. This implies $\beta_0 = 1/28 \approx .035$. If t were the February after a leap year, then we get $(1 + \beta_0)^{-1} \approx m_t/m_{t-12} = 28/29$ also implying $\beta_0 = 1/28$. These relationships also make sense in more general models: if other regression effects are present, they should be explained with other regression variables and so do not affect the interpretation of $1 + \beta_0$, and if there are stochastic nonstationarities present, these should be part of both Z_t and Y_t ($1 + \beta_0$ still reflects the relative magnitude as above). Year-to-year comparisons of other months yield no information about β_0 , and comparing different months with different lengths is more complex since there would be a confounding with seasonal effects.

3. Conclusions

Length of month effects in a time series Y_t can be modeled and adjusted for using a term $\beta_0 m_t$ in a REGARIMA model. With an additive model (a REGARIMA model for Y_t) we can subtract $\hat{\beta}_0(m_t - \bar{m})$ from Y_t to adjust for length of month, where $\hat{\beta}_0$ is the estimate of β_0 . With a multiplicative model (a REGARIMA model for $\log(Y_t)$) we can take

$\log(Y_t) + 1 - \hat{\beta}_0 m_t$, or $Y_t \exp[1 - \hat{\beta}_0 m_t]$ on the original scale, to adjust for length of month. Alternatively, we can adjust for length of month effects by a simple division by m_t and rescaling, i.e. taking $\bar{m} Y_t / m_t$. In this note we have compared and contrasted these two approaches.

In the additive model the two approaches to length of month adjustment will lead to different results, although the resulting models, (1.1) and (1.2) multiplied by \bar{m} , should be approximately consistent with each other. Arguments can be made for both approaches; the choice between the two may be a matter of personal preference.

In the multiplicative model, on the other hand, taking $\bar{m} Y_t / m_t$ has approximately the same effect as taking $Y_t \exp[1 - \hat{\beta}_0 m_t]$ if and only if $\hat{\beta}_0 \approx 1/\bar{m} \approx .035$. This is also the approximate value we would expect for β_0 in the multiplicative model anyway. An argument can be made for preferring division by length of month in this case on the grounds that if modeling length of month effects with a $\beta_0 m_t$ term yields very different results, then believing these results places great faith in the model that is probably unwarranted. Notice that estimates of β_0 in (2.2) or (2.3) can deviate from .035 due to various model failures as well as to simple estimation error. Among possible model failures is omission of needed regression terms from the model, since such omissions are known to bias the estimates of the regression parameters that are in the model.

REFERENCE

- Bell, W. R. and Hillmer, S. C. (1983) "Modeling Time Series with Calendar Variation," Journal of the American Statistical Association, 78, 526-534.