RESEARCH REPORT SERIES
(Statistics #2009-04)

Stock Series Holiday Regressors Generated By Flow Series Holiday Regressors

David F. Findley

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: April 30, 2009
Revised and Reissued: September 30, 2009

Stock Series Holiday Regressors Generated By Flow Series Holiday Regressors

David F. Findley
U.S. Bureau of the Census

Abstract

Stock economic time series such as end-of-month inventories arise from a succession of monthly inflows and outflows, i.e., from an accumulation of monthly net flows. We use the perspective of stocks as accumulations of monthly flows to derive holiday regressors for end-of-month stock series from cumulative sums of flow series holiday regressors. Our focus is on common flow holiday regressor properties that yield simple formulas for the stock regressors. The resulting regressors seem to be the first holiday regressors for stock series. Stock Easter holiday effect regressors obtained in this way have been implemented in X-13-ARIMA-SEATS. Empirical results are shown from their application to U.S. manufacturing inventory series from the M3 Survey.

Key Words: Stock time series; Inventory series; Seasonal adjustment; Calendar effects; Moving holidays; Forecasting, X-13-ARIMA-SEATS.

1 Overview

Stock economic time series, for example end-of-month inventories, arise as the cumulative sum of monthly inflows and outflows–that is, from the accumulation of monthly net flows. Following an approach similar to that used by Bell (1984, 1995) to obtain effective stock series trading day regressors (see also Findley and Monsell, 2009), we regard end-of-month stock series as accumulations of monthly flows and derive holiday regressors for such stock series from cumulative sums of basic flow series holiday regressors. These resulting regressors have a simple form and seem to be the first general moving holiday effect regressors for stock series. For further discussion of stocks and flows, see Wikipedia Contributors (2009).

Section 2 describes properties common to standard flow holiday regressors that are shown, in Section 3 to generate stock regressors that have a simple form along with two properties desirable for seasonal and calendar-effect adjustment: they are level-neutral and free of seasonality. Section 4 specializes their formulas to derive the stock Easter effect regressors generated by the widely used monthly-proportional-effect regressors associated with time intervals preceding Easter. These new Easter effect regressors are used to identify effects of the dates of Easter in some of the U.S. manufacturing inventory series that were identified in Findley and Monsell (2009) as having stock trading day effects.

Then, to provide a broader perspective, in Section 5 regressor examples are presented to illustrate consequences of deriving stock regressors from flow regressors that fail to have all three properties required in Section 3. The concluding Section 6 has two main parts. The first describes an indirect strategy for testing for the presence of flow-induced holiday effects in stock series using only the flow series regressors. The second discusses the situation in which the approach presented for deriving regressors for stock series is applied to log transformed stock series in order to estimate holiday effects in the untransformed series. This is the situation of the empirical study of Section 4.

2 Properties of Flow Regressors Utilized

We start from a monthly flow series regressor $H(t)$ for a holiday whose dates change from year to year. The dates are assumed to repeat every P years (to a good approximation) from some initial year forward in the (Gregorian) calendar. This initial year must not be later than the earliest year of the monthly stock time series for which the holiday's effect is to be estimated. We define the time index t so that

$$t = j + 12(M - 1) \tag{1}$$

for the j-th calendar month of the M-th year under consideration, $1 \leq j \leq 12$, where $M = 1$ denotes the first year of the time interval of the series for which the holiday effect is to be estimated. Preceding years for which the regressor is defined are indexed by $M = 0, -1, \ldots$ and later years by $M > 1$. The regressor $H(t)$ is required to be periodic with period $12P$ for some positive integer P: for all t,

$$H(t + 12P) = H(t). \tag{2}$$

We also require $H(t)$ to have constant annual sums,

$$\sum_{j=1}^{12} H(j + 12(M-1)) = K, \tag{3}$$

with K independent of M.

For example, the $H(t)$ that underlie the principle U.S. moving holiday flow series models of the `regression` spec of X-12-ARIMA are each specified by an interval of length $w \geq 1$ days connected to the date of the holiday. The regressor treats the holiday's effect as occurring in this interval. More precisely, for each $1 \leq j \leq 12$, let $n_j(M) = n_j(M, w)$ denote the number of days of month j that fall in this interval in year M. Then for $t = j + 12(M-1)$, $H(t)$ has the value

$$H(j + 12(M-1)) = \frac{n_j(M)}{w}, \qquad 1 \leq j \leq 12. \tag{4}$$

For each holiday (Easter, Labor Day, Thanksgiving), the values of w considered are such that M-th year's holiday-effects interval lies within year M. Therefore $\sum_{j=1}^{12} n_j(M) = w$, resulting in (3) with $K = 1$ for all M.

In the context of seasonal and holiday effect adjustment, it is desirable to have holiday-effect estimates whose removal from the series does not change the overall level of the series. Also, they should not remove any seasonal effects. This latter insures that all estimated seasonal effects will be described by the seasonal factors produced by the seasonal adjustment procedure, which is applied after adjustment for holiday effects. To achieve a regressor that provides estimates with these two properties, the starting regressor $H(t)$ must usually be modified to obtain a regressor with no long-term level or seasonal components. The twelve individual P year calendar month averages,

$$\bar{H}_j = \frac{1}{P} \sum_{m=M}^{M+P-1} H(j + 12(m-1)), \qquad 1 \leq j \leq 12, \tag{5}$$

whose values do not depend on M because of (2), identify the level plus seasonal component of $H(t)$: For $t = j + 12(M-1)$, the level and seasonally adjusted regressor is given by

$$\tilde{H}(t) = H(t) - \bar{H}_j, \qquad 1 \leq j \leq 12, \tag{6}$$

see Bell (1984) and Findley and Soukup (2000)[1]. The regressor $\tilde{H}(t)$ satisfies

$$\tilde{H}(t + 12P) = \tilde{H}(t), \tag{7}$$

and, from (5), its level plus seasonal component is zero:

$$\frac{1}{P} \sum_{m=M}^{M+P-1} \tilde{H}(j + 12(m-1)) = 0, \qquad 1 \leq j \leq 12. \tag{8}$$

Also, it follows from (3) and (5) that $\sum_{j=1}^{12} \bar{H}_j = (1/P)\, PK = K$. Therefore the annual sums $\sum_{j=1}^{12} \tilde{H}(j + 12(M-1)) = \sum_{j=1}^{12} H(j + 12(M-1)) - \sum_{j=1}^{12} \bar{H}_j$ are zero:

$$\sum_{j=1}^{12} \tilde{H}(j + 12(M-1)) = 0, \tag{9}$$

for each M, a useful property as will be seen.

3 End-of-Month Stock Series Regressors

Following Cleveland and Grupe (1983) and Bell (1984, 1995), we view an end-of-month stock series as an accumulation of consecutive monthly flow series values and derive the regressor for the stock series by accumulating the values of an appropriate flow series regressor. Simple formulas and desirable properties will obtained for flow series regressors $\tilde{H}(t)$ with the properties (7), (8) and (9).

[1]The period 12 function $\eta(t)$ defined by $\eta(j + 12(M-1)) = \bar{H}_j$ for $1 \leq j \leq 12$ and $M \geq M_0$ is the sum of the level component

$$\bar{H} = \frac{1}{12P} \sum_{t=1}^{12P} H(t) = \frac{1}{12} \sum_{j=1}^{12} \bar{H}_j$$

and the seasonal component $\eta(t) - \bar{H}$. The latter is the sum of the six seasonal subperiod components (frequencies $k/12$ cycles per month, $k = 1, \ldots, 6$) of $H(t)$. As a result, the representation of $\tilde{H}(t) = H(t) - \eta(t)$ as a sum of sines and cosines contains none having these frequencies and no constant term (frequency 0 component), see Section 4.2.3 of Anderson (1971). Thus $\tilde{H}(t)$ is level-neutral and free of seasonalilty. Bell (1984) provides an alternative discussion of $\eta(t)$ and $\tilde{H}(t)$, without reference to periodicity, in which the level-neutral property is defined by $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \tilde{H}(t) = 0$. This property follows from the argument used to establish (15) below.

For any convenient $M_0 \leq 1$ and any $t = j + 12(M-1)$, with $1 \leq j \leq 12$ and $M \geq M_0$, we define the stock series regressor $\tilde{H}_S(t)$ generated by $\tilde{H}(t)$ to be

$$\tilde{H}_S(t) = \sum_{u=1+12(M_0-1)}^{t} \tilde{H}(u). \tag{10}$$

Due to (9), this sum simplifies to the sum over months in year M through time t,

$$\begin{aligned} \tilde{H}_S(t) &= \sum_{m=M_0}^{M-1} \sum_{i=1}^{12} \tilde{H}(i + 12(m-1)) + \sum_{i=1}^{j} \tilde{H}(i + 12(M-1)) \\ &= \sum_{i=1}^{j} \tilde{H}(i + 12(M-1)). \end{aligned} \tag{11}$$

($\sum_{m=M_0}^{M-1}$ is the null sum with value zero if $M-1 < M_0$.) Since $\tilde{H}(i + 12(M+P-1)) = \tilde{H}(i + 12(M-1))$ for $1 \leq i \leq j$ by (7), it follows from (11) that $\tilde{H}_S(t)$ is periodic,

$$\tilde{H}_S(t + 12P) = \tilde{H}_S(t). \tag{12}$$

Also, its level plus seasonal component is zero: Indeed, for any $1 \leq j \leq 12$,

$$\sum_{m=M}^{M+P-1} \tilde{H}_S(j + 12(m-1)) = P \sum_{i=1}^{j} \left\{ \frac{1}{P} \sum_{m=M}^{M+P-1} \tilde{H}(i + 12(m-1)) \right\} = 0, \tag{13}$$

for any $M \geq M_0$, due to (8).

However, for annual sums the analogue of (9) does not hold. From (11) and (9),

$$\begin{aligned} &\sum_{j=1}^{12} \tilde{H}_S(j + 12(M-1)) \\ &= \sum_{j=1}^{12} \sum_{i=1}^{j} \tilde{H}(i + 12(M-1)) = \sum_{i=1}^{12} \sum_{j=i}^{12} \tilde{H}(i + 12(M-1)) \\ &= \sum_{i=1}^{12} (13 - i) \tilde{H}(i + 12(M-1)) \\ &= \sum_{i=1}^{11} (12 - i) \tilde{H}(i + 12(M-1)), \end{aligned} \tag{14}$$

the last from (9). This sum can be nonzero and vary with M, see (16) below.

$\tilde{H}_S(t)$ also has the important property of being *level-neutral* in the extended sense that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\tilde{H}_S(t)=0. \tag{15}$$

In fact, the sequence $f(T)=\sum_{t=1}^{T}\tilde{H}_S(t)$, $T\geq 1$ has a stronger pair of properties. For any integers $v\geq 1$ and $1\leq u\leq 12P-1$, we have $f(12Pv)=0$ and $f(u+12Pv)=f(u)$. The first property follows from (13) and the second from $f(u+12Pv)=f(12Pv)+\sum_{t=1}^{u}\tilde{H}_S(t+12Pv)=\sum_{t=1}^{u}\tilde{H}_S(t)$, due to the first property and (12). In particular, $f(T)$ is periodic with period $12P$ and so is bounded. From this (15) follows.

Remark 3.1. If the requirements imposed above on the flow regressor $\tilde{H}(t)$ which generates the stock regressor are weakened, then the stock regressor (10) can fail to have the level-neutral property, as Section 5.1 and the example of Section 5.3 will show.

4 Stock Easter Regressors and their Application

The X-12-ARIMA holiday regressors, all of which are described in Table 4.1 of the X-12-ARIMA Reference Manual (U.S. Census Bureau, 2007), have annual sums of zero, and all but one have the level and seasonally adjusted form (6) of $\tilde{H}(t)$. The exception is the Statistics Canada Easter regressor, denoted `sceaster[w]` in Table 4.1, which will be discussed in more detail in Section 5.3. First we give detailed formulas for the stock Easter holiday regressor generated by the principal flow series Easter regressors, designated as `easter[w]` in Table 4.1.

4.1 Monthly Stock Easter Regressors from `easter[w]`

The principal flow series Easter-effect regressors $H(t)$ of X-12-ARIMA (and TRAMO-SEATS, see Gómez and Maravall, 1997, 2003) have the form (4) with interval lengths $1\leq w\leq 25$. For a given w, the interval consists of the w consecutive days up through the day before Easter. The dates of Easter vary between March 22 and April 25. Therefore, (3) holds with $K=1$ for $H(t)$. For each w, the calendar month averages $\bar{H}_1,\ldots,\bar{H}_{12}$ of (5) are given in Table 7.30[2] of U.S. Census Bureau (2007), and the resulting $\tilde{H}(t)$ is the regressor specified by the variable `easter[w]` of the `regression`

[2]The period P of the current Easter calendar is 5,700,000 years, see Montes (1998). A good approximation to the frequency distribution of the dates of Easter over this very long period can be

spec of X-12-ARIMA. Because $1 \leq w \leq 25$, the regressors $H(t)$ and $\tilde{H}(t)$ are zero except in February, March and April.

Hence, by (11), for month $t = j + 12(M-1)$, the value $\tilde{H}_S(t)$ of the stock series regressor generated by $\tilde{H}(t)$ is given by

$$\tilde{H}_S(j + 12(M-1)) = \begin{cases} 0 & , j = 1 \\ \tilde{H}(2 + 12(M-1)) & , j = 2 \\ \tilde{H}(2 + 12(M-1)) + \tilde{H}(3 + 12(M-1)) & , j = 3 \\ 0 & , 4 \leq j \leq 12 \end{cases}.$$

The final value 0 results from (9).

When $1 \leq w \leq 21$, $H(t)$, $\tilde{H}(t)$ and $\tilde{H}_S(t)$ are zero in February, so $\tilde{H}_S(t)$ is nonzero only in March and the annual sum formula (14) reduces to

$$\sum_{j=1}^{12} \tilde{H}_S(j + 12(M-1)) = \tilde{H}(3 + 12(M-1)).$$

With $w = 15$, for example, Table 7.30 shows that $\bar{H}_3 = .4973$ to the precision shown. To this accuracy,

$$\tilde{H}(3 + 12(M-1)) = \begin{cases} 1 - .4973 = 0.5027, & \text{Easter in March} \\ \frac{1}{15}\max(16 - k, 0) - .4973, & \text{Easter on April } k \end{cases}. \quad (16)$$

4.2 Quarterly Stock Easter Regressors from `easter[w]`

The corresponding results for quarterly series are obtained by using $t = j + 4(M-1)$ in place of (1) and modifying the remaining formulas to be consistent with this change. In the case of the Easter stock regressor formulas of the preceding subsection, for any $1 \leq w \leq 25$ we have

$$\tilde{H}_S(j + 4(M-1)) = \begin{cases} \tilde{H}(1 + 4(M-1) & , j = 1 \\ 0 & , 2 \leq j \leq 4 \end{cases}$$

and $\sum_{j=1}^{4} \tilde{H}_S(j + 4(M-1)) = \tilde{H}(1 + 4(M-1))$ for the generated stock quarterly regressor and its annual sums.

obtained from the dates of Easter for the 500 years 1600-2099 given in Bednarek (2007). The values of $\bar{H}_1, \ldots, \bar{H}_{12}$ given in Table 7.30 of U.S. Census Bureau (2007) are based on this approximate distribution. As long as $\sum_{j=1,12} \bar{H}_j = 1$ and a seasonal differencing used in the regARIMA model (or fixed seasonal regressors), the choice of these values will not change the Easter effect regressor coefficient estimate, because it is obtained from the differenced regressor, see (21). Moderately different choices will result in slightly different seasonal factors being calculated from the calendar-effect adjusted series.

4.3 Results for U.S. Manufacturers Inventory Series

These monthly and quarterly stock Easter regressors are implemented in a soon to be released version of the X-13-ARIMA-SEATS[3] software (X-13A-S for short) as the variables `easterstock[w]`of the `regression` spec. In analogy with what the program does with flow series, the specification `aictest=easterstock` causes the regARIMA model with no Easter effect regressor that is specified for a series to be augmented with an `easterstock[w]` regressor. Each of the three lengths $w = 1, 8$ and 15 is tried. The AIC values of the resulting three models are compared to the AIC value of the model with no Easter-effect regressor. AIC denotes the information criterion of Akaike (1972).

For the situation we consider, suppose the regARIMA model with no Easter effect regressor has a total of h regression and ARMA parameters, all estimated via maximum likelihood from the associated regARMA model for N successive data points obtained by applying the full differencing operator of the ARIMA model (see Section 6.1). With L_N denoting the log of this model's maximized likelihood and $L_N[w]$ the corresponding quantity for the model augmented with the `easterstock[w]` regressor, the AIC values of these models are given by

$$\begin{aligned} AIC &= -2L_N + 2h \\ AIC[w] &= -2L_N[w] + 2(h+1). \end{aligned} \tag{17}$$

Among the four models, X-13A-S follows Akaike's Minimum AIC criterion (MAIC) and selects the one with the smallest AIC[4].

In Findley and Monsell (2009) a total of 19 among 91 inventory series of the U.S. Census Bureau's monthly U.S. Manufacturers' Shipments, Inventories and Orders Survey (the M3 Survey) were determined to have stock trading day effects both by MAIC and by a regARIMA log likelihood-ratio test at the .05 level. Here we present the results of a further investigation of these 19 series based on applying the

[3] X-13-ARIMA-SEATS is an enhanced version of X-12-ARIMA that incorporates SEATS and provides a variety of common diagnostics for comparing X-11 and canonical model-based seasonal decompositions. See Findley (2005) for additional details.

[4] Actually, the program selects the model with the smallest value of the sample-size corrected AIC of Hurvich and Tsay (1989), defined generically by

$$AICC = -2L_N + 2h\left(1 - \frac{h+1}{N}\right)^{-1}$$

for AIC defined by (17). For the series we consider, h is small enough relative to N that the differences between the AIC and $AICC$ values do not affect the choice of model.

`aictest=easterstock` procedure and, if this selects an `easterstock[w]`regressor, applying to it a .05 significance level hypothesis test analogous to that used in Findley and Monsell (2009). The test uses the χ_1^2 asymptotic null hypothesis distribution of $2\{L_N[w] - L_N\}$, see Taniguchi and Kakizawa (2000, p. 61). The selected `easterstock[w]` regressor is accepted only if the AIC difference $\Delta AIC = AIC - AIC[w] = 2\{L_N[w] - L_N\} - 2$ satisfies $\Delta AIC \geq 1.8415$. This is equivalent to $2\{L_N[w] - L_N\} \geq 3.8415$, in accord with $\Pr\{\chi_1^2 \geq 3.8415\} = .05$.

The series of Findley and Monsell (2009) end in October, 2006. Their starting dates vary from January 1992 to January 1995 according to the choice made for regARIMA modeling of each series. These are end-of-month inventory series, with the qualification that adjustments are made to produce approximate end-of-calendar-month values for reporters to the M3 Survey who provide end-of-report-period values for four- or five-week periods instead of for calendar months. For details, see M3 (2008). Of the 19 series, four were determined by the criteria described above to have Easter holiday effects. These series are identified in Table 1.

Table 1. M3 inventory series with significant Easter Holiday Effects

21SFI	Finished Goods Inventories of Wood Products
26SMI	Materials and Supplies Inventories of Plastics and Rubber Products
31ATI	Total Inventories of Iron and Steel Mills and Ferroalloy and Steel Products
31SFI	Finished Goods Inventories of Primary Metals

Table 2 shows the values w of the regressors chosen, their AIC differences from the model without the regressor, the estimated coefficient $\hat{\alpha}$ of $\tilde{H}_S(t)$, its estimated standard error $\hat{\sigma}_\alpha$, and also ratios (18) of (empirical) root mean square out-of-sample forecast error at leads $l = 1, 12$.

The ratios are defined as follows when the data are log transformed prior to modeling, as holds for our examples. For the log-transformed time series data y_t, $1 \leq t \leq T$, given a forecast lead $l \geq 1$ and forecast origin $1 \leq \tau \leq T - l$, let $y_{\tau+l|\tau}$ denote the forecast of $y_{\tau+l}$ obtained from y_t, $1 \leq t \leq \tau$ that is provided by the regARIMA model with no stock Easter effect regressor when its parameters are estimated from y_t, $1 \leq t \leq \tau$. With τ_0 denoting the index of the initial forecast origin (January 2004 in our case), define $RMSE_l = \left\{(T - l - \tau_0 + 1)^{-1} \sum_{\tau=\tau_0}^{T-l} \left(y_{\tau+l} - y_{\tau+l|\tau}\right)^2\right\}^{1/2}$. Let $RMSE_l[w]$ denote the analogous value for the model with the selected `easterstock[w]` regressor. The last two columns of Table 2 give the values of the ratios

$$RMSE_l / RMSE_l[w], l = 1, 12. \tag{18}$$

In parentheses, they also include the corresponding ratios $RMSE_l^{\text{mar}}/RMSE_l^{\text{mar}}[w]$ for just the March forecasts. (March is the month in which the Easter effect regressors are non-zero.) With two slight exceptions for March forecasts, the ratios are larger than 1.0, showing that use of the chosen stock holiday regressors usually provides root mean square forecast improvement at both lags for all four series. For $l = 1$ there are 33 forecasts (3 for March) and for $l = 12$ there are 22 (2 for March). Apart from the two exceptions just mentioned, the root mean square error improvement is greater for the March only forecasts.

Table 2. Easter interval length choice w *and values of* ΔAIC, $\hat{\alpha}$, $\hat{\sigma}_\alpha$ and of *the ratios* $RMSE_l/RMSE_l[w]$ (and $RMSE_l^{\text{mar}}/RMSE_l^{\text{mar}}[w]$) $l = 1, 12$
for the series of Table 1

Series	w	ΔAIC	$\hat{\alpha}$	$\hat{\sigma}_a$	$l = 1$	$l = 12$
21SFI	15	4.011	.025	.0098	1.009 (1.433)	1.031 (1.230)
26SMI	8	3.307	.002	.0008	1.010 (1.187)	1.008 (0.997)
31ATI	15	4.533	.008	.0030	1.010 (0.995)	1.005 (1.058)
31SFI	1	8.551	.017	.0050	1.076 (1.103)	1.030 (1.229)

The mean square error values for (18) can be obtained from X-13A-S (`save = fce` in the `history` spec) as can the forecast error values for each month (`save = fch` in the `history` spec). From the latter, the ratios $RMSE_l^{\text{mar}}/RMSE_l^{\text{mar}}[w]$, $l = 1, 12$ for March forecasts can be computed.

Table 3 of Findley and Monsell (2009) shows that the stock trading day regressor appropriate for these four series is the one-coefficient regressor derived in this reference. Further, it shows that only one of these series, 31ATI, has model residuals with significant Ljung-Box goodness of fit Q statistics when the model with this regressor (and no Easter-effect regressor) is fitted. The same is true when these models are augmented with the holiday regressor indicated in Table 2. For 31ATI, use of this regressor reduces from 15 to 12 the number of significant Q statistics at lags up through 24.

Finally, we present information regarding the ranges of the seasonal, stock trading day, and stock Easter effect adjustment factors for these series, referring to these factors generically as S, TD, and E, respectively. Table 3 reveals that, with the exception of 31SFI, the ranges of the holiday factors,

$$\max_t \left\{ \exp\left(\hat{\alpha} \tilde{H}_S(t) \right) \right\} - \min_t \left\{ \exp\left(\hat{\alpha} \tilde{H}_S(t) \right) \right\}$$

are larger than the ranges of the trading day factors. This is consistent with what is typically found for flow series. Also, the largest range of holiday factors occurs for 21SFI, the series with the largest seasonal factor range. Compared to flow series, the ranges are quite small, with the moderate exception of the seasonal factor range of 21SFI.

Table 3. Minimum and maximum of the seasonal, trading day and Easter effect adjustment factors

Series	$\min S$	$\max S$	$\min TD$	$\max TD$	$\min E$	$\max E$
21SFI	91.35	108.25	99.69	100.31	98.83	101.34
26SMI	98.54	101.41	99.80	100.20	99.59	100.76
31ATI	97.92	102.41	99.72	100.28	99.61	100.44
31SFI	96.56	103.21	99.40	100.61	99.59	101.34

Titova and Monsell (2009) present results for other Census Bureau inventory series found to have statistically significant stock Easter holiday effects.

5 Modifications for Other Situations

We now show by means of several examples that when either (3) fails to hold, or seasonal mean removal (6) is not done, or stocks are measured on a day prior to the last day of the month, then modifications of the approach taken in the preceding section are typically needed to obtain stock regressors satisfactory for moving holiday adjustment.

5.1 Examples in Which Flow Regressor Annual Sums are Not Constant

There can be situations in which the interval associated with a holiday regressor of the form (4) overlaps two different years, with the consequence that annual sums $\sum_{j=1}^{12} H\left(j+12(M-1)\right)$ are not constant, i.e. (3) fails to hold. For example, Lin and Liu (2003) construct regressors (6) starting from (4) to model and adjust for the effects of Chinese lunar calendar holidays. One of these is the Chinese Lunar New Year, which has a 60 year cycle of dates varying between January 21 and February 20, see http://www.webexhibits.org/calendars/calendar-chinese.html. If the preholiday interval associated with increased economic activity has a length of

21 days or more, then it will include one or more days from December of the year preceding any year M in which the holiday falls on January 21. When this happens, $\sum_{j=1}^{12} H(j+12(M-1))$ is less than one. In some other years it will be equal to one, so neither (3) nor (9) will hold. However (assuming the interval's length never exceeds 51 days), the analogues of (3) and (9) will hold for 12 month sums that begin in December and end in November: $\sum_{j=0}^{11} H(j+12(M-1)) = 1$. So, if $H(t)$ has been calculated for $t = 12(M_0 - 1)$, one can define $\tilde{H}_S(t)$ for any $t = j + 12(M-1)$ with $0 \leq j \leq 11$ and $M \geq M_0$ by means of $\tilde{H}_S(t) = \sum_{u=12(M_0-1)}^{t} \tilde{H}(u)$ and obtain the simplified formula $\tilde{H}_S(t) = \sum_{i=0}^{j} \tilde{H}(i+12(M-1))$ together with the properties (13) and (15).

A holiday for which such a simplified stock holiday formula does not exist is Eid ul-Fitr' at the end of the Islamic lunar calendar month of Ramadan. Ramadan moves through all of the calendar months of the Gregorian calendar, see http://www.webexhibits.org/calendars/calendar-islamic.html, and holiday intervals associated with Eid ul-Fitr' occasionally overlap two calendar years. As a result, there is no 12 month span that always includes the holiday interval.

5.2 Further Examples Requiring a Redefinition of "Months" and "Years"

The modification proposed above to accommodate the Chinese Lunar New Year effects can be interpreted as a redefinition of "years" to denote consecutive 12-months intervals starting with December for the purpose of obtaining regressors with desirable properties. Here we describe a useful example for which a similar strategy can be successfully employed.

X-12-ARIMA and X-13A-S provide stock trading day regressors for series for which the stock is always measured on the ω-th day of the month for fixed $1 \leq \omega \leq 31$ or at the end of the month when the month's length is less than ω. Easter regressors with properties analogous to those of Section 4.1 can be obtained for such stocks by redefining the j-th month of the year to start the day after the stock day of the preceding calendar month and to end on the stock day of the j-th calendar month. For example, with $\omega = 15$, the first day of the first newly defined month of year M begins on December 16 of year $M-1$ and this newly defined first month ends on January 15th. If Easter falls on April 17 of year M, this is the second day of the fifth newly defined month. As a consequence, for each interval length w, the values of the proportionality regressors (4) must be recalculated, as must the associated level plus seasonal components (5), before the desired stock regressor values (11) can be obtained.

Titova and Monsell (2009) provide results for some series for which an Easter effect regressor obtained in this way for $\omega = 28$ is favored over the regressors of Section 4.1 and over no holiday regressor.

5.3 Deficiency of Flow Regressors with a Seasonal Component

Here we use the Statistics Canada Easter regressor `sceaster[w]` of X-12-ARIMA to demonstrate that when a flow regressor is not centered on its calendar month averages, the stock regressor it generates by accumulation can have a nonzero level plus seasonal component even when the flow regressor's annual sums (and level component) are zero.

The interval of length w associated with `sceaster[w]` consists of Easter and the preceding $w-1$ days. In year M, it is defined by the number $n(M) = n(M,w)$ of days of March included in this interval. For $1 \leq w \leq 22$ and $1 \leq j \leq 12$, its values are given by

$$H^{SC}(j + 12(M-1)) = \begin{cases} \frac{n(M)}{w} & , j = 3 \\ -\frac{n(M)}{w} & , j = 4 \\ 0 & , j \neq 3,4 \end{cases} , \tag{19}$$

so its annual sums satisfy

$$\sum_{j=1}^{12} H^{SC}(j + 12(M-1)) = 0, \quad M \geq M_0. \tag{20}$$

Hence its level component is zero:

$$\bar{H}^{SC} = (12P)^{-1} \sum_{t=1}^{12P} H^{SC}(t) = (12P)^{-1} \sum_{M=1}^{P} \sum_{j=1}^{12} H^{SC}(j + 12(M-1)) = 0.$$

However, its seasonal component $\bar{H}_1^{SC}, \ldots, \bar{H}_{12}^{SC}$ defined in analogy with (5) is given by $\bar{H}_3^{SC} = P^{-1} \sum_{M=1}^{P} \frac{n(M)}{w} > 0$, $\bar{H}_4^{SC} = -\bar{H}_3^{SC}$ and $\bar{H}_j^{SC} = 0$, $j \neq 3, 4$. Thus $H^{SC}(t)$ has a nonzero seasonal component.

Now consider the stock regressors generated by these regressors, $H_S^{SC}(t) = \sum_{u=1+12(M_0-1)}^{t} H^{SC}(u)$. It follows from (20) that for $t = j + 12(M-1)$, this formula simplifies to $H_S^{SC}(t) = \sum_{i=1}^{j} H^{SC}(i + 12(M-1))$. From (19), it is nonzero only in March, when its value is $n(M)/w$. Hence, it has a nonzero level plus seasonal component, given by $\bar{H}_{S,j}^{SC} = \bar{H}_3^{SC}$ for $j = 3$ and $\bar{H}_{S,j}^{SC} = 0$ for $j \neq 3$, whose level component is positive, $\bar{H}_S^{SC} = (1/12)\sum_{j=1}^{12} \bar{H}_{S,j}^{SC} = \bar{H}_3^{SC}/12$. Consequently, for Easter-effect adjustment of a stock series for whose flow series $H^{SC}(t)$ is appropriate, the regressor $H_S^{SC}(t)$ should be replaced by the adjusted regressor defined at $t = j + 12(M-1)$ by $\tilde{H}_S^{SC}(j+12(M-1)) - \bar{H}_{S,j}^{SC}$. This is also the regressor generated by the level plus seasonally adjusted flow regressor $H^{SC}(j+12(M-1)) - \bar{H}_j^{SC}$, $M \geq M_0, 1 \leq j \leq 12$.

6 Final Remarks

6.1 An Indirect Testing Strategy

To facilitate exploration of a variety of stock effects for a set of time series, it could be helpful to note that, using only the flow regressor $\tilde{H}(t)$ from which the stock regressor $\tilde{H}_S(t)$ of (10) would be generated, it is possible to test a stock series $Y(t)$ for the presence of the effect defined by $\tilde{H}_S(t)$: This is done by testing the flow series $X(t) = Y(t) - Y(t-1)$, $t \geq 2$ for the significance of $\tilde{H}_S(t) - \tilde{H}_S(t-1) = \tilde{H}(t)$.

To carry out this strategy, one must find a regARIMA model for the differenced series $X(t)$ whose regression component includes $\tilde{H}(t)$ and whose associated ARIMA model is invertible, i.e. the zeroes of its moving average polynomial, if it has one, have magnitude greater than 1. This is required by the asymptotic distribution theory of the goodness-of-fit statistics and the test statistics of X-12-ARIMA related to the estimates of the regression coefficients, see Pierce (1971, a, b). (Thus the strategy cannot be used when first differencing induces noninvertibility, i.e. when $Y(t)$ can be well modeled without the use of a first difference of any kind, a very rare situation.)

The estimated coefficient obtained for $\tilde{H}(t)$ as a component of the regression vector $Z_X(t)$ of a regARIMA$(p,d,q)(P,D,Q)_{12}$ model for $X(t)$ coincides with the estimated coefficient of $\tilde{H}_S(t)$ for a regARIMA$(p,d+1,q)(P,D,Q)_{12}$ model for $Y(t)$ with a regression component $Z_Y(t)$ such that $Z_Y(t) - Z_Y(t-1) = Z_X(t)$, for example $Z_Y(t) = \sum_{u=1}^{t} Z_X(t)$. This happens because estimation is done on the fully differenced data modeled as regARMA$(p,q)(P,Q)$ with the fully differenced regressors, and this is the same model for both regARIMA models. For example, if the model for $X(t)$ is

$$(1-B)^d \left(1-B^{12}\right)^D \phi(B)\,\Phi\left(B^{12}\right)(X(t) - \beta Z_X(t)) = \theta(B)\,\Theta\left(B^{12}\right) a(t),$$

its parameters are estimated from the regARMA model of

$$W(t) = (1-B)^d (1-B^{12})^D X(t),$$

$$\phi(B)\,\Phi\left(B^{12}\right)(W(t) - \beta Z_W(t)) = \theta(B)\,\Theta\left(B^{12}\right) a(t), \tag{21}$$

with $Z_W(t) = (1-B)^d (1-B^{12})^D Z_X(t)$. This is also the estimation model for $Y(t)$ since

$$(1-B)^{d+1}\left(1-B^{12}\right)^D \phi(B)\,\Phi\left(B^{12}\right)(Y(t) - \beta Z_Y(t)) = \theta(B)\,\Theta\left(B^{12}\right) a(t),$$

and $(1-B) Z_Y(t) = Z_X(t)$.

When regARIMA modeling of the observed stock series requires the log transformation, then in the formulas above $Y(t) = \log y(t)$ and $X(t) = Y(t) - Y(t-1) = \log(y(t)/y(t-1))$. We next consider log transformed data.

6.2 The Log Transform Case

For many positive-valued stock series, log transformation of the original data is required to obtain an adequately fitting regARIMA model. If the observed stock series $y(t)$ has a holiday effect, but $Y(t) = \log y(t)$ is the modeled series, then $Y(t)$ can be regarded as a stock series with a holiday effect inherited from $y(t)$, which would induce a holiday effect in $X(t) = Y(t) - Y(t-1)$, a flow series (in the log domain). Because $Y(t) = Y(12(M_0-1)) + \sum_{u=1+12(M_0-1)}^{t} X(t)$, if $\tilde{H}(t)$ is a good regressor for the holiday effects of $X(t)$, then $\tilde{H}_S(t)$ defined by (10) can be an effective regressor for the holiday effects of $Y(t)$, resulting in useful holiday effect factors for $y(t)$ of the form $\exp\left(\alpha \tilde{H}_S(t)\right)$, as the forecasting results of Section 4.3 show.

6.3 Vector Regressors

Finally, we note that, although the specific holiday regressor examples considered above are scalar, the general discussion still applies when the regressors are column vectors and their coefficients are conforming row vectors.

Acknowledgments. This document arose in response to a question from Julian Chow of the U.K. Office of National Statistics about how to estimate Easter effects in stock series. I am indebted to William Bell for helpful suggestions and comments on earlier drafts, including a suggestion that stimulated the development of Section 6.1. Brian Monsell enhanced X-13A-S to include the stock series Easter-effect regressors of Sections 4.1 and 4.2 in order to facilitate empirical investigations.

References

[1] Akaike, H. (1973). Information theory and an extension of the likelihood principle. In B. Petrov and F. Czaki (Eds.), Second International Symposium on Information Theory, pp. 267–287. Budapest: Akademia Kiado.

[2] Anderson, T.W. (1971). The Statistical Analysis of Time Series. New York: Wiley.

[3] Bednarek, M. (2007) The Dates of Easter for 500 Years, [Online] http://home.swiftdsl.com.au/~mbednarek//easter. php [2007, November 18].

[4] Bell, W.R. (1984). Seasonal Decomposition of Deterministic Effects. Statistical Research Division. U.S. Bureau of the Census Statistical Research Division Report Number: Census/SRD/RR-84/01. http://www.census.gov/srd/papers/pdf/rr84-1.pdf

[5] Bell, W.R. (1995). Correction to Seasonal Decomposition of Deterministic Effects. Statistical Research Division. U.S. Bureau of the Census Statistical Research Division Report Number: Census/SRD/RR-95/01. http://www.census.gov/srd/papers/pdf/rr95-01.pdf

[6] Cleveland, W.P. and Grupe, M.R. (1983). Modeling Time Series When Calendar Effects Are Present. In Applied Time Series Analysis of Economic Data (ed. A. Zellner) pp. 57-67, Economic Research Report ER-6, U.S. Department of Commerce, Washington: Bureau of the Census.
http://www.census.gov/ts/papers/Conference1983/ClevelandGrupe1983.pdf

[7] Findley, D.F. (2005). Some Recent Developments and Directions in Seasonal Adjustment. Journal of Official Statistics 21, 343-365. http://www.jos.nu/Articles/abstract.asp?article=212343

[8] Findley, D.F. and Monsell, B.C. (2009). Modeling Stock Trading Day Effects Under Flow Day-of-Week Effect Constraints. Journal of Official Statistics (to appear). http://www.census.gov/ts/papers/findleymonselljos2008.pdf

[9] Findley, D.F. and Soukup, R.J. (2000). Modeling and Model Selection for Moving Holidays, 2000 Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 102-107, American Statistical Association: Alexandria. http://www.census.gov/ts/papers/asa00_eas.pdf

[10] Gómez, V. and A. Maravall (1997, 2003). Programs TRAMO and SEATS : Instructions for the User (beta version: June 1997) Working Paper 97001, Ministerio de Economía y Hacienda, Dirrectíon General de Análisis y Programacíon Presupuestaria, Madrid. http://www.bde.es/servicio/software/tramo/manualdos.pdf
2003 Update http://www.bde.es/servicio/software/tramo/updatedos.pdf

[11] Hurvich, C.M. and C. Tsai (1989). Regression and Time Series Model Selection in Small Samples. Biometrika 76, 297–307.

[12] Lin, J.-L. and T.-S. Liu (2003). Modeling Lunar Calendar Holiday Effects in Taiwan. Taiwan Economic Forecast and Policy 33 (1), 1–37. http://www.census.gov/ts/papers/lunar.pdf

[13] M3 (2008). 2008 Instruction Manual For Reporting in Monthly Survey M3 Manufacturers' Shipments, Inventories, and Orders. U.S. Bureau of the Census. http://www.census.gov/indicator/www/m3/contactinfo/instructionman.pdf

[14] Montes, M.J. (1998). Calculation of the Ecclesiastical Calendar. [Online] http://www.smart.net/~mmontes/ec-cal.html

[15] Pierce, D. (1971a). Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors. Biometrika 58, 299–312.

[16] Pierce, D. (1971b). Distribution of Residual Autocorrelations in the Regression Model with Autoregressive-Moving Average Errors, Journal of the Royal Statistical Society Series B 33, 140-146.

[17] Taniguchi, M. and Kakizawa, Y. (2000). Asymptotic Theory of Statistical Inference of Time Series. New York: Springer-Verlag.

[18] Titova, N. and Monsell, B.C. (2009). Detecting Stock Calendar Effects in U.S. Census Bureau Inventory Series. Proceedings of the Joint Statistical Meetings, American Statistical Association: Alexandria.

[19] U.S. Census Bureau (2007). X-12-ARIMA Reference Manual, Version 0.3 http://www.census.gov/ts/x12a/v03/x12adocV03.pdf

[20] Wikipedia Contributors (2009).Stock and flow, Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Stock_and_flow (accessed January 15, 2009).