Journal of Official Statistics, Vol. 25, No. 3, 2009, pp. 415–430

Modeling Stock Trading Day Effects Under Flow Day-of-Week Effect Constraints

David F. Findley[1] *and Brian C. Monsell*[2]

By deriving an invertible linear relation between stock and flow trading day regression coefficients, we show how flow day-of-week effect constraints can be imposed upon the day-of-week-effect component of the stock trading day model of Bell used in X-12-ARIMA. As an application, a new one-coefficient stock trading day model is derived from the constraints that give rise to the one-coefficient weekday-weekend-contrast flow trading day model of TRAMO and X-12-ARIMA. We present summary results and some details of a quite successful application of the new model to the manufacturers' inventory series of the U.S. Census Bureau's M3 Survey. (JEL C87, C82).

Key words: Time series; RegARIMA models; seasonal adjustment; trading day adjustment; X-12-ARIMA; X-13A-S; M3 survey; inventory series.

1. Introduction

Monthly economic time series usually measure accumulations of daily economic activity, for example monthly sales or inventories at month's end. Within the daily activity, there can be a pattern of within-week variations that causes the monthly values to change with the day-of-week composition of the month, e.g., the days that occur five times or the day on which the month ends. If this pattern is stable enough over time and strong enough, it leads to statistically significant effects on the monthly data that can and should be accounted for when interpreting or modeling the data. These day-of-week effects, sometimes together with effects related to month-length, are called trading day or working day effects. Proper modeling of trading day effects generally leads to better ARIMA models and better preadjusted series for the seasonal factor identification procedures of TRAMO-SEATS (Gómez and Maravall 1997, 2003) and of X-12-ARIMA (Findley, Monsell, Bell, Otto, and Chen 1998 and U.S. Census Bureau 2007).

The trading day effect regressors of Bell and Hillmer (1983) for flow series, i.e., monthly accumulations such as total monthly sales, are available in TRAMO-SEATS and X-12-ARIMA as optional components of regARIMA models (regression models with ARIMA disturbances). For modeling end-of-month stock series, such as end-of-month

[1] U.S. Census Bureau, Washington, DC 20233, U.S.A. Email: David.F.Findley@census.gov
[2] U.S. Census Bureau, Washington, DC 20233, U.S.A. Email: Brian.C.Monsell@census.gov
Acknowledgments: The authors are grateful to William Bell, Tucker McElroy and three referees for comments and suggestions that substantially improved this article.
Disclaimer: This article is released to inform and to encourage discussion of research. All opinions expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

inventories, trading day regressors were derived in Cleveland and Grupe (1983) and in the research reports of Bell (1984, 1995) by treating the stock series as an accumulation of consecutive monthly flows.

As will be seen below, the basic day-of-week effect models are specified by six independent coefficients. For flow series, known or inferred properties of individual series sometimes suggest relations among the coefficients that lead to more parsimonious models. The most important example, reviewed in Section 3.1, is the one in which the days Monday through Friday are assumed to contribute equally to the economic activity, and Saturday's contribution is assumed equal to Sunday's. Imposed upon the day-of-week coefficients, these constraints gave rise to the one-coefficient flow trading day model of TRAMO-SEATS and, later, of X-12-ARIMA.

In the next section, we derive an invertible linear relation between flow and stock day-of-week-effect regression coefficients that enables us to show, in Section 3, how flow coefficient constraints transfer to stock coefficients. As an application, in Section 3.1 a new one-coefficient trading day regression model for stock series is derived from the constraints that give rise to the one-coefficient flow series model mentioned above. Section 5 describes how this new model was applied with considerable success to the manufacturers' inventory series of the U.S. Census Bureau's M3 Survey. Section 6 considers other regressors that are applicable when no data transformation is needed for regARIMA modeling and shows how, in this situation, more general flow-coefficient constraints can be implemented using Cleveland and Grupe's stock trading day regressors. A final section presents some conclusions. Until Section 6, the term trading day effects is used as a synonym for day-of-week effects.

2. Formulas Relating Flow and Stock Coefficients

2.1. *The Basic Flow Day-of-Week Effect Model*

With $i = 1, \ldots, 7$ indexing Monday through Sunday and $t = 1, 2, \ldots, T$ indexing the successive months of the time series, let $X_t(i)$ be the number of times the i-th weekday occurs in month t. Then $\sum_{i=1}^{7} \beta_i X_t(i)$ is the basic formula for flow series trading day effects underlying the regression model components of regARIMA models used to estimate such effects; see Bell and Hillmer (1983) and Findley et al. (1998). To derive specialized models for differing situations, $\sum_{i=1}^{7} \beta_i X_t(i)$ is decomposed into day-of-week and length-of-month effects. Setting $\bar{\beta} = 1/7 \sum_{i=1}^{7} \beta_i$, $\tilde{\beta} = \beta_i - \bar{\beta}$ and $m_t = \sum_{i=1}^{7} X_t(i)$ (the length of month t), we have $\beta_i = \tilde{\beta}_i + \bar{\beta}_i$ and

$$\sum_{i=1}^{7} \beta_i X_t(i) = \sum_{i=1}^{7} \tilde{\beta}_i X_t(i) + \bar{\beta} m_t \tag{1}$$

Due to

$$\sum_{i=1}^{7} \tilde{\beta}_i = 0 \tag{2}$$

setting $X_t^*(i) = X_t(i) - X_t(7)$, $1 \leq i \leq 6$ and $X_t^* = \left[X_t^*(1)X_t^*(2). . .X_t^*(6)\right]$, we also have

$$\sum_{i=1}^{7} \tilde{\beta}_i X_t(i) = \sum_{i=1}^{6} \tilde{\beta}_i X_t^*(i) = X_t^* \tilde{\beta} \tag{3}$$

where

$$\tilde{\beta} = \left[\tilde{\beta}_1 \tilde{\beta}_2. . .\tilde{\beta}_6\right]' \tag{4}$$

Our focus in this article is the effect of constraints for $\tilde{\beta}$ on the coefficients of the associated stock series analogue of $\sum_{i=1}^{6} \tilde{\beta}_i X_t^*(i)$, which is derived in the next section. A key property of $\sum_{i=1}^{6} \tilde{\beta}_i X_t^*(i)$ arises from the repetition of the day-of-week calendar, every 28 years if rare corrections to the Gregorian calendar are ignored. This causes the seven long-term means $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} X_t(i)$, $i = 1, \ldots, 7$ to have the same value, effectively the common value of the mean for $T = 12 \times 28 = 336$. Consequently,

$$\lim_{T\to\infty} T^{-1} \sum_{i=1}^{6} \tilde{\beta}_i X_t^*(i) = \sum_{i=1}^{6} \tilde{\beta}_i \left\{ \lim_{T\to\infty} T^{-1} X_t^*(i) \right\} = \sum_{i=1}^{6} \tilde{\beta}_i \cdot 0 = 0 \tag{5}$$

That is, the day-of-week component $\sum_{i=1}^{7} \tilde{\beta}_i X_t(i)$ of (1) has the important property for trading day adjustment of being level neutral in the sense that its long-term mean is zero. See Bell (1984, 1995) for a more general and detailed discussion.

2.2. *The Basic Stock Day-of-Week Effect Model*

Bell (1984, 1995) obtained stock-series trading day regression models by accumulating the monthly flow effects (1):

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \beta_i X_j(i) = \sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_j(i) + \bar{\beta} \sum_{j=1}^{t} m_j \tag{6}$$

To present Bell's day-of-week-effect formula, for $1 \leq k \leq 7$ let $I_t(k) = 1$ if month t ends on the k-th day of the week. Otherwise let $I_t(k) = 0$. Suppose that the k_0-th day of the week immediately precedes the first day of month $t = 1$. We define

$$\gamma_7 = -\sum_{i=1}^{k_0} \tilde{\beta}_i \tag{7}$$

$$\gamma_k = \sum_{i=1}^{k} \tilde{\beta}_i + \gamma_7, \; 1 \leq k \leq 6 \tag{8}$$

Then, with $\bar{\gamma} = 1/7 \sum_{k=1}^{7} \gamma_k$ and $\tilde{\gamma}_k = \gamma_k - \bar{\gamma}$, $1 \leq k \leq 7$, the derivation on pp. 5–7 of Bell (1984) establishes that

$$\sum_{j=1}^{t} \sum_{i=1}^{7} \tilde{\beta}_i X_j(i) = \sum_{k=1}^{7} \gamma_k I_t(k) = \sum_{k=1}^{7} \tilde{\gamma}_k I_t(k) + \bar{\gamma} \tag{9}$$

Due to

$$\sum_{k=1}^{7} \tilde{\gamma}_k = 0 \qquad (10)$$

with $I_t^*(k) = I_t(k) - I_t(7),\ 1 \leq k \leq 6$, we have

$$\sum_{k=1}^{7} \tilde{\gamma}_k I_t(k) = \sum_{k=1}^{6} \tilde{\gamma}_k I_t^*(k) = I_t^* \tilde{\gamma} \qquad (11)$$

for

$$\tilde{\gamma} = \left[\tilde{\gamma}_1\ \tilde{\gamma}_2 \ldots \tilde{\gamma}_6\right] \qquad (12)$$

and

$$I_t^* = \left[I_t^*(1)\, I_t^*(2) \ldots I_t^*(6)\right] \qquad (13)$$

The expression $I_t^* \tilde{\gamma}$ is Bell's stock day-of-week-effect formula. The linear relations between $\tilde{\gamma}$ and $\tilde{\beta}$ will be derived below.

As with the flow trading day regressors, the repetition of the day-of-week calendar results in the regressors $I_t(k)$ having long-term means that do not depend on k, from which it follows that the stock day-of-week effects given by (11) are level-neutral,

$$\lim_{T\to\infty} T^{-1} \sum_{k=1}^{6} \tilde{\gamma}_k I_t^*(k) = \sum_{k=1}^{6} \tilde{\gamma}_k \left\{ \lim_{T\to\infty} T^{-1} I_t^*(k) \right\} = \sum_{k=1}^{6} \tilde{\gamma}_k \cdot 0 = 0 \qquad (14)$$

in analogy with (5). The same is true for $\bar{w}$-th day-of-month stocks defined as follows: for a specified $1 \leq \bar{w} \leq 31$, the stock is for the last day of the month if the month-length is less than $\bar{w}$; otherwise it is for the $\bar{w}$-th day of the month. (Thus, the choice $\bar{w} = 31$ specifies end-of-month stocks). The validity of (9) and (14) for $\bar{w} < 31$ can be obtained by redefining months in Bell's derivation to refer to the intervals between consecutive stock measurements. The stock regressors I_t^* available in X-12-ARIMA are for $\bar{w}$-th day of month stocks for any $1 \leq \bar{w} \leq 31$. The same regression models have been implemented in TRAMO-SEATS for its next release (Maravall 2008).

2.3. *Relations Between the Basic Flow and Stock Coefficients*

We now derive the following invertible linear relations between $\tilde{\gamma}$ and $\tilde{\beta}$.

$$\tilde{\beta} = N\tilde{\gamma} \qquad (15)$$

with

$$N = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix}$$

so

$$\tilde{\gamma} = N^{-1}\tilde{\beta} \tag{16}$$

with

$$N^{-1} = \frac{1}{7}\begin{bmatrix} 1 & -5 & -4 & -3 & -2 & -1 \\ 1 & 2 & -4 & -3 & -2 & -1 \\ 1 & 2 & 3 & -3 & -2 & -1 \\ 1 & 2 & 3 & 4 & -2 & -1 \\ 1 & 2 & 3 & 4 & 5 & -1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \tag{17}$$

To obtain (15), we note first that with

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(8) is equivalent to $L\tilde{\beta} = \left[\gamma_1 - \gamma_7 \ \gamma_2 - \gamma_7 \ldots \gamma_6 - \gamma_7\right]'$. Next, using (10), observe for $k = 1, \ldots, 6$ that $\gamma_k - \gamma_7 = \tilde{\gamma}_k - \tilde{\gamma}_7 = \tilde{\gamma}_k + \sum_{j=1}^{6} \tilde{\gamma}_j = 2\tilde{\gamma}_k + \sum_{j \neq k} \tilde{\gamma}_j$. Thus, with

$$M = L + L' = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 \end{bmatrix}$$

we have $L\tilde{\beta} = M\tilde{\gamma}$. From

$$L^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

we obtain $N = L^{-1}M$ and (15). Next, noting that M is sum of the identity matrix and $\mathbf{1}'\mathbf{1}$, with $\mathbf{1} = [\mathbf{1}, \ldots, \mathbf{1}]$, a standard inverse formula, see Noble (1969, p. 148) for example, yields

$$\mathrm{M}^{-1} = \frac{1}{7}\begin{bmatrix} 6 & -1 & -1 & -1 & -1 & -1 \\ -1 & 6 & -1 & -1 & -1 & -1 \\ -1 & -1 & 6 & -1 & -1 & -1 \\ -1 & -1 & -1 & 6 & -1 & -1 \\ -1 & -1 & -1 & -1 & 6 & -1 \\ -1 & -1 & -1 & -1 & -1 & 6 \end{bmatrix}$$

and therefore (17) and (16) from $\mathrm{N}^{-1} = \mathrm{M}^{-1}\mathrm{L}$.

3. The Effect of Flow-Coefficient Constraints on Stock Coefficients

With stock series, there can be information about the associated flow series that suggests one or more linear constraints on the day-of-week-effect regression coefficients $\tilde{\beta}_j$, $1 \le j \le 6$ in (3) of the form

$$\sum_{i=1}^{6} h_i \tilde{\beta}_i = 0 \tag{18}$$

With $\tilde{\beta}$ as in (4), a set of such constraints can be expressed as

$$H\tilde{\beta} = 0 \tag{19}$$

for some matrix H of full rank (less than six). From (15) and (16), the constraint (19) on $\tilde{\beta}$ is equivalent to the constraint on $\tilde{\gamma}$ given by

$$HN\tilde{\gamma} = 0 \tag{20}$$

As the familiar constrained flow model considered below will illustrate, a natural source of constraints (18) are *contrasts* on the coefficients β_i, $1 \le i \le 7$ of (1), i.e., constraints of the form $\sum_{i=1}^{7} c_i \beta_i = 0$ with $\sum_{i=1}^{7} c_i = 0$. Indeed, for these, the $\tilde{\beta}_i = \beta_i - \bar{\beta}$ satisfy

$$0 = \sum_{i=1}^{7} c_i \tilde{\beta}_i = \sum_{i=1}^{6} c_i \tilde{\beta}_i - c_7 \sum_{i=1}^{7} \tilde{\beta}_i = \sum_{i=1}^{6} (c_i - c_7)\tilde{\beta}_i$$

which yields (18) with $h_i = c_i - c_7$, $1 \le i \le 6$. (Conversely, a constraint $\sum_{i=1}^{6} h_i \tilde{\beta}_i = 0$ on $\tilde{\beta}$ yields the contrast $\sum_{i=1}^{7} c_i \beta_i = 0$ with $c_7 = -1/7 \sum_{j=1}^{6} h_j$ and $c_i = h_i + c_7$, $1 \le i \le 6$).

Silvey (1975, p. 60) outlines a general approach to obtaining the regressors and regression models implied by linear constraints. For (20), this involves adding $r = 6 - \text{rank}\,(H)$ rows to HN that are chosen to obtain an invertible 6×6 matrix J. The decomposition

$$I_t^* \tilde{\gamma} = \left(I_t^* J^{-1}\right) J \tilde{\gamma} \tag{21}$$

then reveals the constrained regressor and its coefficients in the last r rows of $I_t^* J^{-1}$ and $J\tilde{\gamma}$, as we illustrate below. But often the constrained form of $\tilde{\beta}$ is known or easily derived. Then the constrained form of $\tilde{\gamma}$ follows from (16), and this leads via (11) to the constrained stock trading day regression model without matrix inversion, as we illustrate first with a fundamental example.

3.1. *A New Class of One-Coefficient Stock TD Models*

We consider the one-coefficient weekday-weekend-contrast flow day-of-week-effect model of TRAMO and X-12-ARIMA. This arises from equality constraints between the weekday coefficients $\beta_1, \ldots, \beta_5$ and between β_6 and β_7,

$$\beta_1 = \beta_2 = \cdots = \beta_5, \beta_6 = \beta_7 \tag{22}$$

which can be expressed as five contrasts $\beta_i - \beta_{i+1} = 0$, $i = 1, 2, 3, 4, 6$. The constraints (22) and (2) immediately yield

$$\tilde{\beta} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -\frac{5}{2} \end{bmatrix}' \tilde{\beta}_5 \tag{23}$$

Thus, from (16) and (17),

$$\tilde{\gamma} = \frac{1}{7} \begin{bmatrix} 1 & -5 & -4 & -3 & -2 & -1 \\ 1 & 2 & -4 & -3 & -2 & -1 \\ 1 & 2 & 3 & -3 & -2 & -1 \\ 1 & 2 & 3 & 4 & -2 & -1 \\ 1 & 2 & 3 & 4 & 5 & -1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -\frac{5}{2} \end{bmatrix} \tilde{\beta}_5 = \frac{1}{2} \begin{bmatrix} -3 \\ -1 \\ 1 \\ 3 \\ 5 \\ 0 \end{bmatrix} \tilde{\beta}_5$$

which yields $\tilde{\gamma}_5 = 5/2 \tilde{\beta}_5$ as well as the constrained form of $\tilde{\gamma}$,

$$\tilde{\gamma} = \begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} & \frac{1}{5} & \frac{3}{5} & 1 & 0 \end{bmatrix}' \tilde{\gamma}_5 \tag{24}$$

From this, defining

$$D_t = I_t^* \begin{bmatrix} -\frac{3}{5} & -\frac{1}{5} & \frac{1}{5} & \frac{3}{5} & 1 & 0 \end{bmatrix}'$$
$$= -\frac{3}{5} I_t^*(1) - \frac{1}{5} I_t^*(2) + \frac{1}{5} I_t^*(3) + \frac{3}{5} I_t^*(4) + I_t^*(5) \quad (25)$$

we have $I_t^* \tilde{\gamma} = \tilde{\gamma}_5 D_t$, which shows that $\tilde{\gamma}_5 D_t$ is the constrained regression function. From (24), we obtain $\tilde{\gamma}_6 = 0$ (so Saturday is an average day, $\gamma_6 = \bar{\gamma}$); and from (10), $\tilde{\gamma}_7 = -\sum_{i=1}^{6} \tilde{\gamma}_1 = -\tilde{\gamma}_5$.

We now re-derive (25) without using (23) to illustrate the application of (21). The constraints (22) for β are equivalent to (19) for $\tilde{\beta}$, with

$$H = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 2 \end{bmatrix}$$

To obtain (21), we require an invertible matrix J constructed by adding a sixth row to HN, for example

$$J = \begin{bmatrix} 3 & 0 & 1 & 1 & 1 & 1 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(The nonzero value in the added sixth row could have been placed in any column but the sixth to achieve an invertible J). The most cumbersome step with (21) is the calculation of J^{-1}, but this can be done easily and exactly by various programs. (We used Scientific Workplace™). The result is

$$J^{-1} = \frac{1}{35} \begin{bmatrix} 12 & -3 & -10 & -9 & -4 & -21 \\ 9 & 24 & 10 & 2 & -3 & -7 \\ 6 & 16 & 30 & 13 & -2 & 7 \\ 3 & 8 & 15 & 24 & -1 & 21 \\ 0 & 0 & 0 & 0 & 0 & 35 \\ -10 & -15 & -15 & -10 & 15 & 0 \end{bmatrix}$$

Due to (20), $J\tilde{\gamma} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \tilde{\gamma}_5 \end{bmatrix}'$ Thus, using (21), (25) is obtained from the product of I_t^* with the sixth column of J^{-1}.

Software is available from the authors that generates an X-12-ARIMA input file of constrained regressor values (25) for end-of-month and general $\bar{w}$-th day of month stocks. These regressors have been implemented in the not yet released program X-13A-S discussed in Findley (2005) and Monsell (2007).

4. RegARIMA Modeling Considerations

4.1. *Reinterpretation With the Log Transformation*

In the next section, results are presented from applying the regressors I_t^* and D_t with $\bar{w} = 31$ to detect and estimate trading day effects in the inventory series of an important U.S. Census Bureau survey. These series, like most economic time series considered for seasonal adjustment, must be log transformed in order to achieve data that can be successfully modeled with a regARIMA model. When the log transformation is used, the stock trading day regression function (11) must be reinterpreted.

We are treating stock series Z_t, $t \geq 1$, as accumulations $Z_t = \sum_{j=1}^{t} Y_j$ of the values of a flow series Y_t, $t \geq 1$. In the log transformation situation, the day-of-week effect $\sum_{i=1}^{7} \tilde{\beta}_i X_t(i) = \sum_{i=1}^{6} \tilde{\beta}_i X_t^*(i)$ of log Y_t is estimated within a regARIMA model for this transformed series and then exponentiated to obtain day-of-week-effect factors $\exp\left(\sum_{i=1}^{7} \tilde{\beta}_i X_t(i)\right)$ for Y_t; see Bell and Hillmer (1983) and Subsections 1.4 and 3.3 of Findley et al. (1998). From our analysis in Sections 2.2 and 2.3 above, the expression $\sum_{k=1}^{7} \tilde{\gamma}_k I_t(k)$ with coefficients given by (16) and $\tilde{\gamma}_7 = -\sum_{i=1}^{6} \tilde{\gamma}_i$ now describes the level-neutral day-of-week effect of the series $\sum_{j=1}^{t} \log Y_j = \log\left(\prod_{j=1}^{t} Y_j\right)$, rather than that of $\log Z_t = \log\left(\sum_{j=1}^{t} Y_j\right)$. Therefore the estimation of $I_t^* \tilde{\gamma} = \sum_{k=1}^{7} \tilde{\gamma}_k I_t(k)$ within a regARIMA model component for log Z_t requires a new rationale. Observe that, with $k(t)$ denoting the index of the day of week on which the stock is measured in month t, we have $\sum_{k=1}^{7} \tilde{\gamma}_k I_t(k) = \tilde{\gamma}_{k(t)}$. Hence, estimation of this component in log Z_t provides stock-day effect factors for Z_t of the simple, intelligible form $\exp\left(\tilde{\gamma}_{k(t)}\right) \doteq 1 + \tilde{\gamma}_{k(t)}$. The motivation for using the constrained regressor D_t in a regARIMA model for log Z_t is not as clear. Nevertheless, it will be shown that D_t is an important alternative to I_t^* when the log transformation is used.

4.2. *Log Likelihood-Ratio Tests for the Model Comparisons*

The estimation of regARIMA models in X-12-ARIMA and TRAMO-SEATS is done by maximizing log-likelihood functions of Gaussian form. To test whether or not a stock day-of-week regressor should be included in the regARIMA model for a series and, if so, whether or not D_t is to be preferred over the unconstrained regressor I_t^*, we use tests based on differences of the calculated maximum Gaussian log-likelihood values. These likelihood-ratio (LR) tests do not require the modeled time series to be Gaussian. Specifically, given the maximum Gaussian log-likelihood value for an unconstrained stationary time series model and for a model nested within it having d fewer independent parameters, let ΔL denote the difference between the latter value and the former. For the null hypothesis that the nested model is correct, Taniguchi and Kakizawa (2000, p. 61) show, under general data assumptions specified in their Lemma 3.1.1 and Theorem 3.1.2 (for the appropriately differenced data in the case of ARIMA models), that the asymptotic distribution of $-2\Delta L$ is chi-square with d degrees of freedom,

$$-2\Delta L \sim \chi^2_d \tag{26}$$

Let L, L^D, and L^{I*} denote the respective maximum log-likelihood values of the model without day-of-week regressors, the model with D_t, and the model with I_t^*. Then, for $\Delta L = L - L^D$, respectively $\Delta L = L - L^{I^*}$, the null hypothesis of no day-of-week effect is tested using (26) with $d = 1$, respectively $d = 6$. If this null hypothesis is rejected by both tests, then $d = 5$ is used in (26) with $\Delta L = L^D - L^{I^*}$ to test the null hypothesis that the model with D_t is correct in preference to the model with I_t^*. In the study described next, these hypothesis tests were performed with significance level $\alpha = .05$.

5. The Empirical Study

5.1. The Series Considered

The testing procedure just described was applied to the 91 inventory series of the U.S. Census Bureau's monthly U.S. Manufacturers' Shipments, Inventories and Orders Survey (the M3 Survey) starting from the regARIMA models used in seasonal adjustment production in 2006. These production models have outlier regressors but no trading day regressors. The data used ended in October, 2006. The starting dates varied 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).

Table 1 shows the Industry categories indicated by the initial number and letter of the identification codes of the series to which direct reference is made in this section. The final two code letters are to be interpreted as follows: TI – Total Inventories; MI – Materials

Table 1. Some M3 Series Category Codes

11S	Food Products
11A	Grain and Oilseed Milling
21S	Wood Products
22S	Paper Products
22A	Pulp, Paper and Paperboard Mills
23S	Printing
26S	Plastics and Rubber Products
27S	Nonmetallic Mineral Products
31S	Primary Metals
31C	Ferrous Metal Foundries
33S	Machinery
34S	Computer and Electronic Products
34K	Electromedical, Measuring and Control Instrument Manufacturing
35A	Electric Light Equipment Manufacturing
36A	Automobile Manufacturing
36C	Heavy Duty Truck Manufacturing

and Supplies Inventories; WI – Work in Process Inventories; FI – Finished Goods Inventories. The latter three are components of the TI series of the same category.

Among the 91 series, there were 21 for which the starting model, with no trading day regressor, was rejected by a test in favor of an enhanced model with I_t^* or D_t. However, for one of these 21 series, we rejected the only alternative model accepted, with D_t, because an X-12-ARIMA warning message showed that its trading day adjustment led to "visually significant" (v.s.) trading day peaks in the autoregressive spectrum estimates of the last eight years of both the differenced log seasonally adjusted series and the irregular component of the seasonal decomposition. Such peaks did not occur when no trading day modeling and adjustment were done for this series. See Soukup and Findley (1999) and Section 6.1 of U.S. Census Bureau (2007) for background on the spectrum diagnostic and the v.s. criterion. An example spectrum plot is shown in the next subsection. For another series too, we rejected the only alternative model accepted, the model with D_t, based on forecast performance (see Section 5.2.1).

5.2. *Analysis of the Trading Day Regressor Selection*

Here is the breakdown of accepted trading day regressors among the remaining 19 series for which the model with no trading day regressor was rejected in favor of a model with I_t^* or D_t. For three of these series only the unconstrained regressor I_t^* was accepted and for 8 series only the constrained regressor D_t. This left eight series for which both trading day regressors were preferred over none. For these, D_t was always accepted in preference to I_t^* by the LR test with $\Delta L = L^D - L^{I^*}$.

However, for two of these last eight series, we preferred I_t^* over D_t because its use removed all v.s. trading day spectral peaks found when no trading day regressor was used, whereas use of D_t left a v.s. peak of reduced height in the spectrum of the regARIMA model residuals of one series (26SFI) and in the differenced log seasonal adjustments of the other (22SFI). Even with these reclassifications the new one-coefficient regressor is very successful: it is preferred for 14 of the 19 series for which modeling with either I_t^* or D_t was justified by the LR tests without strong contradiction from another diagnostic.

Further, as we detail next, for 13 of these 14 series, at least one other modeling diagnostic gave additional support to the use of D_t rather than no trading day regressor. We will subsequently present similar support for use of I_t^* with the five series for which it was preferred.

5.2.1. Further Results For D_t

Spectrum Diagnostics. With no trading day modeling, 6 of the 14 series for which D_t was chosen had v.s. trading day peaks in the spectrum estimates of the last eight years of regARIMA model residuals, the differenced log seasonally adjusted series, and/or the irregulars. Use of D_t eliminated all of the v.s. peaks for these series (11ATI, 22ATI, 23SMI, 27SFI, 31SFI, 36CTI). For example, with no trading day regressor in its model, all three spectra of the series 31SFI had strong v.s. peaks at the primary trading day frequency (.348 cycles/month) and the secondary trading day frequency (.432 cycles/month). This is illustrated in Figure 1, an overlay plot of the decibel spectrum graphs of two differenced log seasonally adjusted series from X-12-ARIMA, one obtained without trading day

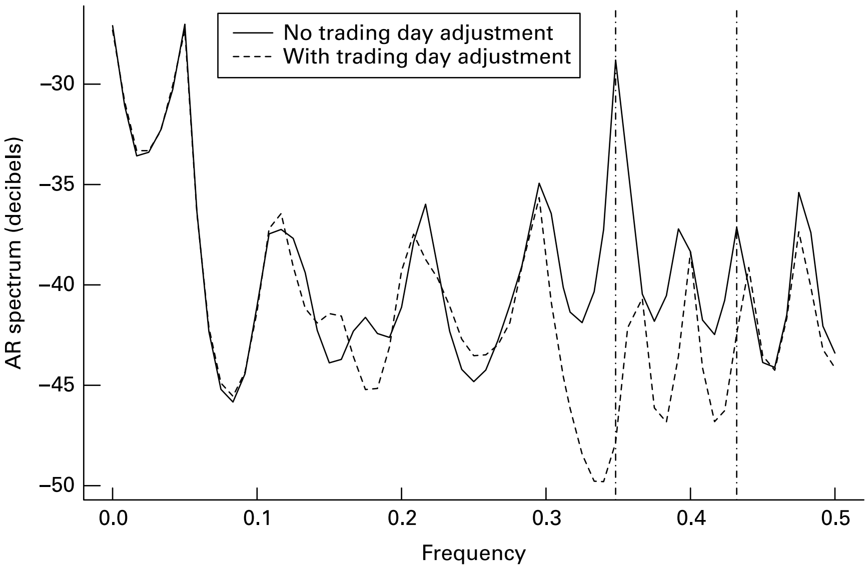

Fig. 1. Spectra of the differenced log seasonal adjustments obtained without and with the trading day adjustment provided by D_t. Use of D_t eliminates both v.s. trading day peaks

adjustment (two v.s. trading day peaks) and the other obtained with adjustment from the preferred trading day regressor D_t (no trading day peaks).

Goodness of Fit. For 8 of these 14 series, the models with no trading day regressor had residuals with one or more Ljung-Box goodness-of-fit Q statistics through lag 24 with a p-value less than .05. Table 3 shows the impact on the p-values of previously significant Q statistics of these series when D_t was added to the regARIMA model. With one exception, use of D_t always reduced the number of significant Qs – to zero for 5 of the 8 series including the three series with improved spectra as well as improved Qs (22ATI, 27FSI, 31FSI). The exceptional series (34SMI) had a single significant Q statistic, at lag 6, that remained significant when D_t was used.

Forecasting. For four of these 14 series, the model without D_t had no significant Q statistics and yielded no v.s. trading day spectral peaks. For the models of these four series with and without D_t, we compared the sample means of the squared errors of lead l out-of-sample forecasts for $l = 1, 12$ using forecast origins from January 2004, onward. To be precise, for a log-transformed time series Z_t, $1 \leq t \leq T$ and forecast lead $l \geq 1$ and forecast origin $1 \leq \tau \leq T$, let $Z_{\tau+l|\tau}$ denote the forecast of $Z_{\tau+l}$ obtained from Z_t, $1 \leq t \leq \tau$ via the regARIMA model without trading day regressor and with parameters estimated from Z_t, $1 \leq t \leq \tau$. With τ_0 denoting the index of the initial forecast origin (January 2004 in our case), define $MSE_l = (T - l - \tau_0 + 1)^{-1}\sum_{\tau=\tau_0}^{T-l}(Z_{\tau+l} - Z_{\tau+l|\tau})^2$. Let MSE_l^D and $MSE_l^{I^*}$ denote the corresponding sample means for the models with D_t and I_t^* respectively. Such sample means are available from X-13A-S. The first four rows of Table 2 give the values of

$$MSE_l / MSE_l^D, \; l = 1, \; 12 \tag{27}$$

Table 2. Values of MSE_l/MSE_l^D

Series	$l = 1$	$l = 12$
21SFI	1.0006	1.0056
26SMI	1.0425	1.0021
31SMI	1.0318	0.9981
35ATI	1.0858	1.0072
11SFI	0.8801	1.0141

for the four series from 33 forecasts with $l = 1$ and 22 with $l = 12$. With one negligible exception for $l = 12$, the values of (27) are all larger than one, showing that use of D_t generally led to at least slightly smaller sample mean squared errors for these four series at leads 1 and 12, thereby providing modest additional support for use of D_t. The measures (27) also led us to discover the second series eliminated from the initial 21 series obtained via the LR tests, 11SFI. For this series, the model with D_t had a substantially larger sample mean squared forecast error at lead 1 than the model without D_t (see the final row of Table 2). Neither model had a significant Q or gave rise to a v.s. peak.

For perspective, the values in Table 2 can be compared to the values of (27) and

$$MSE_l/MSE_l^{I^*}, \; l = 1, 12 \tag{28}$$

in Table 3 for the 15 series having either a v.s. peak or a significant Q or both when no trading day regressor was used. The comparison shows that, often for $l = 1$ and occasionally for $l = 12$, the benefit to forecasting performance of using a preferred trading

Table 3. p-Value data (NoTD/TD) and values of the ratio (27) or (28) for $l = 1, 12$ for the 15 series with one or more significant Qs or v.s. spectrum peaks

Series	TD choice	No. $p_k \le 0.05$	p_{24} if $\le .05$	$\min_k p_k$ if $\le .05$	$l = 1$	$l = 12$
11ATI	D_t	–/–	–/–	–/–	1.0061	0.9930
22ATI	D_t	1/–	–/–	.049/–	1.1717	1.0057
22SFI	I_t^*	5/–	.046/–	.030/–	1.4981	1.0970
23SMI	D_t	–/–	–/–	–/–	0.9621	1.0155
26SFI	I_T^*	–/–	–/–	–/–	1.1946	1.2588
27SFI	D_t	10/–	.049/–	.013/–	1.4967	1.0052
27SMI	D_t	7/6	.004/.009	.004/.009	1.0109	0.9947
31ATI	D_t	18/15	.000/.009	.000/.007	1.0124	1.0513
31CTI	I_T^*	7/–	.009/–	.009/–	1.1041	1.0574
31SFI	D_t	2/–	.018/–	.013/–	1.1937	1.0159
33SFI	D_t	2/1	–/–	.040/.044	1.0424	1.0007
34SMI	D_t	1/1	–/–	.046/.026	1.0550	0.9870
34KTI	I_T^*	–/–	–/–	–/–	1.1085	1.0032
36ATI	I_T^*	22/3	.018/–	.000/.012	1.0497	1.0281
36CTI	D_t	–/–	–/–	–/–	1.0442	0.9960

day regressor is more substantial when the model with no trading day regressors has significant Qs and/or gives rise to v.s. peaks.

Turning to the sizes of the trading day percent adjustment factors, $100 \exp(D_t \tilde{\gamma}_5)$, $1 \le t \le T$, among the 14 series for which D_t was preferred, the factors had the largest range, from 99.28 to 100.73, for the Materials and Supplies Inventories of Printing (23SMI), a series whose seasonal factors (from X-12-ARIMA output table D 10) ranged from 95.90 to 105.00. The trading day factors had the smallest range, from 99.78 to 100.22, for Total Inventories of Pulp, Paper, and Paperboard Mills (22ATI), whose seasonal factors ranged from 98.32 to 101.47.

5.3. Further Results for I_t^*

With the 5 among the 19 series for which the unconstrained regressor I_t^* was preferred, its use always reduced the number of Ljung-Box Qs with p-values less than .05 for the three series that had such Qs, 22SFI, 31CTI and 36ATI (see Table 3). It also reduced the number of v.s. spectral peaks at trading day frequencies among the spectra of the three series that had such peaks, 22SFI, 26SFI and 34KTI, eliminating the only such peak in the case of 26SFI. The largest range of percent adjustment factors $100 \exp\left(I_t^* \tilde{\gamma}\right)$, $1 \le t \le T$, from 98.60 to 101.76, occurred for Total Inventories of Automobiles (36ATI), whose seasonal factors ranged from 90.04 to 106.88. The smallest range, from 99.52 to 100.23, occurred for Finished Goods Inventories of Plastics and Rubber Products (26SFI), whose seasonal factors ranged from 97.44 to 102.38.

The fact that the latter series is one of the two series, among these five, for which the LR test preferred D_t over I_t^* raises the question whether series with D_t preferred by the test tend to have smaller trading day factor ranges than series with I_t^* preferred. Evidence against this hypothesis comes from the other series, 22SFI, whose range with I_t^*, from 99.54 to 100.51, was the second largest among the five series, and whose range with D_t is still larger, from 99.47 to 100.53. (The seasonal factor range of 26SFI becomes shorter with D_t, from 99.69 to 100.31).

In Table 3, p_k is the p-value of the Q statistic at lag k.

6. Additional Regressor Options for Untransformed Series

Although day-of-week effect regressors are the main focus of this article, for completeness we now mention other stock trading day regressor and constraint options that are applicable to series that do not require a transformation for RegARIMA modeling. These options have not been empirically evaluated.

6.1. Regressors Related to Month-Length

In the situation in which Z_t is not transformed, (6) suggests that, in addition to the level-neutral day-of-week effect, the accumulating month-lengths $\sum_{j=1}^{t} m_j$, or an appropriate component thereof, should be considered as an additional stock trading regressor. Bell (1995) considers decompositions of $\sum_{j=1}^{t} m_j$, revising those presented in Bell (1984). For the seasonal adjustment situation, in which seasonal, trend and level effects can be identified by the seasonal adjustment decomposition, only the level-neutral accumulating-leap-year component of $\sum_{j=1}^{t} m_j$, given as the third component of the decomposition (6) of Bell (1995), is a natural candidate regressor. However, this regressor is not implemented in X-12-ARIMA or in the forthcoming release of TRAMO-SEATS. In the case of X-12-ARIMA, the main reasons for its omission are the paucity of series that are not transformed and the lack of a complementary way of expressing and estimating leap-year-related effects in the log transformation situation.

6.2. *Regressors for General Linear Constraints*

We now consider the case in which the linear constraints on β are not contrasts. If the observed series is end-of-month stocks and if it can be well modeled without transformation, then the identity $\sum_{j=1}^{t}\sum_{i=1}^{7}\beta_i X_j(i) = \sum_{i=1}^{7}\beta_i\left\{\sum_{j=1}^{t} X_j(i)\right\}$ reveals that its trading day effects (6) can be estimated by estimating the flow trading coefficients β_i using as regressors $W_t(i) = \sum_{j=1}^{t} X_j(i)$, $1 \le i \le 7$, as Cleveland and Grupe (1983) observed to define their end-of-month stock model. Let the constraints on $\beta = [\beta_1 \beta_2 \ldots \beta_7]'$ be $G\beta = 0$, with G of full rank. If G is enlarged by addition of rows to become an invertible matrix K, then the decomposition $W_t\beta = (W_t K^{-1})K\beta$ with $W_t = [W_t(1) \ldots W_t(7)]$ can be used to obtain the constrained regression model by a procedure analogous to that illustrated for (21) in Section 3.1. But, as we just noted, Bell (1995) identifies components of (6) which should be included in the trend and seasonal components of the data instead of in the trading day component if seasonal adjustment is done, and it might not be possible to extract these unwanted components from $W_t K^{-1}$. This is not a concern if forecasting, not adjustment, is the goal.

7. Conclusions

We expect the results reported in this article for the inventory series of the U.S. Census Bureau's M3 Survey to be broadly typical. For any comparable set of macroeconomic stock time series, usually a substantial percentage will have statistically significant day-of-week effects. The use of a one-coefficient regressor like D_t of (25) is likely to significantly increase the number of series in which such effects are identified by the log likelihood ratio test of Section 4.2. As we have shown, test decisions made concerning the use of I_t^* or D_t can be usefully refined with the aid of goodness-of-fit and spectral diagnostics. Also, measures of out-of-sample forecast performance of the competing models like (27) and (28) above (or the more comprehensive graphical diagnostics presented in Sections 3 and 4 of Findley et al. 1998) can provide further refinement and insight.

8. References

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

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

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

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: U.S. Bureau of the Census. http://www.census.gov/ts/papers/Conference1983/ClevelandGrupe1983.pdf

Findley, D.F. (2005). Recent Developments and Directions in Seasonal Adjustment. Journal of Official Statistics, 21, 343–365.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., and Chen, B.C. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program. Journal of Business and Economic Statistics, 16, 127–177 (with Discussion).

Gómez, V. and Maravall, A. (1997, 2003). Programs TRAMO and SEATS: Instructions for the User. Working Paper 9628. Sevicio de Estudios, Banco de España. http://www.bde.es/servicio/software/tramo/manualdos.pdf 2003 Update http://www.bde.es/servicio/software/tramo/updatedos.pdf

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

Maravall, A. (2008). Personal Communication.

Monsell, B.C. (2007). The X-13A-S Seasonal Adjustment Program. Proceedings of the 2007 Federal Committee on Statistical Methodology Research Conference. http://www.fcsm.gov/07papers/Monsell.II-B.pdf

Noble, B. (1969). Applied Linear Algebra. Englewood Cliffs: Prentice Hall.

Silvey, S.D. (1975). Statistical Inference. London: Halstead Press.

Soukup, R.J. and Findley, D.F. (1999). On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects After Modeling or Adjustment. 1999 Proceedings of the Business and Economic Statistics Section. Alexandria: American Statistical Association, http://www.census.gov/ts/papers/rr9903s.pdf, 144–149.

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

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

Received July 2007
Revised June 2008