# DETECTION AND MODELING OF TRADING DAY EFFECTS

**Raymond J. Soukup and David F Findley, U.S. Census Bureau**
**David Findley, Statistical Research Division U.S. Census Bureau, Room 3000-4, Washington DC 20233-9100**
**david.f.findley@ccmail.census.gov**

We illustrate the roles of three types of diagnostics for detecting or confirming the presence of trading day effects and for selecting among competing trading day models. These are spectral analysis, AIC comparisons, and out-of-sample forecast-error diagnostics. We show merits and limitations of each in the context of a study of four alternative trading day models for fifty-eight Census Bureau time series. We consider both direct estimation from the observations and indirect estimation from the irregulars.

**Key Words: Spectral analysis, regARIMA models, model selection, forecasting, monthly time series**

## 1. INTRODUCTION

Monthly economic time series are often systematically influenced by effects specific to the seven days of the week and when or how often each of the seven days of the week occurs in the month. In flow series (e.g. monthly totals of daily sales), the presence of such an effect is revealed when the monthly values depend in a consistent way on which days of the week occur five times in the month. With retail grocery sales in the U.S., for example, the sales volume is smaller on Mondays, Tuesdays and Wednesdays than on days later in the week. Thus sales in March, say, will be relatively lower in a year in which March has an excess of early weekdays and higher when March has five Thursdays, Fridays and Saturdays. In addition, the average daily effect can give rise to a length-of-month effect and this effect is not completely absorbed into the seasonal component of the series because the length of February is not the same every year. The component of length-of-month effect that is not part of the seasonal component is called the leap year effect.

Recurring day-of-week effects and leap year effects in monthly (or quarterly) economic time series are called *trading day effects*. When these effects are of significant size and are not taken into account, they can significantly impede modeling and forecasting, see Hillmer (1982). Also, like seasonal effects, trading day effects can make it difficult to compare series values across months or to compare movements in one series with movements in other series. For this reason, national statistical offices like the U. S. Census Bureau seek to detect trading day effects in the important macroeconomic series they produce and to estimate and remove any detected effects during the seasonal adjustment process. Details on the construction of trading day models (for flow series) are given in Findley, Monsell, Bell, Otto and Chen (1998), based largely on Young (1965), Cleveland and Devlin (1980) and Bell (1984). We will give a brief overview of the main models used for *multiplicative* adjustments. We assign the indices $j = 1, \ldots, 7$ to Monday, …, Sunday, respectively, and define

$$D_{jt} = \text{number of } j\text{-days in month } t,$$

$$N_t = \sum_{j=1}^{7} D_{jt} = \text{length of month } t,$$

$$N_t^* = \frac{1}{4}(N_t + N_{t-12} + N_{t-24} + N_{t-36}),$$

and

$$LY_t = N_t - N_t^* \ .$$

Note that $N_t^* =$ 28.25 if month $t$ is a February and $N_t^* = N_t$ otherwise. Hence $LY_t$ = 0.75 if $t$ is a leap year February, $LY_t$ = –0.25 if $t$ is a non-leap year February, and $LY_t$ = 0 otherwise.

We will consider four types of flow trading day models for use in regression models with ARIMA errors (regARIMA models). Each can be fit to time series by the U.S. Census Bureau's X-12-ARIMA modeling and seasonal adjustment program (Findley, *et al.*, 1998). The first model, which we will call TD7 (Trading Day with 7 estimated coefficients) in this article, is given by

$$\log Y_t = \beta_o LY_t + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + Z_t . \tag{1}$$

Here $Z_t$ denotes a process with a user-specified ARIMA covariance structure that may have a regression mean function for effects other than trading day effects, such as outlier and holiday effects. Note that the TD7 model has a leap year coefficient and six regression coefficients related to day-of-week activity. The Sunday coefficient (that of $D_{7t}$) is given by $\beta_7 = -\sum_{i=1}^{6} \beta_j$ . (Thus the seven coefficients of the $D_{jt}$ sum to zero.) After estimation, the model (1) provides a day-of-week-effect factor $e^{\beta_j} \cong 1 + \beta_j$ for each $j$-day that occurs five times in month $t$, and, when month $t$ is a February, the factor $e^{\beta_0 LY_t} \cong 1 + \beta_0 LY_t$. The product of the factors is the estimated trading day effect.

In Findley, *et al.* (1998), it is shown that a theoretical model for trading day effects can be derived wherein the parameter $\beta_0$ in (1) is equal to $1/28.25 \cong$ 0.0354. This model, which we will call TD6, has only six estimated trading day coefficients,

$$\log Y_t = \frac{1}{28.25} LY_t + \sum_{j=1}^{6} \beta_j (D_{jt} - D_{7t}) + Z_t . \tag{2}$$

A more parsimonious model, first provided by TRAMO (Gomez and Maravall (1996)), has only one coefficient for the day-of-week effects. This model arises from constraining the Monday to Friday coefficients to have the same value, and also requiring the Saturday and Sunday coefficients to coincide. Since the seven coefficients sum to zero, these constraints yield a single weekday-weekend contrast regressor. With this regressor, one can estimate the coefficient of the leap year variable as in (1) or use the theoretical value given in (2). The resulting models, which we will call the TD2 and TD1 models respectively, are given by

$$\log(Y_t) = \beta_0 LY_t + \beta_{M-F} \left( \sum_{j=1}^{5} D_{jt} - \frac{5}{2} \sum_{j=6}^{7} D_{jt} \right) + Z_t$$

and

$$\log(Y_t) = \frac{1}{28.25} LY_t + \beta_{M-F} \left( \sum_{j=1}^{5} D_{jt} - \frac{5}{2} \sum_{j=6}^{7} D_{jt} \right) + Z_t . \tag{3}$$

We also consider the possibility of having no trading day or leap year regressors in the model, so that the model becomes $\log(Y_t) = Z_t$, which we will call the NOTD model. We will show that the application of an ill-fitting trading day model can lead to worse forecasting performance than the NOTD model.

When trading day effects are estimated from the irregular component of a preliminary seasonal adjustment, rather than directly from the observations $Y_t$, modified versions of these models are used, see Findley *et al.* (1998).

## 2. METHODS FOR DETECTION AND COMPARISON

We consider three methods for detecting or confirming the need for trading day modeling and for comparing the performance of trading day models: 1) analysis of spectra at frequencies associated with trading day effects, 2) comparison of sample-size modified Akaike Information Criterion values, and 3) comparison of historical forecasting performance in recent data via out-of-sample forecast error diagnostics.

### 2.1. Spectral analysis

For series of sufficient length, spectrum estimates can provide very sensitive diagnostics for detecting the presence of a periodic component in a time series. Because trading day effects in monthly series arise from periodic effects that repeat every seven days in the (unobserved) underlying daily data, one can expect spectrum analysis to be useful for detecting trading day effects (including residual trading day effects left or induced by inadequate models), see Cleveland and Devlin (1980) and McNulty and Huffman (1989).

X-12-ARIMA plots estimated spectra of two detrended versions of the seasonally adjusted series and, optionally, also of the residuals of the fitted regARIMA model (the in-sample one-step forecast errors). The program prints a message warning of a “visually significant” trading day peak whenever a spectral amplitude at a frequency associated with trading day effects is sufficiently higher than its immediate neighbors. The “visual significance” criterion is the same for all spectra. The trading day frequencies unitized are 0.348 and 0.432 cycles/month. For more details see Soukup and Findley (2000). Assuming seasonal adjustment is done, the default set of spectra analyzed consists of the spectrum of the *differenced*, log-transformed, seasonally adjusted series and the spectrum of the irregular series (i.e. the “final Henderson trend”-adjusted seasonally-adjusted series). If a fully satisfactory trading day model is used, there should be no significant peaks at any of the trading day frequencies of any of the three spectra, and therefore no warning messages from the default or model residual spectra.

For our study, models and diagnostics are compared for fifty-eight Census Bureau series, forty-seven of which are sales series from the Retail Trade Survey and eleven of which are construction series from the Building Permits Survey. The retail series end in February 1999 and the permits series end in January 2000. The starting dates used were the earliest dates that gave acceptable (or at least the best obtainable) regARIMA goodness of fit diagnostics, usually January of 1987 or 1991 for retail series (earlier in some cases) and various dates in the 1980’s for

| Source of Warnings | Model Used | | | |
|---|---|---|---|---|
| | TD7 | TD6 | TD1 | NOTD |
| Warnings from default and model residual spectra | 2 | 2 | 31 | 49 |
| Warning(s) only from default spectra | 7 | 6 | 4 | 4 |
| Warning only from model residual spectrum | 1 | 1 | 5 | 5 |
| Total | 10 | 9 | 40 | 58 |

Table 1: Numbers of series with visually significant spectral peaks in the default or residuals spectra.

the building permit series. The ARIMA model orders were determined from modeling each series with the TD6 regressors. Table 1 provides a breakdown by model choice of the numbers of series for which at least one significant trading day peak occurred, an indication of model inadequacy. For the NOTD model, a significant spectral peak occurred in at least one of the three spectra of all series. For the TD6 and TD7 models, there are only nine and ten series, respectively, with significant TD peaks. Of these, only two have peaks in both types of spectra, a very reliable indicator of a residual effect. For the others, it is possible that some peaks are spurious, i.e. not an indication of residual trading day effects, see Soukup and Findley (2000). Thus TD6 and TD7 are broadly adequate.

The spectral analysis clearly reveals the need for trading day modeling of most of these series, as well as the inadequacy of the weekday-weekend model TD1 for many (retail) series. However, it does not enable us to choose between two models that leave no peaks in any spectra or that both leave significant peaks. To make such choices, one can compare sample-size adjusted AIC's or the out-of-sample forecast error performance of the models.

### 2.2. AICC comparisons

Let $L_N$ denote the maximized log-likelihood of an estimated regARIMA model fit to $N$ observations (after all differencing and seasonal differencing operations have been performed). Let $p$ denote the number of parameters in the model that can be consistently estimated if $N$ increases without bound. (In practice, this can be interpreted to mean all parameters except coefficients of regressors for outlier effects, such as additive outliers, level shifts, and other interventions.) For this model, a regARIMA analogue of the sample-size adjusted version of Akaike's AIC derived by Sugiura (1978) and Hurvich and Tsay (1989) is defined by

$$AICC_N = -2L_N + 2p\left(\frac{N}{N-p-1}\right).$$

In analogy with Akaike's minimum AIC criterion, among regARIMA models with the same outlier regressors that are fit to the same data (after identical differencing operations), the model with the smallest AICC value is suggested to be the best model.

In our study, outliers were identified by the automatic procedure described in Appendix C of Findley *et al.* (1998), and our experience was that fewer outliers were found for NOTD models than for the other models. This is not surprising because the standard errors calculated for automatic outlier identification tend to be larger when no trading day model is used, with the result that outlier *t*-values are usually smaller in magnitude. To make possible AICC comparisons of the different trading day models, for each series, we combined the outliers obtained automatically for each model type into a superset of regressors and used this superset when obtaining AICC's for the different trading day models and NOTD.

Figure 1 provides AICC comparisons among NOTD, TD1, and TD6 for the 58 series. The AICC values are given relative to the values of AICC for the NOTD model. Therefore, negative values favor inclusion of trading day regressors in the model. TD6 is the preferred models much more often than TD1 or NOTD models for retail series (numbers 1 through 47) but the TD1 and TD2 models are usually favored for the building permit series (numbers 48 through 58 in Figure 1 – TD2 results are shown only for these series). The preference for TD1 and TD2 reflects the fact that permit-issuing offices are closed on Sundays and open on Saturdays for at most a limited number of hours. There are exceptional retail series (38, 39, 40 and 44) for which the TD1 model is preferred. For these series, Saturday and Sunday have rather similar coefficients that are different from the coefficients of the other weekdays. For most retail series, Saturday sales are greater than an average day's whereas Sunday sales are less. For series 7,8, and 55, it is questionable whether there are trading day effects in the series that the models can estimate, because the AICC differences from the NOTD model's AICC values are very small.

Figure 2 is plot of the AICC differences between TD7 and TD6. For most series, these differences have relatively small positive values close to 2.0, indicating that the estimated coefficient of $LY_t$ is statistically close to the theoretical value used by TD6. The six relatively large negative values rather strongly favor the TD7 model, i.e. estimating $\beta_0$. For the retail series 3, 17, 23, and 37, the largest estimate is .05 and the rest are smaller than 1/28.25. For the permits series 48 and 52, the estimates are about .11, indicating that the extra day in leap year Februaries, when $LY_t = .75$, increases the number of permits issued by the factor $e^{.11\times.75} \cong 1.0825$ on average. (Almost the same estimates of $\beta_0$ are obtained with TD2). Such an increase seems implausible without auxiliary evidence that, for these series, most of the February permit activity occurs late in the month. In the next section, we will examine whether estimating $\beta_0$ improves forecasts for these series.

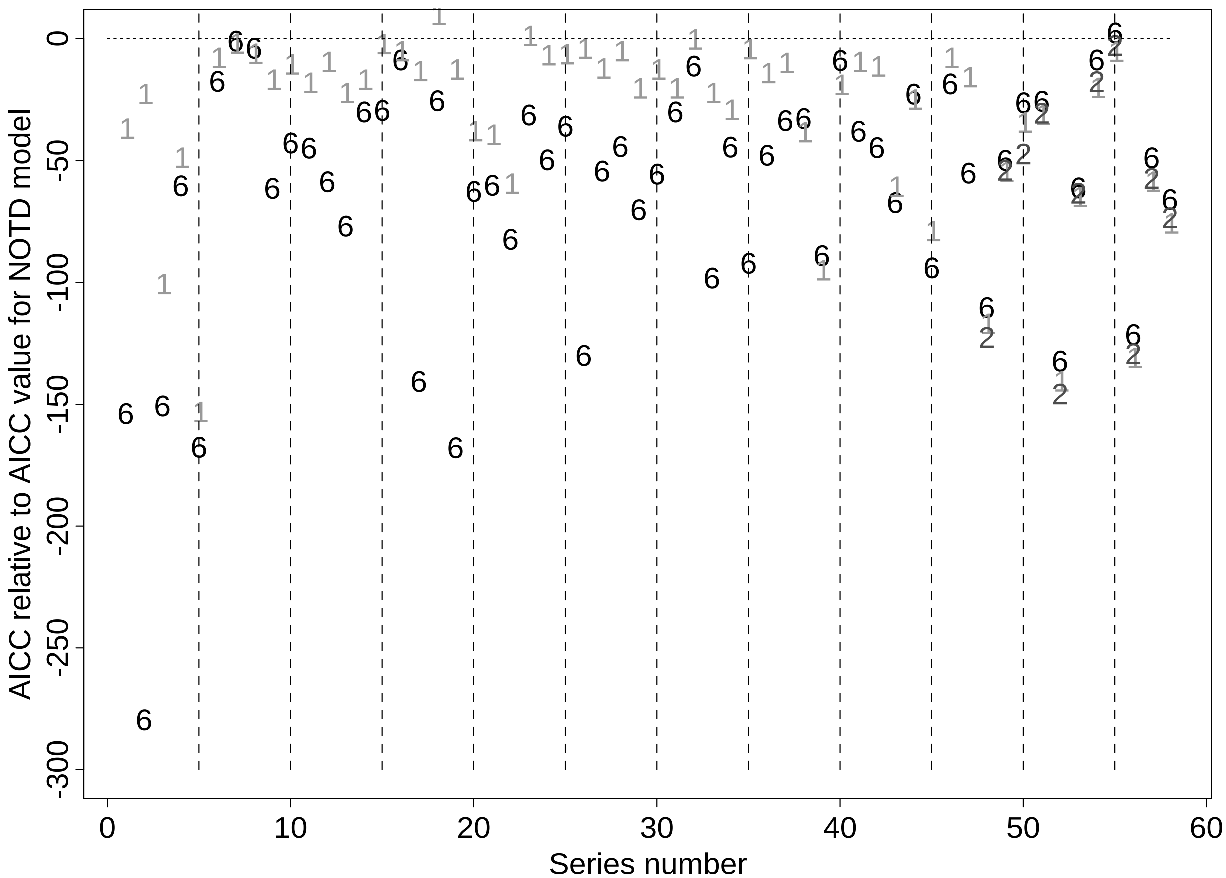


Fig. 1. AICC differences for TD6, TD1, and sometimes TD2 from NOTD for the series in the study, which are numbered on the x-axis from 1 to 58. Series 1-47 are retail sales series and series 48-58 are building permit series. For each series and model, the value obtained by subtracting the AICC of NOTD from of the model's AICC is plotted. Consequently, the ordinate value for the NOTD model appears as zero. For each series, the plotted numbers 6, 2 and 1 show the value of this AICC difference for the models TD6, TD2 and TD1 respectively. A lower (more negative) ordinate indicates a better model. For most series, NOTD is the least preferred model, followed by TD1 and TD2. The amount by which TD6 is preferred over these three models is often very significant.

### 2.3. Out of sample forecast errors

To obtain information about one model's *h*-step-ahead forecasting ability relative to that of another model, the later available time series data can be regarded as future data to be forecasted by each model from the earlier data and forecast errors can be calculated and compared. X-12-ARIMA supports a graphical squared error diagnostic detailed in Findley, *et al.* (1998) that we now utilize to compare the different trading day models for series in the study.

To describe the diagnostic, assume we are interested in $h$ step-ahead forecasting of $Y_t$ for $h \geq 1$. Suppose that a regARIMA model has been proposed for the transformed series $y_t = f(Y_t)$. Let $N_0$ be a number less than $N-h$

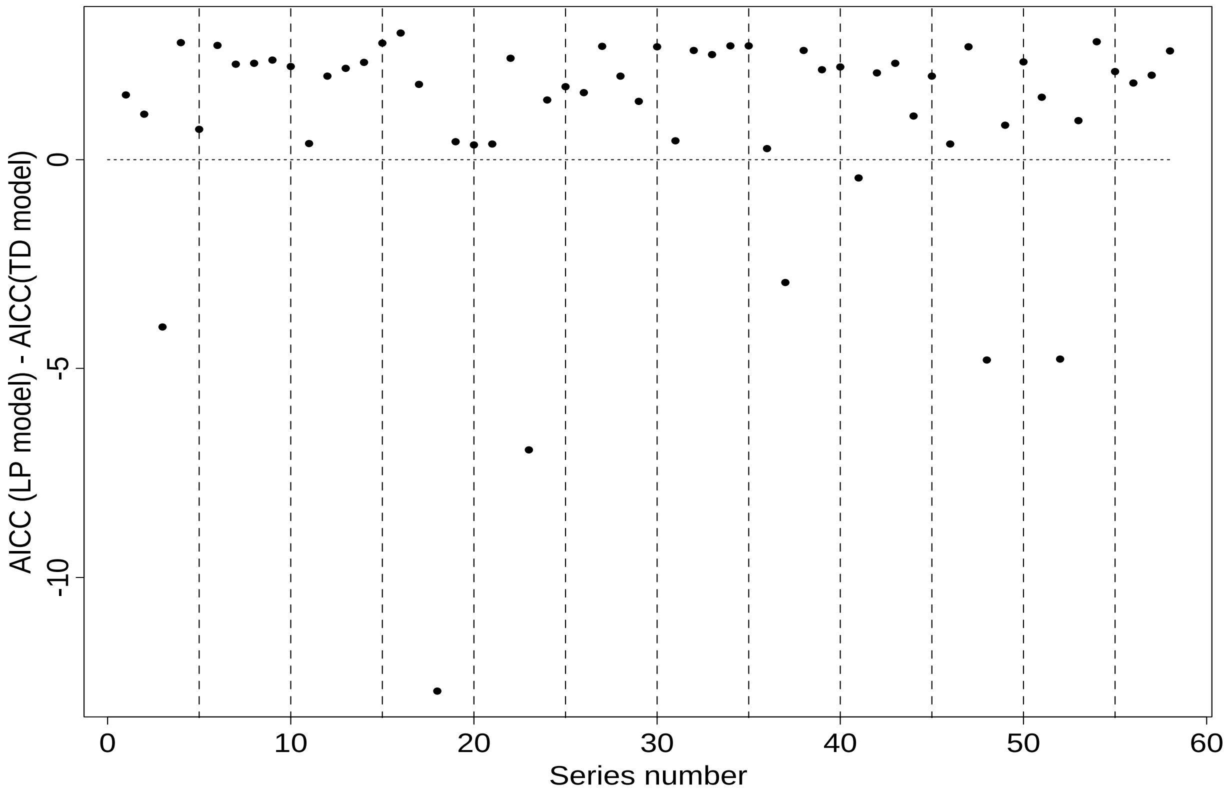


Fig. 2. AICC differences between TD7 (estimated leap year coefficient) and TD6 model (theoretical leap year parameter), plotted as a function of series number. Negative values favor TD7, while positive values favor TD6.

that is large enough that the data $y_t, 1 \le t \le N_0$ can be expected to yield reasonable estimates of the model's coefficients. (For the study, we used $N_0$ = 61.) For each $t$ in $N_0 \le t \le N-h$, let $y_{t+h|t}$ denote the forecast of $y_{t+h}$ obtained by estimating the regARIMA model using only the data $y_s, 1 \le s \le t$ and using this estimated model to forecast $h$ steps from time $t$. Then the *out-of-sample h*-step forecast of $Y_{t+h}$ is defined to be $Y_{t+h|t} = f^{-1}(y_{t+h|t})$. We define the associated *out-of-sample forecast error* (OSFE) by $e_{t+h|t} = Y_{t+h} - Y_{t+h|t}$ (with an appropriate modification of $Y_{t+h|t}$ if some outlier regressor is nonzero at time $t+h$). The main diagnostic calculated by the program is the sequence of accumulating sums of squared OSFE's,

$$SS_{h,M} = \sum_{t=N_0}^{M} e_{t+h|t}^2, M = N_0, \ldots, N-h.$$

Suppose there are two competing models, Model 1 and Model 2, with forecast errors $e_{t+h|t}^{(1)}$ and $e_{t+h|t}^{(2)}$ and with sums $SS_{h,M}^{(1)}$ and $SS_{h,M}^{(2)}$ respectively. Then we plot the normalized version of the differences $SS_{h,M}^{(1)} - SS_{h,M}^{(2)}$, defined by

$$SS_{h,M}^{1,2} = \frac{SS_{h,M}^{(1)} - SS_{h,M}^{(2)}}{SS_{h,N-h}^{(2)} / (N-h-N_0)}, \qquad (4)$$

as a function of $M$, for $N_0 \le M \le N-h$. The recursion formula

$$SS^{1,2}_{h,M+1} = SS^{1,2}_{h,M} + \frac{(e^{(1)}_{M+1+h|M+1})^2 - (e^{(2)}_{M+1+h|M+1})^2}{SS^{(2)}_{h,N-h} / (N-h-N_0)}$$

shows that over intervals of values of *M* where the graph of (4) goes up, the forecast performance of Model 2 is better (has smaller squared forecast errors). Where it goes down, Model 1 is better; and where it has no general direction, neither model's forecast performance dominates. The denominator in (4), the empirical mean squared forecast error of the second model, provides a scale for the interpretation of local and overall movements in each graph. An example of such a graph is given in Figure 3, where the models NOTD and TD6 are compared using one- and twelve-step-ahead OSFE's for series 28 of Figure 1, Retail Sales of Shoe Stores. In this case, NOTD is Model 1 and TD6 is Model 2, and the upward trend in OSFE's for both one- and twelve-step- ahead forecasts shows that the forecast performance of TD6 is persistently better.

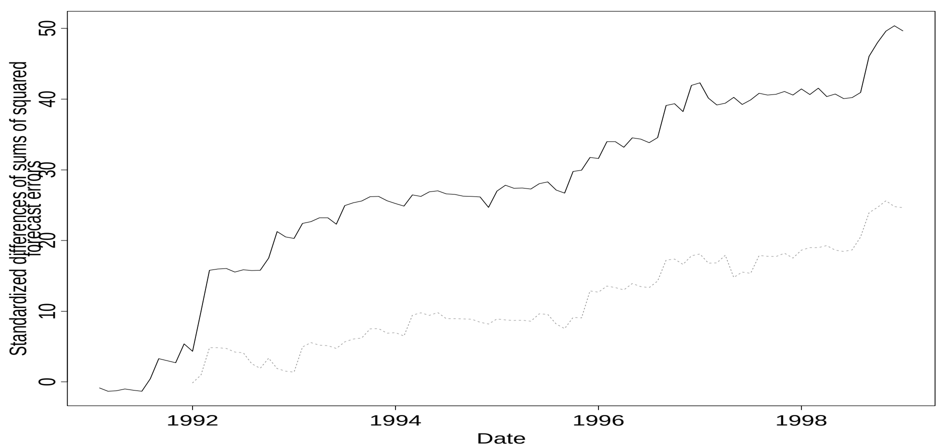


Fig. 3. An example of out-of-sample forecast error analysis (retail sales of shoe stores). The persistent upward trend in the curves related to lead 1 (solid line) and lead 12 (dashed line) forecasts indicates that the regARIMA model with TD6 forecasts persistently better than the NOTD model.

By contrast, Figure 4 shows that for series 8, Retail Sales by Mobile Home Dealers, the forecast performance of TD6 is persistently worse than NOTD. This is a series whose estimated trading day effects are unusually large: the factors range from 0.93 to 1.07. However, the fact that all but one of the (mostly large) trading day coefficient estimates have insignificant *t*-statistics suggests that the estimated factors may be of poor quality, and this suggestion is confirmed by the OSFE analysis. A probable explanation is lack of quality in the data: this series has high sampling error.

The OSFE diagnostic has the important virtue of not requiring the assumption that any of the models being compared is correct. However, it can be inconclusive in that it may suggest a preferred model only over too short a time interval to draw a conclusion, or it may suggest no preferred model at all. In addition, the preferred model can change with the forecast step *h*, in which case the modeler must have a preferred *h* or must label such results inconclusive as we have done in the present study.

Before comparing the performance of TD6, TD7 and TD1 with one another, we compared their forecasting performance against NOTD for $h = 1, 12$ in order to determine the subset of series for which at least one of the three trading day models had predictive ability. The results are shown in Table 2. For all OSFE comparisons involving TD7 (or TD2), the coefficient estimate used for the leap year regressor in forecast calculations is the


Fig. 4. An example of out-of-sample forecast error analysis (retail sales of mobile home dealers). The persistent downward trend in the curves related to lead 1 (solid line) and lead 12 (dashed line) forecasts indicates that the NOTD model forecasts persistently better than the TD6 model.

estimate obtained from the full data span, $1 \leq t \leq N$, because we found that estimation from shorter data spans sometimes yielded highly variable estimates of this coefficient.

*Series with NOTD as preferred model.* The series for which NOTD is preferred by OSFE analysis are series 7,8 and 55 in the AICC comparison plot, Figure 1, which shows that the series 7, 8 and 55 are also ones whose AICC values for trading day models are closest to the AICC of NOTD. Series 8 has already been discussed in connection with Figure 4, and the discussion of the factors given there also applies to series 7 and 55. With NOTD, there are output spectra of series 7,8 and 55 with rather large peaks at the main trading day frequency 0.348 cycles/month, suggesting rather strongly that there are trading day effects in these three series. Further, these peaks no longer appear when the minimum AICC trading day model is used. But for these series, the day-of-week coefficients of the trading day models are large with large standard errors, which may explain why the models are bad for forecasting. They are therefore of doubtful value for trading day adjustment, despite the spectral results.

| OSFE Model Preference | Number of Series | | |
|---|---|---|---|
| | All series | Retail Sales series | Building Permits series |
| TD6, TD7, or TD1 over NOTD | 44 | 35 | 9 |
| NOTD over TD6, TD7 or TD1 | 3 | 2 | 1 |
| Inconclusive | 11 | 10 | 1 |
| Total | 58 | 47 | 11 |

Table 2. Model preferences from out-of-sample forecast error graphs for the three models with trading day regressors compared with the NOTD model.

*Series with TD1, TD6 or TD7 as preferred models.* To determine the numbers of series for which either TD7, TD6 or TD1 has the greatest predictive power, we consider only the subset of forty-four series for which is evident from OSFE plots that one of these models is better than NOTD. The OSFE model preferences from this subset are shown in Table 3. As seen in Tables 2 and 3, the subset contains thirty-five of the forty-seven retail series and nine of the eleven building permits series.

| OSFE Model Preference | Number of Series | | |
|---|---|---|---|
| | All series | Retail Sales series only | Building Permits series only |
| TD1 | 4 | 2 | 2 |
| TD7, TD6 preferred over TD1 | 26 | 26 | 0 |
| No preference among TD7, TD6, TD1 | 14 | 7 | 7 |
| Total | 44 | 35 | 9 |

Table 3: Model preferences from out-of-sample forecast error graphs for the models TD7, TD6 and TD1. The analysis is performed for a subset of 44 series for which one of the three models TD7, TD6, and TD1 is clearly preferable to the NOTD model.

*Retail sales series.* With OSFE analyses of the thirty-five sales series for which a trading day model is favored over NOTD, the most common outcome is that TD7 and TD6 are preferred over TD1. The two retail sales series for which OSFE plots favor the TD1 model are series 39 and 40 in Figure 1 — two of the four series for which the TD1 model has the lowest AICC value. In no case is TD7 preferred over TD6 or vice versa, in contrast to the AICC results of Figure 2. For series where the estimation of the leap year coefficient in the TD7 model affects forecasting behavior in Februaries, it appears equally likely for those forecasts to be better or worse.

*Building permits series.* In the OSFE analysis of the nine permits series for which a trading day model is favored over NOTD, the TD1 model is favored for series 56 and, in equal measure with TD2, is also favored for series 57 over TD7 and TD6. There is no convincing favorite among the four models TD7, TD6, TD2 and TD1 for the remaining seven series. Thus, in particular, although AICC prefers estimating the leap year coefficient for series 48 and 52, doing so provides no advantage (or disadvantage) for forecasting.

An analysis (not presented here) of the absolute percent revisions statistics described in Section 2.2 of Findley *et al.* (1998), which compare concurrent adjustments of historical months with their most recent adjustments, offered some complementary support for the conclusions obtained from OSFE. For series where OSFE strongly favored trading day modeling, trading day adjustment often significantly reduced median revisions in the adjusted series. When there was no preference over NOTD, or even preference for NOTD, trading day adjustment often increased the median revisions by several tenths of a percent. Where there was no OSFE preference among the trading day models TD1, TD6, and TD7, none of these models had consistently better revisions performance than the other two.

## 3. COMPARISON WITH INDIRECT ESTIMATION FROM THE IRREGULAR COMPONENT

We also compared this approach, in which trading day effects are estimated from a regARIMA model for the directly observed data, with the alternative approach of indirect estimation from a regression model of the irregular component of a preliminary seasonal adjustment that is discussed in Findley *et al.* (1998). This alternative is the only approach available in older seasonal adjustment programs that do not have a regARIMA modeling capability. OSFE and revisions results were obtained for the forty-five series for which NOTD was rejected by OSFE for either direct or indirect estimation. According to the model preferences determined above, out-of-sample forecasts from either TD6 or TD1 were obtained, and their forecast errors were compared with those of the corresponding 6- or 1-parameter trading day models for the irregular component. Direct estimation led to persistently smaller squared forecast errors at leads one or twelve for thirty-two series. Indirect estimation was preferred for only one series, leaving twelve series without a preferred estimation method.

We also compared the maximum and median absolute revisions values from direct and indirect adjustments. Among the forty-five series, there were five for which the maximal revision of the indirect adjustment was larger by a percent or more than the maximal revision of the direct adjustment and none for which the reverse was true. There were only two series for which median revisions showed significant differences (0.5%) and, for both, the revisions for indirect estimation were larger. (These two were the only series among the forty-five for which AICC's calculated for the irregular component models favored NOTD.) The fact that the broad forecasting advantage of direct estimation noted in the preceding paragraphs corresponded only occasionally to advantages in the size of revisions may simply indicate that the revisions to the seasonal factors dominate. We considered one other aspect. In years in which the direct modeling provided better forecasts based on limited past data, the trading day coefficients of the indirect models often evolved in somewhat contradictory ways that would dissatisfy analysts. In such cases, the evolution of the direct model coefficients was generally more coherent, probably a consequence of their greater statistical efficiency due to the ARIMA modeling of the covariance structure of the regression error process.

## 4. CONCLUSIONS

We illustrated several approaches to detecting and modeling trading day effects. The diagnostics did not always point to the same conclusions, and we illustrated how inconsistencies can be addressed**.** ***Spectral analysis*** was used to obtain an overall picture of the adequacy of the trading day models. ***AICC comparisons*** provided a quick ranking of the performance of the models. Although spectral analysis suggested that trading day effects are present in all series and that the trading day models with six or seven coefficients can usually remove these effects***, out-of-sample forecast error*** analysis uncovered a few series for which all trading day models considered are so inadequate that it is better not to use them. For the ***retail sales series,*** AICC comparisons usually led to the selection of the six or seven coefficient model. For ***building permit series,*** there was always some support for the more parsimonious one-coefficient weekday-weekend contrast model. For both retail and permits series, in cases where AICC comparisons preferred the seven coefficient model (having an estimated leap year coefficient), neither out-of-sample forecast error analyses nor revision analyses demonstrated any advantage for this model, and with a few series, the value of the estimated leap year coefficient was implausibly large. Thus it did not seem necessary or desirable to use a model with an estimated leap year coefficient. Because of the almost consistently worse forecast results obtained from indirect estimation of trading day effects from the irregular component, the direct approach emphasized in this article appears more sound statistically, also in the sense of better providing the efficient parameter estimates assumed by t- and goodness of fit statistics of regression modeling. The X-12-ARIMA program used to obtain the diagnostics presented can be obtained from *http://www.census.gov/pub/ts/x12a/final/pc/omega.exe.*

*Disclaimer.* This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review by the Census Bureau than do official publications. This report is released to inform interested parties and to encourage discussion

## 5. REFERENCES


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

Cleveland, W. S. and S. J. Devlin (1980), "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods," *Journal of the American Statistical Association*, **75**, pp. 487-496.

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.

Gómez, V. and Maravall, A. (1996), "Programs TRAMO and SEATS, Instructions for the User (Beta Version: September 1996)," Banco de España – Servicio de Estudios, Documento de Trabajo no. 9628 (English version).

Hillmer, S. (1982). "Forecasting Time Series with Trading Day Variation," *Journal of Forecasting,* **1**, pp. 385-395.

Hurvich, C. M., and Tsai, C. L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika,* **76**, pp. 297-307.

McNulty, M. S. and Huffman, W. E. (1989), "The Sample Spectrum of Time Series with Trading Day Variation," *Economics Letters*, **31**, pp**.** 367-370.

Sugiura, N. (1978), "Further analysis of the data by Akaike's information criterion and the finite corrections," *Communications in Statistics* A, **7**, pp. 13-26.

Soukup, R. J. and Findley, D. F (2000), "Using the Spectrum to Automatically Detect Trading Day Effects after Modeling or Seasonal Adjustment," U.S. Census Bureau, Statistical Research Division (in preparation)

Young, A.H. (1965), "Estimating Trading Day Variation in Monthly Economic Time Series," Technical Paper 12, U. S. Department of Commerce, Bureau of the Census, Washington, D.C.