



PERGAMON

Transportation Research Part B 34 (2000) 451–469

---

---

TRANSPORTATION  
RESEARCH  
PART B

---

---

www.elsevier.com/locate/trb

# Adaptive estimation of daily demands with complex calendar effects for freight transportation

Gregory A. Godfrey \*, Warren B. Powell

*Department of Financial Engineering and Operations Research, School of Engineering/Appl. Science,  
Princeton University, Princeton, NJ 08544, USA*

Received 4 September 1997; received in revised form 7 May 1999; accepted 18 May 1999

---

## Abstract

We address the problem of forecasting spatial activities on a daily basis that are subject to the types of multiple, complex calendar effects that arise in many applications. Our problem is motivated by applications where we generally need to produce thousands, and frequently tens of thousands, of models, as arises in the prediction of daily origin–destination freight flows. Exponential smoothing-based models are the simplest to implement, but standard methods can handle only simple seasonal patterns. We propose a class of exponential smoothing-based methods that handle multiple calendar effects. These methods are much easier to implement and apply than more sophisticated ARIMA-based methods. We show that our techniques actually outperform ARIMA-based methods in terms of forecast error, indicating that our simplicity does not involve any loss in accuracy. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Forecasting; Time series; Freight demand; Exponential smoothing

---

## 1. Introduction

We address the problem of forecasting daily demands that arise in large freight transportation applications at a fairly high level of detail. Our most specific issue is the need for *daily* forecasts of freight demand, generally extended for one or two weeks into the future. In addition, we require models that include not only the attributes of origin and destination, but often several other characteristics as well, such as weight and commodity class. These forecasts may be used directly, or as inputs to other optimization models which use the demand forecasts to produce estimates of other activities. The optimization models require demand forecasts at the most detailed level, even

---

\* Corresponding author. Tel.: +1-609-258-5373; fax: +1-609-258-3796.  
*E-mail address:* powell@princeton.edu (W.B. Powell).

when there is a lot of uncertainty surrounding a disaggregate forecast. As a result, we generally face the need of estimating tens of thousands of models, which typically need to be updated every day.

A large less-than-truckload (LTL) motor carrier, for example, may serve 500 cities with potentially 500<sup>2</sup> markets. While in practice such demand matrices are rarely dense, the need to include other attributes can create model databases with as many as 100 000 models. While there is a lot of noise when estimating models at this level of disaggregation, the forecasts may be used in other models which require forecasts of shipments with a vector of attributes such as origin, destination, weight and type. For example, we have used these forecasts in network models which flow individual shipments through the network, and forecasting systems which project both inbound and outbound tonnages by aggregating more detailed origin–destination forecasts. These models need to forecast activities at a very high level of detail (in some cases, we treat these forecasts as random variables, though it is common to use a point estimate). In most cases, we will then calculate statistics from these models that are aggregated upward, such as the total flow over a link or through a node.

Our problem arises in applications where the only data available to estimate the models is historical activities. We can provide an initial estimate of the forecast parameters by training the model using history, and then update the parameters daily (or weekly or monthly) as new information becomes available. In some cases, such as freight flows for an individual shipper, there may be no history, since new shippers may be added to the problem dynamically. We assume that we do not have access to other exogenous economic parameters. For this reason, the method is unlikely to be effective in providing extended forecasts (for example, more than a few months into the future).

The need to estimate so many models limits our ability to use many standard estimation methods such as ARIMA, least-squares or other techniques that require working with a full history of data for each model. We are unable to store a complete history of data for so many models. In addition, we would like to re-estimate these models each day as new information arrives, since these demand patterns are non-stationary. For this reason, we prefer to work with exponential smoothing-based models that require only the actual observation from the previous period and the previous model to produce an updated model. In the paper, we compare our results against ARIMA models, where ARIMA methods serve as a benchmark. We argue that our methods are much simpler than ARIMA, and are even slightly more accurate, allowing us to conclude that the methods here are significantly easier to implement with no corresponding loss in accuracy (actually, there appears to be a slight improvement).

In practice, estimation methods based on exponential smoothing (see the excellent review by Gardner (1985)) are typically based on quarterly, monthly or possibly weekly data. The standard models assume that a single seasonal coefficient describes the periodic demand fluctuation. Daily forecasts, on the other hand, are better described by assuming that a superposition of multiple calendar effects govern the demand process. Some effects are strictly periodic and easy to define such as the day of week and month of year. Others are regular but harder to define such as the beginning and end of month and the end of quarter. Even more difficult are the holiday effects. Holidays may be fixed to a date (Christmas, Independence Day), fixed to a specific day of week (Labor Day, Memorial Day), or vary from year to year (Easter). Also, it may not be clear which days are affected by a holiday. For example, if Christmas falls on a Sunday, then the demand may

drop sharply during the three days before Christmas. On the other hand, if it falls on a Wednesday, then the three-day period from Tuesday to Thursday may have relatively low demand.

For testing purposes, we have real-world data sets consisting of 6 yr of Monday through Friday freight demand consisting of the total demand originating at 25 cities. We have chosen to do our testing at this level of aggregation since there will not be as much noise on a relative basis compared to the more disaggregate origin–destination data. This choice will simplify the comparison of competing techniques. The demands have the following general characteristics: (1) the baseline and trend may change slowly over time; (2) the daily trend effect  $T$  is small relative to the baseline  $S$ ; (3) larger short-term trends appear sporadically over time; and (4) seasonal effects may change the daily demand by as much as  $\pm 25\%$ . On a city-by-city basis, though, the demands exhibit a wide range patterns of trend, seasonality and noise.

We make the following contributions in this paper:

1. We formulate a forecasting model, called damped trend multi-calendar (DTMC) exponential smoothing, that incorporates multiple periodic and non-periodic calendar effects into the standard Brown single season exponential smoothing model. The model combines all the simple features of exponential smoothing, but provides for the complex calendar effects that arise in daily demand forecasting.
2. We present two sets of updating equations for the model, one based on standard exponential smoothing, and a second derived from stochastic optimization principles, which we believe to be completely new. The two methods are shown to be quite similar in form. The alternative derivation is provided in Appendix A.
3. We demonstrate experimentally, using real-world data sets, that the DTMC model provides better forecasts than both single exponential smoothing (about 2% lower average RMSE) and the best fitted seasonal autoregressive integrated moving average (ARIMA) model we could find (about 3% lower average RMSE).

In Section 2, we derive the DTMC model from the Brown single season exponential smoothing model. In Section 3, we document the experimental design details necessary for practical use of the forecasting system. In Section 4, we perform the experimental comparisons using the 25 real-world data sets described previously. Finally, in Section 5, we summarize the major test results and present ideas for future research. In Appendix A, we provide an alternative derivation of our method based on stochastic optimization techniques.

## 2. Exponential smoothing models

In this section, we introduce the Brown single season exponential smoothing model and transform it into one capable of incorporating multiple calendar effects. In addition, we present several properties of exponential smoothing models.

We start with the basic model:

$$X_t = [S_t + T_t \cdot t]I_t + \epsilon_t, \quad (1)$$

where  $t$  corresponds to an individual day. Following standard notation in the exponential literature (Gardner (1985))  $S_t$  is the base component,  $T_t$  is the trend term (which provides an estimate

of the slope of the data),  $I_t$  is the seasonal multiplier on day  $t$ , and  $\epsilon_t$  is a random error component with mean 0 and variance  $\sigma^2$ . At a point in time  $t_0$ , we might have a model:

$$F_{t_0,t} = [S_{t_0} + T_{t_0}(t - t_0)]I_{t_0,t}, \quad (2)$$

where  $F_{t_0,t}$  is the forecast for time  $t > t_0$  made at time  $t_0$ , and  $S_{t_0}$ ,  $T_{t_0}$  and  $I_{t_0,t}$  are estimates of the baseline, trend and seasonal effects as of time  $t_0$ . To understand the role of the baseline and trend, note from Eq. (2) that  $S_{t_0}$  represents the estimate of the intercept of the forecast at  $t = t_0$  and  $T_{t_0}$  is the estimate of the slope (or trend).

Exponential smoothing techniques feature intuitive, easy to implement models with low memory and computing requirements. In spite of this simplicity, the accuracy of short-range forecasts using exponential smoothing has been shown to be similar to that of more complex, resource intensive forecasting systems (Gardner, 1985; Hibon, 1984; Makridakis and Hibon, 1979). Exponential smoothing also provides optimal forecasts for specific demand processes (Harrison, 1967; Gardner, 1985; Montgomery et al., 1990; Muth, 1960; Nerlove and Wage, 1964; Theil and Wage, 1964).

Exponential smoothing methods use a weighted average of the most recent demand and previous forecast. Let  $X_t$  and  $S_t$  be the actual and forecasted demand for time  $t$ . Assume for this example that the demand equals a constant term plus noise. Using the simplest form of exponential smoothing, future forecasts are computed by

$$S_t = \alpha X_{t-1} + (1 - \alpha)S_{t-1}, \quad (3)$$

where  $\alpha$  is a smoothing parameter that determines the weight placed on the most recent demand. The exponential nature follows by taking the infinite expansion of  $S_t$ :

$$S_t = \sum_{i=0}^{\infty} \alpha(1 - \alpha)^i X_{t-i-1}. \quad (4)$$

The weights in Eq. (4) decrease exponentially and sum to one making the sum an exponentially weighted average. Of course, in practice, an infinite demand history is not available, so an initial forecast estimate is used to minimize the truncation error.

We introduce notation for our demand model (1) with baseline, trend, seasonal and noise components. We adopt the notation and model formulations suggested by Gardner (1985) in his review of exponential smoothing research through the early 1980s:

$\alpha$	smoothing parameter for the baseline
$\delta$	smoothing parameter for the seasonal/calendar factors
$\phi$	multiplicative (dampening) factor for trend
$S_t$	smoothed baseline estimate, computed after $X_t$ is observed
$T_t$	smoothed trend estimate at the end of period $t$
$I_t$	smoothed seasonal factor estimate at the end of period $t$
$X_t$	observed value of the time series in period $t$
$m$	number of periods ahead to forecast
$\hat{X}_t(m)$	forecast for $m$ periods ahead from period $t$
$e_t(m)$	$m$ -step ahead forecast error, $e_t(m) = X_t - \hat{X}_{t-m}(m)$

For our experiments, we use only the one-step ahead forecasts, so we denote  $\hat{X}_t(1)$  by  $\hat{X}_t$  and  $e_t(1)$  by  $e_t$ . The methods, though, appear in their general form.

To the single season forecasting equation, we shall add a dampening multiplier,  $\phi$ , to the trend as developed by Gardner and McKenzie (1985) to discount the effect of the trend in the future. Damped trend models produce significantly better long-range forecasts than linear trends that tend to overshoot the future (Gardner and McKenzie, 1989). There is a side benefit for our one-step ahead forecasts in that setting  $\phi = 0$  allows us to effectively “turn-off” the trend for data sets that do not exhibit a trend.

Gardner’s model 7-4 (Gardner, 1985) has a model with damped trend and multiplicative season for one season with period  $p$ . The equation to forecast  $m$  periods ahead from period  $t$  is

$$\hat{X}_t(m) = \left( S_t + \sum_{i=1}^{(m)} \phi^i T_t \right) I_{t-p+(m \bmod p)} \tag{5}$$

The seasonal multipliers represent elements in a cycle such as months of the year ( $I_{t-11}$  to  $I_t$ ). The subscript  $t - p + (m \bmod p)$  refers to the forecasted month.

Having specified a forecasting equation, we need equations to update the baseline, trend and seasonal terms. There are two popular exponential smoothing updating schemes that include trend and seasonal terms. The Holt et al. (1960) and Winters (1960) model uses separate smoothing parameters for the baseline and trend. The Brown (1963) model uses only one parameter for both terms and is a subset of the Holt–Winters model. Holt–Winters has greater flexibility and slightly better accuracy (Gardner and Dannenbring, 1980; Hibon, 1984; Makridakis and Hibon, 1979) due to its two smoothing parameters, but the Brown model reduces the search for the proper smoothing parameters. The Brown model is also critically damped meaning that it responds quickly to changes in the demand without overshooting (McClain, 1974; McClain and Thomas, 1973).

Since the calendar effects and trend multiplier add two more parameters to be determined, we choose to use the Brown model. The updating equations for the baseline, trend and seasonal multipliers, in this case, are (from model 7-4 in Gardner (1985)):

$$S_t = S_{t-1} + \phi T_{t-1} + \alpha(2 - \alpha)e_t/I_{t-p}, \tag{6}$$

$$T_t = \phi T_{t-1} + \alpha(\alpha - \phi + 1)e_t/I_{t-p}, \tag{7}$$

$$I_t = I_{t-p} + \delta[1 - \alpha(2 - \alpha)]e_t/S_t. \tag{8}$$

To illustrate, let  $\phi = 1$  and let  $\beta = \alpha(2 - \alpha)$  and  $\epsilon_t = (X_t - \hat{X}_t)$  (that is, the error is the actual observation minus the forecast). In this case, Eq. (6) reduces to:

$$\begin{aligned} S_t &= [\hat{X}_t + \beta(X_t - \hat{X}_t)]/I_{t-p} \\ &= [(1 - \beta)\hat{X}_t + \beta X_t]/I_{t-p}, \end{aligned} \tag{9}$$

which is to say that our estimate of the baseline,  $S_t$ , as in Eq. (3), is a weighted combination of the old forecast for time  $t$  and the observed demand for time  $t$ . The presence of  $I_{t-p}$  in the denominator simply deseasonalizes the data.

Notice that  $I_t$  replaces  $I_{t-p}$  to continue the sliding window of  $p$  seasonal multipliers. The original Brown equations relate the new estimates in terms of the old estimates and the new demands.

Accordingly, some of the complexity in the smoothing terms above can be attributed to using the error-correcting form of the updating equations.

Next, we extend the Brown model defined by Eqs. (5)–(8) to include multiple calendar effects. We start by introducing additional notation:

- $J$  set of calendar attributes representing weekdays, months, etc.
- $y_j(t)$   $\begin{cases} 1 & \text{if attribute } j \in J \text{ is active in period } t, \\ 0 & \text{otherwise} \end{cases}$
- $I_t^m(j)$  calendar factor for attribute  $j \in J$  for period  $t + m$  as of period  $t$ ,  
 $\exp(\alpha_{j,t} \cdot y_j(t + m))$
- $I_t^m$  calendar factor across all attributes for period  $t + m$  as of period  $t$ ,  
 $\prod_{j \in J} I_t^m(j) = \exp(\sum_{j \in J} \alpha_{j,t} \cdot y_j(t + m))$
- $\alpha_{j,t}$  coefficient used to calculate calendar factor  $I_t^m(j)$
- $\lambda$  smoothing parameter for the residuals

Again, we denote  $I_t^1(j)$  by  $I_t(j)$  and  $I_t^1$  by  $I_t$  for simplicity. The primary change to the updating equations is in the seasonal multipliers. In the Brown model, the multiplier in the forecasting equation corresponds to one element in the cycle. In the new formulation, however, all multipliers appear in the forecast equation as  $\prod_{j \in J} I_t^m(j)$ . If the demand contains day of week and month of year effects, then there are 19 multipliers, 7 for the weekdays and 12 for the months. For each forecast, one weekday and one month is active, so two of the  $y_j(t + m)$  will be one and the rest zero. For example, a Tuesday in August would have  $y_3(t + m) = 1$  and  $y_{15}(t + m) = 1$ . In this case, the aggregate calendar factor is

$$I_t^m = \prod_{1 \leq j \leq 19} I_t^m(j) = I_t^m(3) \cdot I_t^m(15) = \exp(\alpha_{3,t}) \exp(\alpha_{15,t}) = \exp(\alpha_{3,t} + \alpha_{15,t}).$$

In this manner, we can easily add as many calendar attributes as desired (even non-periodic ones such as holidays) without further complicating the model.

The forecasting equation for the DTMC model is

$$\hat{X}_t(m) = \left( S_t + \sum_{i=1}^m \phi^i T_t \right) \prod_{j \in J} I_t^m(j). \tag{10}$$

We use  $I_t^m$  and  $\prod_{j \in J} I_t^m(j)$  interchangeably in the equations that follow. The updating equations for the baseline and trend are exactly the same as Eqs. (6) and (7) except the calendar subscript changes to  $t - 1$ .

The updating equation for the calendar multipliers changes significantly. Instead of updating  $I_t$ , we update  $a_{j,t}$  for each  $j \in J = \{j \in J : y_j(t + m) = 1\}$  where  $J$  is the set of *active* attributes at period  $t + m$ . We do not update the non-active attributes because they are not used in the forecast.

Starting with the updating equation,

$$I_t^0 = I_{t-1}^1 + \delta[1 - \alpha(2 - \alpha)]e_t/S_t, \tag{11}$$

let  $\Delta I_{t-1}$  be the correction term so that  $I_t^0 = I_{t-1}^1 + \Delta I_{t-1}$ . We shall split the correction equally among the active attributes in a multiplicative way. That is, if the aggregate calendar factor should increase by 20% and there are three active attributes, then each of the three active calendar factors

increases by 6.3% ( $1.063^3 = 1.20$ ). Proceeding in this manner, we take the logarithm of each side, factor out  $I_{t-1}^1$ , and split the right side logarithm to get

$$\log(I_t^0) = \log(I_{t-1}^1) + \log\left[1 + \frac{\Delta I_{t-1}}{I_{t-1}^1}\right]. \tag{12}$$

Next, we split the aggregate calendar factors into individual calendar factors for each active attribute and use the fact that the log of a product equals the sum of the logs:

$$\log\left(\prod_{j \in \hat{J}} I_t^0(j)\right) = \log\left(\prod_{j \in \hat{J}} I_{t-1}^1(j)\right) + \log\left[1 + \frac{\Delta I_{t-1}}{I_{t-1}^1}\right], \tag{13}$$

$$\sum_{j \in \hat{J}} \log(I_t^0(j)) = \sum_{j \in \hat{J}} \log(I_{t-1}^1(j)) + \log\left[1 + \frac{\Delta I_{t-1}}{I_{t-1}^1}\right], \tag{14}$$

$$\sum_{j \in \hat{J}} \alpha_{j,t} = \sum_{j \in \hat{J}} \alpha_{j,t-1} + \log\left[1 + \frac{\Delta I_{t-1}}{I_{t-1}^1}\right]. \tag{15}$$

Splitting the correction evenly across the active attributes is now simple. Let  $|\hat{J}|$  be the number of elements in the active attribute set. To update  $a_{j,t}$  for each  $j \in \hat{J}$ , divide the correction term in Eq. (15) by  $|\hat{J}|$  and add it to  $a_{j,t-1}$ . Explicitly, the updating equation is

$$a_{j,t} = a_{j,t-1} + \frac{1}{|\hat{J}|} \cdot \log\left[1 + \delta[1 - \alpha(2 - \alpha)]\left(\frac{e_t}{S_t I_{t-1}^1}\right)\right] \quad \text{for all } j \in \hat{J}. \tag{16}$$

The DTMC model, then, consists of the forecasting Eq. (10) and updating Eqs. (6), (7) and 16. After each update, the calendar coefficients,  $a_{j,t}$ , are normalized for each attribute class (day of week, month of year) to sum to zero. For deseasonalized data, the undamped version ( $\phi = 1$ ) of this model has all of the optimality properties of the Brown model including the rapid response to demand changes without overshooting. The damping multiplier improves long-range forecasting as cited previously and provides a way to turn down the trend part of the model if the demand shows no trend. Finally, the multiple calendar effects allow non-periodic “seasons” such as beginning and end of month and holidays that ARIMA and other fixed period models cannot provide.

### 3. Experimental design

The fundamental research question we want to answer is whether the DTMC model provides better forecasts than a standard procedure such as ARIMA. However, to answer this question, there are specific implementation details that we must first resolve. In particular, four questions drive the experimental design. First, how do we detect changes in the demand process? We evaluate three control chart methodologies to solve this problem. Second, how do we initialize the coefficients of the forecasting model? Third, how do we select the proper smoothing parameters? Finally, how do we use fitted ARIMA models to provide a comparison with our forecasting system? We dedicate the remainder of this section to answer these questions.

We gauge the accuracy of any forecasting technique using the mean-squared error (MSE) of the one-step ahead forecasts. We have chosen to use one-step ahead measurements for three reasons. First, given the nature of our smoothing, forecasts of demands several days in the future will not differ significantly from the one-step ahead forecast, and hence we do not expect a measurable difference in the quality of the results if we compare results within short horizons. Second, given the nature of our daily update, it was extremely efficient to get the one-step ahead error, while it is quite difficult to estimate errors more than one-day in the future. Finally, forecasts more than one or two weeks into the future will introduce errors that reflect that we are using pure history, and do not have access to any other economic data. Our forecasting methodology, when applied to long-term forecasts, will suffer the same weaknesses as any time-series technique which lacks exogenous information.

### 3.1. Control charts

Control charts monitor forecast bias and signal the forecasting models to adjust when errors fall outside the control limits. We evaluate three different tracking signals, the Trigg, exponentially weighted moving average (EWMA), and Shewhart signals in our experiments. In all cases, the tracking signal must respond quickly to permanent changes in the base demand level, but not over-react to temporary demand “spikes”. McClain and Thomas (1973) discuss this *response–variance trade-off* in greater detail.

The equation for the Trigg tracking signal (Trigg, 1964; Trigg and Leach, 1967) is

$$\text{TRIGG} = \frac{\text{Smoothed Forecast Error}}{\text{Smoothed Absolute Error}}, \quad (17)$$

where  $\lambda$  is the smoothing parameter for both error statistics. The signal falls between  $-1$  and  $+1$ , approaching the extremes only when the errors consistently have the same sign, meaning that the forecasts are consistently over or under the actual demand.

Trigg (1964) analyzes the distribution of the tracking signal to determine appropriate control limits. Assuming that the errors have no serial correlation, zero mean, and standard deviation  $\sigma$ , Trigg finds that the standard deviation of the tracking signal is

$$\sigma_{\text{TRIGG}}(\lambda) \approx 1.2 \sqrt{\frac{\lambda}{2 - \lambda}}. \quad (18)$$

However, smoothing both the numerator and denominator introduces correlations that strengthen as  $\lambda$  increases and distort the signal distribution. For this reason, we use simulation to estimate confidence limits under Gaussian errors (see Table 1). The three probabilities correspond

Table 1  
Estimated confidence limits for the Trigg tracking signal via simulation

Prob	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.20$	$\lambda = 0.30$
0.68	0.198	0.289	0.428	0.504
0.95	0.382	0.523	0.720	0.668
0.997	0.528	0.718	0.896	0.764

to the  $\sigma$ ,  $2\sigma$  and  $3\sigma$  limits for a Gaussian distribution. For example, if  $\lambda = 0.1$ , then approximately 95% of the time, the tracking signal will fall within  $(\pm 1.9)(1.2)(0.23) = \pm 0.523$  under Gaussian error conditions.

Next, we consider the exponentially weighted moving average (EWMA) control chart. The EWMA tracking signal, formulated by Roberts (1959), is computed by

$$\text{EWMA} = \frac{\text{Smoothed Forecast Error}}{\text{Estimate of Error Standard Deviation}}. \quad (19)$$

The smoothing parameter in the numerator is  $\lambda$ . Hunter (1986) and Lucas and Saccucci (1990) both describe properties of the EWMA tracking signal. The standard deviation estimate,  $\hat{\sigma}$ , that we use is the square root of a smoothed sum of squared errors with a small, fixed smoothing parameter such as 0.05. Using  $\lambda$  to smooth  $\hat{\sigma}$  introduces correlation difficulties, but a value of 0.05 reduces correlation yet still responds to changes in the error magnitude. Assuming that  $\hat{\sigma}$  is a reasonable estimate of  $\sigma$ , the standard deviation of the EWMA tracking signal is

$$\sigma_{\text{EWMA}}(\lambda) = \sqrt{\frac{\lambda}{2 - \lambda}}. \quad (20)$$

The third chart, the Shewhart, is a special case of the EWMA chart. Hunter (1986) notes that the EWMA has two limiting cases. When  $\lambda = 1.0$ , the signal equals the forecast error divided by the standard deviation estimate. This is known as a Shewhart control chart. Correlations due to the smoothed standard deviation estimate make traditional  $3\sigma$  limits for the forecast error correspond to Shewhart limits of around 2.6. As  $\lambda$  tends to zero, the EWMA signal becomes a cumulative sum (CUSUM) signal. Hunter (1986) and Johnson and Leone (1977) discuss limitations that force the CUSUM implementation to be more difficult than the other signals. For this reason, we do not consider the CUSUM chart.

When a tracking signal falls outside the control limits, our forecasting system needs to adapt automatically to correct the problem. For applications such as ours with a large number of models, the best action is to increase the baseline smoothing coefficient to adjust quickly to the higher or lower demand level. Different applications of this idea appear in Flowers (1980), Shor (1985), Trigg and Leach (1967) and Whybark (1973). We adopt a variation of the Whybark technique. When a control limit is broken, we set the baseline smoothing coefficient,  $\alpha$ , to a high value such as 0.35 for the next period and the trend multiplier and calendar smoothing coefficient to zero to protect the trend and calendar coefficients while the baseline adjusts itself. The smoothing coefficients are reset when the tracking signal crosses a lower threshold, usually equal to the original control limit. This allows the baseline to adjust until the errors are under control (see Fig. 1).

### 3.2. Model initialization

Next, we consider model initialization. The model we use for the experiments includes four calendar effects: day of week, week of month, month of year, and end of quarter. The week of month is broken down into three parts: the first and last seven days of the month and the days in between. This choice was based on a visual analysis of the data. The pattern arises because of a tendency of shippers to push more freight out in the last week, which in turn produces a light

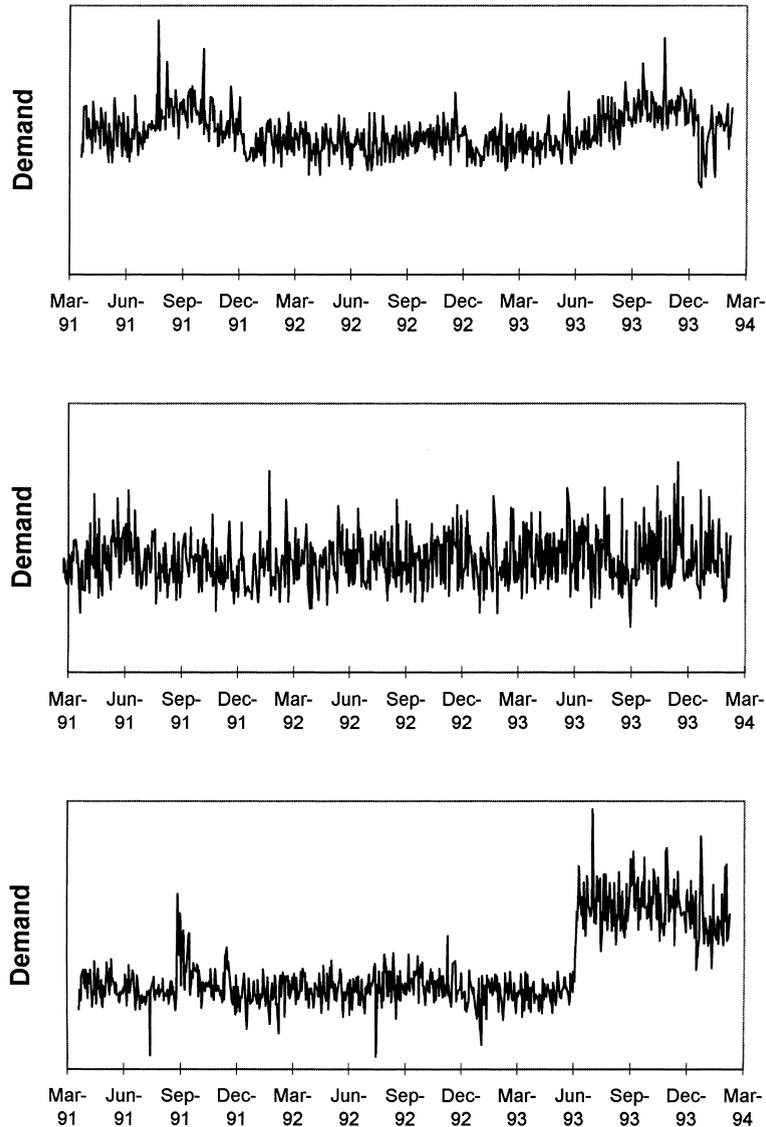


Fig. 1. The effect of a control chart for accelerating the response to a systematic increase in demand.

period in the first week of the next month. The middle two weeks appeared to be relatively stable. End of quarter is defined to be the last seven days of March, June, September and December. There are two attributes for the end of quarter effect, either the day is in the end of quarter or it is not. Each of the four calendar effects, then, has exactly one attribute active for any given day. The calendar effects are normalized separately so that the multipliers average one for each calendar effect set.

Traditionally, seasonal data sets are deseasonalized using a heuristic (Montgomery et al., 1990) or decomposition (Makridakis and Wheelwright, 1978), and then the baseline and trend are

initialized using the deseasonalized data. There are two reasons why we do not follow tradition. First, the calendar effects we model are not easily deseasonalized because they are not necessarily periodic. Second, having thousands of models to initialize makes more sophisticated techniques impractical. Instead, we employ backcasting (Ledolter and Abraham, 1984) and multiple passes through the training portion of the data set to perform the initialization or model fitting.

Each of the 25 real-world locations provides a variety of freight demand characteristics. Fig. 2 presents three cities with different demand attributes. These demand variations allow us to examine the quality of our forecasts under different conditions. We split the 25 demand sets into two parts: the training set (from April 1988 to March 1991) and the testing set (from April 1991 to February 1994). We start by forecasting forward in time from 1988 to 1991 using only the baseline, trend and control chart (if available). After that, we add the calendar effects to the others and backcast (forecast backward in time) from 1991 to 1988, then forecast from 1988 to 1991. Multiple backward and forward passes will lower the training set errors, but after more than one or two passes, the testing set errors increase due to overfitting the training set. The values in Table 2 are typical of the training results for the DTMC model. For this reason, we use only one backward and forward pass in the experiments that follow (observe that regression techniques that provide the best set of parameters on a training set are comparable to running a larger number of training passes, and may be expected to provide even worse results).

### 3.3. Smoothing parameter selection

To find a good set of smoothing parameters,  $\alpha$ ,  $\delta$  and  $\phi$ , we adopt the modified grid search found in Gardner and McKenzie (1985). Using eight different parameter triplets forming a cube, we compare the one-step ahead MSE for each. The triplet with the lowest MSE forms the center of the next, proportionately smaller, cube. The search continues until the cube size falls below some tolerance. Since we conduct a grid search for each model, each city finds the set of parameters most appropriate for its demand characteristics. For the experiments, we use (low, high)

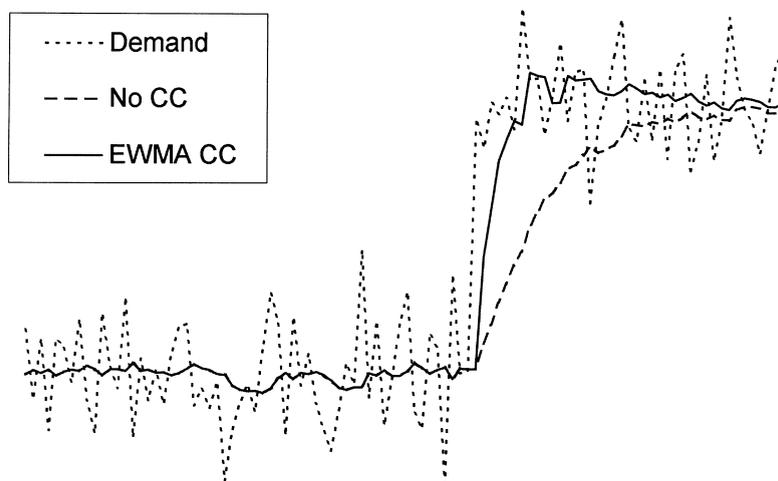


Fig. 2. Representative demand characteristics from three different cities.

Table 2  
Effects of multiple training passes on root mean-squared error (RMSE)

Number of back and forth passes	Training avg. RMSE	Testing avg. RMSE
1	11 859	10 925
2	11 691	10 939
3	11 679	10 955
4	11 641	10 966

limits of (0.02, 0.20) for  $\alpha$ , (0.03, 0.20) for  $\delta$ , and (0.0, 1.0) for  $\phi$ . The tolerances are 0.005, 0.01 and 0.05, respectively. More sophisticated search techniques appear in Berry and Bliemel (1974), Buffa and Taubert (1972) and Flowers (1980).

### 3.4. Handling zero demand

There are two approaches for handling observations of zero demand. In our work, we have found that sometimes it is necessary to ignore observations of zero, while in others, we have to treat the situation of no observations as a zero, and build this into the forecast. In this paper, the lack of any observations on a given day is treated as a zero and the parameters are updated accordingly. Caution must be used here since some disaggregate models can have very small baselines. Demand can arise in infrequent clumps. Since there may be thousands of such models which, when aggregated upward, account for a significant amount of the total, such models cannot be routinely ignored. However, we do use a tolerance level to determine when a model should be discarded.

We have used this method in a rail application where some freight movements can be sporadic. In this setting, we forecast only positive observations, and depend on a different process to predict zero observations. While it is important to be sensitive to this issue, a detailed treatment is beyond the scope of this paper.

### 3.5. ARIMA model selection

Finally, we use seasonal ARIMA models to use as a benchmark against the DTMC model. Standard ARIMA references include Box and Jenkins (1970), Box et al. (1978) and Pankratz (1983). After compiling MSE statistics over the training set for all 5-day season ARIMA models up to order two, ARIMA(2, 1, 1)(1, 0, 1)<sub>5</sub> had the lowest MSE. In standard backshift notation, the ARIMA(2, 1, 1)(1, 0, 1)<sub>5</sub> can be expressed by

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_5 B^5)(1 - B)\hat{z}_t = (1 - \Theta_5 B^5)(1 - \theta_1 B)a_t, \quad (21)$$

where  $\phi_1$ ,  $\phi_2$ ,  $\theta_1$ ,  $\Phi_5$  and  $\Theta_5$  are fitted parameters,  $\hat{z}_t$  is the forecasted demand,  $a_t$  is a white noise term, and  $B$  is a multiplicative backshift operator. Any variable multiplied by  $B$  has its time subscript shifted backward by the power of  $B$ ; that is,  $B^k z_t = z_{t-k}$ .

We may also express the ARIMA(2, 1, 1)(1, 0, 1)<sub>5</sub> model as an explicit function of the previously observed demands and noise terms:

$$\begin{aligned}
z_t - (1 + \phi_1)z_{t-1} + (\phi_1 - \phi_2)z_{t-2} + \phi_2z_{t-3} - \Phi_5[z_{t-5} - (1 + \phi_1)z_{t-6} + (\phi_1 - \phi_2)z_{t-7} + \phi_2z_{t-8}] \\
= a_t - \theta_1a_{t-1} - \Theta_5(a_{t-5} - \theta_1a_{t-6}).
\end{aligned} \tag{22}$$

This ARIMA model serves as the benchmark against which all experimental RMSE's are normalized. In addition, for the testing portion of the demand set, the ARIMA parameters are re-optimized daily. Re-optimizing daily is impractical for the thousands of models encountered in reality, but for the experiments we run, this provides better ARIMA forecasts for comparison. As an aside, higher order ARIMA models were considered, but none had a RMSE more than 1% lower than ARIMA(2, 1, 1)(1, 0, 1)<sub>5</sub> over the training set. In all, nearly 100 different ARIMA models were tested.

#### 4. Experimental comparisons

We present results for three different sets of experiments described below.

- We examine the validity of modeling multiple calendar effects and conclude that using only the day of week effect gives results slightly better than the best ARIMA model and using multiple effects drops the average RMSE another 2%.
- We examine the relationship between trend-adjusted models and the need for control charts and conclude that dampening the trend is superior to using a full trend in these relatively flat demand sets and that control charts have little effect on the training sets, but give a slight improvement for the testing sets.
- We test the effect of each of the control charts under a range of control parameters. EWMA charts give the most accurate and consistent results. Trigg charts also improve accuracy, but are more sensitive to the control parameters. The Shewhart charts provide no improvement.

First, we examine the value of modeling multiple calendar effects. For this set of experiments, we do not use a control chart, and we set the value of the trend multiplier to zero to turn-off the trend portion of the model. We justify these assumptions in the next set of experiments. We use the single backward and forward pass with grid search technique described in Section 3. The results of these experiments averaged over all cities appear in Table 3. This table shows the accuracy of models incorporating different calendar effects. We show the error for both the training data and the testing data. The errors are summarized two ways. The first is the RMSE ratio, giving the RMSE for our model over the RMSE obtained using an ARIMA(2, 1, 1)(1, 0, 1)<sub>5</sub> model. The second is the percent of time that our forecast produced an error smaller than the ARIMA method. The second method is less susceptible to the occasional large error that may produce a larger RMSE, disguising the possibly large number of observations where a method actually produced a better forecast.

We conclude that modeling day of week effects is the best *single* calendar effect to model. In fact, this is the same 5-day period used by the ARIMA model and the results are comparable. However, the combination of day, week and month yields RMSE statistics that are 2.5% lower than ARIMA on average and beat the ARIMA model in 21 of 25 cities (84%) over the training and testing sets. The improvement over the ARIMA model is even more pronounced considering the ARIMA parameters are re-optimized daily and the DTMC parameters are optimized only

Table 3

Average RMSE ratio and percent of demand sets for which exponential smoothing with calendar effect is superior for both training and testing sets against the ARIMA(2, 1, 1)(1, 0, 1)<sub>5</sub> model

Calendar effects	Exponential smoothing models versus ARIMA			
	Training set		Testing set	
	Average RMSE ratio	Percent better than ARIMA	Average RMSE ratio	Percent better than ARIMA
No calendar	1.015	24	1.021	0
Day of week only	0.996	60	0.996	60
Week of month only	1.001	48	1.005	36
Month of year only	1.015	28	1.023	16
End of quarter only	1.016	16	1.023	0
Day and week	0.978	76	0.976	84
Day, week, and month	0.976	84	0.974	84
Day, week, month, quarter	0.978	80	0.976	76

over the training set. Based on these results, the day, week and month calendar effects appear in all subsequent experiments.

Next, we examine the relationship between the trend multiplier,  $\phi$ , and control charts. When  $\phi = 1$ , the model includes the full linear trend. When  $\phi = 0$ , the trend in the forecasting model is forced to zero. These extreme cases are appropriate for data sets with strong and non-existent trends, respectively. However, the grid search in Section 3 finds the optimal  $\phi$  for each city. We examine the results for each of these three values of  $\phi$  with and without a control chart. The control chart we use for this experiment is the EWMA with  $\lambda = 0.10$ ,  $\alpha_{cc} = 0.25$ , low = high = 2.5. The results appear in Table 4.

The optimal value of  $\phi$  is near zero for most of the cities, so it is no surprise that the damped trend results are comparable to the “no trend” results. The full trend model has a significantly higher RMSE than the other two models because most cities have no noticeable daily trend, but the full trend model tries to force a trend on the data. The control chart has limited value with the training sets, but provides a slight improvement on the testing sets, especially with the full trend model. Since the damped trend model is nearly identical to the no trend model, we set  $\phi = 0$  for

Table 4

Relationship between control chart and trend multiplier

		No trend $\phi = 0$		Damped trend		Full trend $\phi = 1$	
		Avg. RMSE ratio	Better than ARIMA (%)	Avg. RMSE ratio	Better than ARIMA (%)	Avg. RMSE ratio	Better than ARIMA (%)
No CC	Training	0.976	84	0.976	84	0.989	64
	Testing	0.974	80	0.975	80	0.994	64
EWMA CC	Training	0.975	84	0.974	84	0.986	64
	Testing	0.971	84	0.972	84	0.988	76

Table 5  
Summary of control chart results with multiple passes

Control chart	1 back/forth pass average		3 back/forth pass average	
	RMSE ratio	Percent better (%)	RMSE ratio	Percent better (%)
EWMA training	0.9752 ± 0.0009	80 ± 6	0.9692 ± 0.0012	88 ± 6
EWMA testing	0.9727 ± 0.0011	82 ± 2	0.9739 ± 0.0010	81 ± 4
Shewhart training	0.9838 ± 0.0051	66 ± 7	0.9785 ± 0.0054	70 ± 8
Shewhart testing	0.9822 ± 0.0061	70 ± 7	0.9846 ± 0.0063	67 ± 8
Trigg training	0.9786 ± 0.0042	73 ± 9	0.9729 ± 0.0047	78 ± 10
Trigg testing	0.9773 ± 0.0054	77 ± 8	0.9786 ± 0.0055	75 ± 9
Training	0.9792 ± 0.0051	73 ± 10	0.9734 ± 0.0055	79 ± 11
Testing	0.9774 ± 0.0060	76 ± 8	0.9790 ± 0.0064	74 ± 9

all cities in subsequent experiments and reduce the grid search from eight points to four at each stage.

The final set of experiments examine the individual control chart parameters. A detailed comparison of the different control charts is not given for reasons of space, but the important results are summarized in Table 5. The EWMA control chart is the clear winner among the control charts. The EWMA has the highest accuracy, lowest standard deviations and tightest consistency across parameter values. These qualities stand out even more for the testing sets, indicating how well the EWMA chart would perform in practice.

As a final experiment, we were interested in knowing the effect of multiple training passes. We ran the same control chart experiment as before with three back-and-forth passes through the training set instead of one. The results appear in Table 5. The two-pass result, not quoted, falls midway between the one and three-pass values. While the three-pass RMSE decreases by 0.5% for the training set, the testing set RMSE increases by 0.1–0.2% for each of the control charts. This highlights why evaluations based on training set statistics can be deceptive when model overfitting is present.

## 5. Conclusions

The DTMC model we propose incorporates multiple periodic and non-periodic calendar effects in an original, but intuitive way. It has proved itself, in practice, to provide better forecasts than the best fitted ARIMA models re-optimized daily and single season exponential smoothing. In addition, we have shown that the EWMA control chart complements the DTMC model well. While the control chart does not significantly improve the forecasts on a day-to-day basis, it provides additional stability at little computational cost.

We need to emphasize that the ARIMA models we have employed are not practical for the large applications that have motivated our research. We would have been happy if our results were the same or slightly worse than the ARIMA models, which require sophisticated software and extensive historical datasets. Our method is easy to code and the models are easy to update on a daily basis with modest storage requirements. That our methods also outperform sophisticated ARIMA techniques, even if by a small margin (based on comparisons of RMSE statistics) is an

important result. More significantly, ARIMA methods, as well as classical exponential smoothing methods, are not designed for the irregular patterns created by calendar effects. It would be possible to estimate a dummy variable regression model, but this would require having access to all the historical data at the same time, a requirement that is not feasible in our context.

Ideas for further investigation include finding better ways to operate the multiple passes during the training phase. Specifically, we want a better way of initializing the calendar multipliers that works well even for non-periodic effects. Second, we are interested in improving the response to a control tracking signal falling outside its control limits. Finally, we would like to know the benefits/drawbacks of re-optimizing the model weekly, monthly, quarterly or yearly.

### Acknowledgements

This research was supported in part by grant AFOSR-F49620-93-1-0098 from the Air Force Office of Scientific Research. The authors would particularly like to acknowledge the comments of Frank Koppelman which helped to clarify the presentation.

### Appendix A. An alternative derivation of updating equations based on stochastic gradients

Although the derivation of the DTMC updating Eqs. (6),(7) and (16) based on the Brown model is correct, we felt it would be worthwhile if we could derive the updating equations from a pure optimization perspective. As we build the notational foundation, we will progress toward an original derivation based on subgradient optimization.

We start with a one-step ahead forecasting model consisting of only a baseline and calendar coefficients. Without loss of generality, the results can be extended to include a trend though the equations become more cumbersome. In addition, we will assume a stationary demand process. We use the following forecasting model for period  $k$ ,

$$F(\Theta) = \exp\left(\theta_0 + \sum_{j \in J} \theta_j y_{jk}\right) = \exp\left(\sum_{j \in \tilde{J}} \theta_j y_{jk}\right) = \exp(\Theta \cdot \bar{y}^k), \quad (23)$$

where  $\Theta$  is a vector whose first component describes the baseline via  $\exp(\theta_0)$  and whose other components describe the calendar effects used to forecast the demand at period  $k$ . If we let  $j = 0$  be the first component of  $\Theta$ , then we can define  $\tilde{J} = \{0\} \cup J$ , set  $y_{0k} = 1$  for all  $k$ , and simplify the forecasting model to be the last equality in Eq. (23).

Let  $(\Omega, \mathcal{F}, \mathcal{P})$  be the underlying probability space for the demand  $D$ , where  $\Omega$  is a set of elementary outcomes,  $\mathcal{F}$  is a set of events, and  $\mathcal{P}$  is a probability measure on  $\mathcal{F}$ . Since we assume the demand is stationary, then  $\mathcal{P}$  remains constant over time and the optimal value for the vector  $\Theta$  becomes independent of time.

Let  $D(\omega)$  be the realization of the demand for the outcome  $\omega \in \Omega$ . Define  $H(\Theta, \omega)$  to be the squared forecast error for the realization  $D(\omega)$ . That is,

$$H(\Theta, \omega) = (D(\omega) - F(\Theta))^2 = \hat{\epsilon}^2. \quad (24)$$

Our goal, then, is to find  $\Theta$  to minimize the expected value of  $H(\Theta)$ :

$$\min_{\Theta} E_{\omega} H(\Theta, \omega) = \min_{\Theta} \int_{\omega \in \Omega} H(\Theta, \omega) \, d\mathcal{P}(\omega). \tag{25}$$

Let  $\hat{H}(\Theta)$  represent the realization  $H(\Theta, \omega)$  of a particular outcome. We will need in our derivation the following two equations:

$$\nabla \hat{H}(\Theta) = -2\hat{\epsilon} \nabla F(\Theta) = -2\hat{\epsilon} F(\Theta) \hat{y}, \tag{26}$$

$$|\hat{y}|^2 = \sum_{j \in \tilde{J}} y_j^2 = \sum_{j \in \tilde{J}} y_j \equiv |\tilde{J}|, \tag{27}$$

The last equality in Eq. (26) follows from taking the gradient of Eq. (23), and the second equality of Eq. (27) follows from all  $y_j$  being either 0 or 1.

We use stochastic gradients to update  $\Theta$  (see Ermoliev, 1988). Let  $\Theta^k$  be our estimate of  $\Theta$  at period  $k$ . Then the updating equation is

$$\begin{aligned} \Theta^{k+1} &= \Theta^k - s^k \cdot \nabla \hat{H}(\Theta^k) \\ &= \Theta^k + s^k \cdot 2\hat{\epsilon}^k F(\Theta^k) \hat{y}^k, \end{aligned} \tag{28}$$

where  $s^k$  is a predetermined step size that is positive with infinite sum and finite squared sum. To perform scaling, we use the step size rule developed for non-differentiable optimization (see Shor, 1985):

$$s^k = \frac{\eta^k \cdot (H_U^k - H_L^k)}{|\nabla H(\Theta^k)|^2}, \tag{29}$$

where  $H_L^k$  and  $H_U^k$  are estimated lower and upper bounds on  $H(\Theta^k)$  and  $\eta^k \in [0, 2]$  such that the conditions on  $s^k$  hold. In our case,  $H_L^k = 0$  and  $H_U^k = (\hat{\epsilon}^k)^2$  are appropriate estimates.

For a particular outcome, the random vector  $\nabla \hat{H}(\Theta)$  may or may not provide a proper direction of descent. However, the only conditions that need to hold are that  $\nabla \hat{H}(\Theta)$  has an expectation that is a subgradient of  $H$  at  $\Theta$  and that the norm of  $\nabla \hat{H}(\Theta)$  has bounded expectation. All that remains is to use Eqs. (26) and (27) to compute  $|\nabla H(\Theta)|^2$ :

$$\begin{aligned} |\nabla H(\Theta)|^2 &= |-2\hat{\epsilon} F(\Theta) \hat{y}|^2 \\ &= 4\hat{\epsilon}^2 F(\Theta)^2 |\hat{y}|^2 \\ &= 4\hat{\epsilon}^2 F(\Theta)^2 |\tilde{J}|. \end{aligned} \tag{30}$$

Therefore, we can simplify the updating Eq. (28) to get,

$$\begin{aligned} (\Theta)^{k+1} &= (\Theta)^k + \frac{\eta^k \cdot ((\hat{\epsilon}^k)^2 - 0)}{4(\hat{\epsilon}^k)^2 F(\Theta^k)^2 |\tilde{J}^k|} \cdot 2\hat{\epsilon}^k F(\Theta^k) \hat{y}^k \\ &= \Theta^k + \frac{\hat{y}^k}{|\tilde{J}^k|} \cdot \frac{\eta^k}{2} \cdot \frac{\hat{\epsilon}_k}{F(\Theta)^k}. \end{aligned} \tag{31}$$

If the sequence  $\eta^k$  has infinite sum and finite squared sum, then under mild conditions on the boundedness of  $F(\Theta)^k$  and the error term  $\hat{\epsilon}^k$  we can prove convergence to the optimal solution with probability one (Ermoliev, 1988).

We now compare the new updating equations with the old ones. Applying the approximation  $\log(1+x) \approx x$  for small  $x$  (which is reasonable considering the smoothing coefficient and relative errors are usually small) to the DTMC updating Eq. (16) for calendar coefficients, we get

$$a_{j,t} = a_{j,t-1} + \frac{1}{|\hat{J}|} \cdot \delta[1 - \alpha(2 - \alpha)] \cdot \left( \frac{e_t}{S_t I_{t-1}^1} \right) \quad \text{for all } j \in \hat{J}. \quad (32)$$

We conclude that the two calendar updating Eqs. (31) and (32) agree since each have similar proportion value ( $\hat{y}^k/|\hat{J}^k|$  versus  $1/|\hat{J}|$  for  $j \in \hat{J}$ , 0 otherwise), smoothing parameter ( $\eta^k/2$  versus  $\delta[1 - \alpha(2 - \alpha)]$ ), and relative error terms ( $\hat{\epsilon}^k/F(\Theta)^k$  versus  $e_t/S_t I_{t-1}^1$ ). With the proper manipulations, the baseline updating equations agree similarly. Thus, we conclude that the independent derivations produce similar updating equations.

## References

- Berry, W.L., Bliemel, F.W., 1974. Selecting exponential smoothing constants: an application of pattern search. *International Journal of Production Research* 12, 483–499.
- Box, G.E.P., Hunter, W.G., Hunter, J.S., 1978. *Statistics for Experimenters*. Wiley, New York.
- Box, G.E.P., Jenkins, G.M., 1970. *Time Series Analysis: Forecasting and Control*. Holden-Day, New York.
- Brown, R.G., 1963. *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice-Hall, Englewood Cliffs, NJ.
- Buffa, E.S., Taubert, W.H., 1972. *Production–inventory Systems: Planning and Control*, revised ed. Irwin, Homewood, IL.
- Ermoliev, Y., 1988. Stochastic quasi-gradient methods. In: Y. Ermoliev, R. Wets (Eds.), *Numerical Techniques for Stochastic Optimization*. Springer, Berlin.
- Flowers, A.D., 1980. A simulation study of smoothing constant limits for an adaptive forecasting system. *Journal of Operations Management* 2, 85–94.
- Gardner Jr., E.S., 1985. Exponential smoothing: the state of the art. *Journal of Forecasting* 4 (1), 1–28.
- Gardner Jr., E.S., Dannenbring, D.G., 1980. Forecasting with exponential smoothing: Some guidelines for model selection. *Decision Sciences* 11, 370–383.
- Gardner Jr., E.S., McKenzie, E., 1985. Forecasting trends in time series. *Management Science* 31 (10), 1237–1246.
- Gardner Jr., E.S., McKenzie, E., 1989. Seasonal exponential smoothing with damped trends. *Management Science* 35 (3), 372–376.
- Harrison, P.J., 1967. Exponential smoothing and short-term sales forecasting. *Management Science* 13, 821–842.
- Hibon, M., 1984. Naive moving average exponential smoothing, and regression methods. In: *The Forecasting Accuracy of Major Time Series Methods*, chapter 8. Wiley, New York, pp. 239–244.
- Holt, C.C., Modigliana, F., Muth, J.F., Simon, H.A., 1960. *Planning Production, Inventories, and Work Force*. Prentice-Hall, Englewood Cliffs, NJ.
- Hunter, J.S., 1986. The exponentially weighted moving average. *Journal of Quality Technology* 18 (4), 19–25.
- Johnson, N.L., Leone, F.C., 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences*, 2nd ed. Wiley, New York.
- Ledolter, J., Abraham, B., 1984. Some comments on the initialization of exponential smoothing. *Journal of Forecasting* 3, 79–84.
- Lucas, J.M., Saccucci, M.S., 1990. Exponentially weighted moving average control schemes: properties and enhancements with discussion. *Technometrics* 32 (1), 1–30.

- Makridakis, S., Hibon, M., 1979. Accuracy of forecasting: an empirical investigation with discussion. *Journal of the Royal Statistical Society* 142, 97–145.
- Makridakis, S., Wheelwright, S.C., 1978. *Interactive Forecasting: Univariate and Multi-variate Methods*, 2nd ed. Holden-Day, San Francisco, CA.
- McClain, J.O., 1974. Dynamics of exponential smoothing with trend and seasonal terms. *Management Science* 20 (9), 1300–1304.
- McClain, J.O., Thomas, L.J., 1973. Response–variance trade-offs in adaptive forecasting. *Operations Research* 21 (2), 554–568.
- Montgomery, D.C., Johnson, L.A., Gardiner, J.S., 1990. *Forecasting and Time Series Analysis*, 2nd ed. McGraw-Hill, New York.
- Muth, J.F., 1960. Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association* 55, 299–306.
- Nerlove, M., Wage, S., 1964. Some observations on adaptive forecasting. *Management Science* 10, 207–224.
- Pankratz, A. 1983. *Forecasting with Univariate Box–Jenkins Models*. Wiley, New York.
- Roberts, S.W., 1959. Control chart tests based on geometric moving averages. *Technometrics* 1, 239–250.
- Shor, N.Z., 1985. *Minimization Methods for Non-differentiable Functions*. Springer, Heidelberg.
- Theil, H., Wage, S., 1964. Some observations on adaptive forecasting. *Management Science* 10, 198–206.
- Trigg, D.W., 1964. Monitoring a forecasting system. *Operational Research Quarterly* 15, 271–274.
- Trigg, D.W., Leach, A.G., 1967. Exponential smoothing with an adaptive response rate. *Operational Research Quarterly* 18, 53–59.
- Whybark, D.C., 1973. A comparison of adaptive forecasting techniques. *The Logistics and Transportation Review* 9 (1), 13–26.
- Winters, P.R., 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* 6, 324–342.