# MODELS OF THE HUMAN FORECASTING BEHAVIOR

A Note on the Relationship between the Exponential
Smoothing Method and the Bayesian Method

TAKEHIKO MATSUDA and MITSUHARU SEKIGUCHI

*Tokyo Institute of Technology*

## Abstract

In a relatively stationary environment, the human forecasting is very similar to the exponential smoothing model. But the exponential smoothing method does not seem to have any persuasive structural basis. In this paper, we discuss the meaning of the smoothing coefficient, referring to the Bayesian method, which is the most rational method under a stationary uncertainty, and conclude that the exponential smoothing method is a rational model if the time-series observations are large in number and the conditional probability of the environment is linearly dependent upon the former posterior estimation.

## 1. Introduction

Forecasting is an important part of the human decision making. We usually decide upon a certain action on the basis of a forecast. If you forecast that it is likely to rain this afternoon, then you will decide

to take an umbrella with you. For such forecasting, there are two types; structured and non-structured.

A structured forecasting is such that, when a forecasting is needed, we can choose the data to use and deduce the forecast from them by means of the network of the cause-and-effect relationships that are hypothesized to be governing the situation and somehow verified to a reasonable extent. Following is the case of a structured forecasting. Suppose that the weather forecast for today has failed and it rained. From this fact we can easily forecast that the waiting line for taxies at the railroad terminal stations will be very long. This forecast is based upon a series of reasonably well-established empirical relationships.

When such structuredness develops to an extreme, we will reach the state where we can completely predict the event to follow, if we can get all the necessary information. This type of forecasting may be called predetermination or predestination. For example, the hours of sunrise and sunset are completely forecastable when the necessary data concerning the mechanism of the solar system are given.

A non-structured forecasting, on the other hand, is concerned with such circumstances where no legitimate data exist or, even if such data may be available, their mutual relationships are unknown. In a practical situation, this type of forecasting will take the form of judgement or the form of hunch; this is, "I don't know the reason, but I feel it will happen."

Such classification of forecasting into the "structured" and the "non-structured" is simply to consider the extremes of a continum. Actual forecasting activities often have both the features to a certain extent and should perhaps be placed somewhere upon the continum. In this paper, analysis of the network of the cause-and-effect relationships which enable us to deduce the forecast from the data is not treated. Instead, the relationship between uncertainty and forecasting is discussed, when the forecasting structure is given.

## 2.  Rational Models of Forecasting

When we have already accumulated some experiences and can guess the relationship between the environment and the outcome, what is the best method to forecast the next outcome?

Let $E=\{e_1, \cdots, e_n\}$ be a set of environments and $X=\{x_1, \cdots, x_m\}$ be a set of outcomes.   We know the relationship $R(E, X)$ only in the probabilistic sense, and we have some past time-series data $(E(t); X(t))=(e(1), \cdots, e(t); x(1), \cdots, x(t))$.   Given such conditions, our problem is specified as to how we forecast the next outcome $x(t+1)$.   In this paper, we do not consider any rewards or penalties in connection with the correctness of the forecast.

Let $P_t'(e_i)$ denote the prior probability with which the environment $e_i$ is to occur at the beginning of the $t$-th period.   Then,

$$\sum_{i=1}^{n} P_t'(e_i)=1.0 \qquad \text{for all } t.$$

And similarly, let $P(x_j|e_i)$ denote the conditional probability which represents the $e_i{\rightarrow}x_j$ relationship, and thus

$$\sum_{j=1}^{m} P(x_j|e_i)=1.0 , \qquad i=1, \cdots, n$$

In this case, the rational forecast of the outcome for period $t$ will be the one that has the maximum likelihood,

$$\hat{x}(t)=x_k$$

where $k$ satisfies

(2-1)          $$\sum_{i} P(x_k|e_i)P(e_i)=\max_{j}\sum_{i} P(x_j|e_i)P(e_i)$$

Now, how can we get the prior probability at the beginning of period $t$?   If we assume that the environment does not change, then it will be most rational to set the prior probability equal to the posterior probability for period $t-1$ [4]; that is,

(2.2)    $P_t'(e_i)=P''_{t-1}(e_i)$        for all   $e_i\in E$ ,

where $P''_{t-1}(e_i)$ is the posterior probability of the environment $e_i$ at the end of period $t-1$. And we can get the posterior probability for period $t$ by the Bayes's theorem.

$$P''_t(e_i)=P''_t(e_i|x(t))=\frac{P(x(t)|e_i)P'_t(e_i)}{\sum\limits_{j=1}^{n}P(x(t)|e_j)P'_t(e_j)}$$

In the case of the continuous events, where the environment and the outcome can be expressed by continuous parameters, the above expression can be rewritten using the probability density functions as follows:

$$v'_{t+1}(e)=\frac{h(x(t)|e)v'_t(e)}{\int_E h(x(t)|e)v'_t(e)de}$$

where $v'_t(e)$ is the prior p.d.f. of the environment and $h(x(t)|e)$ is the conditional likelihood that the outcome $x(t)$ may occur under a certain environment $e$ at period $t$, then $\int_x h(x(t)|e)dx(t)=1.0$ .

Given a conditional likelihood function and a prior probability distribution function, what is the best forecast? There are three ways of thinking. First, for $\hat{x}(t+1)$ take the outcome which gives the maximum likelihood; namely,

(2-3)     $\hat{x}_1(t+1)=x^*$

where $x^*$ satisfies

$$\int_E h(x^*|e)v'_{t+1}(e)de=\max_x\int_E h(x|e)v'_{t+1}(e)de .$$

Second, take the expected value; that is,

(2-4)     $\hat{x}_2(t+1)=\int_X x\int_E h(x|e)v'_{t+1}(e)dedx .$

Third, considering the emotional aspect, optimistic or pessimistic, take the expected values modified by the variance, that is,

(2-5)      $\hat{x}_3(t+1) = \int_X x \int_E h(x|e)v'_{t+1}(e)dedx$

$$+ \alpha \left( \int_X (x-\bar{x})^2 \int_E h(x|e)v'_{t+1}(e)dedx \right)^{\frac{1}{2}},$$

where $\bar{x} = \hat{x}_2(t+1)$ and $\alpha$ is a parameter which represents the attitude of the forecaster. If $\alpha$ is equal to 0, the case is same with (2-4); otherwise, we can say that the forecast is an optimistic one or a pessimistic one. But we will not discuss the difference among these types any more.

As the general example, we consider the normal process in which the mean is unknown but the variance is known. The first type forecast and the second type forecast mentioned above give the same forecast; $\hat{x}_1(t+1) = \hat{x}_2(t+1)$. The prior probability density function $v'_t(e)$ and the conditional likelihood function $h(x|e)$ are;

$$v'_t(e) = \frac{1}{\sqrt{2\pi}\,\sigma'_t} \exp\left\{ -\frac{(e-\mu'_t)^2}{2\sigma'^2_t} \right\}$$

$$h(x|e) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{(x-e)^2}{2\sigma^2} \right\}$$

then, the posterior probability density function is

$$v''_t(e) = \frac{1}{\sqrt{2\pi}\sqrt{\dfrac{\sigma'^2_t\sigma^2}{\sigma'^2_t+\sigma^2}}} \exp\left\{ -\frac{\left(e-\dfrac{\sigma'^2_t x_t+\sigma^2\mu'_t}{\sigma^2+\sigma'^2_t}\right)^2}{2\cdot\dfrac{\sigma^2\sigma'^2_t}{\sigma^2+\sigma'^2_t}} \right\}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma''_t} \exp\left\{ -\frac{(e-\mu''_t)^2}{2\sigma''^2_t} \right. .$$

Since we assume that $v'_{t+1}(e) = v''_t(e)$,

(2-6)      $\mu'_{t+1} = \mu''_t = \dfrac{\sigma'^2_t x_t+\sigma^2\mu'_t}{\sigma^2+\sigma'^2_t}$

(2-7)      $\sigma'^2_{t+1} = \sigma''^2_t = \dfrac{\sigma^2\cdot\sigma'^2_t}{\sigma^2+\sigma'^2_t}$  .

From (2-7), $\sigma^2/\sigma'^2_{t+1} = 1 + \sigma^2/\sigma'^2_t$ . And by taking into consideration the time-series characteristics, we obtain

$$\frac{\sigma^2}{\sigma'^2_{t+1}} = t + \frac{\sigma^2}{\sigma'^2_1} \ .$$

Now that the prior probability distribution at period 1 can be interpreted as estimating the environment without any past experience, we can reasonably assume that $\sigma'^2_1 \gg \sigma^2$ [1]. So,

$$\sigma^2/\sigma'^2_1 \doteqdot 0 \quad \text{and} \quad \sigma^2/\sigma'^2_{t+1} = t \ .$$

Thus

(2-8) $\qquad \sigma'^2_{t+1} = \sigma^2/t$

and

(2-9) $\qquad \mu'_{t+1} = \frac{1}{t} \sum_{k=1}^{t} x_k \ .$

This is the same result with the well-known case, where the problem is to estimate the mean by using the data of sample size $t$ taken from the population which is normally distributed with a known variance.

### 3. A Model that Discounts the Past Information

The preceding forecasting model has a prominent characteristic in that all the past data are used with the same weight, $1/t$. This means that if we are informed, or know by experience, that the environment does not change, such an equal weight forecasting method would give the optimum. But an introspection into the human forecasting behavior would easily suggest that the consideration of the time-series characteristic have much influence upon the forecast. Therefore, the use of such a forecasting method that discounts the past information is likely to give a better approximation to the human forecasting process.

---

1) For the prior distribution without any information, Lindley [2] takes the $[-\infty, \infty]$ uniform distribution.

Let $\{x_1, \cdots, x_t\}$ be the time-series data which have already been experienced or observed, and $f$ be the forecasting function. Then,

$$\hat{x}_{t+1} = f(x_1, \cdots, x_t)$$

Taking all the elements of the time-series data as independent variables, we might characterize a model that discounts the past information as follows,

$$|\partial f/\partial x_1|_\xi < |\partial f/\partial x_2|_\xi < \cdots < |\partial f/\partial x_t|_\xi \qquad (\xi \in X)$$

and $|\partial f/\partial x_i|/|\partial f/\partial x_t|$ $(=\beta_i)$ represents the amount of discount for period $i$.

In the simplest case, where $\beta_i = \beta^{t-i}$ *i.e.* the rate of discount per period is constant $\beta(0 \le \beta \le 1)$, the forecast for period $t+1$ will be

(3. 1)
$$\hat{x}_{t+1} = \sum_{\tau=1}^{t} \beta^{t-\tau} x_t \bigg/ \sum_{\tau=1}^{t} \beta^{t-\tau}$$

$$= \frac{1}{\sum_{\tau=0}^{t-1} \beta^\tau} x_t + \beta \cdot \frac{\sum_{\tau=0}^{t-2} \beta^\tau}{\sum_{\tau=0}^{t-1} \beta^t} \cdot \hat{x}_t \ .$$

If $t$ is so large that $\beta^t$ is nearly 0, then, approximately

$$\frac{\sum_{\tau=0}^{t-2} \beta^\tau}{\sum_{\tau=0}^{t-1} \beta^\tau} = 1 \quad \text{and} \quad \sum_{\tau=0}^{t-1} \beta^\tau = \frac{1}{1-\beta} \ .$$

Thus, the model can be rewritten as

(3. 2)
$$\hat{x}_{t+1} = (1-\beta) x_t + \beta \hat{x}_t$$
$$= \alpha x_t + (1-\alpha) \hat{x}_t$$

where $\alpha = 1-\beta$. Notice that this is equivalent to the exponential smoothing model.

## 4. Structural Aspects of the Discounted Information Model

In choosing the smoothing constant $\alpha$, many researchers have re-

course to the simulation study. For instance, Winters [5] varies these coefficients from 0 to 0.7, and selects the best ones. But such procedures gives no positive ground for the validity of the smoothing coefficients thus selected, for the reason is not based upon the structural or physical features of the process. In what follows, we propose an explanation as to what ground the smoothing coefficient $\alpha$ or the time discount factor $\beta$ is founded upon, by clarifying the structural characteristics of the discounted information model.

The pattern of a forecasting process may be shown by the following steps;

1. prior estimation of the environment for the period concerned
2. forecasting
3. getting the outcome
4. reflecting the realized environment
5. return to the step 1.

If we assume that a man behaves rationally in his forecasting, then where in this process does the discounting of the past information take place? We have already mentioned these four steps, 1 through 4, but as for the feedback process, step 5, we have made an assumption that if there were no environmental change, then the prior estimation for the subsequent period would be equal to the posterior estimation for the current period.

In case of a human forecasting activities, however, this assumption does not seem to hold exactly, and should be modified into a weaker assumption such that the environment in the next period does not drastically vary from that of the present period. This weaker assumption means that the prior estimation of the next period is not exactly the same as the posterior estimation of this period, but rather takes a form with more vagueness.

Consider again the foregoing normal case. The posterior distribution is

$$v''_t(e) = \frac{1}{\sqrt{2\pi}\ \sigma''_t} \exp\left\{-\frac{(e-\mu''_t)^2}{2\sigma''^2_t}\right\} .$$

Then, the prior distribution at period $t+1$ takes a form:

$$v'_{t+1}(e) = \frac{1}{\sqrt{2\pi}\ \sqrt{\sigma''^2_t+\eta_t^2}} \exp\left\{-\frac{(e-\mu''_t)^2}{2(\sigma''^2_t+\eta_t^2)}\right\} .$$

This means that there exists some doubt as to whether the environment in the next period is the same as in this period. The situation may be expressed by a probability function, stating that if the environment in this period is $e'$ then the environment in the next period will be $e$, the connecting relationship between $e$ and $e'$ being such as follows;

$$g_t(e|e') = \frac{1}{\sqrt{2\pi}\ \eta_t} \exp_t\left\{-\frac{(e-e')^2}{2\eta_t^2}\right\} ,$$

so,

(4.1)
$$v'_{t+1}(e) = \int_{-\infty}^{\infty} g_t(e|e')v''_t(e')de'$$

$$= \frac{1}{\sqrt{2\pi}\ \sigma'_{t+1}} \exp\left\{-\frac{(e-\mu'_{t+1})^2}{2\sigma'^2_{t+1}}\right\}$$

where,

$$\sigma'^2_{t+1} = \sigma''^2_t + \eta_t^2 \qquad \text{and} \qquad \mu'_{t+1} = \mu''_t .$$

If we assume $\eta_t^2 = \lambda\sigma''^2_t$, $(\lambda > 0)$, then

(4.2)
$$\sigma'^2_{t+1} = (1+\lambda)\sigma''^2_t = \frac{1}{\beta}\sigma''^2_t$$

where $\beta = 1/(1+\lambda) < 1$.

From the equation (2-7)

$$\sigma''^2_t = \frac{\sigma^2}{1 + \dfrac{\sigma^2}{\sigma'^2_t}}$$

then using (4-2), we obtain

$$(4\text{-}3) \qquad \frac{\sigma^2}{\sigma'^2_{t+1}} = \beta\left(1 + \frac{\sigma^2}{\sigma'^2_t} \cdots\right)$$

$$= \sum_{\tau=1}^{t} \beta^\tau + \beta^t \frac{\sigma^2}{\sigma'^2_1}$$

If we assume $\sigma'^2_1 \gg \sigma^2$, then

$$(4\text{-}4) \qquad \sigma'^2_{t+1} = \sigma^2 \Big/ \sum_{\tau=1}^{t} \beta^\tau$$

and from (4-4) and (2-6)

$$\mu'_{t+1} = \mu''_t$$

$$= \left(1 \Big/ \sum_{\tau=1}^{t} \beta^\tau\right) \cdot x_t + \beta \cdot \left(\sum_{\tau=1}^{t-1} \beta^\tau \Big/ \sum_{\tau=1}^{t} \beta^\tau\right) \cdot \mu'_\tau .$$

Notice that this result corresponds closely with (3-1), the model that discounts the past information, discussed in section 3.

If $t$ is sufficiently large, then

$$\sigma'^2_{t+1} = \frac{1-\beta}{\beta} \sigma^2$$

$$= \frac{1}{\beta} \sigma''^2_t$$

$$(4\text{-}5) \qquad \therefore \ \sigma''^2_t = (1-\beta)\sigma^2 .$$

From the assumption and (4-5)

$$\eta_t^2 = \lambda \sigma''^2_t$$
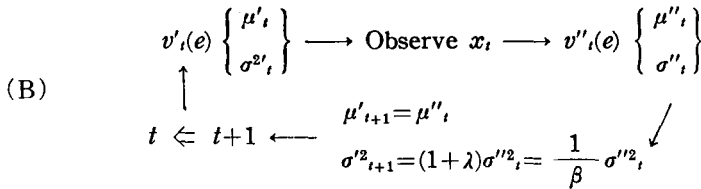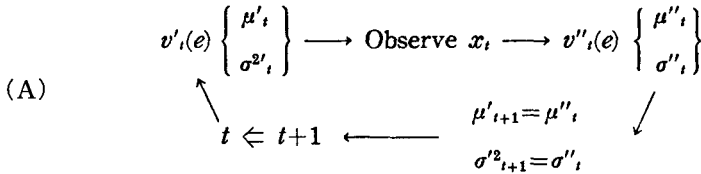
$$= \frac{(1-\beta)^2}{\beta} \sigma^2 .$$

Thus the ratio of the degree of doubt about the environmental change to the degree of uncertainty of the process itself can be expressed as;

$$(4\text{-}6) \qquad \frac{\eta_t^2}{\sigma^2} = \frac{(1-\beta)^2}{\beta} = \frac{\alpha^2}{1-\alpha} = \frac{\lambda^2}{1+\lambda} .$$

Hence, the exponential smoothing method is the same as the dis-

counted information model, and the smoothing constant $\alpha$ corresponds closely with the discount ratio $\beta$. This is to say that the posterior esti-mation of the environment from some past observations is augmented with the parameter of vagueness $\lambda$ in formulating the prior estimation for the next period. This $\lambda$ can be interpreted as the degree of doubt about the stationarity of the environment through periods.

In summary, there are some structural differences between the Bayesian forecasting model and the exponential-smoothing forecasting model. The differences would seem to exist in the connecting link be-tween the posterior estimation and the prior estimation. Next chart A expresses the Bayesian type forecasting, and chart B represents the smoothing type forecasting.

(A)

$$v'_t(e) \begin{Bmatrix} \mu'_t \\ \sigma^{2\prime}_t \end{Bmatrix} \longrightarrow \text{Observe } x_t \longrightarrow v''_t(e) \begin{Bmatrix} \mu''_t \\ \sigma''_t \end{Bmatrix}$$

$$t \Leftarrow t+1 \longleftarrow \quad \begin{matrix} \mu'_{t+1} = \mu''_t \\ \sigma'^2_{t+1} = \sigma''_t \end{matrix}$$

(B)

$$v'_t(e) \begin{Bmatrix} \mu'_t \\ \sigma^{2\prime}_t \end{Bmatrix} \longrightarrow \text{Observe } x_t \longrightarrow v''_t(e) \begin{Bmatrix} \mu''_t \\ \sigma''_t \end{Bmatrix}$$

$$t \Leftarrow t+1 \longleftarrow \quad \begin{matrix} \mu'_{t+1} = \mu''_t \\ \sigma'^2_{t+1} = (1+\lambda)\sigma''^2_t = \dfrac{1}{\beta}\sigma''^2_t \end{matrix}$$

## 5. A Case of Linear Regression

In the preceding section we assumed that the environment did not change, and had a constant parameter. In this section, we assume that the environment varies and is characterized by a variable $x_t$, and the outcome $y_t$ which we wish to forecast has a linear relationship with $x_t$, expressed as $y_t = \alpha + \beta x_t + \varepsilon_t$, where $\alpha$ and $\beta$ are unknown parameters and $\varepsilon_t$ is a random variable distributed with $N(0, \sigma^2)$. We want to for-

mulate the expressions of sequential estimation for the regression parameters $\alpha$ and $\beta$.

First we formulate the expressions by the Bayesian method which gives the same result with the least squares method, if we assume the prior distribution to be uniform or its variance to be very large. Second, we formulate a case of discounting by a similar method used as before. Third, we consider the case of time series where $x_i = t$, discounting the past data, and we will show that this model is the same with the double exponential smoothing method, if $t$ is very large.

( i ) Bayesian method

Basic assumptions ;

1. A pair of observed data $(x, y)$ has a relationship ;

$$y = \alpha + \beta x + \varepsilon$$

where $\alpha$ and $\beta$ are unknown parameters, and $\varepsilon$ is a random variable from $N(0, \sigma^2)$.

2. Each pair of the data is observed with a discrete time interval, and so the pair $(x_n, y_n)$ represents the observation in period $n$.

3. The estimated values of $\alpha$, $\beta$ have such probability density function, such as

$$v(\alpha, \beta | X_n, Y_n) = h(\alpha | \beta, X_n, Y_n) \cdot \pi(\beta | X_n, Y_n)$$

where $X_n = (x_1, \cdots, x_n)$ and $Y_n = (y_1, \cdots, y_n)$ are all observed data, and

$$h(\alpha | \beta, X_n, Y_n) \propto \exp\left\{ -\frac{(\alpha - a_n)^2}{2\lambda_n^2} \right\}$$

$$\pi(\beta | X_n, Y_n) \propto \exp\left\{ -\frac{(\beta - b_n)^2}{2\eta_n^2} \right\}$$

4. The prior distribution is equal to the posterior distribution in the preceding period.

5. The prior distribution in period 1 has a very large variance, so that $\sigma^2/\lambda_0^2 = 0$ and $\sigma^2/\eta_0^2 = 0$.

Then, the prior distribution of $(\alpha, \beta)$ in period $n+1$ is

$$v(\alpha, \beta | X_{n-1}, Y_{n-1}, x_n, y_n) \propto \rho(x_n, y_n | \alpha, \beta) \cdot v(\alpha, \beta | X_{n-1}, Y_{n-1})$$

$$\propto \exp\left\{-\frac{(y_n - \alpha - \beta x_n)^2}{2\sigma^2}\right\} \exp\left\{-\frac{(\alpha - a_{n-1})^2}{2\lambda^2_{n-1}}\right\}$$

$$\exp\left\{-\frac{(\beta - b_{n-1})^2}{2\eta^2_{n-1}}\right\}$$

$$h(\alpha | \beta, x_n, y_n, X_{n-1}, Y_{n-1}) \propto \rho(x_n, y_n | \alpha, \beta) \cdot h(\alpha | \beta, X_{n-1}, Y_{n-1})$$

$$\propto \exp\left\{-\frac{\sigma^2 + \lambda^2_{n-1}}{2\sigma^2 \lambda^2_{n-1}} \left(\alpha - \frac{\lambda^2_{n-1}(y_n - \beta x_n) + \sigma^2 a_{n-1}}{\sigma^2 + \lambda^2_{n-1}}\right)^2\right\}$$

Therefore the estimate of the parameter $\alpha$ is

(5.1) $$\hat{\alpha} = a_n = \frac{\lambda^2_{n-1}(y_n - \beta x_n) + \sigma^2 a_{n-1}}{\sigma^2 + \lambda^2_{n-1}}$$

(5.2) $$a_n = \frac{\left(y_n + \dfrac{\sigma^2}{\lambda^2_{n-1}} Y_{n-1}\right) - \beta\left(x_n + \dfrac{\sigma^2}{\lambda^2_{n-1}} X_{n-1}\right)}{1 + \dfrac{\sigma^2}{\lambda^2_{n-1}}}$$

$$= Y_n - \beta X_n$$

where

(5-3) $$Y_n = \sum_{i=1}^{n} y_i / n$$

(5-4) $$X_n = \sum_{i=1}^{n} x_i / n ,$$

and

$$\lambda^2_n = \frac{\sigma^2 \lambda^2_{n-1}}{\sigma^2 + \lambda^2_{n-1}}$$

$$= \frac{\sigma^2}{1 + \dfrac{\sigma^2}{\lambda^2_{n-1}}}$$

$$(5\text{-}5) \qquad \therefore \quad \frac{\sigma^2}{\lambda^2_n} = n + \frac{\sigma^2}{\lambda^2_0} \doteqdot n \ .$$

And the conditional likelihood of $\beta$ is expressed as,

$$\pi(\beta | x_n, y_n, X_{n-1}, Y_{n-1}) := \int_{-\infty}^{\infty} v(\alpha, \beta | x_n, y_n, X_{n-1}, Y_{n-1}) d\alpha$$

$$\propto \exp\left\{ - \frac{\eta^2_{n-1}(x_n - X_{n-1})^2 + \lambda^2_{n-1} + \sigma^2}{2\eta^2_{n-1}(\sigma^2 + \lambda^2_{n-1})} \cdot \right.$$

$$\left. \left( \beta - \frac{\eta^2_{n-1}(x_n - X_{n-1})(y_n - Y_{n-1}) + (\sigma^2 + \lambda^2_{n-1})b_{n-1}}{\eta^2_{n-1}(x_n - X_{n-1})^2 + \lambda^2_{n-1} + \sigma^2} \right)^2 \right\}$$

Thus, the estimate of the parameter $\beta$ is

$$(5\text{-}6) \qquad \hat{\beta} = b_n = \frac{(x_n - \bar{X}_{n-1})(y_n - \bar{Y}_{n-1}) + \dfrac{\sigma^2 + \lambda^2_{n-1}}{\eta^2_{n-1}} b_{n-1}}{(x_n - X_{n-1})^2 + \dfrac{\sigma^2 + \lambda^2_{n-1}}{\eta^2_{n-1}}}$$

where, from (5-3)

$$\sigma^2 + \lambda^2_{n-1} = \frac{n}{n-1} \sigma^2$$

and

$$\eta^2_n = \frac{\eta^2_{n-1}(\sigma^2 + \lambda^2_{n-1})}{\eta^2_{n-1}(x_n - X_{n-1})^2 + \lambda^2_{n-1} + \sigma^2}$$

$$= \frac{\dfrac{n}{n-1} \sigma^2}{(x_n - \bar{X}_{n-1})^2 + \dfrac{n}{n-1} \cdot \dfrac{\sigma^2}{\eta^2_{n-1}}}$$

$$(5\text{-}7) \qquad \therefore \quad \frac{\sigma^2}{\eta^2_n} = \frac{n-1}{n}(x_n - \bar{X}_{n-1})^2 + \frac{\sigma^2}{\eta^2_{n-1}}$$

$$= \sum_{k=1}^{n} x^2_k - \frac{1}{n} \left( \sum_{k=1}^{n} x_k \right)^2$$

$$= S^2_n$$

Then,

(5-8)
$$b_n = \frac{(x_n - \bar{X}_{n-1})(y_n - \bar{Y}_{n-1}) + \dfrac{n}{n-1} S^2_{n-1} b_{n-1}}{(x_n - \bar{X}_{n-1})^2 + \dfrac{n}{n-1} S^2_{n-1}}$$

$$= \frac{\sum\limits_{k=1}^{n} \left( x_k - \dfrac{\sum\limits_{i=1}^{n} x_i}{n} \right) \left( y_k - \dfrac{\sum\limits_{i=1}^{n} y_i}{n} \right)}{S^2_n}$$

These results are exactly the same as the result of the least squares method.

( ii ) Discounting past data

Here, a discounting factor $\gamma$ is used in the process of calculating the prior distribution. We put $\gamma$ in the equations of the variance, which gives

$$\lambda^2_n = \frac{1}{\gamma} \cdot \frac{\sigma^2}{1 + \dfrac{\sigma^2}{\lambda^2_{n-1}}}$$

$$\eta^2_n = \frac{1}{\gamma} \cdot \frac{\eta^2_{n-1}(\sigma^2 + \lambda^2_{n-1})}{(x_n - \bar{X}_{n-1})^2 \eta^2_{n-1} + \lambda^2_{n-1} + \sigma^2}$$

Then,

(5-9)
$$\frac{\sigma^2}{\lambda^2_n} = \sum_{\tau=1}^{n} \gamma^\tau$$

and

$$\frac{\sigma^2}{\eta^2_n} = \gamma \cdot \left[ \frac{\sum\limits_{k=1}^{n-1} \gamma^k}{\sum\limits_{k=0}^{n-1} \gamma^k} \left\{ (x_n - \bar{X}_{n-1})^2 + \frac{\sum\limits_{k=0}^{n-1} \gamma^k}{\sum\limits_{k=1}^{n-1} \gamma^k} \cdot \frac{\sigma^2}{\eta^2_{n-1}} \right\} \right]$$

$$= \sum_{j=2}^{n} \gamma^{n-j-1} \frac{\sum\limits_{k=1}^{j-1} \gamma^k}{\sum\limits_{k=0}^{j-1} \gamma^k} \{(x_j - \bar{X}_{j-1})^2\}$$

where

(5-10)
$$X_j = \frac{x_j + \frac{\sigma^2}{\lambda^2_{n-1}} X_{j-1}}{1 + \frac{\sigma^2}{\lambda^2_{n-1}}}$$

$$= \sum_{k=1}^{j} \gamma^{j-k} x_k \Big/ \sum_{k=0}^{j-1} \gamma^k \ ,$$

similarly

$$Y_j = \sum_{k=1}^{j} \gamma^{j-k} y_k \Big/ \sum_{k=0}^{j-1} \gamma^k$$

(5-11) $\quad \therefore \quad \dfrac{\sigma^2}{\eta^2_n} = \gamma \left\{ \sum_{j=1}^{n} \gamma^{n-j} x^2_j - \dfrac{\left( \sum_{j=1}^{n} \gamma^{n-j} x_j \right)^2}{\sum_{j=0}^{n-1} \gamma^j} \right\}$

$$= \gamma \cdot S_{xx}(n)$$

Thus we get the estimate of $\beta$ at period $n$ as follows;

(5-12) $\quad b_n = \dfrac{(x_n - X_{n-1})(y_n - Y_{n-1}) + \left( \sum_{j=0}^{n-1} \gamma^j \Big/ \sum_{j=0}^{n-2} \gamma^j \right) S_{xx}(n-1) \cdot b_{n-1}}{(x_n - X_{n-1})(y_n - Y_{n-1}) + \left( \sum_{j=0}^{n-1} \gamma^j \Big/ \sum_{j=0}^{n-2} \gamma^j \right) S_{xx}(n-1)}$

$$= \dfrac{\sum_{k=1}^{n} \gamma^{n-k} x_k y_k - \left( \sum_{k=1}^{n} \gamma^{n-k} x_k \right) \left( \sum_{k=1}^{n} \gamma^{n-k} y_k \right) \Big/ \sum_{j=0}^{n-1} \gamma^j}{S_{xx}(n)}$$

This is the same with discounted least squares method.[2]

---

2) We use the word "discounted" whenever we can not treat the past observed data in the same manner with the latest one and therefore, in order to use these data effectively, we must give them smaller weights than to the latest.

Suppose that we observed some pairs of data $\{x_\tau, \ y_\tau\}$ ($\tau = 1, \cdots, t$; $t$ is the latest period of observation) and from these data we want to estimate the regression parameters $\alpha$ and $\beta$. In the usual case, we solve the following problem;

To get $\hat{\alpha}$, $\hat{\beta}$ which minimize $\sum_{\tau=1}^{t} (y_\tau - \alpha - \beta x_\tau)^2$. In comparison, the problem in the discounted least squares method is;

To get $\hat{\alpha}_t$, $\hat{\beta}_t$ which minimize $\sum_{\tau=1}^{t} \gamma^{t-\tau} (y_t - \alpha - \beta x_\tau)^2$, where $\gamma$ is the discounting factor and $0 < \gamma < 1$.

(iii)   A case of time dependent

If we replace $x_t$ with $t$ in the foregoing model, we get a model of time dependent.   The observed data $\{y(\tau)\}$ ; $(\tau = t_1, \cdots, n)$ are considered to have the following structure

$$y(\tau) = \alpha + \beta\tau + \varepsilon\tau$$

From (5-12), we can estimate $\beta$ as

$$b_n = \frac{\sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}\tau y(\tau) - \left(\sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}\tau\right)\left(\sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}y(\tau)\right)\bigg/ \sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}}{\sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}\tau^2 - \left(\sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}\tau\right)^2 \bigg/ \sum\limits_{\tau=t_1}^{n} \gamma^{n-\tau}}$$

If we consider the case of an infinite number of data; that is, $t_1 \to \infty$. Then,

$$\sum_{\tau=-\infty}^{n} \gamma^{n-\tau} = \frac{1}{1-\gamma}$$

$$\sum_{\tau=-\infty}^{n} \gamma^{n-\tau}\tau = \frac{n-\gamma n-\gamma}{(1-\gamma)^2}$$

$$\sum_{\tau=-\infty}^{n} \gamma^{n-\tau}\tau^2 = \frac{1}{(1-\gamma)^3}[\{n(1-\gamma)-\gamma\}^2 + \gamma]$$

(5-13)         $$b_n = \frac{(1-\gamma)^3}{\gamma}\left\{\sum_{}^{n} \gamma^{n-\tau}\tau y(\tau) - \left(n-\frac{\gamma}{1-\gamma}\right)\sum_{}^{n} \gamma^{n-\tau}y(\tau)\right\}$$

and

(5-14)         $$a_n = \bar{Y} - b_n \cdot \bar{n}$$

$$= (1-\gamma)\left[\sum_{}^{n} \gamma^{n-\tau}y(\tau)\left\{1 + \frac{(1-\gamma)^2}{\gamma}\left(n-\frac{\gamma}{1-\gamma}\right)^2\right\}\right.$$

$$\left. - \frac{(1-\gamma)^2}{\gamma}\left(n-\frac{\gamma}{1-\gamma}\right)\sum_{}^{n} \gamma^{n-\tau}\tau y(\tau)\right].$$

From these $a_n$ and $b_n$, we can forecast the next $\hat{y}(n+1)$ as,

(5-15)         $$\hat{y}(n+1) = a_n + b_n \cdot (n+1)$$

$$= \frac{(1-\gamma)^2}{\gamma}\left\{\sum_{}^{n} \gamma^{n-\tau}\tau y(\tau) - \left(n-\frac{2\gamma}{1-\gamma}\sum_{}^{n} \gamma^{n-\tau}y(\tau)\right)\right\}.$$

Brown [1] has proposed to use the double exponential smoothing when the time-series data are looked upon as linear. Suppose that the true process is lenear,

$$x(t) = \alpha + \beta \cdot t$$

and let $x_t$ be an observed datum, then the smoothed statistics are

$$S_t(x) = (1-\gamma)x_t + \gamma S_{t-1}(x)$$
$$S^2_t(x) = (1-\gamma)S_t + \gamma S^2_{t-1}(x) \ .$$

Then, the estimates of the coefficients are

$$\hat{\alpha}(t) = 2S_t(x) - S^2_t(x)$$
$$\hat{\beta}(t) = \frac{1-\gamma}{\gamma}[S_t(x) - S^2_t(x)]$$

and the forecast for $\tau$ periods ahead is

$$\hat{x}_\tau(t) = \hat{\alpha}(t) + \tau\hat{\beta}(\tau)$$
$$= \left(2 + \frac{\gamma\tau}{1-\gamma}\right)S_t(x) - \left(1 + \frac{\gamma\tau}{1+\gamma}\right)S^2_t(x) \ .$$

These statistics $S_t(x)$ and $S^2_t(x)$ can be rewritten as,

$$S_t(x) = (1-\gamma)\sum^t \gamma^{t-k}x_k$$
$$S^2_t(x) = (1-\gamma)^2\left\{(n+1)\sum^t \gamma^{t-k}x_k - \sum^t \gamma^{t-k}kx_k\right\} \ .$$

$$\therefore \quad \beta(t) = \frac{(1-\gamma)^3}{\gamma}\left[\sum^t \gamma^{t-k}kx_k - \left(t - \frac{\gamma}{1-\gamma}\right)\sum^t \gamma^{t-k}x_k\right\} \ .$$

This is the same as (5-13). For $a_n$, we can get the same result by taking into account the difference of time.

Thus, we can conclude that the smoothing method is the same with the discounting method with infinite data, and therefore has a sound structural basis.

# ReEFERENCES

[ 1 ]   Brown, R.G., *Forecasting and Prediction of Discreate Time Series*, Prentice-Hall, 1962.

[ 2 ]   Lindley, D.V., *Introduction to Probability and Statistic*, Cambridge University Press, 1965.

[ 3 ]   Raiffa, H. and Schlaifer, R., *Applied Statistical Decision Theory*, The M.I.T. Press, 1961.

[ 4 ]   Savage, L.J., *The Foundations of Statistics*, John Wiley & Sons, Inc., 1954.

[ 5 ]   Winters, P.R., "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, Vol. 6, No. 3, 1960.