ECONOMETRIC FORECASTING VIA DISCOUNTED LEAST SQUARES

Robert A. Agnew

Corporate Development Department
FMC Corporation
Chicago, Illinois

ABSTRACT

Simple direct smoothing formulas are derived for updating coefficient estimates and forecasts in a discounted least squares model. These formulas are the natural extensions of R. G. Brown's well-known smoothing formulas to a general econometric setting with arbitrary explanatory time series. The recursive updating process and its forecast error properties are illustrated via a simple, yet realistic numerical example.

1. INTRODUCTION

Discounted least squares (DLS) is a generalization of the familiar ordinary least squares (OLS) model of econometrics. DLS retains the basic OLS assumption of uncorrelated homoscedastic disturbances, but it introduces artificial heteroscedasticity by discounting past squared deviations in the least squares minimization. In effect, the procedure simply gives more weight to the most recent observations. OLS is a special case of DLS where the discount factor is equal to one.

Despite the ostensible artificiality of DLS, the technique actually provides considerable flexibility in model (coefficient) updating. The discount factor can be considered as a convenient lever to adjust the dynamism of the updating process. As a practical matter, the technique works well in some real-world, short-term forecasting applications, in the sense that it improves on OLS and compares favorably with other generalized techniques.

Although DLS doesn't seem to have generated much interest among econometricians, it does form the basis for exponential smoothing models. These smoothing models relate a dependent series to specified regular functions of time, and they yield neat, direct steady-state formulas for simultaneous coefficient updating and origin translating. This steady-state updating technique was originated by Brown [1]; a more modern, accessible reference is Montgomery and Johnson [3]. Brown's specialized version of DLS is widely known and used, but to our knowledge his approach hasn't been extended explicitly to a more general setting with arbitrary explanatory time series, although the concept of recursive updating is not new (see Odell and Lewis [4]). The objective of this paper is to explicitly extend Brown's formulas to a general econometric setting and to illustrate their potential practical value via a simple, yet realistic example.

DLS is, of course, a special case of generalized least squares (GLS), and hence the solution of a DLS model is conceptually routine, in that the normal equations are straightforwardly constructed and solved via matrix inversion. Unlike arbitrary GLS models, however, the DLS normal equations can be reduced to a simple direct form for coefficient updating, and matrix inversion can be avoided after initialization. We derive these direct "smoothing" formulas in Section 3 after reviewing the general DLS structure in Section 2. Section 4 reviews Brown's specialized version of DLS and provides a conceptual link to his steady-state smoothing formulas. In Section 5, we illustrate the use of DLS in a simple econometric model; although the particular model is somewhat crude, it provides us with a device to benchmark DLS forecast error performance. Section 6 elaborates on the key issue of selecting a specific value for the discount factor.

2. THE DISCOUNTED LEAST SQUARES MODEL

Let $y(1), \ldots, y(t)$ be observations of a time series for periods $1, \ldots, t$; form the t-dimensional column vector $\mathbf{y}(t) = (y(1), \ldots, y(t))'$. Let $\mathbf{x}(1), \ldots, \mathbf{x}(t)$ be a time series of explanatory n-dimensional column vectors for periods $1, \ldots t$; form the $t \times n$ matrix $\mathbf{X}(t) = [\mathbf{x}(1), \ldots, \mathbf{x}(t)]'$. Then, we have the usual linear model formulation

$$\mathbf{y}(t) = \mathbf{X}(t)\boldsymbol{\beta} + \mathbf{u}(t) \tag{1}$$

where $\mathbf{u}(t) = (u(1), \ldots, u(t))'$ is the t-dimensional disturbance vector and $\boldsymbol{\beta}$ is the unknown n-dimensional coefficient vector. The coefficient vector $\boldsymbol{\beta}$ is assumed to be time invariant.

As usual, we assume

$$\begin{aligned} &E(\mathbf{u}(t)) = \mathbf{0} \\ &E(\mathbf{u}(t)\mathbf{u}(t)') = \sigma^2\mathbf{I}(t) \end{aligned} \tag{2}$$

where σ^2 is an unknown scalar, $\mathbf{0}$ is the t-dimensional zero column vector, and $\mathbf{I}(t)$ is the $t \times t$ identity matrix. Rather than forming the usual (OLS) normal equations, however, we choose to minimize the discounted sum-of-squares $(\mathbf{y}(t) - \mathbf{X}(t)\mathbf{b}(t))'\mathbf{W}(t)(\mathbf{y}(t) - \mathbf{X}(t)\mathbf{b}(t))$ where $\mathbf{W}(t) = \text{diag}\{\delta^{t-1}, \ldots, \delta, 1\}$ is a diagonal $t \times t$ weight matrix incorporating the discount factor $\delta \in (0, 1]$. This minimization yields the DLS normal equations

$$\mathbf{X}(t)'\mathbf{W}(t)\mathbf{X}(t)\mathbf{b}(t) = \mathbf{X}(t)'\mathbf{W}(t)\mathbf{y}(t) \tag{3}$$

where $\mathbf{b}(t)$ is the DLS estimate of $\boldsymbol{\beta}$ at period t. Equation (3) can be written equivalently

$$\mathbf{b}(t) = \mathbf{S}(t)^{-1}\mathbf{v}(t) \tag{4}$$

where $\mathbf{v}(t) = \mathbf{X}(t)'\mathbf{W}(t)\mathbf{y}(t)$ and $\mathbf{S}(t) = \mathbf{X}(t)'\mathbf{W}(t)\mathbf{X}(t)$ is assumed nonsingular.

The effect of the weight matrix in the discounted sum-of-squares minimization is to put more emphasis on the most recent residuals. What we are doing is compensating for deficiencies in the static linear model (1); i.e., we don't really expect $\boldsymbol{\beta}$ to be time invariant so we are placing greater weight on the most recent observations.

Given that we have solved the normal equations (4) for $\mathbf{b}(t)$, we denote our forecast of $y(t + k)$, the observation k periods hence, by

$$\hat{y}_k(t + k) = \mathbf{x}(t + k)'\mathbf{b}(t) \tag{5}$$

assuming $\mathbf{x}(t+k)$ is known; in most econometric situations, of course, $\mathbf{x}(t+k)$ is itself a forecast derived by some other means. Equation (5) can be written equivalently

$$\hat{y}_k(t) = \mathbf{x}(t)'\mathbf{b}(t-k) \tag{6}$$

for $t > k$, and we define the k-period forecast error by

$$e_k(t) = y(t) - \hat{y}_k(t). \tag{7}$$

We will also refer in Sections 5 and 6 to the absolute error $a_k(t) = |e_k(t)|$ and the absolute percentage error $p_k(t) = 100\, a_k(t)/\hat{y}_k(t)$.

This section has outlined the basic assumptions and mechanics of DLS. From here on, our focus becomes computational. Our goal is to put the elements of equation (4) into a recursive format, which is why we have explicitly attached the time index to all of our variables.

3. RECURSIVE SMOOTHING FORMULAS

To develop recursive formulas, we first partition $\mathbf{X}(t)$, $\mathbf{y}(t)$, and $\mathbf{W}(t)$ as follows.

$$\mathbf{X}(t)' = [\mathbf{X}(t-1)',\ \mathbf{x}(t)]$$

$$\mathbf{y}(t)' = [\mathbf{y}(t-1)',\ y(t)]$$

$$\mathbf{W}(t) = \begin{bmatrix} \delta\mathbf{W}(t-1), \mathbf{0} \\ \mathbf{0}', \quad 1 \end{bmatrix}$$

where the variables indexed by $t-1$ are defined as in Section 2 over the first $t-1$ observations and $\mathbf{0}$ is the $(t-1)$-dimensional zero column vector.

Ordinary matrix multiplication yields the following recursive formulas for $\mathbf{S}(t)$ and $\mathbf{v}(t)$.

$$\begin{aligned} \mathbf{S}(t) &= \mathbf{X}(t)'\mathbf{W}(t)\mathbf{X}(t) \\ &= \delta\mathbf{X}(t-1)'\mathbf{W}(t-1)\mathbf{X}(t-1) + \mathbf{x}(t)\mathbf{x}(t)' \\ &= \delta\mathbf{S}(t-1) + \mathbf{x}(t)\mathbf{x}(t)' \end{aligned} \tag{8}$$

$$\begin{aligned} \mathbf{v}(t) &= \mathbf{X}(t)'\mathbf{W}(t)\mathbf{y}(t) \\ &= \delta\mathbf{X}(t-1)'\mathbf{W}(t-1)\mathbf{y}(t-1) + \mathbf{x}(t)y(t) \\ &= \delta\mathbf{v}(t-1) + \mathbf{x}(t)y(t). \end{aligned} \tag{9}$$

Equations (8) and (9) are the basic smoothing formulas for the elements of (4). However, the smoothing is still *indirect* and calculation of the coefficient estimate vector $\mathbf{b}(t)$ in (4) still requires matrix inversion at each update.

To obtain a *direct* smoothing formula, we note that

$$\begin{aligned} (\delta\mathbf{S}(t-1) + \mathbf{x}(t)\mathbf{x}(t)')\mathbf{b}(t) &= \delta\mathbf{v}(t-1) + \mathbf{x}(t)y(t) \\ &= \delta\mathbf{S}(t-1)\mathbf{b}(t-1) + \mathbf{x}(t)\,(e_1(t) + \mathbf{x}(t)'\mathbf{b}(t-1)) \end{aligned}$$

and upon rearranging terms that

$$\begin{aligned} \mathbf{S}(t)(\mathbf{b}(t) - \mathbf{b}(t-1)) &= \mathbf{x}(t)e_1(t) \\ \mathbf{b}(t) &= \mathbf{b}(t-1) + \mathbf{z}(t)e_1(t) \end{aligned} \tag{10}$$

where $\mathbf{z}(t) = \mathbf{S}(t)^{-1}\mathbf{x}(t)$. Equation (10) is a direct smoothing formula for updating $\mathbf{b}(t)$, although the calculation still involves matrix inversion at every stage.

Now (following Rao [5], problem 2.8, p. 33),

$$\mathbf{S}(t)^{-1}(\delta\mathbf{S}(t-1) + \mathbf{x}(t)\mathbf{x}(t)') = \mathbf{I}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. Hence,

$$\begin{aligned}\delta\mathbf{S}(t)^{-1}\mathbf{S}(t-1) &= \mathbf{I} - \mathbf{S}(t)^{-1}\mathbf{x}(t)\mathbf{x}(t)' \\ &= \mathbf{I} - \mathbf{z}(t)\mathbf{x}(t)'\end{aligned}$$

and upon post-multiplying each side by $\mathbf{S}(t-1)^{-1}$, we get

$$\mathbf{S}(t)^{-1} = \delta^{-1}(\mathbf{I} - \mathbf{z}(t)\mathbf{x}(t)')\mathbf{S}(t-1)^{-1} \tag{11}$$

so that $\mathbf{S}(t)^{-1}$ is obtainable directly from $\mathbf{S}(t-1)^{-1}$ via simple matrix multiplication once $\mathbf{z}(t)$ is known. To get a formula for $\mathbf{z}(t)$ in terms of $\mathbf{S}(t-1)^{-1}$, we post-multiply both sides of (11) by $\mathbf{x}(t)$ to obtain

$$\mathbf{z}(t) = \delta^{-1}(\mathbf{I} - \mathbf{z}(t)\mathbf{x}(t)')\mathbf{S}(t-1)^{-1}\mathbf{x}(t)$$

and upon rearranging terms we get

$$\mathbf{z}(t) = (\delta + \mathbf{x}(t)'\mathbf{S}(t-1)^{-1}\mathbf{x}(t))^{-1}\mathbf{S}(t-1)^{-1}\mathbf{x}(t). \tag{12}$$

To summarize, once we have obtained $\mathbf{S}(t_0)^{-1}$ and $\mathbf{b}(t_0)$ for some base period t_0 from (4), updates can be performed directly, without matrix inversion, via formulas (10), (11) and (12) together.

In this era of rapid, economical computation, avoidance of matrix inversion is not so important as it once was. Furthermore, single-equation econometric models usually don't involve large numbers of explicit coefficients because multicollinearity tends to render the estimation process unstable and meaningless; and, of course, multicollinearity presents the same problem in DLS as it does in OLS. Nevertheless, the simple, direct updating process embodied in (10), (11), and (12) is appealing and very easy to implement, especially in a matrix-oriented computer language like APL. The updating process is also very economical in terms of data storage requirements. Furthermore, equation (10) provides a conceptual bridge to the well-known steady-state smoothing formulas of R.G. Brown [1].

4. RELATIONSHIP TO BROWN'S STEADY-STATE FORMULAS

Brown [1] worked out steady-state results for the special case where $\mathbf{x}(t) = \mathbf{A}^t\mathbf{x}(0)$ for some fixed $n \times n$ transition matrix $\mathbf{A}$ and suitable starting vector $\mathbf{x}(0)$; see also Montgomery and Johnson [3] for a complete derivation. This special case of DLS includes all independent variables which are regular functions of time (e.g., polynomial and trigonometric functions of time), and it allows reduction of the direct smoothing equations to a very neat steady-state form. On the other hand, it is clear that general econometric modeling is not encompassed by this simplified framework.

Brown's special case can, of course, be treated within our general setup of Section 3, but there is no apparent simplification of the results or saving of update effort. Brown, however, obtained simplified results via an "origin shift" of the coefficients

$$\dot{\mathbf{b}}(t) = (\mathbf{A}')^t\mathbf{b}(t) = (\mathbf{A}^t)'\mathbf{b}(t). \tag{13}$$

Then, since $(\mathbf{A}')^{-t} = (\mathbf{A}^{-t})'$, we get

$$\hat{y}_k(t+k) = \mathbf{x}(t+k)'\mathbf{b}(t) = \mathbf{x}(t+k)'(\mathbf{A}')^{-t}\dot{\mathbf{b}}(t) \tag{14}$$
$$= (\mathbf{A}^{-t}\mathbf{x}(t+k))'\dot{\mathbf{b}}(t) = \mathbf{x}(k)'\dot{\mathbf{b}}(t)$$

so that forecasts depend only on the transformed coefficient estimate vector $\dot{\mathbf{b}}(t)$ and the "shifted" explanatory vector $\mathbf{x}(k)$.

The direct smoothing equation (10) becomes

$$(\mathbf{A}')^{-t}\dot{\mathbf{b}}(t) = (\mathbf{A}')^{-(t-1)}\dot{\mathbf{b}}(t-1) + \mathbf{z}(t)e_1(t). \tag{15}$$

$$\dot{\mathbf{b}}(t) = \mathbf{A}'\dot{\mathbf{b}}(t-1) + \dot{\mathbf{z}}(t)e_1(t)$$

where $\dot{\mathbf{z}}(t) = (\mathbf{A}')^t\mathbf{z}(t) = (\mathbf{A}^t)'\mathbf{z}(t)$. Now,

$$\dot{\mathbf{z}}(t) = (\mathbf{A}')^t\mathbf{S}(t)^{-1}\mathbf{x}(t) = (\mathbf{A}')^t\mathbf{S}(t)^{-1}\mathbf{A}^t\mathbf{x}(0) \tag{16}$$

$$= \dot{\mathbf{S}}(t)^{-1}\mathbf{x}(0)$$

$$\dot{\mathbf{S}}(t) = \mathbf{A}^{-t}\mathbf{S}(t)\,(\mathbf{A}')^{-t} = \mathbf{A}^{-t}\mathbf{S}(t)\,(\mathbf{A}^{-t})' \tag{17}$$
$$= \mathbf{A}^{-t}\left\{\sum_{j=1}^{t}\delta^{t-j}(\mathbf{A}^j\mathbf{x}(0))\,(\mathbf{A}^j\mathbf{x}(0))'\right\}(\mathbf{A}^{-t})'$$
$$= \sum_{j=1}^{t}\delta^{t-j}(\mathbf{A}^{-(t-j)}\mathbf{x}(0))\,(\mathbf{A}^{-(t-j)}\mathbf{x}(0))'$$
$$= \sum_{j=0}^{t-1}\delta^{j}(\mathbf{A}^{-j}\mathbf{x}(0))\,(\mathbf{A}^{-j}\mathbf{x}(0))'.$$

For a well-behaved transition matrix $\mathbf{A}$ and $\delta \in (0, 1)$, the limit $\dot{\mathbf{S}}(\infty) = \lim_{t\to\infty}\dot{\mathbf{S}}(t)$ exists and is generally approached rapidly. Hence, equation (15) rapidly takes the steady-state form

$$\dot{\mathbf{b}}(t) = \mathbf{A}'\dot{\mathbf{b}}(t-1) + \dot{\mathbf{z}}(\infty)e_1(t) \tag{18}$$

where $\dot{\mathbf{z}}(\infty) = \dot{\mathbf{S}}(\infty)^{-1}\mathbf{x}(0)$. This is a very neat result in that the matrix $\dot{\mathbf{S}}(\infty)$ is easily computed for a given discount factor and need only be inverted once to get the steady-state multiplier $\dot{\mathbf{z}}(\infty)$; then, given a starting vector $\dot{\mathbf{b}}(0)$, steady-state updates can occur routinely via equation (18). In practice, an initial, untransformed DLS estimate $\mathbf{b}(t_0)$ is obtained via equation (4) and $\dot{\mathbf{b}}(t_0) = (\mathbf{A}')^{t_0}\,\mathbf{b}(t_0)$ is used as $\dot{\mathbf{b}}(0)$ to start the steady-state update process (18); otherwise, the starting vector $\dot{\mathbf{b}}(0)$ is chosen judgmentally.

Besides its mathematical simplicity, the steady-state update process (18) has the additional desirable feature of very limited data storage requirements. Once computed, $\dot{\mathbf{z}}(\infty)$ is a fixed parameter vector and $\mathbf{A}$ is a fixed parameter matrix; besides these fixed parameters, one need only maintain the most recently updated coefficient estimate vector $\dot{\mathbf{b}}(t)$. It is worth noting that the generalized recursive procedure of Section 3 is just as parsimonious in terms of data storage requirements, although more routine updating is necessary (e.g., $\mathbf{S}(t)^{-1}$ and $\mathbf{z}(t)$, as well as $\mathbf{b}(t)$). The following two examples illustrate both Brown's steady-state updating process and the more general recursive updating process of Section 3.

EXAMPLE: SIMPLE EXPONENTIAL SMOOTHING

In this example, we assume $E(y(t)) = \beta$ for all t, where β is an unknown scalar constant. Hence, $n = 1$ and we have the simple scalar relationships $x(t) = A = 1$, $\dot{b}(t) = b(t)$, $\dot{S}(t) = S(t)$, and $\dot{z}(t) = z(t)$ for all t. Assuming $\delta \in (0, 1)$ and $\alpha = 1 - \delta$, we have

$$S(t) = \sum_{j=1}^{t} \delta^{t-j} = \sum_{j=0}^{t-1} \delta^{j} = (1 - \delta^{t})\alpha^{-1} \rightarrow \alpha^{-1}$$

$$z(t) = S(t)^{-1} \rightarrow \alpha.$$

Hence, the transient update equation ((10) or (15)) takes the form

$$b(t) = b(t-1) + (1 - \delta^{t})^{-1}\alpha e_1(t)$$

while the steady-state update equation (18) takes the familiar form

$$b(t) = b(t-1) + \alpha e_1(t).$$

In this simple situation $\hat{y}_k(t + k) = b(t)$ for all $k \geqslant 1$, so our forecasts for all future periods at time t are identical.

EXAMPLE: EXPONENTIAL SMOOTHING WITH TIME TREND

In this example, we assume

$$E(y(t)) = \beta_1 + \beta_2 t = \mathbf{x}(t)'\boldsymbol{\beta}$$

for all t, where $n = 2$, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ is an unknown vector consisting of intercept β_1 and trend (or slope) β_2, and $\mathbf{x}(t) = (1, t)'$. In Brown's setup, $\mathbf{x}(t) = \mathbf{A}^{t}\mathbf{x}(0)$ where $\mathbf{x}(0) = (1, 0)'$ and

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

It follows that

$$\mathbf{A}^{j} = \begin{bmatrix} 1 & 0 \\ j & 1 \end{bmatrix}$$

and $\mathbf{A}^{j}\mathbf{x}(0) = (1, j)'$ for any real integer j. Assuming $\delta \in (0, 1)$ and $\alpha = 1 - \delta$, we get

$$\dot{\mathbf{S}}(t) = \sum_{j=0}^{t-1} \delta^{j} \begin{bmatrix} 1 & -j \\ -j & j^2 \end{bmatrix} = \begin{bmatrix} f_0(t) & -f_1(t) \\ -f_1(t) & f_2(t) \end{bmatrix}$$

where

$$f_0(t) = \sum_{j=0}^{t-1} \delta^{j} = (1 - \delta^{t})\alpha^{-1} \rightarrow \alpha^{-1}$$

$$f_1(t) = \sum_{j=0}^{t-1} j\delta^{j} = (\delta f_0(t) - t\delta^{t})\alpha^{-1} \rightarrow \delta\alpha^{-2}$$

$$f_2(t) = \sum_{j=0}^{t-1} j^2\delta^{j} = ((1 + \delta) f_1(t) - t(t-1)\delta^{t})\alpha^{-1} \rightarrow \delta(1 + \delta)\alpha^{-3}.$$

Then,

$$\dot{\mathbf{z}}(t) = \dot{\mathbf{S}}(t)^{-1}\mathbf{x}(0)$$
$$= (f_0(t)f_2(t) - f_1(t)^2)^{-1}(f_2(t), f_1(t))' \to (\alpha(1+\delta), \alpha^2)'.$$

Hence, the transient update equation (15) takes the form

$$\dot{b}_1(t) = \dot{b}_1(t-1) + \dot{b}_2(t-1) + \dot{z}_1(t)e_1(t)$$
$$\dot{b}_2(t) = \dot{b}_2(t-1) + \dot{z}_2(t)e_1(t)$$

where

$$\dot{z}_1(t) = (f_0(t)f_2(t) - f_1(t)^2)^{-1}f_2(t)$$
$$\dot{z}_2(t) = (f_0(t)f_2(t) - f_1(t)^2)^{-1}f_1(t)$$

and the steady-state update equation (18) takes the familiar form

$$\dot{b}_1(t) = \dot{b}_1(t-1) + \dot{b}_2(t-1) + \alpha(1+\delta)e_1(t)$$
$$\dot{b}_2(t) = \dot{b}_2(t-1) + \alpha^2 e_1(t).$$

Since $\mathbf{z}(t) = (\mathbf{A}^{-t})'\dot{\mathbf{z}}(t) = (\dot{z}_1(t) - \dot{z}_2(t)t, \dot{z}_2(t))'$, we have $z_1(t) \simeq \dot{z}_1(\infty) - \dot{z}_2(\infty)t = \alpha(1+\delta) - \alpha^2 t$ and $z_2(t) \simeq \alpha^2$ for t sufficiently large. Hence, the general update equation (10) takes the form

$$b_1(t) = b_1(t-1) + (\dot{z}_1(t) - \dot{z}_2(t)t)e_1(t)$$
$$b_2(t) = b_2(t-1) + \dot{z}_2(t)e_1(t)$$

which indicates that the intercept update process would become increasingly volatile over time. This volatility, however, is immaterial to the forecasting process. Since $\dot{\mathbf{b}}(t) = (\mathbf{A}^t)'\mathbf{b}(t) = (b_1(t) + b_2(t)t, b_2(t))'$, the forecast for the k-th future period at time t is identically

$$\hat{y}_k(t+k) = b_1(t) + b_2(t)(t+k)$$
$$= (b_1(t) + b_2(t)t) + b_2(t)k$$
$$= \dot{b}_1(t) + \dot{b}_2(t)k.$$

Brown's steady-state approach is very attractive for DLS models which depend on regular functions of time. We think the smoothing formulas of Section 3 provide a straightforward extension of Brown's approach to general econometric models with arbitrary explanatory time series. The second example above indicates that the general update process (10) is likely to be less volatile over the long run if systematic growth components are removed from the various series beforehand by first-differencing or some other means. We have not found this to be an important point in practical econometric work, however. The example in the next section illustrates a simple model where business cycle factors predominate and drive the coefficient (forecast) update process.

5. A SIMPLE ECONOMETRIC EXAMPLE

To illustrate the efficacy of DLS in an econometric context, we put $y(t)$ equal to total U.S. new passenger car unit sales in quarter t and $\mathbf{x}(t) = (x_1(t), x_2(t))'$, where $x_1(t) \equiv 1$ and $x_2(t)$ is the motor vehicles and parts personal consumption component of real Gross National Product (1972\$) in quarter t. Both $y(t)$ and $x_2(t)$ are seasonally adjusted, annualized rates from the Bureau of Economic Analysis, U.S. Department of Commerce. The overall time interval ranges from the first quarter of 1970 to the last quarter of 1980. We use the 12 quarter interval from 1970:1 to 1972:4 to initialize the model. We use the 32 quarter interval from 1973:1 to 1980:4 to monitor the update process and to measure *ex post* forecast errors for 1, 2, 3, and 4 quarters ahead. The simulated ex post forecasts incorporate actual historical observations for the consumption series $x_2(t)$; in reality, future values would have to be estimated (ex ante) from some type of quarterly GNP forecasting model with the requisite level of disaggregation.

Now clearly the time series $y(t)$ and $x_2(t)$ should be highly correlated since most new passenger car sales (aside from business and government purchases) are imbedded in the broader personal consumption series, which also includes a net used car component, small trucks, other recreational vehicles, service parts, and so forth. On the other hand, the relationship between the two series is far from exact and may change over time, cyclically and perhaps secularly. Moreover, the new car sales series is widely followed as a cyclical indicator, and a simple mechanism to generate forecasts from a GNP component is very desirable.

Table 1 summarizes the average signed, absolute, and absolute percentage forecast errors 1, 2, and 4 quarters ahead (as defined in Section 2) for discount factors ranging from .1 to 1.0; the omitted 3 quarter ahead results are analogous. Our focus on simple arithmetic error measures is intuitive. Despite the least-squares basis of DLS, we think most people conceptualize forecast errors in these simple arithmetic terms.

The average errors for a first-order autoregressive (AR1) model are also included in Table 1. This GLS model (see Johnston [2], Chapter 9, and Theil [6], Section 6.3) is commonly used in econometric work when the OLS Durbin-Watson statistic turns out exceptionally small, as it did in this case over most of the simulated ranges. The AR1 solution procedure is also commonly known as the Cochrane-Orcutt technique.

Table 2 contains detailed quarter-by-quarter coefficient adjustments and forecasts for the case $\delta = .5$. Table 3 contains the full range of input data and intermediate results for that same case $\delta = .5$. This example is computationally very simple, so we just iterated the indirect equations (8), (9), and (4) to get updated coefficient estimates $\mathbf{b}(t) = (b_1(t), b_2(t))'$ for each quarter t. Table 2 and 3 entries have been suitably rounded for presentation purposes.

It is clear from the simulation results that OLS is significantly inferior to both AR1 and alternative DLS models in terms of average absolute forecast error, and furthermore that AR1 is inferior to the best DLS models with discount factors in the .5 neighborhood. Since the DLS solution procedure is also much simpler than AR1, the approach certainly seems worthy of consideration in many short-term forecasting situations where AR1 is routinely applied.

It is interesting to note in Table 1 that DLS forecast bias generally increases with the forecast horizon for a fixed discount factor and with the discount factor for a fixed forecast horizon. On the other hand, the average absolute forecast error, in this particular example, is approximately minimized across all the forecast horizons with discount factor in the .5 neighborhood.

TABLE 1 — *Summary Forecast Error Analysis*

	Average 1973:1-1980:4			Average 1973:2-1980:4			Average 1973:4-1980:4		
δ	e_1	a_1	p_1	e_2	a_2	p_2	e_4	a_4	p_4
.1	−.09	.56	6.7	.06	.82	22.4	−.27	.86	9.3
.2	−.05	.39	4.6	−.05	.48	5.5	−.33	.57	5.7
.3	−.04	.31	3.5	−.08	.35	3.8	−.34	.51	5.1
.4	−.05	.27	2.8	−.11	.30	3.1	−.36	.49	4.9
.5	−.07	.25	2.6	−.15	.30	3.0	−.40	.49	4.8
.6	−.11	.26	2.6	−.20	.31	3.1	−.45	.50	4.9
.7	−.17	.29	3.0	−.27	.36	3.6	−.54	.58	5.7
.8	−.29	.37	3.7	−.41	.45	4.5	−.67	.68	6.7
.9	−.51	.53	5.3	−.62	.63	6.2	−.84	.84	8.1
1.0(OLS)	−.84	.84	8.1	−.92	.92	8.7	−1.09	1.09	10.2
AR1	−.19	.35	3.5	−.32	.50	5.0	−.64	.74	7.0

TABLE 2 — *Detailed Results for* $\delta = .5$

Quarter	t	$b_1(t)$	$b_2(t)$	$x_2(t)$	$y(t)$	$\hat{y}_1(t)$	$\hat{y}_2(t)$	$\hat{y}_3(t)$	$\hat{y}_4(t)$
1972:4	12	2.67	0.158	56.5	11.6	NA	NA	NA	NA
1973:1	13	2.32	0.164	60.7	12.3	12.2	NA	NA	NA
1973:2	14	2.12	0.170	58.0	12.1	11.8	11.8	NA	NA
1973.3	15	−0.58	0.213	55.6	11.1	11.6	11.5	11.4	NA
1973.4	16	−5.92	0.306	51.7	9.8	10.5	10.9	10.8	10.8
1974:1	17	−2.47	0.244	47.9	9.3	8.7	9.6	10.3	10.2
1974:2	18	−3.45	0.262	47.9	9.0	9.2	8.7	9.6	10.3
1974:3	19	−3.32	0.257	49.9	9.4	9.6	9.7	9.4	10.1
1974:4	20	−4.18	0.274	42.3	7.4	7.5	7.6	7.8	7.0
1975:1	21	−2.86	0.249	43.7	8.2	7.8	7.9	8.0	8.2
1975:2	22	−3.76	0.264	44.2	7.7	8.2	7.9	8.0	8.1
1975:3	23	−0.68	0.194	49.8	8.9	9.4	9.5	9.5	9.5
1975:4	24	−0.01	0.179	52.2	9.3	9.4	10.0	10.1	10.1
1976:1	25	0.77	0.163	56.9	10.0	10.2	10.3	11.3	11.3
1976:2	26	1.62	0.145	57.0	9.8	10.0	10.2	10.4	11.3
1976:3	27	1.79	0.142	57.4	9.9	9.9	10.1	10.3	10.4
1976:4	28	0.83	0.160	58.0	10.2	10.0	10.0	10.2	10.4
1977:1	29	0.10	0.173	63.5	11.1	11.0	10.8	10.8	11.1
1977:2	30	−0.75	0.188	63.1	11.2	11.0	10.9	10.7	10.8
1977:3	31	0.38	0.168	63.2	10.9	11.2	11.0	11.0	10.7
1977:4	32	1.71	0.146	64.1	11.0	11.2	11.3	11.2	11.1
1978:1	33	0.72	0.161	62.4	10.7	10.8	10.9	11.0	10.9
1978:2	34	0.19	0.169	68.0	11.7	11.7	11.6	11.8	12.1
1978:3	35	0.22	0.168	65.6	11.2	11.3	11.3	11.3	11.4
1978:4	36	3.27	0.119	66.8	11.1	11.4	11.5	11.5	11.5
1979:1	37	2.66	0.130	66.4	11.4	11.2	11.4	11.4	11.4
1979:2	38	1.85	0.142	59.4	10.3	10.4	10.4	10.2	10.2
1979:3	39	2.82	0.128	60.8	10.7	10.5	10.6	10.5	10.4
1979:4	40	−1.71	0.196	60.3	9.8	10.6	10.4	10.5	10.5
1980:1	41	−2.48	0.210	62.1	10.6	10.5	10.8	10.7	10.7
1980:2	42	−1.98	0.202	47.0	7.5	7.4	7.5	8.8	8.5
1980:3	43	−1.28	0.192	51.5	8.8	8.4	8.3	8.4	9.4
1980:4	44	−1.17	0.189	54.6	9.1	9.2	9.0	9.0	9.0

TABLE 3 — *Input Data and Intermediate Results for* $\delta = .5$

Quarter	t	$x_2(t)$	$y(t)$	$S_{11}(t)$	$S_{12}(t)$	$S_{22}(t)$	$v_1(t)$	$v_2(t)$
1970:1	1	39.0	8.8	1.00	39.0	1,521	8.8	343
1970:2	2	40.3	9.1	1.50	59.8	2,385	13.5	538
1970:3	3	40.2	9.0	1.75	70.1	2,808	15.8	631
1970:4	4	33.1	6.9	1.88	68.1	2,500	14.8	544
1971:1	5	42.6	10.0	1.94	76.7	3,065	17.4	698
1971:2	6	44.1	9.9	1.97	82.4	3,477	18.6	786
1971:3	7	46.1	10.4	1.98	87.3	3,864	19.7	872
1971:4	8	49.3	10.6	1.99	93.0	4,362	20.4	959
1972:1	9	49.7	10.5	2.00	96.2	4,651	20.7	1,001
1972:2	10	51.2	10.7	2.00	99.3	4,947	21.1	1,048
1972:3	11	52.2	10.8	2.00	101.8	5,198	21.3	1,088
1972:4	12	56.5	11.6	2.00	107.4	5,791	22.3	1,199
1973:1	13	60.7	12.3	2.00	114.4	6,580	23.4	1,346
1973:2	14	58.0	12.1	2.00	115.2	6,654	23.8	1,375
1973:3	15	55.6	11.1	2.00	113.2	6,418	23.0	1,305
1973:4	16	51.7	9.8	2.00	108.3	5,882	21.3	1,159
1974:1	17	47.9	9.3	2.00	102.1	5,235	20.0	1,025
1974:2	18	47.9	9.0	2.00	98.9	4,912	19.0	944
1974:3	19	49.9	9.4	2.00	99.4	4,946	18.9	941
1974:4	20	42.3	7.4	2.00	92.0	4,262	16.8	783
1975:1	21	43.7	8.2	2.00	89.7	4,041	16.6	750
1975:2	22	44.2	7.7	2.00	89.0	3,974	16.0	715
1975:3	23	49.8	8.9	2.00	94.3	4,467	16.9	801
1975:4	24	52.2	9.3	2.00	99.4	4,958	17.8	886
1976:1	25	56.9	10.0	2.00	106.6	5,717	18.9	1,012
1976:2	26	57.0	9.8	2.00	110.3	6,107	19.2	1,065
1976:3	27	57.4	9.9	2.00	112.5	6,348	19.5	1,101
1976:4	28	58.0	10.2	2.00	114.3	6,538	20.0	1,142
1977:1	29	63.5	11.1	2.00	120.6	7,301	21.1	1,276
1977:2	30	63.1	11.2	2.00	123.4	7,632	21.7	1,345
1977:3	31	63.2	10.9	2.00	124.9	7,810	21.8	1,361
1977:4	32	64.1	11.0	2.00	126.6	8,014	21.9	1,386
1978:1	33	62.4	10.7	2.00	125.7	7,901	21.6	1,361
1978:2	34	68.0	11.7	2.00	130.8	8,574	22.5	1,476
1978:3	35	65.6	11.2	2.00	131.0	8,591	22.5	1,473
1978:4	36	66.8	11.1	2.00	132.3	8,758	22.3	1,478
1979:1	37	66.4	11.4	2.00	132.6	8,788	22.6	1,496
1979:2	38	59.4	10.3	2.00	125.7	7,922	21.6	1,360
1979:3	39	60.8	10.7	2.00	123.6	7,658	21.5	1,330
1979:4	40	60.3	9.8	2.00	122.1	7,465	20.5	1,256
1980:1	41	62.1	10.6	2.00	123.2	7,589	20.9	1,286
1980:2	42	47.0	7.5	2.00	108.6	6,003	17.9	996
1980:3	43	51.5	8.8	2.00	105.8	5,654	17.8	951
1980:4	44	54.6	9.1	2.00	107.5	5,808	18.0	972

Moreover, there is little difference in performance for discount factors in the fairly broad range .3 to .7, indicating robustness with respect to choice of this judgmental parameter. For very small discount factors, the coefficients evidently adjust rapidly to eliminate bias but the resulting procedure is excessively volatile. For very large discount factors (e.g., OLS), the coefficients adjust slowly and the resulting procedure, although smoother, is quite biased. For mid-range discount factors (around .5 in this example), the resulting procedure adjusts rapidly enough to keep bias small while also providing sufficient stability to yield relatively small average absolute errors; i.e., it is a good blend of dynamism and stability.

Table 2 shows how the coefficients and forecasts fluctuated over the simulation interval 1973:1 to 1980:4 for the quasi-optimum case $\delta = .5$. Note that the slope estimate $b_2(t)$ is far more stable than the intercept estimate $b_1(t)$. Note also the characteristic pattern of negative intercepts during the 1973-75 recessionary contraction, a pattern commencing once again in late 1979. The forecasts are, for the most part, pretty close to the actual new car sales.

This crude forecasting model can probably be improved econometrically. For example, we have ignored the possible incremental impact of strikes and interest rate fluctuations on new car sales. We have also made no attempt to differentiate domestics from imports. Furthermore, we are well aware that, as a practical matter, bias is often reduced in OLS models by adjusting the forecasts with judgmental "add factors". The purpose of this example is merely to indicate the potential value of DLS.

6. CHOOSING THE DISCOUNT FACTOR

Whenever DLS is employed in any form, choice of the discount factor becomes an important issue. In the example of Section 5, ex post simulation showed $\delta = .5$ to be a good choice. This is not always the case, however. We've analyzed other cases where greater inherent volatility in the dependent series caused a larger discount factor to be approximately optimum. In any event, this kind of ex post simulation analysis is obviously a good idea when plenty of historical data exists. Otherwise, one must resort to judgmental criteria.

From a judgmental standpoint, the author would tend to focus on discount factors between .3 and 1.0 (OLS) to maintain some degree of stability in the coefficient update process. If OLS on the historical data yields a very low Durbin-Watson statistic, one would naturally choose a lower discount factor in this range to enhance the dynamism of the process. In this case, DLS provides an alternative to the ubiquitous AR1 procedure.

Another judgmental criterion is the "moving average equivalence" principle (see Montgomery and Johnson [3], p. 52). The idea is to equate (very approximately) a DLS model with an equivalent "rolling" OLS model, over the most recent m historical periods, in terms of the "average age of the data" $\sum_{j=0}^{m-1} j/m = (m-1)/2$. Based on simple exponential smoothing, we get the equivalence relationship $\delta = (m-1)/(m+1)$ or $m = (1+\delta)/(1-\delta)$; this equivalence is fairly crude when extended to general DLS models, but it is nonetheless appealing. For example, if we wanted the average age of our data to be 3 periods, we would choose $\delta = .5$; for 4 periods, $\delta = .6$; and for 9 periods, $\delta = .8$. In many short-term forecasting problems with quarterly data, an average age of 2 to 8 quarters seems intuitively appealing; this corresponds to $.33 \leqslant \delta \leqslant .78$.

Another possibility is to adjust the discount factor itself over time. Trigg and Leach [7] introduced an intuitive technique to adjust the discount factor in response to a buildup of systematic forecast errors (see also Montgomery and Johnson [3], pp. 175-179). Specifically, they make δ a decreasing function of $|\bar{e}_1(t)|/\bar{a}_1(t)$, where $\bar{e}_1(t)$ and $\bar{a}_1(t)$ are simple exponentially smoothed values of the one-period-ahead signed and absolute forecast errors respectively; this adjustment decreases δ to increase responsiveness when there is a buildup of signed forecast error (or bias) in relation to the mean absolute forecast error. The Trigg-Leach adaptive technique may induce instability into some DLS models, and we haven't tested it in a general context. Nevertheless, the technique is widely employed as an enhancement to simple exponential smoothing, its volatility is easily constrained, and the general approach seems like an avenue worth exploring in practical forecasting situations.

As we've seen, there is no hard and fast rule for choosing a discount factor, although it appears that good values can be estimated in specific situations and there are even possibilities for adjusting the factor adaptively. The fact that DLS incorporates a judgmental factor should not be considered a negative. The discount factor merely adds practical flexibility.

REFERENCES

[1] Brown, R.G., *Smoothing, Forecasting and Prediction of Discrete Time Series* (Prentice-Hall, Englewood Cliffs, NJ, 1962).

[2] Johnston, J., *Econometric Methods* 2nd Edition (McGraw-Hill, New York, 1972).

[3] Montgomery, D.C. and L.A. Johnson, *Forecasting and Time Series Analysis* (McGraw-Hill, New York, 1976).

[4] Odell, P.L. and T.O. Lewis, "Best Linear Recursive Estimation," *Journal of the American Statistical Association*, *66*, 893-896 (1971).

[5] Rao, C.R., *Linear Statistical Inference and Its Applications* 2nd Edition (John Wiley & Sons, New York, 1973).

[6] Theil, H., *Principles of Econometrics* (John Wiley & Sons, New York, 1971).

[7] Trigg, D.W. and A.G. Leach, "Exponential Smoothing with an Adaptive Response Rate," *Operational Research Quarterly*, *18*, 53-59 (1967).