Sankhyā : The Indian Journal of Statistics
1986, Volume 48, Series B, Pt. 3, pp. 401-426

A STOCHASTIC APPROXIMATION ALGORITHM FOR A CLASS OF NONLINEAR DYNAMICAL ECONOMETRIC SYSTEM

By SUKESH K. GHOSH
University of Waterloo, Canada

SUMMARY. This article proposes a novel approach to estimation of a stochastic differential equation system in economics. An algorithm is developed and applied, along the lines of Dvoretzky (1956) and engineering practices, using discrete point approximations. In it the probability becomes *exactly* one (rather than *almost* one) that the estimate tends to the true parameter value with increasing iterations.

1. Introduction

There seems to be a revival of interests in the continuous time econometric models propounded by Koopmans (1950) and recommended by, for instance, Phillips (1959), and Bergstrom (1966). This is clear from recent economists' efforts to fit stochastic differential equations to economic systems. Different approaches have been followed. One of the approaches is that of approximating the continuous model by a computationally simple form preserving the linearity of the original model. Examples are Sargan's (1974) and Wymer's[1] (1972) work in both of which variables are in absolute form but which differ in the degree of the differential equation. For it is a differential equation of the first degree in the endogenous variables in the first work, while it is that of a general degree in the second work. Another approach initiated by Phillips (1974, 803–19) replaces the continuous model by a discrete one, which, in effect, is exact, based on Lagrangean interpolation of exogenous variables between data points. However, the model is nonlinear in parameters, not in variables[2] and the predetermined variables are, as in Wymer (1972), in absolute form. The most recent approach is pioneered by Robinson and exemplified in two of his very powerful papers (1976a) and

AMS (1984) *subject classification :* 62P20.

Key words and phrases : Stochastic discrete equations iterations 'sure convergence".

[1] The empirical application of the theory (Wymer, 1972) was the subject matter of a later paper by Wymer (1973).

[2] Phillips (1972) earlier estimated a small example of a stochastic differential equation system, non-linear in parameter only, by an exact discrete method without special emphasis on the exogenous variables. It should be added that the method of numerical differentiation used in Phillips (1974) has also been used by Bergstrom and Wymer (1974) in their paper.

(1976b). In these a system of simultaneous linear differential equations involving more than first degree differentials in both the endogenous *and*, which is remarkable, exogenous variables is replaced by a discrete approximation "that is more conveniently handled in the frequency domain" by means of Fourier transformations. The second paper, more specifically, suggests a frequency-domain class of instrumental variables estimators for *all or part* of an open system of linear differential equations with computationally preferred properties in those situations where these can be used.

The purpose of this paper is to present yet another approach to estimation of a stochastic differential equation system. We suggest a stochastic approximation algorithm for a class of non-linear dynamical econometric system. The non-linearity[3] is in variables[4], not in parameters, unlike Phillips' (1972, 1974) nonlinearity (in parameters, not in variables). Discrete point approximation of a very simple nature is tried including that commonly used in economics, i.e., differentials replaced by differences. Thus our approximations are similar to, although not exactly the same as those used in, for instance, (Bergstrom, 1966, pp. 176; Wymer, 1972, 567–8; Phillips, 1972, pp. 1032)[5]. The virtue of our approximations is in their ability to unfold the adaptive nature of economic systems. The algorithm proposed (Section 2.1 and Section 2.2) is one in which, as the number of iterations increases, under certain general conditions the probability becomes *exactly* one (rather than *almost* one) that the estimate tends to the true parameter

---

[3] The *form* of the nonlinear equation, presented in (1) below, is chosen on the basis of experience of the economic system and may be validated by an equation or equations, containing functions without parameters, which constitute qualitative statements about the system to be identified. Given this form, the model structure is fully known except for a parameter vector $a = (a_1, \ldots, a_s, a_{s+1}, \ldots, a_S)^T$, T being the transpose (see (7)), embedded in the structure. This structure, although linear in parameters, is nonlinear in variables (see equation (6)). The non-linearity makes the transition back to the original model (equation (1)) difficult, if not impossible. This explains why, after having derived the approximation, no attempt is made in this paper to go back to the original model.

When however, such a lack of attempt is not vindicated and future research can perhaps improve the situation, limited justification for it may still be advanced in terms of those much less complex situations, e.g., linear (non-differential equation) systems, where transition from reduced form to the original structual equations is not possible and yet the reduced forms do serve a number of useful purposes.

[4] This kind of non linearity in variables is like that used in Fisher (1961) in a non-differential equation context.

[5] These references approximate definite integrals by discrete point values.

value.[6] At this point, the estimate is error free, thus no need arises for statements on the variance of estimate to be made, variances being zero, unlike usual econometric practices. For instance, a small simulated example on consumption (Section 2.3) shows that only 98 iterations done in 2.06 seconds produced parameter estimates with error as low as 0·0000072, or virtually nothing. This algorithm is a direct application of some stochastic convergence results of Dvoretzky (1956) as exemplified in the engineering literature in especially the contributions of Ho and Lee (1965), Lee (1964), Sakrison (1967), Saridis and Stein (1968) and DeFigueiredo and Netravalli (1970, 1971).

One important feature of our work is assuming measurement errors in variables at the point of their actual calculation, i.e., at discrete points, given by discrete random process characteristics. The form of these errors is arbitrarily (though commonly) assigned however. These errors arise from discrepancies between theoretical concepts and their representative measures. Examples of these concepts may be capital, income, profit rate, money - some of the basic ingredients in, for instance, a model of capital accumulation, money, and growth in the dynamic context. Memory is still perhaps not hazy about a fierce capital controversy initiated by the Cambridge school which went on for quite sometime in one of the most recent pasts of our economic thinking.

2. The models

2.1. *The simplest model with one endogenous, one exogenous, and one error variable and a general order dynamics.* Consider a non-linear recursive economic system characterized by an endogenous variable x, an exogenous variable z, a random input $\bar{u}$ and an s-th order dynamics of the form :

$$\frac{d^s x}{dt^s} = f\left(x, \frac{dx}{dt}, \ldots, \frac{d^{s-1}x}{dt^{s-1}}\right) + \bar{b}z + \bar{u} \qquad \ldots \quad (1)$$

all the variables x, z and $\bar{u}$ referring to a time point k to be denoted by $x(k)$, $z(k)$ and $\bar{u}(k)$. The choice of the form of f, as indicated in footnote 3 of the

[6] A word of caution must be sounded here. The convergence takes no account of possible discrete approximation bias, which will generally not vanish however large the number of iterations might be. But the extent of such bias will be similar to that obtained by, for instance, Wymer. Since Wymer (1972) in his paper discusses this in detail, we leave it out here.

It is perhaps interesting to add that based on discrete approximation, the estimates may not be consistent for any residual process but they may be perferred because there may not be any reason to take 'pure noise' sufficiently seriously to justify the added cost of estimating the 'exact' model.

introduction, is fixed and given by qualitative characteristics of the system, but its structure does depend on parameters a's to be explained below. $\bar{b}$ is the coefficient for $\bar{z}$. Using discrete approximations to the differentials generating the new variables y, z and u and the new coefficient b for $\bar{b}$, we have the approximate[7] model :

$$y(k+1) = \begin{bmatrix} 0 : I \\ \cdots\cdots \\ a^T \end{bmatrix} y(k) + \begin{bmatrix} 0 \\ \cdots\cdots\cdots\cdots \\ bz(k) + \iota u(k) \end{bmatrix} \quad \dots \quad (2)$$

where y is a column vector of S elements, 0 a null column vector of $S-1$ elements, I an identity matrix of size $S-1$, a^T is a row vector of S elements, T being the symbol of transpose, $z(k)$ and $u(k)$ are scalars, and ι a column vector of S 1's, and

$$S = s + \binom{s+1}{2} + \binom{s+2}{3} + \dots + \binom{s+p-1}{p} \quad \dots \quad (3)$$

$$y(k) = (y^1(k), y_2(k), \dots, y^s(k), y^{s+1}(k), \dots, y^S(k))^T \quad \dots \quad (4)$$

$$b = (b_1, b_2, \dots, b_s, b_{s+1}, \dots, b_S)^T \quad \dots \quad (5)$$

$$\begin{aligned} \psi(y(k)) &= a^T\phi(y(k)) \; (= a^T y(k)) \\ &= \sum_{i_1=1}^{s} a_{i_1} y^{i_1}(k) + \sum_{i_1=1}^{s} \sum_{i_2=1}^{i_1} a_{i_1 i_2} y^{i_1}(k) y^{i_2}(k) \\ &\quad + \dots + \sum_{i_1=1}^{s} \sum_{i_2=1}^{i_1} \dots \sum_{i_p=1}^{i_{p-1}} a_{i_1 i_2 \dots i_p} y^{i_1}(k) \dots y^{i_p}(k) \end{aligned} \quad \dots \quad (6)$$

$$\begin{aligned} a &= (a_1, \dots, a_s, a_{11}, a_{21}, \dots, a_{ss}, \dots, a_{11\cdots1}, a_{21\cdots1}, \dots, a_{ss\cdots s})^T \\ &= (a_1, \dots, a_s, a_{s+1}, \dots, a_S)^T \end{aligned} \quad \dots \quad (7)$$

$$\begin{aligned} \phi(y(k)) = y(k) &= (y^1(k), \dots, y^s(k), y^1(k)y^1(k), \dots, y^s(k)y^s(k), \\ &\underbrace{y^1(k) \dots y^1 k)}_{p \text{ times}}, \dots, \underbrace{y^s(k) \dots y^s(k)}_{p \text{ times}})^T. \end{aligned} \quad \dots \quad (8)$$

What was a general nonlinearity $f(\ )$ in the differentials has been replaced by that of a specific class, that is, a polynomial Ψ in equivalent discrete time

[7] See Appendix A for details, and for the relationship between x and y at discrete points. Note that non-linearity increases the number of discrete points over linearity. For example, in our method the number of such points is S (see (3) above), which is larger than the number s to be used in a linear situation such as De Figueiredo and Netravalli (1970). However, it remains true that the first s points are the most crucial.

observations $y(k)$[8, 9] involving the parameters a's. This will be more clear below.

Assume scalar linear measurements of endogenous variables $y(k)$ and exogenous variables $z(k)$ subject to errors $v(k)$ and $\zeta(k)$:

$$\bar{Y}(k) = h^T y(k) = (0\ 0\ \ldots\ 1)y(k) \qquad \ldots\ (9)$$

$$Y(k) = \bar{Y}(k)+v(k) \qquad \ldots\ (10)$$

$$Z(k) = z(k)+\zeta(k). \qquad \ldots\ (11)$$

The model in the discrete approximation form can now be interpreted as follows. It is assumed that the transformed system is forced by a control function $z(k)$, $k = 0, 1, \ldots$, which is a sample realization of a stationary stochastic process and by the disturbance $u(k)$, $k = 0, 1, \ldots$, a similar process. The system dynamics characterized by the matrix

$$\begin{bmatrix} 0 & \vdots & I \\ \cdots & \cdots & \cdots \\ & a^T & \end{bmatrix}$$

is assumed stable[10] and the initial state reaction in y is assumed to have subsided before estimation beings. Thus $y(k)$, $k = 0, 1, \ldots$, is also a sample function of a stationary process. The measurements used for estimation are $Y(k)$ and $Z(k)$, $k = 0, 1, \ldots$, the noise contaminated system output and input

---

[8] For this particular replacement, the system (2)−(8) has a property: $\phi(w+y) = \xi(w) + \chi(w)\phi(y)$

for all w and admissible values of y, when the elements of the S-vector $\xi(w)$ and the coefficients of the $S \times S$ matrix $\chi(w)$ depend only on w (and not on y). (Once the vector ϕ is chosen, these elements and coefficients become known functions of w). It can be easily proved that with $\phi\ (y\ (k))$ as defined in (8), which follows from the representation of $\psi(y(k))$ by a finite number of terms of the Taylor Series expansion of $\psi(y(k))$ around $y(k) = 0$, the system (2)−(8) possesses the above property. An alternative approximation would be by projection of $\psi(y(k))$ into a subspace of its space.

[9] Instead of the typical representation of $\psi(y(k))$ as in (6) with $\phi(y(k))$ as in (8), one might have a representation by trigonometric polynomials in the variables $y(k)$, i.e.,

$$\psi(y(k)) = a^T\phi(y(k))$$

$$a = (a_1, a_2, \ldots, a_S)^T$$

$$\phi(y)(k)) = (\sin y(k), \cos y(k), \ldots, \sin Sy(k), \cos Sy(k))^T.$$

The detailed treatment given to (2)−(8) would equally apply to the system based on the above substitutions. However this would involve the lengthy use of some standard trigonometric relations. The interested reader may find some useful hints in (Netravalli and De Figueiredo, 1971) in connection with a simpler situation.

[10] For this, what is required is a set of characteristic roots of the matrix that lie within the unit circle. For more on stability of the continuous and discrete approximation models, see Appendix B.

as it were. We can assume, though without loss of generality, that the structure of (2) involving $y(k+1)$, $y(k)$ and $z(k)$ is controllable and that involving $y(k+1)$, $y(k)$ and $\bar{Y}(k)$ is observable.[11] We further assume :

(i) $z(k)$[12], $u(k)$, $v(k)$ and $\zeta(k)$ are mutually independent sequences of statistically independent random variables. $u(k)$, $v(k)$ and $\zeta(k)$ are random noise elements with variances σ_u^2, σ_v^2, and σ_ω^2 and means given by $E(u^s(k))^t = \alpha_t$; $E(v^t(k)) = \beta_t$; $E(\zeta(k)) = 0$; $k, t \in J$.[13]

(ii) $$\lambda_0 = \inf |\lambda_{k+S-1}| > 0$$

where λ_{k+S-1} is the minimum characteristic root of

$$\begin{bmatrix} E(Cy(k)y^T(k)/y(k-1) = \xi) + HE(Z^cZ^{cT})H^T & HE(Z^cZ^{cT}) \\ E(Z^cZ^{cT})H^T & E(Z^cZ^{cT}) \end{bmatrix}$$

for any ξ, where Z^c is defined in (16) below and C and H are the coefficient matrices of $y(k)$ and Z^c in the reduction or expansion of $\bar{Y}^c$ (see (16)).[14,15]

(iii) The means $E(z(k))(=\tau)$, $E(u^S(k))(=\alpha)$, and $E(v(k))\,(=\beta)$ are known[16], and second, third, and fourth moments of $z(k)$, $u(k)$, $v(k)$ and $\zeta(k)$ are finite.

The following lemma will be immediately useful.

Lemma 1 : *The system as specified by* (2) *and* (9) *is equivalent to the following difference equation* :

$$\begin{aligned} &\bar{Y}(k+S) - a_S\bar{Y}(k+S-1) - \ldots - a_1\bar{Y}(k) \\ &= b_S^0 z(k+S-1) + b_{S-1}^0 z(k+S-2) + \ldots + b_1^0 z(k) \\ &+ d_S^0 u(k+S-1) + d_{S-1}^0 u(k+S-2) + \ldots + d_1^0 u(k) \qquad \ldots \quad (12) \end{aligned}$$

---

[11] For more on the controllability and observability of system structures, see Ho and Lee [1965, 93-110].

[12] Even though z is a random variable, its dependence on k produces a sequence of random variables.

[13] J is the set of non-negative integers.

[14] For more hints on C and H, see footnote 21 and Appendix C. The detailed structure of these matrices is very cumbersome. Also it is not directly needed in this paper. So we leave it out here, even though one can easily work it out. Notice that Assumption (ii) validates the use of the convergence results.

[15] This resembles requiring the underlying matrix of Phillips' [1972, 1025-8] Minimum Distance Estimators to have the smallest characteristic root with a positive sign.

[16] The assumption that the mean values are known is perhaps no more unreasonable than the one that they are zero used in many econometric contexts.

where $b^0 = Tb;$

$d^0 = T(1\ 1\ \dots\ 1)^{\tau};$

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & a_1 & 0 \\ \cdot & \cdot & & & & a_1 & a_2 & 0 \\ \cdot & \cdot & & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \cdot & \cdot & \cdot \\ \cdot & a_1 & a_2 & a_3 & \dots & a_{S-3} & a_{S-2} & 0 \\ a_1 & a_2 & a_3 & a_4 & \dots & a_{S-2} & a_{S-1} & 0 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 1 \end{bmatrix} \quad \dots \ (13)$$

Proof: Since

$$\bar{Y}(k) = (0\ 0\ \dots\ 0\ 1)y(k)^{17}$$

so

$$y^S(k) = \bar{Y}(k).$$

Now $$\bar{Y}(k+1) = (0\ 0\ \dots\ 1)\left\{\begin{bmatrix}(0 : I)y(k) \\ \dots\dots\dots\dots \\ \psi(y(k))\end{bmatrix} + \begin{bmatrix} 0 \\ \dots\dots\dots\dots \\ bz(k)+u(k)\end{bmatrix}\right\}.$$

Thus $$a_1y^1(k)+\dots+a_Sy^S(k) = \bar{Y}(k+1)-b_Sz(k)-u(k)$$

in which y^i's ($i = 1, 2, \dots, S$) represent the successive elements of the ϕ vector (see (8) and S is as in (3)).

Similarly, $\bar{Y}(k+2)$ can be expressed as follows:

$$a_S(a_1y^1(k)+\dots+a_Sy^S(k))+(a_1y^2(k)+\dots+a_{S-1}y^S(k))$$
$$= \bar{Y}(k+2)-(a_1b_1+\dots+a_Sb_S)z(k)-b_Sz(k+1)-(a_1+\dots+a_S)u(k)-u(k+1).$$

As for $\bar{Y}(k+3)$, after some manipulations we get

$$0 = \bar{Y}(k+3)-a_S\bar{Y}(k+2)-a_{S-1}\bar{Y}(k+1)-b_Sz(k+2)-(a_1b_1+\dots+a_{S-1}b_{S-1})$$
$$\times z(k+1)-(a_1b_2+\dots+a_{S-2}b_{S-1})z(k)-u(k+2)-(a_1+\dots+a_{S-1})$$
$$\times u(k+1)-(a_1+\dots+a_{S-2})u(k)-(a_1y^3(k)+\dots+a_{S-2}y^S(k)).$$

Thus, finally, after an overwhelming amount of tedious algebra, we get

$$\bar{Y}(k+S)-a_S\bar{Y}(k+S-1)-\dots-a_1\bar{Y}(k)$$
$$= b_S[z(k+S-1)-a_Sz(k+S-2)-\dots-a_2z(k)]+(a_1b_1+\dots+a_Sb_S)$$
$$\times z(k+S-2)+\dots+(a_1b_{S-2}+\dots+a_3b_S)z(k+1)+(a_1b_{S-1}+a_2b_S)z(k)$$
$$+[u(k+S-1)-a_Su(k+S-2)-\dots-a_2u(k)]+(a_1+\dots+a_S)u(k+S-2)$$
$$+\dots+(a_1+a_2+a_3)u(k+1)+(a_1+a_2)u(k). \quad \dots \ (14)$$

[17] The proof that follows is a nonlinear extension of the linear case as treated by Lee (1964, 90·3).

From above, we have

$$b_S^0 = b_S$$
$$b_{S-1}^0 = -b_S a_S + (a_1 b_1 + \ldots + a_S b_S)$$
$$b_{S-2}^0 = -b_S a_{S-1} + (a_1 b_2 + \ldots + a_{S-1} b_S)$$
$$\vdots \qquad \vdots$$
$$b_1^0 = -b_S a_2 + (a_1 b_{S-1} + a_2 b_S)$$
$$d_S^0 = 1$$
$$d_{S-1}^0 = -a_S + (a_1 + \ldots + a_S)$$
$$d_{S-2}^0 = -a_{S-1} + (a_1 + \ldots + a_{S-1})$$
$$\vdots \qquad \vdots$$
$$d_1^0 = -a_2 + (a_1 + a_2)$$

which can be cast in the matrix form (13). Thus is equation (12) derived from (14) using (13). Q.E.D.

The main theme of this note is then presented in the following theorem :

Theorem : *Under the condition and assumptions of this section, the stochastic approximation algorithm for the recursive estimation of the parameter vector*

$$\theta^T = [a^T b_0^T] \qquad \ldots \quad (15)$$

is as given in (26). *The parameter estimates provided by this algorithm are such that with probability one, these converge asymptotically to the true values, and the mean square error of estimates converges to zero.*[18]

Proof : Define

$$Z^{C^T}(k+S-1) \triangleq [z(k), \ldots, z(k+S-1)]$$
$$\bar{Y}^{c^T}(k+S-1) \triangleq [Y(k), \ldots, Y(k+S-1), z(k), \ldots, z(k+S-1)]$$
$$V^T(k+S-1) = [V_1^T(k+S-1) \; \vdots \; V_2^T(k+S-1)] \triangleq [v(k), \ldots, v(k+S-1);$$
$$\zeta(k), \ldots, \zeta(k+S-1)] \qquad \ldots \quad (16)$$
$$U^T(k+S-1) \triangleq [u(k), \ldots, u(k+S-1)]$$
$$W^T(k+S-1) \triangleq [\bar{Y}^{c^T}(k+S-1) + V^T(k+S-1)]$$
$$= [Y(k), \ldots, Y(k+S-1); Z(k), \ldots, Z(k+S-1)].$$

[18] See footnote 6 for some caveat.

Then, from (10) and (12), solve for $Y(k+S)$ thus :

$$\bar{Y}(k+S) = \bar{Y}^{c^T}(k+S-1)\theta + U^T(k+S-1)d^0 \quad \dots \quad (17)$$

$$Y(k+S) = W^T(k+S-1)\theta + \eta(k+S) \quad \dots \quad (18)$$

where[19]

$$\eta(k+S) = v(k+S) - W^T(k+S-1)\theta + U^T(k+S-1)d^0. \quad \dots \quad (19)$$

For a stochastic approximation algorithm, one may minimize the criterion :

$$E(\eta(k+S))^2 = E(Y(k+S) - W^T(k+S-1)\theta)^2 \quad \dots \quad (20)$$

with respect to θ and build the estimated gradient

$$G = W(k+S-1)Y(k+S) - W(k+S-1)W^T(k+S-1)\hat{\theta}$$

into the algorithm such that for the linear regression function

$$E[G \mid \hat{\theta}(k-1)] = g(\hat{\theta}(k-1)).$$

The root occurs at $\hat{\theta}(k-1) = \theta$. These considerations lead to certain specific measurements for $Y(k+S)$ and $W(k+S-1)$, all of which are implied (in the present case) in G and then in the algorithm. These measurements[20] are given by :

$$E(Y(k+S)) = E(\psi(y(k+S-1))) + \alpha_1 + \beta_1 + b_{ST} \quad \dots \quad (21)$$

and

$$Y^c(k+S-1) = \tilde{A}^* y(k+S-1) + \phi(V_1(k+S-1)) \quad \dots \quad (22)$$

where

$$\tilde{A}^* = \begin{bmatrix} \tilde{A}^*_{11} & 0 & \dots & 0 \\ \tilde{A}^*_{21} & \tilde{A}^*_{22} & \dots & 0 \\ \tilde{A}^*_{p1} & \tilde{A}^*_{p2} & \dots & \tilde{A}^*_{pp} \end{bmatrix} \quad \dots \quad (23)$$

and

$\tilde{A}^*_{jj}$ is an identity matrix of size $\binom{s+j-1}{j} \times \binom{s+j-1}{j}$ for $1 \leqslant j \leqslant p$;

and $\tilde{A}^*_{ij}$ is a $\binom{s+i-1}{i} \times \binom{s+j-1}{i}$ matrix with either zero or measurement

---

[19] $\eta(k+S)$ below has the following statistical properties :

$E(\eta(j)) = 0 \qquad j = 0, 1, 2 \,..$

$E(\eta(i)\,\eta(j)) = \begin{cases} \text{finite} & |i-j| \leqslant S \\ 0 & |i-j| > S \end{cases}$

$E(W(j-1)\,\eta(j)) \neq 0 \qquad j = 1, 2, \,..$

[20] The underlying calculations are based on an extension of De Figueiredo and Netravalli (1970, 204-5) and Netravalli and De Figueiredo (1971, 20-30). See Appendix C for some hints.

noise components (V_1's) as its terms. Typically, a term, for instance, of $\tilde{A}^*_{21}$ of $\tilde{A}^*$ is

$$\tilde{A}^*_{21} = \begin{bmatrix} 2V_1^1(k+S-1) & 0 & \ldots\ 0 \ldots & 0 \\ V_1^2(k+S-1) & V_1^1(k+S-1) & \ldots\ 0 \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 2V_1^2(k+S-1) & \ldots\ 0 \ldots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \ldots\ \cdot \ldots & 2V_1^s(k+S-1) \end{bmatrix} \quad \ldots \quad (24)$$

It is assumed for the sake of the convergence proof that

(iv) $E(A^*)$ is a known and constant nonsingular matrix, and

$E(A^*_{ij}\, A^*_{kl})$ are known constants for relevant i, j, k and l.

The algorithm[21] is given below, with $\hat{\theta}(k-1)$ the estimate of θ at the $(k-1)$-th stage, and ρ a sequence of non-negative numbers :

$$\hat{\theta}(k+s) = \hat{\theta}(k-1)+\rho\left(\frac{k-1}{s+1}\right)\{W(k+S-1)[Y(k+S)-W^T(k+S-1)\hat{\theta}(k-1)]\} \quad \ldots \quad (25)$$

for $k = 1,\ s+2,\ 2s+3, \ldots$, which spacing ensures the uncorrelatedness of the disturbance terms. It is assumed that :

(v) $$\sum_1^\infty \rho(j) = \infty$$

$$\sum_1^\infty \rho^2(j) < \infty$$

again for the sake of the convergence proof. This algorithm may be shown to converge to the minimizing value $\hat{\theta}$ of $E(\eta(k+S))^2$ of (20). But $\hat{\theta}$ will not

[21] The algorithm (see (25) above) says that the estimate of the vector θ at time $k+S$ based on the receipt of the new information $Y(k+S)$ has two parts : (i) the old estimate at time $(k-1)$; (ii) a correction factor proportional to the difference between the actual measurement $Y(k+S)$ and the estimated measurement $W^T(k+S-1)\,\hat{\theta}(k-1)$ based on the old estimate. Since the residue between the actual and predicted measurements arises from the errors of estimate and also from the random terms $u(t)$, the part $\left(\frac{\rho(k-1)\,W}{s+1}\right)$ can be taken to be a weighting factor which fairly spreads the residue in line with the confidence one places on it. (See page 15 for more on ρ. An alternative way to look at (25) above is to consider it as a gradient scheme which goes in the direction of minimizing the instantaneous estimation error of the measurement. Note that (25) has θ built into it which includes measuring ψ by $Y(k+S)$: equation (21), and $W(k+S-1)$ by $Yc(k+S-1)$: (10) and (22).

be unbiased for θ. For an unbiased estimator one obtains the following algorithm,[22,23] (the dimensions, $S\times S$, $2S\times 2S$, $S\times 1$, or $2S\times 1$, of some matrices or vectors are shown below their symbols and/or signs):

$$\hat{\theta}(k+s) = \hat{\theta}(k-1)+\rho\left(\frac{k-1}{s+1}\right)\Bigg\{W(k+S-1)[Y(k+S)-W^T(k+S-1)\hat{\theta}(k-1)]$$

$$+\begin{bmatrix} \sigma_v^2 I & \vdots & 0_{S\times S} \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0_{S\times S} & \vdots & \sigma_w^2 I \end{bmatrix}_{2S\times 2S}\hat{\theta}(k-1)-\begin{bmatrix} \sigma_u^2\, Dd^0 \\ \cdots\cdots \\ 0_{S\times 1} \end{bmatrix}\Bigg\}_{2S\times 1} \quad \dots \quad (26)$$

in which the bias term, with reference to (18), is

$$E(W(k+S-1)\eta(k+S-1)) = -\begin{bmatrix} \sigma_v^2\ I & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & \sigma_w^2\ I \end{bmatrix}\theta+\begin{bmatrix} \sigma_u^2\ Dd^0 \\ \cdots\cdots \\ 0 \end{bmatrix} \quad \dots \quad (27)$$

and

$$D = \begin{bmatrix} 0 & 0 & 0 & 0\ ..\ 0 & 0 \\ 1 & 0 & 0 & 0\ ..\ 0 & 0 \\ \overset{S}{\Sigma}\ a_i & 1 & 0 & 0\ ..\ 0 & 0 \\ a_S\ \overset{S}{\Sigma}\ a_i+\overset{S-1}{\Sigma}\ a_i & \overset{S}{\Sigma}\ a_i & 1 & 0\ ..\ 0 & 0 \\ \{(a_{S-1}+a_S^2)\overset{S}{\Sigma}a_i+a_S\overset{S-1}{\Sigma}a_i+\overset{S-2}{\Sigma}\ a_i\} & \{a_S\overset{S}{\Sigma}a_i+\overset{S-1}{\Sigma}\ a_i\} & \overset{S}{\Sigma}a_i & 1\ ..\ 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \pi & \cdot & \cdot & \cdots 1 & 0 \end{bmatrix} \quad \dots \quad (28)$$

where

$$\pi = \{a_4+a_5a_S+a_6(a_{S-1}+a_S^2)+a_7(a_{S-2}+2a_{S-1}a_S+a_S^3)$$
$$+a_8(a_{S-3}+2a_{S-2}a_S+3a_{S-1}^2a_S+a_{S-1}^2+a_S^4+\dots)\}\overset{S}{\Sigma}\ a_i$$
$$+\{a_5+a_6a_5+a_7(a_S^2+a_{S-1})+a_S(a_{S-2}+2a_Sa_{S-1}+a_S^3+\dots)\ \overset{S-1}{\Sigma}\ a_i$$
$$+\dots+\overset{S-(S-3)}{\Sigma}\ a_i.$$

[22]This is an extension of the algorithm developed by Saridis and Stein (1968, 522-3) in the linear case. See Appendix C for some hints.

[23]For the practicability of this algorithm from the point of view of economic data availability, see footnote 27 to follow.

Specifically, the matrix D arises as the coefficient of $U(k+S-1)$ in the expansion[24] of $W(k+S-1)$. Its structure is triangular, bounded on one side by a hypotenuse of unities. Crucial in this structure are the elements of the first column. There is a simple rule to obtain these elements however. For instance, the element in a particular position is obtained by successively multiplying by $a_S, a_{S-1}, \ldots$ the elements preceding it but below the element of unity, adding these results, and finally adding to this sum the sum total of a number of a's. The number of these a's is the maximum S less the number of positions the particular element is below the $\sum_1^S a_i$ position. For example, the fifth element is obtained by multiplying the fourth and the third elements by a_S and a_{S-1} respectively, adding these products and finally adding to them $\sum^{S-2} a_i$, a sum of a's equal in number to S minus 2 which is the number of positions the fifth element is below the third element. The third element is $\sum^S a_i$.[25]

It is a simple extension of Saridis and Stein (1968, 52-3) to show that the conditions required for a convergence proof of Dvoretzky (1956, Sec. 8) are satisfied for the algorithm in (26). Thus

$$E|\hat{\theta}[\underset{i\to\infty}{i}(S+1)]-\theta|^2 \to 0$$

and

$$\Pr\{\underset{i\to\infty}{\hat{\theta}}[i(S+1)] = \theta\} = 1.$$

2.2. *The generalized model with many variables and general order dynamics.* The analogue of (2) in the case of m endogenous variables and n exogenous variables is:

$$\underset{mS\times 1}{y(k+1)} = \left[\begin{array}{c:c} \underset{m(S-1)\times m}{0} & \underset{m(S-1)\times m(S-1)}{I} \\ \hdashline \underset{m\times m}{A_1} \quad \underset{m\times m}{A_2} \ \ldots & \underset{m\times m}{A_S} \end{array}\right] \underset{mS\times 1}{y(k)} + \left[\begin{array}{c} 0 \\ \hdashline Bz(k)+u(k) \end{array}\right]$$

(An example of this structure in the purely linear case is to be found in equation (2) of Wymer, 1972). A's are the analogues of a's of (2), even though they are now matrices. They jointly signify dynamic simultaneity. B is a

[24] As one expands $W(k+S-1)$ there will b two other coefficient matrices, one, i.e., C which is the coefficient of $y(k)$ and the other, i.e., Π of $Z^e(k+S-1)$, both of which can be easily calculated just as D. An exact knowledge of these coefficients will not be needed except when one wants to see if Dvoretzky's (1956, Section 8) conditions hold for the convergence of the algorithm. See especially the last paragraph of this section for more on the convergence.

[25] On the whole, once the elements of the first column of D are obtained, those of the other columns are derived very easily as shown in the body of D itself. .

matrix of dimension $mS\times n$. The dimensions, e.g., $m(S-1)\times m$ of the other matrices or vectors are shown below their symbols. ι is a column vector of mS elements.

The algorithm of (26) can be suitably modified to incorporate the above changes in the scope of the problem. $\hat{\theta}$ will be of dimension $(mS+nS)\times m$; W of dimension $(mS+nS)\times 1$; Y of dimension $1\times m$; $\begin{bmatrix}\sigma_v^2 I & \vdots & 0\\ \cdots & \vdots & \cdots\\ 0 & \vdots & \sigma_u^2 I\end{bmatrix}$ of dimension $(mS+nS)\times(mS+nS)$; D[26] of $mS\times mS$; and d of $mS\times m$. The last follows from the dimension $mS\times mS$ of T and $mS\times m$ of d (for $d^0 = Td$). Finally the 0 matrix under $\sigma_u^2\ Dd^0$ will have the dimension $nS\times m$.

2.3. *Testing of the algorithm on a simple simulated case.* We shall conclude this article by a simple test of the algorithm presented in equation (26). We shall assume no exogenous variable in the approximate model equation (2). The value of s will be assumed to be 2. Thus S, which is equal to $s+\binom{s+1}{2}+\ldots+\binom{s+p-1}{p}$, will be 5. As a result of these simplifications, equation (26) will reduce to

$$\hat{\theta}(k+2) = \hat{\theta}(k-1)+\rho\left(\frac{k-1}{3}\right)\{W(k+4)[Y(k+5)-W^T(k+4)\hat{\theta}(k-1)] \\ +\sigma_v^2 I\hat{\theta}(k-1)-\sigma_u^2 Dd^0\}$$

where $k = 1, 4, 7, 10, \ldots$.

In above, $W(k+4)$ is a column vector of elements $Y(k)$, $Y(k+1)$, ..., $Y(k+4)$; $W^T(k+4)$ is the transpose of $W(k+4)$; I is an identity matrix of size 5×5; d^0 is a column vector of elements a_1, a_1+a_2, $a_1+a_2+a_3$, $a_1+a_2+a_3+a_4$, and 1; and D is a matrix of elements as follows:

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \sum_1^5 a_i & 1 & 0 & 0 & 0 \\ a_5\sum_1^5 a_i+\sum_1^4 a_i & \sum_1^5 a_i & 1 & 0 & 0 \\ (a_4+a_5^2)\sum_1^5 a_i+a_5\sum_1^4 a_i+\sum_1^3 a_i & a_5\sum_1^5 a_i+\sum_1^4 a_i & \sum_1^5 a_i & 1 & 0 \end{bmatrix}$$

[26] The structure of D will change in the following ways: The single elements O's and 1's will respectively be 0 matrices and 1 (identity) matrices; and a's will be the various A matrices. Similar matrix implications apply to the matrix T whose new dimension is discussed below.

From (9) and (10), $Y(k)$ is obtained from $y(k)$ by the relation

$$\bar{Y}(k) = (0 \quad 0 \quad 0 \quad 0 \quad 1)y(k)$$

and

$$Y(k) = \bar{Y}(k)+v(k).$$

Thus the last element of $y(k)$ becomes the effective observation on $\bar{Y}(k)$. Due to this, from (2),

$$\begin{aligned}\bar{Y}(k+1) &= a_1 y^1(k)+a_2 y^2(k)+a_3[y^1(k)]^2+a_4[y^1(k)y^2(k)]+a_5[y_2(k)]^2+u(k)\\ &= \psi(y(k))+u(k)\end{aligned}$$

and, iteratively, we get other values, e.g., $\bar{Y}(k+2)$, $\bar{Y}(k+3)$, ... from similar relations.[27]

In the simulation experiment, we proceed as follows: We choose ·6, ·4, ·4, —·2, ·25 for the actual values of a's in $\theta = (a_1\, a_2\, a_3\, a_4\, a_5)^T$. Given these, and using the normal probability table, we can find out actual values of $\bar{Y}(k)$'s from, for instance, the following relation

$$\bar{Y}(k+1) = \psi(y(k))+\alpha+N\sigma_u$$

where N (= 1·96) has been chosen to be the 5% probability point of the normal variable.[28] Of course, we need knowledge about $y^1(k)$ and $y^2(k)$, $\alpha\ (= E(u(k))$, $\sigma_u\ (= \sqrt{V(u(k))}$, and $v(k)$ before $\bar{Y}(k+1)$ and then $Y(k+1)$ can be determined. For this we assume :

$$\begin{aligned} y^1(k) &= ·2132\\ y^2(k) &= ·2134\\ \alpha &= -0·02\\ V(u(k)) &= 0·02\\ v(k) &= 0·25\\ V(v(k)) &= 0·33\end{aligned}$$

We have borrowed $y^1(k)$ and $y^2(k)$ from Phillips' (1973, p. 1038) sample data on consumption calculated as the mathematical expectation (for the

[27] It should be clear that the algorithm needs only a limited number of basic observations like $y^1(k)$ and $y^2(k)$ based on which any number of $\bar{Y}(k+1)$ for various values of k may be obtained. Such observations as $y^1(k)$ and $y^2(k)$ are ordinarily available from economic time series so that data availability should not be any bigger problem for the algorithm than for any other econometric estimation method.

[28] We have tried 30 other points of the standard normal error variable N randomly chosen from page 484 of Random Normal Numbers of the Chemical Rubber Company *Handbook of Tables for Probability and Statistics*, (2nd Ed.) edited by W. H. Beyer. The results are briefly reported in footnote 30.

equilibrium consumption level) of the first twenty five and the second twnety five observations respectively, scaled by the factor 10^{-2} for computational conveniences. The other assumed data, although arbitrarily chosen, bear some resemblance with those used by Saridis and Stein (1968, p. 518) and Netravalli and DeFigueiredo (1971, p. 331). Note that a knowledge of $\sigma_v^2 (= V(u(k))$ is needed to complete computation of the algorithm.

Going back to the algorithm, we come across the weighting factor $\rho\left(\frac{k-1}{s+1}\right)$. This can be replaced by an objective and appropriate matrix $P(k+s)$ given recursively by

$$
\begin{aligned}
P(k+s) = {} & P(k-1) - P(k-1)W(k+S-1) \\
& \times \{W^T(k+S-1)P(k-1)W(k+S-1)+1]^{-1} \\
& \times W^T(k+S-1)P(k-1).
\end{aligned}
$$

This represents a version of the least squares estimator given in Lee (1964, 49–59). It has been proved (Ho, 1963, 152–154) that $P(k)$ behaves asymptotically as $1/k$. Therefore, assumption (v) is fulfilled in the limit and convergence of the algorithm can still be achieved in this least squares case.

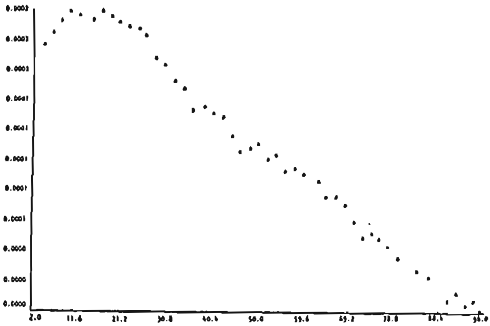

Fig. 1.

The computational work for the algorithm was done on the IBM 360 of the University of Waterloo. The initial estimates $\hat{\theta}(0)$ for $k = 1$ were the simple least squares estimates based on 50 simulated observations generated from the actual parameter values stated before. It took only 2·06 seconds with 98 iterations before the algorithmic estimates converged to the actual parameter values with the associated error as low and negligible as 0·0000072. The path of convergence is shown in the accompanying graph of plotted points (Figure 1) where errors (defined as maximum absolute differences between exact and computed parameter values) are shown along the vertical axis and the number of iterations along the horizontal axis.[29]

The success of the algorithm as revealed from the empirical test above should not mislead one into believing that the time taken would always be the small. Indeed, with multiple exogenous variables introduced into a network of recursive relationships involving multiple endogenous variables such as those that might be relevant in some economic contexts, the time taken might be more than the 2·06 seconds of the above limited example. For it is sometimes contended that the convergence of stochastic algorithms is generally not as fast as one would want it to be.[30] But one should not forget that the limiting algorithm is error proof for practical purposes. Thus even if it may take somewhat longer, it can perhaps be tolerated. However there are ways by which the computational time can be reduced.

One is by suitably choosing the initial parameter estimates; others may be by suitably choosing the variables so the associated measurement errors are not large, and so on. Extensive research is needed on these and other aspects of the algorithms in relation to economic problems.

3. Conclusion

The paper has been on the "Sure" convergence algorithm. It has developed that both theoretically and, within the limits of the simulated example, empirically. Due to its recursive nature it is capable of generating data to feed any number of iterations, given even a very limited number of initial data. In the simulated example, only two initial (time-series) y values (see p. 14 for values of $y^1(k)$ and $y^2(k)$) were used. Yet these two made possible

[29] The relationship of errors to selected numbers of iteration is to be found in Table 1 to follow.

[30] As mentioned in footnote 28, we have tried 30 other values for N, the standard normal error variable, defined in page 14, chosen from Random Normal Numbers Table. We like to report here only the broad results. For values of N ranging between −2.098 and 0.464, the maximum absolute error or difference between actual and computed parameter values ranged between less than 0.0002 and 0.0005 at 2,000 iterations requiring on the average five seconds of computational time. Thus the increase in the number of iterations indicates that the convergence process tends to be slow. However, the computational time taken is still small.

2,000 iterations of the algorithm as mentioned in footnote 30. Indeed there could be any larger number of iterations run for the sake of larger precision of results. Thus the small sample difficulty of economic problems is

TABLE 1. MAXIMUM ABSOLUTE DIFFERENCES BETWEEN ACTUAL AND COMPUTED PARAMETER VALUES AT SELECTED ITERATIONS FOR A VALUE OF THE STANDARD NORMAL ERROR VARIABLE = 1.96

no. of iterations	maximum absolute error
1	0.00283
10	0.00019
20	0.00018
30	0.00016
40	0.000126
50	0.000109
60	0.000091
70	0.000060
80	0.000038
90	0.000016
98	0.0000072

not a problem with this discrete-point algorithm, and since, as in engineering applications, the algorithm has to be iterated until the error practically converges to zero, and the algorithm under the given conditions surely converges, there does not seem to be a need for a statement on variances of estimates to be made. Of course this does not mean that such a statement can not be made but that that would take us beyond the scope of this paper.

Appendix

A. A representation for (1) is obtained by having $y_1 = x$ and writing (1) as:

$$\begin{aligned} Dy_1 &= (y_2-y_1)/\alpha_0+(\bar{\delta}_1\alpha_0^S\bar{z}+\alpha_0^S\bar{u})/\alpha_0 \\ Dy_2 &= (y_3-y_2)/\alpha_0+(\bar{\delta}_2\alpha_0^S\bar{z}+d_0^S\bar{u})/\alpha_0 \\ &\vdots \\ Dy_{S-1} &= (y_S-y_{S-1})/\alpha_0+(\bar{\delta}_{S-1}\alpha_0^S\bar{z}+\alpha_0^S\bar{u})/\alpha_0 \\ Dy_S &= R(y_1, y_2, \ldots, y_S, \bar{z}, \bar{u})/\alpha_0 \end{aligned} \qquad \ldots \quad (A.1)$$

where $D \equiv d/dt$, α_0 is constant, R is a function of its arguments such that it conforms to (1), and the relationship between S and s is as given in (3) of the text, i.e., $S = s+\binom{s+1}{2}+\binom{s+2}{3}+\ldots+\binom{s+p-1}{p}$. The problem is to determine R. (The superscripted y's of the text are replaced by the subscripted y's to cope with the detailed use of symbols here).

From (A.1), write,

$$(D+1/\alpha_0)y_1-y_2/\alpha_0 = (\bar{b}_1\alpha_0^S\bar{z}+\alpha_0^S\bar{u})/\alpha_0$$

$$(D+1/\alpha_0)y_2-y_3/\alpha_0 = (\bar{b}_2\alpha_0^S\bar{z}+\alpha_0^S\bar{u})/\alpha_0$$

$$\vdots$$

$$(D+1/\alpha_0)y_{S-1}-y_S/\alpha_0 = (\bar{b}_{S-1}\alpha_0^S\bar{z}+\alpha_0^S\bar{u})/\alpha_0$$

$$\alpha_0 Dy_S = R.$$

Eliminate $y_2, \ldots, y_S$ and get, because of $D(\bar{b}_i\bar{z}+\bar{u}) = 0$, $i = 1, 2, \ldots, S-1$,

$$\alpha_0^S D(D+1/\alpha_0)^{S-1}y_1 = R. \qquad \ldots \quad \text{(A.2)}$$

Then

$$R = \alpha_0^S D^S y_1+\{\alpha_0^S D(D+1/\alpha_0)^{S-1}y_1-\alpha_0^S D^S y_1\}$$

$$= -\alpha_0^S f(y_1, Dy_1, \ldots, D^{S-1}y_1)+\alpha_0^S\bar{b}\bar{z}+\alpha_0^S\bar{u}$$

$$+\{\alpha_0^S D(D+1/\alpha_0)^{S-1}-\alpha_0^S D^S\}y_1.$$

To express R in terms of $y_1, y_2, \ldots$, note from (A.1) that

$$Dy_1 = (y_2-y_1)/\alpha_0+(\bar{b}_1\alpha_0^S\bar{z}+\alpha_0^S\bar{u})$$

$$D^2y_1 = D(y_2/\alpha_0)-D(y_1/\alpha_0) = (y_3-y_2)/\alpha_0^2-(y_2-y_1)/\alpha_0^2$$

$$= (y_3-2y_2+y_1)/\alpha_0^2 \quad \text{etc., etc.}$$

In general

$D \equiv (E-1)/\alpha_0$ where E is the advancing operator, i.e., $Ey_k = y_{k+1}$.

Then,

$$D^k y_1 = (E-1)^k/\alpha_0^k y_1$$

$$= (1/\alpha_0^k)\left\{y_{k+1}-\binom{k}{1}y_k+\ldots\right\};\ k = 0, 1, \ldots, S-1.$$

Hence,

$$R = -f\left(y_1, \left(\frac{E-1}{\alpha_0}\right)y_1, (\bar{b}_1\alpha_0^S\bar{z}+\alpha_0^S\bar{u})/\alpha_0 \left(\frac{E-1}{\alpha_0}\right)^2 y_1, \ldots, \left(\frac{E-1}{\alpha_0}\right)^{S-1} y_1\right)$$

$$+\alpha_0^S b_S z+\alpha_0^S u+\{(E-1)E^{S-1}-(E-1)^S\}y_1$$

$$= -f\left(y_1, \left(\frac{E-1}{\alpha_0}\right)y_1, (b_1 z+u)/\alpha_0 \left(\frac{E-1}{\alpha_0}\right)^2 y_1, \ldots, \left(\frac{E-1}{\alpha_0}\right)^{S-1} y_1\right)$$

$$+b_S z+u+\left\{(S-1)y_S-\binom{S}{2}y_{S-1}+\binom{S}{3}y_{S-2}-\ldots-(-1)^S y_1\right\} \quad \ldots \quad \text{(A.3)}$$

where

$$\alpha_0^S \bar{b}_i = b_i,\ i = 1, 2, \ldots, S$$
$$\bar{z} = z$$
$$\alpha_0^S \bar{u} = u.$$

We have assumed in the text that the nonlinearity of the function f is such that (A.3) ultimately reduces to

$$R = a^T y(k) + bz(k) + u(k)$$

where T is the symbol for transpose and a and $y(k)$ are as defined in (7) and (8) respectively of the text.

To get difference equation approximation, replace the following value at the t-th time point:

$$Dy_k(t)$$

by

$$\{y_k(t+1) - y_k(t)\}/\alpha_0$$

in equation (A.1). This gives, in view of the equation in (A.1) for $k = 1, 2, \ldots, S-1$,

$$y_k(t+1) - y_k(t) \cong y_{k+1}(t) - y_k(t)$$

or,

$$y_k(t+1) \cong y_{k+1}(t).$$

Thus, (A.1) leads to

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{S-1} \\ y_S \end{bmatrix}_{t+1} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ & & & & \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_S \end{bmatrix}_t + \begin{bmatrix} b_1 z + u \\ b_2 z + u \\ \vdots \\ \vdots \\ R' \end{bmatrix}$$

where the subscripts $t+1$ and t indicate the time points at which things are recorded and

$$\begin{aligned} R' &= R + y_S(t) \\ &= -f + \left\{ S y_S - \binom{S}{2} y_{S-1} + \ldots - (-1)^S y_1 \right\} \\ &\quad + b_S z + u \\ &= \psi + b_S z + u \end{aligned}$$

where ψ is as defined in (6) of the text. The above may be cast into the form of equation (2) of the text at the point of time $t = k$.

B. We should perhaps expand on the stability point in relation to the continuous as well as the discrete approximation models. The stability of nonlinear differential and difference equations is a very complex matter and even if the continuous model is stable, it is not obvious that the discrete approximation will be and vice versa. However the following observation can be made which is applicable to our case.

From our Appendix equation (A.1), with $R = a^T y + bZ + u$, we have

$$Dy = \frac{1}{\alpha_0}\begin{bmatrix} -1 & 1 & 0 & . & . & 0 & 0 \\ 0 & -1 & 1 & . & . & 0 & 0 \\ \vdots & & & & & & \\ 0 & . & . & . & . & -1 & 1 \\ a_1 & . & . & . & . & a_{S-1} & a_S \end{bmatrix} y + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{bz+u}{\alpha_0} \end{bmatrix} \quad \ldots \quad \text{(i)}$$

$$= \frac{1}{\alpha_0} Ay + F, \text{ say.}$$

The approximating difference scheme replaces Dy by $\{y(t+1)-y(u)\}/\alpha_0$. So (i) gives

$$\frac{y(t+1)-y(t)}{\alpha_0} = \frac{1}{\alpha_0} Ay(t)+F$$

or

$$y(t+1) = (I+A)y(t)+\alpha_0 F. \quad \ldots \quad \text{(ii)}$$

The relationship between the stability of (i) and that of (ii) is now a question of: Do the eigen-values of A (in (i)) have negative real parts? And do the eigen-values of $I+A$ (in (ii)) lie within the unit circle?

But the eigen-values of $I+A$ are just "I+(eigen-values of A)". Hence stability for (ii) is assured if eigenvalues of A lie in the unit circle "shifted left by unity".

Note, however, that if eigen-values associated with (i) are $\lambda_1, \lambda_2, \ldots$, (i.e., these are the eigen-values of "$\frac{1}{\alpha_0} A$"), then the eigen-values of "A" are just $\alpha_0\lambda_1, \alpha_0\lambda_2, \ldots$, and, presumably, α_0 is small to justify the approximation

$$Dy \cong \frac{y(t+1)-y(t)}{\alpha_0}.$$

Hence, if $\lambda_1, \lambda_2, \ldots$, lie in left-half plane (so (i) is stable) *and* α_0 is sufficiently small, the eigen-values of A (which are $\alpha_0\lambda_1, \alpha_0\lambda_2, \ldots$) will lie in the shifted circle, so (ii) will be stable too.

C. Equation (21) is established as follows :

$$E(Y(k+S)) = E[\bar{Y}(k+S)+v(k+S)]$$

$$= E[(0\ 0 \ldots 1)y(k+S)]+E(v(k+S))$$

$$= E[(\psi(y_{k+S-1})+b_S z(k+S-1)+u(k+S-1)]+E(v(k+S))$$

(using (2)–(8), especially (2) and (6))

$$= E(\psi(y_{k+S-1}))+b_S\tau+\alpha_1+\beta_1$$

(using assumption (iii)). Q.E.D.

Equation (22) is established as follows.

Let,

$$t = k+S-1$$

$$Y^c(k+S-1) = [Y(k)Y(k+1)Y(k+S-1)]^T \quad \ldots \quad \text{(B.1)}$$

$$V_1(k+S-1) = [v(k)v(k+1) \ldots v(k+S-1)]^T \quad \ldots \quad \text{(B.2)}$$

$$Q_1(y_t) = [y^1(t)\ y^2(t) \ldots y^s(t)]^T \quad \ldots \quad \text{(B.3)}$$

$$Q_2(y_t) = [y^1(t)y^1(t)\ y^2(t)y^1(t) \ldots y^s(t)y^s(t)]^T \quad \ldots \quad \text{(B.4)}$$

$$\vdots$$

$$Q_p(y_t) = [\underbrace{y^1(t)\ldots y^1(t)}_{p \text{ times}}\ \underbrace{y^2(t)y^1(t)\ldots y^1(t)}_{p \text{ times}} \ldots \underbrace{y^s(t)\ldots y^s(t)}_{p \text{ times}}]^T \quad \ldots \quad \text{(B.5)}$$

By referring to (4) and (6), we have

$$y(t) = [Q_1(y_t)\ Q_2(y_t) \ldots Q_p(y_t)]^T \quad \ldots \quad \text{(B.6)}$$

$$= \phi(y_t). \quad \ldots \quad \text{(B.7)}$$

Now $Q_i(\cdot)$ is a $\binom{s+i-1}{i} \times 1$ vector, $1 \leqslant i \leqslant p$. Also refer to the definitions of $\bar{A}^*_{jj}$ and $\bar{A}^*_{ij}$ immediately following equation (23) of the text.

From (B.1), (B.2) or (16) and (B.3), and (9) and (10), we have,

$$Q_1(Y^C(k+S-1)) = Q_1\left(((0\ 0 \ldots 1) \begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+S-1) \end{bmatrix} + \begin{bmatrix} v(k) \\ v(k+1) \\ \vdots \\ v(k+S-1) \end{bmatrix} \right)$$

$$= Q_1(y(k+S-1))+Q_1(V_1(k+S-1))$$

$$= \bar{A}^*_{11}\ Q_1(y(k+S-1))+Q_1(V_1(k+S-1))$$

(since $\tilde{A}^*_{11} = I$, which is an identity matrix of size :

$$\binom{s+1-1}{1} \times \binom{s+1-1}{1} = s \times s).$$

It should be clear now that

$$Y^{c^1}(t)\,Y^{c^1}(t) = y^1(t)y^1(t) + 2V_1^1(t)y^1(t) + V_1^1(t)V_1^1(t)$$

where, for instance, $Y^{c^1}(t)$ is the first observation on the vector $Y^c(t)$.

Similarly,

$$Y^{c^2}(t)\,Y^{c^1}(t) = y^2(t)y^1(t) + V_1^2(t)y^1(t) + V_1^1(t)y^2(t) + V_1^2(t)V_1^2(t)$$

$$\vdots \qquad\qquad \vdots$$

$$Y^{c^s}(t)\,Y^{c^s}(t) = y^s(t)y^s(t) + 2V_1^s(t)y^s(t) + V_1^s(t)V_1^s(t).$$

It follows from the above equations and from (8.3)–(8.5) that

$$Q_2(Y^c(t)) = \tilde{A}^*_{21}Q_1(y(t)) + \tilde{A}^*_{22}Q_2(y(t)) + Q_2(V_1(t))$$

where $\tilde{A}^*_{21}$ is as given in (25) of the text. By an exactly similar procedure, we derive

$$Q_i(Y^c(t)) = \sum_{j=1}^{i} \tilde{A}^*_{ij}\, Q_j(y(t)) + Q_i(V_1(t)).$$

for $2 < i \leqslant p$, where $\tilde{A}^*_{ij}$ is a previously defined matrix with either zero or measurement noise elements as its terms. Generalization of the above result produces equation (22) of the text. However, there one must remember that $Y^c(k+S-1)$ receives the same vector representation as $y(t)$ of (B.6) or (4) and $\phi(V_1(k+S-1))$ as $\phi(y_t)$ of (B.7) or (8).

Lastly we consider the bias of the estimate. The bias of estimate, from (25), arises on account of non-vanishing values of

$$E\{W(k+S-1)[Y(k+S) - W^T(k+S-1)\hat{\theta}(k-1)]\}$$

which, from (18), is equal to

$$E(W\eta)$$

writing, for instance, $\eta(k+S)$ as η, $W(k+S-1)$ as W.

Now, from (16) and (10), and Assumption (i),

$$E(W\eta) = E\{(\bar{Y}^c + V)[v(k+S) - W^T\theta + U^T d^0]\}$$

$$= -\begin{bmatrix} \sigma_v^2 I & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \sigma_w^2 I \end{bmatrix} \theta + E(\bar{Y}^c U^T d^0).$$

Next, we express $\bar{Y}^c$ in terms of y's, z's, and u's for the computation of $E(\bar{Y}^c U^T d^0)$. In this expression, terms involving y's and z's do not finally matter because these variables and u's are independent by assumption and thus the terms involving the expectation of these variables and u's will vanish. This being so, we focus our attention on the component u's in the expansion of $\bar{Y}^c$.

Since

$$\bar{Y}(k) = (0\ 0\ \ldots\ 0\ 1)y(k).$$

so

$$y^s(k) = \bar{Y}(k). \qquad \ldots \quad (C.1)$$

Again, since

$$\bar{Y}(k+1) = (0\ 0\ \ldots\ 0\ 1)\left\{\begin{bmatrix} (0 & \vdots & I)y(k) \\ \cdots & \cdots & \cdots \\ & \psi(y(k)) & \end{bmatrix} + bz(k) + cu(k)\right\}.$$

so

$$a_1 y^1(k) + \ldots + a_s y^s(k) = \bar{Y}(k+1) - b_s z(k) - u(k). \qquad \ldots \quad (C.2)$$

Also from proof of Lemma 1, we have

$$a_S a_1 y^1(k) + (a_1 + a_S a_2) y^2(k) + \ldots + (a_S^2 + a_{S-1}) y^S(k)$$

$$= \bar{Y}(k+2) - (a_1 b_1 + \ldots + a_S b_S) z(k)$$

$$- b_S z(k+1) - (a_1 + \ldots + a_S) u(k) - u(k+1). \qquad \ldots \quad (C.3)$$

We can go on, similarly obtaining expressions for $\bar{Y}(k+3), \ldots,$ and $\bar{Y}(k+S-1)$, and arrange these $\bar{Y}$'s to form $\bar{Y}^c\ (k+S-1)$ as defined in (16) written as

$$\bar{Y}^c(k+S-1) = \begin{bmatrix} \bar{Y}(k) \\ \cdots \\ \bar{Y}(k+S-1) \\ \cdots\cdots \\ Z^c(k+S-1) \end{bmatrix}$$

to be now expressed, after using (C.1), (C.2), (C.3), ..., as

$$\bar{Y}^c(k+S-1) = \begin{bmatrix} Cy(k) + HZ^c(k+S-1) + DU(k+S-1) \\ \cdots\cdots \\ Z^c(k+S-1) \end{bmatrix}$$

where

$$C = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_S \\ a_S a_1 & a_1 + a_S a_2 & \cdots & a_S^2 + a_{S-1} \\ a_{S-1}a_1 + a_S^2 a_1 & a_{S-1}a_2 + a_S a_1 + a_S^2 a_2 & \cdots & a_{S-2} + a_{S-1}a_S + a_{S-1}a_S + a_S^3 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

$$H = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ b_S & 0 & \cdots & 0 \\ \sum^{S} a_i b_i & b_S & \cdots & 0 \\ a_s \sum a_i b_i + \sum^{s-1} a_i b_{i+1} & \sum a_i b_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

and D, worked out relatively fully, is as shown in (28) of the text. (For a mechanical rule of determining the elements of D, see the paragraph following (28)). Note that we show only some of the elements of the matrices C and H for a basis as to how these elements are computed, but not all elements, since we do not need them anyway. As mentioned before, terms involving C and H drop out in the calculation of $E(W\eta)$ continued below.

$$E(W\eta) = -\begin{bmatrix} \sigma_e^2 I & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \sigma_w^2 I \end{bmatrix} \theta + \begin{bmatrix} \sigma_u^2 D d^0 \\ \cdots \\ 0 \end{bmatrix}$$

which is equation (27). This is subtracted from the terms inside the second brackets on the right hand side of (25) to produce equation (26). The adjustment gets rid of the bias.

Acknowledgement

The author is grateful to Peter Ponzo of Applied Mathematics and T. Prasad of Civil Engineering Departments of the University of Waterloo for various assistances, and to Clifford Wymer of the London School of

Economics for initial encouragement. However, the author alone is to blame for errors. The author also gratefully acknowledges the financial grants received from the Canada Council and the University of Waterloo in connection with the computer work of Section 2.3.

References

Albert, A. E., and L. A. Gardner, Jr. (1967): *Stochastic Approximation and Nonlinear Regression*, M. I. T. Press Res. Mono, No. 42.

Bergstrom, A. R. (1966): Non-recursive models as discrete approximations to systems of stochastic differential equations. *Econometrica*, 34, 173-182.

Bergstrom, A. R. and Wymer, C. R. (1974): A model of disequilibrium neoclassical growth with application to the U. K. Read at the *European meeting of the Econometric Society*.

De Figueiredo, R. J. P., and Netravalli A. N. (1970): Parameter estimation: System identification. *Symposium on Nonlinear Estimation and Its Application*, IEEE, 202-206.

Dvoretzky, A. (1956): On stochastic approximation, in *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*, vol. 1, Univ. of California Press, Berkeley, Calif.

Fisher, F. M. (1961): Identifiability criteira in nonlinear systems. *Econometrica*, Oct., Vol, 574-589.

Ho, Y. C. (1963): On the stochastic approximation method and optimal filtering theory, *J. Math. Anal. and Appl.*, 6, 152-154.

Koopmans, T. C. (1950): Models involving continuous time variables, in *Statistical Inference in Dynamic Economic Models*, T. C. Koopmans, ed., Cowles Commission for Research in Economics, Monograph 10, New York: John Wiley and Sons.

Ho, T. C., and Lee, R. C. K. (1965): Identification of linear dynamic systems, *Information Control*, 8, 93-110.

Lee, R. C. K. (1964): *Optimal Estimation, Identification and Control*, M. I. T. Press Res. Mono, No. 28.

Netravalli, A. N. and De Figueiredo J. P. (1971): On the identification of nonlinear dynamical system. *IEEE Transactions on Automatic Control*, Vol. AC-16, No. 1, 28-36.

Phillips, A. W. (1959): The estimation of parameters in systems of stochastic differential equations, *Biometrika*, 46, 67-76.

Phillips, P. C. B. (1972): The structural estimation of a stochastic differential equation system. *Econometrica*, 40, No. 6, 1021-1042.

——— (1974): The estimation of some continuous time models. *Econometrica*, 42, 803-823.

Robinson, P. M. (1976a): The estimation of linear differential equations with constant coefficients, *Econometrica*, 44, No. 4, 751-764.

——— (1976b): Instrumental variables estimation of differential equations, *Econometrica*, 44, No. 4, 765-776.

Sakrison, D. J. (1967): The use of stochastic approximation to solve the system identification problem, *IEEE Trans. Automat. Contr.*, Vol. AC-12, 563-567.

Sargan, J. D. (1974): Some discrete approximations to continuous time stochastic models, *Jour. Royal Statist. Soc.*, Series B, 36, 74-90.