USC M8616

MODELLING RESEARCH GROUP

DEPARTMENT OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA
90089-0152

MODELLING RESEARCH GROUP

SEQUENTIAL NONLINEAR ESTIMATION WITH NONAUGMENTED PRIORS

ROBERT KALABA

LEIGH TESFATSION

MRG WORKING PAPER #R8616

DEPARTMENT OF ECONOMICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA
90089-0152

SEQUENTIAL NONLINEAR ESTIMATION WITH NONAUGMENTED PRIORS

ROBERT KALABA

LEIGH TESFATSION

MRG WORKING PAPER #R8616

Revised April 15, 1987

ABSTRACT

How might the basic compatibility of theory and observations be tested for nonlinear dynamic processes without imposing arbitrary stochastic restrictions on the process state variables? The present paper proposes a "flexible least cost" approach. For each possible estimated state sequence x, let $c_D(x)$ and $c_M(x)$ denote the costs incurred for deviations away from the prior dynamic and measurement specifications, respectively. The greatest lower bound for the set of cost vectors $\big(c_D(x), c_M(x)\big)$ associated with all possible estimated state sequences x gives the locus of minimal attainable dynamic and measurement costs. The estimated state sequences which attain the "cost-efficiency frontier" indicate the possible ways the actual process could have developed over time in a manner minimally incompatible with the prior dynamic and measurement specifications. An algorithm is developed for the exact sequential updating of the least-cost state estimates as the duration of the process increases and additional observations are obtained.

MRG Working Paper No. 8616
Revised: April 15, 1987

SEQUENTIAL NONLINEAR ESTIMATION WITH NONAUGMENTED PRIORS[1]

Robert Kalaba[2] and Leigh Tesfatsion[3]

Abstract. How might the basic compatibility of theory and observations be tested for nonlinear dynamic processes without imposing arbitrary stochastic restrictions on the process state variables? The present paper proposes a "flexible least cost" approach. For each possible estimated state sequence x, let $c_D(x)$ and $c_M(x)$ denote the costs incurred for deviations away from the prior dynamic and measurement specifications, respectively. The greatest lower bound for the set of cost vectors $(c_D(x),c_M(x))$ associated with all possible estimated state sequences x gives the locus of minimal attainable dynamic and measurement costs. The estimated state sequences which attain this "cost-efficiency frontier" indicate the possible ways the actual process could have developed over time in a manner minimally incompatible with the prior dynamic and measurement specifications. An algorithm is developed for the exact sequential updating of the least-cost state estimates as the duration of the process increases and additional observations are obtained.

Keywords. Nonlinear State Estimation; Nonaugmented Priors; Exact Sequential Solution; Invariant Imbedding.

1. INTRODUCTION

Suppose noisy observations are obtained on a process modelled in terms of nonlinear dynamic and measurement specifications. How might the basic compatibility of the theory and the observations be tested without introducing arbitrary stochastic restrictions?

The following "flexible least cost" approach is proposed. Associated

---

[1] A preliminary version of this paper [Ref.1] was presented at the June 1986 Summer Econometric Society Meeting at Duke University. The authors are grateful to E. Massoumi, S. Mittnik, H. C. Quirmbach, A. Zellner, and especially to R. Huss for helpful comments.

[2] Professor, Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089-1451.

[3] Associate Professor, Modelling Research Group, Department of Economics, University of Southern California, Los Angeles, CA 90089-0152.

with each possible estimate x for the sequence of successive process states are two types of residual modelling error. First, x may fail to satisfy the prior dynamic specifications. Second, x may fail to satisfy the prior measurement specifications.

Let $c_D(x)$ denote the cost assigned to x for deviations away from the prior dynamic specifications, and let $c_M(x)$ denote the cost assigned to x for deviations away from the prior measurement specifications. Let P denote the set of cost vectors $(c_D(x), c_M(x))$ corresponding to all possible estimated state sequences x. By construction, the greatest lower bound for P yields the locus of minimal attainable dynamic and measurement costs, conditional on the given set of observations. Estimated state sequences which attain this "cost-efficiency frontier" indicate how the actual process could have developed over time in a manner minimally incompatible with the prior dynamic and measurement specifications.

The present paper develops the basic conceptual ideas underlying the flexible least cost approach. Section 2 describes the generation of the cost-efficiency frontier and corresponding flexible least cost state estimates for a process modelled in terms of general nonlinear dynamic and measurement priors, using squared residual error sums to measure incompatibility costs. Section 3 presents an algorithm for the exact sequential updating of the flexible least cost state estimates as the length of the process increases and additional observations are obtained. The incorporation of stochastic restrictions and the use of more generally specified cost functions are taken up in section 4. Section 5 relates the flexible least cost approach to previous work on nonlinear estimation by the present authors and by other researchers in systems science and engineering. Section 6 relates the flexible least cost approach to nonlinear estimation techniques currently used in economics and statistics.

A detailed theoretical investigation of the flexible least cost approach is undertaken in Ref. 2 for a linear regression problem with time-varying regression coeffients (states). Squared residual error sums are used for the dynamic and measurement costs. A FORTRAN program for this application is presented and validated in Ref. 3. The cost-efficiency frontier and corresponding flexible least cost estimates for the time-varying regression coefficients are generated for a variety of simulated nonlinear motions in the true underlying regression coefficients.

2. FLEXIBLE LEAST COST: AN ILLUSTRATIVE EXAMPLE

The present section describes the generation of the cost-efficiency frontier and corresponding flexible least cost state estimates for a process modelled in terms of general nonlinear dynamic and measurement priors. Squared residual error sums are used to measure incompatibility costs. Stochastic restrictions and more general incompatibility measures are considered below in section 4.

Suppose a process is modelled by a system of nonlinear dynamic equations

$$x_{t+1} = F(x_t) + u_t \,, \quad t = 0,\ldots,T-1 \,, \tag{1a}$$

where x_t is an n-dimensional column vector of state variables and u_t is an n-dimensional column vector of residual dynamic errors, $n \geq 1$. Suppose also that observations obtained on the state vectors are modelled by a system of nonlinear measurement equations

$$y_t = H(x_t) + v_t \,, \quad t = 0,\ldots,T \,, \tag{1b}$$

where y_t is an m-dimensional column vector of observations and v_t is an m-dimensional column vector of residual measurement errors, $m \geq 1$.[4]

Finally, suppose the residual dynamic and measurement errors u_t and v_t are believed ex ante to be small, but no additional properties are specified for them. The prior theoretical specifications concerning the generation of the observations $y_0,\ldots,y_T$ might thus be cast in the following form:

Prior Dynamic Specifications:

$$x_{t+1} - F(x_t) \approx \underset{\sim}{0}, \quad t = 0,\ldots,T-1 ; \tag{2a}$$

Prior Measurement Specifications:

$$y_t - H(x_t) \approx \underset{\sim}{0}, \quad t = 0,\ldots,T . \tag{2b}$$

The basic problem at hand is assumed to be the reconciliation of theory with observations; i.e., does there exist any sequence of state vectors $x_0,\ldots,x_T$ which satisfy the prior theoretical specifications (2) in an acceptable approximate sense for the realized sequence of observation vectors $y_0,\ldots,y_T$? How might such an estimated state sequence be found?

One possible approach is as follows. Associated with each possible estimated state sequence $x = (x_0,\ldots,x_T)$ are two types of residual modelling error. First, x may fail to satisfy the prior dynamic specifications (2a). Second, x may fail to satisfy the prior measurement specifications (2b). Suppose the cost assigned to x for the first type of error is measured by the sum of squared residual dynamic errors

---

[4]Since the dynamic and measurement functions $F(\bullet)$ and $H(\bullet)$ are permitted to be general nonlinear functions, unknown parameters can be incorporated into the state vectors x_t. Time-dependence of the dynamic and measurement functions can be handled by including the time variable t as a component of each state vector. It is assumed that preliminary scaling and transformations have been carried out as appropriate.

$$c_D(x;T) \;=\; \sum_{t=0}^{T-1} |x_{t+1} - F(x_t)|^2 \,, \tag{3a}$$

and the cost assigned to x for the second type of error is measured by the sum of squared residual measurement errors

$$c_M(x;T) \;=\; \sum_{t=0}^{T} |y_t - H(x_t)|^2 \,. \tag{3b}$$

Define the <u>(time-T) cost possibility set</u> to be the collection

$$P(T) \;=\; \{\, c_D(x;T),\, c_M(x;T) \,:\, x \in E^{n[T+1]} \,\} \tag{4}$$

of all possible configurations of dynamic and measurement costs attainable at time T, conditional on the given observations $y_0,\ldots,y_T$.

If the prior theoretical specifications (2) are correct, then, by construction, the dynamic and measurement costs associated with the actual sequence of state vectors will be approximately zero. In general, however, the greatest lower bound $P_F(T)$ for the cost possibility set P(T) will be bounded away from the origin in E^2. By construction, the greatest lower bound $P_F(T)$ gives the locus of minimal dynamic and measurement costs attainable at time T, conditional on the given set of observations. That is, given any estimated state sequence yielding a point on $P_F(T)$, there exists no other estimated state sequence which simultaneously lowers both types of cost. Hereafter $P_F(T)$ will be referred to as the <u>(time-T) cost-efficiency frontier</u>. [See figure 1(a).]

How might the cost-efficiency frontier be found? In analogy to the usual procedure for tracing out Pareto-efficiency frontiers, a parameterized family of minimization problems is considered.

Thus, let $\mu \geq 0$ and $T \geq 0$ be given, and let each possible estimated state sequence $x = (x_0,\ldots,x_T)$ be assigned an <u>incompatibility cost</u>

$$C(x;\mu,T) \quad = \quad \mu c_D(x;T) \; + \; c_M(x;T) \tag{5}$$

consisting of the μ-weighted average of the associated dynamic and measurement costs.[5] Suppose a unique state sequence minimizes the incompatibility cost (5) for each $\mu > 0$ and each $T \geq T^*$ for some given $T^* \geq 0$. Let this unique state sequence be denoted by

$$x^f(\mu,T) \quad = \quad (\; x_0^f(\mu,T), \;\ldots\;, \; x_T^f(\mu,T) \;) \;. \tag{6}$$

If $\mu = 0$, let (6) denote any sequence of state estimates which minimizes the dynamic cost $c_D(x;T)$ subject to $c_M(x;T) = 0$. Hereafter, (6) will be referred to as the <u>flexible least cost (FLC) solution, conditional on μ and T</u>.[6]

Finally, let the dynamic and measurement costs corresponding to the FLC solution (6) be denoted by

---

[5] When a least-squares formulation such as (5) is used as the incompatibility measure, a common reaction is that the analysis is implicitly relying on normality assumptions for residual error terms. To the contrary, the incompatibility measure is intended to represent the investigator's assessment of the <u>cost</u> associated with various possible deviations between theory and observations; it bears no necessary relation to any intrinsic stochastic properties of the residual error terms. Thus, for example, the use of the least squares incompatibility measure (5) only indicates that dynamic or measurement deviations of equal magnitude are specified to be equally costly, not that these deviations are anticipated to be symmetrically distributed around zero. More general specifications for the incompatibility measure are considered in section 4.1, below.

[6] More generally, one could define (6) to be the <u>set</u> of state sequences which minimize (5) for given μ and T. However, if the minimizing state sequence is nonunique for some positive μ and T, then the cost-efficiency frontier $P_F(T)$ could have concave portions. In this case the procedure given for generating the cost-efficiency frontier would have to be modified, since the evaluation of the minimized cost function (5) for μ varying over $[0,+\infty)$ would only partially trace out the frontier.

$$c_D(\mu,T) \;=\; c_D(x^f(\mu,T);T)\;; \tag{7a}$$

$$c_M(\mu,T) \;=\; c_M(x^f(\mu,T);T)\;. \tag{7b}$$

By construction, a point (c_D,c_M) in E^2 lies on the cost-efficiency frontier $P_F(T)$ for $T \geq T^*$ if and only if there exists some $\mu \geq 0$ such that $(c_D,\ c_M) = (\ c_D(\mu,T),\ c_M(\mu,T)\ )$. The cost-efficiency frontier $P_F(T)$ thus takes the form

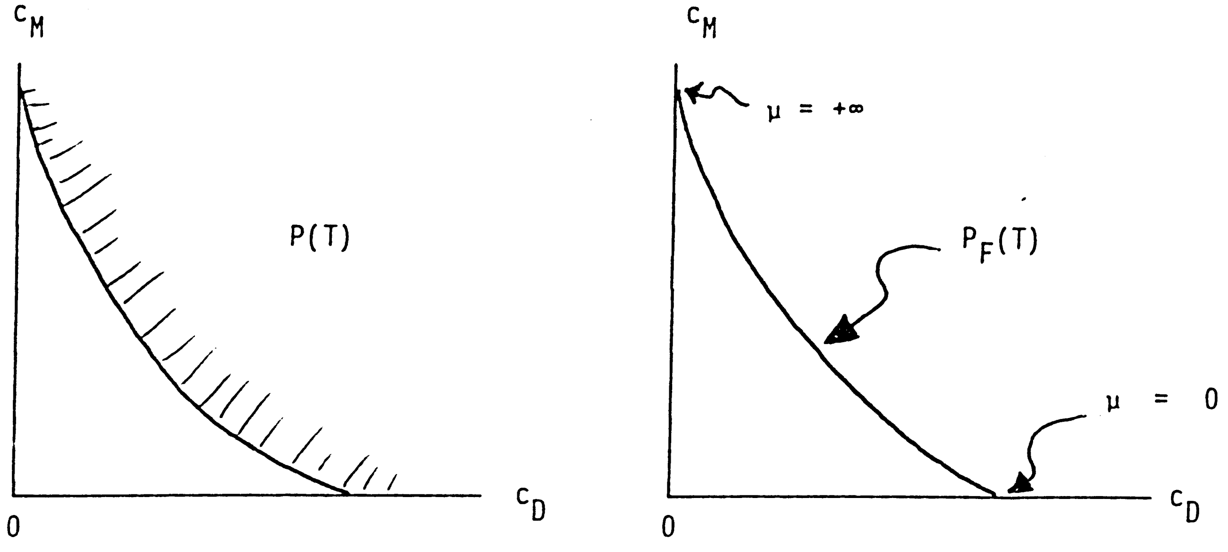
$$P_F(T) \;=\; \{\ c_D(\mu,T),\ c_M(\mu,T)\ :\ 0 \leq \mu < \infty\ \}\;. \tag{8}$$

The cost-efficiency frontier $P_F(T)$ is qualitatively depicted[7] in figure 1(b). As $\mu \rightarrow 0$, the incompatibility cost function (5) ultimately places no weight on the prior dynamic specifications (2a); i.e., c_M is minimized with no regard for c_D. Thus, c_M can generally be brought down to zero, and the corresponding value for c_D will be large. Conversely, as $\mu \rightarrow \infty$, the incompatibility cost function (5) places absolute priority on the prior dynamic specifications (2a); i.e., c_M is minimized subject to $c_D = 0$.

The next section of the paper develops a procedure for the exact sequential minimization of the FLC incompatibility cost function (5).

---

[7] In each of the numerical experiments undertaken in Ref. 3, the cost-efficiency frontier is adequately traced out by evaluating the costs (7) over a rough grid of μ-points increasing by powers of ten. All of the numerically generated frontiers display the convex shape qualitatively depicted in figure 1(b).



(a): Cost Possibility Set P(T) (b): Cost-Efficiency Frontier $P_F(T)$

FIGURE 1

3. EXACT SEQUENTIAL MINIMIZATION OF THE INCOMPATIBILITY COST FUNCTION

Direct minimization of the FLC incompatibility cost function (5) is a formidable task when nonlinearities are present. In section 3.1 a procedure is developed for the exact sequential minimization of (5) as the duration of the process increases and additional observations are obtained. The original n[T+1]-dimensional problem of minimizing (5) with respect to $x_0,\ldots,x_T$ is thus decomposed into a sequence of T+1 minimization problems each of dimension n, a significant computational reduction. Using the basic updating equations for the sequential minimization of (5), it is shown in sections 3.2 through 3.4 that FLC filtered and smoothed state estimates can be generated at each time T together with one-step-ahead state predictions. Practical implementation of the basic updating equations is considered in section 3.5.

3.1 Basic Updating Equations

Let $\mu > 0$ and $t \geq 1$ be given. Define the total cost of the estimation process at time t-1, conditional on the state estimates $x_0,\ldots,x_t$ for times 0 through t, to be the μ-weighted sum of squared residual dynamic and measurement errors

$$J(x_0,\ldots,x_t;\mu,t-1) = \mu \sum_{s=0}^{t-1} |x_{s+1} - F(x_s)|^2 + \sum_{s=0}^{t-1} |y_s - H(x_s)|^2 . \quad (9)$$

Let $\emptyset(x_t;\mu,t-1)$ denote the smallest total cost of the estimation process at time t-1, conditional on the state estimate x_t for time t. Thus, for $t \geq 1$,

$$\emptyset(x_t;\mu,t-1) = \inf_{x_0,\ldots,x_{t-1}} J(x_0,\ldots,x_t;\mu,t-1) . \quad (10)$$

By construction, the function $\emptyset(\bullet;\mu,t-1)$ defined by (10) is bounded below over its domain $E^{n[t+1]}$. Thus, by the principle of iterated infima, $\emptyset(\bullet;\mu,t-1)$ satisfies the recurrence relation

$$\emptyset(x_{t+1};\mu,t) = \inf_{x_t}[\ \mu|x_{t+1} - F(x_t)|^2 + |y_t - H(x_t)|^2 + \emptyset(x_t;\mu,t-1)\ ] \quad (11a)$$

for all x_{t+1} in E^n and $t \geq 0$.

The recurrence relation (11a) is initialized by assigning a prior cost-of-estimation $\emptyset(x_0;\mu,-1)$ to each x_0 in E^n. Given the cost specification (5), this prior cost-of-estimation is trivially given by

$$\emptyset(x_0;\mu,-1) \equiv 0 . \quad (11b)$$

In general, however, $\emptyset(x_0;\mu,-1)$ could summarize the cost of being in state x_0 at time 0 conditional on everything that is known about the process prior to obtaining an observation y_0 at time 0.

The recurrence relation (11a) can be given a dynamic programming interpretation. At any current time t the choice of a state estimate x_t incurs three types of cost, conditional on an anticipated state estimate x_{t+1} for time t+1. First, the state estimate x_t may fail to satisfy the prior dynamic specification (2a). The cost incurred for this dynamic error is $\mu|x_{t+1} - F(x_t)|^2$. Second, the state estimate x_t may fail to satisfy the prior measurement specification (2b). The cost incurred for this measurement error is $|y_t - H(x_t)|^2$. Third, a cost $\emptyset(x_t;\mu,t-1)$ is incurred by choosing the state estimate x_t at time t, based on everything that is known about the process through time t-1.

These three costs together comprise the total cost of choosing a state estimate x_t at time t, conditional on an anticipated state estimate x_{t+1} for time t+1. Minimization of this total cost with respect to x_t thus yields the cost $\emptyset(x_{t+1};\mu,t)$ incurred by choosing the state estimate x_{t+1} at time t+1, based on everything that is known about the process through time t.

3.2. Filtered State Estimates

Suppose the cost function (5) is minimized by a unique sequence of state vectors for each process length $T \geq T^*$ for some given $T^* \geq 0$. Then the following exact sequential procedure can be used to generate the FLC filter estimate $x_T^f(\mu,T)$ for the state vector x_T at each time $T \geq T^*$.

In storage at time T = 0 is the function $\emptyset(x_0;\mu,-1)$, defined for all x_0 in E^n. A first observation vector y_0 is obtained. In preparation for the next time 1, determine and store for each x_1 in E^n the function value

$$\emptyset(x_1;\mu,0) = \inf_{x_0} [\ \mu|x_1 - F(x_0)|^2 + |y_0 - H(x_0)|^2 + \emptyset(x_0;\mu,-1)\ ] . \quad (12)$$

If $0 = T^*$, the FLC estimate for the state vector x_0 at time 0, based on the single observation vector y_0, is given by

$$x_0^f(\mu,0) = \arg\left[\min_{x_0} [\ |y_0 - H(x_0)|^2 + \emptyset(x_0;\mu,-1)\ ] \right] . \quad (13)$$

In storage at time $T \geq 0$ is the function $\emptyset(x_T;\mu,T-1)$, defined for all x_T in E^n. An additional observation vector y_T is obtained. In preparation for the next time T+1, determine and store for each x_{T+1} in E^n the function value

$$\emptyset(x_{T+1};\mu,T) = \inf_{x_T} [\ \mu|x_{T+1} - F(x_T)|^2 + |y_T - H(x_T)|^2 + \emptyset(x_T;\mu,T-1)\ ] . \quad (14)$$

If $T \geq T^*$, the FLC estimate for the state vector x_T at time T, based on the observation vectors $y_0,\ldots,y_T$, is given by

$$x_T^f(\mu,T) = \arg\left[\min_{x_T} [\ |y_T - H(x_T)|^2 + \emptyset(x_T;\mu,T-1)\ ] \right] . \quad (15)$$

It is easily established that (15) does yield the FLC filter estimate for the state vector x_T at each time $T \geq T^*$. By assumption, the total cost of the estimation process at time T, conditional on the state estimates $x_0,\ldots,x_T$, is measured by the incompatibility cost function (5). This cost function can equivalently be written as

$$|y_T - H(x_T)|^2 \quad + \quad J(x_0,\ldots,x_T;\mu,T-1) \ , \tag{16}$$

where the function $J(\bullet;\mu,T-1)$ is defined as in (9). The simultaneous minimization of the cost function (5) with respect to the state vectors $x_0,\ldots,x_T$ can thus be equivalently expressed as

$$\min_{x_0,\ldots,x_T} \left[|y_T - H(x_T)|^2 \quad + \quad J(x_0,\ldots,x_T;\mu,T-1) \right]$$

$$= \quad \min_{x_T} \left[|y_T - H(x_T)|^2 \quad + \quad \min_{x_0,\ldots,x_{T-1}} \quad J(x_0,\ldots,x_T;\mu,T-1) \right]$$

$$= \quad \min_{x_T} \left[\ |y_T - H(x_T)|^2 \quad + \quad \emptyset(x_T;\mu,T-1) \ \right] . \tag{17}$$

However, expression (17) is precisely the term in large square brackets in (15).

3.3. Smoothed State Estimates

FLC smoothed state estimates can also be generated sequentially, using the recurrence relation (14). For example, consider the problem of obtaining the FLC smoothed estimate for the state vector x_T at time $T \geq T^*$ as the length of the process increases from T to T+1 and an additional observation vector y_{T+1} is obtained. In preparation for time T+1, the function value $\emptyset(x_{T+1};\mu,T)$ in (14) has been calculated and stored for each x_{T+1} in E^n. As a by-product of this calculation, the (unique) minimizing x_T as a function of x_{T+1} has been obtained. Suppose this functional relationship has also been stored. Let this functional relationship be denoted by

$$x_T \quad = \quad S_T(x_{T+1}) \ . \tag{18}$$

The FLC filter estimate for the state vector x_{T+1} at time T+1 is given by

$$x^f_{T+1}(\mu,T+1) \;=\; \arg\left[\min_{x_{T+1}} [\,|y_{T+1} - H(x_{T+1})|^2 + \emptyset(x_{T+1};\mu,T)\,]\right] . \qquad (19)$$

The FLC smoothed estimate for the time-T state vector x_T, based on the observation vectors $y_0,\ldots,y_{T+1}$ for times 0 through T+1, is then given by

$$x^f_T(\mu,T+1) \;=\; S_T(\,x^f_{T+1}(\mu,T+1)\,) . \qquad (20)$$

More generally, given $T \geq T^*$, and given any fixed time t, $0 \leq t \leq T$, the FLC smoothed estimate $x^f_t(\mu,T+1)$ for the time-t state vector x_t based on the observation vectors $y_0,\ldots,y_{T+1}$ for times 0 through T+1 is found as follows. In storage at time T+1 are the previously calculated functions

$$\begin{aligned} x_t &= S_t(x_{t+1}) ; \\ &\;\;\vdots \\ x_T &= S_T(x_{T+1}) , \end{aligned} \qquad (21)$$

together with the function $\emptyset(x_{T+1};\mu,T)$. The FLC filter estimate $x^f_{T+1}(\mu,T+1)$ for the state vector x_{T+1} at time T+1 is determined as in (19). The FLC smoothed estimate for the state vector x_t at time t is then found by solving equations (21) for x_t in reverse order, starting from the terminal condition $x_{T+1} = x^f_{T+1}(\mu,T+1)$.

3.4. One-Step-Ahead Prediction

Finally, the FLC one-step-ahead prediction for the state vector x_{T+1} at time T+1, conditional on the observation vectors $y_0,\ldots,y_T$ for times 0 through $T \geq T^*$, is determined as follows. Since no observation vector for time T+1 is yet available, the selection of a predicted value for x_{T+1} only incurs a cost if this predicted value deviates from the value generated directly through the dynamic function $F(\bullet)$. By assumption, this cost takes the form

$$|x_{T+1} - F(x_T^f(\mu,T))|^2 . \qquad (22)$$

Thus, the FLC one-step-ahead prediction for x_{T+1} is given by

$$x_{T+1}^f(\mu,T) = F(x_T^f(\mu,T)) . \qquad (23)$$

3.5. Practical Implementation

In Ref. 3 it is seen for the case of linear dynamic and measurement equations that the minimizations with respect to x_T required in (14) and (15) can be carried out exactly by simple calculus, resulting in explicit exact updating equations for the successive state estimates. For the more challenging nonlinear problem, unless special problem features can be exploited, these minimizations will typically have to be done by some type of search process. Although seemingly in equations (14) and (15) a search over the entire space E^n is required, in actuality the search can be limited to certain regions.

Specifically, after a sufficient number of observations have been obtained, the grid of values for x_T in (15) can be limited to a neighborhood of $F(x_{T-1}^f(\mu,T-1))$ and the grid of values for x_T in (14) can be limited to some region containing both $F^{-1}(x_{T+1})$ and $x_T^f(\mu,T)$. Also, in (14) the evaluation of $\emptyset(x_{T+1};\mu,T)$ can be limited to values of x_{T+1} in a neighborhood of $F(x_T^f(\mu,T))$. In effect, then, at each time T one should have in storage not only selected values of $\emptyset(x_T;\mu,T-1)$ but also the value of the previous FLC filter estimate $x_{T-1}^f(\mu,T-1)$.

Towards the beginning of the observation process, special properties of the dynamic and measurement specifications would presumably have to be exploited to reduce the required searches.

4. GENERALIZATIONS

4.1 Incorporation of Prior Stochastic Specifications

Section 2 illustrates the FLC approach for a process modelled in terms of prior dynamic and measurement specifications. However, prior stochastic specifications can also be incorporated.

For example, suppose the dynamic equations for the process in section 2 are believed ex ante to be subject to serially independent identically distributed random shocks ϵ_t with zero mean and finite covariance matrix. Specifically, suppose in place of (1a) one has

$$x_{t+1} = F(x_t) + \epsilon_t + u_t , \quad t = 0,\ldots,T-1 , \tag{24}$$

where again the residual dynamic errors u_t are believed to be small but are otherwise unrestricted. The prior theoretical specifications (2) might then be modified as follows:

Prior Dynamic Specifications:

$$x_{t+1} - F(x_t) - \epsilon_t \approx \underset{\sim}{0} , \quad t = 0,\ldots,T-1 ; \tag{25a}$$

Prior Measurement Specifications:

$$y_t - H(x_t) \approx \underset{\sim}{0} , \quad t = 0,\ldots,T ; \tag{25b}$$

Prior Stochastic Specifications:

$$[\sum_{t=0}^{T-1} \epsilon_t]/T \approx \underset{\sim}{0} . \tag{25c}$$

The incompatibility cost function (5) should then be modified to take into account the modified prior theoretical specifications (25). For example, the modified cost function might take the form[8]

$$\mu_1 \sum_{t=0}^{T-1} |x_{t+1} - F(x_t) - \epsilon_t|^2/T \;+\; \mu_2 \sum_{t=0}^{T} |y_t - H(x_t)|^2/T \;+\; \mu_3 \,|[\sum_{t=0}^{T-1} \epsilon_t]/T|^2 \,, \qquad (26)$$

where μ_1, μ_2, and μ_3 are positive scalar weights, normalized to sum to one.

The cost function (26) is then minimized with respect to both the state variables $(x_0,\ldots,x_T)$ and the shock terms $(\epsilon_0,\ldots,\epsilon_{T-1})$. For any given set of observations, evaluation of the minimized cost function (26) over the simplex of normalized penalty weights $\{\mu_1,\mu_2,\mu_3\}$ yields the cost-efficiency frontier, i.e., the minimal attainable squared residual error sums corresponding to the three possible sources of theoretical error listed in (25)--dynamic, measurement, and stochastic.

The above example illustrates how a moment restriction might be incorporated into the incompatibility cost function by penalizing the deviation between the empirical moment and the ex ante specified population moment. The cost function could also impose a penalty for various other types of residual stochastic error. For example, one might wish to consider independence restrictions, or restrictions to particular classes of distributions.

The FLC approach can thus be viewed as a generalization of the standard "minimum distance" technique for empirical distribution estimation in which penalties are imposed for deviations between general features of empirical

---

[8]Each <u>type</u> of constraint in (25)--dynamic, measurement, and stochastic --is given <u>equal</u> weight in the incompatibility cost function (26) apart from the effects of the differential penalty weights μ_i. Thus, the squared residual dynamic and measurement error sums are now divided by T. More general penalty weighting schemes could of course be considered. As before, it is assumed that preliminary scaling and transformations have been carried out as appropriate.

distributions and their assumed population counterparts. The crucial difference is that the FLC incompatibility cost function is not focused exclusively on stochastic restrictions; rather, all prior theoretical specifications for a process--dynamic, measurement, and stochastic--are simultaneously considered.

The FLC approach thus provides a way to examine the consequences of any proposed asymmetric treatment of theoretical restrictions. For example, suppose the cost-efficiency frontier indicates that a high cost must be paid in terms of squared residual dynamic and measurement errors in order to bring the sum of squared residual stochastic errors close to zero. A rigid adherence to the prior stochastic specifications might then seem unreasonable.

4.2 Nonquadratic Incompatibility Measures

Sequential solutions can be obtained in principle for more general types of processes and for more general incompatibility measures than considered in section 2.

For example, suppose the prior dynamic specifications (2a) are generalized to

$$F(x_{t+1},x_t) \approx \underset{\sim}{0} , \quad t = 0,\ldots,T-1 ; \tag{27a}$$

and the prior measurement specifications (2b) are generalized to

$$H(y_t,x_t) \approx \underset{\sim}{0} , \quad t = 0,\ldots,T . \tag{27b}$$

The incompatibility measure corresponding to (27), to be minimized with respect to $x_0,\ldots,x_T$, might take the form

$$\mu \sum_{t=0}^{T-1} V_t(F(x_{t+1},x_t)) + \sum_{t=0}^{T} W_t(H(y_t,x_t)) , \tag{28}$$

where each of the real-valued functions $V_t(\bullet)$ and $W_t(\bullet)$ maps 0 into 0, and is strictly increasing with respect to the absolute value of its argument.

Alternatively, suppose an investigator believes ex ante that the true state vectors x_t satisy the prior dynamic specifications (2a) and, in addition, a certain scalar function $H(x_t)$ of these state vectors exhibits maximal positive correlation with a scalar sequence of observations y_t. The incompatibility measure corresponding to these prior theoretical specifications, to be minimized with respect to $x_0,\ldots,x_T$, might take the form

$$\mu \sum_{t=0}^{T-1} |x_{t+1} - F(x_t)|^2 - \sum_{t=0}^{T} y_t H(x_t) . \tag{29}$$

Algorithms for sequentially minimizing the incompatibility measures (28) and (29) can be constructed in a manner analogous to the construction in section 3.

5. RELATIONSHIP TO PREVIOUS WORK IN SYSTEMS SCIENCE AND ENGINEERING

The idea of forming a cost-of-estimation function as a suitably weighted sum of squared residual dynamic and measurement errors was stressed by R. Sridhar, R. Bellman, and other associates in a series of studies focusing on a class of continuous-time nonlinear filtering problems arising in rigid body dynamics. [See Refs. 4-6.]

The calculus of variations was originally applied to obtain minimal cost solutions. However, emphasis soon shifted to sequential solutions, and use was made of parallel ideas developed in the theory of invariant imbedding for obtaining solutions of nonlinear two-point boundary value problems as functions of interval length and boundary condition parameters. Since the resulting sequential filtering equations were typically difficult to implement, Sridhar proposed a quadratic approximation technique for solving these equations. Subsequently, Sugisaka and Sagara [Ref. 7] showed that significant improvements could be obtained using higher-order approximations. The applicability of the Sridhar filter to diverse problems arising in biomedical and mechanical engineering, e.g., blood glucose regulation and beam buckling, is discussed in Ref. 8 [Chapters 4 and 7].

Building on this work, exact sequential filtering equations are obtained in Ref. 9 for a nonlinear two-point boundary value problem associated with a scalar, discrete-time analog of the basic Sridhar continuous-time filtering problem. A subsequent paper [Ref. 10] introduces and tests a tabular method for the numerical solution of these filtering equations. A companion paper [Ref. 11] generalizes the results of Ref. 9 to a class of discrete-time nonlinear estimation problems arising in economics.

In Refs. 9-11 the exact sequential filtering equations are obtained by converting a two-point boundary value problem associated with the first-order necessary conditions into an initial value problem. However, in Ref. 12 it is shown that one does not have to proceed through the first-order necessary conditions.

Specifically, two algorithms are developed and validated in Ref. 12 for the exact sequential minimization of the cost-of-estimation function as the duration of the process increases and new observation vectors are obtained. The first algorithm proceeds by an imbedding on the process length and the final state vector. The second algorithm proceeds by an imbedding on the process length and the final observation vector. Each algorithm generates optimal (least cost) filtered and smoothed state estimates, together with optimal one-step-ahead state predictions. The state-imbedding algorithm forms the basis for the sequential cost minimization method presented in section 3 of the present paper.

6. RELATIONSHIP TO PREVIOUS WORK IN ECONOMICS AND STATISTICS

In economics the processes of interest are complex dynamic social systems not easily subject to controlled experiments. The currently dominant viewpoint, initiated in great part by work at the Cowles Foundation in the nineteen forties and fifties [see Ref. 13], is that economic processes are inherently stochastic processes which must be understood through statistical methods. Econometrics, the application of statistical techniques to the analysis of economic systems, is an outgrowth of this viewpoint. Nevertheless, a small but growing number of economists now argue that an exclusive reliance on statistical methods for the estimation and testing of economic models may be premature.

For example, economists involved in applied general equilibrium modelling have stressed that an exclusive reliance on statistical methods may only be possible for models so simplified that many (or even most) of the interesting features are removed. [See, for example, Refs. 14-15.] Empirical studies by these researchers typically involve the calibration of nonlinear deterministic theory-based models to benchmark data sets. Other economists, particularly those involved in the analysis of chaotic processes, have forcefully argued that deterministic nonlinear dynamical models can produce solution paths that resemble realizations for stochastic processes and hence possess the properties typically seen in empirical time-series data. These properties might therefore be directly derivable from relevant deterministic theoretical relationships, without the need to introduce arbitrary stochastic restrictions. [See, for example, Refs. 16-19.]

Objective investigation of these competing viewpoints requires model estimation and testing techniques with flexible information requirements which allow both stochastic and deterministic nonlinear models to be examined for their compatibility with empirical evidence. The techniques currently advocated in economics and statistics do not appear to have the needed flexibility, since the stochasticity of the underlying process is taken for granted.

More precisely, hypothesis tests typically involve the construction of a criterion function for measuring the goodness-of-fit of competing models. Although current hypothesis testing techniques differ in terms of the goodness-of-fit criterion functions they advocate, seemingly all postulate the existence of an underlying conditional probability distribution governing observations when the null hypothesis is true. In addition, in order to evaluate power, most studies making use of these techniques also postulate the existence of a family of conditional probability distributions governing observations when the null hypothesis is false. Thus, extensive familiarity with and use of these techniques leads naturally to the view that observations on physical and social processes are realizations for stochastic processes, and that legitimate hypothesis testing cannot be conducted outside of a stochastic framework.

For example, recent studies focusing on hypothesis testing for nonlinear models place heavy emphasis on the availability of a likelihood function. The popular likelihood ratio test compares the values of the maximized likelihood function for restricted and unrestricted versions of a model. If the values of the likelihood differ enough, this is taken as evidence that the restrictions fail to hold in reality. Within the maximum likelihood framework, various other closely related approaches have also been devised for measuring goodness-of-fit: e.g., the score (Lagrange multiplier) test, the Wald test, and various Hausman tests. [See, for example, Refs. 20-21.]

Recently there has also been a resurgence of interest among economists in the M-estimation (generalized maximum likelihood) framework first introduced by Huber [Ref. 22]. This approach constructs goodness-of-fit criterion functions in terms of certain auxiliary criterion functions which have zero expectation under the null hypothesis. The M-estimation framework encompasses maximum likelihood and method of moments techniques, among others. [See Ref. 23.] Finally, Bayesian methods based on posterior distributions also continue to be advocated. [See Refs. 24-25.]

The methods described in Refs. 20-25 might all be loosely classified under the heading of "robust statistics", i.e., the statistical study of approximate parametric models. [See Ref. 26.] Another branch of the statistical literature has focused on the development of nonparametric methods for fitting smooth functional forms to noisy data generated in accordance with some underlying (partially) unknown probability distribution. [See Refs. 27-29.] The goodness-of-fit criterion functions used by these methods typically consist of weighted averages of squared residual measurement and "smoothness" errors, where the weights are chosen according to a statistical criterion--e.g., minimization of the true mean square error averaged over the data points.

In summary, the estimation techniques currently predominant in economics and statistics require a stochastic framework to be maintained throughout. No consideration is given to the full range of error trade-offs inherent in a given data set when an attempt is made to fit the data set to theoretical specifications. In contrast, the flexible least cost approach provides a way to investigate these trade-offs both directly, through generated state estimates, and indirectly through the cost-efficiency frontier. It thus appears to be a potentially useful approach for researchers interested in examining a broader spectrum of models than current estimation techniques permit.

REFERENCES

1. Kalaba, R., and Tesfatsion, L., Flexible Least Squares, Modelling Research Group Working Paper No. 8616, Department of Economics, University of Southern California, April 1986.

2. Kalaba, R., and Tesfatsion, L., Time-Varying Regression Via Flexible Least Squares, Modelling Research Group Working Paper No. 8633, Department of Economics, University of Southern California, July 1986.

3. Kalaba, R., Rasakhoo, N., and Tesfatsion, L., Time-Varying Regression Via Flexible Least Squares: Numerical Implementation, Modelling Research Group Working Paper, Department of Economics, University of Southern California, April 1987.

4. Bellman, R., Kagiwada, H., Kalaba, R., and Sridhar, R., Invariant Imbedding and Nonlinear Filtering Theory, Journal of the Astronautical Sciences, Vol. 13, pp. 110-115, 1966.

5. Detchmendy, D., and Sridhar, R., Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems, Transactions of the ASME, pp. 362-368, 1966.

6. Kagiwada, H., Kalaba, R., Schumitsky, A., and Sridhar, R., Invariant Imbedding and Sequential Interpolating Filters for Nonlinear Processes, Journal of Basic Engineering, Vol. 91, pp. 195-200, 1969.

7. Sugisaka, M., and Sagara, S., A Nonlinear Filter With Higher-Order Weight Functions Via Invariant Imbedding, International Journal of Control, Vol. 21, pp. 801-823, 1975.

8. Kalaba, R., and Spingarn, K., Control, Identification, and Input Optimization, Plenum Press, New York, 1982.

9. Kalaba, R., and Tesfatsion, L., An Exact Sequential Solution Procedure for a Class of Discrete-Time Nonlinear Estimation Problems, IEEE Transactions on Automatic Control, Vol. AC-26, pp. 1144-1149, 1981.

10. Kalaba, R., Spingarn, K., and Tesfatsion, L., A Sequential Method for Nonlinear Filtering: Numerical Implementation and Comparisons, Journal of Optimization Theory and Applications, Vol. 34, pp. 541-559, 1981.

11. Kalaba, R., and Tesfatsion, L., A Least-Squares Model Specification Test for a Class of Dynamic Nonlinear Economic Models With Systematically Varying Parameters, Journal of Optimization Theory and Applications, Vol. 32, pp. 538-567, 1980.

12. Kalaba, R., and Tesfatsion, L., Exact Sequential Filtering, Smoothing, and Prediction for Nonlinear Systems, Journal of Nonlinear Analysis, 1987, to appear.

13. Hildreth, C., The Cowles Foundation in Chicago, 1939-1955, Springer-Verlag, N. Y., 1986.

14. Mansur, A., and Whalley, J., Numerical Specification of Applied General Equilibrium Models: Estimation, Calibration, and Data, in Applied General Equilibrium Analysis, Edited by H. Scarf and J. Shoven, Cambridge University Press, Cambridge, pp. 69-127, 1984.

15. Shoven, J., and Whalley, J., Applied General-Equilibrium Models of Taxation and International Trade, Journal of Economic Literature, Vol. 22, pp. 1007-1051, 1984.

16. Blatt, J. M., On the Econometric Approach to Business-Cycle Analysis, Oxford Economic Papers, Vol. 30, pp. 292-300, 1978.

17. Day, R., The Emergence of Chaos From Classical Economic Growth, Quarterly Journal of Economics, pp. 201-213, May 1983.

18. Brock, W., Distinguishing Random and Deterministic Systems: Abridged Version, Journal of Economic Theory, Vol.40, pp. 168-193, 1986.

19. Barnett, W., and P. Chen, The Aggregation-Theoretic Monetary Aggregates Are Chaotic and Have Strange Attractors, Discussion Paper, Department of Economics, University of Texas at Austin, Austin, TX, 1986.

20. Engle, R., Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, Chapter 13 in Handbook of Econometrics, Vol. 2, Edited by Z. Griliches and M. Intriligator, Elsevier Science Publishers, Amsterdam, 1984.

21. Ruud, P. A., Tests of Specification in Econometrics, Econometric Reviews, Vol. 3, pp. 211-242, 1984.

22. Huber, P., The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Edited by L. M. LeCan and J. Neyman, University of California Press, Berkeley, 1967.

23. Tauchen, G., Diagnostic Testing and Evaluation of Maximum Likelihood Models, Journal of Econometrics, Vol. 30, pp. 415-443, 1985.

24. Zellner, A., Bayesian Econometrics, Econometrica, Vol. 53, pp. 253-269, 1985.

25. Leamer, E., Specification Searches, John Wiley & Sons, New York, 1978.

26. Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W., Robust Statistics, John Wiley & Sons, New York, 1986.

27. Tukey, J., Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977.

28. Craven, P., and Wahba, G., Smoothing Noisy Data With Spline Functions, Numerische Mathematik, Vol. 31, pp. 377-403, 1979.

29. Rice, J., and Rosenblatt, M., Smoothing Splines: Regression, Derivatives, and Deconvolution, Annals of Statistics, Vol. 11, pp. 141-156, 1983.