



ELSEVIER

Computational Statistics & Data Analysis 21 (1996) 193–214

COMPUTATIONAL  
STATISTICS  
& DATA ANALYSIS

# A multicriteria approach to model specification and estimation<sup>☆</sup>

Robert Kalaba<sup>a</sup>, Leigh Tesfatsion<sup>b</sup> \*

<sup>a</sup>Departments of Electrical and Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, USA

<sup>b</sup>Department of Economics and Department of Mathematics, Iowa State University, Ames, IA 50011-1070, USA

Received February 1992; revised December 1994

---

## Abstract

In decision theory, incommensurabilities among conflicting decision criteria are typically handled by multicriteria optimization methods such as Pareto efficiency and mean-variance analysis. In econometrics and statistics, where conflicting model criteria replace conflicting decision criteria, probability assessments are routinely used to transform disparate model discrepancy terms into apparently commensurable quantities. This tactic has both strengths and weaknesses. On the plus side, it permits the construction of a single real-valued measure of theory and data incompatibility in the form of a likelihood function or a posterior probability distribution. On the minus side, the amalgamation of conceptually distinct model discrepancy terms into a single real-valued incompatibility measure can make it difficult to untangle the true source of any diagnosed model specification problem. This paper discusses recent theoretical and empirical work on a multicriteria “flexible least squares” (FLS) approach to model specification and estimation. The basic FLS objective is to determine the “cost-efficient frontier,” that is, the set of estimates that are minimally incompatible with a specified set of model criteria. The relation of this work to previous work in econometrics, statistics, and systems science is also clarified.

*Keywords:* Model specification; Set-valued estimation; Multicriteria decision making; Efficiency; Vector-valued optimization; Flexible least squares

---

---

<sup>☆</sup> This work was partially supported by NIH Grant No. DK 33729 and has been presented at meetings of the Econometric Society, the Society for Economic Dynamics and Control, the Midwest Econometrics Group, and the IC<sup>2</sup> Institute. A preliminary abridged version of this paper appears in the IC<sup>2</sup> proceedings volume [20]. The authors are grateful to the editor and two anonymous referees for helpful comments.

\* Corresponding author. E-mail: tesfatsi@iastate.edu

## 1. Introduction

Why have multicriteria decision making (MCDM) techniques played only a minor role in econometric and statistical methodology to date?

On the surface, this minor role is surprising. Every postulated theoretical relation is almost surely false. A cross-sectional function for household demand may be misspecified as linear rather than nonlinear. Dynamic relations may be misspecified because, for example, a wealth accumulation function omits an important variable. Measurement errors stemming from imprecise measuring instruments may not be additive, or, even if additive, they may not be normally distributed. The important point is that conceptually distinct types of theoretical relations are false for conceptually distinct reasons. Consequently, model specification and estimation would seem, intrinsically, to be a *multicriteria* decision problem. Any model will typically entail various conceptually distinct types of model specification error, and a researcher undertaking the estimation of the model would presumably want *each* type of error to be small.

The apparent explanation for the minor MCDM role is that standard econometric and statistical techniques routinely require researchers to cast their inference problems in an all-encompassing stochastic framework. As will be discussed more carefully in subsequent sections, the actual data generating process is assumed to be describable by means of some well-defined probability distribution either objectively, i.e., apart from any observer, or subjectively, as a coherent reflection of a researcher's beliefs. Within this all-encompassing stochastic framework, discrepancy terms arising from model misspecification are interpreted as random quantities governed by joint probability distributions. The determination of the separate and joint behavior of the theoretical variables in relation to process observations can then be analyzed in terms of a likelihood function or a posterior probability distribution. The problem of reconciling imperfect theory with observations is thus transformed into the problem of determining the most probable parameter values for a stochastic model whose structural form is assumed to be correctly and completely specified.

What are the strengths and weaknesses of this standard approach? On the plus side, it provides a powerful and elegant way in which to scale and weigh disparate sources of information. All discussion of theoretical variables is conducted in terms of assumed joint probability relations, so that a common level of abstraction is achieved. This permits the construction of a single *real-valued* measure of incompatibility (goodness of fit) between theory and observations, e.g., the construction of a likelihood function. To use an analogy from decision theory, it is as if the preferences of decision makers with potentially conflicting objectives could always be represented in aggregate form by a single real-valued utility function.

On the minus side, it forces an inferential study to proceed under the generally false presumption of correct model specification. This standard "null hypothesis" is to be employed even when a researcher is fully aware that he has resorted to conventional or otherwise arbitrary probability assessments for model discrepancy terms. Residuals (estimates for the model discrepancy terms) can of course

subsequently be subjected to various diagnostic procedures to check for model misspecification. Yet the fact remains that all incompatibilities between theory and observations, whatever their actual source, are forced to reveal themselves as inconsistencies between postulated and empirical *probability* relations; the cross-sectional, dynamic, or measurement relations tend to be pushed into the background or lost sight of entirely through various analytical manipulations. Untangling the true source of a diagnosed specification problem can thus be difficult.

In Refs. [14–19] the problem of model specification and estimation is re-examined from a multicriteria perspective. A framework is developed which encompasses a broad range of views concerning the appropriate interpretation and treatment of model discrepancy terms. On the one hand, conceptually distinct discrepancy terms can be considered without amalgamation, as illustrated by the “flexible least squares” (FLS) approach. The basic FLS objective is to determine the set of estimates that are “cost efficient” in the sense that no other estimates yield uniformly smaller discrepancy terms. Alternatively, when appropriate, joint probability assessments can be used to achieve a complete amalgamation of the discrepancy terms into a single real-valued measure of theory and data incompatibility.

Section 2 illustrates the FLS approach for a time-varying linear estimation problem in which a researcher is unable or unwilling to provide probability assessments for model discrepancy terms. Section 3 contrasts the FLS handling of this problem with the standard inferential approach in which probability assessments for discrepancy terms are assumed to be available. A variety of FLS simulation studies and empirical applications are reviewed in Section 4. A more general multicriteria framework for model specification and estimation is outlined in Section 5. Section 6 discusses the relationship of this multicriteria framework to previous uses of multicriteria methods in econometrics, statistics, and systems science. Final remarks are given in Section 7.

## 2. The FLS Approach: An Illustrative Example

Suppose scalar observations  $y_1, y_2, \dots, y_T$  have been obtained on a process at successive time points  $1, 2, \dots, T$ . The basic estimation objective is to understand the way in which the process has evolved over the course of the observation period.

The state of the process at each time  $t$  is described by an  $N \times 1$  column vector  $x_t$  of unknown process attributes. For example, for a time-varying linear regression problem,  $x_t$  might simply be a listing of the time  $t$  regression coefficients. For an economic growth problem,  $x_t$  might include stocks of real and financial assets available at time  $t$ , together with various structural parameters characterizing the objectives and constraints faced by firms and households.

The relationship between the observation  $y_t$  and the state vector  $x_t$  at each time  $t$  is postulated a priori to be approximately linear. In addition, the evolution of the state vector  $x_t$  – although not well understood a priori – is postulated to be gradual

in the sense that  $x_t$  undergoes at most a small change from one observation time to the next. These prior postulates of approximately linear measurement and gradual state evolution are modelled as follows:

**Measurement Relations [Approximate Linearity].**

$$y_t - h'_t x_t \approx 0, \quad t = 1, \dots, T, \quad (1)$$

where  $h'_t$  is a  $1 \times N$  row vector of known exogenous variables.

**Dynamic Relations [Gradual State Evolution].**

$$x_{t+1} - x_t \approx 0, \quad t = 1, \dots, T - 1. \quad (2)$$

In accordance with the basic estimation objective, suppose an attempt is now made to determine all possible estimates  $\hat{X}_T = (\hat{x}_1, \dots, \hat{x}_T)$  for the state sequence  $X_T = (x_1, \dots, x_T)$  that are minimally incompatible with the given theoretical relations (1) and (2), conditional on the given observation sequence  $Y_T = (y_1, \dots, y_T)$ . The multicriteria nature of this estimation problem is seen as follows. Two conceptually distinct types of discrepancy terms can be associated with each possible state sequence estimate  $\hat{X}_T$ . First, the choice of  $\hat{X}_T$  could result in nonzero measurement discrepancy terms  $y_t - h'_t \hat{x}_t$  in (1). Second, the choice of  $\hat{X}_T$  could result in nonzero dynamic discrepancy terms  $\hat{x}_{t+1} - \hat{x}_t$  in (2). In order to conclude that the theoretical relations (1) and (2) are in reasonable agreement with the observations, each type of discrepancy would have to be small in some sense.

Suppose a measurement cost  $c_M(\hat{X}_T, Y_T, T)$  and a dynamic cost  $c_D(\hat{X}_T, Y_T, T)$  are separately assessed for the two disparate types of discrepancy terms entailed by the choice of a state sequence estimate  $\hat{X}_T$ . These costs represent the degree to which nonzero discrepancy terms are viewed as undesirable. For illustration, suppose these costs take the form of sums of squared discrepancy terms, implying that positive and negative discrepancies are viewed as equally undesirable. More precisely, for any given state sequence estimate  $\hat{X}_T$ , let the measurement cost associated with  $\hat{X}_T$  be given by

$$c_M(\hat{X}_T, Y_T, T) = \sum_{t=1}^T [y_t - h'_t \hat{x}_t]^2, \quad (3)$$

and let the dynamic cost associated with  $\hat{X}_T$  be given by

$$c_D(\hat{X}_T, Y_T, T) = \sum_{t=1}^{T-1} [\hat{x}_{t+1} - \hat{x}_t]' D [\hat{x}_{t+1} - \hat{x}_t], \quad (4)$$

where  $D$  is a suitably selected positive definite scaling matrix.<sup>1</sup>

<sup>1</sup> The scaling matrix  $D$  can be specified so that the FLS estimates obtained below for the state vectors  $x_t$  are essentially invariant to the choice of units for the components of the exogenous vectors  $h_t$ . See Ref. [40, Footnote 3].

If the prior beliefs (1) and (2) concerning the measurement and dynamic relations hold true with absolute equality, then selecting the actual state sequence  $X_T$  as the state sequence estimate would result in zero values for both  $c_M$  and  $c_D$  – the “ideal” cost point in the terminology of Yu [45, p. 67]. In all other cases, each potential state sequence estimate  $\hat{X}_T$  will entail positive measurement and/or dynamic costs. Nevertheless, not all of these state sequence estimates are equally interesting. In particular, a state sequence estimate  $\hat{X}_T$  that is dominated by another estimate  $X_T^*$ , in the sense that  $X_T^*$  yields a lower value for one type of cost without increasing the value of the other, should presumably be excluded from consideration.

Attention is therefore focused on the set of undominated state sequence estimates. Such estimates are referred to as *flexible least squares* (FLS) estimates. Each FLS estimate shows how the process state vector could have evolved over time in a manner minimally incompatible with the prior measurement and dynamic relations (1) and (2). Without additional modelling criteria, restricting attention to any proper subset of the FLS estimates is an arbitrary decision. Consequently, the FLS approach envisions the generation and consideration of a representative sample of the FLS estimates in order to determine the similarities and divergencies displayed by these potential state sequences. The similarities might be used to construct more structured hypotheses regarding the measurement and evolution of the state vector. The divergencies reflect the uncertainty inherent in the problem formulation regarding the true nature of the underlying process.

Define the *cost possibility set* to be the collection

$$C(T) = \{c_D(\hat{X}_T, Y_T, T), c_M(\hat{X}_T, Y_T, T) \mid \hat{X}_T \in R^{TN}\} \tag{5}$$

of all possible configurations of dynamic and measurement costs attainable at time  $T$ , conditional on the given observation sequence  $Y_T$ . In analogy to the usual Pareto-efficient frontier, the *cost-efficient frontier*  $C^F(T)$  is then defined to be the collection of all undominated cost vectors  $c = (c_D, c_M)$  in  $C(T)$ , i.e., all cost vectors  $c$  in  $C(T)$  for which there exists no other cost vector  $c^*$  in  $C(T)$  satisfying  $c^* \leq c$  with  $c^* \neq c$ . Formally, letting  $\text{vmin}$  denote vector minimization,

$$C^F(T) = \text{vmin } C(T). \tag{6}$$

By construction, then, the cost-efficient frontier is the collection of all cost vectors associated with the FLS state sequence estimates.

If the  $N \times T$  matrix  $[h_1, \dots, h_T]$  has full rank  $N$ , the cost-efficient frontier  $C^F(T)$  is a strictly convex curve in the  $c_D$ – $c_M$  plane giving the locus of vector-minimal costs attainable at time  $T$ , conditional on the given observations. In particular, as depicted in Figure 1,  $C^F(T)$  reveals the measurement cost  $c_M$  that must be paid in order to achieve a zero dynamic cost  $c_D$ , i.e., time-constant state vector estimates.

Once the FLS estimates and the cost-efficient frontier have been determined, three different levels of analysis can be used to investigate the degree to which the theoretical relations (1) and (2) are incompatible with the observations  $y_1, \dots, y_T$ .

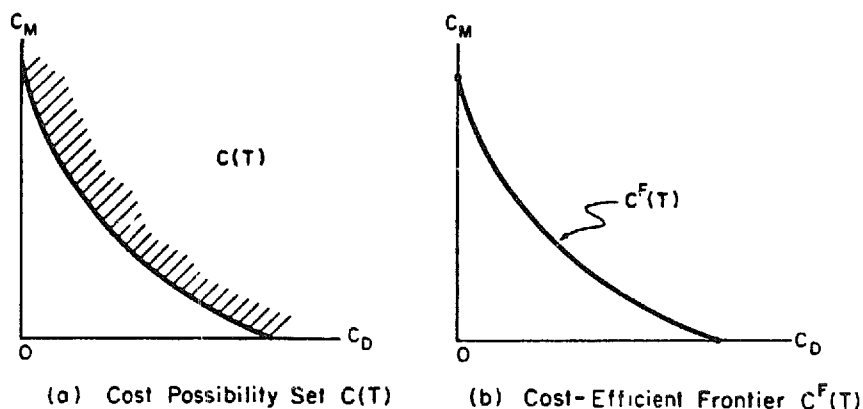


Fig. 1.

First, one can determine the efficient attainable trade-off between the measurement and dynamic costs  $c_M$  and  $c_D$  at any point  $\mu$  along the cost-efficient frontier, where  $\mu$  denotes the slope of the frontier multiplied by  $-1$ ; i.e.,  $\mu \equiv -dc_M/dc_D$ . Second, one can generate the FLS estimates whose cost vectors correspond to a rough grid of  $\mu$ -points spanning the frontier. Each of these FLS estimates yields a possible time path for the actual state vector, and summary descriptive statistics (e.g., average value and standard deviation) constructed for these estimates can be used to indicate the extent to which the state vector evolves over time. Finally, the time-paths traced out by the FLS estimates can be directly examined for evidence of systematic and possibly idiosyncratic time variations in individual state variables that are difficult to discern from summary statistical characterizations. Various simulation and empirical studies making use of this three-stage FLS analysis are discussed in Section 4, below.

In summary, the basic FLS objective is to characterize the set of all state sequence estimates that achieve vector-minimal incompatibility between process observations and imperfectly specified theoretical relations, whatever form these theoretical relations might take. Although probability relations can be incorporated along with other types of theoretical relations (see [14, 19]), they do not play a distinguished role. Indeed, as illustrated above, they may be absent altogether. In contrast, commonly used statistical estimation techniques such as maximum a posteriori (MAP) and maximum likelihood estimation are point estimation techniques that attempt to determine the most probable state sequence estimate for a stochastic model whose structure is assumed to be correctly and completely specified. The crucial distinction between the two approaches lies in the use of probability theory to transform potentially disparate model discrepancy terms into apparently commensurable quantities.

The next section illustrates this distinction by re-formulating the state estimation problem (1) and (2) in accordance with standard statistical practice.

### 3. Standard Approach to the Problem in Section 2

Suppose scalar observations  $y_1, \dots, y_T$  obtained on a process are pos.ulated to be approximately linearly related to a sequence of state vectors  $x_1, \dots, x_T$ . The prior measurement relations take the following form.

#### Measurement Relations [Approximate Linearity].

$$y_t = h'_t x_t + v_t, \quad t = 1, \dots, T, \tag{7}$$

where  $x_t$  denotes an  $N \times 1$  column vector of unknown state variables,  $h'_t$  denotes a  $1 \times N$  row vector of known exogenous variables, and  $v_t$  denotes a scalar measurement discrepancy term.

If no restrictions are placed on the discrepancy term  $v_t$ , then Eq. (7) is simply a defining relation for  $v_t$ . That is,  $v_t$  is a slack variable, and Eq. (7) is true by definition whether or not an approximately linear relation exists between  $y_t$  and  $x_t$  in actuality. The slack variable  $v_t$  depends on everything affecting  $y_t$  that is not captured by the term  $h'_t x_t$  – that is, everything unknown, or not presumed to be known, about how  $y_t$  might depend on higher order terms in  $x_t$ , on missing variables, and so forth. To give content to the prior of “approximately linear measurement,” the discrepancy term  $v_t$  must further be restricted to be small in some sense.

Suppose in addition to (7) that the state vector  $x_t$  is assumed to evolve gradually over time. The prior dynamic relations take the following form.

#### Dynamic Relations [Gradual State Evolution].

$$x_{t+1} = x_t + w_t, \quad t = 1, \dots, T - 1, \tag{8}$$

where the  $N \times 1$  vector  $w_t$  denotes a dynamic discrepancy term.

As before, if no restrictions are placed on the discrepancy term  $w_t$ , then Eq. (8) simply defines  $w_t$  to be a slack variable incorporating everything unknown, or not presumed to be known, about how the differenced state vector  $[x_{t+1} - x_t]$  depends on higher-order terms in  $x_t$ , on missing variables, and so forth. Consequently, as it stands, Eq. (8) is true regardless of the actual relation between  $x_{t+1}$  and  $x_t$ . To give content to the prior of “gradual state evolution,” the discrepancy term  $w_t$  must further be restricted to be small in some sense.

If no additional theoretical relations are introduced, the estimation problem described above is simply an alternative representation for the multicriteria estimation problem outlined in Section 2. Each possible estimate for the state sequence  $(x_1, \dots, x_T)$  entails two conceptually distinct apple-and-orange types of discrepancy terms – measurement and dynamic – and a researcher undertaking this estimation would presumably want *each* type of discrepancy to be small.

However, standard econometric and statistical techniques invariably do introduce a third type of theoretical relation at this point in the description of

an estimation problem: namely, probability relations restricting discrepancy terms. Consider, for example, the following commonly assumed relations implying that the measurement and dynamic discrepancy terms  $v_t$  and  $w_t$  in (7) and (8) are random quantities with known probability density functions (PDFs) governing both their individual and joint behavior.

### Probability Relations.

$$(v_t) \text{ and } (w_t) = \text{mutually and serially independent processes;} \quad (9)$$

$$(\text{PDF for } v_t) = P_v, \quad t = 1, \dots, T; \quad (10)$$

$$(\text{PDF for } w_t) = P_w, \quad t = 1, \dots, T - 1; \quad (11)$$

$$x_t \text{ distributed independently of } v_t \text{ and } w_t \text{ for each } t, \quad (12)$$

$$(\text{PDF for } x_t) = P_x. \quad (13)$$

Since (7) and (8) are still interpreted as equations in the usual exact mathematical sense,  $v_t$  and  $w_t$  now appear in these equations as commensurable “disturbance terms” impinging on correctly specified theoretical relations. The previous interpretation for  $v_t$  and  $w_t$  as apple-and-orange discrepancy terms incorporating everything unknown about the measurement and dynamic aspects of the process is thus dramatically altered.

Once the commensurability of the discrepancy terms  $w_t$  and  $v_t$  is assumed, a single real-valued measure of theory and data incompatibility can be constructed. Specifically, combining the measurement relations (7) with the probability relations (9)–(13) permits the derivation of a probability density function  $P(Y_T|X_T)$  for the observation sequence  $Y_T = (y_1, \dots, y_T)$  conditional on the state sequence  $X_T = (x_1, \dots, x_T)$ . Combining the dynamic relations (8) with the probability relations (9)–(13) permits the derivation of a “prior” probability density function  $P(X_T)$  for  $X_T$ . The joint probability density function for  $X_T$  and  $Y_T$  then takes the form

$$P(X_T, Y_T) = P(Y_T|X_T) \cdot P(X_T). \quad (14)$$

The joint probability density function (14) elegantly combines the two distinct sources of theory and data incompatibility – measurement and dynamic– into a single *real-valued* measure of incompatibility for any considered state sequence  $X_T$ .

As detailed in [43], an objective commonly assumed for estimation problems described by relations of the form (7)–(13) is maximum a posteriori (MAP) estimation, i.e., the determination of the state sequence  $X_T$  that maximizes the posterior probability density function  $P(X_T|Y_T)$ . Since the observation sequence  $Y_T$  is assumed to be given, this objective is equivalent to determining the state sequence  $X_T$  that maximizes the product of  $P(X_T|Y_T)$  and  $P(Y_T)$ . In accordance with Bayesian rules of probability theory,

$$P(X_T|Y_T) \cdot P(Y_T) = P(Y_T|X_T) \cdot P(X_T), \quad (15)$$



where, as earlier explained, the right-hand expression can be evaluated using the relations (7)–(13). Determining a MAP state sequence is thus equivalent to determining a state sequence which minimizes the real-valued incompatibility cost function

$$c(\mathbf{X}_T, \mathbf{Y}_T, T) = -\log[P(\mathbf{Y}_T|\mathbf{X}_T)P(\mathbf{X}_T)]. \quad (16)$$

In summary, what ultimately has been accomplished by the augmentation of the measurement and dynamic relations (7) and (8) with the probability relations (9)–(13)? *The multicriteria problem of achieving vector-minimal incompatibility between imperfectly specified theoretical relations and process observations has been transformed into the single-criterion problem of determining the most probable state sequence for a stochastic model whose structure is assumed to be correctly and completely specified.*

One basic objection to this standard estimation approach is that it entails an interpretation for the discrepancy terms that is at odds with the originally specified priors (1) and (2). In particular, the time-trend smoothness prior (2) is replaced with the prior of a random walk, even though these two priors represent different conceptualizations for the movement of the underlying state vectors. The time-trend prior (2) postulates that successive state vectors evolve gradually from one time period to the next, a movement that might be captured by a straight line or a sine wave, for example. In contrast, the random walk model implies that “error terms” are persistently accumulated in successive state vectors, resulting in a nonstationary process exhibiting jagged discontinuities between successive state vectors.

It is sometimes countered that this distinction is unimportant if the *variances* of the random walk error terms are anticipated to be small. However, as stressed in recent macroeconomic work, e.g., [34], the dynamic properties of a time-trend model are altogether different from the dynamic properties of a random walk model, however one models the variances of these error terms. Consequently, “small discrepancy terms” and “small error term variances” are not conceptually interchangeable descriptions. In particular, for initial diagnostic checks of poorly understood structures, the probabilistic assumption of “small variances” can be an overly restrictive concept (cf. [35]).

Another important objection to the standard estimation approach is that the probability relations (9)–(13) imply that  $w_t$  and  $v_t$  are governed by a well-defined joint probability distribution and hence are cardinally comparable. For many processes it is hard to maintain this assumption in a publicly credible way. For example, the observations  $y_1, \dots, y_T$  might be the outcome of a nonreplicable experiment, implying that probability assessments for the discrepancy terms  $w_t$  and  $v_t$  cannot be put to an objective test. Alternatively, as stressed in Section 2, the theoretical relations (7) and (8) might represent tentatively held conjectures concerning a poorly understood process, or a linearized set of relations obtained for an analytically intractable nonlinear process. In this case it is questionable whether the discrepancy terms are governed by any meaningful probability relationships. A researcher might then have to resort to specifications determined largely by

convention if he is forced to provide a probabilistic characterization for the discrepancy terms.

A third objection to the standard estimation approach is that conceptually distinct discrepancy terms are amalgamated into a single real-valued incompatibility measure such as (16). This amalgamation makes it difficult to detect and correctly sort out which aspects of the model, if any, are seriously misspecified. There is of course no way to determine from the single real-valued measure (16) that a serious specification error has occurred, e.g., in the dynamic relations (8) rather than the measurement relations (7). In fact, (16) is constructed under the premise that no specification error has occurred, and there is no way to use it per se to check for any kind of modelling difficulty. Rather, subsequent tests must be conducted to check whether the data appear to be anomalous with respect to the given model specification, or whether other plausible model specifications exist that make the data appear less anomalous.

A further difficulty here, as detailed in Ref. [19, Section 5.1], is that standard diagnostic procedures force all incompatibilities between theory and observations to reveal themselves as incompatibilities between theoretically anticipated probability relations and empirically determined statistical properties. For example, suppose the dynamic relations (8) are fundamentally misspecified because the true dynamic dependence of  $x_{t+1}$  on  $x_t$  is highly nonlinear. Using standard diagnostic tests on the dynamic residual terms  $\hat{w}_t \equiv [\hat{x}_{t+1} - \hat{x}_t]$ , a researcher would presumably perceive that the properties of these residuals are at odds with the probability relations assumed for  $w_t$  in (9) and (11). The tendency of the researcher might then be to concentrate on modifying the probability assumptions for  $w_t$  to improve statistical fit – e.g., to replace serial independence with first-order serial correlation, or to assume that  $w_t$  has a time-varying covariance matrix – rather than to think more carefully about the actual physical or behavioral relationships connecting  $x_{t+1}$  to  $x_t$ .

These three objections to the standard estimation approach – potentially distorted priors, inappropriate and potentially misleading assumptions of cardinal comparability, and the confounding of conceptually distinct discrepancy terms – would be of purely academic interest if treating discrepancy terms as commensurable random disturbance terms constituted the only way to obtain estimates for unknown process states. However, Section 2 suggests to the contrary that an alternative multicriteria treatment of discrepancy terms is also feasible for this purpose.

#### 4. FLS Simulation and Empirical Studies

In the previous two sections a case is made for the conceptual desirability of a multicriteria FLS approach to the estimation of process states for processes whose properties are poorly understood a priori and hence whose descriptions incorporate potentially significant specification errors. Not yet examined, however,

is the extent to which the FLS approach permits the recovery of *accurate* information about process states. The present section briefly reviews a number of simulation and empirical studies that have addressed this issue.

Ref. [17] undertakes an FLS analysis of a time-varying linear regression problem, a special case of (1) and (2) in which the time  $t$  state vector  $x_t$  denotes the vector of time  $t$  regression coefficients and the time  $t$  exogenous vector  $h_t$  denotes the vector of time  $t$  regressor variables. The basic estimation objective is to determine whether the regression coefficients have exhibited any systematic time-variation over the course of the observation period.

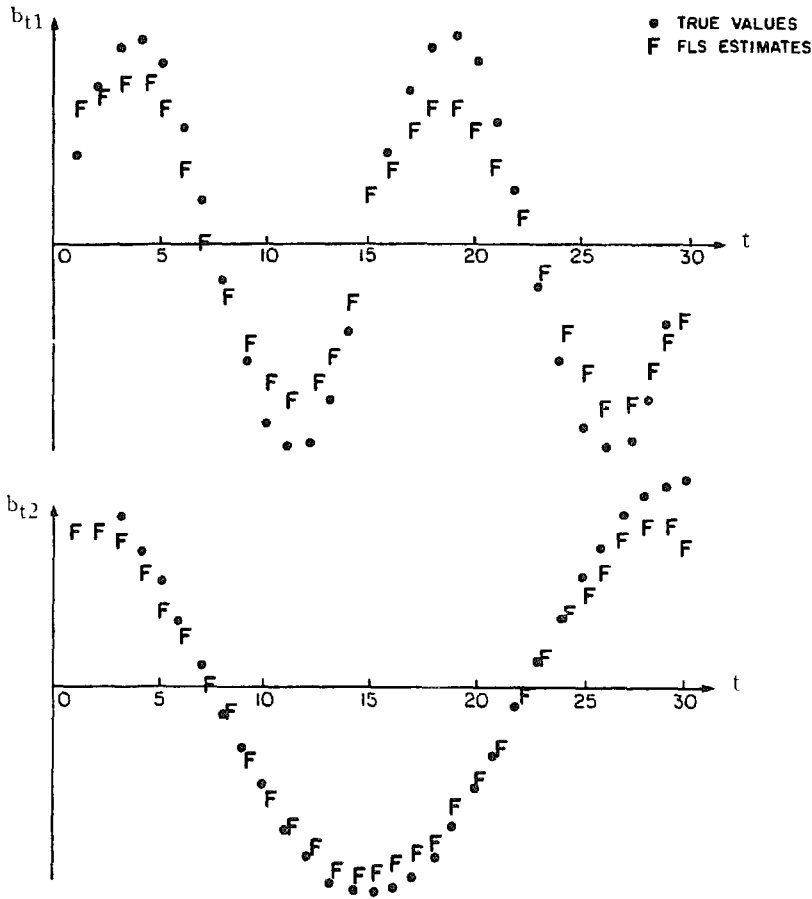
A Fortran program for generating the FLS estimates is provided in Ref. [17], together with an explanation of the program logic.<sup>2</sup> Various FLS simulation experiments making use of this program are reported and graphically depicted in [13, 17]. These experiments demonstrate the ability of the FLS method to track and recover linear, quadratic, sinusoidal, and elliptical motions in the true underlying regression coefficients, despite noisy observations, and relying only on prior measurement and dynamic relations of the form (1) and (2). Indeed, the motions are recovered with good qualitative accuracy all along the FLS frontier.

For example, experiments were carried out for which the components of the true two-dimensional coefficient (state) vectors  $x_t = (b_{t1}, b_{t2})$  were simulated to be sinusoidal functions of  $t$ . The first component,  $b_{t1}$ , moved through two complete periods of a sine wave over the interval of time from  $t = 1$  to  $t = 30$ , and the second component,  $b_{t2}$ , moved through one complete period of a sine wave over this same time interval. Each observation  $y_t$  was generated in accordance with the linear regression model  $y_t = h'_t x_t + v_t$ , where the components of the regressor vector  $h'_t$  were taken to be deterministic cyclic functions of  $t$  and the components of the measurement discrepancy term  $v_t$  were independently generated from a pseudo-random number generator for a normal distribution  $N(0, 0.5)$ .

As depicted in Fig. 2, the FLS estimates for  $b_{t1}$  and  $b_{t2}$  closely tracked the true values for these coefficients both qualitatively and quantitatively at the point  $\mu = 1$  along the cost-efficient frontier. As  $\mu$  was increased from 1 to 1000 by powers of ten, the FLS estimates were pulled steadily inward toward the zero dynamic cost (ordinary least squares) solution,  $(b_{t1}, b_{t2}) = (0.03, 0.04)$  for  $t = 1, \dots, 30$ . Nevertheless, for each  $\mu$ , the two-period and one-period sinusoidal motions of the true coefficients were still reflected. Thus, 60 coefficients were recovered from only 30 observations, with good qualitative accuracy, all along the cost-efficient frontier.

Although these simulation experiments indicate that the FLS estimates are able to track smooth motions in the regression coefficients, the question remains whether discontinuous motions cause the FLS method to fail. This issue arose in the FLS money demand study [40], for the focus of the study concerned possible

<sup>2</sup>This FLS program for time-varying linear regression has recently been incorporated into the statistical package SHAZAM; see [44], or email [info@shazam.econ.ubc.ca](mailto:info@shazam.econ.ubc.ca) for information. See also [18] for a more general FLS Fortran program, GFLS, applicable for systems characterized by approximately linear measurement and dynamic relations



Sine Wave Experiment with Parameter Values  $\sigma=0.05, \mu=1, \text{ and } T=30$

Fig 2.

step-function breaks in money demand regression coefficients. Various simulation experiments were therefore conducted in which the components of the true regression coefficients were shifted idiosyncratically at various points in time. Surprisingly, using only measurement and smoothness priors analogous to (1) and (2), the FLS estimates were able to track and recover these step-function shifts with good qualitative accuracy all along the cost-efficient frontier despite the absence of any prior knowledge concerning the timing, number, and magnitude of the shifts. Indeed, the larger the magnitude of the shifts, the better the accuracy of the estimates.

To understand this seeming paradox, consider what happens if an underlying true linear regression coefficient  $\beta_{it}$  undergoes a single step-function shift from its current value  $b$  to a new value  $b'$  at some time  $t = t'$ . If the FLS estimate  $\hat{\beta}_{it}$  for  $\beta_{it}$  is equal to  $b$  for  $t < t'$ , and if it remains at  $b$  over the remainder of the observation period from  $t'$  to  $T$  despite the shift in  $\beta_{it}$  at  $t = t'$ , then the result is an accumulation

of measurement costs over  $t'$  to  $T$ ; and the larger the magnitude of the shift, the larger the accumulation of measurement costs. On the other hand, if  $\hat{\beta}_t$  were likewise to shift from  $b$  to  $b'$  at  $t = t'$ , the result would be a one-time dynamic cost but *no* subsequent accumulation of measurement costs. Thus, cost-minimization considerations will generally dictate that the FLS estimates should shift in response to shifts in the underlying true coefficients as long as the shifts are spaced sufficiently far apart and do not occur close to the final observation time  $T$ .

Given the promising nature of these shift simulation results, the FLS method was next used in [40] to undertake an empirical money demand investigation. Measurement and dynamic relations analogous to (1) and (2) were used to model US money demand over the volatile period 1959: Q2–1985: Q3. In particular, no prior information regarding possible shift times was used in the FLS estimation procedure. The time paths traced out by the FLS coefficient (state) estimates were found to exhibit a clear-cut downward shift in 1974, during the time of the first OPEC oil price shock, at each tested point along the cost-efficient frontier. This finding was in accordance with previous ordinary least squares (OLS) studies of US money demand that had investigated the possibility of a 1974 shift in the money demand regression coefficients using variants of the Chow test and recursive least squares.

In addition, however, the FLS results in [40] also indicated the presence of systematic idiosyncratic time variations in the regression coefficients – e.g., a sharp and steady decline in the coefficient for the inflation rate – which Chow tests and recursive least squares are not designed to detect. Moreover, the “unit root” nonstationarity problem reported in these previous OLS money demand studies was seen to disappear once the FLS coefficient estimates were allowed to exhibit even small amounts of time variation in accordance with the dynamic smoothness prior (2).

A number of other empirical FLS studies have recently appeared that suggest the potential usefulness of FLS as a diagnostic tool. For example, Dorfman and Foster [6] use FLS to develop a new measure of productivity change. They assume that measurement errors are independently and identically distributed random variables whereas the coefficients characterizing the production relation evolve slowly over time in an unknown deterministic manner. Under these assumptions they are able to provide a statistical interpretation for their FLS coefficient estimates and hence also for their FLS measure of productivity change.

Dorfman and Foster then apply their FLS productivity measure to US agricultural data for the period 1948–1983. They compare the FLS measure with two more traditional measures that assume time-constant production function parameters—total factor productivity, and a measure of technical change based on the elasticity of production with respect to time. They find (pp. 286–288) that the FLS measure is more stable than these latter measures in the sense of having a smaller variance around a constant percentage growth rate. Interestingly, the FLS measure also produces considerably lower estimates of productivity growth than the total factor productivity measure and generally higher estimates of productivity growth than the elasticity measure.

In [30], Lütkepohl uses FLS to obtain detailed information on the variability of individual coefficients for a US money demand relation specified in error-correction form. He shows (Fig. 1, p. 735) that all long-run coefficients are relatively stable over the 33-year period 1954–1987, with the least stable being the coefficient on the short-term interest rate (proxied by the discount rate on 91-day Treasury bills) and the most stable being the coefficient on transactions volume (proxied by real GNP). On the other hand, the short-run coefficients on rates of change in the interest rate, transactions volume, and the general price level (proxied by the GNP deflator) are considerably more volatile than the long-run coefficients over this same period. He concludes (p. 742) that these FLS findings are consistent with a financial innovations explanation of money demand instability over this period.

In a different study, Lütkepohl and Herwartz [31] generalize the FLS time-varying linear regression method developed in [17, 18] by allowing for anticipated seasonal periodicities as well as for time trends. They first undertake a study of their generalized FLS algorithm for three artificially generated time series, each having a seasonal pattern, in which a variable is linearly dependent on its value in some past time. In the first model, the intercept and slope coefficients are both time invariant; in the second model, these coefficients are both periodic; and in the third model, the intercept is periodic and undergoes a structural shift in mid-sample whereas the slope coefficient is time invariant. They show (Tables 1 and 2) that their generalized FLS measure is able to detect the time invariance of the coefficients in the first model and the coefficient periodicities in the second and third models, as well as the structural shift for the third model.

Lütkepohl and Herwartz then use their generalized FLS measure to study actual consumption and income time series data for the (West) German economy, with the goal of detecting specific types of coefficient variations and identifying any coefficients that appear to be time-invariant. They interpret their FLS findings for income as evidence in favor of a model in which the intercept is periodic and remaining coefficients are time invariant; and they interpret their FLS findings for consumption as evidence in favor of a model with time-invariant coefficients for first and second-order lagged terms but with periodically varying coefficients for the intercept and higher-order lagged terms.

Finally, Schneider [36] carries out an extensive comparative study between maximum likelihood (EM and scoring) and FLS time-varying linear regression methods, where the latter is characterized (p. 192) as a descriptive variant of Kalman filtering that constitutes a “simple but powerful tool of exploratory data analysis.” He first applies FLS as a preliminary descriptive stability test to a standard Goldfeld-type model of money demand for (West) Germany. As depicted in his Figs. 14.2–14.7 (pp. 206–208), he concludes (p. 211) that only the coefficients for the short-term and long-term interest rates and the 90-day swap rate exhibit a distinct time-varying behavior. In particular, he notes that the behavior of the short-term interest rate is particularly remarkable: an apparent stabilization from 1974 onward that coincides with the date when the German central bank officially switched from an interest-rate target regime to a money target regime. [In subsequent discussion (p. 212) he notes that an FLS argument can also be made for

a step change in the swap rate in 1974, the introduction date of flexible exchange rates.] In support of these conclusions, he notes that the patterns in the individual paths of the FLS coefficient estimates persist over a large portion of the cost-efficient frontier.

Schneider next checks what paths for the regression coefficients are picked out by maximum likelihood (ML) when the descriptive dynamic and measurement costs  $c_D$  and  $c_M$  are reinterpreted as elements of a likelihood function generated from a random walk model for the regression coefficients. First cautioning that little is known about the sampling distributions of ML estimators for this time-varying linear regression model, he concludes (p. 212) that the movements in the coefficients for the short-term interest rate and the swap rate appear to be significantly identified to be time-varying at a type I error level of 1%. The estimates for the variances are low, reflecting the fact that the random walk model spreads out the time variation over the entire sample period. Although the FLS-apparent step changes in the short-term interest rate and swap rate are thus considerably smoothed, the ML-estimated paths for the individual coefficients nevertheless exhibit the same general features as the FLS-estimated paths. This is seen in his Figs. 14.8–14.13 (pp. 214–217) depicting two-standard-deviation bands about the means of proxied *a posteriori* distributions for the regression coefficients, conditional on ML estimates for remaining structural parameters.

## 5. Generalizations

In previous sections it is shown how FLS can be used to investigate the basic incompatibility of theory and data for processes characterized by approximately linear measurement relations and gradual state evolution. In this section we describe a more general multicriteria approach to estimation developed in [19]. We also suggest how the latter approach might be recast in the form of a utility maximization problem subject to a budget constraint.

Consider a situation in which a sequence  $Y_T = (y_1, \dots, y_T)$  of noisy observations  $y_t$  has been obtained on some process of interest. The basic objective is to learn about the sequence of states  $X_T = (x_1, \dots, x_T)$  through which the process has passed.

Suppose the degree to which each possible state sequence estimate  $\hat{X}_T$  is incompatible with the given observation sequence  $Y_T$  is measured by a  $K$ -dimensional vector  $c(\hat{X}_T, Y_T, T)$  of incompatibility costs. These costs may represent penalties imposed for failure to satisfy criteria *conjectured* to be true (theoretical relations), and also penalties imposed for failure to satisfy criteria *preferred* to be true (objectives). Let  $C(T)$  denote the set of all incompatibility cost vectors  $c = c(\hat{X}_T, Y_T, T)$  corresponding to possible state sequence estimates  $\hat{X}_T$ . The *cost-efficient frontier*, denoted by  $C^F(T)$ , is then defined to be the collection of undominated cost vectors  $c$  in  $C(T)$ . That is, a cost vector  $c$  in  $C(T)$  is an element of  $C^F(T)$  if and only if there exists no other cost vector  $c^*$  in  $C(T)$  satisfying  $c^* \leq c$  with  $c^* \neq c$ .

By construction, the state sequence estimates  $\hat{X}_T$  whose cost vectors attain the cost-efficient frontier are characterized by a basic efficiency property: For the given observations, no other possible state sequence estimate yields lower incompatibility cost with respect to each of the  $K$  modelling criteria included in the incompatibility cost vector. Each of these state sequence estimates thus represents one possible way the actual process could have evolved over time in a manner minimally incompatible with the prior theoretical relations and objectives.

The basic multicriteria estimation problem can be summarized as follows:

**The Basic Multicriteria Estimation Problem:** Given a process length  $T$ , an observation sequence  $Y_T$ , and a vector-valued incompatibility cost function  $c(\cdot, Y_T, T)$ , determine all possible state sequence estimates  $\hat{X}_T$  that vector-minimize the incompatibility cost  $c(\hat{X}_T, Y_T, T)$ . That is, determine all possible state sequence estimates  $\hat{X}_T$  whose cost vectors  $c(\hat{X}_T, Y_T, T)$  attain the cost-efficient frontier  $C^F(T)$ .

The cost-efficient frontier  $C^F(T)$  can be obtained by means of a multicriteria extension of the usual scalar dynamic programming equations.<sup>3</sup> Consider the estimation problem at any intermediate time  $t$ . Suppose a  $K$ -dimensional vector  $c(\hat{X}_t, Y_t, t)$  of incompatibility costs can be associated with each  $t$ -length state sequence estimate  $\hat{X}_t = (\hat{x}_1, \dots, \hat{x}_t)$ , conditional on the sequence of observations  $Y_t = (y_1, \dots, y_t)$ . Let  $C(\hat{x}_t, t)$  denote the set of all cost vectors  $c(\hat{X}_t, Y_t, t)$  attainable at time  $t$ , conditional on the time- $t$  state estimate being  $\hat{x}_t$ ; and let  $C^F(\hat{x}_t, t)$  denote the cost-efficient frontier for  $C(\hat{x}_t, t)$ . Given certain regularity conditions, it is shown in [19] that the state-conditional frontier at any intermediate time  $t$  is mapped into a state-conditional frontier at time  $t + 1$  in accordance with a vector-valued recurrence relation having the form

$$C^F(\hat{x}_{t+1}, t+1) = \text{vmin} \left( \bigcup_{\hat{x}_t} [C^F(\hat{x}_t, t) + \Delta c(\hat{x}_t, \hat{x}_{t+1}, y_{t+1}, t+1)] \right), \quad (17)$$

where  $\text{vmin}$  denotes vector-minimization and  $\Delta c(\cdot)$  denotes a vector of incremental costs associated with the state transition  $(\hat{x}_t, \hat{x}_{t+1})$ . The cost-efficient frontier at the final time  $T$  is then given by

$$C^F(T) = \text{vmin} \left[ \bigcup_{\hat{x}_T} C^F(\hat{x}_T, T) \right]. \quad (18)$$

Three well-known state estimation algorithms are derived in [19] as single-criterion special cases of the multicriteria recurrence relations (17) and (18): namely, the Kalman filter [21], the Viterbi filter [8, 41], and the Larson–Peschon filter [26] for sequentially generating maximum a posteriori (MAP) probability estimates. In

<sup>3</sup>General multicriteria dynamic programming algorithms have previously been developed by a variety of other researchers. See, for example, Ref. [29].



addition, an algorithm for sequentially generating the FLS estimates for the problem discussed in Section 2, above, is derived as a bicriteria special case of (17) and (18).

Finally, it is interesting to note that the basic multicriteria estimation problem outlined above can be recast as a problem of utility maximization subject to constraint. That is, one can include in the cost vector only those costs corresponding to criteria conjectured to be true (i.e., theoretical relations), so that the resulting cost-efficient frontier depicting the feasible efficient trade-offs among model discrepancy terms is analogous to a “budget constraint.” One could then superimpose on this frontier the indifference curves for a researcher’s “utility function” that assigns a utility value to each possible configuration of costs (discrepancy terms), thus permitting for that researcher the selection of a unique “best” model specification along the frontier. In this way it might be possible to separate the subjective selection of a model based on properties preferred by individual researchers from the more objective identification of model specifications that are efficient with regard to possible trade-offs among discrepancy terms.

## 6. Relation to Previous Work

Roughly stated, multicriteria decision making (MCDM) is the study of decision situations in which one or more agents with potentially conflicting objectives must somehow decide on the implementation of an action. Due in large part to the seminal work of Charnes and Cooper, Yu, Zeleny and others dating back to the early nineteen sixties, MCDM has now become an established interdisciplinary field that cuts across the boundary lines separating operations research, management science, systems science, computer science, applied mathematics, psychology, and many other disciplines. See, for example, Refs. [7, 24, 38, 45–47].

The duality between decision making (control) and estimation (system identification) for single-criterion optimization problems has been known for over 30 years ([21, p. 42]). Surprisingly, however, the interconnections between *multicriteria* decision making and *multicriteria* estimation have yet to be systematically explored.

Some use of multicriteria methods has of course occurred in statistical inferential studies. Multicriteria methods have traditionally been used to describe the trade-off between Type I and Type II errors. In addition, multicriteria methods have been used to describe the trade-off between bias and variance (fidelity and smoothness) which some estimation procedures entail. See, for example, the discussion of ridge trace procedures in [12, pp. 915–916], the discussion of smoothing splines in [42], and the discussion in [9] of penalized likelihood methods for the location and probabilistic evaluation of “bumps” in estimated probability densities.

One also finds instances in which researchers have advocated using multicriteria methods for other types of estimation purposes. For example, in the systems literature, Benedict and Bordner [3] proposed a bicriteria estimation algorithm for

a class of radar tracking problems. Moreover, various researchers have proposed using bicriteria methods for handling the dual objectives of system optimization and system identification which arise for “dual control” problems, i.e., for problems in which an agent is attempting to control a system at the same time he is attempting to learn about its characteristics. See, for example, [10, 25]. In the MCDM literature, both Narula and Wellington [33] and Zeleny [46, pp. 469–471] have proposed the use of multicriteria methods for linear regression analysis. Also, Charnes and Cooper have developed a “data envelopment analysis” method for the estimation of the Pareto-efficient frontier of an empirically determined multi-input/multi-output production function. The method has been used to classify organizations using the same kinds of inputs and outputs either as efficient or inefficient; see [1]. Other potential applications of this method are discussed in [4, 37].

In the econometrics literature, Leamer [27, pp. 141–170] introduces the notion of an “information contract curve” in the space of regression coefficients to discuss regression selection strategies in the case in which only the contours (iso-density surfaces) of the prior probability density function and the likelihood function are known. Specifically, the information contract curve is a locus of points giving all feasible estimates for the regression coefficient vector which are efficient relative to two potentially conflicting criteria: maximization of the prior probability density function specified as a contour map; and maximization of the sample-conditioned likelihood function specified as a contour map.

In subsequent work (see [28]), Leamer proposes a more general “global sensitivity analysis” for investigating the sensitivity of posterior distribution inferences to alternative choices of prior probability distributions. A related line of work on “set-valued filtering” has been developed in the systems science and statistics literatures; see, for example, [39, Section V.B; 32]. These studies argue that unique estimates cannot be inferred from data sets when, for whatever reason, a data analyst is unable to use probability assessments to fully scale and weigh disparate sources of information in the form of a uniquely specified posterior probability distribution.

Many statisticians, econometricians, and systems scientists are either unwilling or unable to undertake a complete scalarization of their estimation problems in the form of a posterior probability distribution. Nevertheless, rather than considering the sensitivity of inferences to alternative prior probability distributions, the majority of these researchers instead rely on ordinary least squares and maximum likelihood methods for initial estimation purposes, followed by subsequent diagnostic testing to check for model misspecification.

Hendry and Richard [11] have attempted to systematize the latter model specification procedure. They formulate various model design criteria which they believe to be of particular relevance for econometric modelling. For any one model, these model design criteria could be construed as constituting an incompatibility cost vector  $c$  in the sense of Section 5, above. However, Hendry and Richard do not attempt to determine the trade-offs among the criteria in accordance with any systematic multicriteria (vector optimization) procedure.

Rather, as is standard in the diagnostic testing literature, Hendry and Richard advocate the sequential application of their model design criteria, opening themselves to the usual criticism (see, e.g., [12, pp. 869–870]) that the choice of a final model might depend upon the particular order of application.<sup>4</sup> One way to interpret this path-dependence criticism is to note that Hendry and Richard may simply be ending up at one among many possible points on a frontier of models that are all equally acceptable (efficient) relative to their postulated set of criteria. In other words, assuming that the various criteria represent an over-identifying set of constraints, a systematic multicriteria treatment of the modelling problem would necessarily lead to a *set* of efficient models rather than to a uniquely determined specification.

In summary, although the MCDM literature has apparently not had much of an impact on econometric and statistical procedures to date, some preliminary steps toward a full-blown multicriteria approach have been taken. Leamer considers the trade-offs between a prior and a data-based conception of a best estimate. Hendry and Richard formalize a set of potentially conflicting model design criteria which they argue will be objectively meaningful to other researchers. In terms of the general multicriteria framework outlined in Section 5, above, the differences separating these two approaches reduce to a different dimension  $K$  for the basic cost vector  $c$ , a different idea concerning which model criteria should be included in  $c$ , and a different degree of recognition that conflicting model criteria result in *set*-valued inferences in the form of a nondegenerate cost-efficient *frontier* of alternative models.

Our work on multicriteria estimation has its roots in “Sridhar filtering.” In a series of studies initiated in the mid-nineteen sixties focusing on continuous-time rigid-body dynamics (see, e.g., Refs. [2, 5] ), R. Sridhar and other associates explored the idea of forming a cost-of-estimation function as a weighted sum of squared dynamic and measurement discrepancy terms. In Refs. [14, 15] we extend this previous work by considering a broader class of models and by deriving exact filtering equations for the determination of the cost-minimizing solutions. In a related study, Kohn and Ansley [22] discuss the relation between the use of Bayesian smoothness priors for state-space smoothing and the use of a Sridhar-type penalized least squares criterion function with quadratically specified dynamic and measurement costs to achieve optimal function smoothing. However, as in the earlier Sridhar studies, the cost-of-estimation functions in these studies are still formulated with uniquely specified penalty weights.

The basic FLS approach, introduced in [16], instead focuses attention on a cost vector  $(c_D, c_M)$  incorporating separate penalty costs for dynamic and measurement discrepancy terms. This permits the construction of a “cost-efficient frontier,” a curve in a two-dimensional cost plane that provides an explicit way to determine the efficient trade-offs between dynamic and measurement discrepancy terms. Since

---

<sup>4</sup>The observation that the “decision path” can affect a final choice is also well known in the MCDM literature; see, e.g., [23]

the costs indicate the relative *undesirability* of various discrepancy term patterns rather than any intrinsic properties of the discrepancy terms per se, quadratic cost specifications – while useful for tractability – are in no sense required. As indicated in Section 5 of this paper, we now view the original FLS formulation as a special case of a more general multicriteria estimation framework in which the cost vector  $c$  can incorporate whatever modelling criteria are deemed relevant for the problem at hand.

## 7. Concluding Remarks

This paper suggests that multicriteria methods such as FLS provide a systematic way to approach the estimation of processes whose descriptions embody potentially significant specification errors. The heart of the FLS approach is the recognition that conflicting model criteria result in *set-valued* inferences in the form of a nondegenerate cost-efficient *frontier* of alternative model specifications. The power and elegance achieved by the usual scalarization through the introduction of probabilistic assumptions is impressive; but when doubt exists concerning the appropriateness of these assumptions, FLS offers a contending conceptual alternative.

## References

- [1] Banker, R.D., A. Charnes and W.W. Cooper, Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Sci.*, **30** (1984) 1078–1092.
- [2] Bellman, R., H. Kagiwada, R. Kalaba and R. Sridhar, Invariant imbedding and nonlinear filtering theory, *J. Astronaut. Sci.*, **13** (1966) 110–115.
- [3] Benedict, T.R. and G.W. Borden, Synthesis of an optimal set of radar track-while-scan smoothing equations, *IRE Trans. Automat. Control*, **7** (1962) 27–32.
- [4] Charnes, A., W.W. Cooper, B. Golany, L. Seiford and J. Stutz, Foundations of data envelopment analysis for Pareto–Koopmans efficient empirical production functions, *J. Econometrics*, **30** (1985) 91–107.
- [5] Detchmendy, D. and R. Sridhar, Sequential estimation of states and parameters in noisy nonlinear dynamical systems, *J. Basic Engrg.*, **88** (1966) 362–368.
- [6] Dorfman, J. and K. Foster, Estimating productivity changes with flexible coefficients, *Western J. of Agricult. Econom.*, **16** (1991) 280–290.
- [7] Dyer, J.S., P.C. Fishburn, R.E. Steuer, J. Wallenius, and S. Zionts, Multiple criteria decision making, multiattribute utility theory: the next ten years, *Management Sci.*, **38** (1992) 645–654.
- [8] Forney, G.D., Jr., The Viterbi algorithm, *Proc. of the IEEE* (1973) 268–278.
- [9] Good, I.J. and R.A. Gaskins, Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data, *J. Amer. Statist. Assoc.*, **75** (1980) 42–56, followed by comments, 56–73.
- [10] Haimes, Y.Y., L.S. Lasdon and D.A. Wismer, On a bicriteria formulation of the problems of integrated system identification and system optimization, *IEEE Trans. Systems Man Cybernetics* **1** (1977) 296–297.
- [11] Hendry, D.F. and J.-F. Richard, The econometric analysis of economic time series, *Internat. Statist. Rev.*, **51** (1983) 111–164.

- [12] Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.C. Lee, *The Theory and Practice of Econometrics* (Wiley, New York, 1985)
- [13] Kalaba, R., N. Rasakhoo and L. Tesfatsion, A Fortran program for time-varying linear regression via flexible least squares, *Comput. Statist. and Data Anal.*, **7** (1989) 291–309.
- [14] Kalaba, R. and L. Tesfatsion, A least-squares model specification test for a class of dynamic nonlinear economic models with systematically varying parameters, *J. Optim. Theory Appl.* **32** (1980) 538–567
- [15] Kalaba, R. and L. Tesfatsion, An exact sequential solution procedure for a class of discrete-time nonlinear estimation problems, *IEEE Trans. Automat. Control*, **26** (1981) 1144–1149.
- [16] Kalaba, R. and L. Tesfatsion, The flexible least squares approach to time-varying linear regression, *J. Econom. Dynamics Control*, **12** (1988) 43–48.
- [17] Kalaba, R. and L. Tesfatsion, Time-varying linear regression via flexible least squares, *Comput. and Math Appl.*, **17** (1989) 1215–1245.
- [18] Kalaba, R. and L. Tesfatsion, Flexible least squares for approximately linear systems, *IEEE Trans. Systems, Man Cybernet.*, **20** (1990) 978–989.
- [19] Kalaba, R. and L. Tesfatsion, An organizing principle for dynamic estimation, *J. Optim. Theory Appl.* **64** (1990) 445–470
- [20] Kalaba, R. and L. Tesfatsion, A multicriteria approach to dynamic estimation, in: R.H. Day and P. Chen (Eds.), *Nonlinear Dynamics and Evolutionary Economics*, (Oxford University Press, NY, 1993) 289–300.
- [21] Kalman, R.E., A new approach to linear filtering and prediction problems, *Trans. of the ASME J Basic Engrg.*, **82** (1960) 35–45.
- [22] Kohn, R. and C.F. Ansley, Equivalence between Bayesian smoothness priors and optimal smoothing for function estimation, in: C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Models* (Marcel Dekker, New York, 1988) 393–430.
- [23] Korhonen, P., H. Moskowitz and J. Wallenius, Choice behavior in interactive multiple criteria decision making, *Ann. of Oper. Res.*, **23** (1990) 161–179
- [24] Korhonen, P., A. Lewandowski and J. Wallenius (Eds.) *Multiple Criteria Decision Support*, Lecture Notes in Economics and Mathematical Systems, Vol. 356, (Springer, Berlin, 1991)
- [25] Koussoulas, N.T., Multiobjective optimization in adaptive and stochastic control, in: C.T. Leondes (Ed.) *Control and Dynamic Systems*, Vol. 25 (Academic Press, New York, 1987)
- [26] Larson, R.E. and J. Peschon, A dynamic programming approach to trajectory estimation, *IEEE Trans. Automat. Control*, **11** (1966) 537–540.
- [27] Leamer, E., *Specification Searches* (Wiley, New York, 1978)
- [28] Leamer, E., Sensitivity analyses would help, in: C. Giangeri (Ed.), *Modelling Economic Series* (Clarendon Press, Oxford, 1990).
- [29] Li, D. and Y.Y. Haimes, The envelope approach for multiobjective optimization problems, *IEEE Trans. Systems, Man Cybernetics*, **17** (1987) 1026–1038; for Errata Corrige, see *Ibid.*, **18** (1988) 332.
- [30] Lütkepohl, H., The sources of the U.S. money demand instability, *Empirical Econom.*, **18** (1993) 729–743.
- [31] Lütkepohl, H. and H. Herwartz, Specification of varying coefficient time series models via generalized flexible least squares, Working Paper No. 9311, Institut für Statistik & Ökonometrie, Humboldt-Universität zu Berlin (1993).
- [32] Morrell, D.R., Epistemic utility estimation, *IEEE Trans. Systems, Man Cybernet.*, **23** (1993) 129–140.
- [33] Narula, S.C. and J.F. Wellington, Linear regression using multiple criteria, in: G. Gandel and T. Gal (Eds.) *Multiple Criteria Decision Making and Applications* (Springer, New York, 1980) 266–277.
- [34] Nelson, C.R. and C.I. Plosser, Trends and random walks in macroeconomic time series: some evidence and implications, *J. Monetary Econom.*, **10** (1982) 139–162.
- [35] Ruspini, E.H., Approximate reasoning: past, present, and future, *Inform. Sci.*, **57**–**58** (1991) 297–317

- [36] Schneider, W., Stability analysis using Kalman filtering, scoring, EM, and an adaptive EM method, in: P. Hackl and A. H. Westlund (Eds.) *Economic Structural Change: Analysis and Forecasting* (Springer, New York, 1991) Ch. 14, pp. 191–221.
- [37] Seiford, L.M. and R.M. Thrall, Recent developments in DEA: the mathematical programming approach to frontier analysis. *J. Econometrics*, **46** (1990) 7–38.
- [38] Steuer, R.E., *Multiple Criteria Optimization: Theory, Computation, and Application* (Wiley, New York, 1986).
- [39] Stirling, W.C. and D.R. Morrell, Convex Bayes decision theory, *IEEE Trans. on Systems, Man Cybernetics*, **21** (1991) 173–183.
- [40] Tesfatsion, L. and J. Veitch, U.S. money demand instability: a flexible least squares approach, *J. Econom. Dynamics Control*, **14** (1990) 151–173.
- [41] Viterbi, A.J., Error bounds of convolutional codes and an asymptotically optimal decoding algorithm, *IEEE Trans. Inform. Theory*, **13** (1967) 260–269.
- [42] Wahba, G., *Spline Models for Observational Data* (SIAM, Philadelphia, 1990).
- [43] West, M. and J. Harrison, *Bayesian Forecasting and Dynamic Models* (Springer, New York, 1989).
- [44] White, K.J., et al., *SHAZAM User's Reference Manual* (McGraw-Hill, New York, 1995).
- [45] Yu, P.L. *Multiple-Criteria Decision Making: Concepts, Techniques, and Extensions* (Plenum Press, New York, 1985).
- [46] Zeleny, M., *Multiple Criteria Decision Making* (McGraw-Hill, New York, 1982).
- [47] Zionts, S., The state of multiple criteria decision making: past, present, and future, in: A. Goicoechea, L. Duckstein, and S. Zionts, *Multiple Criteria Decision Making, Proc. of the 9 Internat. Conf.: Theory and Applications in Business, Industry, and Government*, (Springer, New York, 1992) 33–43.