HYPOTHESIS, ESTIMATION, AND VALIDATION OF DYNAMIC SOCIAL MODELS -- ENERGY DEMAND MODELING

by

David Walter Peterson

S.B., Massachusetts Institute of Technology
(1969)

S.M., Massachusetts Institute of Technology
(1972)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June, 1975

Signature of Author.....
Department of Electrical Engineering, May 16, 1969

Certified by.....
Thesis Supervisor

Accepted by.....
Chairman, Departmental Committee on Graduate Students

HYPOTHESIS, ESTIMATION, AND VALIDATION OF DYNAMIC SOCIAL MODELS -- ENERGY DEMAND MODELING

by

David Walter Peterson

ABSTRACT

The first part of this study establishes a unified set of techniques for the estimation and validation of nonlinear, dynamic, time-varying models, with special reference to models and data typical of the social sciences. The techniques are called Full-Information Maximum Likelihood via Optimal Filtering (FIMLOF). The basics of FIMLOF are adapted from advanced engineering techniques of system identification and verification; additions and variations are created to deal with special problems of cross sectional data, unknown initial conditions, missing data, and bad data points. FIMLOF is demonstrated to apply to a wide range of model and data structures outside the scope of traditional techniques.

The feasibility of FIMLOF is examined in the context of simulation experiments, as a preliminary exercise. The methods and software are seen to yield expected theoretical results on a simple first order system. FIMLOF is also seen to work correctly on a high order, nonlinear system where other methods fail.

The FIMLOF techniques have been implemented in a user-oriented computer program, called the General Purpose System Identifier and Evaluator (GPSIE), in order to make the tools of FIMLOF more accessible, both to engineers and social scientists.

The second part of the study describes the use of FIMLOF (via GPSIE) in the development of a model of fuel demand in the residential-commercial sector of the U.S. In addition to the estimation of parameters, the methods are shown to yield sensitive validity tests, and insight for model improvement. A new technique for bad data identification and robust estimation is demonstrated.

THESIS SUPERVISOR: Fred C. Schweppe
TITLE: Professor of Electrical Engineering

ACKNOWLEDGEMENTS

Of the people who contributed to this study and its background, the following deserve special mention.

Professor Fred C. Schweppe provided the initial inspiration for the subject of the thesis, and was both a guiding light and a sounding board throughout. I salute his skill and integrity as a scholar, advisor, and counselor; I thank him for his technical help and encouragement. Many of the ideas used in this study are to some major degree his, especially in the areas of likelihood computation and bad data detection.

Dr. Martin L. Baughman supported the effort throughtout, and provided the motivating application problem, which in turned shaped much of the substance of GPSIE. His advice was astute and practical.

Professor Edwin Kuh provided a remarkably broad knowledge of current practice in econometrics, and demonstrated a gift for insight and criticism, always constructive. He also provided contact with other helpful people, including G.S. Madalla, Richard Becker, and David Kendrick.

Thanks are also due Bradley Schraeder and Alan Graham for uncovering errors in the GPSIE software, and suggesting improvements.

CONTENTS

Chapter 1

OVERVIEW and MOTIVATION

The process of modeling may be thought of as a three-stage iteration:

1. Hypothesize a mathematical structure
2. Estimate unknown parameters
3. Test the model for consistency with data

If the model fails step (3), the modeler returns to step (1) to hypothesize a new structure.

This thesis develops and implements some advanced methods for performing steps (2) and (3) of the above iteration. The methods are based on the method of full information maximum likelihood, via optimal filtering (FIMLOF). FIMLOF, and related techniques, allow the steps of estimation and validation to be performed for a much wider range of model forms than has been possible using traditional tools. Therefore, the methods also contribute to step (1) of the iteration, by removing traditional constraints on model specification. Figures 1-1 and 1-2 give an overview of the capabilities of FIMLOF and the types of models to which the FIMLOF methods apply.

Although the the new techniques can be used to improve on or check the results of other statistical methods, the

PARAMETER ESTIMATION:

- o Estimate all parameters in a model, including characteristics of driving and measurement noise.
- o Estimate some parameters, taking others as known a priori.

CONFIDENCE TESTS (VALIDATION):

- o Test for consistency between model and data.
- o Choose among alternate models (hypothesis testing)

FEASIBILITY:

- o Determine if the estimation of parameters in a model is possible, using given data
- o Determine the kind and amount of data needed to estimate unknown parameters to a given accuracy (via experiments with simulation data).

STATE ESTIMATION, SIMULATION, AND FORECASTS:

- o Estimate unknown inputs and the state trajectory
- o Simulate the model (deterministic or noise-driven)
- o Compute confidence bounds on forecasts

BAD DATA DETECTION AND ROBUST ESTIMATION:

- o Detect presence of bad data
- o Locate specific bad data points
- o Eliminate the bad data from the sample
- o Automatic compensation for bad data: robust estimation

Figure 1-1

Capabilities of FIMLOF and GPSIE

FIMLOF and GPSIE Operate Under
Conditions Of:

1. Nonlinearities in model dynamics
2. Nonlinear measurement functions.
3. Measurement error (errors in variables)
4. Mixed sampling intervals (can, for example, estimate a weekly model, using monthly and yearly data.
5. Missing data (without sacrificing other data at the same sample time).
6. Models with unmeasured endogenous variables.
7. Cross-sectional, time series mixed data.
8. Unknown characteristics of equation errors and measurement noise.

Figure 1-2

Applicability of FIMLOF and GPSIE

most promising contribution of FIMLOF is its ability to deal statistically with model and data structures which lie outside the range of applicability of the other methods.

The methods were developed as part of a larger effort [Baughman,1974] to model the demand for fuels in the residential commercial sector of the U.S. economy. There were three motivations which led to the development:

1. Models were being hypothesized, based on economic and engineering theory, which were not linear in the parameters, and which could therefore not be estimated with ordinary least squares (OLS), or other traditional single-equation methods.
2. Much of the available data on fuel consumption and fuel prices seemed to contain significant amounts of measurement error (errors in the variables); some evidence [Senge,1974] indicated that OLS and GLS could be sensitive to measurement error, yielding invalid and misleading results.
3. In the long run, it was felt desirable to be able to add to the models expectations and other variables, for which no direct measurements were available. The new methods showed promise, from their application in engineering situations, of being able to deal with models containing unmeasured endogenous variables.

This thesis reviews the method of full information

maximum likelihood via optimal filtering [Schweppe, 1965], develops and extends the method for application to dynamic social models. The operation of the FIMLOF methods is demonstrated both in controlled simulation experiments and on real data in the modeling of energy demand.

The success of the resulting techniques is promising enough that the software used in this thesis has been generalized and made publicly available, in the form of a program called the General Purpose System Identifier and Evaluator (GPSIE). GPSIE is described in [Peterson, 1974].

Chapter 2 of this thesis describes the mathematics of the FIMLOF techniques, their relation to other methods, and their intuitive basis.

Chapter 3 reviews some key examples of the experimental results used to test the validity of the new methods. Results using GPSIE are compared with Senge's experiments [Senge, 1974].

Chapter 4 discusses some possible implications of the FIMLOF techniques on the hypothesis of model structures.

Chapters 5 and 6 describe the application of FIMLOF to models of energy demand in the residential commercial sector. Chapter 5 describes some results on a preliminary structure which was rejected when it failed to pass various consistency tests. Chapter 6 describes a somewhat more successful result with a revised model.

Chapter 7 gives a brief summary of some of the key

results of the work.

Chapter 2

FIMLOF FOR NONLINEAR DYNAMIC SOCIAL MODELS

This chapter describes full-information maximum likelihood via optimal filtering (FIMLOF) for use in estimation and consistency testing of dynamic social models. Section 2.1 defines a standard form and nomenclature for the class of nonlinear stochastic dynamic models to which FIMLOF may be applied. Section 2.2 establishes the generality of the standard form -- it is shown that a great variety of model forms can be reduced to the standard form. Section 2.3 develops the mathematics of FIMLOF as applied to the standard form. Section 2.4 gives an intuitive interpretation of the mathematics of FIMLOF, and describes its relation to other methods. Section 2.5 lists several useful properties of the likelihood surface and filter time series at the maximum likelihood point. These properites are used as the basis of several tests of consistency between models and data. Section 2.6 develops new methods for the detection and identification of bad data, suggests robust estimation algorithms based on the methods. Section 2.7 concludes the chapter with a description of the implementation of FIMLOF in a computer program called the General Purpose System Identifier and Evaluator (GPSIE).

2.1 Standard Model Form.

GPSIE deals with models which can be reduced to (or suitably approximated by) the following form:

$$\underline{x}(n) = \underline{f}(\underline{x}(n-1), \underline{u}(n), \underline{w}(n), n)$$

$$\underline{z}(n) = \underline{h}(\underline{x}(n), \underline{v}(n), n)$$

$$n = 1 \ . \ . \ . \ N,$$

where

$\underline{x}(n)$ is the vector (dimension Kx) of state variables of the system. State variables may be informally defined as the "memory" of the system. The future behavior of the model depends only on the present values of the state variables and the future values of exogenous inputs. The examples and guidelines of Section 2.2 should clarify the idea for most readers not already familiar with state variables; for further details, the reader may consult [DeRusso,1965] or [Minsky,1965]. The "initial conditions" of the state variables, $\underline{x}(0)$, are assumed to be uncertain, with expected value $\underline{x}_0$ and covariance $\underline{\Psi}$.

$\underline{z}(n)$ is the vector (dimension Kz) of measurement data at time index n. "n" is simply the index of time-points for which there is data -- it is not assumed that the data is evenly distributed in time.

$\underline{f}$ is a vector-valued nonlinear function, called the

state function. It may, of course, be linear.

$\underline{u}(n)$ is a vector function of n (time), called the exogenous input vector, or the control input. Its dimension is Ke; it may be absent, in which case Ke=0.

$\underline{w}(n)$ is an unknown vector input, assumed to be a white, Gaussian (normal) process. The mean of $\underline{w}(n)$ is assumed to be zero; its covariance matrix is called $\underline{Q}(n)$. Note that the dimension of $\underline{w}$ may differ from that of $\underline{x}$.

$\underline{h}$ is a vector-valued (dimension Kz), nonlinear function, called the measurement function.

$\underline{v}(n)$ is an unknown vector-valued measurement error, assumed to be a white, zero-mean, Gaussian process. Its covariance matrix is called $\underline{R}(n)$. Note that the dimension of $\underline{v}$ may differ from that of $\underline{z}$.

The standard model form is summarized in Figure 2-1.

STATE EQUATIONS:

$$\underline{x}(n) = \underline{f}\left[\underline{x}(n-1),\ \underline{u}(n),\ \underline{w}(n),\ n\right]$$

MEASUREMENT EQUATIONS:

$$\underline{z}(n) = \underline{h}\left[\underline{x}(n), \underline{v}(n),\ n\right]$$

INDEX OF DATA SAMPLES:

$$n = 1, \ldots N$$

INITIAL CONDITIONS:

$$\underline{x}(0) = N\left[\underline{x}_o, \underline{\Upsilon}\right] \qquad *$$

EQUATION ERRORS (DRIVING NOISE):

$$\underline{w}(n) = N\left[\underline{0},\ \underline{Q}(n)\right]$$

MEASUREMENT ERRORS

$$\underline{v}(n) = N\left[\underline{0},\ \underline{R}(n)\right]$$

LINEARIZATION ABOUT ESTIMATED STATE:

$$\underline{\tilde{F}}(n) = \left.\frac{\partial \underline{f}}{\partial \underline{x}}\right|_{\hat{\underline{x}}(n-1|n-1)} \qquad \underline{\tilde{H}}(n) = \left.\frac{\partial \underline{h}}{\partial \underline{x}}\right|_{\hat{\underline{x}}(n|n-1)}$$

$$\underline{\tilde{Q}}(n) = \left(\frac{\partial \underline{f}}{\partial \underline{w}}\right) \underline{Q}(n) \left(\frac{\partial \underline{f}}{\partial \underline{w}}\right)'$$

$$\underline{R}(n) = \left(\frac{\partial \underline{h}}{\partial \underline{v}}\right) \underline{R}(n) \left(\frac{\partial \underline{h}}{\partial \underline{v}}\right)'$$

} partials evaluated at $\underline{w}(n) = \underline{0}$, and $\underline{v}(n) = \underline{0}$.

(Partials for use in optimal filter)

* $N\left[\underline{m}, \underline{C}\right]$ denotes a normal, white process with vector mean $\underline{m}$ and covariance matrix $\underline{C}$.

Figure 2-1

Standard Model Form

2.2 Generality of the Standard Form.

The following subsections list common model types, and discuss how each can be expressed in the standard form of Section 2.1. The treatment here is brief; for more details, the reader is referred to Chapter 3 of [Schweppe,1973].

2.2.1 STATIC SYSTEM.

By "static system" we mean single or multiple measurements $\underline{z}$ of a constant $\underline{x}$. The corresponding standard form is achieved by setting $\underline{w}(n)=0$ for all n and making $\underline{f}$ the identity function. Then we have

$$\underline{x}(n) = \underline{x}(n-1) = \underline{x} \ , \quad \text{and}$$

$$\underline{z}(n) = \underline{h}[\underline{x}(n), \underline{v}(n)] = \underline{h}[\underline{x}, \underline{v}(n)] \ .$$

2.2.2 MISSING OBSERVATIONS.

Suppose that at sample (or samples) n, some of the components of the measurement vector $\underline{z}(n)$ are missing. The situation of missing observations arises often and in many guises:

1. Changed measuring methods. This case, which often arises in social data, means that some of the measured variables are available only after a given point in time.
2. Different freqencies of sampling. Some variables, for example, might be measured monthly, others

yearly.

3. Multiplexed measurements. In physical systems, we may have a single measuring device which looks at different variables or groups of variables in rotation. For example, we might have data on one variable for odd n, and another variable for even n.

4. Simple missing observations. Due to loss of records, typographical errors, or malfunctioning sensors, some data points (possibly scattered) may be simply unavailable, or may be considered too unreliable to use.

The missing data phenomena are all, in essence, a special case of a time-varying measurement function $\underline{h}$. If, for example, no data is available at n=3, then $\underline{h}(3)=\underline{0}$. Similarly, if the first component of $\underline{z}(5)$ is missing, then $h(5)=0$. Methods for dealing with missing observations are discussed more fully in later sections.

2.2.3 NO MEASUREMENTS AT ALL OF SOME VARIABLES.

This case is already inherent in the standard form. Note that $\underline{z}(n)$ need not be of the same dimension as $\underline{x}(n)$. The general measurement function $\underline{h}$ also allows a single measurement (component of $\underline{z}$) to be a function of several components of $\underline{x}$.

2.2.4 PERFECT MEASUREMENTS.

If some of the data is presumed to be error-free, simply set the appropriate elements of $\underline{R}$ to zero, indicating zero measurement error for the corresponding component of $\underline{z}$.

2.2.5 MODEL ERRORS.

The input uncertainty $\underline{w}(n)$ may represent not only exogenous random inputs to the system (weather, etc.), but may also be used to represent the uncertainty associated with the model formulation. Components of $\underline{w}$ added to an equation may be used to represent the (lack of) precision of that equation.

2.2.7 KNOWN INITIAL CONDITIONS.

In some cases, $\underline{x}(0)$ can be assumed to be known perfectly. In this case, one sets $\underline{\Psi} = \underline{0}$. For single components of $\underline{x}(0)$ known perfectly, the corresponding row and column of $\underline{\Psi}$ should be set to zero.

2.2.8 COMPLETELY UNKNOWN INITIAL CONDITIONS.

To approximate the case of unknown initial conditions, it is usually best to set $\underline{\Psi}$ very large. For most cases, this method suffices. If, however, an unusually precise formulation is required, $\underline{x}(0)$ may be modeled as if it were known perfectly, but with the unknown components to be estimated as unknown parameters. In this case, the corresponding components of $\underline{\Psi}$ should be set to zero.

2.2.9 NONZERO-MEAN UNCERTAINTY.

If it is desired to model $\underline{w}(n)$ and $\underline{v}(n)$ as non-zero mean, simply incorporate the mean values as part of the exogenous input $\underline{u}(n)$. For example, take the scalar model

$$x(n) = a*x(n-1) + u(n) + \tilde{w}(n) ,$$

where

$$E\{\tilde{w}(n)\} = k$$

By suitable definition of u(n) and w(n), we get the standard form

$$x(n) = a*x(n-1) + u(n) + w(n)$$

where

$$u(n) = \tilde{u}(n) + k$$

$$w(n) = \tilde{w}(n) - k$$

2.2.10 NON-WHITE UNCERTAINTY.

If, as is often the case, one wishes to model $\underline{w}(n)$ or $\underline{v}(n)$ as time-correlated processes, this can be done by modeling, for example, $\underline{\tilde{w}}(n)$ as the output of a dynamic subsystem driven by a white process $\underline{w}(n)$. A common choice for the subsystem is the first order linear model

$$\tilde{w}(n) = p*\tilde{w}(n-1) + w(n) ,$$

where p may be set a priori or may be included as an unknown parameter to be estimated. Such subsystems must then be incorporated in the state function part of the standard-form model, by including $\tilde{w}(n)$ as a component of the state vector $\underline{x}(n)$.

2.2.11 NON-GAUSSIAN UNCERTAINTY.

Less common than the non-white case, but still possible, is the case in which $\underline{w}(n)$ or $\underline{v}(n)$ is modeled as a non-Gaussian process. In general, the desired effect can be achieved or approximated by taking, for example, $\underline{\tilde{w}}(n)$ to be the output of a nonlinear subsystem driven by white, Gaussian noise.

2.2.12 COMPLETELY UNKNOWN INPUT.

Sometimes it is desirable to model the input disturbance $\underline{w}(n)$ as a completely unknown (Fisher) process. An excellent approximation to this condition can often be made by setting the appropriate components of $\underline{Q}(n)$ to be very large. The rigorous way would be to treat $\underline{w}(n)$ as known and estimate all N*Kw components as unknown parameters. This method, however, is usually too expensive to be practical.

2.2.13 UNCERTAIN MEASUREMENT OF EXOGENOUS INPUTS.

The standard form assumes that the model-input vector $\underline{u}(n)$ is known perfectly. In some cases, however, $\underline{u}(n)$ is instead measured with error. Consider a model in which the measurements of the exogenous input is

$$\underline{\tilde{u}}(n) = \underline{u}(n) + \underline{e}(n),$$

where $\underline{u}(n)$ is the true, but unknown, input, and $\underline{e}(n)$ is the corresponding measurement error. Assume, for example, a model of the form

$$\underline{x}(n) = \underline{A}(n)\underline{x}(n-1) + \underline{u}(n) + \underline{w}(n)$$

In this case, we get the standard form

$$\underline{x}(n) = \underline{A}(n)\underline{x}(n-1) + \tilde{\underline{u}}(n) + \tilde{\underline{w}}(n),$$

where

$$\tilde{\underline{u}}(n) = \underline{u}(n) + \underline{e}(n)$$

$$\tilde{\underline{w}}(n) = \underline{w}(n) - \underline{e}(n) .$$

Similar transformations can be made whenever the measurement of $\underline{u}(n)$ is invertable; that is, if the true input $\underline{u}(n)$ can be expressed as a function of the measured input $\tilde{\underline{u}}(n)$ and the measurement error $\underline{e}(n)$.

2.2.14 LAGGED VARIABLES AND AUTOREGRESSIVE MODELS.

Often, models contain lagged variables. For example, consider the scalar model

$$x(n) = a*x(n-1) + b*x(n-3) + w(n)$$

where x(n-3) is x(n) lagged three time steps. Such models are converted to standard form by augmenting the state vector and state function to include a simple model of the lagging process. In the example, the equivalent standard form is the three-state model

$$x1(n) = a*x1(n-1) + b*x2(n-1) + w(n)$$

$$x2(n) = x3(n-1)$$

$$x3(n) = x1(n-1)$$

Substitution quickly demonstrates that x2(n-1) is equivalent to x1(n-3).

2.2.15 CONTINUOUS DYNAMICS.

Continuous-time dynamics (differential equations, rather than difference equations) must always be approximated as discrete-time systems if they are to be simulated on a digital computer. A wide variety of integration schemes exist to simulate the continuous dynamics between data points.

2.2.16 REGIONAL (CROSS-SECTIONAL) DATA.

Often (especially in social modeling), data from several separate but similar systems are used to estimate parameters in a single model. For example, a model of water use in part of the U.S. might be applied to ten years of data collected separately in each of 15 counties. The model might then be a function of both time and location. Such a model can be converted to the standard form by concatenating the data and being careful to make all the appropriate model changes at the transition from one region to the next. To make the job easier for the user, GPSIE includes options for the handling of cross-sectional data in convenient form.

2.2.17 MOVING AVERAGE MODELS (LAGGED INPUTS).

Consider a typical moving-average model of the type

$$x(n) = a*w(n) + b*w(n-1) + c*w(n-2) .$$

Such models can usually be reduced to the standard form by state augmentation:

$$x1(n) = b*x2(n-1) + c*x3(n-1) + a*w(n)$$

$$x2(n) = w(n)$$

$$x3(n) = x2(n-1)$$

2.2.18 IMPLICIT STATE FUNCTION.

In economics, one often encounters equations of the form:

$$\underline{f}(\underline{x}(n), \underline{x}(n-1) \ldots) = 0 ,$$

in which the state vector $\underline{x}(n)$ cannot be expressed as an explicit function of its previous value x(n-1) and the exogenous inputs $\underline{u}(n)$ and $\underline{w}(n)$. While the FIMLOF method requires that $\underline{x}(n)$ and $\frac{\partial \underline{x}(n)}{\partial \underline{x}(n-1)}$ be computed from $\underline{x}(n-1)$, the method does not require that the computation be analytically explicit. Thus, in the extreme case where explicit expressions for neither $\underline{x}(n)$ nor $\frac{\partial \underline{x}(n)}{\partial \underline{x}(n-1)}$ can be derived, both can be computed numerically. For example, $\underline{x}(n)$ could be computed by Newton's method [see any book on numerical analysis], and the partial could be computed by taking finite perturbations in $\underline{x}(n)$ and using Newton's method to compute the corresponding $\underline{x}(n)$ vectors, from which the partial may be approximated via finite differences:

$$\left(\frac{\partial \underline{x}(n)}{\partial \underline{x}(n-1)}\right)_{jk} \approx \frac{x_j(n, x_k(n-1)+\Delta) - x_j(n)}{\Delta}$$

The FIMLOF method, as a side benefit, often removes the need to use implicit formulations in economic models (see Chapter 4).

2.3 Mathematics of FIMLOF.

This section summarizes the computation of the log likelihood function for the standard form model defined in Section 2.1. For a more detailed treatment, see [Schweppe,1973,pp 286-288].

The likelihood function is computed one point at a time. That is, for each different model (or, equivalently, each different value of the vector of unknown parameters) the data are processed once to yield the log likelihood of that model. To find the maximum likelihood parameters, then, the parameters are varied according to a hill-climbing algorithm (nonlinear optimization).

The computation of the log likelihood for a given model is as follows: define

$$\underline{z}_n \triangleq \begin{bmatrix} \underline{z}(1) \\ \vdots \\ \underline{z}(n) \end{bmatrix}$$

For any set of unknown parameters (or, equivalently, for an hypothesized model j), $\underline{z}_{n_j}$ is a random vector with probability density function $p_j(\underline{z}_n)$. The log likelihood function $\xi(n)$ is defined as

$$\xi(n) = \ln\{p_j(\underline{z}_n)\}$$

By Bayes' rule,

$$p_j(\underline{z}_n) = p_j(\underline{z}_{n-1}) * p_j(\underline{z}_n | \underline{z}_{n-1})$$

or, taking the log of both sides,

$$\xi_j(n) = \xi_j(n-1) + \ln\{p_j(\underline{z}_n|\underline{z}_{n-1})\}, \quad \xi_j(0) = 0 \qquad (*)$$

It remains to compute $\ln\{p_j(\underline{z}(n)|\underline{z}_{n-1})\}$; the computation is performed by the optimal filter. Define

$$\hat{\underline{z}}(n|n-1) = E\{\underline{z}(n)|\underline{z}_{n-1}\}$$

$$\underline{\delta}_z(n) = \underline{z}(n) - \hat{\underline{z}}(n|n-1)$$

$$\underline{\underline{\Sigma}}_z(n|n-1) = E\{\underline{\delta}_z(n)\underline{\delta}_z'(n)\}$$

Under the assumption that the probability densities are gaussian, we have $p_j(\underline{z}(n)|\underline{z}_{n-1}) =$

$$\left[(2\pi)^{Kz}\det\{\underline{\underline{\Sigma}}_z(n|n-1)\}\right]^{-1/2} \exp\left\{-\frac{1}{2}\underline{\delta}_z'\underline{\underline{\Sigma}}_z^{-1}(n|n-1)\underline{\delta}_z(n)\right\}$$

where KZ is the dimension of z(n). Substituting this result into (*), we get $2\xi_j(n) = 2\xi_j(n-1) - K_z\ln(2\pi)$

$$-\ln\left[\det\{\underline{\underline{\Sigma}}_z(n|n-1)\}\right] - \underline{\delta}_z'(n)\underline{\underline{\Sigma}}_z^{-1}(n|n-1)\underline{\delta}_z(n)$$

The remaining undefined terms $\hat{\underline{z}}(n|n-1)$ and $\underline{\underline{\Sigma}}_z(n|n-1)$ are computed by the optimal filter, the equations of which are given in Figure 2-3.

For more details, see [Schweppe,1973] and [Kalman,1960].

2.4 Intuitive Interpretation and Relation To Other Methods.

This section gives a simple example of the method outlined mathematically in the above section, to explain the intuitive justification for the FIMLOF method, and to give some insight into the relation between FIMLOF and OLS, WLS, and other standard econometric methods.

predicted state: $\underline{\hat{x}}(n|n-1) = \underline{f}[\underline{\hat{x}}(n-1|n-1), \underline{u}(n)]$

pred. measurement: $\underline{\hat{z}}(n|n-1) = \underline{h}[\underline{\hat{x}}(n|n-1)]$

residuals: $\underline{\delta}_z(n|n-1) = \underline{z}(n) - \underline{\hat{z}}(n|n-1)$

predicted state covariance: $\underline{\underline{\Sigma}}_x(n|n-1) = \underline{\underline{\tilde{F}}}(n)\,\underline{\underline{\Sigma}}_x(n-1|n-1)\,\underline{\underline{\tilde{F}}}'(n) + \underline{\underline{\tilde{Q}}}(n)$

predicted meas. covariance: $\underline{\underline{\Sigma}}_z(n|n-1) = \underline{\underline{\tilde{H}}}(n)\,\underline{\underline{\Sigma}}_v(n|n-1)\,\underline{\underline{\tilde{H}}}'(n) + \underline{\underline{\tilde{R}}}(n)$

normalized predicted measurement residuals: $\underline{\tilde{\delta}}_z(n|n-1) = \sqrt{\underline{\underline{\Sigma}}_z}^{-1}\,\underline{\delta}_z(n|n-1)$

updated state covariance: $\underline{\underline{\Sigma}}_x(n|n) = \left[\underline{\underline{\Sigma}}_x^{-1}(n|n-1) + \underline{\underline{\tilde{H}}}'(n)\,\underline{\underline{\tilde{R}}}^{-1}(n)\,\underline{\underline{\tilde{H}}}(n)\right]^{-1}$

filter gain: $\underline{K}(n) = \underline{\underline{\Sigma}}_x(n|n-1)\,\underline{\underline{\tilde{H}}}'(n)\,\underline{\underline{\Sigma}}_z^{-1}(n|n-1)$

updated state estimate: $\underline{\hat{x}}(n|n) = \underline{\hat{x}}(n|n-1) + \underline{K}(n)\,\underline{\delta}_z(n|n-1)$

log likelihood: $\xi(n) = \xi(n-1) - \frac{1}{2}\ln\left[\det\{\underline{\underline{\Sigma}}_z(n|n-1)\}\right]$

$$-\frac{1}{2}\,\underline{\delta}_z'(n|n-1)\,\underline{\underline{\Sigma}}_z^{-1}(n|n-1)\,\underline{\delta}_z(n|n-1)$$

initial conditions: $\underline{\hat{x}}(0|0) = \underline{x}_0$, $\underline{\underline{\Sigma}}_x(0|0) = \underline{\underline{\Psi}}$,

$\xi(0) = 0$.

Figure 2-3

Equations for the Optimal Filter

Consider the data graphed in Figure 2-4, and suppose we hypothesize the data to have been generated by the system

$$x(n) = rx(n-1)$$

$$z(n) = x(n)$$

where r is an unknown parameter we wish to estimate from the data. Since the hypothesized system is completely deterministic, and the data of FIgure 2-4 is "clean," we may do a good job of estimating r by simply taking the ratio between a few successive values of z(n). However, the purpose of this example is to illustrate, in a simple environment, a more roundabout way of estimating r which will work not only in the simple case, but in much more complicated situations. The essence of the more complicated method, which we shall develop into FIMLOF, is to guess a value for r and simulate the equations, and measure the error between the simulated data $\hat{z}(n)$ and the actual data z(n) of figure 2-4. The estimate of r is then chosen as the guessed value which minimizes the error between the actual and simulated data.

There are several ways of doing the simulation; we here discuss three ways. One is obvious but quickly leads to trouble in less ideal situations, the second is equivalent to ordinary least squares, and the third will be FIMLOF.

2.4.1 NAIVE SIMULATION.

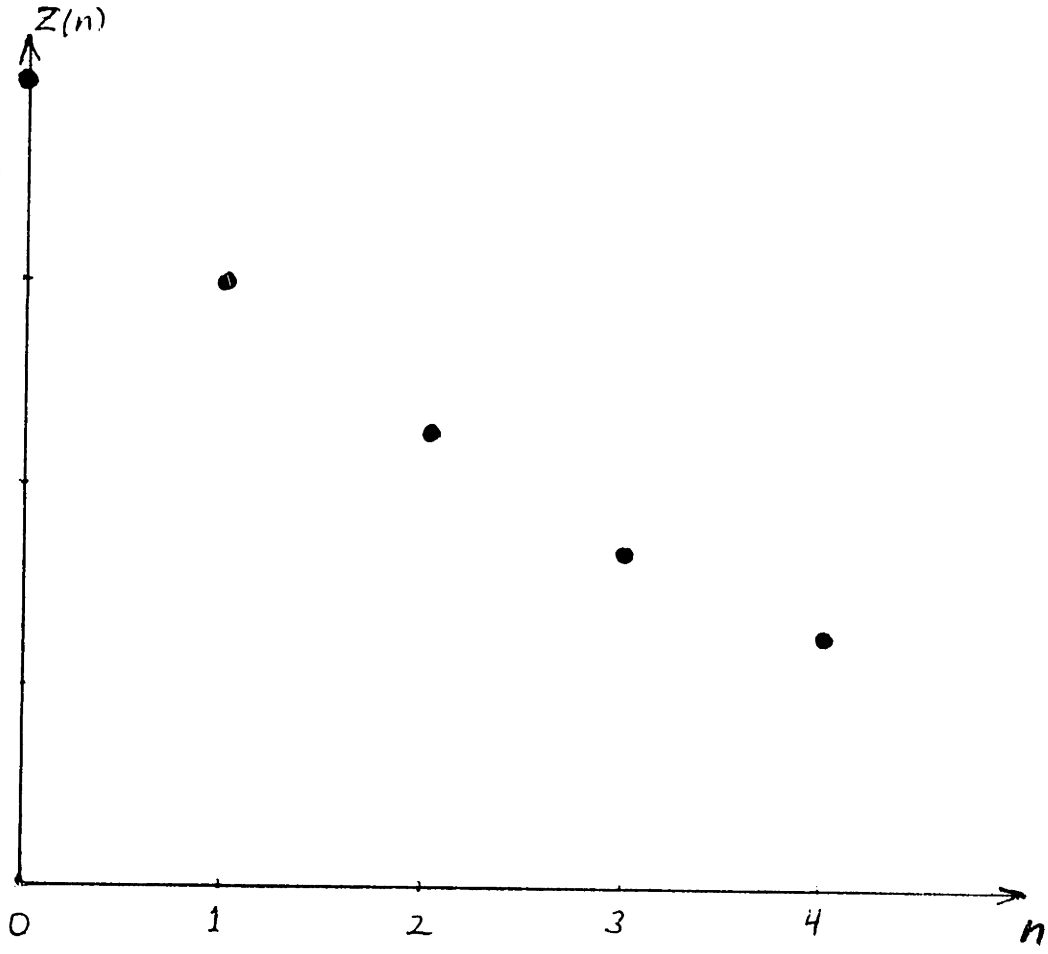

Figure 2-4

Data For Example Estimation

Let us assume that we initialize the system at the first data point z(1) and perform the iteration

$$\hat{x}(n) = \hat{r}\hat{x}(n-1)$$

$$\hat{z}(n) = \hat{x}(n)$$

for n=1...N. We may now measure the error by taking the sum of the squared errors at each data point. The NS sum of squared errors, also called the loss function, is denoted as J:

$$J = \sum_{n=1}^{N} [z(n)-\hat{z}(n)]$$

In this nonpathological case, the error will be zero if we guess r correctly. We may guess close to the right r perhaps more efficiently and methodically by using a good hill-climbing algorithm, but the essential idea is to guess until we are satisfied that no smaller error can be found.

However, in less sanguine cases, such as the system

$$x(n) = rx(n-1) + w(n), \quad w(n)=N[0,Q]$$

the naive simulation method may give minimum error for a completely wrong guess of r, since the real system may "drift" away from the deterministic trajectory, under the influence of the driving noise w(n). We shall carry the argument further after introducing a method that works better.

2.4.2 ORDINARY LEAST SQUARES (OLS).

When driving noise w(n) is introduced, we can in general obtain better estimates of r by re-initializing the

system at each data point. That is, we use the iteration

$$\hat{x}(n) = \hat{r}z(n-1)$$

$$\hat{z}(n) = \hat{x}(n)$$

and compute the OLS summed error (loss function) J' as

$$J' = \sum_{n=1}^{N} [z(n)-\hat{z}(n)]$$

In the case with driving noise $w(n)$ and covariance Q introduced above, it is easy to see that as n increases, the covariance of $x(n)$ (and $z(n)$) approach a steady state value of

$$E\{x^2(n)\} = E\{z^2(n)\} = Q/(1-r^2)$$

Assuming that $r<1$, the naive simulation (NS) method will yield, for large n, negligible $\hat{x}(n)$, and the covariance of $x(n)$ will be close to its limiting value of $Q/(1-r^2)$. Call the n for which the system is approximately in steady state $\tilde{n}$, and let us compute the contribution to the error terms for large $n>>\tilde{n}$ under the two methods of estimation. We define $E\{\Delta J(n)\}$ as the expectation of the increment in the loss function J due to the n_th sample of data $z(n)$.

For the OLS case, we have

$$E\{\Delta J'(n)\}ols = E\{(r-\hat{r})^2 z^2(n-1) + w^2(n) + 2w(n)(r-\hat{r})z(n-1)\}$$

$$= (r-\hat{r})^2 E\{x^2(n-1)\} + Q$$

$$= (r-\hat{r})^2 Q/(1-r^2) + Q$$

It is clear from the last expression above that the expectation of the OLS loss function J' is minimized by $\hat{r}=r$, and that any $\hat{r}\neq r$ adds to the loss function by a stochastic amount with constant positive expectation. Thus the

estimate $\hat{r}$ which minimizes J' is consistent.

For the naive simulation (NS) case, a similar analysis yields

$$E\{\Delta J(n)\}ns = E\{[z(n) - \hat{r}^n x(0)]^2\}$$

$$= E\{x^2(n)\} - 2E\{\hat{r}^n x(0)x(n)\} + E\{\hat{r}^{2n} x^2(0)\}$$

$$= Q/(1-r^2) - 2x^2(0)\hat{r}^{2n} + \hat{r}^{2n} x^2(0)$$

$$= Q/(1-r^2) - \hat{r}^{2n} x^2(0)$$

But for large n, this reduces to

$$E\{\Delta J(n)\}ns = Q/(1-r^2), \qquad n>>1$$

which is independent of $\hat{r}$! Thus, for noise driven systems, the naive simulation method in effect ignores most of the data, while the OLS method yields a consistent, unbiased estimate.

However, the OLS method also breaks down when measurement noise v(n) ("error in the variables") is present in the system [Wonnacott,1970]:

$$x(n) = rx(n-1) + w(n)$$

$$z(n) = x(n) + v(n)$$

The next section describes how FIMLOF incorporates the desirable features of both NS and OLS to compensate simultaneously for w(n) and v(n).

2.4.3 FIMLOF.

The intuitive interpretation of FIMLOF presented here is best understood by first putting the NS and OLS methods in a common framework. Each method consists of the

following iteration:

1. Compute a predicted measurement $\hat{z}(n|n-1)$ by simulating the model one time step from time n-1 to n.
2. Square and accumulate the error term (residual): $z(n)-\hat{z}(n|n-1)$.
3. Reinitialize the model to prepare for the next simulation step (set x(n) in the model to some $\hat{x}(n|n)$), and go to step 1.

The only difference between NS and OLS is in Step 3 of the above iteration. NS "reinitializes" by leaving the model state at the value obtained in the previous simulation, ignoring the data z(n). That is, NS omits step 3 of the iteration. In contrast, OLS ignores the previous state of the model completely, and reinitializes the model at z(n). Thus, in OLS, Step 3 of the iteration is completely data-based, while for NS it is model-based.

The advantage of OLS over the naive simulation method (NS) may be interpreted as follows: when a large w(n) abruptly moves the system far from its former state, the naive simulation method "stays put" and thereby accumulates large fallacious errors; whereas the OLS method, via its reinitialization at the data point, "follows" the system as it drifts under the influence of w(n).

But with the addition of measurement error v(n) in the above system, OLS may run into a problem similar to that

encountered by NS in Section 2.4.2. That is, a large v(n) may trick OLS into reinitializing at a z(n) far away from the true state of the system x(n), leading to large errors, even with perfect parameter guesses.

Thus, we are in a dilemma. The naive simulation method does not go far enough in taking into account the data by repositioning the model; but OLS may go too far, reinitializing at z(n) which may be far from the true state of the system. The essence of FIMLOF is to strike an optimum balance between the two extremes, by basing the reinitialization on explicit assumptions about the variances of the two error sources. If the equation error w(n) has large variance Q, and the measurement error v(n) has small variance R, then FIMLOF behaves like OLS, ignoring the prediction of the model and repositioning at the new data z(n). In the opposite extreme, with large R and small Q, FIMLOF behaves much like the naive simulation method, relying on the model predictions to maintain a steady course through the error-prone data. The reader may verify these limiting cases by examining the equations fo Figure 2-3 for large Q and for large R. In the general case, where w(n) and v(n) have arbitrary variances, the equations of the optimal filter reinitialize the system at the best Bayesian estimate of x(n), given all previous information z(1)...z(n).

AT this point, the reader may object that the prior

knowledge of the variances of v(n) and w(n) used to make the Bayesian estimate may be unavailable a priori. In this case, one simply adds the variances of v(n) and w(n) to the list of unknown parameters to be estimated. An important feature of FIMLOF is the ability to extract from the data information (when it is present) about any aspect of the standard model form, including structural specification (through parameterization of structure) and the characteristics of the error processes.

2.5 Consistency Tests.

The FIMLOF method yields some powerful statistics for evaluating parameter estimates and overall consistency of the estimated (or hypothesized) model with the data. The next two subsections describe these tests. Section 2.5.1 discusses tests that derive directly from the likelihood surface; Section 2.5.2 describes some especially valuable tests that are based on the optimal filtering mathematics.

2.5.1 USE OF THE LIKELIHOOD SURFACE.

Two properties of the log likelihood surface are useful in locating and interpreting parameter estimates.

1. At the global maximum of the log likelihood surface, which determines the FIML estimates of unknown parameters, the hessian matrix of second partials, as computed by finite differences, gives an approximation to

the Fisher information matrix. The square roots of the main diagonal terms of the inverted hessian, then, measure the standard deviation confidence bounds of the parameter estimates. One may be tempted to use these confidence bounds to compute T statistics for the parameter estimates. But in the nonlinear framework, such statistics could be misleading, since a zero parameter can be quite significant and important. For example, consider the following model with a single unknown parameter θ:

$$x(n) = \theta f(x(n-1)) + (1-\theta)g(x(n-1))$$

Here, an estimate of $\theta=0$ is very significant indeed, since the estimate indicates that the structure defined by the function g is the most likely. So, for example, we might estimate θ to be .01, with standard deviation .05. The T statistic would be 0.2 which might be interpreted as an insignificant estimate, where in the context of the model, the estimate and its standard deviation imply very significant conclusions. In conclusion, the FIMLOF method is capable of computing a traditional statistic, but in the general context of nonlinear systems, the statistic itself, under traditional informal interpretation, may be misleading.

2. A little known but highly useful fact is that at the maximum, the vertical change in the likelihood surface corresponding to a two standard deviation horizontal displacement from the maximum corresponds to a change in the

log likelihood of -2.

The derivation is based on the often valid approximation [Edwards, 1972] that the log likelihood surface at its global maximum (if it exists) is roughly quadratic. We approximate the surace, then, by a Taylor series, discarding all terms above the quadratic. Let $\xi(\underline{\alpha})$ be the log likelihood evaluated for the vector of parameter values $\underline{\alpha}$. Let $\underline{\alpha}^*$ be the values of the parameters which maximize ξ. Then by the Taylor series approximation,

$$\xi(\underline{\alpha}) = \xi(\underline{\alpha}^*) + \left.\frac{\partial \xi}{\partial \underline{\alpha}}\right|_{\underline{\alpha}^*} (\underline{\alpha} - \underline{\alpha}^*) + \frac{1}{2} (\underline{\alpha} - \underline{\alpha}^*)' \left.\frac{\partial^2 \xi}{\partial \underline{\alpha}^2}\right|_{\underline{\alpha}^*} (\underline{\alpha} - \underline{\alpha}^*)$$

But at the maximum of the likelihood function, the gradient is zero, and the hessian of the log likelihood serves as the negative of the information matrix, or error covariance matrix of the estimated parameters $\underline{\underline{\Sigma}}_{\alpha}$. Thus,

$$\xi(\underline{\alpha}) - \xi(\underline{\alpha}^*) = -\frac{1}{2}(\Delta\underline{\alpha})' \underline{\underline{\Sigma}}_{\alpha} (\Delta\underline{\alpha}) \quad ,$$

where $\underline{\underline{\Sigma}}_{\alpha}$ is the error covariance matrix of the parameter estimates $\hat{\underline{\alpha}}$, and $\Delta\underline{\alpha}$ is the displacement vector $\underline{\alpha} - \underline{\alpha}^*$. Now let $\underline{\alpha}$ be two standard deviations from $\underline{\alpha}^*$. Then by definition, $\underline{\alpha}$ lies on the "two-sigma" error ellipsoid defined by

$$\Delta\underline{\alpha}' \underline{\underline{\Sigma}}_{\alpha} \Delta\underline{\alpha} = 2^2 = 4, \text{ or}$$

$$\xi(\underline{\alpha}) - \xi(\underline{\alpha}^*) = -2.$$

The "two-sigma" log likelihood change of -2 is especially useful, because it is an absolute difference, independent of units of measure in the data, scaling factors, and computer word lengths. The "two sigma two"

result is of use both as a check on the standard deviation estimates on the parameters, but also as a basis for stopping rules in the hill climbing routines. If, after a "major" iteration of a search routine (what is "major" depends on the search algorithm) the log likelihood has not increased by something of the order of magnitude of 2, it may be a sign that most of the gains to be made have been made and that further searching will achieve only insignificant adjustments of the parameter estimates. Certainly, for example, once the maximum has been "boxed in" (as when a linear search has passed over the maximum and seeks to locate the maximum more precisely), little is to be gained from trying to gain another .01 increase in the log likelihood (of course, for some search algorithms which require very precise linear searches, it may be desirable to push the search accuracy to the limits of the computer word length). But in general, the "two-sigma-two" property makes the likelihood surface easier to maximize than an arbitrary surface.

2.5.2 CONFIDENCE TESTS FROM THE OPTIMAL FILTER.

When the model used in the optimal filter is a "true" model, the normalized predicted residual process $\underline{\tilde{\delta}}_2(n|n-1)$ of the filter should be a white process of constant unit variance and normal distribution. Since these properties of the residual process are not used directly in maximizing the

log likelihood, they provide an independent test of model validity. The resulting tests are especially useful, since they are 1) sensitive to small errors in model specification, and 2) are not guaranteed to pass, even at the global maximum of the likelihood surface. Furthermore, the filter residuals, unlike the residuals from OLS, should be white, normal, and constant unit variance even for nonlinear, time-varying, and cross sectional models. Therefore, if used properly, these tests can be applied in all uses of FIMLOF.

Two methods are useful in testing for the whiteness of the residual process. First, a Durbin-Watson statistic[45] may be computed for each component of the measurement vector $\underline{z}(n)$:

$$DW_i = \sum_{n=2}^{N} \frac{\left[\tilde{\delta}_{z_i}(n) - \tilde{\delta}_{z_i}(n-1)\right]^2}{\tilde{\delta}^2_{z_i}(n)}$$

A second test is useful both for comparison with the Durbin-Watson statistics, but also for checking to see if the unit variance property is present in the residuals. The test consists of computing several correlation matrices on the residuals, defined as follows:

$$\underline{R}(j) = \frac{1}{N-j} \sum_{n=1}^{N-j} \underline{\tilde{\delta}}_z(n)\, \underline{\tilde{\delta}}_z'(n+j)$$

If the model is "true" then R(0) should be approximately equal to the identity matrix, and all other R(j) should be approximately equal to the zero matrix. The standard

deviations of each component of the various R(j) can be easily computed:

$$\sigma_{R(0)_{ii}} = \sqrt{\frac{2}{N}} \quad \text{(main diagonal elements of } R(0)\text{)}$$

$$\sigma_{R(0)_{ik},\, i\neq k} = \sqrt{\frac{1}{N}} \quad \text{(off-diagonal " " ")}$$

$$\sigma_{R(j)_{ik}} = \sqrt{\frac{1}{N} - \frac{j}{N^2}} \quad \text{(all elements of } R(j),\ j\neq 0\text{)}$$

The derivations are as follows. For the diagonal terms of R(0),

$$\text{var}\{R(0)_{ii}\} = E\{[(1/N)\sum_{n=1}^{N} r^2(n) - 1]^2\}$$

$$= 3N/N^2 + (2/N^2)\binom{N}{2} - (2N/N) + 1$$

$$= 3/N + [(N-1)/N] - 2 + 1 = 2/N$$

For the off-diagonal terms of R(0),

$$\text{var}\{R(0)_{ij}\},\ i\neq j, = E\{[(1/N)\sum_{n=1}^{N} r_i(n)r_j(n)]^2\}$$

$$= (1/N^2)\sum_{n=1}^{N} E\{r_i^2(n)r_j^2(n)\}$$

$$= (1/N^2)N = 1/N$$

For all terms of $\underline{R}(j)$, j>1, the following holds:

$$\text{var}\{R(j)_{ik}\},\ j>1, = E\{[(1/N)\sum_{n=1}^{N-j} r_i(n)r_k(n-j)]^2\}$$

$$= (1/N^2)(N-j)$$

$$= (1/N) - (j/N^2)$$

This last variance becomes 1/N if the circular correlation is used, where the last j-1 residuals are used as lagged residuals for the first j-1 terms of the full summation n=1....N.

Note that in the general format of the nonlinear model and data structure, the number of samples of data for each component of $\underline{z}$ may be different (the index n refers to

existing samples, not to time intevals!), so that in the general case, few components of the R(j) would have the same standard deviations. Thus, it is useful to normalize the $\underline{R}(j)$ matrices, and print out matrices $\underline{P}(j)$ in which each component of $\underline{P}(j)$ gives the number of standard deviations by which the corresponding component of the corresponding $\underline{R}(j)$ differs from its expected value:

$$\underline{P}(j)_{ik} = (R(j)_{ik} - E\{R(j)_{ik}\}) / \sigma_{R(j)_{ik}}$$

The resulting matrices $\underline{P}(j)$ provide a complete and readable test statistic, which has a uniform interpretation, regardless of data sampling distributions, model formats, or scaling factors.

In practice, it is often seen that $\underline{P}(0)$, $\underline{P}(1)$, $\underline{P}(2)$, and $\underline{P}(3)$ provide a sufficient test of the residual process. The sensitivity of the test is such that, for "real" data, the test should be considered passed if the terms of $\underline{P}(j)$ have absolute magnitudes of less than four or five. Experiments with monte carlo simulation data, in which the same model used to generate the data is used in the optimal filter, it is not uncommon to see components of $\underline{P}(j)$ greater than 2. Most experience to date indicates that the test is quite sensitive.

The correlation matrices $\underline{R}(j)$, however, should not be discarded; they are particularly useful for diagnostic analysis in case a validity test fails. The main diagonal terms of the $\underline{R}(j)$, for example, constitute a spectral

analysis of the residual process. The $\underline{R}(j)$, may lend insight into where to improve a model, and may be more easily interpreted for cross-sectional data, where the residuals may contain a bias component which is a white process among the regions, but constant within each region. An example of such analysis is given in Chapter 6.

The white, normal, and constant unit variance property of the normalized predicted residuals $\underline{\tilde{\delta}}_z(n|n-1)$ leads to another powerful test of model validity: the sum of the squared normahized predicted residuals (this sum is referred to as the SUMSQ statistic in this thesis and in GPSIE) is a chi square variable with known mean and variance:

$$\text{SUMSQ} = \sum_{n=1}^{N} \underline{\tilde{\delta}}_z'(n)\ \underline{\tilde{\delta}}_z(n) = \sum_{n=1}^{N} \underline{\delta}_z'(n)\ \underline{\Sigma}_z^{-1}(n)\ \underline{\delta}_z(n)$$

Thus, when a (possibly local) maximum of the likelihood surface is found, a comparison of SUMSQ with its expected value provides a quick indication of whether the global maximum has been found.

Examples of these tests in application to real data are found in Chapter 6.

2.6 Bad Data Detection and Robust Estimation.

The process of validation applies to data as well as to models. It is a mistake in any field to estimate or test a model using data that has not been examined for typographical errors, sensor failures, and other common sources of bad data. We so far in this presentation have

assumed that the data was correct, and that the model required estimation and validation. In this section we take the complementary approach of assuming that the model is reasonably accurate, but that there are scattered errors in the data.

We define "bad data" in the following way. First, consider errors in the measurement data $\underline{z}(n)$. Since we hypothesize that the "correct" data contains error $\underline{v}(n)$, we think of bad data as containing errors which which deviate strongly from the statistical model defining $\underline{v}(n)$ as a normal white process with covariance $\underline{R}(n)$. Specifically, if the normal data vector is

$$\underline{z}(n) = \underline{h}[\underline{x}(n),\underline{v}(n),n]$$

as in the measurement data of the standard model form of Figure 2-3, then a bad data point is defined by the following model:

$$\underline{z}(n) = \underline{h}[\underline{x}(n),\underline{v}(n),n] + \underline{e}_j$$

where $\underline{e}_j$ is a column vector of all zeros, except for the j-th component, which is one. The scalar b is assumed to be large enough in absolute value so as to clearly depart from the expected range of error introduced by $\underline{v}(n)$. Errors which are small enough to be consistent with the characteristics of the $\underline{v}(n)$ process are called noisy data, rather than bad data; we shall discuss the distinction later in Section 2.6.4.

FIMLOF makes possible a variety of powerful techniques

for bad data detection and identification. The bad data detection problem is to discover that a bad data point exists; the bad data identification problem is to isolate the location of the bad data point to a particular component of a particular data vector. The methods to be discussed below are all based on defining residual processes such that the residual component corresponding to the bad data point will be readily aparent to the casual observer or to a watchful computer program. Specific methods and their residual processes are derived in the following subsections.

2.6.1 NORMALIZED PREDICTED MEASUREMENT RESIDUALS (NPMR).

The normalized predicted measurement residuals $\underline{\tilde{\delta}}_z(n|n-1)$ (henceforth sometimes referred to as NPMR) are the residuals used to compute the correlation matrices $\underline{P}(j)$ of Section 2.5.2. Under an accurate model and data errors conforming to the assumptions of the model, the NPMR process should have constant unit variances under all conditions, including time-varying models, cross-sectional data, nonlinear dynamics, etc. A bad data point will usually, then, reveal its presence (but not location) via one or more components of $\underline{\tilde{\delta}}_z(n|n-1)$ with absolute values of greater than, say, 3 or 4. Unfortunately, there is no guarantee that the large residual components of NPMR will correspond to the source of the bad data, since each component of the normalized residual vector is a linear function of all

components of the unnormalized residuals, in the general case. Thus, the only practical response upon detecting an exceptional component of NPMR is to examine all the data at the appropriate sample time(s), hoping that the error in the original data will be aparent. There are two problems with this procedure:

1. The Bad data point may not be readily apparent in tables or graphs of the original data, since the bad data derives not from an absolutely large error, but from an error large compared with the modeled uncertainty. Thus, a 5% error in data assumed or estimated to be .1% accurate would be a bad data point. But if the dynamics of the system or vagaries of driving noise are of the order of 5% of the data values, it would be difficult, if not impossible to see the bad point in a graph of the single series. Only meticulous comparison of many data graphs would reveal the problem.
2. The second problem with the manual procedure of tracking down bad data from large components of NPMR is that the procedure cannot be readily mechanized. Thus, human intervention is required; such intervention may be impossible in some situations, and in any case should be optional rather than required.

In conclusion, the procedure of scanning the NPMR for

components with absolute values greater than 3 or 4 is useful for bad data detection, especially in relatively small samples, but does not allow bad data identification, nor can it be reduced to a robust estimation algorithm, except for the rather crude and unnecessary method of discarding all data at the sample time for which bad data is detected.

The following two sections describe residual processes which can be used for both bad data detection and bad data identification.

2.6.2 NORMALIZED UPDATED RESIDUALS (NUMR and NUSR).

This section defines two residual processes useful for both bad data detectiion and identification. The normalized updated measurement residuals NUMR reveal and locate bad data in the measurement vectors $\underline{z}(n)$. The normalized updated state residuals NUSR reveal and locate bad data associated with individual components of the state vector (primarily via exogenous input data). The treatment here is brief and informal; for mathematical details of the derivation of NUMR and NUSR, see Appendix C of this thesis.

The NUMR process is also denoted as $\underline{r}_z(n|n)$, and is defined by the following equation:

$$\underline{r}_z(n|n) = [\mathrm{diag}(\underline{\Sigma}_z^{-1})]^{-\frac{1}{2}} \underline{R}^{-1} [\underline{z}(n) - \hat{\underline{z}}(n|n)]$$

Similarly, the NUSR process is denoted as $\underline{r}_x(n|n)$, and is defined as:

$$\underline{r}_x(n|n) = [\text{diag}\{\underline{H}'\underline{\Sigma}_z(n|n-1)\underline{H}\}]^{-1/2}\underline{\Sigma}_x^{-1}(n|n-1)[\hat{\underline{x}}(n|n-1) - \hat{\underline{x}}(n|n)]$$

The NUMR and NUSR processes interact with each other, and must be considered together. They have two useful properties:

1. First, both NUMR and NUSR have constant unit variance. That is, each component of NUMR and NUSR has a constant standard deviation of one, under all circumstances, as long as the model is valid and the data conforms to the model. Thus, one or more components of NUMR or NUSR with absolute value greater than 3 or 4 is a reliable detector of the existence of a bad data point somewhere at the corresponding sample time.

2. In addition, it can be shown (Appendix C) that at a sample time involving a bad data point, the component of NUMR or NUSR with the maximum absolute value corresponds to the component of $\underline{z}(n)$ or $\underline{x}(n)$ which is in error. For example, a typographical error in the first component of $\underline{z}(3)$ may, depending on the model structure, cause several components of both $\underline{r}_z(n|n)$ and $\underline{r}_x(n|n)$ to exceed the acceptable limit (say, 4). But the component with the largest absolute value identifies the specific component of $\underline{z}(3)$ or $\underline{x}(3)$ which contains the typographical error. (In the case of $\underline{x}(n)$, the error might be in an exogenous

input to the equation determining the component of $\underline{x}(n)$).

These properties of NUMR and NUSR make it possible to write computer programs which automatically identify and delete bad data points, and allow the efficient screening of data sets for questionable entries. As their use in Chapter 6 shows, they can also be useful in model validation and model improvement.

2.6.3 ROBUST ESTIMATION.

Several features of FIMLOF make possible some new methods of robust estimation, closely related to the above ideas. We here define robust estimation to be an estimation method which is insensitive to occaisonal large errors in the data (bad data points). The new methods take advantage of the residual processes described above, and of the ability of FIMLOF to deal with missing data.

One method of robust estimation using FIMLOF consists of a two-step iteration:

1. Identify bad data points, using NUMR and NUSR.
2. Convert the bad data points to missing data points and reestimate.

A less rigorous way is to weight the exceptionally large residuals by some small fraction in computing the likelihood function. The unit variance property of the three residual processes discussed above make the weighting schemes more

effective than when they are applied to the data itself or to residual processes with nonconstant variance. The properties of NUMR and NUSR help avoid the common pitfall of downweighting some good data along with the bad.

2.7 Implementation: GPSIE.

The various FIMLOF-based techniques discussed in this chapter have been implemented in a computer program called the General Purpose System Identifier and Evaluator (GPSIE). GPSIE is a large precompiled program which, when coupled with a small program describing the particular model of interest, can be used to load data, compute likelihoods via optimal filtering, search for maxima in the likelihood function, compute the validity statistics discussed above, and plot the results. GPSIE embodies a large number of options for dealing with special cases and for maintaining efficient computation in various circumstances. For more details on GPSIE, see [Peterson,1974] and [Peterson,1975].

Chapter 3

FEASIBILITY OF FIMLOF AND GPSIE

The techniques described in Chapter 2 are relatively new. What evidence is there that the methods work? This chapter cites some FIMLOF results from engineering, and describes simulation work done by the author to test the methodology of FIMLOF as implemented in GPSIE. Section 3.1 describes some results for engineering systems using FIMLOF or approximations thereof. Section 3.2 gives some results of GPSIE on a simple first order system, to illustrate some of the techniques used to check out the software. Section 3.3 gives the results of the application of GPSIE to noisy data simulated by a high order, nonlinear dynamic model of a firm. The results are compared with those obtained by Senge [Senge, 1974] on the same model and data, using OLS and GLS. Section 3.4 concludes the chapter with some informal conclusions on the reliability of FIMLOF and GPSIE.

3.1 Early Results In Engineering.

The mathematics of FIMLOF were first developed by Schweppe [Schweppe,1965] in the context of engineering communications and control systems. Applications of the

full FIMLOF method, however, have been few, possibly because of the computational costs, and the availability of special-case simplifications, such as those used by Masiello [Masiello,1973]. The application by Moore and Schweppe [Moore,1972] on the estimation of unknown parameters for the control of nuclear power plants provides an instructive case study, in which some light is shed on problems of aggregation and sources of numerical error in the FIMLOF method. The FIMLOF estimation of aircraft parameters by Mehra and Tyler [Mehra,1973] provides another confirmation of the feasibility of the methods in the engineering context.

The applications to date in the engineering field have been for systems which differ in several ways from social system:

1. Engineering systems are usually modeled as continuous differential equations, via Newton's laws, with the structure well specified a priori. The major question in structural specification is not so much equation form, but level of aggregation. In contrast, econometric modelers, because of the paucity of theoretical specification of equation form, are "cursed" with great flexibility in choosing model structure.
2. Data for engineering systems tends to be available over many time increments, in contrast to the few

time increments available for social systems. For example, wind tunnel data or flight-recorder data may be avaiilable for thousands of seconds, while economic data extends for a few tens or hundreds of time steps of one month or one year. The value of the data for estimation and validation depends on the model structure (particularly the time constants, or eigenvalues of the state transition matrix), the number of pime steps sampled, and the distribution of the samples over time. However, it is a safe generalization that engineering data is usually of greater effective length than data from social systems.

3. Counterbalancing the above feature of data availability, a much broader selection of measured variables may be available for social models. In social systems, most model variables have some kind of data available, while in engineering systems, it is not unusual to have only a small fraction of the endogenous variables measured.

3.2 Results On A First-Order System.

The following FIMLOF results on a first-order, linear system illustrate some of the power of the method, and also serve as a confirmation of the GPSIE software, since for this simple system, analytical estimates of the hessian of

the log likelihood surface are available for comparison with the numerical results.

The following system is used to generate simulation data, using random number generators to supply the stochastic variables:

$$x(n) = s*x(n-1) + w(n)$$

$$z(n) = x(n) + v(n)$$

$$x(0) = 3$$

$$w(n) = N[0,q]$$

$$v(n) = N[0,r]$$

$$s=.75,\ q=1,\ r=1$$

In interpreting the following results, it is important to keep in mind the extreme amounts of noise present, both in the dynamic equation, and in the measurements z(n). In general, as the following figures confirm, the errors are larger in magnitude than the variables themselves. Figure 3-1 shows a graph of the "true" (as opposed to estimated) values of x(n) for n=1...21; Figure 3-2 shows the same samples of the measurements z(n). It should be emphasized that the estimations to follow in no way have access to the "true" data shown in Figure 3-1. The only data available to the filtering equations are the data z(n) shown in Figure 3-2.

Two estimations were performed, using 100 and 1000 data points (the first 31 of which are shown in Figure 3-2). GPSIE was used to estimate values of parameters s,q, and r.

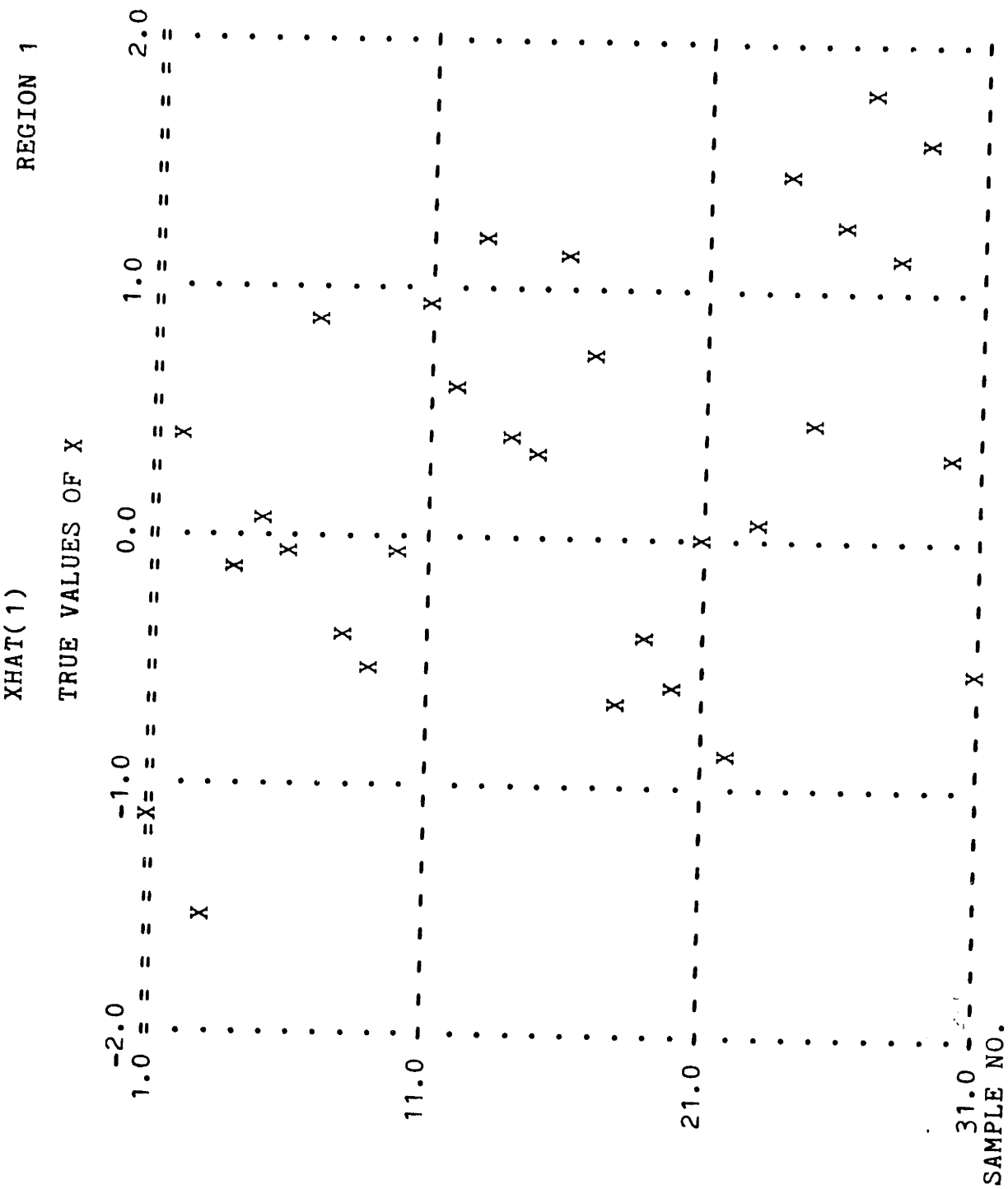

Figure 3-1

True Values of the State Variable (Inaccessible)

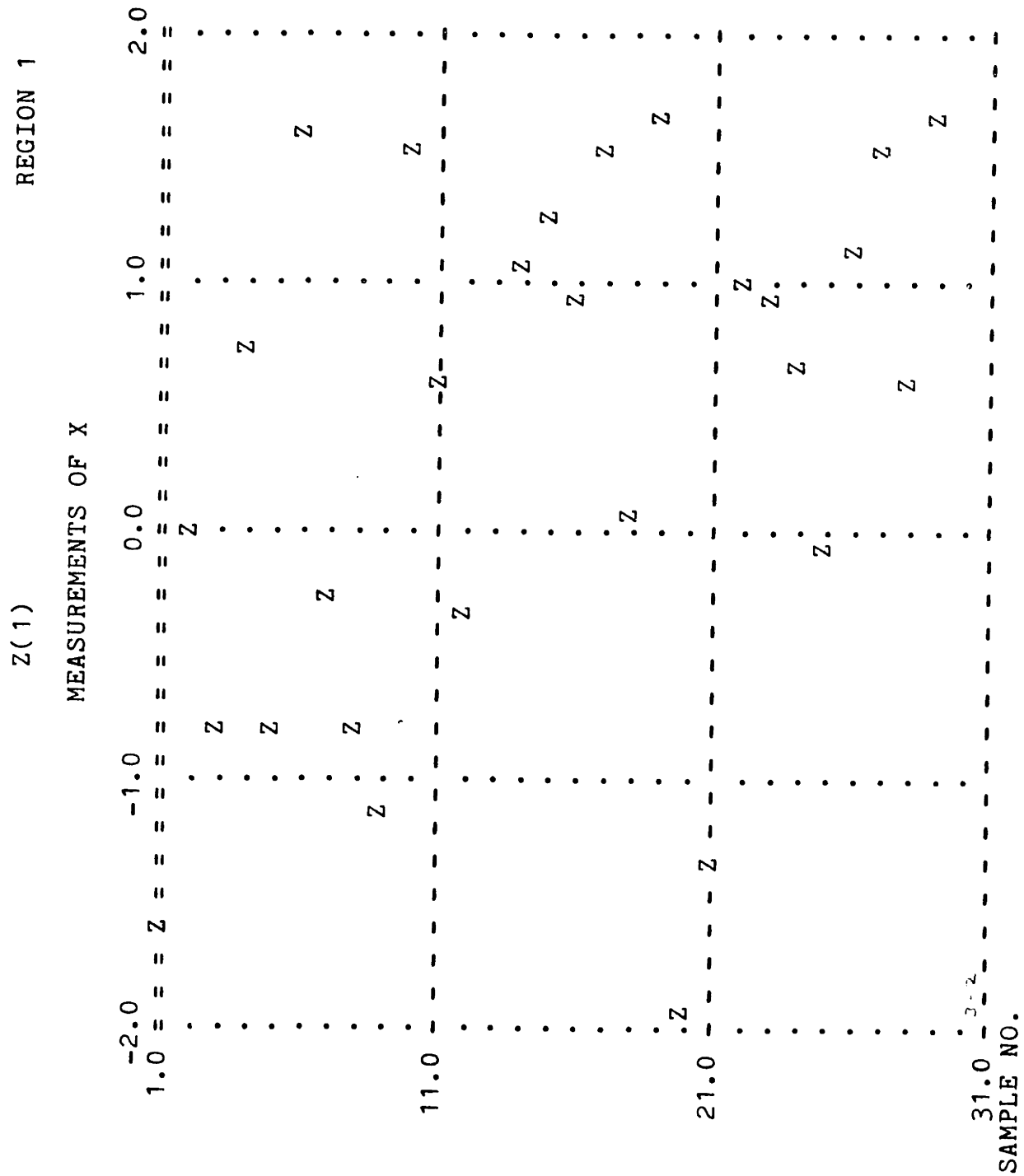

Figure 3-2

Noisy Measurements of the State Variable

(Used for Estimation)

Since simulation data was used, the estimates may be compared with the "true" values of the parameters; since the system has a simple structure, we may derive analytically an approximation of the hessian (second partial) of the likelihood surface with respect to the single parameter s, and compare the resulting value with those computed by GPSIE.

Figure 3-3 shows the results for the 1000 data-point sample. The estimates of all three parameters are quite close. Note the ability of FIMLOF to operate correctly in the face of "errors in the variables" -- that is, measurement error in variables on the right-hand side of the difference equations, without the use of instrumental variables. The FIMLOF algorithm is also capable of using information from previous time steps, via lagged endogenous variables, to separately estimate the variances of the equation error and measurement error. Note that these estimates of variances are max like estimates of parameters in an explicit stochastic model, rather than "sample variances" in the conventional sense. The lower precision of the variance estimates, as compared with the structural parameter estimate, is typical of max like estimates of dynamic systems.

The validity statistics shown in Figure 3-3 strongly support, as might be expected, the hypothesis that the model is consistent with the data. The Durbin-Watson statistic is

no. of data points		1000	
parameters	s	q	r
true values	.75	1	1
estimates	.75	.97	1.1
std. errors	.04	.20	.17
log likelihood		-932.88	
SUMSQ		989	
exp. val. of SUMSQ		997	
std. dev. for SUMSQ		44.7	
Durbin-Watson statistic		2.01	
normalized correlation matrices (no. of std. dev. from exp. val.):			
P(0)		-0.3	
P(1)		-0.1	
P(2)		1.0	
P(3)		-0.1	
std error of s at true values of parameters		.024	
expected std. error of s at true values of parameters		.021	

Figure 3-3

Estimation and Validation Results,
First-Order System,
1000 Data Points.

excellent, as are the normalized correlation matrices (the normalized correlation test is considered "good" if all the terms have absolute values less than about 4; if all terms have absolute values less than 2, as is the case here, the test passes easily, both for whiteness and unit variance). The final two terms in Figure 3-3 provide a check on the GPSIE software. For the s parameter in this particular system, one can compute an asymptotic expected value for the standard error of the parameter s, as measured by the inverse of the second partial derivative of the log likelihood surface with respect to s, evaluated at the true values of the parameters. In this case, the expected value is .021, and the numerical result is .024. Considering the many approximations involved, the match is good. This test, of course, has limited usefulness in "real" estimation situations, where the expected value of the standard error cannot be computed analytically, and where the true values of the parameters are not available. In such cases, the second partial derivatives (hessian matrix) of the log likelihood surface are computed at the maximum likelihood point (as opposed to the unknown "true" point). Figure 3-3 shows that even in the best of circumstances, the approximation should only be taken as an order of magnitude estimate of the standard error, since the standard error of s at the max like point is .04; at the "true" parameter values, .024.

In summary, the results shown in Figure 3-3 indicate that the technique is numerically sound, and that accurate estimates can be made in the face of considerable equation error and measurement error, without resort to instrumental variables, and without prior knowledge of the relative or absolute sizes of the error processes.

Figure 3-4 shows results analogous to those of Figure 3-3, but for only 100 data points. Considering the domination of the system by error processes (see Figures 3-1 and 3-2), the results are excellent. The estimates of the error variances are quite approximate, as indicated by the standard errors, but the structural parameter s is estimated quite accurately, and the validity statistics strongly support the hypothesis that the model is consistent with the data.

3.3 Feasibility Tests--High-Order, Nonlinear Systems.

The tests in the preceding section lend support to the FIMLOF method and the GPSIE software, but the only conditions of the tests which could not be considered benign are the large error variances. In this section, we conduct a similar test, for a more realistic model. The conditions of the experiment are chosen to allow comparison with the work of Senge [Senge,1974]. Senge estimated parameters from noisy simulation data generated by a nonlinear, dynamic model of the firm [Forrester,1967]. The model consists of 9

no. of data points	100		
parameters	s	q	r
true values	.75	1	1
estimates	.75	.76	.94
std. errors	.16	.69	.55
log likelihood	-85.0		
SUMSQ	101		
exp. val. of SUMSQ	97		
std. dev. for SUMSQ	14		
Durbin-Watson statistic	2.05		
normalized correlation matrices (no. of std. dev. from exp. val.):			
P(0)	-0.1		
P(1)	0.1		
P(2)	.3		
P(3)	-1.4		
std. error of s at true values of parameters	.082		
expected std. error of s at true values of parameters	.066		

Figure 3-4

Estimation and Validation Results
First-Order System,
100 Data Points.

dynamic diffence equations, defining 9 state variables, 7 of which are measured. Senge simulated the model, using random number generators to introduce both equation errors and measurement errors (errors in the variables), in differing amounts. In the particular experiment compared below, Senge introduced equation errors ranging from 6% to 60% of the mean of the endogenous variables, and obtained excellent estimates of the 13 system parameters, using OLS and GLS. However, when Senge introduced 10% measurement error on the 7 measured endogenous variables, he obtained large errors in the parameter estimates. Figure 3-5 shows the results Senge obtained, compared with the estimates obtained with GPSIE under the same conditions. The results indicate that FIMLOF techniques, as implemented in GPSIE, may yield accurate results, even in the presence of system nonlinearities and measurement error which may cause difficulties with simpler estimation techniques. The OLS estimates shown in Figure 3-5 would require reestimation by means of addtional data or some other method; the FIMLOF technique leads somewhat more directly to correct estimates.

The above comparison should be interpreted only as a (somewhat trivial) confirmation of the computational integrity of GPSIE, since no violation of the mathematical basis of FIMLOF was introduced. The failure of OLS under the circumstances is less surprising, since the method assumes accurate measurement of the equation right-hand-side

parameter name	true value	GPSIE est.	OLS estimate
SEM	400	392	4349
SED1	-.0281	-.029	-.430
SED2	-.0295	-.0295	.096
SED3	.00228	.00228	-.0074
PCF1	.61782	.615	3.7117
PCF2	-.13244	-.132	-.74891
CEF1	-.0698	-.0693	.03966
CEF2	.12442	.1245	-.14609
CEF3	-.08138	-.0813	.13853
CEF4	.027704	.02704	-.03144
DRAT	1	.97	1.3
SAT	20	19.85	18.5

Figure 3-5

Comparison of Estimation Techniques for
Ninth-Order Nonlinear System with Errors In Variables

variables, which were measured with 10% error in the experiment. Senge is continuing his research to deal with more realistic conditions; it is expected that a future area for productive research would be to extend such comparisons of estimation techniques to gain an understanding of relative merits and efficiencies, and to sketch the boundaries of the conditions under which various methods fail.

Chapter 4

SOURCES OF MODEL STRUCTURE

This chapter discusses the first step in the three-step process of model building: hypothesizing a model structure. The hypothesis of a model structure is seldom discussed in print; the model structure is often taken as given, and analysis proceeds from that point. We discuss the building of model structure here for two reasons:

1. The techniques of estimation and hypothesis testing discussed in this work apply to the general class of nonlinear dynamic models. That is, the models need be neither linear nor linear in the parameters. For these linear models, the distinction between structure and parameters is clear, but for the general class of models discussed in this thesis, the distinction between structure and parameter becomes fuzzy. In nonlinear models, parameters may take on qualities commonly associated with structure.
2. The second reason for discussing the formulation of model structure is that the ability to deal with a wide class of structures leads to methods of model formulation which entail more freedom than is commonly available to econometric model builders. The methods developed in this thesis

give the econometric model builder more freedom of choice both in what variables to include in the model, and the way in which the variables interact.

4.1 Structure vs. Parameters: Estimating Structure.

Before discussing methods for hypothesizing model structure, we shall define what we mean by "structure", vs. "parameters." For time-invariant linear systems, the distinction is clear: the system must be (by definition) of the form

$$\underline{x}(n) = \underline{A}\underline{x}(n-1) + \underline{B}\underline{u}(n) + \underline{w}(n)$$

where $\underline{A}$ and $\underline{B}$ are constant matrices, $\underline{x}(n)$ is the state vector at time n, $\underline{u}(n)$ is a vector of known inputs, and $\underline{w}(n)$ is a white, normal process of mean $\underline{0}$ and covariance $\underline{Q}$. In this case, the parameters are simply the constant coefficients of the matrices $\underline{A}$, $\underline{B}$, and $\underline{Q}$. A similar definition can be made in the case of systems which are linear in the parameters. For example, in the system

$$\underline{y} = \underline{X}\underline{b}$$

where $\underline{y}$ is a vector of outputs, $\underline{X}$ is a matrix of variables which may be functions of exogenous inputs and of lagged values of $\underline{y}$; $\underline{b}$ is defined as the vector of parameters.

In general nonlinear systems, however, we shall have to offer a more general definition of parameters:

A <u>parameter</u> in a nonlinear system is a constant exogenous input to the system.

By the above definition, a parameter in a nonlinear system may enter the system in any nonlinear way. The parameter may be known or unknown, but it is always asserted (in this thesis) to be a constant whose value is not determined by the rest of the system. Note that the above definition is superficially incompatible with the idea of "time varying" parameters discussed in the literature; see Section 4.4 for a reconciliation. It is obvious that very few model constructs can be asserted to be "genuinely" constant, but we define parameters as those things chosen by the modeler to be represented as constant over the time base treated by the model.

The above definition may seem reasonably clear, but it has the following drawback, for which there may be no remedy: in nonlinear systems, parameters may take on qualities usually associated with structure. For example, consider the system

$$\underline{x}(n) = \theta\underline{f}(\underline{x}(n-1)) + (1-\theta)\underline{g}(\underline{x}(n-1))$$

By any reasonable definition, θ would be considered a parameter in the above equation. But θ determines the structure of the system. If $\theta = 0$, then the system has the structure determined by the function $\underline{f}$; if $\theta = 1$, then the system structure is determined by the function $\underline{g}$.

Therefore, we must be careful in applying the usual connotations to the terms "structure" and "parameter" when dealing with nonlinear systems. By the above sort of model

building, we may estimate parameters which de facto result in the estimation of structure.

The estimation of structure (as in estimating θ in the above system) may be thought of as a kind of continuous hypothesis test. The maximum-likelihood value of θ may be thought of as selecting the most likely structure from the range of structures implied in the above equation.

It is important to reemphasize here that neither parameters nor structure can be usefully estimated from merely numerical data and thin air. Estimation always entails a choice from a range of alternatives. A well hypothesized model defines a range of plausible alternatives consistent with the purposes of the study at hand. The ability to estimate structure increases, rather than decreases, the role of experience, logic, and theory in model building.

4.2 Hypothesizing Model Structure.

Most of the tools of estimation commonly used by econometricians today restrict the model builder to structures which are linear in the parameters. The problem of structure hypothesis, then is reduced to two axes of freedom: 1) what variables to include in the model, and 2) what functional manipulations to perform on the data before they are introduced into the linear structure.

The usual techniques also require that every variable in

a model must measured by an available data stream of a common sampling frequency, coupled with the automatic structuring of the model in the linear in the parameters format, leads to the possibility of model formulation dominated by the constraints of the estimation techniques.

The techniques described in this thesis extend the freedom of the model builder to include the following kinds of considerations and choices:

1. Variables (endogenous) may be included in the model for which there is no data. For example, expectations and other difficult to measure model constructs may be explicitly included in the model.

2. Data of differing sample frequency may be applied to the estimation and evaluation of a single model, without the use of interpolation or deletion of data. For example, monthly, quarterly, and yearly data may all be used at once to estimate parameters in a single econometric model.

3. The time interval of the model need not correspond to all or any of the data sampling intervals. For example, a weekly model may be estimated using monthly and yearly data. Thus, numerical instabilities arising from large time steps may be avoided. Continuous models may be estimated using

discrete data.

4. Measurement errors in data may be taken into account without the use of instrumental variables; the characteristics of measurement error may be estimated.

5. Unknown parameters may enter in the model formulation in any nonlinear way. The modeler need not spend time attempting to approximate the theoretically correct model by a linear form.

The above freedoms in the specification of model structure may raise questions about where model structure comes from. Lurking in the background of the model building process are the influences of plausibility tests, practical experience, and theoretical considerations, which lead the researcher to try out some regressor combinations suggested by theoretical causality, and to reject other combinations as unreasonable a priori. These "background" considerations and limitations, used to constrain the options considered in model formulation, become the dominant considerations in modeling under the structural freedoms introduced by FIMLOF. Therefore, it is appropriate to discuss the various sources of model structure.

There are three sources of model structure: 1) Explicit theory, 2) Limitations imposed by data and numerical techniques, and 3) Implicit theory.

Explicit theory consists of general model structures

which are commonly accepted as useful or "true." Newton's laws of mechanics provide the standard example of an explicit theory, widely used in engineering and the physical sciences for the formulation of dynamic models. Unfortunately, the truism holds that explicit theories as general and useful as Newton's are lacking in the social sciences.

Limitations imposed by data and numerical techniques, as discussed above, have had a considerable constraining influence on models in the social sciences. One of the purposes of this work is to reduce these limitations. As mathematics, computational power, and the collection of data continue to develop, we may expect these limitations to diminish further, and become less influential in the formulation of models in the social sciences.

Implicit theory consists of the informal and intuitive plausibility tests used by both physical and social scientists to select and reject model formulations a priori. Implicit theory may be thought of as a collection of (often unstated) guidelines, ranging from simple consistency to matters of purpose and scientific taste. For example, a common guideline in model formulation is the implicit theory that the model should reduce to the proper limiting cases, when parameters are pushed to extreme values. To some extent, such guidelines are trite and obvious; from an other point of view, they lack precision. Discussions of such

ill-defined bases of model formulation are understandably avoided in the literature, but as the limitations imposed by data and numerical techniques are lessened, it will become increasingly important to recognize and state our implicit theories. A tentative list of informal theories is included as Appendix B of this thesis to illustrate the idea.

4.3 Simultaneous Equations (Implicit Functions).

Econometric models sometimes contain simultaneous formulations which reduce to the implicit dynamic form

$$\underline{g}[\underline{x}(n),\underline{x}(n-1),\ldots] = 0,$$

instead of the standard form, in which $\underline{x}(n)$ can be explicitly computed:

$$\underline{x}(n) = \underline{f}[\underline{x}(n-1),\ldots].$$

Although FIMLOF techniques can be used directly on the implicit formulation (see Section 2.2.18), the greater freedom of model formulation and data processing allowed by FIMLOF may reduce the need for implicit formulations in the first place.

For example, the implicit form may arise from the assumption that an economic system is in equilibrium. But the economic market being modeled, like most markets, may be prone to cycles and exogenous shocks (disturbances) which make it difficult to assume *a priori* an equilibrium condition. Under FIMLOF, it is possible to model the full disequilibrium structure of the market, and, via parameter

estimation in a nonlinear structure, determine the degree to which equilibrium is achieved. This approach is made possible by the ability of FIMLOF to estimate models which 1) have several time steps between data points, so that, for example, a model with weekly equilibrating dynamics could be estimated on monthly and yearly data; and 2) contain unmeasured variables, such as expectations and perception lags, which may cause disequilibrium.

4.4 Time-Varying Parameters.

The definition of parameters as constant exogenous inputs (Section 4.1) seems to be at variance with the notion of time-varying parameters, such as

$$x(n) = bx(n-1) + w(n),$$

where b may vary over time in some unpredictable fashion. This subsection describes two approaches to the time-varying parameter problem which are facilitated by FIMLOF techniques.

4.4.1 PARAMETERIZATION OF PARAMETERS.

One approach to the time-varying parameter problem is to parameterize the parameter variation. For example, one may replace the above first order system with the following second order system, in which the time-varying parameter becomes a state variable:

$$x(n) = b(n)x(n-1) + w(n)$$

$$b(n) = \theta b(n) + v(n)$$

Where the approach is to use FIMLOF to estimate θ, and use optimal filtering to track b(n) over time. The contribution of FIMLOF is in the estimation of the parameter θ, as the above system is neither linear in the parameters nor fully measured (no data, obviously, is available on b(n)). The computational feasibility of the approach is in this case untried, but the full-information aspects of FIMLOF suggest that the above approach might be a productive area of enquiry.

4.4.2 FAULT DETECTION.

An alternative approach to time-varying parameters would be to model the parameter as piecewise constant, estimate its value over a relatively short period of data, and then monitor the validity statistics from the optimal filter to look for indications that the model no longer is consistent with the data. When the validity test fails, "fault detection" is said to have occurred, and reestimation of the parameter(s) is called for. Note that this method can be used both with reference to the original time varying parameters, such as b above, or on second-order parameters, such as θ. For an example of this approach, see [Moore,1972].

Chapter 5

ESTIMATION AND REJECTION OF A PRELIMINARY MODEL

This chapter describes the use of GPSIE in estimating parameters and testing the validity of a preliminary model of fuel demand in the residential commercial sector. After much exploration, the model was rejected. Section 5.1 gives a brief overview of the model structure and its background. Section 5.2 describes the estimation results and validity statistics, which led to the rejection of the model.

5.1 Preliminary Model Structure.

The model structure described here is a tentative structure developed by Baughman and Joskow [Baughman, 1973]. The model describes the demand for natural gas, fuel oil, and electricity in the residential commercial sector of the U.S., as a function of fuel prices, income per capita, and capital costs of oil and gas furnaces. The dynamic structure of the model is defined by the following three equations:

$$E(n) = B*E(n-1) + Fe*MSD(n)$$

$$G(n) = B*G(n-1) + Fg*MSD(n)$$

$$O(n) = B*O(n-1) + Fo*MSD(n)$$

where E(n), G(n), and O(n) are the demands for electricity,

gas, and oil in year n, measured in BTU's. The B coefficient represents the lag effects in fuel demand, due to two sources: finite depreciation times for fuel-consuming appliances, and the psychological and perception delays inherent in consumer decisions. Thus, for example, B=.9 would, in a rough way, correspond to an appliance lifetime of about 10 years, since ceteris paribus one tenth of demand would disappear each year unless replaced by the installation of new or replacement appliances. Included in the 10 years are psychological effects, so that a given B would imply a lifetime somewhat longer than the physical lifetime of the appliances.

MSD (Market Sensitive Demand) represents the number of BTU's of fuel demand "up for grabs" each year. MSD is the sum of replacement demand and new demand from growth. The replacement demand is defined as

$$(1-B)E(n-1) + (1-B)G(n-1) + (1-B)O(n-1)$$

which is equivalent to assuming that all depreciated appliances are replaced, possibly by an equivalent appliance burning a different fuel. The growth in demand (which may be negative) is due to changing population and income growth.

The equations for the fuel market shares Fe, Fg, Fo are in a logit-type form:

$$Fe = 1/TM$$

$$Fg = GM/TM$$

$$Fo = OM/TM$$

where

$$OM = \exp(-AA*Po + AB*Pe + AC)$$

$$GM = \exp(-AD*Pg + AB*Pe + AE)$$

$$TM = 1 + GM + OM$$

where TM, GM, and OM are defined only for convenience of notation; AA,AB,...AE are parameters to be estimated; and Po, Pe, and Pg are fuel prices.

The preliminary model defined above meets several basic tests of model specification. For example, the fuel market shares F_i always sum to one, and a fuel whose price becomes very large is reduced to zero market share. While relatively simple in form, the structure embodies some key features of the dynamics of interfuel competition.

Note that the preliminary model is not, however, linear in the parameters. The model can be converted to a linear in the parameters form only be setting B to zero, and normalizing the gas and oil equations by the electricity demand:

$$G(n)/E(n) = Fg/Fe = GM$$

$$O(n)/E(n) = Fo/Fe = OM$$

Taking the natural logarithm of both sides yields the linear form

$$\ln\{O(n)/E(n)\} = -AA*Po(n) + AB*Pe(n) + AC$$

$$\ln\{G(n)/E(n)\} = -AD*Pg(n) + AB*Pe(n) + AE$$

Thus, we achieve a linear-in-the-parameters model, but only

by eliminating the dynamics. Setting B=0 is equivalent to assuming that the original model is always in equilibrium, or to asserting that fuel demands will repond immediately (within one year) with full effect to changes in fuel prices. For example, the resulting model asserts that existing stocks of gas furnaces, water heaters, and clothes driers will have little or no influence on next years' consumption of natural gas.

The unreasonableness of the static, linear in the parameters model was one of the prime motivations for developing GPSIE. The next section describes the estimation work on the above preliminary model, which led to the rejection of the model and to the improved model described in Chapter 6.

5.2 Estimation of the Preliminary Model.

Two kinds of estimations were performed on the preliminary model before GPSIE was applied to it. First, the model was estimated in its linear form (with B=0), using single equation techniques (OLS and WLS). In addition, estimation was attempted for the model with B>0 using simultaneous equation estimation. The latter method is a technique much like the "OLS the hard way" method described in Chapter 2. The fuel consumption variables are initialized at the data n_th data point, and are simulated forward one time period. The residuals (differences between

the simulated variables at the (n+1)_th period and the data at period n+1) are squared, added to the accumulated total of squared residuals, and the equations are reinitialized at the (n+1)_th data values, and the iteration is repeated. The sum of the squared residuals is then minimized by nonlinear optimization (inverse hill climbing, or unconstrained minimization). Unfortunately, this method yielded an estimate of B greater than one, an a priori absurd value.

The strategy with GPSIE was to first hold B=0 while estimating the remaining parameters, as a check on the OLS estimates for the static model. Then GPSIE would be used to estimate all the parameters (including B) in unconstrained maximization of the likelihood function. Validity tests were performed at important checkpoints.

A PL/I subroutine was written, defining the preliminary model in GPSIE-compatible format, with the potential search space of parameters B, AA, AB, AC, AD, and AE.

GPSIE requires assumptions about the measurement error in the data, and the equation errors. These were initially supplied as parameters based on judgement. It was intended to estimate these error variances by including them as unknown parameters, but the preliminary model was rejected, as we shall see, on other grounds before that stage of investigation was reached. The standard deviation error of each of the three fuel-split equations was assumed to be 5%.

The standard errors in the fuel consumption data were assumed to be (based on extensive experience with the data) to be 5% for the electricity data, 15% for the natural gas data, and 40% for the oil data. The rather high error estimate for the oil data was well justified. The oil consumption data (as supplied by the Bureau of Mines) is known to suffer from considerable errors, due to erratic reporting, and due to occaisonal confusion in the field as to whether to report the consumption as commercial or industrial. The error is especially noticeable in states where consumption of fuel oil is low, such as the midwest. Figure 5-1 gives a good example of the kinds of seemingly large errors that can occur in such data. It is difficult indeed to believe that the consumption of oil in Texas fell by half in 1961, stayed fairly constant for several years, and then abruptly doubled in 1969. Thus, the large initial estimates of measurement error for oil.

First, the six parameters to be estimated were set to the values estimated via OLS:

parameter	OLS value
B	0
AA	4.84
AB	.253
AC	5.12
AD	1.13
AE	.857

At these initial OLS estimates, the log likelihood was

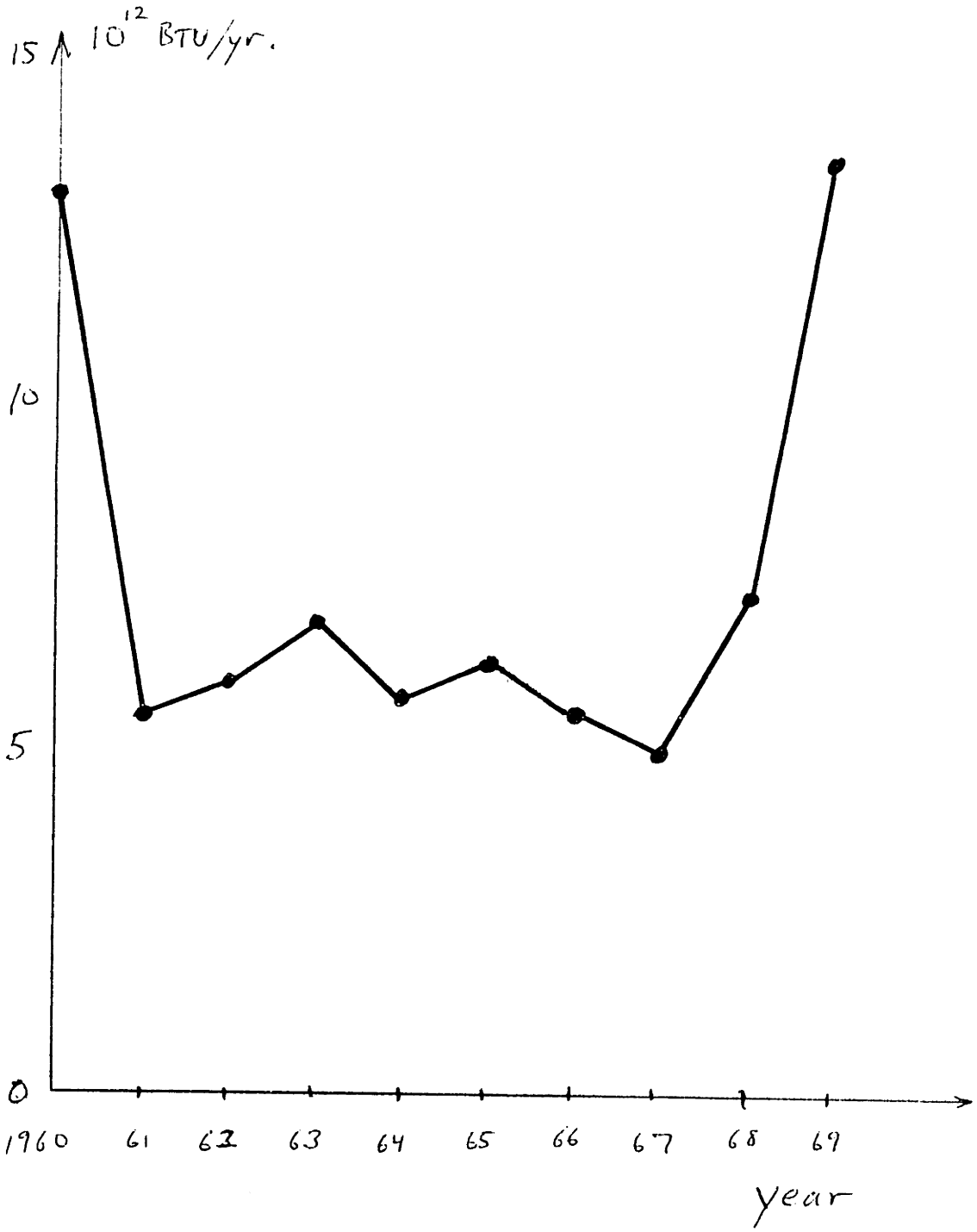

Figure 5-1

Oil Consumption in Texas, 1960-1969
Residential and Commercial Sectors
(Source: "Shipments of Fuel Oil and Kerosine,"
Bureau of Mines Mineral Industry Survey)

computed using GPSIE. It was -43,934. The sum of the squared normalized residuals SUMSQ was 77,900, but its expected value was only 1470, with a standard deviation of 54. The huge discrepancy between SUMSQ and its theoretical value is a strong indication of inconsistency between the model (with the OLS parameters) and the data. The large SUMSQ suggests that either the model is inaccurate, yielding large residuals, or that the equation and measurement errors have been understated. The generous error assumptions make the latter possibility unlikely, so we take the SUMSQ validity test as a strong negative, in consonance with the prior judgement that the dynamic effects of appliance stocks are important in fuel consumption. In addition, the correlation matrices $\underline{P}(j)$ (see Chapter 2) were computed to test the whiteness of the normalized residuals. The whiteness statistics, like the SUMSQ test, indicated a gross inconsistency between model and data. Of the 36 elements of the P matrices, the smallest term was 73, indicating a cross-correlation 73 standard deviations higher than its expected value. The other terms were even worse, ranging to a maximum error of 2,084 standard deviations.

Next, various searches were performed to find parameters of higher likelihood. The parameter B was held constant at various values, and the other parameters were varied to find the maximum likelihood, given the fixed value of B. In all cases, the gains in likelihood were small

compared to the gains to be found by varying B itself. For example, with B held at zero, the other parameters as a direct comparison with the OLS results. Although this procedure yielded 50% reductions in the estimates of AB and AD, the log likelihood increased only about 10%, to -39,846. The SUMSQ statistic at this point had been drasticaly reduced to 2941, but this was still 27 standard deviations away from its expected value of 1470. The whiteness matrices still had no terms within the acceptance band of 4 standard deviations, and the maximum term was 1581 standard deviations. These results were strong evidence against the hypothesis B=0.

To illustrate the great sensitivity of the system to changes in B, consider the results of a one dimensional search along B, holding the other parameters constant at the values obtained in the search with B=0. The log likelihood surface with respect to B was found to have a smooth, unimodal shape with a clear maximum at B=1.05! At this point, the log likelihood was -6421, with SUMSQ statistic of 2941 (off by "only" 27 standard deviations). The whiteness test matrices were also greatly improved, with 33% of the terms falling within the 4-sigma acceptable band, and a maximum deviatio of 50 standard deviations. In spite of the seemingly absurd value of B=1.05, these statistics indicated a strong preference over other values of B. For example by holding B constant at the plausible value of B=[illegible], and

searching over the other parameters, the maximum log likelihood was 12,909, a factor of 2 worse. At B=.9, the search over the other parameters converged at a log likelihood of -9,100.

Finally, all six parameters were searched simultaneously, leading to a maximum log likelihood of -5,524. The final parameter estimates were:

parameter	OLS estimate	FIMLOF estimate
B	(0)	1.09
AA	4.84	5.29
AB	.253	2.42
AC	5.12	6.44
AD	1.13	2.53
AE	.857	1.71

At the FIMLOF parameter values, the SUMSQ was 1165, off 5 standard deviations from its expected value of 1464, a significant discrepancy but not beyond hope, especially since a search over the error variances would be warranted. The whiteness test was similarly close, if not completely satisfactory: 81% of the terms were within the 4-sigma band, and the maximum error was 10 sigma.

There remained, however, the mystery of the large B. Why would a physically absurd value be estimated as most likely to be consitent with the data? A solution to the mystery was suggested when the hessian of the log likelihood surface was computed at the estimated point. The hessian turned out to be singular, because perturbations along the

parameter AB (coefficient on the Electricity price) created zero change in the likelihood. Further exploration revealed that, within machine precision, all values of AB above the "estimated" value of 2.42 were equally likely. After some examination, a quick calculation revealed that the maximum likelihood values of AB had in common the feature that they were large enough to give electricity a zero market share at historical prices. Further examination revealed that the two seemingly absurd results combined to yield a reasonable picture. A large AB and a B>0 cause the model to reduce to the following limiting case:

$$OM >> 0$$

$$GM >> 0$$

$$Fe = 0$$

$$TM = OM + GM$$

$$E(n) = B*E(n-1)$$

$$G(n) = B*G(n-1) + Fg*(1-B)*[E(n-1)+G(n-1)+O(n-1)+growth]$$

$$O(n) = B*O(n-1) + Fo*(1-B)*[E(n-1)+G(n-1)+O(n-1)+growth]$$

Note that the equation for electricity demand has become completely uncoupled from the exogenous inputs and from the rest of the system. For the estimated parameters, electricity demand grows by a constant fraction (9%) each year, no matter what happens to the prices of any of the fuels. Gas and oil are thereby left to fight over the remaining market sensitive demand, which is equal to

$$(1-B)*[E(n-1)+G(n-1)+O(n-1)+growth]$$

However, since B>1, the market sensitive demand is negative. Both gas and oil, like electricity, are given, by B>0 a constant percentage increase in demand, but then some of that increase may be taken away, via the negative market-sensitive demand.

In retrospect, the result is internally consistent. The only consumption data of high confidence is the electricity consumption series. These are without exception, smooth exponential growth curves, growing without fail at about 9% per year. Thus, the model matches the electricity data almost perfectly. The model may do a much poorer job at matching the noisy oil and gas data, but GPSIE has been "told" (and rightly so) not to take the oil and gas data as seriously as the accurate electricity data. Thus, to some extent, the gas and oil equations, with their backwards response to prices, are weighted out of the likelihood computation.

At this point, a host of improvements in the model and interesting estimation experiments suggest themselves. Enough flaws had been uncovered in the basic model structure, however, to justify repairing the structure before indulging in more estimation and validity testing.

The "failure" of the preliminary model was included here because it illustrates some of the power and insight of the FIMLOF method, by which one may gain not only tests of

model validity but also some idea of how to remedy the situation.

Chapter 6 presents a reformulated model which stands up better under scrutiny.

Chapter 6

RESULTS ON IMPROVED FUEL DEMAND MODEL

Following the difficulties with the preliminary model described in Chapter 5, Baughman and Joskow formulated a new structure for fuel demands in the residential commercial sector. The structure is shown below. An equation is shown below. To eliminate estimations based on growth in total demand modeled as a completely unknown input, an equation describing total energy demand has been added:

$$\frac{\text{energy}}{\text{population}} = \exp\left\{AT + BT * \frac{\text{income}}{\text{popula't.}} + CT * (\text{min. temp.}) + DT * \frac{\text{popula't.}}{\text{area}} + ET * (\text{av. price})\right\} * \left(\frac{\text{energy}(-1)}{\text{pop.}(-1)}\right)^{FT} \qquad (6.1)$$

The above equation (6.1) defines total energy demand per capita, as a function of income per capita, minimum annual temperature, population density, and average fuel prices. The dynamic effects of appliance stocks and perception delays are incorporated in the model, but in a nonlinear,, exponential expression, rather than the linear lag of the preliminary model. Simulation experiments have shown that the dynamic behavior of the new lag formulation is similar to the previous formulation, but the new formulation has the

advantage of being linear in the parameters, through the following transformation:

$$\ln\left(\frac{\text{energy}}{\text{pop.}}\right) = AT + BT * \frac{\text{income}}{\text{pop.}} + CT * (\text{min. temp}) + DT * \frac{\text{pop.}}{\text{area}} + ET * (\text{av. price}) + FT * \ln\left(\frac{\text{energy}(-1)}{\text{pop}(-1)}\right) \quad (6.2)$$

Since the above equation determines total fuel demand, only two more equations are required to determine the separate demands for gas, oil, and electricity. The following equations determine the relative consumption levels for the three fuels:

$$\ln\left(\frac{\text{gas}}{\text{elect.}}\right) = A + C * \ln\left(\frac{\text{gas price}}{\text{elect. price}}\right) + D * (\text{max temp}) + F * (\text{min. temp}) + H * \ln\left(\frac{\text{gas}(-1)}{\text{elect.}(-1)}\right) \quad (6.3)$$

$$\ln\left(\frac{\text{oil}}{\text{elect.}}\right) = B + C * \ln\left(\frac{\text{oil price}}{\text{elect. price}}\right) + E (\text{max temp}) + G * (\text{min. temp}) + H * \ln\left(\frac{\text{oil}(-1)}{\text{elect.}(-1)}\right) \quad (6.4)$$

The variables in the revised model are defined as follows. Gas consumption is measured in "effective BTU's", which are defined as one-half the actual number of BTU's of

gas consumed, in order to make the units of gas consumption commensurate with the measure of electricity consumption. The "effective BTU" formulation represents heat value wasted in gas furnaces and water heaters, through venting of combustion products. The formulation also has the effect of making gas prices more comparable with electricity prices, since electricity generation involves a similar heat loss in generation and transmission processes. Similarly, oil consumption is measured in effective BTU's, equal to half the actual BTU content of oil consumed. Electricity consumption is measured in BTU's as the heat content of actual electric sales. All fuel prices are then measured in dollars per effective BTU. Population is measured as the number of people per region; personal income in dollars per year in the corresponding regions. Areas are simply the number of square miles in each region. The maximum temperature is defined as the average temperature in the warmest 3 months of the year in the region of interest; minimum temperatures are the average temperature in the coolest 3 months of the year. The average fuel price used in equation (6.2) is defined as

$$\text{av. price} = \frac{\text{gas} * (\text{gas price}) + \text{oil} * (\text{oil price}) + \text{elect.} * (\text{elect. pr.})}{\text{gas} + \text{oil} + \text{electricity}} \qquad (6.2)$$

As in the preliminary model, the data used to estimate the model is a mixture of time series and cross sectional

data. For the revised model, we used data from 1967 to 1972, measured annually, in the 48 continental United States, plus separate data for Washington, D.C. (See Section 5.2 for a discussion of the interpretation of the use of cross sectional data).

For estimation of the above equations using GPSIE, the dynamics of the model were represented in the log linear form shown in equations (6.2) to (6.4). That is, the state vector of the dynamic system was defined as

$$\underline{x}(n) = \begin{bmatrix} \ln\left(\frac{\tilde{G}(n) + \tilde{O}(n) + \tilde{E}(n)}{\text{population}(n)}\right) \\ \ln\left(\tilde{G}(n)/\tilde{E}(n)\right) \\ \ln\left(\tilde{O}(n)/E(n)\right) \end{bmatrix} \tag{6.6}$$

The data used to estimate the model, however, was left in unaltered form. That is, the data vectors of measurement data $\underline{z}(n)$ were defined as

$$\underline{z}(n) = \begin{bmatrix} G(n) \\ O(n) \\ E(n) \end{bmatrix} \tag{6.7}$$

where the consumptions are measured in actual BTU's consumed, rather than the "effective BTU's" defined above. The formulation therefore required a somewhat involved measurement function

$$\underline{z}(n) = \underline{h}[\underline{x}(n), \underline{v}(n)]$$

to relate the state variables to the unprocessed measurement data. The measurement function is

$$gas = z_1 = \frac{pop.*e^{x_1}e^{x_2}}{1+e^{x_2}+e^{x_3}}$$

$$oil = z_2 = \frac{pop.*e^{x_1}e^{x_3}}{1+e^{x_2}+e^{x_3}} \quad .$$

$$elect. = z_3 = \frac{pop*e^{x_1}}{1+e^{x_2}+e^{x_3}}$$

A tempting alternative would be to process the data and use measurement data which correspond to the state vector. The more complex route was taken to allow direct comparison of the likelihoods computed for the revised model with any other model based on the same data. The log-linear form was maintained to allow the use of WLS estimation, but other models, via estimation with GPSIE, might be in a more natural form, with state variables which are more simply related to consumption measurements. One advantage of the maximum likelihood method is the ability to directly compare any two likelihoods computed on the same measurement data $\underline{z}(n)$, no matter how disparate the forms of the explanatory models or the nature of the exogenous inputs to those models.

6.2 First Estimation -- A False Start.

The revised model of Section 6.1 was completed with a simple model of equation errors and measurement errors, and estimated using GPSIE. The resulting estimates, while not so absurd as those for the preliminary model of Chapter 5, seemed to be of unreasonable magnitude, and several problems were encountered with the validity tests. This "false start" with the revised structure is briefly described in this section because it provides the motivation for several of the features and procedures described in Chapter 2 and applied in the remaining sections of this chapter.

The equation errors were formulated as in the preliminary model, as normal, white processes of zero mean and standard deviations which were fixed fractions of the initial conditions of the left hand variables. That is,

$$\sigma_{w_i} = b_i \, x_i(0)$$

where $\underline{x}(n)$ is the system state variable, $\underline{w}(n)$ are the equation errors, and $\underline{b}$ are the parameters which determine the proportionality of the errors.

The measurement errors were similarly hypothesized as additive normal white processes, of standard deviations proportional to the measurement data corresponding to the initial conditions in each region. That is, based on 1967 data. The equations were:

$$\sigma_{v_i} = c_i \, z_i(0) \quad [\underline{z}(0) = 1967 \text{ data}]$$

The resulting model contained 20 parameters to be estimated. Fourteen of the parameters describe the structure of the model dynamics; six parameters described the equation errors and measurement errors described above. The fourteen structural parameters were initially set to the values estimated using WLS [Baughman & Joskow,1975]; the error parameters were set based on prior judgement as follows. Standard deviations of each equation error was taken as 10% of the initial conditions of the corresponding state variable. The error in measuring gas consumption was taken as 2.5% of the 1967 consumption in each region, oil measurement error was taken as 5%, and electricity measurement error as 0.5% of 1967 consumption.

Maximum-likelihood estimates for the parameters were then computed, using a Powell search [Zangwill,1967]. Figure 6-1 compares the initial parameter values and the max like estimates, along with some of the key validity statistics. Although the search improved the likelihood by a factor of three, there are several suspicious features which deny the validity of the resulting model.

First, the SUMSQ statistic is 2/3 its expected size, nearly 30 standard deviations too small. The small size of SUMSQ is an indication that something may be wrong with the structure or the estimates of the stochastic parts of the

NO.	NAME	DEFINITION	WLS	FIMLOF
1	AT	additive const., total energy	-1.54	-1.86
2.	BT	coeff. on pers. income, total energy	2.89	2.22
3	CT	temperature coeff., total energy	-1.20	-1.21
4	DT	population density coeff., total energy	.937	1.73
5	ET	coeff. average price, total energy	-4.88	-14.4
6	FT	coeff. lagged term, total energy	.839	.783
7	A	additive constant, G/E	.070	.084
8	B	additive constant, O/E	.208	.127
9	C	price coefficient, both equations	-.137	-.067
10	D	max. temp. coeff., G/E	-1.50	-1.23
11	E	max. temp. coeff., O/E	-2.20	-2.20
12	F	min. temp. coeff., G/E	-2.20	-2.02
13	G	min. temp. coeff., O/E	-6.30	-5.58
14	H	lagged term coeff., both equations	.897	.923
15	SDFZ1	% std. dev., gas consumption data		.123
16	SDFZ2	% std. dev., oil consumption data		.284
17	SDFZ3	% std. dev., electricity cons. data		.0005
18	EQEF1	% std. error, total energy equation		.02
19	EQEF2	% std. error, G/E equation		.415
20	EQEF3	% std. error, O/E equation		.360

Figure 6-1

Preliminary Estimates for Revised Model

model (equation and measurement errors), since the computed covariance of the residuals is apparently larger than the sample variance of the residuals. Similarly, the residuals themselves may be, for some reason, too small. As we shall see in Sections 6.3 and 6.4, both effects were present in this case. The tests on the whiteness of the residuals cast still more doubt on the estimated model. The Durbin Watson statistics [Durbin,1955] for gas, oil, and electricity were 1.91, 0.63, and 0.46, indicating that the oil and electricity residuals were far from being a white process. In contrast, however, the normalized correlation matrices, as shown in Figure 6-2, seemed superficially to indicate acceptable properties-- 94% of the terms were within 4 standard deviations of their expected values, and the maximum error was 4.4 standard deviations. However, the largest errors were found on the main diagonal of P(0), and all were strongly negative. These results were all consistent with the hypothesis that the residual covariance matrix was too large, in spite of the excellent convergence of the Powell search at the estimated values shown in Figure 6-1.

A further reason to doubt the estimated results was found in the parameter estimates themselves. The estimate of the parameter ET in particular seemed too large, as it implied an elasticity of total energy demand with respect to price of 1.8, via the calculation shown in Figure 6-3.

DURBIN-WATSON STATISTICS:

1.91 0.63 0.46

WHITENESS TEST MATRICES:
(NO. OF STD. DEV. FROM EXPECTED VALUES)

P0:

-4.4	1.9	0.7
1.9	-4.0	0.4
0.7	0.4	-2.9

P1:

2.1	-0.1	0.1
-1.0	2.7	1.3
-1.1	1.2	3.8

P2:

-1.3	-1.5	-0.9
-1.3	1.6	1.6
-0.0	1.8	2.6

P3:

0.6	-0.6	0.2
-1.1	1.9	0.5
-0.8	0.7	2.6

Figure 6-2

Whiteness Tests for Preliminary Estimates

$$\text{elasticity} \triangleq \frac{\partial \text{energy}}{\partial \text{price}} \cdot \frac{\text{price}}{\text{energy}}$$

$$\text{energy} = \text{pop.} * e^{K} e^{\frac{ET * \text{price}}{1 - FT}},$$

$$\text{where} \quad K = \frac{AT + BT * \frac{\text{income}}{\text{pop}} + CT * (\text{min temp}) + DT \left(\frac{\text{pop}}{\text{area}}\right)}{1 - FT}$$

$$\frac{\partial \text{energy}}{\partial \text{price}} = \text{pop} * e^{K} * e^{\frac{ET * \text{price}}{1 - FT}} * \left(\frac{ET}{1 - FT}\right)$$

$$= \text{energy} * \left(\frac{ET}{1 - FT}\right)$$

$$\therefore \text{elasticity} = \frac{ET * \text{price}}{1 - FT}.$$

Figure 6-3

Calculation of Elasticity of Energy vs. Price

An additional trouble with the preliminary estimates was discovered in looking at scatter plots of fuel consumption vs. price. Several "outliers" were found (data points lying far outside the distribution determined by the majority of data points). Although such points constituted as small fraction of the total data, it was feared that the large apparent anomalies of these few points might be enough to distort the results.

The following three sections discuss the analysis of these various validity tests, and describe the improvements made in the model and data as a result.

6.3 Detection and Response To Bad Data.

The scatter plots mentioned in the last section precipitated the search for major errors in the data, but additional methods were required to efficiently locate the problem data. Scatter plots are inherently limited to two dimensions, and the equations of the model related several data series at once. Thus, three additional approaches were used to search for bad data points.

Farst, the normalized predicted measurement residuals (NPMR) for the three measurement variables were plotted, using GPSIE. These residuals, given correct data and an accurate model, should have a constant variance of one. All residuals of magnitude greater than 1.5 were marked as suspicious, and the data at the corresponding year and region were examined. (The unusually tight standard of 1.5

standard deviations was made necessary by the fact that the residuals seemed to be overnormalized by too large a covariance matrix). A glance at the plotted residuals also indicated that the variance was not constant across the regions (indicating cross sectional heteroskedasticity). This observation prompted the corrections discussed in Section 6.5.

Second, a similar analysis was performed, using the NUMR and NUSR residuals described in Section 2.6. These normalized updated residuals proved effective in locating several bad data points which were not apparent from the other methods used.

Third, all the data series were plotted and examined visually, to check for unusual time variation in a single series. This method is the least susceptable to automation, but it proved to be effective, both in analyzing data errors and in developing a feel for the system being modeled.

All suspicious data points found by the above methods were checked against their sources. The errors were found to fall into three categories:

1. Many errors were simple typographical mistakes, which were corrected to match the original data source.
2. Some apparent errors in the data for New England and the District of Columbia were due to inappropriate distribution of aggregate data. For

example, gas consumption prior to 1969 was reported only for the combined total of Maine, New Hampshire, and Vermont. In the data files, these totals had been arbitrarily divided equally among the three states. Similarly, some of the data for Washington D.C. and Maryland were the result of dividing aggregate data for the combined regions by two. These difficulties were remedied by re-aggregating the data into 46 regions by combining Maine, New Hampshire and Vermont, and by combining Maryland with the District of Columbia. Extensive variables were simply added; intensive variables, such as temperatures, varied so little among the adjoining regions that averaging them presented no problem.

3. Several remaining suspicious data points could not be attributed to any overt error, but nevertheless seemed unreasonable. For example, reported oil consumption in Georgia follows an orderly trend, except for the single year 1969, which seems abnormally low (see Figure 6-5). Such anomalies were especially numerous in the oil consumption data where the market share of oil was low. In cases where the discrepancy was especially large, the data point was assumed to be the result of some mistake in the reporting process, and was

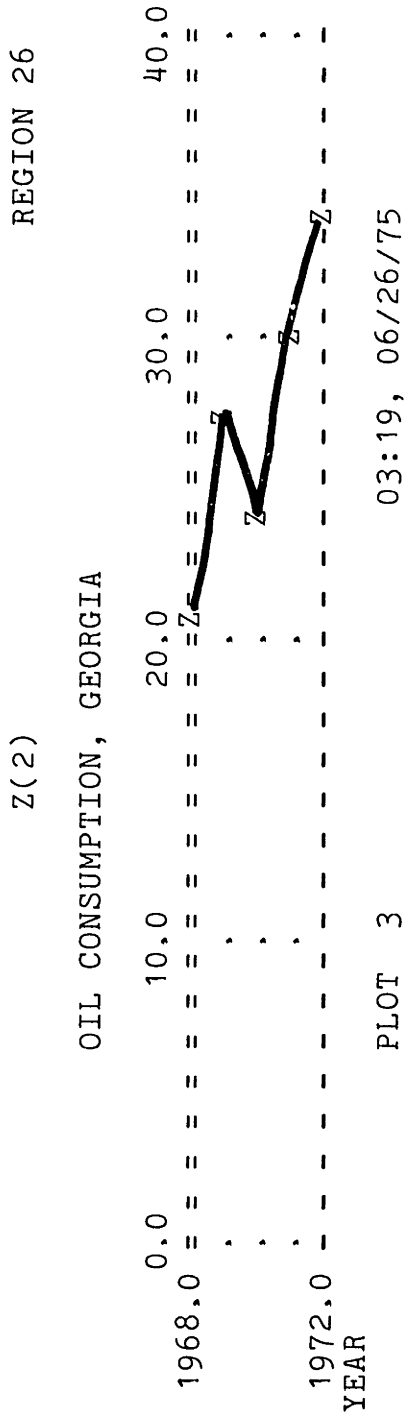

Figure 6-5

Reported Oil Consumption in Georgia

replaced with the NODATA code which signals GPSIE to skip the data point. One advantage of the FIMLOF method is that single components of the measurement vector $\underline{z}(n)$ may be deleted without sacrificing remaining components of the vector at that sample time. Six such data points were deleted. There remained, however, an apparently disproportional variation in reported gas and oil consumption in the small-consuming states. This variation led to a reformulation of the modeling of measurement error, as described in Section 6.5.

6.4 Initial Conditions.

The mathematics of the optimal filter used in FIMLOF assumes that the initial values of the state variables are random variables with known mean and covariance. In practice, however, the initial conditions are completely unknown, aside from the available data. The usual practice in engineering applications is to set the initial conditions to match closely the first piece of available data, and perhaps set the initial covariance large enough so as to express the modeler's "ignorance" of the true initial conditions.

Almost any scheme, however, which uses the first sample of output data $\underline{z}(1)$ to set the initial conditions will cause the residuals at the first sample to be zero. First, the

initial conditions must be fairly close to the true initial conditions, so that the linearization of the filter will not be about a trajectory far from the true state of the system. This requirement rules out, for example, setting the initial conditions to some arbitrary number unrelated to the system at hand, such as setting $\underline{x}(0)=\underline{0}$. Thus the residuals are likely to be somewhat small, since it is desirable to start the filter "on track". Second, one may "cheat" by, for example, initializing the filter at the first data sample, but performing an integration to generage a typical sized prediction error at the first sample. Thus, we set $\underline{x}(0)=\underline{z}(1)$, and then integrate to get $\hat{\underline{x}}(1|0)$ and $\hat{\underline{z}}(1|0)$, which will in general be different from $\underline{z}(1)$, because of the dynamics of the system. However, in this case, any estimation will tend toward a system that changes as little as possible in the first time step, unless the user expresses low confidence in the first state estimate by setting the initial covariance of the state to be large. But the large covariance of the initial state will cause the computation of a large covariance for the first predicted residuals $\_(1|0)$, and the resulting normalized residuals will again be close to zero. Thus, if the data used as $\underline{z}(1)$ for computing the first residual is also used in some reasonable way for setting the initial conditions of the optimal filter, the contribution of $\underline{z}(1)$ to the likelihood computation will be close to zero, and the first normalized

predicted measurement residual, rather than being a random vector with unit variance, will almost certainly be close to zero.

Now in a typical engineering application of the optimal filter, the above consideration will be of no practical concern, since the initial predicted residual vector will be only one out of, perhaps, several hundred or thousand residuals. Neither the likelihood computation nor the whiteness tests will be significantly affected. But in a situation such as the energy data discussed in this Chapter, we are using many short time series, and every fifth residual vector is zero! Thus the SUMSQ statistic described in Section 6.2 is inherently biased downward (although not enough to account entirely for its abnormally low value), and the whiteness and correlation statistics are distorted.

Therefore, the application of FIMLOF to social modeling requires either revised definitions of the various validity statistics to take into account the above problem with initial conditions under cross-sectional data or short time series, or some better way must be found to set the initial conditions of the filter. We have opted for the latter approach here, and recommend it strongly to others. The remainder of this section describes various ways of setting the filter initial conditions which avoid the creation of zero residuals.

6.4.1 INITIAL CONDITIONS, USING Z(0).

One method for overcoming the above difficulties is to set aside the first sample vector of measurement data for use only in setting the initial conditions of the filter. The first data vector is then called $\underline{z}(0)$, and is not used directly in the likelihood computations. That is, $\underline{z}(0)$ is used to determine $\hat{\underline{x}}(0|0)$, but residuals are computed, as before, only for $\underline{z}(1)...\underline{z}(N)$. The essence of the method is to use $\underline{z}(0)$ to convert a completely unknown $\underline{x}(0)$ into a Bayesian $\underline{x}(0)$ with an estimated mean and covariance. Assume, for example, that

$$\underline{z}(0) = \underline{H}\underline{x}(0) + \underline{v}(0),$$

$$\underline{x}(0) \text{ is completely unknown}$$

$$\underline{v}(0) \text{ is } N[0,\underline{R}]$$

where the relation may be exact, or may derive from linearization of the nonlinear standard form. If $\dim\{\underline{z}(0)\}\geq\dim\{\underline{x}(0)\}$ then we may compute

$$\hat{\underline{x}}(0|0) = [\underline{H}'\underline{R}^{-1}\underline{H}]^{-1}\underline{H}'\underline{R}^{-1}\underline{z}(0), \text{ and}$$

$$\underline{\Psi} = \text{cov}\{\hat{\underline{x}}(0|0)\} = [\underline{H}'\underline{R}^{-1}\underline{H}]^{-1}$$

The resulting initial conditions are not truly Bayesian, since we have only estimates of the mean and covariance of $\underline{x}(0)$, rather than *a priori* known mean and covariance. However, the result is clearly superior to the alternatives discussed above, and in practice is usually a sufficiently close approximation.

The method described above is the one used for the energy demand model of this Chapter. In this case, $\underline{H}=\underline{I}$, so the equations reduce to

$$\hat{\underline{x}}(0|0) = \underline{z}(0)$$

$$\underline{\Psi} = \underline{R}(0)$$

This method, however, breaks down if $\dim\{\underline{z}(0)\}<\dim\{\underline{x}(0)\}$, since $[\underline{H}'\underline{R}^{-1}\underline{H}]$ is then singular. The estimate also fails, of course if $\underline{H}'\underline{R}^{-1}\underline{H}$ is singular for any other reason.

6.4.2 INITIAL CONDITIONS, USING Z(1)...Z(N).

In case the method of Section 6.4.1 fails to yield estimates of the desired accuracy, Two additional methods are available which take advantage of all the available data. Instead of separating out the first sample vector as $\underline{z}(0)$, the first vector remains labeled $\underline{z}(1)$, and one of the following two approaches is used:

1. The most direct and accurate approach is to set $\underline{\Psi}=\underline{0}$ and include the components of $\underline{x}(0)$ as unknown parameters to be estimated.

2. A second approach that uses all the data to estimate the unknown initial conditions is to use the smoothing form of the optimal filter, such that $\hat{\underline{x}}(0|N)$ is computed in each pass through the data [Schweppe,1973], [Jazwinski,1970].

6.5 Modeling Measurement Error.

In searching for bad data points, it was observed that the data for gas consumption and oil consumption seemed to be excessively noisy in those states where the consumption of the fuel in question was low. Thus, for example, in the midwest and southwest, where fuel oil is seldom used, the oil data seemed to be particularly unreliable. In the WLS estimates of the model, it was found that the use of OLS (unweighted data) produced residuals of noticeable heteroskedasticity. The situation was improved noticeably by weighting all the consumption data by the square root of the sum of oil and gas consumption. Thus,

$$G'(n) = G(n)*SQRT[G(n)+O(n)]$$

$$O'(n) = O(n)*SQRT[G(n)+O(n)]$$

$$E'(n) = E(n)*SQRT[G(n)+O(n)]$$

This adjustment of the data, used in the WLS estimates, was based on the assumption that the variance in the consumption data was due to the variance of a multinomial decision process [Theil,1967], [Schweppe,1974], in which the fuel prices and other exogenous inputs determine the probabilities of the multinomial process, rather than the decision being modeled as deterministic. A similar argument led to a change in the modeling of the measuement error in the context of FIMLOF. For the estimation with GPSIE, we observed that the large variance in reported consumption in the small consuming states was unlikely to be due to

multinomial variance in the appliance choice decisions. To test this hypothesis, the theoretical variance of the data was derived for a region with a small number of decisions. The smallest number of decisions for which we could find data was the installation of oil furnaces in Kansas. In 1968, there were 12,350 oil furnaces in use in the residential sector. There were 487 decisions to install oil burners, including new homes (40), replacements (312), and conversions (135). The total market share of oil in the residential sector in Kansas is 4%. A reasonable assumption to make (since this computation is only to determine the rough order of magnitude of the binomial variance) is that the fraction of all decisions made for furnaces in 1968 was also 4% for oil. This assumption is equivalent to asserting that the market is roughly in equilibrium. Then the total number of decisions would be 487/.04, or about 12,000 multinomial trials, 4% of which went to oil. Under a multinomial model of fuel choice, the variance of the number of decisions for oil is proportional to the total number of trials:

$$\text{var}\{\text{no. of oil decisions}\} = 12000 * f * (1-f)$$

where f is the probability of choosing oil, which is (roughly) 4%. This gives a variance in the number of decisions for oil of 480, for a standard deviation of 22. Since the reported consumption of oil depends on the total number of furnaces in use, rather on just the new and

replaced furnaces, this standard deviation should be normalized by the 12,350 furnaces in use in Kansas in 1968, yielding a standard error percentage of 0.2%! Thus, we conclude that the large apparent variation in reported oil consumption is probably not due to the variance associated with the multinomial process model of fuel choice.

However, a good argument can be made that the variation in reported oil consumption can be attributed to errors in the data collection process. The oil consumption data, for example, is based on the collection of forms by the Bureau of Mines, as published in the BOM Mineral Industry Surveys. The submission of these forms is in itself known to be a somewhat random process. Forms may be lost or delayed; consumption of an industrial firm may be reported as commercial consumption, etc. The forms are collected not from end users, but from producers and distributers, usually at a rather aggregate level. Thus, if one models the submission of forms to the Bureau of Mines as a random process goverened by a binomial decision (unconscious?) of whether to submit an accurate form in a given period. The number of trials each period is much smaller for this decision, and could yield significant variance in the reported consumption figures, even if the probability of submitting an accurate form is high. Thus, it is a reasonable hypothesis to model the measurement error in gas and oil consumption as having a variance proportional to the

size of consumption. This model is compared with the two alternatives in Figure 6-6. The vagaries of reporting consumption were assumed to be much less for electricity, where the standard deviation was maintained as proportional to consumption, as before.

The constants of proportionality between the variance and the consumption figures were included as unknown parameters to be estimated by GPSIE. Initial guesses of the parameters were computed by visual estimates of the sample standard deviations of several repersentative graphs of the consumption data.

6.6 Filter Convergence.

Although experiments with simulation data indicated that there were no numerical problems with the filter and that the model structure was identifiable, an additional numerical test was performed to make sure the filter was behaving properly. The motivation for the test was the observation that the measurement function of the model which relates consumption data to the state variables was highly nonlinear. Thus, the estimated state about which the measurement function is linearized for the Riccatti equation is important -- a small error in the estimated state might lead to a significant error in the linearization matrices. To test for the presence of linearization error, a local iteration in the filtering equations was introduced, in

WLS MODEL: ** MULTINOMIAL MODEL USED TO MODEL EQUATION ERROR VARIANCE AS PROPORTIONAL TO CONSUMPTION; MEASUREMENT ERROR ASSUMED ZERO.

PRELIMINARY FIMLOF MODEL: ** FOR BOTH EQUATION ERROR AND MEASUREMENT ERROR, STD. DEVIATION TAKEN AS PROPORTIONAL TO CONSUMPTION.

FINAL FIMLOF MODEL: ** VARIANCE OF MEASUREMENT ERROR TAKEN AS PROPORTIONAL TO CONSUMPTION; STD. DEVIATION OF EQUATION ERROR TAKEN AS PROPORTIONAL TO CONSUMPTION.

Figure 6-6

Models of Variation in Gas and Oil Consumption

order to insure that the best estimate of the system state was used as the linearization point. The iteration is due to Breakwell and is taken from [Jazwinski,1970]. The optimal filter with the local iteration is sometimes called the "Iterated Extended Kalman Filter." To test for the possibility of errors in the likelihood computation due to measurement nonlinearity, the likelihoods computed with and without the iteration were compared for several values of the unknown parameters. The maximum change in the log likelihood due to the introduction of the local iteration was less than two tenths of one percent, indicating that the measurement nonlinearity was not causing difficulty.

6.7 Final Estimation Results.

The changes and improvements in the model and data described in Sections 6.3 - 6.6 were implemented. The model was linked with GPSIE, and the log likelihood of the parameters from the WLS estimation was evaluated. The log likelihood of the WLS estimates was -2,543, which indicates a factor of two improvement over the log likelihood of the same parameters under the conditions of Section 6.2. The differences between the two situations are 1) a new model of the measurement error of gas and oil, 2) corrected data files, and 3) an improved way of setting initial conditions in the filter, using 1967 consumption data. The SUMSQ statistic at the initial parameters, however, was 1,541,

which is 23 standard deviations greater than the expected value of SUMSQ=670.

To locate the maximum-likelihood parameter values, a Powell search was initiated, starting from the WLS parameter estimates. The search converged at the parameter values shown in Figure 6-9. The log likelihood of the parameter estimates was -1,155, a major improvement over the initial parameter values obtained by WLS. Furthermore, the SUMSQ statistic at the max like point was 680, quite close to the expected value of 670 (the error of 10 is less than one third of a standard deviation).

The whiteness statistics shown in Figure 6-10 indicate significant improvement over the corresponding statistics from the estimates of Section 6.2, before the improvements in model and data. The main diagonal terms of the correlation matrix P(0) indicate that the normalized residual series have sample variances close to the expected value of one (indicating that the SUMSQ statistic is good not only in the aggregate, but also on a component basis). However, the relatively large main diagonal terms of the remaining normalized matrices indicate considerable serial correlation of the individual residual series. The Durbin-Watson statistics confirm the presence of serial corelation in each of the fuels. For a single time series, these statistics would lead one to reject the present model and look for missing lag effects operating over about 5 time

No.	Name	Definition	WLS	FIMLOF
1	AT	additive constant, total energy	-1.54	-1.54
2	BT	coeff. on pers. income, total energy	2.89	2.81
3	CT	temp. coeff., total energy	-1.20	-1.62
4	DT	coeff. population density, total en.	.937	.113
5	ET	price coeff., total energy eq.	-4.88	-4.16
6	FT	coeff. of lagged term, total energy	.839	.839
7	A	additive constant, G/E equation	.070	.038
8	B	additive constant, O/E equation	.208	.038
9	C	price coeff., both fuel split eq.	-.137	-.029
10	D	max. temp. coeff., G/E equation	-1.50	-1.23
11	E	max. temp. coeff., O/E equation	-2.20	-1.47
12	F	min. temp. coeff., G/E equation	-2.20	-1.71
13	G	min. temp. o		
13	G	min. temp. coeff., O/E equation	-6.30	-5.71
14	H	coeff. of lagged term, both fuel eq.	.897	.965
15	VARFZ1	var. of meas. error, G	--	.45
16	VARFZ2	var. of meas. error, O	--	.86
17	SDFZ3	std. dev. meas. error, E	--	.003
18	EQEF1	fraction std. dev., total energy eq.	--	.01
19	EQEF2	fraction std. dev., G/E equation	--	.06
20	EQEF3	fraction std. dev., O/E equation	--	.06
		log likelihood	-2543	-1155
		SUMSQ (exp. = 670)	1541	680

Figure 6-9

Parameter Estimates, Final Model

R(0):	.92	.01	-.15
	.01	1.1	0
	-.15	0	.96
R(1):	.32	-.03	-.16
	-.07	.53	-.05
	-.04	.15	.48
R(2):	.24	-.04	-.17
	-.04	.41	.09
	-.02	.15	.34
R(3):	.16	-.09	.12
	.05	.24	.07
	.05	.21	.29

Correlation Matrices R(j)

P(0)	-.9	.1	-2.2
	.1	1.1	0
	-2.2	0	-.4
P(1)	4.8	-.5	-2.5
	-1.	8.1	-.8
	-.6	2.3	7.3
P(2)	3.7	-.6	-2.6
	-.6	6.2	1.4
	-.3	2.3	5.1
P(3)	2.4	-1.4	-1.8
	.8	3.7	1.0
	-.7	3.2	4.4

Normalized Correlation Matrices P(j) (No. of Std. Deviations error in R(j))

DURBIN-WATSON Statistics:

Gas: 1.05 Oil: 0.71 Electricity: .88

Figure 6-10

Whiteness Statistics for Final Model

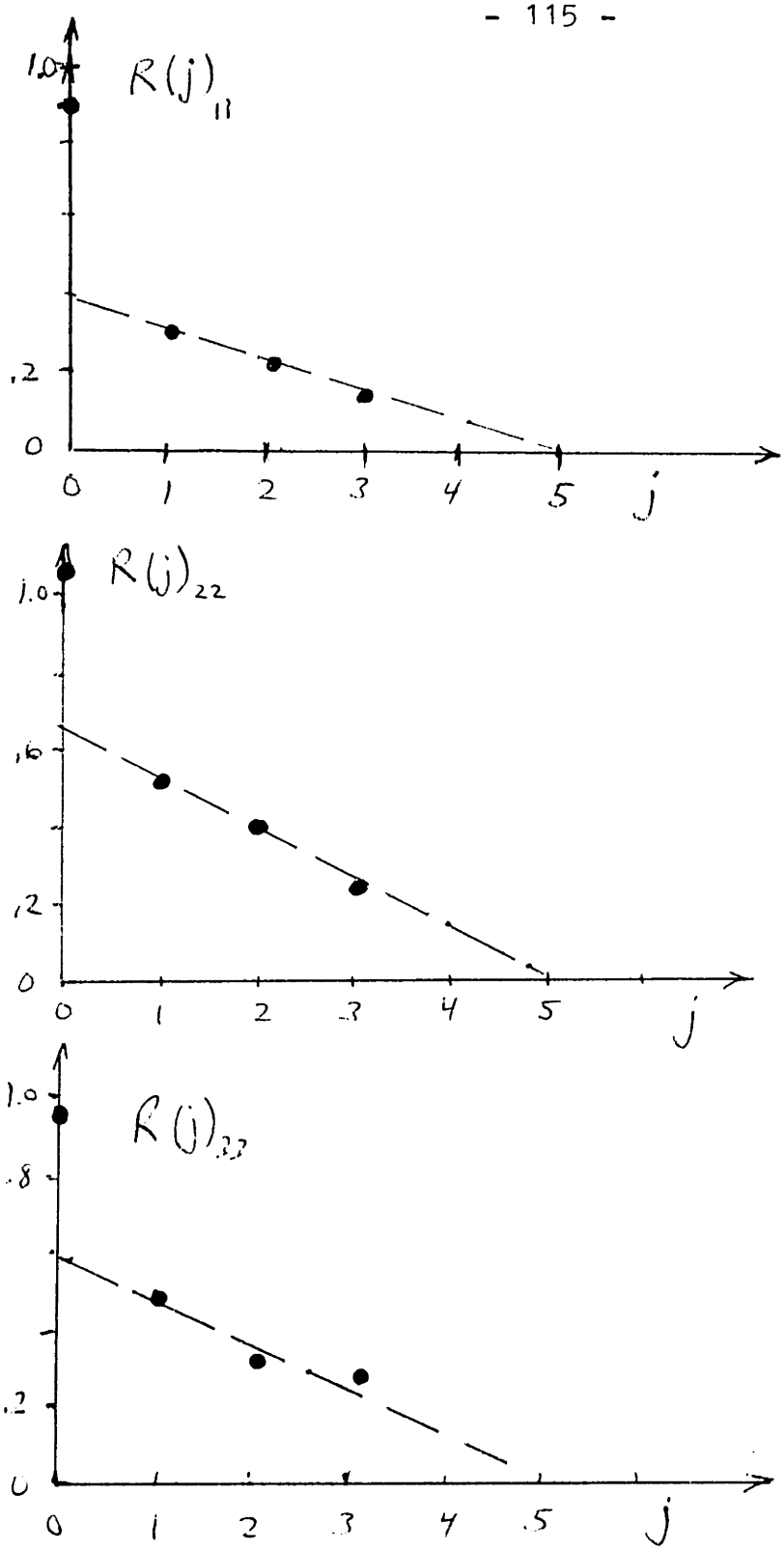

Figure 6-11

Spectral Analysis of the Residuals

steps (see Figure 6-11). However, in this case, an alternative explanation is obvious -- each 5 time steps, the sequence of residuals enters a new region, and acquires a new bias, unrelated to the bias of the region preceding. For zero time lag, one gets essentially the sample variance of the residual component series. Let

$$r(n) = b(n) + d(n)$$

where r(n) is the residual (scalar component) at sample n, b(n) is the bias of the region containing r(n), and d(n) is the "delta" component which varies within the region. The bias component b(n) is assumed to be uncorrelated across regions, but constant within regions; thereore,

$$E\{b(n_j)b(n_j-k)\} = \begin{cases} E\{b(n)b(n)\}, \text{ for } j<k \\ 0, \text{ otherwise} \end{cases}$$

where n_j is the j_th sample in a given region. Now by definition, the correlation matrix main diagonal component of R(m) is

$$r(n)r(n-m)$$

but the expectation of the above term is

$$E\{r(n)r(n-k)\} = E\{b(n)b(n-k)\} + E\{d(n)d(n-k)\}$$

since E{d(n)}=0. Now since the data in question has 5 samples per region, it is clear that

$$E\{r(n)r(n-k)\} = \begin{cases} 1, & k=0 \\ (4/5)E\{b(n)b(n)\}, & k=1 \\ (3/5)E\{b(n)b(n)\}, & k=2 \\ (2/5)E\{b(n)b(n)\}, & k=3 \\ (1/5)E\{b(n)b(n)\}, & k=4 \\ 0, & k>4 \end{cases}$$

For the gas consumption residuals, it appears that the sample variance of b(n) is .4,; for oil, .67, and for electricity, .6 (these numbers were computed as 5/4 times the 1-step lagged correlation). Thus, roughly half of the residual variance is due to cross sectional errors, and half due to time series errors. In both cases, the error is a combination of both equation error and measurement error.

In conclusion, the residuals exhibit serial correlation, but all the serial correlation can be explained by the cross sectional nature of the data. Thus, the model can be considered to pass the whiteness of residuals test; the excellent SUMSQ statistic and variances of individual residual components combine to lend great support to the idea that the model is consistent with the data.

Chapter 7

CONCLUSIONS

In this thesis, a variety of techniques for the estimation of parameters and the testing of consistency of dynamic models with respect to time series data are developed. Adopted from their engineering origins and developed for application to social and econometric models, the techniques are referred to here under the generic label, Full-Information Maximum Likelihood via Optimal Filtering (FIMLOF). With suitable modifications and additions, it has been found that FIMLOF estimation and validation techniques work well in the realm of social modeling, despite the differences in purpose, model structure, and data between social and engineering systems. The FIMLOF techniques extend to a wide class of nonlinear, dynamic, stochastic model forms the application of Fisher's principle of maximum likelihood:

> Everything else being equal, we choose the system which gives the highest chance to the facts observed. [1]

In particular, the experimental and theoretical developments leading to this thesis demonstrate that the important tasks of parameter estimation and validation can be performed under conditions of

- Nonlinearity of system dynamics and measurement functions
- Unmeasured variables and mixed sampling intervals
- Highly corrupted data, including "errors in variables"
- Cross-sectional data
- Short time sequences of data

The FIMLOF techniques have been implemented in a user-oriented computer program, the General Purpose System Identifier and Evaluator (GPSIE). The GPSIE software enables the modeler to concentrate on model construction and evaluation, without getting bogged down in the writing of special purpose codes for filtering, likelihood evaluation, etc. In addition to making available some new techniques, GPSIE may make it easier for engineers and social scientists to use and explore parts of the FIMLOF complex which, while not new, are still relatively unknown and unavialable without great investment in study and programming. GPSIE is a step toward making the powerful tools commonplace.

In additior to extending maximum likelihood methods to a new range of model structures, the FIMLOF techniques may provide a useful check on other results, when applied to special-case model structures (such as OLS applied to linear-in-the-parameters models).

The FIMLOF methods allow the extension of standard validity tests to situations where the tests were not previously applicable; for example, Durbin-Watson statistics may be computed for time-varying models and for cross-sectional data. Additional tests are also made possible, as well as new ways of interpreting the old tests.

The residual process of the optimal filter has been shown to lead to some powerful methods of bad data detection and identification. The Normalized Updated Measurement Residuals (NUMR) and Normalized Updated State Residuals (NUSR), in addition, show promise of being useful for automated elimination of bad data, and robust estimation.

Although not explicitly discussed in the thesis, it should also

be pointed out that the FIMLOF and GPSIE techniques may also contribute to social modeling by making routine the computation of statistical confidence bounds on forecasts made with dynamic models. The confidence bounds may serve not only as a further test of model validity (in evaluating accuracy of forecasts), but also may help in pinpointing areas most productive for the interpretation or improvement of a given model.

Perhaps the most important contribution of FIMLOF and GPSIE to social modeling is the degree to which the methods reduce the constraints on model formulation often associated with parameter estimation and confidence tests. It should be emphasized that the methods presented here, while powerful, should not be taken as automatic, sure-fire tests of model validity. FIMLOF is "full information" only with respect to numerical data, taken in the context of the specific model structure chosen by the model builder. It is the obligation of the model builder to create a model structure consistent not only with numerical data, but also with prior knowledge, accepted theory, and internal consistency. No matter how sophisticated the numerical methods, the most likely model, by numerical standards alone, is the "black box" model with so many parameters that it can be tuned to duplicate all the numerical time series. The FIMLOF techniques extend and complement, but do not supplant, the application of logic and experience. It is hoped that social modeling may become more and more dominated by logical consistency and realism, as the constraints due to the mathematics of estimation and validation are lessened.

BIBLIOGRAPHY

1. Anderson, Kent, Residential Demand for Electricity: Econometric Estimates for California and the United States, Rand Corporation, January 1972.

2. Astrom, K.J., "On the Achievable Accuracy in Identification Problems," IFAC 1967.

3. Astrom, K.J. and S. Wenmark, "Numerical Identification of Stationary Time Series," Sixth International Instruments and Measurements Congress, Stockholm, September 1964.

4. Balestra, P., The Demand for Natural Gas in the United States, Amsterdam, North Holland, 1967

5. Baughman, M.L., Dynamic Energy System Modeling -- Interfuel Competition, Ph. D. Thesis, M.I.T. Dept. of Electrical Engineering, 1972.

6. Baughman, M.L.,"The Effects of Fuel Prices On Residential Appliance Choice In the United States," Working paper, M.I.T. Energy Lab, 1975.

7. Baughman, M.L., P.L. Joskow, and F.S. Zerhoot, "Interfuel Substitution in the Consumption of Energy in the United States," Part I: Residential and Commercial Sector, Draft, May 10, 1974, M.I.T. Energy Laboratory.

8. Brunner, R.D., and G.D. Brewer, Organized Complexity, New York, Free Press, 1971.

9. Chow, Gregory, "Tests of the Equality Between Two Sets of Coefficients in Two Linear Regressions," Econometrica, Vol 28, 1960.

10. DeRusso, P.M., and R.J. Roy, State Variables For Engineers, John Wiley & Sons, 1965.

11. Dhrymes, P.J. et. al., "Criteria For Evaluation of Econometric Models," Annals of Economic and Social Measurement, 1/3, 1972.

12. Fisher, F.M., "Causation and Specification in Economic Theory and Econometrics," Synthese 20, D. Reidel Co., Dordrecht, Holland, 1969.

13. Fisher, F. and C. Kaysen, A Study in Econometrics: The Demand for Electricity in the United States, Amsterdam, North Holland, 1962.

14. Fisher, R.A., "On the Mathematical Foundations of Theoretical Statistics," Phil. Trans. Roy. Soc. London, 222, 1922, p. 309.

15. Fletcher, R. and M.J.D. Powell, "A rapidly convergant Descent Method for Minimization," The Computer Journal, Vol. 6, 1967

16. Forrester, J.W., Industrial Dynamics, M.I.T. PRess, Cambridge, Mass. 1961.

17. Forrester, J.W., "Market Growth as Influenced by Capital Investment, Industrial Management Review, Vol. 9, No. 2, Winter, 1968.

18. Edwards, A.W.F., Likelihood, Cambridge University Press, 1972.

19. Houthakker, H., and Taylor, L., Consumer Demand in the United States, 1929-1970, Cambridge, Harvard Universtiy Press, 1972.

20. Jazwinski, A.H., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

21. Kashyap, R.L., "Maximum Likelihood Identification of Stochastic Linear Dynamic Systems," IEEE Trans. Automatic Control, February, 1970.

22. Maddala, G.S. and A.S. Rao, "Tests for Serial Correlation in Regression Models with Lagged Dependent Variables and Serially Correlated Errors," Econometrica, Vol. 41, No. 4, 1973.

23. McAvoy, P. and R. Pindyck, "Alternative Regulatory Policies for Dealing With the Natrual Gas Shortage," The Bell Journal of Economics and Management Science, Autumn, 1973.

24. Mehra, R.K., "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation," AIAA Journal, Vol. 9, no. 1, January 1971.

25. Mehra, R.K., "An Innovations Approach to Fault Detection and Diagnosis in Dynamic Systems," Automatica, Vol. 7, 1971.

26. Mehra, R.K., "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Transactions on Automatic Control, Vol. AC-15, No. 2, April 1970.

27. Mehra, R.K., "Identification in Control and Econometrics; Similarities and Differences," Annals of Economic and Social Measurement, Vol. 3, No. 21, January, 1974.

28. Mehra, R.K., and J.S. Tyler, "Case Studies in Aircraft Parameter Identification," from Identificaton and System Parameter Estimation (ed. P. Eykhoff), American Elsevier Publishing Co., New York, 1973.

29. Minsky, Marvin L., Computation: Finite and Infinite Machines, Prentice-Hall, 1967.

30. Moore, R. and F.C. Schweppe, "Adaptive Control for Nuclear Power Plant Load Changes," Fifth World Congress of IFAC, 1972.

31. Naylor, T.H., et. al., "Spectral Analysis of Data Generated by Simulation Experiments with Econometric Models," Econometrica, April 1969.

32. Peterson, D.W., "Proposed Stochastic Models of Energy Demand: A Brief Discussion," M.I.T. Energy Lab internal memo, February 26, 1974.

33. Peterson, D.W., and F.C. Schweppe, "Code for a General Purpose System Identifier and Evaluator (GPSIE)," IEEE Trans. Auto. Control, December, 1974.

34. Price, C. "An Analysis of Divergence Problems in the Kalman Filter," IEEE Trans. Auto. Control, AC-13, December, 1968.

35. Schweppe, F.C., Uncertain Dynamic Systems, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

36. Schweppe, F.C., "Evaluation of Likelihood Functions for Gaussian Signals," IEEE TRans. Inform. Theory, IT-11, No.1, January 1965.

37. Schweppe, F.C., "Random Comments and Associated Mathematical Gymnastics Related to Fuel Share Modeling," M.I.T. Energy Lab internal memo, February 19, 1974.

38. Schweppe, F.C., "Chapter IV in the Continuing Drama of Probabilistic Fuel-Energy Modeling," M.I.T. Energy Lab internal memo, March 4, 1974.

39. Schweppe, F.C., "Algorithms for Estimating a Re-Entry Body's Position, Velocity, and Ballistic Coefficient in Real Time or for Post Flight Analysis," 1964-4 Lincoln Lab, Lexington, Mass., M.I.T., January, 1965.

40. Schweppe, F.C., "State Space Evaluation of the Battacharyya Distance Between Two Gaussian Processes," Inform. Control, Vol 11, No. 3, September 1967.

41. Senge, P.M., "Evaluating the Validity of Econometric Methods for Estimation and Testing of Dynamic Systems," System Dynamics Group Memo D-1944-2, Alfred P. Sloan School of Management, M.I.T., February 12, 1974. (Draft working paper)

42. Senge, P.M., "An Experimental Evaluation of Generalized Least Squares Estimation," System Dynamics Group Memo D-1944-6, Alfred P. Sloan School of Management, M.I.T, November 14, 1974.

43. Stanford Research Institute, "Patterns of Energy Consumption in the United States," January 1972.

44. Theil, H., Statistical Decomposition Analysis, Amsterdam, North Holland, 1972.

45. Theil, H., Principles of Econometrics, New York, John Wiley & Sons, 1971.

46. Turchin, V.F., et. al, "The Use of Mathematical-Statistics Methods in the Solution of Incorrectly Posed Problems," Uspekhi Fizicheskikh Nauk (Soviet Physics Uspekhi), Vol 102, November, 1970.

47. Wonnacott, R.J. and T.H. Wonnacott, Econometrics, John Wiley & Sons, New York, 1970.

48. Wilks, S., Mathematical Statistics, New York, John Wiley & Sons, 1962.

49. Zangwill, W.I., "Minimizing A Function Without Calculating Derivatives," TheComputer Journal, Vol. 10, 1967.

50. Handschin, E. et. al., "Bad Data Analysis for Power System State Estimation," Electric Power Systems Engineering Laboratory Report, 1974, M.I.T., Cambridge, Massachusetts.

51. Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems," Journal Basic Eng., Ser. D, Vol. 82, March 1960.

52. Stravinsky, I., Poetics of Music, Vintage Books, Random House, N.Y. 1956.

Appendix A

GPSIE

GPSIE (General Purpose System Identifier and Evaluator) is a precompiled program which consists of 1) matrix routines for computing points of the log likelihood function, via optimal filtering (extended Kalman filtering, with various iterations and other options special cases and difficult problems), 2) several nonlinear-programming algorithms for maximizing the likelihood function over the space of unknown parameters, 3) auxilliary computation algorithms for confidence bounds on estimates and forecasts, robust estimation, error detection, etc., and 4) control logic for handling input, output, simulation, options, and error recovery.

For each model, the user writes a subprogram which describes the model of interest, its dimensions, and its linearization. The user subprogram (for the first version of GPSIE) must be written in PL/I, but subsquent versions are planned which will be FORTRAN-based. The user subprogram is compiled, linked with the precompiled GPSIE, and loaded. GPSIE then accesses the user subprogram to learn the dimensions of the model and data, the initial guesses of any unknown parameters, and user options.

A.1 Features and Options.

SEARCH OPTIONS. Numerical maximization of the likelihood function often requires versatility of approach. GPSIE includes as options

the following search algorithms: Newton-Raphson, Gauss-Newton, Davidon-Fletcher-Powell, and Powell (Zangwill modification), as well as a manual search, in which the user specifies the sequence of parameter values whose likelihood is to be evaluated.

LEAST-SQUARES INITIALIZATION. It is often helpful to first approximate the maximum-likelihood solution via a least-squares approach, which is computationally cheaper. GPSIE includes least-squares capaabilities as options.

STEADY-STATE FILTER. It is sometimes desirable to assume the optimal filter is in steady state after a given sample; either because it is in steady-state, or to simplify computation early in a search (much as in the least-squares initializations). GPSIE can compute the steady-state filter gains "off-line", or may be set to assume steady state during specified portions of the data. In addition, for time-invariant linear systems, the filter gain and residual covariance matrix may be estimated directly as unknown parameters.

CROSS-SECTIONAL or REGIONAL DATA. Data, especially in social systems, is often available from several systems operating in parallel and sharing the same unknown parameters. GPSIE can process such data conveniantly, through an option for dealing with mixed time-series cross sectional data.

SIMULATION CAPABILITY. GPSIE includes noise generators and control logic for simulating the model of interest and storing the results. Deterministic simulation, with computation of confidence bounds, may also be done for forecasting and error analysis.

COMPUTATION OF DERIVATIVES. GPSIE can compute both gradietn and HEssian of the log-likelihbod function (with respect to unknown parameters) by finite-differences, in order to handle the most general case. An option is included for a statistical approximation to the Hessian, which is computationally more efficient, but which works well only under relatively benign conditions.

MISSING DATA and VARIABLE SAMPLING INTERVALS. GPSIE contains featues for the convenient labeling of data points as "NODATA", to be skipped over. This feature is not only useful for missing data, robust estimation, and error detection, but also allows convenient use of data of mixed sampling intervals, or of sampling intervals different from the integration time-step of the model.

A PRIORI INFORMATION ON PARAMETERS. Parameters may be taken as completely unknown; unknown, with a priori mean and variance; or a priori known; or any combination thereof.

A.2 Major Limitations. In spite of its generality, GPSIE has two major limitations, inherent in the use of the extended Kalman filter:

1) In nonlinear systems, excessive noise in the data or initial conditions, or a poorly observable system structure, may allow the estimated state of the system to drift too far from the true state. Under such conditions, the linearization of the filter may become invalid. GPSIE includes local iteration options to test for and compensate for divergence of the filter, but extreme forms may invalidate the computations.

2) The variable-dimensioning feature of GPSIE allows it to handle systems of any reasonable size, but requirements of computer time or storage may obviously become extravagant for some systems. For example, estimation of 10 - 20 parameters for a system with 5 - 10 state variables may be expected to cost about $100. However, the costs are significant only when repeated passes through the data are required to search for maximum-likelihood parameters. Even relatively large systems may be treated as a priori known (no estimation of parameters) and tested for consistency with data, forecast with confidence bounds, etc. Large systems may also be treated suboptimally by splitting them into sectors of tractable size. This method may sacrifice little if high-quality data is available for the variables which are common to the sectors (the "linking" variables).

Appendix B

INFORMAL TESTS OF MODEL VALIDITY

> ...Human activity must impose limits upon itself. The more art is controlled, limited, worked over, the more it is free.... Whatever diminishes constraint diminishes strength. The more constraints one imposes, the more one frees one's self of the chains that shackle the spirit. 52 Stravinsky, 1956

Mathematical model building operates under two distinct kinds of constraints: 1) Constraints due to axioms of estimation mathematics, computer size, costs, etc., and 2) Constraints of logic, consistency, "scientific taste," etc. The former constraints are generally considered nuisances, and much work (including this thesis) is devoted to reducing their influence. As the efforts to reduce the first kind of constraints progress, the model builder is constrained instead by the second kind. The second kind of constraints, however, are sometimes no constraints at all. They are ill-defined, and every modeler is privileged to a unique set of them.

The purpose of this appendix is to offer an example list of informal tests of model validity, to illustrate what is meant by the second kind of constraint, discussed above and in Chapter 4. The tests range from trivial to sublime and grandiose. It is not implied that the author's models pass all these tests.

The discussion of the tests is couched in terms of models of social systems, but the generalization to other kinds of modeling is

clear.

B.1 Pragmatic Tests.

Models, it may be asserted, should serve a useful purpose. Hence, there are some pragmatic tests which ask whether the model achieves that purpose. These "pragmatic" tests can usually be directly applied only after the model has been used (implemented), and even then evaluation is uncertain, to say the least. However, thinking about these tests during the process of making the model helps one maintain an attitude toward purposeful, high-quality work.

TEST 1: Does the modeling activity lead to improved real-world systems (the ultimate pragmatic test).

TEST 2: Does the model correctly predict the direction of change resulting from policy changes?

TEST 3: Does the model correctly predict the extent of change resulting from policy changes?

(Note that tests 2 and 3 can be applied in two ways during the process of model building and implementation:

1. The model should be consistent with any policy changes that have occurred historically in the system under consideration.

2. Policy changes suggested by the model may be implemented on a trial basis, in limited (controlled?) experiments.

B.2 Tests of Purpose.

A model's validity can be judged only with respect to a purpose. Many of the tests listed here can be applied only in the context of a clear and explicit purpose.

TEST 4: Does the model have a clear purpose? Specifically, what are the symptoms (variables) that the model is to be used to explain or modify? What experiments or policies do we want the model to predict or explain the results of? The test of purpose is a major constraint on the scope of a model. A clear purpose may provide a priori criteria for excluding various ideas and variables from the model; with no purpose to the model, any observed feature of reality is fair game for inclusion.

TEST 5: The model should be formulated from a well defined perspective. For example, a model of a firm from the perspective of an investment banker will be different from a model of the same firm from the perspective of a production manager. This test provides a constraint on the level of aggregation. Just as a model without a well-defined purpose (Test 4) is apt to become too "wide" (dealing with too many things), a model without a well-defined perspective is likely to become too "deep" (going into too much detail).

TEST 6: Is the model addressed to important questions and problems? This test, like Test 1, is not usually associated with the notion of validity. It instead represents an attitude, or a judgement of "scientific taste." However, it is a test often applied (especially to other people's models).

B.3. Tests of Variables and Structure.

Mechanisms in the model should correspond with "real" mechanisms.

TEST 7: Every variable and parameter should have a precise definition relating to things in the "real world." This definition

is intended to eliminate nameless parameters and other hard-to-evaluate constructs from the model. The ideal is to have a transparent and communicable structure, rather than a "black box" approach. (however, in some applications, and for some purposes, the black box approach may be best -- another instance of the subjectivity and inconsistency of informal tests).

TEST 8: Look at the constants in the model: are they really constant for the purposes of this model? No real-world parameter is "really" constant. Making it so in the model is an assertion that it changes slowly, compared with the dynamic behavior of the model; or that changes in the parameter will not affect the conclusions to be drawn from the study.

TEST 9: Consider the exogenous variables in the model, if any. Are they, in fact, independent of all the other variables in the model? Imagine that some or all of the endogenous variables are forced to zero; is the exogenous variable still plausible under these circumstances? Similarly, imagine system levels to take on huge values.

TEST 10: For each relationship, what is the corresponding real-world mechanism? Seek to witness or sample the actual operation of the system being modeled; avoid relying on verbal generalizations only.

TEST 11: Look at critical points in functions and relationships; points of steepest slope, inflection points, asymptotes, dips, abrupt changes in slope or curvature should all have physical meanings and plausible justifications.

B.4 Tests of Sub-Sector Behavior.

By sub-sector, we mean individual equations or groups of equations which are closely connected in the model. Each sub-sector or equation may be thought of as having inputs and outputs, which may not be the case for a complete model with no exogenous inputs. However, the same tests may be applied in the latter case, if parameters are thought of as "inputs" (see Section 4.1), and endogenous variables as "outputs."

TEST 12: Imagine or simulate changes in the input variables, especially sudden or large changes. The magnitude, transient character, asymptotic behavior, and time phasing of the resulting changes in the outputs should make sense and should be consistent with past observations and time-series data, where appropriate.

TEST 13: The behavior of equations and sub-sectors should be defined and should make sense for extreme values of input variables. Are the outputs reasonable when the inputs are zero? When the inputs are abnormally large? The behavior in the extremes, coupled with an assumption or observation of smoothness, may be a useful constraint on the actual, mid-range operating region.

B.5 Tests of Model Behavior.

TEST 14: The simulation behavior of the model should not differ in any significant way from that of the real system, when judged within the purpose and perspective of the model. Note that the duplication of historical time series via free-running simulation may not be part of this test, depending on the purpose and aggregation of the model.

For example, a cyclic model (say of commodity cycles), in which exogenous shocks, such as weather changes are modeled as random disturbances, may exhibit the same sort of cyclic behavior as the real system, but the exact phase of the cycles will vary according to the random inputs. Thus, the real system and model are liable to show relative drift, sometimes being 108^{o} out of phase, even for a "perfect" model. However, the FIMLOF consistency tests still hold.

TEST 15: Are the time-phase relationships among the variables correct?

TEST 16: Does the model, under simulation, do anything interesting (unexpected or contrary to the obvious)? If so, why? Does the reason make sense in terms of the real system?

TEST 17: Look at the abruptness of changes in variables, both absolute and relative to changes in other variables, as at turning points, etc. Do these correspond with those seen historically in the real system?

TEST 18: Is the model insensitive to variations in parameters and inputs which the real system has been historically insensitive to?

B.6 Conclusions.

The above list of informal model-validity tests is included only as an illustration of the existence of such tests. In their present form, they are, of course, vague, contradictory, and didactic. One might wish them to be complete, consistent, well-defined, and objective. Making them so is left as an exercise for the reader.

Appendix C

BAD DATA DETECTION AND IDENTIFICATION FOR DYNAMIC SYSTEMS

This appendix gives the formal derivation of the Normalized Updated Measurement Residuals (NUMR) and Normalized Updated State Residuals (NUSR) discussed in Chapter 2 and used in Chapter 6 for the identification of bad data points, due to major typographical errors, etc. The derivation holds strictly for linear dynamic systems; for use in FIMLOF and GPSIE, the linear model is taken to represent a nonlinear system, linearized about the trajectory defined by the optimal filter. Thus, the technique applies approximately to nonlinear systems under the same conditions discussed above for the extended Kalman filter and the locally-iterated filters implemented in GPSIE.

The usual filter equations (Figure 2-3) can be interpreted in the following way (so far assuming no bad data points): both the measurement data $\underline{z}(n)$ and the predicted state estimate $\hat{\underline{x}}(n|n-1)$ may be thought of as noisy measurements of the true state $\underline{x}(n)$. Since all information from $\underline{z}(1)...\underline{z}(n)$ is contained in $\underline{z}(n)$ and $\hat{\underline{x}}(n|n-1)$, $\underline{x}(n)$ must be considered completely unknown.

$$\underline{z}(n) = \underline{\underline{H}}\underline{x}(n) + \underline{v}(n)$$

$$\hat{\underline{x}}(n|n-1) = \underline{x}(n) + \underline{\delta}_x(n|n-1),$$

where

$$\text{Cov}\{\underline{v}(n)\} = E\{\underline{v}(n)\underline{v}'(n)\} = \underline{R}(n), \text{ and cov } \underline{\delta}_x(n|n-1) = \underline{\underline{\Sigma}}_x(n\ n-1)$$

We reduce the above equations to the following simplified but equivalent form:

$$\underline{y} = \underline{A}\underline{x} + \underline{\varepsilon}, \text{ where}$$

where

$$\underline{y} = \begin{bmatrix} \underline{z}(n) \\ \hat{\underline{x}}(n|n-1) \end{bmatrix} \qquad \underline{A} = \begin{bmatrix} \underline{H} \\ \underline{I} \end{bmatrix} \qquad \underline{x} = \underline{x}(n)$$

and

$$\text{cov}\{\underline{\varepsilon}\} = E\{\underline{\varepsilon}\,\underline{\varepsilon}'\} = \begin{bmatrix} \underline{R} & \underline{0} \\ \underline{0} & \underline{\Sigma}_x(n|n-1) \end{bmatrix} = \underline{\Omega}$$

Consider the problem of estimating $\underline{x}$ given $\underline{y}$, and checking for bad data (via a large component of _). Using Fisher estimation logic (see 35 , p. 134), since $\underline{x}$ in this context is completely unknown, we get

$$\hat{\underline{x}} = (\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\underline{y} \ .$$

We define the standard (unnormalized) residuals $\underline{r}$ as

$$\begin{aligned} \underline{r} &= \underline{y} - \underline{A}\hat{\underline{x}} \\ &= \left[\underline{I} - \underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\right]\left[\underline{A}\underline{x} + \underline{\varepsilon}\right] \\ &= \left[\underline{I} - \underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\right]\underline{\varepsilon} \end{aligned} \qquad \text{(C.1)}$$

We shall define the normalized residuals to be some linear function of the *updated* residuals $\underline{r}$. Denote the normalized residuals by $\underline{r}_n$.

$$\underline{r}_n = \underline{M}\,\underline{r} = \underline{W}\,\underline{\varepsilon} \qquad \text{(C.2)}$$

where

$$\underline{W} = \underline{M}\left[\underline{I} - \underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\right] \qquad \text{(C.3)}$$

The objective now is to choose the matrix $\underline{M}$ so as to facilitate bad data detection and identification (where bad data is to be defined below). A suitable $\underline{M}$ must satisfy the following two conditions:

Condition I (unit variance)

$$E\{\underline{r}_n \underline{r}'_n\} = 1 \qquad \text{for all sample times, except for}$$

bad data points.

Condition II (identification of bad data)

Define $\underline{W}_j$ as the j^{th} column of $\underline{W}$, and $\underline{W}_{jk}$ as the k^{th} element of $\underline{W}_j$. Then we require $|\underline{W}_{jj}| > |W_{jk}|$, $\quad \forall\ k \neq j$.

We define "bad data" as a large component of $\underline{\varepsilon}$, such that $\underline{\varepsilon}$ may be approximated as

$$\underline{\varepsilon} = \alpha \underline{e}_j \ , \qquad |\alpha| \ \text{large}$$

where

$\underline{e}_j$ = column vector of zeros, except for the j^{th} element, which is 1.

Then

$$\underline{r}_n = \underline{W}_j \alpha \qquad \text{for a bad data point.}$$

Bad data detection will occur when one or more of the components of $\underline{r}_n$ is large, compared to its unit standard deviation. That is,

$$|\underline{r}_{n\ j}| > \sim 3 \qquad \text{for some } j.$$

Condition II is motivated by the desire to achieve bad data identification via the additional criterion that when bad data is detected, the largest $|\underline{r}_{n\ j}|$ identifies the bad component of $\underline{\varepsilon}$.

Now, from (C.2),

$$\text{cov}\{\underline{r}_n\} = E\{\underline{r}_n \underline{r}_n'\} = \underline{M}\left[\underline{I} - \underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\right]\underline{\Omega}$$
$$\times \left[\underline{I} - \underline{\Omega}^{-1}\underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\right]\underline{M}'$$
$$= \underline{M}\left[\underline{\Omega} - \underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\right]\underline{M}'$$
$$= \underline{M}\,\underline{\Omega}\left[\underline{\Omega}^{-1} - \underline{\Omega}^{-1}\underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\right]\underline{\Omega}\,\underline{M}' \qquad (C.4)$$

From (C.3),

$$\underline{W} = \underline{M}\,\underline{\Omega}\left[\underline{\Omega}^{-1} - \underline{\Omega}^{-1}\underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1}\right] \qquad (C.5)$$

Define

$$\underline{\Gamma} = \underline{\Omega}^{-1} - \underline{\Omega}^{-1}\underline{A}(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1}\underline{A}'\underline{\Omega}^{-1} \qquad (C.6)$$

So (C.4 and C.5) become

$$E\{\underline{r}_n \underline{r}_n'\} = \underline{M}\,\underline{\Omega}\,\underline{\Gamma}\,\underline{\Omega}\,\underline{M}' \qquad (C.7)$$

and

$$\underline{W} = \underline{M}\,\underline{\Omega}\,\underline{\Gamma}$$

Define

$$\underline{D}_{\Gamma} = \text{diag}\{\underline{\Gamma}\} = \begin{bmatrix} \Gamma_{11} & & \underline{0} \\ & \ddots & \\ & & \Gamma_{JJ} \\ \underline{0} & & \ddots \end{bmatrix}$$

and

$$\underline{D}_{\Gamma}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\Gamma_{11}}} & & \underline{0} \\ & \ddots & \\ & & \frac{1}{\sqrt{\Gamma_{JJ}}} \\ \underline{0} & & \ddots \end{bmatrix}$$

We can now satisfy Conditions I and II if $\underline{M}$ is defined as

$$\underline{M}\,\underline{\Omega} = \underline{D}^{-\frac{1}{2}}$$

or

$$\underline{M} = \underline{D}^{-\frac{1}{2}}\,\underline{\Omega}^{-1} \tag{C.9}$$

which yields (via (C.7))

$$E\left\{\underline{r}_n\,\underline{r}'_n\right\} = \begin{bmatrix} 1 & \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}\Gamma_{22}}} & \cdots \\ \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}\Gamma_{22}}} & 1 & \\ \vdots & & \end{bmatrix}$$

and from (C.8)

$$\underline{W} = \begin{bmatrix} \frac{\Gamma_{11}}{\sqrt{\Gamma_{11}}} & \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}}} & \cdots \\ \frac{\Gamma_{12}}{\sqrt{\Gamma_{22}}} & \frac{\Gamma_{22}}{\sqrt{\Gamma_{22}}} & \\ \vdots & & \end{bmatrix} \tag{C.10}$$

The $\underline{W}$ defined by (C.10) can easily be shown to satisfy Condition II. For example,

$$\underline{W}_{11} = \frac{\Gamma_{11}}{\sqrt{\Gamma_{11}}}$$

$$\underline{W}_{1k}{}^{n} = \frac{\Gamma_{1k}}{\sqrt{\Gamma_{kk}}}$$

The fact that $\underline{\Gamma}$ is positive semidefinite implies that

$$\Gamma_{11}\Gamma_{kk} \geq \Gamma_{1k}^2$$

$$\sqrt{\Gamma_{11}}\sqrt{\Gamma_{kk}} \geq |\Gamma_{1k}|$$

$$\frac{\Gamma_{11}}{\sqrt{\Gamma_{11}}} \geq \frac{|\Gamma_{1k}|}{\sqrt{\Gamma_{kk}}}$$

$$\text{so} \quad |W_{11}| \geq |W_{1k}| \qquad \text{for all k.}$$

Now we return to the original notation:

$$\underline{A} = \begin{bmatrix} \underline{H} \\ \underline{I} \end{bmatrix}, \qquad \underline{\Omega} = \begin{bmatrix} \underline{R} & \underline{0} \\ \underline{0} & \underline{\Sigma}_x(n|n-1) \end{bmatrix}$$

$$\underline{\Omega}^{-1}\underline{A} = \begin{bmatrix} \underline{R}^{-1}\underline{H} \\ \underline{\Sigma}_x(n|n-1) \end{bmatrix} \qquad \underline{A}'\underline{\Omega}^{-1} = \begin{bmatrix} \underline{H}'\underline{R}^{-1} & \underline{\Sigma}_x^{-1}(n|n-1) \end{bmatrix}$$

$$\underline{A}'\underline{\Omega}^{-1}\underline{A} = \underline{H}'\underline{R}^{-1}\underline{H} + \underline{\Sigma}_x^{-1}(n|n-1)$$

$$(\underline{A}'\underline{\Omega}^{-1}\underline{A})^{-1} = \underline{\Sigma}_x(n|n)$$

so

$$\underline{\Gamma} = \begin{bmatrix} \underline{R}^{-1} - \underline{R}^{-1}\underline{H}\underline{\Sigma}_x(n|n)\underline{H}'\underline{R}^{-1} & \sim \\ \sim & \underline{\Sigma}_x^{-1}(n|n-1) \\ & - \underline{\Sigma}_x^{-1}(n|n-1)\underline{\Sigma}_x(n|n)\underline{\Sigma}_x^{-1}(n|n-1) \end{bmatrix}$$

or

$$\underline{\Gamma} = \begin{bmatrix} \underline{\Sigma}_z^{-1}(n|n-1) & \underline{\Gamma}_{12} \\ \underline{\Gamma}_{21} & \underline{H}'\underline{\Sigma}_z^{-1}(n|n-1)\underline{H} \end{bmatrix}$$

where

$$\underline{\Sigma}_z(n|n-1) = \underline{H}\,\underline{\Sigma}_x(n|n-1)\,\underline{H}' + \underline{R}$$

The equations of Section 2.6.2 follow, with

$$\underline{r}_n = \begin{bmatrix} \underline{r}_z(n|n) \\ \underline{r}_x(n|n) \end{bmatrix}$$

Note that $\underline{D}_x^{-\frac{1}{2}}$ is not always defined. For example, let the j^{th} column of $\underline{H}$ be the zero vector $\underline{0}$, indicating that the j^{th} state variable is not directly measured. Then the j^{th} diagonal term of $\underline{D}_x^{-\frac{1}{2}}$ is formally equal to 1/0. But note that $\underline{r}_n(n/n) = \underline{D}_\Gamma^{-\frac{1}{2}}\,\underline{\Omega}^{-1}\,\underline{r}(n)$, and for the normalized updated state residual NUSR $\underline{r}_x(n|n)$,

$$\begin{aligned} \underline{r}_x(n|n) &= \underline{D}_x^{-\frac{1}{2}}\underline{\Sigma}_x^{-1}(n|n-1)\left[\hat{\underline{x}}(n|n-1) - \hat{\underline{x}}(n|n)\right] \\ &= -\underline{H}'\underline{\Sigma}_z^{-1}(n|n-1)\,\underline{\delta}_x(n|n-1)\ . \end{aligned}$$

Thus, for elements of $\underline{x}(n)$ not directly measured (via a zero column of $\underline{H}$), the corresponding elements of $\underline{\Sigma}_x^{-1}(n|n-1)\left[\hat{\underline{x}}(n|n-1) - \hat{\underline{x}}(n|n)\right]$ are zero. It is not hard, then to show that the corresponding component of $\underline{r}_x(n\ n)$ is indeterminate.

The physical interpretation of the above situation is as follows. If x_j is not directly measured, then there is no redundancy on its

measure, the only information coming from the indirect "measurement" $\underline{\hat{x}}(n|n-1)$. But the derivation of $\underline{r}_{\underline{x}}(n|n)$ depends on the availability of redundant information about $\underline{x}(n)$ and the errors $\underline{\xi}(n)$.

Thus, when a component of $\underline{r}_{\underline{x}}(n|n)$ is indeterminate, via the nonexistence of all terms of $\underline{D}_{\underline{x}}^{-\frac{1}{2}}$, then the undefined components of $\underline{r}_{\underline{x}}(n|n)$ should be set to any arbitrary number. In GPSIE, they are set to zero.

David W. Peterson was born in Ashland, Kansas on April 2, 1946. He received the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, Massachusetts, in 1969 and 1972, respectively.

From 1969 to 1972 he was on the Research Staff of the Sloan School of Management, M.I.T., applying simulation and control theory to the design of top management policies for industrial corporations, and lecturing on the application and implementation of system dynamics. Since 1972 he has been a Research Assistant in Electrical Engineering at M.I.T., associated with the Energy Laboratory. His interests include the modeling of social systems, parameter estimation, filtering, and stochastic control. He has applied the above techniques to areas of corporate management and energy policy.

He is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.