

AD-A141 728

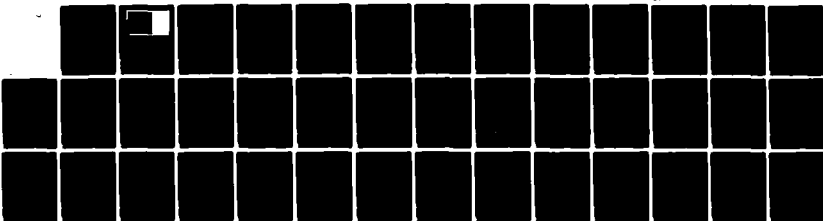
THE EFFECT OF AGGREGATION IN NONLINEAR REGRESSION(U)  
WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER  
M L GOLDBERG APR 84 MRC-TSR-2669 DAAG29-80-C-0041

1/1

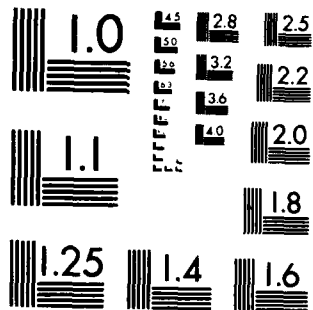
UNCLASSIFIED

F/G 12/1

NI



END  
DATE  
FILMED  
7-84  
DTIC



MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

2

AD-A141 728

MRC Technical Summary Report #2669

THE EFFECT OF AGGREGATION IN  
NONLINEAR REGRESSION

Miriam L. Goldberg

**Mathematics Research Center  
University of Wisconsin—Madison  
610 Walnut Street  
Madison, Wisconsin 53705**

April 1984

(Received March 12, 1984)

**DTIC FILE COPY**

Approved for public release  
Distribution unlimited

Sponsored by  
U. S. Army Research Office  
P. O. Box 12211  
Research Triangle Park  
North Carolina 27709

**DTIC**  
ELECTE **D**  
JUN 01 1984  
**S**  
**E**

84 05 31 065

- 1 -  
UNIVERSITY OF WISCONSIN-MADISON  
MATHEMATICS RESEARCH CENTER

THE EFFECT OF AGGREGATION IN NONLINEAR REGRESSION

Miriam L. Goldberg

Technical Summary Report #2669  
April 1984

ABSTRACT

It is often convenient to apply a nonlinear model developed for a single unit to data representing the average response for several units. The relationship between the parameters of the aggregate model and those of the individual units is investigated in a geometrical framework, for a general nonlinear model. The aggregation effect derived is closely related to the bias in nonlinear estimation, as given by Box (1971). Unlike the bias, however, the aggregation effect may be of comparable magnitude to the standard error of the estimates. The theoretical results derived are verified empirically in an application to a model of residential energy consumption.

AMS (MOS) Subject Classifications: 62J02, 62P99

Key Words: Nonlinearity, Regression, Aggregation, Nondifferentiable, Energy, Bias, Curvature

Work Unit Number 4 (Statistics and Probability)

---

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by joint funding from the Ford Foundation's State Environmental Management Program and the New Jersey Department of Energy.

### SIGNIFICANCE AND EXPLANATION

Analysis of complex physical or economic systems often involves the use of a model which is a nonlinear function of certain coefficients to be estimated from data. The same basic model may describe the behavior of several different units, such as households or individuals; however, the values of the coefficients vary from unit to unit. To determine the model coefficients it may be impractical to obtain data from each of the individuals. Instead, the available data represent averages over all the individuals. When the basic model used is nonlinear, estimates based on these aggregate data will be systematically different from those which would be obtained by averaging estimates based on the individual-unit data. Understanding this difference is important if the aggregate results are to be interpreted as representative of the group of individuals. This paper derives a formula for the systematic difference, and gives guidelines for determining if this effect is important in a given application.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/ _____	
Availability Codes	
Dist	Avail and/or Special
A-1	



---

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

Miriam L. Goldberg

## 1. INTRODUCTION

Aggregate data are collected and analyzed in a wide variety of contexts. Macroeconomic and sociological studies provide obvious examples, but in the physical sciences as well aggregate data may be relied on as a matter of necessity or convenience. A reasonable approach to analyzing data representing the average response of several units is to fit the data to the same model which describes a single unit's response. However, if this model is nonlinear, the resulting parameter values are different from those which would be obtained by analyzing the units separately and then averaging. Understanding this difference is important if an individual unit is to be compared to the "typical" embodied in the aggregate.

In this paper, a simple physical model of household energy demand provides the reference point for an exploration of the effect of aggregation in a nonlinear least squares regression. The model and aggregation are described in the remainder of this introductory section. Section 2 then lays out the geometry of the general nonlinear least squares regression, following Bates and Watts (1980). Discussion of biases begins in Section 3 with a review of M. J. Box's (1971) calculation of the bias for the single-unit

---

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by joint funding from the Ford Foundation's State Environmental Management Program and the New Jersey Department of Energy.

case (i.e., without aggregation). In Section 3, Box's work is interpreted in the geometrical context established in Section 2. Finally, in Section 4 the aggregation effect is derived and related to Box's results for the single unit, again in terms of geometry. The application of the main result in Section 4 to the model which motivated this study is presented in Section 5.

### 1.1. The Residential Energy Consumption Model

This study is the outgrowth of a broad-range investigation of energy consumption in houses which has taken place over the past ten years. (See, eg. Socolow (1978) and Dutt et al (1982).) A basic model employed in the household energy studies describes a house's daily fuel consumption  $Y_m$  during period (or month)  $m$  as a simple function of the  $N_m$  daily outdoor temperature readings  $T_{mj}$ :

$$Y_m = \alpha + \beta X_m(\tau) + \epsilon_m \quad (1.1)$$

with

$$X_m(\tau) = \frac{1}{N_m} \sum_{j=1}^{N_m} (\tau - T_{mj})_+ + \epsilon_m. \quad (1.2)$$

The subscript '+' in (1.2) indicates that the term in parentheses is set to zero if negative. The truncated sum  $X_m(\tau)$  is the average degree-days per day, base  $\tau$ . This truncation is the source of the nonlinearity in the model (1.1).

Physically, the intercept  $\alpha$  represents baseload consumption, such as for water heating and cooking, the slope

$\beta$  is a heating rate (or effective heat loss rate) and the reference temperature  $\tau$  is the outdoor temperature at which the furnace first comes on. The parameters  $\alpha$ ,  $\beta$ , and  $\tau$  are estimated by the method of least squares, using data  $Y_m$  from utility billing records for the house. Changes in these parameters, when correctly interpreted, reflect the impact of conservation measures undertaken in the house.

An overall conservation index  $\Gamma$  is obtained from the parameters  $\alpha$ ,  $\beta$ , and  $\tau$  as

$$\Gamma = 365(\alpha + X_0(\tau)) \quad (1.3)$$

where  $X_0(\tau)$  is the average base- $\tau$  degree-days per day, over a base period of several years. We call the index  $\Gamma$  the normalized annual consumption.

Aggregate versions of (1.1) and (1.3) have been used to analyze state- and utility- wide natural gas consumption in the residential heating sector (Fels and Goldberg, 1982 and Goldberg and Fels, 1982). These analyses represent an aggregation over many thousands of houses, for which individual data are not available. Another potential application of the aggregate model is for the "scorekeeping" of large-scale retrofit programs. If the model (1.1) is to be employed in assessing the total savings achieved in a large number of houses, then modeling the aggregate can simplify the scorekeeping task. One purpose of this study is to determine the adequacy of such a shortcut.

Assume, then, that we have  $U$  different housing units, each of whose consumption is described by (1.1) with parameters  $\alpha^u$ ,  $\beta^u$ ,  $\tau^u$ , and errors  $\epsilon_m^u$ ,  $u=1,2,\dots,U$ . Defining the aggregate consumption  $Y$  as the simple average of consumption in the  $U$  houses, we have

$$Y_m = \frac{1}{U} \sum_u (\alpha^u + \beta^u X_m(\tau^u) + \epsilon_m^u). \quad (1.4)$$

We will fit the aggregate  $Y_m$  to (1.1). The question to be answered in this paper is, how are the resulting parameter estimates related to the individual parameters  $\alpha^u$ ,  $\beta^u$ , and  $\tau^u$ ? The answer will be derived for a general nonlinear model, then applied to the energy example of our illustration.

## 2. THE GEOMETRY OF NONLINEAR LEAST SQUARES

The general nonlinear model with unknown  $p$ -dimensional parameter  $\theta$  can be written in matrix form as

$$Y = \eta(\theta) + \epsilon \quad (2.1)$$

$$E(\epsilon) = 0$$

$$E(\epsilon'\epsilon) = \sigma^2 I.$$

In (2.1),  $Y$ ,  $\eta$ , and  $\epsilon$  are  $M$ -dimensional vectors, such that  $\eta_m$ , the  $m$ th component of  $\eta$ , depends on observations  $X_m$  as well as on  $\theta$ . For the heating model (1.1),  $\theta = [\alpha \ \beta \ \tau]'$ , and  $\eta(\theta) = \alpha + \beta X(\tau)$ , where  $X$  is the  $M$ -dimensional vector with components  $X_m$  given by (1.2).

For the general model, as  $\theta$  ranges over the parameter space  $\Theta$ , the function  $\eta$  sweeps out a  $p$ -dimensional expectation surface  $S$  in the  $M$ -dimensional sample space:

$$S = \{\eta(\theta) : \theta \in \Theta\}. \quad (2.2)$$

We assume the function  $\eta$  to be twice differentiable, and define

$$\dot{\eta}(\theta) = \frac{\partial \eta}{\partial \theta'} = \left[ \frac{\partial \eta_m}{\partial \theta_j} \right]_{M \times p} \quad (2.3)$$

$$\ddot{\eta}(\theta) = \frac{\partial^2 \eta}{\partial \theta \partial \theta'} = \left[ \frac{\partial^2 \eta_m}{\partial \theta_i \partial \theta_j} \right]_{M \times p \times p} \quad (2.4)$$

(Throughout this paper, an expression in square brackets indicates a array with components given by the subscripted expression, such that  $m=1,2,\dots,M$ ;  $i,j=1,2,\dots,p$ .)

We then write the tangent vector at  $\eta(\theta_0)$  in the direction  $v$  as

$$t_v = \dot{\eta}(\theta_0)v.$$

The acceleration vector  $a_v$  is given by

$$a_v = v' \ddot{\eta}(\theta)v, \quad (2.5)$$

where the  $m$ th component of the  $M$ -vector  $a_v$  is the quadratic form in  $v$  given by the  $m$ th face of  $\ddot{\eta}$ . A measure of the nonlinearity of the model is provided by Bates' and Watts' (1980) relative curvature  $\gamma_v$ , defined as

$$\gamma_v = \rho \frac{|a_v|}{|t_v|^2} \quad (2.6)$$

where the scale factor

$$\rho = \sqrt{p\sigma^2}. \quad (2.7)$$

Below, we refer to the relative curvature  $\gamma$  simply as the curvature.

The acceleration can be decomposed into two components  $a_v^{||}$  and  $a_v^\perp$ , which are respectively parallel and normal to the tangent plane at  $\eta(\theta_0)$ . Substituting the tangential  $a_v^{||}$  and normal  $a_v^\perp$  for  $a_v$  in (2.6) gives analogous definitions of the tangential and normal curvatures  $\gamma_v^{||}$  and  $\gamma_v^\perp$ , respectively. Noting that the tangential acceleration component is caused by the parameterization chosen, while the normal component is independent of this choice, Bates and Watts refer to the tangential  $\gamma_v^{||}$  as the "parameter-effects" curvature, and to the normal  $\gamma_v^\perp$  as the "intrinsic curvature." As will be seen in the next section, biases in parameter estimates relate directly to the parameter-effects curvature  $\gamma_v^{||}$ , while biases in estimating  $\eta(\theta)$ , the conditional expectation of  $Y$  given the data  $X$ , relate to the intrinsic curvature  $\gamma_v^\perp$ .

### 3. ESTIMATION ERRORS FOR THE SINGLE UNIT NONLINEAR MODEL

For the general nonlinear model (2.1) with continuously differentiable  $\eta$ , the least squares estimate  $\hat{\theta}$  is the

solution to the normal equation

$$\dot{\eta}'(\theta)(Y - \eta(\theta)) = 0. \quad (3.1)$$

From this formula, Box (1970) shows that the error in the least squares estimate  $\hat{\theta}$  is given to second order in  $\varepsilon$  by

$$\hat{\theta} - \theta = (\dot{\eta}'\dot{\eta})^{-1}\dot{\eta}'\varepsilon + q \quad (3.2)$$

where  $q$  is quadratic in  $\varepsilon$ . The derivative  $\dot{\eta}$  is evaluated at the true  $\theta$ . To the order of this approximation, the bias in  $\hat{\theta}$  is given by the expectation of  $q$ , which Box derives as

$$b = E(q) \quad (3.3)$$

$$= -\frac{1}{2} (\dot{\eta}'\dot{\eta})^{-1}\dot{\eta}' [E\text{tr}(\ddot{\eta}_m(\dot{\eta}'\dot{\eta})^{-1}\sigma^2)].$$

Here,  $\ddot{\eta}_m$  denotes the  $m$ th face of  $\ddot{\eta}$ . Since  $\dot{\eta}$  is evaluated at the true  $\theta$ , the first-order term has zero expectation. Of course, for  $\eta$  linear in  $\theta$ , the first-order term in (3.2) is the usual linear regression error term, and there are no higher order terms.

To see the implications of Box's formula, we expand  $\eta(\hat{\theta})$  and  $\dot{\eta}(\hat{\theta})$  in a Taylor series about  $\theta$ . With the results given by (3.2) and (3.3), this expansion yields, again to  $O(\sigma^2)$ ,

$$E(\eta(\hat{\theta}) - \eta(\theta)) \quad (3.4)$$

$$\approx \dot{\eta}(\theta)E(\hat{\theta} - \theta) + \frac{1}{2}E\{[(\hat{\theta} - \theta)'\ddot{\eta}_m(\hat{\theta} - \theta)]\}$$

$$\begin{aligned}
&= \dot{\eta}(\theta)E(q) + \frac{1}{2} E([e' \dot{\eta}(\dot{\eta}'\dot{\eta})^{-1} \dot{\eta}'_m (\dot{\eta}'\dot{\eta})^{-1} \dot{\eta}' \varepsilon]) + O(\sigma^3) \\
&\approx -\frac{1}{2} \dot{\eta}(\dot{\eta}'\dot{\eta})^{-1} \dot{\eta}' [tr(\dot{\eta}'_m (\dot{\eta}'\dot{\eta})^{-1} \sigma^2)] + \frac{1}{2} [tr(\dot{\eta}'_m (\dot{\eta}'\dot{\eta})^{-1} \sigma^2)] \\
&\approx \frac{1}{2} (I-P) [tr(\dot{\eta}'_m (\dot{\eta}'\dot{\eta})^{-1} \sigma^2)].
\end{aligned}$$

In (3.4),  $P = \dot{\eta}(\dot{\eta}'\dot{\eta})^{-1} \dot{\eta}'$  is the projection matrix onto the tangent plane at  $\theta$ . Noting from (3.2) that

$$E((\hat{\theta} - \theta)(\hat{\theta} - \theta)') = (\dot{\eta}'\dot{\eta})^{-1} \sigma^2 + O(\sigma^3), \quad (3.5)$$

we find

$$\begin{aligned}
E(\eta(\hat{\theta}) - \eta(\theta)) & \quad (3.6) \\
&= \frac{1}{2} (I-P) [tr(\dot{\eta}'_m E((\hat{\theta} - \theta)(\hat{\theta} - \theta)'))]: \quad O(\sigma^3).
\end{aligned}$$

Thus, the expected error in the fit  $\eta(\hat{\theta})$  is orthogonal to the tangent plane at  $\eta(\theta)$ .

Equations (3.3) and (3.6) can be expressed in simpler form by writing Box's bias approximation in terms of the accelerations defined by (2.5). Equation (3.3) then becomes

$$\begin{aligned}
E(\hat{\theta} - \theta) &\approx -\frac{1}{2} (\dot{\eta}'\dot{\eta})^{-1} \dot{\eta}' E(a) \quad (3.7) \\
&= -\frac{1}{2} (\dot{\eta}'\dot{\eta})^{-1} \dot{\eta}' E(a^{||}),
\end{aligned}$$

where

$$a = (\hat{\theta} - \theta)' \ddot{\eta} (\hat{\theta} - \theta)$$

is the acceleration in the direction  $\hat{\theta} - \theta$ . Similarly, the orthogonality condition (3.6) can be written as

$$E(\eta(\hat{\theta}) - \eta(\theta)) \approx \frac{1}{2} E(a^\perp). \quad (3.8)$$

Thus, as noted by Bates and Watts (1980), the bias in  $\hat{\theta}$  depends only on parameter-effects nonlinearity. By contrast, the bias in the fit  $\eta(\hat{\theta})$  depends only on intrinsic nonlinearity.

We can also motivate the latter result heuristically, as follows. Since the response vector  $Y$  has a distribution spherically symmetric about  $\eta(\theta)$ , the expected value of the projection  $\eta(\hat{\theta})$  of  $Y$  onto a symmetric expectation surface  $\eta(\theta)$  will have no component in the tangent plane. Such symmetry is implicit in the second-order approximation, which describes any planar cross-section of the surface by a parabola centered at  $\eta(\theta)$ .

The situation is illustrated for a simulated data set in Fig. 1a. For a one-dimensional model defined by

$$\eta(\theta) = [\exp(\theta X_m)],$$

The figure shows the projection of the expectation surface onto the plane determined by the two vectors  $\dot{\eta}(\theta)$  and  $\ddot{\eta}(\theta)$ . For the particular data set shown, the nonlinearity is moderately large; the root-mean-square parameter-effects and intrinsic curvatures are, respectively,  $\gamma_{rms}^{\parallel} = 0.4$  and  $\gamma_{rms}^{\perp} = 0.3$ . For purposes of this illustration, the estimated parameters  $\hat{\theta}$  and  $\hat{\sigma}^2$  were assumed to represent the true values.

Over the range of values displayed, the expectation surface is indeed quite symmetrical, and nearly parabolic. Thus, the quadratic approximation appears to be quite good. The expected point of the fit  $\eta(\hat{\theta})$  as given by (3.6) is indicated in the figure by the '+' sign slightly offset from the curve, on a line perpendicular to the tangent line.

Above the expectation surface, against the same tangent axis, the probability density function is shown for the tangential component of the random error  $\epsilon$ . The symmetry of this density function, which implies the orthogonality condition (3.6), also means that in terms of a uniform coordinate system on the tangent plane, the parameter estimate is unbiased. That is, if equally-spaced points in  $\theta$  were mapped under  $F\eta$  to equally-spaced points in the tangent plane, the least-squares estimate  $\hat{\theta}$  would be unbiased.

The nonuniform coordinate system imposed on the tangent line is indicated by the tick marks, representing evenly-spaced values of  $\theta$ . The compression of these marks for large values of  $\theta$  reflects the parameter-effects non-linearity. As a result of this compression, an error  $\epsilon$  in the positive direction along the tangent line will cause an error of much larger magnitude in  $\hat{\theta}$  than an equal error  $\epsilon$  in the negative direction. Hence, despite the intrinsic symmetry of the surface  $\eta(\theta)$  and the errors  $\epsilon$ , the parameter estimate  $\hat{\theta}$  is biased.

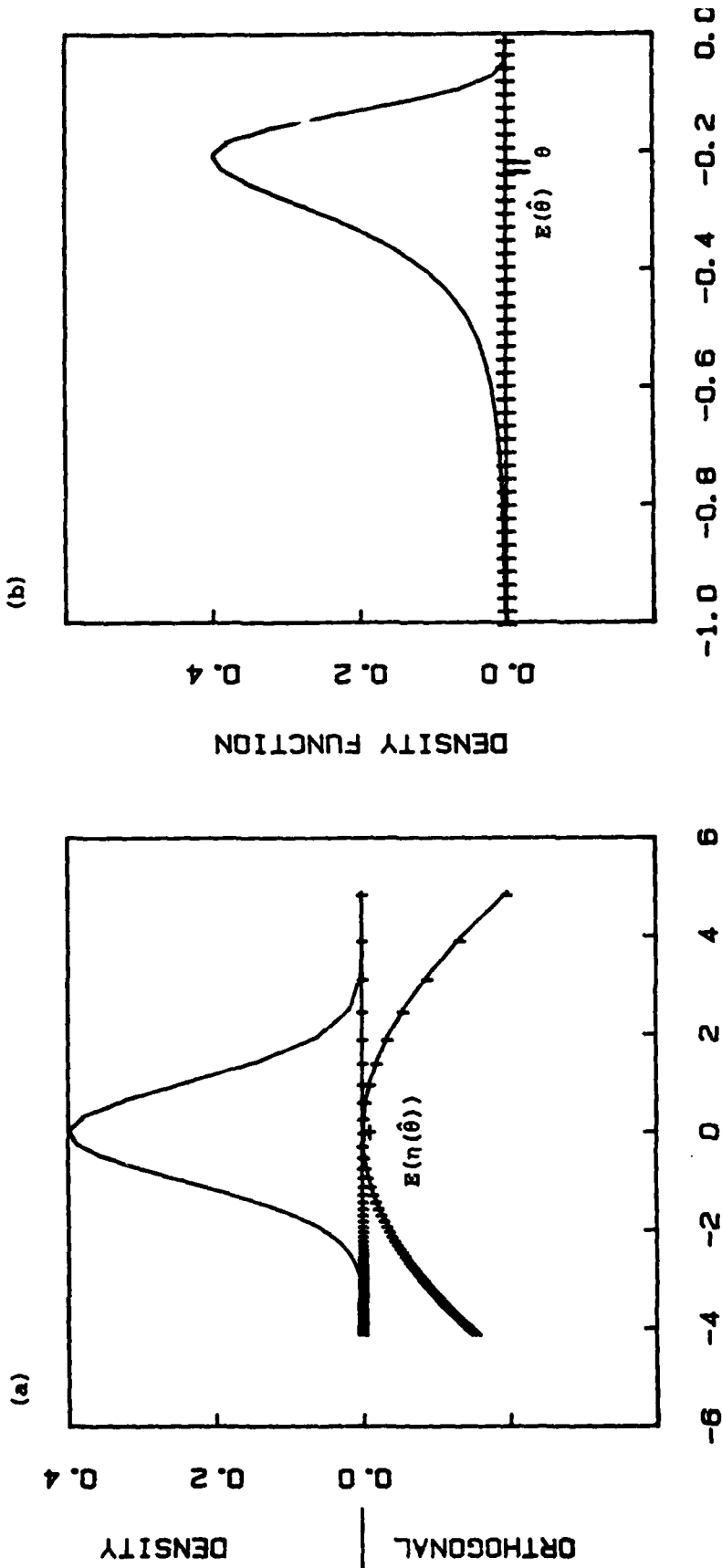
The bias induced by the parameterization can be seen by mapping the tangent line and its overlying density back to the parameter space  $\theta$ , as shown in Fig. 1b. This mapping corresponds to stretching and squeezing the horizontal axis so that the tick marks are equally spaced, with the result that the density function for  $\hat{\theta}$  is skewed. Thus, the expected value of  $\hat{\theta}$ , as determined from Box's formula (3.3), is seen in the figure to be offset from the true value  $\theta$ , which is the mode of the distribution.

The same framework used to illustrate Box's bias in Fig. 1 will help to describe the aggregation effect developed below. We first explore the errors introduced by aggregation, then consider the more general model misspecification problem of which aggregation is a special case.

#### 4. THE AGGREGATE NONLINEAR MODEL

##### 4.1. General Framework

Section 1 describes the aggregation problem for the residential energy consumption model (1.1). For the general case, we assume a population of  $U$  different units, each of whose responses  $Y^u$  can be described by (2.1), but with different values  $\theta^u$  of the unknown parameter. We cannot observe any individual unit's response, but at each time  $m$  we observe the average of the  $U$  responses. Thus,



TANGENT PLANE COORDINATE                      PARAMETER SPACE COORDINATE

Figure 1. Box's Bias Illustrated by a One-Parameter Exponential Model.

- a) The approximately parabolic expectation surface is projected onto the plane of the tangent and acceleration vectors. Superimposed over the tangent plane is the Gaussian probability density function for the tangential component of the error vector  $\epsilon$ .
- b) The density function mapped back to the parameter space is skewed, resulting in the indicated bias.

$$Y = \frac{1}{U} \sum_u (\eta(\theta^u) + \epsilon^u) \quad (4.1)$$

$$= \bar{\eta} + \epsilon,$$

with

$$\epsilon = \frac{1}{U} \sum_u \epsilon^u \quad (4.2)$$

and

$$\bar{\eta} = \frac{1}{U} \sum_u \eta(\theta^u).$$

We define  $\theta$  and  $D$  as the population mean and dispersion of parameter values  $\theta^u$ :

$$\theta = \frac{1}{U} \sum_u \theta^u \quad (4.3)$$

$$D = \frac{1}{U} \sum_u (\theta^u - \theta)(\theta^u - \theta)'. \quad (4.4)$$

We could equally well assume the  $\theta^u$  to be a random sample from a population with mean  $\theta$  and variance-covariance matrix  $D$ . Since it is possible to obtain the results we are interested in with reference to the particular individuals observed, i.e. without appealing to expectations, we do so. This approach avoids any confusion which might be caused by having random units as well as random disturbances  $\epsilon$ . The expectation operator  $E$  will refer in all cases to expectations over the set of possible disturbances  $\epsilon$ .

Since  $\bar{\eta}$  is a function of up to  $pU$  different parameters, and the individual responses are unobservable, we take

a simplified model, fitting

$$Y = \eta(\theta) + e. \quad (4.5)$$

The least squares estimate  $\hat{\theta}$  is found by solving the normal equation (3.1), which can now be written as

$$\dot{\eta}(\hat{\theta})'(\bar{\eta} + \varepsilon - \eta(\hat{\theta})) = 0. \quad (4.6)$$

#### 4.2. Asymptotic Aggregation Effect

By setting  $\varepsilon = 0$  and solving (4.6) for  $\theta$ , we find the aggregate value  $\theta^A$  for which the least squares estimate  $\hat{\theta}$  is a consistent estimator:

$$\dot{\eta}(\theta^A)'(\bar{\eta} - \eta(\theta^A)) = 0. \quad (4.7)$$

That is, the "asymptotic estimate"  $\theta^A$  corresponds to the point where the discrepancy

$$z = \eta(\theta^A) - \bar{\eta}$$

is normal to the expectation surface. Of course,  $\eta(\theta^A)$  is the point on the expectation surface closest to  $\bar{\eta}$ .

Following Box (1971) we find the difference  $\delta$  between the average parameter value  $\theta$  and the asymptotic estimate  $\theta^A$  by expanding  $\bar{\eta}$  in the normal equation (4.7) about the asymptotic  $\theta^A$ :

$$0 \approx \dot{\eta}'(\eta(\theta^A) + \dot{\eta}(\frac{1}{U} \sum_u (\theta^u - \theta^A))) \quad (4.8)$$

$$\begin{aligned}
& + \frac{1}{2} \frac{1}{U} \sum_u (\theta^u - \theta^A)' \ddot{\eta} (\theta^u - \theta^A) - \eta(\theta^A) \\
& = \dot{\eta}' (\dot{\eta} \delta + \frac{1}{2} \frac{1}{U} \sum_u (\theta^u - \theta)' \ddot{\eta} (\theta^u - \theta) \\
& \quad + \frac{1}{2} \delta' \ddot{\eta} \delta) \\
& = \dot{\eta}' (\dot{\eta} \delta + \frac{1}{2} [\text{tr}(\ddot{\eta}_m D)]) + \frac{1}{2} \delta' \ddot{\eta} \delta,
\end{aligned}$$

where

$$\delta = \theta - \theta^A.$$

Assuming  $\delta \delta' \ll D$ , (4.8) gives, to first order in  $D$ ,

$$0 = (\dot{\eta}' \dot{\eta}) \delta + \frac{1}{2} \dot{\eta}' [\text{tr}(\ddot{\eta}_m D)]. \quad (4.9)$$

Hence,

$$\delta \approx -\frac{1}{2} (\dot{\eta}' \dot{\eta})^{-1} \dot{\eta}' [\text{tr}(\ddot{\eta}_m D)]. \quad (4.10)$$

Note that the approximation (4.10) is valid provided the value  $\delta$  it yields satisfies the above assumption that  $\delta \delta' \ll D$ .

Recalling (3.5), we see that the aggregation effect  $\delta$  given by (4.10) is the same as the bias  $b$  given by (3.3), with the population variance  $D$  replacing the estimation variance  $\sigma^2 (\dot{\eta}' \dot{\eta})^{-1}$ . If we adopt the shorthand notation

$$a_u = a(\theta^u - \theta) \quad (4.11)$$

$$= (\theta^u - \theta)' \dot{\eta} (\theta^u - \theta) |_{\theta}$$

and define

$$\bar{a} = \frac{1}{U} \sum_u a_u, \quad (4.12)$$

then by arguments similar to those used in the previous section we have

$$\delta \approx -\frac{1}{2} (\dot{\eta}' \dot{\eta})^{-1} \dot{\eta}' \bar{a} \quad (4.13)$$

$$= -\frac{1}{2} (\dot{\eta}' \dot{\eta})^{-1} \dot{\eta}' \bar{a}^{\parallel};$$

$$\eta(\theta) - \bar{\eta} \approx \frac{1}{2} \bar{a}; \quad (4.14)$$

$$\eta(\theta^A) - \bar{\eta} \approx \frac{1}{2} \bar{a}^{\perp}. \quad (4.15)$$

Thus, the difference between  $\eta(\theta)$ , the model at the average parameter  $\theta$ , and the model average  $\bar{\eta}$  is approximated by the second-order term  $(1/2) \bar{a}$  (in (4.14)). The difference between the asymptotic estimate  $\eta(\theta^A)$  and  $\bar{\eta}$  is approximated by the normal component of the acceleration term,  $(1/2) \bar{a}^{\perp}$  (in (4.15)). The aggregation effect  $\delta$  is approximated by the coefficients, with respect to the columns of  $\dot{\eta}$ , of the tangential component  $(1/2) \bar{a}^{\parallel}$  (in (4.13)). Note that even if  $\eta(\theta)$  is an exact quadratic, (4.13) and (4.15) are still only approximate, because we have ignored products of first and second derivatives with terms of order  $\delta^2$ .

The effect of aggregation is illustrated in Fig. 2, for the same data set displayed in Fig. 1. For this illustration, we have assumed that the random error  $\varepsilon = 0$ , and taken the dispersion  $D$  of actual parameter values  $\theta^u$  to

be identical to the dispersion matrix of estimates

$\sigma^2(\dot{\eta}'\dot{\eta})^{-1}$  from the original example. The point  $\bar{\eta}$  shown in Fig. 2a is that given by the approximation (4.15).

We have further assumed that the values of  $\theta^u$  are symmetrically distributed about their mean value  $\theta$ . This assumption is useful for illustration, and will be natural in many situations. Figure 2a shows how the points  $\eta(\theta^u)$  on the expectation surface are asymmetrically distributed, as a result of the nonuniform mapping from the parameter space  $\theta$ . Thus, the average sample-space point  $\bar{\eta}$  is on a line orthogonal to the surface at  $\eta(\theta^A)$ , not at  $\eta(\theta)$ . The difference between the projection and the point corresponding to the average  $\theta$  is translated back to the parameter space in Fig. 2b.

Figures 1 and 2 illustrate the similarity between the bias and the aggregation effect. Comparing the corresponding formulae (3.3) and (4.10) shows that the relative importance of bias and aggregation depends on the relative magnitude of the estimation variance  $\sigma^2(\dot{\eta}'\dot{\eta})^{-1}$  and the population variance  $D$ .

The estimation variance is conveniently represented by the standard radius  $\rho$  defined by (2.7). This radius measures the approximately spherical spread of estimates  $\eta(\hat{\theta})$  about their expected value: according to the linear approximation

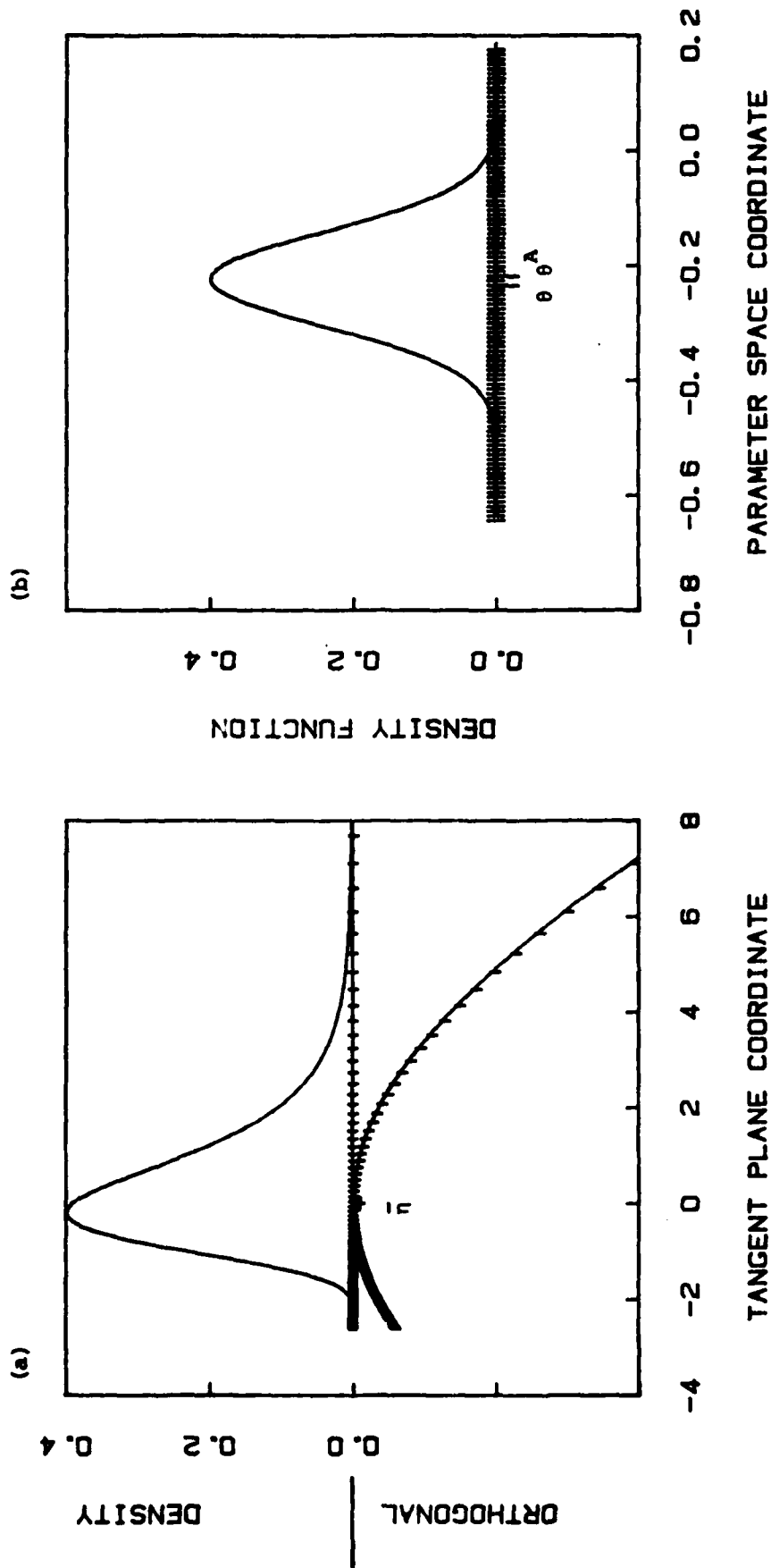


Figure 2. Aggregation Effect Illustrated by a One-Parameter Exponential Model.

a) The probability density function shows the distribution of the tangential components of sample space points  $\eta(\theta)$  corresponding to different units  $u$ .

b) In the parameter space, the distribution of parameter values  $\theta^u$  in the population is assumed symmetric. The nonlinear mapping to the sample space results in the indicated discrepancy between the average parameter  $\theta$  and the asymptotic ( $\epsilon = 0$ ) estimate  $\theta^A$ .

$$E(|\eta(\hat{\theta}) - \eta(\theta)|^2) \approx \rho\sigma^2. \quad (4.16)$$

The relative curvature  $\gamma$  which incorporates the radius  $\rho$  as a normalizing constant indicates the potential magnitude of the bias, but not of the aggregation effect. The latter depends upon the dispersion of expectation points  $\eta(\theta^u)$ , which is not in general spherically symmetric, even to a first approximation. Hence, no single radius which might be defined by analogy with (4.16) adequately describes this spread. In comparing the bias with the aggregation effect, the important consideration is the spread of the estimates compared to the spread of population values, not overall, but in the directions of large parameter-effects curvature.

When the bias and aggregation effects are of comparable magnitude, the combined effects of dispersion of values  $\theta^u$  and nonzero random disturbances  $\varepsilon$  must be considered. Fortunately, as will be seen, the two effects can be treated as additive.

#### 4.3. Finite Sample Bias for the Aggregate Nonlinear Model

Equation (3.3) gives Box's expression for the bias due to nonlinearity alone for a nonlinear least squares estimator. We will derive a similar expression for the nonlinearity bias in the presence of model misspecification. The particular misspecification we are concerned with is the

averaging described in Section 4.1. However, our development will be somewhat more general.

Assume as above

$$Y = \bar{\eta} + \varepsilon.$$

$$E(\varepsilon) = 0.$$

$$\text{Var}(\varepsilon) = \sigma^2 I.$$

$$z = \bar{\eta} - \eta(\theta^A).$$

$$\dot{\eta}(\theta^A)' z = 0. \quad (4.17)$$

This is the same formulation as used above, but now  $\bar{\eta}$  may be regarded as any function which does not lie on the expectation surface  $S$  defined by  $\eta(\theta)$ , the function fit to  $Y$ . We consider the effect of  $\varepsilon$  on the least squares estimate of  $\theta^A$ .

The normal equation (3.1) can now be written as

$$\dot{\eta}'(\hat{\theta})((\bar{\eta} + \varepsilon) - \eta(\hat{\theta})) = 0. \quad (4.18)$$

Expanding (4.16) about  $\theta^A$  yields

$$0 = (\dot{\eta} + (\hat{\theta} - \theta^A)' \ddot{\eta})' \quad (4.19)$$

$$x(\bar{\eta} + \varepsilon - (\eta + \dot{\eta}(\hat{\theta} - \theta^A) + \frac{1}{2}(\hat{\theta} - \theta^A)' \ddot{\eta}(\hat{\theta} - \theta^A)))|_{\theta^A}$$

Since

$$Y = \eta(\theta^A) + z + \varepsilon,$$

it is appropriate to treat the total discrepancy  $z + \varepsilon$  as

a single quantity producing errors in the estimate of  $\theta^A$ .

Thus, following Box, we assume

$$\hat{\theta} - \theta^A \approx C(z+\epsilon) + q, \quad (4.20)$$

where  $C$  is a  $p \times M$  matrix and the  $p$ -dimensional  $q$  is quadratic in  $z + \epsilon$ . Equation (4.17) then becomes (evaluating all derivatives at  $\theta^A$ ):

$$0 = (\dot{\eta} + (z+\epsilon)'C'\ddot{\eta} + q'\ddot{\eta})' \quad (4.21)$$

$$\begin{aligned} & \times (z+\epsilon - \dot{\eta}C(z+\epsilon) - \dot{\eta}q - \frac{1}{2}(z+\epsilon)'C'\ddot{\eta}C(z+\epsilon) \\ & + O((z+\epsilon)^3)). \end{aligned}$$

Equating to zero successive powers of  $z + \epsilon$ , we have

$$(\dot{\eta}' - \dot{\eta}'\dot{\eta}C)(z+\epsilon) = 0 \quad (4.22)$$

so that

$$C = (\dot{\eta}'\dot{\eta})^{-1}\dot{\eta}'. \quad (4.23)$$

To second order,

$$\begin{aligned} 0 &= (z+\epsilon)'C'\ddot{\eta}'(I-P)(z+\epsilon) - \dot{\eta}'\dot{\eta}q \\ &\quad - \frac{1}{2}\dot{\eta}'(z+\epsilon)'C'\ddot{\eta}C(z+\epsilon) \\ &= \epsilon'C'\ddot{\eta}'z + \epsilon'C'\ddot{\eta}'(I-P)z - \dot{\eta}'\dot{\eta}q \\ &\quad - \frac{1}{2}\dot{\eta}'\epsilon'C'\ddot{\eta}C\epsilon. \end{aligned} \quad (4.24)$$

Hence,

$$q = (\dot{\eta}'\dot{\eta})^{-1}(\epsilon'C'\ddot{\eta}'z + \epsilon'C'\ddot{\eta}'(I-P)\epsilon) \quad (4.25)$$

$$- \frac{1}{2} \dot{\eta}' \varepsilon' C' \dot{\eta} C \varepsilon).$$

Box has shown that the second term in parentheses has zero expectation. Thus, to the order of our approximations, the bias is given by

$$E(\hat{\theta} - \theta^A) \approx E(C(z+\varepsilon) + q) \quad (4.26)$$

$$= E(q)$$

$$\approx - \frac{1}{2} (\dot{\eta}' \dot{\eta})^{-1} \dot{\eta}' [\text{tr}(\dot{\eta} \dot{\eta}'^{-1} \sigma^2)]$$

which is identical to (3.3). That is, the first approximation to the random-error induced bias of  $\hat{\theta}$  is the same as for the correctly specified model. The result is reasonable: (3.3) shows the bias has no dependence on the error component normal to the tangent plane at  $\eta(\theta)$ , and the effect of misspecification is to add a normal error component  $z$  to  $\eta(\theta^A)$ .

Thus, to first order, the effects of aggregation and random errors are additive. Combining (3.3) and (4.10) we have the expected error in the aggregate estimate  $\hat{\theta} - \theta$  as

$$E(\hat{\theta} - \theta) \approx \frac{1}{2} [\text{tr}(\dot{\eta} \dot{\eta}'^{-1} (D - \sigma^2 (\dot{\eta}' \dot{\eta})^{-1}))]. \quad (4.27)$$

The asymptotic estimate  $\theta^A$  corresponds to the point  $\eta(\theta^A)$  on the expectation surface  $S$  which comes closest to  $\bar{\eta}$ , the expected value of the aggregate  $Y$ . The point  $\eta(\theta^A)$  is invariant under reparameterizations, which is not true of  $\eta(\theta)$ . For many purposes, then, the consistently estimated  $\theta^A$  and not the population average  $\theta$  better

characterizes the aggregate. On the other hand, the population mean  $\theta$  may have a more natural interpretation than the asymptotic estimate  $\theta^A$ , which is defined only in terms of the regression model. In such a case, the net effect represented by (4.27) is of interest.

## 5. BIAS ESTIMATION FOR NONDIFFERENTIABLE MODELS

### 5.1. Considerations for General Nondifferentiable Models

The explicit formulas developed in the previous sections to estimate systematic effects apply only to smooth models. We will consider first how to apply Box's bias formula to a segmented model. Similar reasoning will then extend the aggregation effect formula (4.10) to the energy model which motivated its development.

The bias given by (3.3) is based on a quadratic approximation  $\eta$  to the model function  $\eta$ . In fact, (3.3) may be interpreted as the bias in the least squares estimate  $\hat{\theta}$  for the model function  $\eta$ . This bias estimate is also valid for the estimate  $\hat{\theta}$  for the original function  $\eta$  to the extent that  $\eta$  approximates  $\eta$  over the region of interest.

As described above, the approximation is found from a Taylor's series expansion of  $\eta$ . However, Box's derivation is equally valid for any quadratic approximation  $\eta$  such that

$$\eta(\theta) = \eta(\theta) + R_1 \quad (5.1)$$

and wherever  $\dot{\eta}$  is defined

$$\dot{\eta}(\theta) = \dot{\eta} + R_2$$

where the remainders  $R_1$  and  $R_2$  are of small order. For an arbitrary  $\eta$ , it may be difficult to prove that  $R_1$  and  $R_2$  are small enough to be neglected in the derivation described in the previous section. But if the approximate model  $\eta(\theta)$  and derivative  $\dot{\eta}$  are close to the original  $\eta(\theta)$  and  $\dot{\eta}(\theta)$  over a wide region, and the estimate  $\tilde{\theta}$  for the approximate model is close to the estimate  $\hat{\theta}$  for the original, then the bias estimated for  $\tilde{\theta}$  should give a good indication of the bias in  $\hat{\theta}$ .

Box's derivation makes use of the fact that the least squares estimate for a continuously differentiable model is a solution to the normal equation (4.6). For the nondifferentiable model, however, least squares estimates that fall at points of discontinuity of  $\dot{\eta}$  are in general not solutions to the normal equation. Nevertheless, the arguments just given apply equally well to a nondifferentiable model. That is, with the interpretation of (3.3) as the bias in  $\tilde{\theta}$  for the approximate model, Box's formula requires only that  $\tilde{\theta}$  be a solution to the normal equation for the approximate model.

To apply this formula to a particular nondifferentiable model of interest, then, we must produce a quadratic approx-

imation  $\hat{\eta}$  which is close to the original model  $\eta$  over any region of interest, and for which the least squares estimate  $\hat{\theta}$  is close to the original estimate  $\hat{\theta}$ . The problem of constructing such an approximation is next addressed for the energy model (1.1).

## 5.2. Applications to the Energy Model

A quadratic approximation to the energy model  $\alpha_1 + \beta X(\tau)$ , which is nondifferentiable with respect to one parameter  $\tau$  only, was found simply by fitting a quadratic to  $X(\tau)$ , coinciding with  $X(\tau)$  at  $\tau = \hat{\tau}$ :

$$\tilde{X}_m(\tau_i) = X_m(\hat{\tau}) + b_m(\tau_i - \hat{\tau}) + c_m(\tau_i - \hat{\tau})^2. \quad (5.2)$$

The approximation  $\tilde{X}$  then replaced the actual degree-day variable  $X$  in (1.1).

For each month  $m$ , the coefficients  $b_m$  and  $c_m$  were found by the method of least squares for a set of evenly spaced reference temperatures  $\tau_i$ . Details of the fitting procedure are described in previous work by the author (Goldberg, 1982 and 1983). In all cases, the smooth  $\tilde{X}(\tau)$  and derivative  $\dot{\tilde{X}}(\tau)$  were quite close to the original  $X(\tau)$  and  $\dot{X}(\tau)$  over a range of values of  $\tau$  corresponding to several standard errors of the estimate  $\hat{\tau}$ .

Applying the bias and aggregation effect formulae (3.3) and (4.10) to the energy model (1.1) for regional aggregates

requires a modification even before the smoothing just discussed. The reason is that meters are read on different days throughout the month. Hence, when (1.1) is averaged over a utility region or state, the degree-days  $X_m(\tau^u)$  for different households  $u$  correspond to different one-month periods, as well as to different reference temperatures  $\tau^u$ . To account for this billing lag, the degree-day variable  $X_m$  for utility or state aggregates in month  $m$  is defined as

$$X_m(\tau) = \quad (5.3)$$

$$\frac{\sum_{j=1}^{N_{m-1}} j (\tau - T_{m-1,j})_+ + \sum_{j=1}^{N_m} (N_m - j) (\tau - T_{mj})_+}{\sum_{j=1}^{N_{m-1}} j + \sum_{j=1}^{N_m} (N_m - j)}$$

Total monthly sales per household  $Y_m$  is then fit to (1.1) with  $X_m$  defined by (5.3) rather than (1.2).

Equation (1.1) was fit by this method for a total of 75 different aggregate data sets, corresponding to different combinations of utility regions and time periods. Each aggregate was for residential gas heating customers only. Details are given in Goldberg (1982 and 1983).

### 5.3. Bias Estimates for the Smoothed Energy Model

The effect of using the quadratic approximation  $\tilde{X}$  in place of  $X$  in (1.1) is summarized in Table 1 for the 75 aggregate data sets. The table shows the median bias  $b$

Table 1. MEDIAN NONLINEARITY BIAS  $b$  FOR 75 AGGREGATE DATA  
SETS FIT TO THE ENERGY MODEL

Parameter $\theta_j$	Bias $b = E(\tilde{\theta}_j - \theta_j)$	Smoothing Error $\tilde{\theta}_j - \hat{\theta}_j$	Standard Error s.e. $(\hat{\theta}_j)$
BASELOAD $\alpha$ (Th/cu-d)	-0.003	-0.002	0.108
HEAT RATE $\beta$ (Th/cu-°Fd)	0.0003	-0.0001	0.0089
REFERENCE TEMPERATURE $\tau$ (°F)	0.01	0.02	1.27
NORMALIZED ANNUAL CONSUMPTION $\Gamma$ (Th/cu-yr)	0	30	20

The bias  $b$  was computed for the smoothed energy model using (3.3). Shown for each quantity is the median value found in 75 fits of aggregate data sets to (1.1). Abbreviations are Th for Therms, cu for customer, d for day, and yr for year.

for the parameter estimate  $\tilde{\theta}$  using  $\tilde{X}$ . Also shown are the median soothing error (the difference between  $\tilde{\theta}$  and the estimate  $\hat{\theta}$  for the original model) and the median standard error for  $\hat{\theta}$  based on a linearization of the model (1.1).

The bias formula for normalized annual consumption  $\tilde{f}$  was obtained by re-expressing the model  $\eta(\alpha, \beta, \tau)$  as

$$[\xi_m(\Gamma, \beta, \tau)] = [\Gamma + \beta(X_m(\tau) - X_o(\tau))],$$

then applying (3.3). To calculate normalized annual consumption  $\tilde{f}$  for the approximate model, the degree-day norm  $X_o$  was also smoothed around  $t$  via (5.2).

For all four parameters  $\alpha$ ,  $\beta$ ,  $\tau$ , and  $\Gamma$ , Table 1 shows that the difference between the estimate  $\tilde{\theta}$  for the smoothed model and  $\hat{\theta}$  for the original is of the same order of magnitude or greater than the nonlinearity bias estimate for  $\tilde{\theta}$ . On the other hand, both the error due to using the approximation and the bias for the approximate model are much smaller than the standard error in all cases. Thus, even taking the absolute difference between  $\tilde{\theta}$  and  $\hat{\theta}$  as an upper bound on the magnitude of the bias in  $\hat{\theta}$ , this bias will be inconsequential for most purposes. In particular, neither comparisons between engineering calculations and parameters found empirically by fitting (1.1) nor comparisons among empirical measurements will be distorted by the least squares estimation procedure.

It has generally been known, and was demonstrated by Box in producing (3.3), that nonlinearity biases are a second order effect compared with the standard error of the estimate in a least squares regression. However, both the conventional wisdom and Box's result were based on smooth models. Here, we have shown that they are essentially valid for the nondifferentiable model of the present study. The methodology used is generalizable to other nondifferentiable models.

#### 5.4. Aggregation Effects for the Regional Aggregate

To apply the aggregation formula (4.10), we need not only a smooth expectation function  $\eta$ , but also an estimate of the population dispersion matrix  $D$ . For the parameters of the energy model (1.1), an estimate of  $D$  was based on the estimated parameters  $\hat{\theta}^u$  from a sample of 71 individual gas-heated houses in New Jersey. As described by Dutt et al (1982), these houses were selected to represent different housing types in the state. The aggregation formula as applied to the energy model and the estimation of the dispersion matrix  $D$  for this case are discussed more fully in Goldberg (1982).

Table 2 summarizes the results of applying the aggregation formula (4.10) to the energy model (1.1). The table shows the median estimated difference  $\delta$  between each

Table 2. MEDIAN AGGREGATION EFFECT  $\delta$  FOR 75 DATA SETS FIT  
TO THE ENERGY MODEL

Parameter $\theta_j$	Aggregation Effect $\delta_j = \bar{\theta}_j - \theta_j^A$	$\frac{\bar{\theta}_j - \theta_j^A}{\text{s.e.}(\hat{\theta}_j)}$	Least Squares Estimate $\hat{\theta}_j$
BASELOAD $\alpha$ (Th/cu-d)	-0.037	-0.30	1.35
HEAT RATE $\beta$ (Th/cu-°Fd)	0.0033	0.44	0.224
REFERENCE TEMPERATURE $\tau$ (°F)	-0.28	-0.25	64.1
NORMALIZED ANNUAL CONSUMPTION $\Gamma$ (Th/cu-yr)	0	-0.01	1543

The aggregation effect  $\delta$  was computed for the smoothed energy model using (4.10). See caption to Table 1.

population mean  $\alpha$ ,  $\beta$ ,  $\tau$ , and  $\Gamma$  and the corresponding aggregate parameter , for the same group of aggregate data sets summarized in Table 1. As a point of reference, the median value found for each coefficient is also indicated. The table shows all the aggregation effects to be quite small compared with the typical parameter values.

The second column of the table shows the median ratio of the aggregation effect to the standard error of the aggregate estimate. From this comparison, it is seen that the estimated difference between the aggregate  $\theta^A$  and the population mean  $\theta$  is generally small compared with the uncertainty in the aggregate estimate  $\hat{\theta}^A$ , though not as small as the biases summarized in Table 1.

As noted in the previous section, the aggregate  $\theta^A$  may be a more meaningful summary statistic for the population than is the mean  $\theta$ . It is extremely important, however, that the results of scorekeeping analyses be easily understood. The close correspondence between the coefficient  $\theta^A$  estimated by the aggregate analysis and the population mean  $\theta$  is therefore an important feature of the energy analysis methodology.

##### 5.5. Comparison of Empirical and Theoretical Aggregation Effects.

We have just seen that the aggregation formula (4.10) allows us to give a simple interpretation of the aggregate parameter estimates for (1.1). Another application of the formula is in justifying a shortcut approach for large-scale retrofit programs. If essentially equivalent results might be obtained by analyzing aggregate data only, rather than analyzing each house individually, measurement costs might be cut significantly. It should be noted that the primary cost savings would come from the reduction in data collection and handling required, rather than in computation.

Above, we found for the state aggregate that the difference between the theoretical aggregate parameters  $\theta^A$  and the population means  $\theta$  was slight. For normalized annual consumption  $\Gamma$ , the parameter of major interest in assessing a retrofit program, the aggregation effect was typically less than one tenth as large as the standard error, and always less than one half as large.

To test the aggregation formula empirically, and to determine whether similar results would be found for a small group of houses, (4.10) was evaluated for one of the modules of the Modular Retrofit Experiment described by Dutt et al (1982). The results are shown in Table 3. Also shown in this table is the empirically determined aggregation effect, the difference between the estimate  $\hat{\theta}^A$  for the aggregate and the average  $\bar{\hat{\theta}}$  of the estimates for the 17 individual houses.

Table 3. COMPARISON OF THEORETICAL AND EMPIRICAL AGGREGATION EFFECTS

Parameter $\theta_j$	Theoretical Difference $\delta_j = \theta_j^A - \bar{\theta}_j$	Empirical Difference $\bar{\theta}_j^A - \bar{\theta}_j$	Median Standard Error of Individual Estimates $s.e.(\hat{\theta}_j^u)$	Standard Error of Aggregate Estimate $s.e.(\hat{\theta}_j^A)$	Aggregate Parameter Value $\hat{\theta}_j^A$
INTERCEPT $\alpha$ (Th/cu-d)	0.060	0.027	0.34	0.28	1.41
HEATING RATE $\beta$ (Th/cu-°F)	-0.0033	-0.0056	0.025	0.023	0.266
REFERENCE TEMPERATURE, $\tau$ (°F) (Assuming $Cov(\beta^u, \tau) = 0$ )	0.14	0.40	2.97	2.44	61.8
NORMALIZED ANNUAL CONSUMPTION $\Gamma$ (Th/cu-yr)	1.86	2.30	72.1	58.8	1620

Results are shown for aggregate and individual fits of (1.1) to consumption data for 17 houses in the Elizabethtown module of the Modular Retrofit Experiment (Dutt et al., 1982).

The theoretical estimate of the aggregation effect agrees with that found empirically only in order of magnitude. Recall, though, that we are comparing the theoretical estimate of the expected difference  $E(\hat{\theta}^A - \theta)$  between the aggregate and mean parameters for the smoothed model with the empirical difference  $\hat{\theta}^A - \bar{\theta}$  for the original. In addition, we have only a rough estimate of the dispersion  $D$  of actual parameter values  $\theta^u$  for the 17 houses. Thus, the basic agreement between the theoretical and empirical values is quite reassuring.

More importantly in terms of energy analysis, both the theoretical and the empirical results indicate only slight differences between the aggregate parameter and the population mean. The differences are on the order of 10% as large as the median standard error of individual estimates. That is, the effect of aggregating is small compared to the uncertainty in individual estimates. For the index  $\Gamma$ , the difference between the mean and aggregate values is less than 4% of the aggregate standard error and less than 0.2% of the aggregate estimate.

The standard deviation of estimates  $\hat{\Gamma}^u$  across the population is much larger than the standard error of the aggregate estimate  $\hat{\Gamma}^A$ . That is, while the aggregate parameter is very well-determined, the aggregate analysis alone does not indicate the variability of  $\Gamma$  within the popula-

tion. Relying solely on aggregate data would preclude obtaining such information. However, if the mean value  $\Gamma$  is the only parameter of interest the aggregate analysis is entirely sufficient.

## 6. CONCLUSIONS

We have seen that the bias  $b$  for the single-response case and the asymptotic aggregation effect  $\delta$  are very similar, with the estimation variance and population variance of the true values playing parallel roles. Both the aggregation effect and Box's bias are induced by the parameter-effects nonlinearity; in the case of bias, a symmetric distribution of fits  $\eta(\hat{\theta})$  translates into an asymmetric distribution of estimates  $\hat{\theta}$ , while in the aggregation case, a symmetric distribution of true parameter values translates into an asymmetric distribution of points  $\eta(\theta^u)$ . When the bias and aggregation effect are of comparable magnitude, the two can be treated as additive.

An important difference between the bias and the aggregation effect is their behavior as the number of observations increases. For the bias, the estimation variance  $\sigma^2(\hat{\eta}'\hat{\eta})^{-1}$  decreases with increasing sample size. As a result, the magnitude of the bias decreases. Furthermore, the accuracy of Box's approximation (3.3) based on the assumption of small  $\epsilon$ , improves.

For the aggregation effect, on the other hand, the dispersion  $D$  of true values  $\theta^u$  is fixed. Therefore, increasing the sample size will neither improve the accuracy of the approximation (4.10) nor reduce the magnitude of the aggregation effect. An important consequence is that in a sufficiently large sample the aggregation effect may be comparable in magnitude to the standard error of the estimate. By contrast, to the accuracy of the second order approximation, the bias is always of smaller order than the standard error. Thus, while the bias caused by nonlinearity can reasonably be ignored in most applications, the effect of aggregation may be quite important.

#### ACKNOWLEDGEMENT

The author thanks Peter Bloomfield and Margaret Fels, supervisors of the dissertation on which this work is based.

#### REFERENCES

1. Bates, D. M. and Watts, D. G. (1980). "Relative Curvature Measures of Nonlinearity." J.R.S.S. B 42(1):1-25.
2. Box, M. J. (1971). "Bias in Nonlinear Estimation" (with discussion). J.R.S.S. B 32:171-201.
3. Dutt, G. S., Lavine, M., Levi, B., and Socolow, R. (1982). "The Modular Retrofit Experiment: Exploring the House Doctor Concept." Princeton University Center for Energy and Environmental Studies Report No. 130.
4. Fels, M. F. and Goldberg, M. L. (1982). "Measuring Household Fuel Consumption on the Standard Living Cycle." Energy 7:489-504.
5. Goldberg, M. L. (1983). "Measures of Nonlinearity for Segmented Regression Models," University of Wisconsin Mathematics Research Center, Madison, WI, Technical Summary Report No. 2549.
6. Goldberg, M. L. (1982). A Geometrical Approach to Nondifferentiable Regression Models as Related to Methods for Assessing Residential Energy Conservation. Princeton University Center for Energy and Environmental Studies Report No. 142. (As Ph.D. Thesis for Department of Statistics.)
7. Goldberg, M. L. and Fels, M. F. (1982). "Post-embargo Conservation: A New Jersey Case Study." Princeton University Center for Energy and Environmental Studies Report No. 143.
8. Socolow, R. H. (ed.), (1978). Saving Energy in the Home. Ballinger Pub. Co., Cambridge.

