A GENERALIZED LINEAR MODEL WITH "GAUSSIAN" REGRESSOR VARIABLES

David R. Brillinger
University of California, Berkeley

ABSTRACT

A model in which the conditional expected value of a response variate is an unknown nonlinear function of an unknown linear combination of regressor variates is considered. It is shown that in the case that the regressors are stochastic and jointly Gaussian, or are deterministic and quasi-Gaussian, the ordinary least squares estimates provide useful estimates of the coefficients of the linear combination up to an arbitrary multiplier. The cases of both conditional and unconditional inference are investigated.

KEY WORDS: *Gaussian regressors, Generalized linear model, Multiple regression, Quasi-Gaussian regressors.*

1. INTRODUCTION

Multiple regression is one of the most powerful of statistical techniques. The procedure has been given numerous justifications and interpretations. The traditional approach to it rests on a linear model

$$y_j = \alpha + \beta x_j + \varepsilon_j \ , \qquad (1.1)$$

with the y_j, x_j, $j=1,\ldots,n$ observed, with α, β unknown parameters of interest, with the ε_j zero mean error variates, with the x_j p column-vectors, and with β a p row-vector.

Prepared with the support of National Science Foundation Grant PFR-7901642. I would like to thank Paul Feigin for a number of helpful comments.

Letting

$$\overline{x} = \sum_{j=1}^{n} x_j/n \ , \tag{1.2}$$

the ordinary least squares estimate, $\hat{\beta}$, satisfies

$$\hat{\beta} \sum_j (x_j - \overline{x})(x_j - \overline{x})^\tau = \sum_j y_j(x_j - \overline{x})^\tau \tag{1.3}$$

with τ denoting the operation of matrix transposition. In some circumstances the entries of β have causal interpretations, though these must be exercised cautiously (see Box, 1966 and Mosteller and Tukey, 1977, Chapter 13). It seems that substantive scientists have gotten more service out of ordinary least squares estimates than the narrow assumptions of the traditional approach might lead one to suspect possible. In many of these situations it is not the actual value of the coefficients that is of interest, rather it is their relative values, which are somehow measuring the relative importance of the regressor variates of interest. In this paper it is demonstrated that, in the case where the regressors are jointly "Gaussian," the ordinary least squares estimates have a working interpretation for a broader class of models then one might have imagined. The solution, $\hat{\beta}$, of (1.3) is shown to provide an estimate of β in the model

$$y_j = g(\alpha + \beta x_j) + \varepsilon_j \ , \tag{1.4}$$

up to an unknown constant of proportionality. The practical implication is that if the regressors are chosen to be Gaussian, or happen to be approximately so, then despite the possible presence of an unknown nonlinearity, $\hat{\beta}$ still reflects the relative importance of the regressor variates.

After computing $\hat{\beta}$, one may go on to prepare a scatter plot of the points $(\hat{\beta}x_j, y_j)$, $j=1,\ldots,n$ and look for a functional form for $g(\cdot)$. Alternatively, one might compute a nonparametric estimate of $g(u)$ by smoothing the y_j values with $\hat{\beta}x_j$ near u.

It is the usual statistical practice to examine the sampling properties of the least squares estimate conditional on the x values that come to hand. Both the unconditional and conditional distributions are investigated in the paper. Interesting questions arise in the present context, because the fact that the xs are Gaussian is an integral part of the study. It will be seen that it is not convenient to construct confidence regions conditional on a realization of a Gaussian sequence; however, useful regions may be constructed if $x_1, x_2, \ldots$ is a deterministic quasi-Gaussian sequence of a particular sort.

The paper further investigates the extent to which the results require an assumption of normality and describes an application of the results to an identification problem in neurophysiology and an estimation problem in economics.

2. AN ELEMENTARY LEMMA

The whole basis of the procedure is the following simple result given in Brillinger (1977).

Lemma 1. Let (U,V) *be bivariate normal with* U *nondegenerate. Let* $g(\cdot)$ *be a measurable function with* $E\{|g(U)|\}$ *and* $E\{|g(U)U|\} < \infty$. *Then*

$$\operatorname{cov}\{g(U),V\} = \operatorname{cov}\{U,V\} \operatorname{cov}\{g(U),U\}/\operatorname{var} U \ . \tag{2.1}$$

Proof. One has $E\{V|U\} = \mu + \Theta U$ with $\Theta = \operatorname{cov}\{U,V\}/\operatorname{var} U$. Now

$$\operatorname{cov}\{g(U),V\} = \operatorname{cov}\{g(U),E\{V|U\}\} = \Theta \operatorname{cov}\{g(U),U\} \ ,$$

giving the result.

That the regression of V on U is linear is key to the result. It is perhaps worth noting that for $g(\cdot)$, an almost differentiable function (defined in Stein, 1981) satisfying $E\{|g'(U)|\} < \infty$, one may write

$$\operatorname{cov}\{g(U),U\}/\operatorname{var} U = E\{g'(U)\} \tag{2.2}$$

This last is an identity that Stein (1981) makes use of in his construction of improved estimates of the mean of a multivariate Gaussian.

Now consider the model (1.4) with x_j Gaussian of covariance matrix Σ and ε_j independent of x_j. Then, from (2.1),

$$\text{cov}\{y_j,x_j\} = \beta\Sigma\ \text{cov}\{g(U),U\}/\text{var } U \tag{2.3}$$

with $U = \alpha + \beta x_j$. (Here $\text{cov}\{y,x\} = E\{(y - \mu_y)(x - \mu_x)^\tau\}$.) The linear regression coefficient of y on x is proportional to β of expression (1.4). Provided $\text{cov}\{g(U),U\} \neq 0$, the constant of proportionality will not be 0. If consistent estimates of $\text{cov}\{y,x\}$ and Σ are constructed, then a consistent estimate of β (up to an arbitrary multiplier) may be constructed. The details of the estimate are presented in the next section for the unconditional case.

3. UNCONDITIONAL INFERENCE

The estimate of interest is the ordinary least squares estimate defined by (1.3). Its properties will be investigated when the variates are related by $y_j = g(\alpha + \beta x_j) + \varepsilon_j$ and when the x_j are Gaussian.

Assumption I. $x_1, x_2, \ldots$ *are independent normals with mean* μ_x *and nonsingular covariance matrix* Σ . $\varepsilon_1, \varepsilon_2, \ldots$ *are independent of the* xs *and have finite variance* σ^2. $E\{x_j^\tau x_j|g(\alpha + \beta x_j)|^2\} < \infty$ *for* $j = 1,2,\ldots$.

From expressions (1.3) and (2.3) one can see that, almost surely, the ordinary least squares estimate $\hat\beta$ tends to $\text{cov}\{y,x\}\Sigma^{-1} = k\beta$, where

$$k = \text{cov}\{g(\alpha + \beta x),\alpha + \beta x\}/\text{var}\{\alpha + \beta x\} . \tag{3.1}$$

That is, $\hat\beta$ is a strongly consistent estimate of β, up to a constant, k, of proportionality. For $\hat\beta$ to be useful, one needs $k \neq 0$.

Turning to the question of the asymptotic distribution of $\hat{\beta}$, set

$$h(x) = g(\alpha + \beta x) - \gamma - \delta x \tag{3.2}$$

where $\delta = k\beta$ and $\gamma = E\{g(\alpha + \beta x) - \delta x\}$. Then one has

Theorem 1. Suppose Assumption I is satisfied. Let $y_j = g(\alpha + \beta x_j) + \varepsilon_j$, $j = 1,2,\ldots$. *Let* $\hat{\beta}$ *be given by (1.3) and* k *by (3.1). Then* $\sqrt{n}(\hat{\beta} - k\beta)$ *is asymptotically normal with mean* $\mathbf{0}$ *and covariance matrix*

$$\sigma^2\Sigma^{-1} + \Sigma^{-1}E\{h(x)^2(x - \mu_x)(x - \mu_x)^\tau\}\Sigma^{-1} . \tag{3.3}$$

This theorem may be demonstrated using a result of Freedman (1981). The proof is presented in the Appendix. In the case that $g(\cdot)$ is a linear function, the second term in (3.3) will be absent and one has the usual expression for the asymptotic covariance matrix of a least squares estimate.

For the estimate $\hat{\beta}$ to be of practical use, one needs some estimate of its covariance matrix. Several general methods are available for obtaining the latter: the delta method, the jackknife, and the bootstrap. $\hat{\beta}$ is a function of U-statistics, hence the use of the jackknife estimate of the covariance matrix is justified by the results of Arvesen (1969). With a further assumption of $E\{|g(\alpha + \beta x)|^4\} < \infty$, the use of the bootstrap estimate is justified by the results of Freedman (1981). The delta method estimate will now be constructed.

Write expression (1.3) as

$$[\hat{\mu}\ \hat{\beta}] \sum_j \begin{bmatrix} 1 \\ x_j \end{bmatrix} [1\ x_j^\tau] \frac{1}{n} = \sum_j y_j [1\, x_j^\tau] \frac{1}{n} \tag{3.4}$$

or $[\hat{\mu}\ \hat{\beta}]\mathbf{A} = \mathbf{B}$. Here $\mathbf{A}$ and $\mathbf{B}$ are means of (matrix-valued) sample values. As $\mathbf{A}$ and $\mathbf{B}$ are means, the variances and covariances of all their entries may be estimated directly, by the usual expressions. Now if $\mathbf{A}_0$, $\mathbf{B}_0$ denote the expected values of $\mathbf{A}$ and $\mathbf{B}$ respectively, then one has the perturbation expansion

$$[\hat{\mu}\ \hat{\beta}] = BA^{-1} = B_0A_0^{-1} + (B - B_0)A_0^{-1} - B_0A_0^{-1}(A - A_0)A_0^{-1} + \dots . \tag{3.5}$$

This gives $\hat{\beta}$ as an (approximate) linear function of A and B, whose covariance matrix may now be estimated using the estimates of the variances and covariances of the entries of A and B and replacing A_0, B_0 by A, B respectively.

Having an approximation to the large sample distribution of $\hat{\beta}$ and an estimate of its covariance matrix, one can go on to construct approximate confidence intervals, test hypotheses, and the like.

A concern with these results, however, is that they are unconditional — averaging over all realizations of the xs. Yet in practice, $x_1,\dots,x_n$ will usually be ancillary and one would like to carry out inference conditional on its value at hand. The next section considers this issue.

4. CONDITIONAL INFERENCE

Let $X = \{x_1, x_2,\dots\}$ denote the sequence of regressor variables. This section is concerned with inferences conditional on X. To begin, consider the case where $x_1, x_2,\dots$ are independent realizations of a p-variate normal with mean μ_x and covariance matrix Σ. Directly from expression (1.3) one has

$$E\{\hat{\beta}|X\} = \sum_j g(\alpha+\beta x_j)(x_j - \bar{x})^\tau \left[\sum_j (x_j - \bar{x})(x_j - \bar{x})^\tau\right]^{-1} = \mu_n(X) \tag{4.1}$$

$$\mathrm{Var}\{\hat{\beta}|X\} = \sigma^2\left[\sum_j (x_j - \bar{x})(x_j - \bar{x})^\tau\right]^{-1} = \sigma^2 S_n(X) . \tag{4.2}$$

The variance is the usual least squares expression. In the case that $g(\cdot)$ is linear, the conditional expected value is β; however it will generally be different from β or $k\beta$. A question of interest is how close may it be expected to be to $k\beta$?

The asymptotic conditional distribution of $\hat{\beta}$ is normal. Specifically one has:

Theorem 2. Suppose Assumption I is satisfied. Then almost surely

$$\text{Prob}\{\sqrt{n}(\hat{\beta} - \mu_n(X))S_n(X)^{-1/2}/\sigma \leq b|X\} \to \Phi(b_1)\ldots\Phi(b_p) \quad (4.3)$$

as $n \to \infty$, *where* $b = (b_1,\ldots,b_p)$ *and* $\Phi(\cdot)$ *is the standard normal cumulative.*

This result provides information concerning the deviations of $\hat{\beta}$ from $\mu_n(X)$ for a given X. However, one is interested in the deviations of $\hat{\beta}$ from $k\beta$. The next lemma indicates that while $\mu_n(X) - k\beta = o_{a.s.}(1)$, it is not generally $o_{a.s.}(1/\sqrt{n})$ and so (4.3) is not of great use in conditional inference questions concerning β. Theorem 2 is proved in the Appendix of the paper.

Lemma 2. Let $x_1, x_2 \ldots$ *be independent normals with mean* μ_x *and nonsingular covariance matrix* Σ. *Suppose* $E\{x_j^\tau x_j|g(\alpha+\beta x_j)|^2\} < \infty$, $j = 1,2,\ldots$. *Then*

$$E\{\hat{\beta}|X\} = \mu_n(X) = k\beta + \frac{1}{\sqrt{n}}W + o_{a.s.}(\frac{1}{\sqrt{n}}) , \quad (4.4)$$

where k *is given by (3.1) and* W *is normal with mean* 0 *and covariance matrix*

$$\Sigma^{-1}E\{h(x)^2(x - \mu_x)(x - \mu_x)^\tau\}\Sigma^{-1} , \quad (4.5)$$

$h(x)$ *being given by (3.2).*

The deviations of $\mu_n(X)$ from $k\beta$ are seen to be of order $1/\sqrt{n}$, generally. One implication of this is that the result (4.3) cannot be used to construct approximate confidence regions for $k\beta$. Some other approach is needed. The lemma is proved in the Appendix.

As the lemma and discussion make clear, for a typical realization of the Gaussian process X, $\mu_n(X)$ does not tend to $k\beta$ rapidly enough to be useful. Consider expression (4.1). The term

$$\sum_j g(\alpha + \beta x_j)x_j/n$$

may be considered an approximation to the integral, or expected value, $E\{g(\alpha + \beta x)x\}$. This suggests that by choosing a sequence $x_1, x_2, \ldots$ corresponding to a clever numerical integration rule, one might be able to have $E\{\hat{\beta}|X\}$ closer to $k\beta$ than $O_{a.s.}(1/n)$. This does turn out to be possible.

Halton (1960) has demonstrated the existence of a sequence of points $u_1, u_2, \ldots$ in the unit cube $[0,1]^p$ with the property that

$$D_n = \sup_{I \epsilon J} \left| \frac{\#(u_1, \ldots, u_n \ \epsilon \ I)}{n} - \mu(I) \right| = O(n^{-1}(\log n)^p) \qquad (4.6)$$

where J is the family of all subintervals of $[0,1]^p$ and where $\mu(I)$ is the Lebesgue measure of I. (A computer algorithm for generating the sequence is given in Halton and Smith, 1964.) The usefulness of this sequence is that for a function, f, with variation, $V(f)$, in the sense of Hardy and Krause (see Neiderreiter, 1978, p. 967), one has

$$\left| \frac{1}{n} \sum_{j=1}^{n} f(u_j) - \int f(u)du \right| \leq V(f)D_n = O(n^{-1}(\log n)^p) \qquad (4.7)$$

for bounded $V(f)$. The sequence $u_1, u_2, \ldots$ may be said to be quasi-uniform. Writing $u_j = (u_{j1}, \ldots, u_{jp})$ and $x_j = (x_{j1}, \ldots, x_{jp})$ with $x_{jk} = \Phi^{-1}(u_{jk})$, the sequence $x_1, x_2, \ldots$ may be said to be quasi-Gaussian. Letting $h(x_j) = f(u_j)$, one has from (4.7),

$$\left|\frac{1}{n} \sum_{j=1}^{n} h(x_j) - \int h(x)\phi(x_1)\ldots\phi(x_p)dx\right| = O(n^{-1}(\log n)^p) \qquad (4.8)$$

for $h(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p))$ of bounded variation with $\phi(\cdot)$ denoting the standard normal density. One might say that for $p > 1$, quasi-Monte Carlo techniques exist that outperform naive Monte Carlo.

Returning to the question of the estimation of $\hat{\beta}$ of the model (1.4), suppose now that the values of the regressors may be chosen by the experimenter. Suppose he takes $x_1, x_2, \ldots$ to be the above quasi-Gaussian sequence. Consider $\hat{\beta}$ satisfying

$$\hat{\beta} \sum_j x_j x_j^\tau = \sum_j y_j x_j^\tau \ . \tag{4.9}$$

(There is no need to correct for the mean with this sequence.) One has

$$\begin{aligned} E\{\sum_j y_j x_j^\tau/n &= \sum_j g(\alpha + \beta x_j) x_j^\tau/n \\ &= \int g(\alpha + \beta x) x^\tau \phi(x_1) \ldots \phi(x_p) dx + O(n^{-1}(\log n)^p) \\ &= k\beta + O(n^{-1}(\log n)^p) \end{aligned}$$

from (4.8), provided g is of bounded variation as required. Similarly,

$$\sum_j x_j x_j^\tau/n = I + O(n^{-1}(\log n)^p) \ .$$

In summary, for the deterministic quasi-Gaussian sequence indicated above, one has

$$E\ \hat{\beta} = k\beta + O(n^{-1}(\log n)^p) \ . \tag{4.10}$$

The variance-covariance matrix of β is $\sigma^2[\sum_j x_j x_j^\tau]^{-1}$, and hence the conclusion of Theorem 2 becomes

$$\mathrm{Prob}\{(\hat{\beta} - k\beta)[\sum_j x_j x_j^\tau]^{1/2}/\sigma \leq b\} \rightarrow \Phi(b_1) \ldots \Phi(b_p) \ . \tag{4.11}$$

Once an estimate of σ is at hand, approximate confidence regions for $k\beta$ may be constructed using (4.11).

With an estimate of $g(\cdot)$, σ^2 may be estimated from the residuals of the fit. Various nonparametric estimates of a regression function are available. A bibliographic review of these is given in Collomb (1981). In the present context one might form

$$\hat{g}(u) = \sum_{j=1}^{n} y_j W_n(u - \hat{\beta} x_j) / \sum_{j=1}^{n} W_n(u - \hat{\beta} x_j) \tag{4.12}$$

for example, with W_n a sequence of weight functions becoming concentrated at 0 as n increases. For large n, $\hat{g}(u)$ may be expected to be near $g(\alpha + u/k)$. The error variance may be estimated by

$$\hat{\sigma}^2 = \sum_{j=1}^{n} [y_j - \hat{g}(\hat{\beta}x_j)]^2/n \; . \tag{4.13}$$

A procedure for constructing approximate confidence regions for $k\beta$ has been set down.

5. DISCUSSION

Section 3 discussed inference in the unconditionsl case when $x_1, x_2, \ldots$ was any realization of a sequence of independent p-variate normals. Section 4 developed inference for the case that $x_1, x_2, \ldots$ was a very particular deterministic sequence (that was quasi-Gaussian). It would appear that the latter conditional inference procedure is the preferred one — as is the case in the usual (linear) regression situation — since $x_1, \ldots, x_n$ is generally an ancillary statistic. Lehmann (1981) comments on some aspects of ancillaries and conditional inference.

If the form of the function $g(\cdot)$ is known, then one will be able to determine other estimates of β, for example, the maximum likelihood. These other estimates may be expected to be more efficient. There have been at least two studies in which the ordinary least squares estimate has been compared with the maximum likelihood estimate. In both cases it has been found to perform well, even when the xs were not Gaussian.

Greene (1981) considered the model

$$y_j = \max\{0, \alpha + \beta x_j + \varepsilon_j\} \tag{5.1}$$

with the εs independent normals of mean 0 and variance σ^2. He derived both the ordinary least squares and the maximum likelihood estimate of β for a set of data from a study of female labor supply. Here y was the number of hours worked in a survey year. The xs are listed in Table 1. (Eight of them are dummy variables.) The estimate, $\hat{\beta}$, has been standardized to $\hat{\beta}\hat{\beta}^\tau = 1$. The proportion of nontruncated observations was .460.

There is close agreement between the results of least squares and maximum likelihood. This occurs despite some of the xs having far from normal distributions. Greene (1981) is able to construct an estimate for of (3.1), since $g(\cdot)$ is known, and so obtain an estimate of β itself.

Table 1

Variable	Maximum Likelihood	Least Squares
x_1 = small child	-.4140	-.3831
x_2 = health	-.5072	-.4472
x_3 = other income	.0005	.0008
x_4 = wage	.5156	.6053
x_5 = south	.2953	.2989
x_6 = farm	-.2266	-.2218
x_7 = urban	.0554	.0523
x_8 = age	.0097	.0094
x_9 = education	.0113	.0125
x_{10} = rel. wage	.1438	.1346
x_{11} = 2nd marriage	.0127	.0143
x_{12} = mean divorce prob.	.2416	.2381
x_{13} = high divorce prob.	.2906	.2652

Brillinger and Segundo (1979) present an example of a successful application of the estimation procedure discussed in this paper, in a more complicated situation. A neuron was stimulated by a fluctuating current, causing it to fire every so often. The stimulating current was taken to be stationary Gaussian. In the classic model of neuron firing, the input current $X(t)$ is filtered to form the membrane potential

$$U(t) = \int_0^{B(t)} a(u)X(t-u)du$$

with $B(t)$ the time elapsed at time t since the neuron last fired. The neuron then fires next when $U(t)$ crosses an approximately constant threshold. It is of interest to estimate the function $a(\cdot)$ of (5.1) and to confirm the presence of a threshold.

A time series analog of the procedure considered in this paper was applied to experimental data consisting of a record of the current taken as input and the times at which the neuron fired. Strictly speaking, the model is not appropriate here because of correlation introduced by $B(t)$ being present in (5.1). A maximum likelihood procedure was developed to deal with this difficulty. It was found that the results of the procedure of this paper were quite consistent with the maximum likelihood results. In the principal experiment, the input current was taken to be Gaussian. In a second experiment, the input current was taken to have a uniform distribution. Figure 1 gives the time series analog of the regression estimate of $a(\cdot)$ when $X(t)$ is Gaussian. Figure 2 gives it for $X(t)$ uniform. The two estimates are surprisingly close, suggesting that the procedure may be robust.

In a part of the study analagous to the estimation of the function $g(\cdot)$, the nonlinearity was estimated and found to have a threshold character. Sampling fluctuations of the estimates were estimated by splitting the data up into a number of segments and estimating the parameters separately for each segment, rather than attempting to use any of the procedures of Section 3.

6. A PARTIAL CONVERSE

The development of the results of this paper made essential use of an assumption of normality for the xs. A question of some interest is whether there is any other distribution leading to similar results. The following theorem indicates that normality is required for regressor variates of one important type.

Theorem 3. Let the p*-variate* x *be of the form* $a + b\varepsilon$ *with* a *a* p*-vector,* b $p\times p$*, and nonsingular, and the entries of* ε *independent, identically distributed of mean* 0*, and finite nonzero variance. Let* Σ *denote the covariance matrix of* x*. Suppose* $p > 1$ *and*

$$\mathrm{cov}\{g(\beta x),x\} = k\beta\Sigma \tag{6.1}$$

for some $\beta \neq 0$*, some* $k \neq 0$*, and all* $g(u)$ *of the form* $\exp\{itu\}$*,* t *real-valued. Then,* x *is normally distributed.*

This theorem is proved in the Appendix. This result is far from a converse; however, it does suggest strongly that normal regressors will prove the most useful.

7. CONCLUDING REMARKS

So far, the work of this paper has been predicated on the assumption (1.4) of a model with an additive error. When the xs were Gaussian and independent of the error, this model led to the relationship $\mathrm{cov}\{y,x\} = k\beta\ \mathrm{var}\ x$, on which the estimation procedure proposed was based. In fact, this relationship follows from the weaker assumption that

$$E\{y|x\} = g(\alpha + \beta x)\ . \tag{7.1}$$

The estimation procedure is now seen to be of use in a broader class of situations. Consider, for example, the binomial response (or regression) model. Here $y = 1$ or 0 with

$$\mathrm{Prob}\{y = 1|x\} = g(\alpha + \beta x) \tag{7.2}$$

with $g(\cdot)$ normal for the probit model and logistic for the logit model. From what has gone before in the paper, one sees that if $g(\cdot)$ is unknown and x is Gaussian, then β may be estimated, up to a constant of proportionality, by ordinary least squares. As a second example, consider the Cox (1972) model of proportional hazards. This involves a random variate y (a survival time), and associated covariates x, with

$$\mathrm{Prob}\{y \le t|x\} = 1 - [1 - F_0(t)]^{\exp\{\beta x\}} ,$$

$F_0(\cdot)$ being an unknown c·d·f. (This class of models is sometimes referred to as the class of Lehmann alternatives, introduced in Lehmann, 1953.) It is clear that, when the expected value exists, $E\{y|x\} = g(\beta x)$, for some $g(\cdot)$. If x is taken to be Gaussian and associated ys recorded, then the procedure of this paper allows the estimation of $k\beta$. As a final example, one has linear regression with a censored dependent variate; for example,

$$\begin{aligned} y &= \alpha + \beta x + \varepsilon && \text{if the right-hand side is nonnegative} \\ &= 0 && \text{otherwise} \end{aligned}$$

Such models are discussed in Green (1981), Nelson (1981), and references given therein. It is clear that $E\{y|x\} = g(\alpha + \beta x)$ and that ordinary least squares estimates are of use in the Gaussian case once again.

APPENDIX

Proof of Theorem 1. By writing expression (1.3) in the form (3.4), without the $1/n$'s one has $[\hat{\mu}\ \hat{\beta}]$ of the form of the statistic $\beta(n)$ considered on page 1219 of Freedman (1981). His result gives the asymptotic normality of $\hat{\beta}$. His expression for the asymptotic covariance matrix may be manipulated to give (3.3).

Proof of Theorem 2. Expand the equations (1.3) to the form (3.4) once again, that is, to the form of the usual normal equations of multiple regression. If $\hat{\Theta} = [\hat{\mu}\ \hat{\beta}]$ and $\Theta_X = E\{\hat{\Theta}|X\}$, this gives

$$(\hat{\Theta} - \Theta_X) \sum_j \begin{bmatrix} 1 \\ x_j \end{bmatrix} [1\ x_j^{\tau}] = \sum_j \varepsilon_j [1\ x_j^{\tau}] .$$

This corresponds to standard multiple regression with the regression coefficient 0. That $\hat{\Theta} - \Theta_X$ is asymptotically normal under the stated conditions of the xs and εs is shown in Miller (1974).

Proof of Lemma 2. Taking $\sigma = 0$ in Theorem 1, it follows that $\sqrt{n}(\mu_n(X) - k\beta)$ is asymptotically normal with mean 0 and covariance matrix (4.5). That $\mu_n(X)$ then has the representation (4.4) follows from a theorem of Skorokhod (1956) (see also Wichura, 1970).

Proof of Theorem 3. One can assume $a = E\{x\} = 0$. Then, from (6.1),

$$E\{g(\beta x)x^{\tau}\} = kE\{\beta x x^{\tau}\} \ . \qquad (A.1)$$

Set $L_p = \beta x$ and $L_j = \gamma_j x$ with the γ_j chosen so that $L_1, \ldots, L_p$ are mutually uncorrelated. Multiplying (A.1) by γ_j^{τ}, one has

$$E\{g(\beta x)L_j\} = kE\{L_p L_j\} = 0$$

and so $E\{g(L_p)L_j\} = 0$. From Lemma 1.1.1 of Kagan et al (1973), this last gives $E\{L_j | L_p\} = 0$. That x is necessarily normal now follows from Theorem 5.5.3 of Kagan et al or Theorem 2 of Cacoullos (1967).

REFERENCES

ARVESEN, J.N. (1969), "Jackknifing U-Statistics," *Annals of Mathematical Statistics*, 40, 2076-2100.

BOX, G.E.P. (1966), "Use and Abuse of Regression," *Technometrics*, 8, 625-629.

BRILLINGER, D.R. (1977), "The Identification of a Particular Nonlinear Time Series System," *Biometrika*, 64, 509-515.

______________, and SEGUNDO, J.P. (1979), "Empirical Examination of the Threshold Model of Neuron Firing," *Biological Cybernetics*, 35, 213-220.

CACOULLOS, T. (1967), "Characterizations of Normality by Constant Regression of Linear Statistics on Another Linear Statistic," *Annals of Mathematical Statitistics*, 38, 1894-1898.

COLLOMB, G. (1981), "Estimation Non-Parametrique de la Regression: Revue Bibliographique," *International Statistical Review*, 49, 75-94.

COX, D.R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society*, Ser. B, 34, 187-220.

FREEDMAN, D.A. (1981), "Bootstrapping Regression Models," *Annals of Statistics*, 9, 1218-1228.

GREENE, W.H. (1981), "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," *Econometrica*, 49, 505-513.

HALTON, J.H. (1960), "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals," *Numerische Mathematik*, 2, 84-90.

__________ and SMITH, G.B. (1964), "Algorithm 247: Radical-Inverse Quasi-Random Point Sequence |G5|," *Communications Association of Computing Machinery*, 7, 701-702.

KAGAN, A.M., LINNIK, Yu.V., and RAO, C.R. (1973), *Characterization Problems in Mathematical Statistics*, New York: Wiley.

LEHMANN, E.L. (1953), "The Power of Rank Tests," *Annals of Mathematical Statistics*, 24, 23-43.

__________ (1981), "An Interpretation of Completeness and Basu's Theorem," *Journal of the American Statistical Association*, 76, 335-340.

MILLER, R.G. (1974), "An Unbalanced Jackknife," *Annals of Mathematical Statistics*, 2, 880-891.

MOSTELLER, F.M., and TUKEY, J.W. (1977), *Data Analysis and Regression*, Reading, Ma.: Addison-Wesley.

NELSON, F.D. (1981), "A Test for Misspecification in the Censored Normal Model," *Econometrica*, 49, 1317-1330.

NIEDERREITER, H. (1978), "Quasi-Monte Carlo Methods and Pseudo-Random Numbers," *Bulletin American Mathematical Society*, 84, 957-1041.

SKOROKHOD, A.V. (1956), "Limit Theorems for Stochastic Processes," *Theory of Probability and Its Applications*, 1, 261-290.

STEIN, C.M. (1981), "Estimation of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 9, 1135-1151.

WICHURA, M. (1970), "On the Construction of Almost Uniformly Convergent Random Variables With Given Weakly Convergent Image Laws," *Annals of Mathematical Statistics*, 41, 284-291.

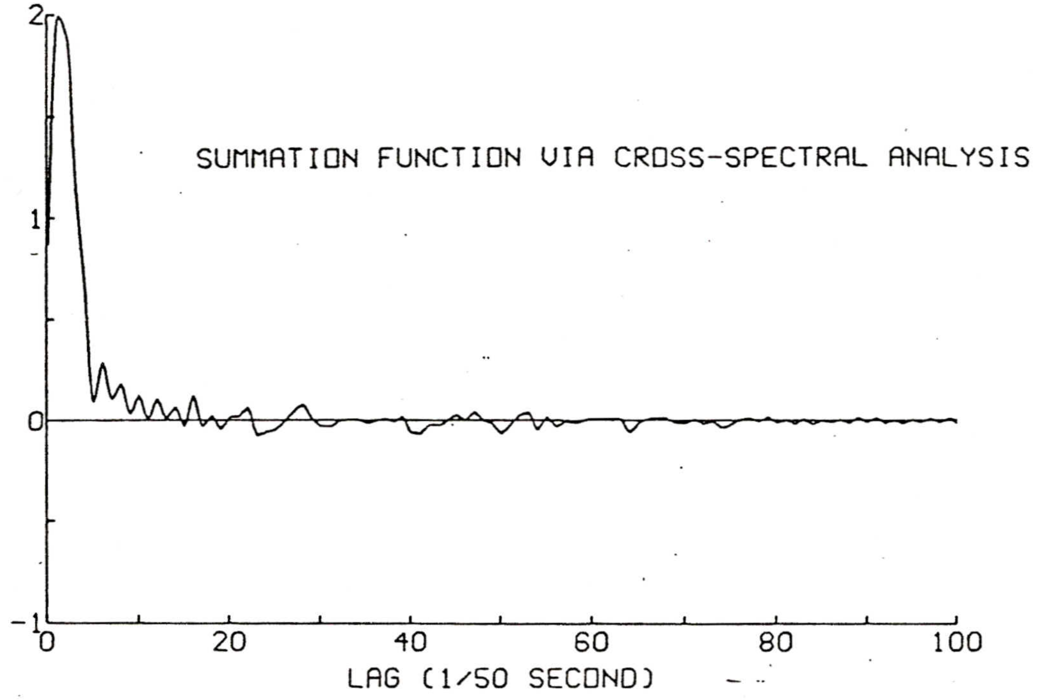

Figure 1. Estimate of the summation function $a(\cdot)$ obtained with Gaussian input.

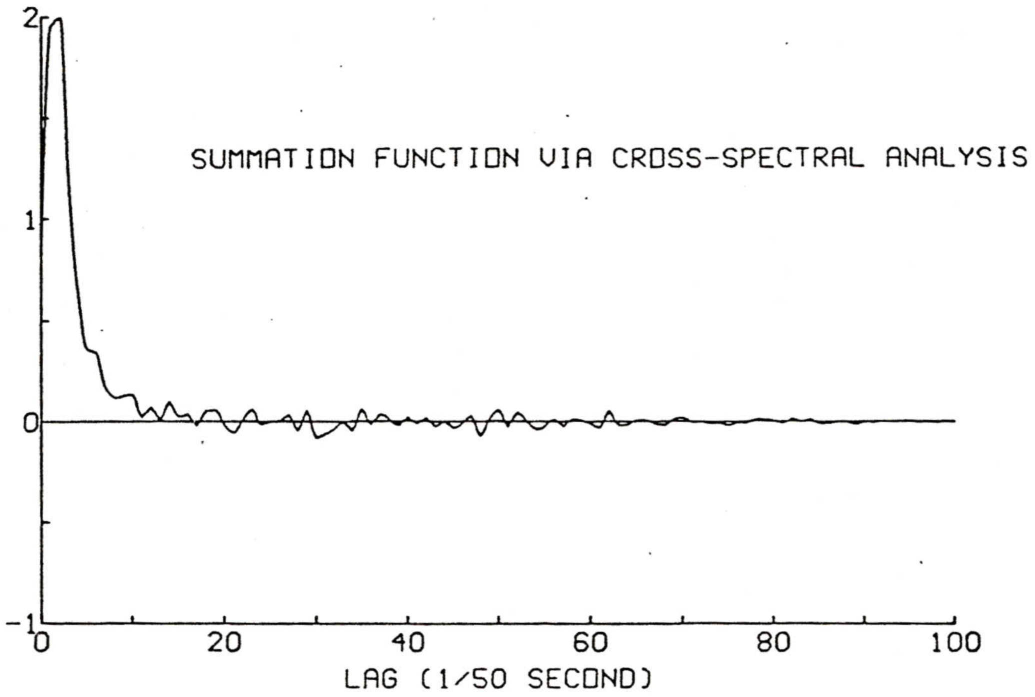

Figure 2. Estimate of the summation function a(•) obtained with uniform input.

29.

A FESTSCHRIFT FOR ERICH L. LEHMANN

In Honor of His Sixty-Fifth Birthday

Editors
Peter J. Bickel
Kjell A. Doksum
J. L. Hodges, Jr.
University of California, Berkeley

Wadsworth International Group
Belmont, California
A Division of Wadsworth, Inc.

(1982)