Econometric Theory, **10**, 1994, 116–129. Printed in the United States of America.

SYMMETRY, REGRESSION DESIGN, AND SAMPLING DISTRIBUTIONS

ANDREW CHESHER
AND
SIMON PETERS
University of Bristol

When values of regressors are symmetrically disposed, many *M*-estimators in a wide class of models have a reflection property, namely, that as the signs of the coefficients on regressors are reversed, their estimators' sampling distribution is reflected about the origin. When the coefficients are zero, sign reversal can have no effect. So in this case, the sampling distribution of regression coefficient estimators is symmetric about zero, the estimators are median unbiased and, when moments exist, the estimators are exactly uncorrelated with estimators of other parameters. The result is unusual in that it does not require response variates to have symmetric conditional distributions. It demonstrates the potential importance of covariate design in determining the distributions of estimators, and it is useful in designing and interpreting Monte Carlo experiments. The result is illustrated by a Monte Carlo experiment in which maximum likelihood and symmetrically censored least-squares estimators are calculated for small samples from a censored normal linear regression, Tobit, model.

1. INTRODUCTION

Since exact distributions of econometric estimators are often hard to derive, Monte Carlo experiments are frequently used to study the behavior of estimators and the quality of approximations to their sampling distributions. Since most econometric models involve covariates, it is necessary to specify covariate designs when a Monte Carlo experiment is conducted. This aspect of Monte Carlo experimentation is rarely given prominence when experiments are reported, perhaps because it is believed that covariate design has a relatively minor influence on the relevant properties of estimators.

In fact, covariate design, in conjunction with parameter values, can have a spectacular effect on the shapes of the exact distributions of estimators

Financial support was provided by ESRC grant number B000232150. An earlier version of this paper was circulated in 1988. We are grateful to Sir David Cox, Brendon McCabe, Peter Phillips, and Richard Smith and to two anonymous referees for helpful comments.

0266-4666/94 $5.00 + .00

whose first-order asymptotic approximate distributions have shapes that are invariant under changes in covariate design. Consequently, Monte Carlo experiments should be designed so as to reveal the impact of covariate design. Unfortunately, in the great majority of reported experiments, few designs are studied and in many studies only one covariate design is used. This can severely limit the applicability of the results of these experiments. The results given here demonstrate the importance of covariate design and provide information concerning the nature of covariate design effects.

The main result of this paper is as follows. Suppose that a covariate design is symmetrically disposed around a central point. Let X be a vector of covariates and let $X = x^0$ denote this central point, which in some cases may not itself be a point in the design. In a symmetric design, each point in the design with a nonzero value, $X = x - x^0$, can be matched with another with the value $X = -(x - x^0)$. When a covariate design is symmetric in this sense, reversal of the signs of regression-type coefficients associated with X causes the sampling distributions of a wide class of coefficient estimators to be reflected about the origin. This reflection also occurs when there are other, asymmetrically disposed covariates, Z, as long as each pair of points with $X = \pm(x - x^0)$ is associated with a common value of Z.

In the special case in which regression coefficients are zero, reversing the signs of regression coefficients can have no effect on the distributions of estimators. Consequently, when designs are symmetric and regression coefficients are zero, the sampling distributions of many estimators are symmetric about zero and so the estimators are median unbiased. Cases in which regression coefficients are zero are of particular interest because they arise when considering the null distributions of Wald and score test statistics to detect omitted regressors.

Asymmetric designs do not necessarily result in asymmetric sampling distributions. For example, the distribution of the least-squares estimator is symmetric whenever the conditional distribution of the response variate is symmetric, regardless of the covariate design. However, there are commonly used models and estimators for which an asymmetric design does cause sampling distributions to be asymmetric and the effect can be dramatic, as the example of Section 3 involving the maximum likelihood Tobit estimator shows.

The symmetry condition on the covariate design is restrictive, but very many Monte Carlo experiments reported in the literature use designs that satisfy the condition. Examples are designs in which covariates' values are chosen to be spaced at equal intervals, or as expected order statistics from symmetric distributions. The reflection property allows the results of experiments using such covariate designs to be extended to new values of regression parameters. Further, it implies that where a Monte Carlo experiment based on a symmetric covariate design generates skewed sampling distributions, the skewness can be eliminated by setting regression coefficients to zero, and reversed by reversing their signs. There are other incidental uses of the result. For example, it provides a useful check on complicated calculations such as those involved in developing asymptotic expansions of certain econometric estimators (see, for example, Chesher, Peters, and Spady [4]).

The reflection result is stated and proved in Section 2. The result is unusual because unlike the result concerning the symmetry of the sampling distribution of the seemingly unrelated regression equation estimator described in Kakwani [8], the many related results described by Andrews [1] and the reflection results given by Cryer, Nankervis, and Savin [6], there is *no* assumption of symmetry in the distribution of the *response* variates.

Section 3 illustrates the results of this paper and shows the potential magnitude of design effects. Monte Carlo estimates of the sampling distributions of maximum likelihood and symmetrically censored least-squares estimators in a left censored linear regression (Tobit) model are presented. These demonstrate the substantial skewness in finite sample distributions that can be induced solely by moving from a symmetric to an asymmetric covariate design. They also show that substantial skewness can be induced even when the covariate design is symmetric merely by moving regression coefficients away from zero.

The results of this paper concern sampling distributions of estimators conditional on covariate values. They are especially relevant to the design and analysis of Monte Carlo experiments where it is common to find fixed covariate designs and frequently the symmetric designs studied here. In applied econometric work, covariate designs are usually not chosen purposively and symmetric designs are rare. However, even though many researchers will never work with symmetric designs, the results of this paper are relevant to them because they show how sensitive the exact finite sample distributions of commonly used econometric estimators can be to covariate design and parameter values.

There are cases in applied work when covariate values are sampled from symmetric or nearly symmetric distributions, for example, when an additive central limit theorem applies to the process generating the covariates. Then a result analogous to the one given here can be useful (see Chesher [3]), namely, that reversal of the sign of regression coefficients causes sampling distributions of their estimators *marginal with respect to the covariates* to be reflected around the origin. This result is also relevant to the interpretation of Monte Carlo experiments in which covariate values are sampled anew at each replication.

2. THE REFLECTION PROPERTY

First, a class of covariate designs is defined. Then a class of models for the conditional distribution of a response variate given values of covariates is introduced and a class of estimators is described. Finally, the reflection property is stated and proved.

2.1. Covariate Designs

Two types of covariate vectors are distinguished. One type, X, has a symmetric design, reflected about a central point, $X = x^0$. The other, Z, has a replicated design, taking identical values as the X covariates are reflected. This means that the n covariate values can be labeled in such a way that

$$\tilde{x}_i = -\tilde{x}_{n+1-i}, \qquad z_i = z_{n+1-i}, \qquad i = 1, \ldots, [n]/2,$$

where $\tilde{x}_i = x_i - x^0$, and $[n] = n + 1$ if n is odd and $[n] = n$ if n is even. In some applications, the covariates Z will be absent, or constant and equal to 1.

Two types of parameters are distinguished: those attached to the symmetrically disposed covariates, elements of a matrix, β, conformable with X; and the remaining parameters, elements of a matrix, δ, which may be associated with other covariates or appear as "nuisance parameters," perhaps indexing scale or distributional shape.

2.2. The Distribution of Response Variates

The response variates associated with the n points in the covariate design are denoted by $Y_1, \ldots, Y_n$. They may be vector valued. They are assumed to be mutually independently distributed given the vectors of values of covariates $x_1, \ldots, x_n$, and $z_1, \ldots, z_n$, with proper conditional distribution functions: $F_i(y_i | \tilde{x}_i, z_i, \tilde{x}_i\beta, \delta)$, $i = 1, \ldots, n$. As before, $\tilde{x}_i = x_i - x^0$ is the ith design point expressed as a deviation from the center point for the design. The distribution functions may depend upon the center point of the design, x^0, but this is not made explicit in the notation.

It is essential that β appear only in the distribution functions through the matrix product $(x - x^0)\beta = \tilde{x}\beta$. The symmetrically disposed covariates X may influence the distribution function in other ways, but if they do then the distribution functions must be even functions of x in the sense that

$$F_i(y | \tilde{x}, z, c, \delta) = F_i(y | -\tilde{x}, z, c, \delta) \qquad \text{for all } i,\ y,\ x,\ z,\ c, \text{ and } \delta,$$

where c is a potential value of $\tilde{x}\beta$.

In many cases, the distribution functions will not vary with i, but if they do, then for all values of their arguments, they must satisfy

$$F_i(y | \tilde{x}, z, \tilde{x}\beta, \delta) = F_{n+1-i}(y | \tilde{x}, z, \tilde{x}\beta, \delta), \qquad i = 1, \ldots, [n]/2.$$

The assumptions set out above encompass a very wide class of models. The response variates, Y_i, can be discrete, continuous, or mixed, so models for censored and grouped data are included. Since the Y_i's can be vector valued, multivariate models such as econometric simultaneous equations models are included. There are no restrictions on the way in which a subset of the covariates, Z, affects the response variate. The covariates X have to affect the response variate through $(X - x^0)\beta$, but they can also have other effects. For example, censored heteroskedastic regression models with $Y^* = X\beta + u$, $Y = \max(Y^*,0)$, $\operatorname{var}(u \mid X = x) = g(x,\delta)$, are included as long as $g(x,\delta)$ is an even function of $x - x^0$. Many models involving covariates that appear in the econometric and statistical literature are contained in the class of models defined by the assumptions set out above.

2.3. A Class of Estimators

We consider M-estimators (Huber [7]), $\hat{\beta}$ and $\hat{\delta}$, which are unique solutions to the estimating equations:

$$\sum_{i=1}^{n} \psi_{j,i}(y_i,\tilde{x}_i,z_i,\tilde{x}_i\hat{\beta},\hat{\delta}) = 0, \qquad j = 1,\ldots,J,$$

where J is the total number of parameters. The estimating equations may also depend on x^0, but this is not made explicit in the notation.

In many cases, the functions $\psi_{j,i}$ will not vary with i, but if they do, then for all values of their arguments, they must satisfy

$$\psi_{j,i}(y,\tilde{x},z,\tilde{x}\beta,\delta) = \psi_{j,n+1-i}(y,\tilde{x},z,\tilde{x}\beta,\delta), \qquad i = 1,\ldots,[n]/2.$$

In many cases, the estimating equations will depend on the symmetrically disposed covariates only through $\tilde{x}_i\beta$. If the covariates have other influences, then for some set of fixed nonzero constants, $\lambda_1,\ldots,\lambda_J$, and for all j, y, x, z, c, and δ, where c is a potential value of $\tilde{x}\beta$, the following condition must hold.

$$\psi_{j,i}(y,\tilde{x},z,c,\delta) = \lambda_j\psi_{j,i}(y,-\tilde{x},z,c,\delta), \ i = 1,\ldots,[n]/2.$$

If in the conditional distribution functions, β does appear only in conjunction with x, as required above, then for many estimators—maximum likelihood estimators, for example—β will necessarily appear in the estimating equations only in conjunction with x.

A very wide class of estimators is encompassed by these assumptions. It contains many M-estimators in well- and misspecified models, including maximum likelihood estimators, linear and nonlinear two- and three-stage least-squares estimators, and least absolute deviation estimators, and it includes semiparametric estimators like Cox's [5] estimator for proportional hazard models and Powell's [10] symmetrically trimmed and symmetrically censored least-squares estimators.

The reflection property is given in the following theorem.

THEOREM. *Under the assumptions set out above, the sampling distribution of the estimator $\hat{\beta}$ when $\beta = -b$ is a reflection around the origin of the sampling distribution of $\hat{\beta}$ when $\beta = +b$ and the sampling distribution of $\hat{\delta}$ is invariant under sign changes in β, in the sense that for all values of β and δ and sets of matrix pairs, A, conditional on the values of the covariates*:

$$P[\{\hat{\beta},\hat{\delta}\} \in A \,|\, \beta = +b,\ \delta = d] = P[\{-\hat{\beta},\hat{\delta}\} \in A \,|\, \beta = -b,\ \delta = d].$$

A proof of the theorem follows. Throughout, probabilities are conditional on the values taken by the covariates.

Proof. Let s denote a realization of $Y_1, \ldots, Y_n$ and let $\hat{\beta}(s)$ and $\hat{\delta}(s)$ denote the solution to the estimating equations at s. Define the operator $\mathbb{C}(s)$ which changes a realization, s, by interchanging the ith and $(n + 1 - i)$th values of y, $i = 1, \ldots, [n]/2$. When the operator $\mathbb{C}(\cdot)$ is applied to a set of realizations, it acts as just described on each member of the set. We first show that $\hat{\beta}(s) = -\hat{\beta}(\mathbb{C}(s))$ and $\hat{\delta}(s) = \hat{\delta}(\mathbb{C}(s))$.

Write the estimating equations at the realization s in the following manner, which embodies the symmetry property of the covariate design.

$$\begin{aligned}\sum_{i=1}^{[n-1]/2} & \psi_{j,i}(y_i,\tilde{x}_i,z_i,\tilde{x}_i\hat{\beta}(s),\hat{\delta}(s)) \\ & + 1_{\{n\ \mathrm{odd}\}}\psi_{j,[n]/2}(y_{[n]/2},\tilde{x}_{[n]/2},z_{[n]/2},\tilde{x}_{[n]/2}\hat{\beta}(s),\hat{\delta}(s)) \\ & + \sum_{i=1}^{[n-1]/2} \psi_{j,i}(y_{n+1-i},-\tilde{x}_i,z_i,-\tilde{x}_i\hat{\beta}(s),\hat{\delta}(s)) = 0, \qquad j = 1, \ldots, J. \qquad \textbf{(1)}\end{aligned}$$

The second term appears only if the sample size is odd.

At the associated realization $\mathbb{C}(s)$, the estimating equations are

$$\begin{aligned}\sum_{i=1}^{[n-1]/2} & \psi_{j,i}(y_{n+1-i},\tilde{x}_i,z_i,\tilde{x}_i\hat{\beta}(\mathbb{C}(s)),\hat{\delta}(\mathbb{C}(s))) \\ & + 1_{\{n\ \mathrm{odd}\}}\psi_{j,[n]/2}(y_{[n]/2},\tilde{x}_{[n]/2},z_{[n]/2},\tilde{x}_{[n]/2}\hat{\beta}(\mathbb{C}(s)),\hat{\delta}(\mathbb{C}(s))) \\ & + \sum_{i=1}^{[n-1]/2} \psi_{j,i}(y_i,-\tilde{x}_i,z_i,-\tilde{x}_i\hat{\beta}(\mathbb{C}(s)),\hat{\delta}(\mathbb{C}(s))) = 0, \qquad j = 1, \ldots, J. \\ & \qquad \textbf{(2)}\end{aligned}$$

The assumptions concerning the estimating equations ensure that (2) is solved at $\hat{\beta}(\mathbb{C}(s)) = -\hat{\beta}(s)$ and $\hat{\delta}(\mathbb{C}(s)) = \hat{\delta}(s)$ because if these values are substituted in (2) then it resembles (1) except that the order of the two summations is reversed and there will be innocuous scale factors present if there are factors λ_j which are not equal to one.

Let A_+ and A_- be the sets of realizations of $Y_1, \ldots, Y_n$, for which, respectively, $\{\hat{\beta},\hat{\delta}\}$ and $\{-\hat{\beta},\hat{\delta}\}$ fall in the set of matrix pairs, A. The argument above implies that $A_- = \mathbb{C}(A_+)$. The assumption concerning the conditional distribution function of the response variates ensures that for any set of realizations, say Z,

$$P[Z \mid \beta = +b,\ \delta = d] = P[\mathbb{C}(Z) \mid \beta = -b,\ \delta = d],$$

since interchanging y_i and y_{n+1-i} leaves the values taken by the distribution functions F_i unchanged once $\tilde{x}'b$ is replaced by $\tilde{x}'(-b)$. In particular,

$$P[A_+ \mid \beta = +b,\ \delta = d] = P[\mathbb{C}(A_+) \mid \beta = -b,\ \delta = d],$$

which expressed in terms of $\hat{\beta}$ and $\hat{\delta}$ is

$$P[\{\hat{\beta},\hat{\delta}\} \in A \mid \beta = +b,\ \delta = d] = P[\{-\hat{\beta},\hat{\delta}\} \in A \mid \beta = -b,\ \delta = d]. \qquad \blacksquare$$

3. DISCUSSION AND ILLUSTRATION

The theorem has some interesting implications. For example, it implies that when $\beta = 0$, $P[\hat{\beta} \in B] = P[-\hat{\beta} \in B]$ for all sets of matrices, B, so that in symmetric designs, $\hat{\beta}$ is symmetrically distributed around zero when β is in fact zero. Another implication is that when $\beta = 0$, $P[(\hat{\beta})^j\hat{\delta} \leq a] = P[(-\hat{\beta})^j\hat{\delta} \leq a]$, so that for odd values of j, the distribution of $\hat{\beta}^j\hat{\delta}$ is symmetric about zero. Setting j equal to 1 it follows that when $\beta = 0$, the covariance of $\hat{\beta}$ and $\hat{\delta}$ is zero if it exists.

The implications of the theorem are well illustrated using a Monte Carlo experiment described by Powell [10] who considers a censored regression model with covariates x_1 and x_2, in which

$$Y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$$Y = \max(Y^*,0).$$

In the experiment, $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 0$, the values of x_1 are equally spaced in an interval $[-q,q]$ chosen so that the variance of x_1 over the design is 1, and values of x_2 alternate between -1 and 1 as x_1 increases. Powell [10] performs 201 Monte Carlo replications, in each one simulating 200 realizations of Y using pseudorandom i.i.d. $N(0,1)$ errors and computes maximum likelihood (ML) and symmetrically censored least-squares (SCLS) estimates. With this number of replications, it is not possible to measure with accuracy the departure of sampling distributions from symmetry. So two larger, 5000-replication, Monte Carlo experiments were conducted. In one, ML estimates for samples of size 20 were computed. In the other, SCLS estimates in samples of size 200 were computed. The sample size was reduced to 20 for experiments involving the ML estimator so that departures from symmetry would not be masked by the operation of the central limit theorem. This was not necessary in the case of the SCLS estimator, which is difficult to compute in samples this small. In all other respects, the experiments followed Powell's [10] design. The results are summarized in Table 1.[1]

Table 1. Summary statistics, 5000 Monte Carlo replications, two covariate censored normal regression model with 50% censoring

	ML Estimators (Sample Size 20)			SCLS Estimators (Sample Size 200)		
Estimators of:	β_0	β_1	β_2	β_0	β_1	β_2
Coefficient values	0	1	0	0	1	0
Median	.0072	.9969	−.0036	−.0391	1.0368	.0017
Mean (m_1)	−.0354	1.0322	−.0069	−.1707	1.1234	−.0005
Std deviation ($m_2^{1/2}$)	.3606	.3471	.2724	.5162	.3979	.1316
Asy std deviation	.3145	.3130	.2624			
Skewness ($m_3/m_2^{3/2}$)	−1.27	0.89	−0.05	−3.61	2.64	0.00
	(.128)	(.066)	(.069)	(.355)	(.252)	(.101)
Kurtosis ($m_4/m_2^2 - 3$)	4.18	1.89	0.77	23.75	13.86	1.94
	(1.03)	(.256)	(.295)	(5.01)	(2.82)	(.410)
Correlations: $\hat\beta_0$		−.55	.01		−.96	.02
$\hat\beta_1$	−.55		−.06	−.96		−.03
$\hat\beta_2$	.01	−.06		.02	−.03	
$\hat\sigma^2$	−.40	.26	−.02			

m_i is the ith central moment across 5000 replications. "Asy std deviations" are $n^{-1/2}$ times the asymptotic standard deviations of $n^{1/2}(\hat\beta_i - \beta_i)$ developed from the information matrix under the assumption that the covariate design is replicated as the sample size, n, increases. $\hat\sigma^2$ is the ML estimator of the error variance. Figures in parentheses are jackknife estimates of the standard errors of the skewness and kurtosis measures.

The model and estimators satisfy the assumptions of the theorem and the covariate design is symmetric about $x_1 = 0$, $x_2 = 0$. Let the estimators, ML or SCLS, be denoted by $\hat\beta_1$ and $\hat\beta_2$. From the theorem, it follows that if β_1 were -1, then the joint sampling distributions of $\hat\beta_1$ and $\hat\beta_2$ would be reflections around zero of those studied by Powell for which $\beta_1 = +1$ and $\beta_2 = 0$, while the sampling distributions of the intercept estimators, $\hat\beta_0$, which can be thought of as being associated with a replicated "covariate" always equal to 1, would be unchanged. If β_1 were zero, then the joint sampling distributions of $\hat\beta_1$ and $\hat\beta_2$ would be symmetric. So the extent to which these distributions deviate from symmetry shows the amount of skewness that is caused solely by β_1 deviating from zero.

Even though β_1 is nonzero, the Monte Carlo experiment also illustrates the symmetry result given at the beginning of this section. This is because successive alternating values of the symmetrically disposed covariate x_2 are associated with almost identical values of x_1 so that values associated with $x_2 = +1$ are close to those associated with $x_2 = -1$. Consequently, the covariate x_1 is almost in the category of replicated covariates, Z, defined earlier, and the theorem leads one to expect that the sampling distributions of estimators of β_2 will be close to symmetric.

The skewness coefficients and the relative magnitudes of means and medians shown in Table 1[2] indicate that the distributions of estimators of β_0 and β_1 are, respectively, negatively and positively skewed while the distributions of estimators of β_2 show negligible skewness. Figures in parentheses are jackknife estimates of the standard errors of the estimates of the standardized cumulants. The cumulant estimates are quite variable despite the large scale of this experiment, but it is quite clear that the variations in the values of the skewness measures do reflect real differences in the shapes of the sampling distributions of the alternative estimators. The correlations between estimators of β_2 and estimators of the other parameters are very small.

Figure 1 shows quantile-quantile (QQ) plots of the ML and SCLS estimates of β_1 and β_2, relocated and scaled so that over the 5000 replications, the linearly transformed values of each estimator have zero mean and unit variance. The graphs are constructed by plotting quantiles of the Monte Carlo replicates against corresponding quantiles from the standard normal distribution, which is the first-order asymptotic approximation to these estimators' sampling distributions.

The asymmetry in the sampling distributions of the estimators of β_1 is very obvious. It would not be present were the true value of β_1 to be zero. The finite sample distributions of the estimators of β_2 deviate very little from symmetry. The distribution of the ML estimator of β_2 is slightly long tailed but it is remarkably close to a normal distribution given that the sample size is only 20 and that on average, 50% of the realizations are censored.[3] Even with a sample size of 200, the normal approximation to the distribution of the SCLS estimator is extremely poor. The estimator is very long tailed. The skewness induced by β_1 being nonzero is much greater for this estimator than for the ML estimator in much smaller samples.

Departures from symmetry can be isolated from other aspects of distributional shape and are very clearly revealed in the "symmetry plots" shown in Figures 2(a) and 2(b). These show quantiles of, respectively, ML and SCLS estimators of β_1 and β_2 expressed in standard deviation units as absolute deviations from medians, values associated with quantiles above the median plotted against values associated with corresponding quantiles below the median. Let $\tilde{\beta}_i^{(j)}, j = 1, \dots, N$ be the values of $\hat{\beta}_i$ obtained in N Monte Carlo replications, expressed in standard deviation units and arranged in ascending order, and let $\hat{\beta}^M$ be the median value obtained. The symmetry plot is generated by plotting points with coordinates

$$\{-(\tilde{\beta}_i^{(j)} - \hat{\beta}^M), (\tilde{\beta}_i^{(N-j+1)} - \hat{\beta}^M)\}, \qquad j = 1, \dots, [N-1].$$

A trail of points close to the 45° line indicates an almost symmetric distribution. Paths, respectively, above or below the 45° line indicate, respectively,

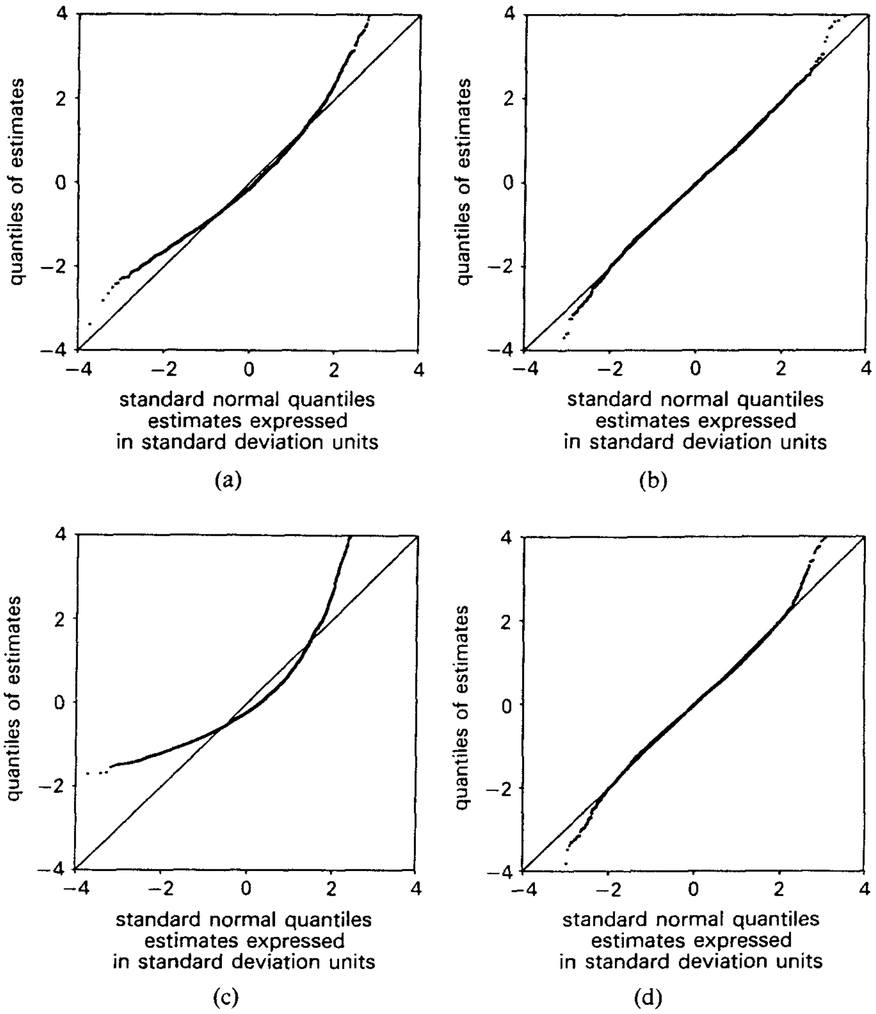

FIGURE 1. (a) ML estimator of β_1. (b) ML estimator of β_2. (c) SCLS estimator of β_1. (d) SCLS estimator of β_2.

positively or negatively skewed distributions. The positive skewness in the distributions of the estimators of β_1 is very obvious in Figure 2(a). A tiny amount of skewness is detectable in the distributions of the estimators of β_2 shown in Figure 2(b). It arises because the design for x_1 is not exactly replicated across the +1 and −1 points in the x_2 design.

So far only a symmetric design has been studied. All asymmetry in sampling distributions has arisen because parameter values deviate from zero.

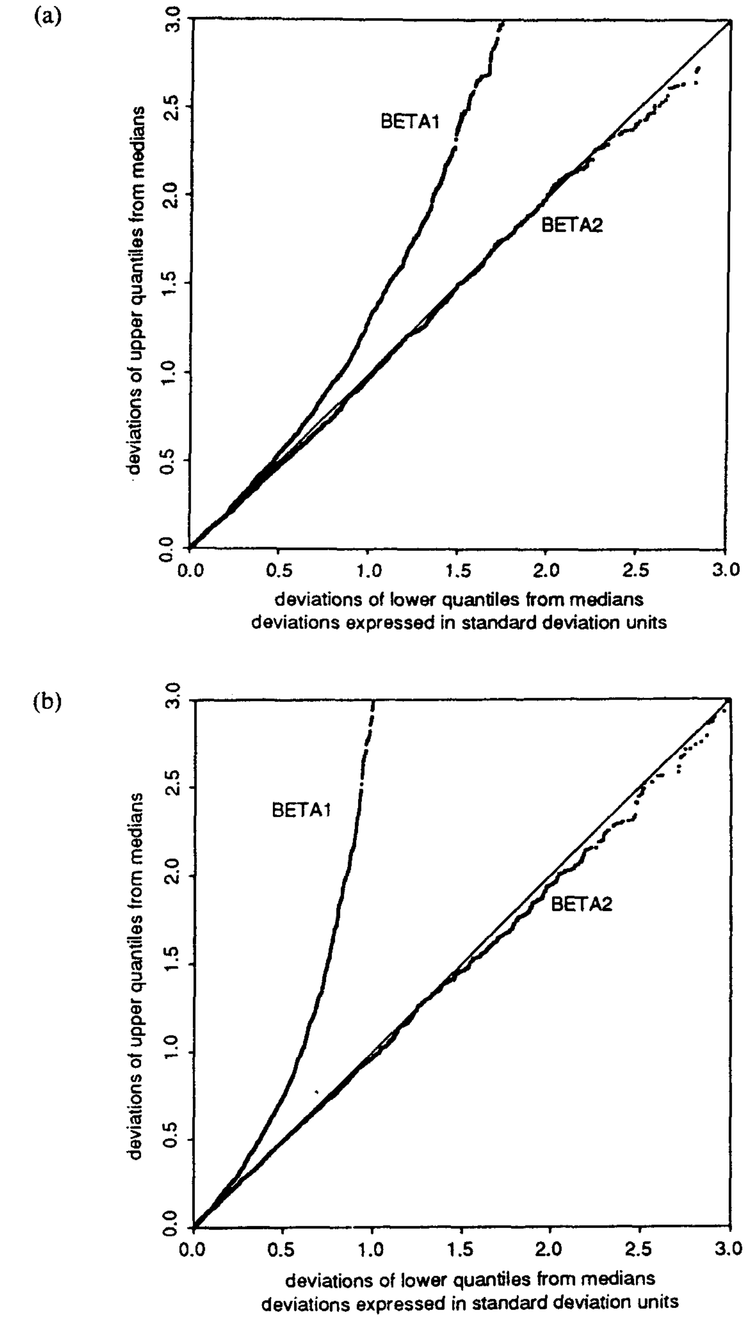

Figure 2. (a) Deviations of quantiles from medians: ML. (b) Deviations of quantiles from medians: SCLS.

What is the effect of altering the design so that it is asymmetric? To answer this question, the Monte Carlo experiment involving the ML estimator was performed again with just one change, namely, that the design for the binary covariate x_2 was altered by moving a single point so that the design became asymmetric. Specifically, the value of x_2 corresponding to the lowest value of x_1 was changed from -1 to -10. Even though the true value of β_2 is zero, this has a dramatic effect, clearly revealed in the symmetry plot shown in Figure 3. The trail of points labeled "Asymmetric $X2$" arises when this asymmetric design is used. There is clearly very substantial positive skewness.

The design for x_2 can be brought back to symmetry by pushing the x_2 value corresponding to the second lowest value of x_1 to $+10$. The result is the trail of points in Figure 3 labeled "Symmetric $X2$." Restoring symme-

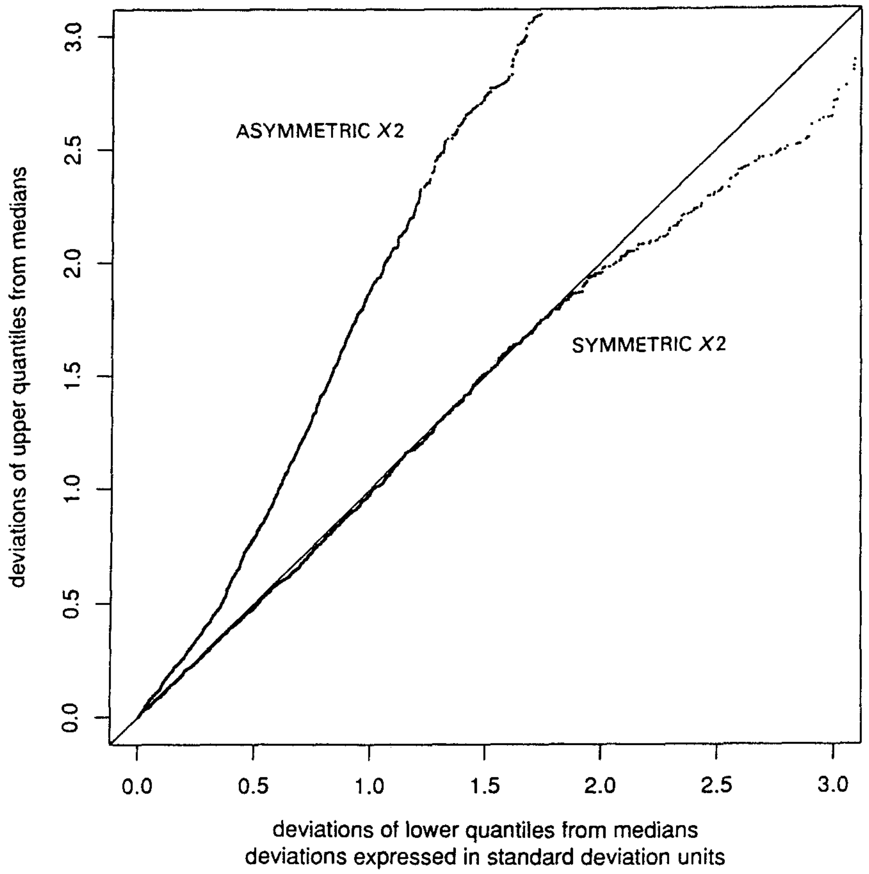

Figure 3. Deviations of quantiles from medians: ML x_2 design perturbed.

try to the design for x_2 restores approximate symmetry to the sampling distribution of the ML estimator of β_2. Again, some slight skewness remains because the x_1 design is not exactly replicated across the positive and negative points in the x_2 design.

4. CONCLUDING REMARKS

It is very common to find symmetric covariate designs in published Monte Carlo studies. For example, Moolgavkar and Venzon [9] report the results of Monte Carlo experiments examining Cox's [5] estimator for proportionate hazard models with linear relative risk, $1 + \beta z$, depending on a single covariate z. In one set of experiments, the values of the covariate are expected uniform order statistics; in another, expected normal order statistics—in each case rescaled to span the interval [0,1]. The model, estimators, and designs satisfy the assumptions of the theorem, so the reflection result applies, and when $\beta = 0$ so that the covariate is ineffective, the sampling distribution of the Cox estimator of β must be symmetric. Moolgavkar and Venzon's results with $\beta = 0$ do suggest a symmetric sampling distribution. They report a mean of 0.04, a median of -0.02, and a standard deviation of 0.38 over 1000 replications with a sample size of 100 at each replication. They remark that "distributional properties appear to be worse with increasing true value of the parameter β" (Moolgavkar and Venzon [9], p. 47). It is evident from their graphs that increasing skewness is the problem, as we would expect, given the results of this paper.

It is clear that values taken by covariates can have a major influence on finite sample properties of estimators. The point has been made before by, for example, Box and Watson [2] and Weisberg [12], yet it is rare to find Monte Carlo experiments that pay adequate attention to covariate design. Many reported Monte Carlo studies that give the impression that first-order asymptotic approximations perform well can in fact only be regarded as showing that there are covariate designs (namely, those studied) in which the approximations are adequate. Unfortunately, covariate design is an unwieldy factor to vary in a Monte Carlo experiment, and a view concerning the range and types of designs that are relevant is essential when designing a Monte Carlo study. The results of this paper can aid the choice of appropriate designs to examine.

NOTES

1. Normal pseudorandom numbers were obtained by applying Press et al.'s [11] version of Marsaglia's polar method to uniform pseudorandom numbers obtained with Wichman and Hill's [13] portable generator. Maximum likelihood estimates were calculated using the method of scoring with, as starting points, least-squares estimates obtained from uncensored data. Calculations were performed in double precision arithmetic on a Sun SPARCstation 2 running SUN-OS 4.1.2.

2. It is possible for censoring to create configurations of realizations for which the ML or the SCLS estimators of one or both coefficients are unbounded or indeterminate. For example, if all realizations associated with $x_2 = -1$ are censored but there are sufficient uncensored realizations at $x_2 = +1$, then the Tobit likelihood function is maximized at $\hat{\beta}_2 = +\infty$. In Powell's design with a sample size of 20, such configurations are quite rare. In 5000 Monte Carlo replications with a sample size of 20, the probability of finding no configurations of realizations leading to unbounded or indeterminate ML estimators is around 0.5. The corresponding probability for samples of size 200 is very close to 1. In fact, no configurations leading to indeterminate or unbounded estimators arose in the two experiments reported in Table 1. However, the figures reported there should be interpreted as applying to the sampling distributions of estimators conditional on their values being determinate and finite.

3. Symmetry is not the only feature of covariate design that influences the quality of first-order asymptotic approximations. The covariate x_2 is almost uncorrelated with x_1 and has a balanced design. If leverage points are introduced into the design, then the ML estimator of β_2 can exhibit very substantial supernormal kurtosis though if the design is symmetric, it remains almost symmetrically distributed.

REFERENCES

1. Andrews, D.W.K. A note on the unbiasedness of feasible GLS, quasimaximum likelihood, robust, adaptive and spectral estimators of the linear model. *Econometrica* 54 (1986): 687–698.
2. Box, G.E.P. & G.S. Watson. Robustness to non-normality of regression tests. *Biometrika* 49 (1962): 93–106.
3. Chesher, A.D. A reflection property of M estimators. Discussion Paper No. 90/282, Department of Economics, University of Bristol, 1990.
4. Chesher, A.D., S. Peters & R. Spady. Approximations to the distributions of heterogeneity tests in the censored normal linear regression model. Discussion Paper No. 89/240, Department of Economics, University of Bristol, 1989.
5. Cox, D.R. Regression models and life tables. *Journal of the Royal Statistical Society, Series B* 34 (1972): 187–220.
6. Cryer, J.D., J.C. Nankervis & N.E. Savin. Mirror image and invariant distributions in ARMA models. *Econometric Theory* 5 (1989): 36–52.
7. Huber, P.J. *Robust Statistical Procedures.* Philadelphia: Society for Industrial and Applied Mathematics, 1977.
8. Kakwani, N.C. The unbiasedness of Zellner's seemingly unrelated regression equation estimators. *Journal of the American Statistical Association* 62 (1967): 141–142.
9. Moolgavkar, S.H. & D.J. Venzon. Confidence regions for parameters of the proportionate hazard model: A simulation study. *Scandinavian Journal of Statistics* 14 (1987): 43–56.
10. Powell, J.L. Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* 54 (1986): 1435–1460.
11. Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press, 1986.
12. Weisberg, S. Comment on "Some large sample tests for non-normality in the linear regression model" by H. White and G.M. MacDonald. *Journal of the American Statistical Association* 75 (1980): 28–31.
13. Wichman, B.A. & I.D. Hill. Algorithm AS183: An efficient portable pseudo-random number generator. *Applied Statistics* 31 (1982): 188–190.