Forecasting Performance of Models Using the Box-Cox Transformation

David M. Smallwood and James R. Blaylock

Abstract

The authors examine the small sample properties and forecasting performance of estimators in models using the Box-Cox transformation via a Monte Carlo experiment They develop a simple estimator for the expected value of the untransformed dependent variable They show that the sign and magnitude of the transformation parameter influence the precision of the estimators and the forecasting performance These results support previous research At different values of the transformation parameter, smaller variances of the parameter estimators do not necessarily imply improved goodness of fit for the model

Keywords

Forecasting performance, transformations, Box-Cox, flexible functional forms

Economic theory usually provides few details as to the specific functional relationships among variables in econometric models Therefore, the applied research economist is often forced to choose among many competing functional forms using noneconomic criteria Flexible functional forms, because they minimize this subjective aspect of model construction, are becoming increasingly popular as a tool to discriminate among competing models' specifications One frequent approach for adding flexibility to models is to incorporate the monotonic transformation introduced by Box and Cox (*3*)[1] Models incorporating the Box-Cox transformation allow researchers to discriminate statistically among many commonly used functional forms including the log, inverse, quadratic, and linear forms (*1, 4, 6, 8, 12, 14*) However, the Box-Cox transformation places additional burdens on the researcher in terms of the complexity of estimating and interpreting model parameters compared with ordinary-least-squares models Furthermore, the small sample properties of the estimators and the forecast performance of models incorporating a transformed dependent variable are not well known (*11*)

Several papers have addressed estimation procedures and interpretation of parameters in Box-Cox models, but only Spitzer (so far as we know) has addressed the small sample properties and forecast performance Estimating parameters in models employing the Box-Cox transformation involves maximizing a complex nonlinear likelihood function Spitzer has outlined several procedures that can be used to accomplish this task (*12*) Procedures for interpreting parameters in Box-Cox models are discussed by Spitzer (*12*), Blaylock and Smallwood (*1, 2*), and others

The small sample properties of the estimators are particularly important for the applied researcher For example, how well do the estimators perform when one has only 30, or perhaps 60, observations? Can one use the standard t-test to test hypotheses about the model parameters? Although the maximum likelihood properties are well known, they apply only asymptotically, and the small sample properties are analytically intractable

Spitzer has investigated the small sample properties of the Box-Cox estimators via Monte Carlo methods (*11*) However, as he notes, his results are tempered by the small number of replications (50) per model Furthermore, he touches on forecast per-

The authors are agricultural economists with the National Economics Divison, ERS

[1]Italicized numbers in parentheses refer to items in References at the end of this article

formance only as a secondary issue [2] Thus, several important questions, such as the calculation of the expected value of the untransformed dependent variable and out-of-sample forecast performance, are not addressed

This article has two major objectives Our first objective is to expand Spitzer's Monte Carlo study (*11*) by using twice the number of replications to provide more reliable information on the small sample properties of estimators in Box-Cox type models The small sample properties of estimators have important implications for economists, and the properties of these estimators ultimately extend to the end use of the model For example, consider the importance of a parameter's variance Spitzer has suggested that, in highly nonlinear functional specifications such as the Box-Cox, the sign and magnitude of the transformation parameters may affect the estimation of the variances for all parameters in the model If so, one must exercise extreme care in performing hypothesis tests, especially if the tests imply or embody policy or program implications

Our second objective is to investigate the ability of transformed models to forecast the original (untransformed) variable and to forecast outside the sample used for estimation Forecast performance is important in evaluating flexible functional forms because a common fear is that they fit an individual sample too well, including the random peculiarities Limited numbers of observations in analyses of live data often prevent extensive testing outside the period of fit Monte Carlo studies, in contrast, provide a unique opportunity for testing this aspect of model performance The evaluation of forecasting performance in a scientifically controlled environment using simulated data provides the applied economist with valuable insights into the strengths and limitations of the Box-Cox technique

To accomplish these objectives, we conducted a Monte Carlo experiment using the general framework set forth by Spitzer (*11*) The general model consisted of three variables and five parameters including a single Box-Cox parameter The model was used to generate some 100 samples of observations which were used for estimation and forecast evaluation This was done for five alternative Box-Cox parameter values and for two sample sizes In contrast to Spitzer (*11*), each data sample contained 10 observations for use in forecast evaluation that were not used to estimate the model parameters

We discuss the Box-Cox transformation and the method of estimation, and we outline model construction and data generation We then discuss estimation and forecasting performance results and briefly summarize our research findings

Box-Cox Transformations

The Box-Cox transformation for any positive variable W is defined as

$$W(\lambda) = (W^{\lambda} - 1)/\lambda \qquad \lambda \neq 0 \tag{1}$$
$$= \ln(W), \qquad \lambda = 0$$

where λ is a parameter to be estimated The transformation is typically applied to models of the form,

$$Y_i(\lambda) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik}(\lambda) + \epsilon_i,\ i = 1, 2, \quad , N \tag{2}$$

The linear and logarithmic models are special cases of equation 2 when λ is equal to 1 and zero, respectively (*14*)

Assume that under the appropriate transformation, the ϵ_i' s are independently and normally distributed with zero mean and constant variance, that is, $N(0, \sigma^2)$ The likelihood function can then be written as

$$L(\beta, \sigma, \lambda) = (2\pi\sigma^2)^{-N/2} J \exp\ [- \sum_{i=1}^{N} \{Y_i(\lambda) - \beta_0 - \sum_{k=1}^{K} \beta_k X_{ik}(\lambda)\}^2/2\sigma^2\ ] \tag{3}$$

where J denotes the Jacobian for the transformation from $Y_i(\lambda)$ to the observed Y_i

$$J = \prod_{i=1}^{N} |\partial Y(\lambda)/\partial Y| = \prod_{i=1}^{N} Y_i^{\lambda - 1} \tag{4}$$

[2]Spitzer's method (*11*) of deriving forecasts is shown in this article to be incorrect

The log-likelihood function can be written as

$$LL = -(N/2)\ln \tilde{\sigma}^2 + (\lambda - 1)\sum_{i=1}^{N} \ln(Y_i) \qquad (5)$$

where $\tilde{\sigma}^2 = \sum_{i=1}^{N} \tilde{\epsilon}_i^2/N$ is the estimated variance of ϵ_i

We used the Fletcher-Powell algorithm with analytically computed first derivatives to maximize the log-likelihood function[3] We used the fundamental statistical relationship that the asymptotic covariance matrix of a maximum likelihood estimator is equal to the inverse of the covariance matrix of the gradient of the likelihood function when estimating the asymptotic covariance matrix of the parameters *(10)*

Zarembka has shown that the distribution of the error term cannot be strictly normal in models where the dependent variable is Box-Cox transformed because a power transformation can be applied only to positive variables *(14)*.

Specifically, if $\lambda > 0$, then $-1/\lambda < Y(\lambda) < \infty$, if $\lambda = 0$, then $-\infty < Y(\lambda) < \infty$, and if $\lambda < 0$, then $-\infty < Y(\lambda) < -1/\lambda$ Consequently, the magnitude and sign of λ affect the range of the dependent variable However, Draper and Cox *(5)*, Zarembka *(14)*, and Spitzer *(12)* have shown empirically that so long as the distribution of the error term is reasonably symmetric and the probability of large negative values of the error term is low, normality may be a good approximation

Model Construction and Data Generation

Following Spitzer *(11)*, we specified the models as

$$Y(\lambda) = 9\,0 - 1\,5X_1(\lambda) + 0\,5X_2(\lambda) + \epsilon \qquad (6)$$

where $\lambda = -1\,5, -1\,0, -0\,15, 1\,0, 1\,5$, respectively, for the five models These values of λ were selected to represent a large range of possible transformation parameters because the size and sign of the parameter may be important in determining the shape and location of the sampling distribution of the coefficient estimates and may affect the forecasting ability of the models In addition, the $X_i(\lambda)$'s were constructed such that.

$$\beta_1^2 \operatorname{var}(X_1(\lambda)) = \beta_2^2 \operatorname{var}(X_2(\lambda))$$

This condition was set so that each variable has equal importance in explaining the variance of $Y(\lambda)$

The method used to generate values of $X_1(\lambda)$ and $X_2(\lambda)$ appears in the appendix. All models were estimated for 100 samples of sizes 30 and 60 ($N = 30, 60$) An additional 10 observations were also generated for each sample for use in evaluating the out-of-sample forecasting performance The equation error term was generated from a population that was independently distributed as $N(0, \sigma^2)$, where $\sigma^2 = 0\,4263$ The value of σ^2 was chosen to yield a residual variation equal to 5 percent of the total variation in $Y(\lambda)$

Tukey *(13)* and Box and Cox *(3)* argue that the purpose of a transformation is to increase the degree of approximation to which three desirable properties for statistical analysis hold In particular, they argue that transformations may lead to a more nearly linear model, may stabilize the error variance, and/or may lead to a model for which a normally distributed error term is acceptable Of course, a transformation may increase the degree of approximation to two or more of these properties simultaneously The true models are constructed such that all three properties hold simultaneously The estimated models should, therefore, seek out the transformation parameter that stabilizes the error variance and normalizes the error distribution

The calculation of unbiased predicted values in transformed models requires that special attention be given to the error term Transformed dependent variables make predictions more difficult because one is interested in predicting the expected value of the original (untransformed) dependent variable rather than the transformed one One derives the simplest predictor of Y_i, and probably the predictor most often used, by first noting that

$$E\left[Y_i(\lambda)\right] = \beta_o + \sum_{k=1}^{K} \beta_k X_{ik}(\lambda) \qquad (7)$$

[3] Spitzer used a modified Newton technique for estimation *(11)*

and then solving for Y_i.

$$Y_i = \left[1 + \lambda \left[\beta_0 + \sum_{k=1}^{K} \beta_k X_{ik}(\lambda)\right]\right]^{1/\lambda}, \quad \lambda \neq 0$$

$$= \exp\{\beta_0 + \sum_{k=1}^{K} \beta_k X_{ik}(\lambda)\}, \quad \lambda \to 0 \qquad (8)$$

However, the expressions in equation (8) are equal to the expected value of Y_i only in the case of the linear model For $\lambda \neq 1$, the expressions are biased estimators[4] These formulas are biased because the expected value of a nonlinear function is not equal to the nonlinear inverse function of the expected value (7) In other words, the error term cannot be dropped from equation (7) before expectations are taken

A simple approximation to the expected value of the original (untransformed) dependent variable can be derived as follows First, define the model in terms of the transformed dependent variable as

$$Z_i = (Y_i^{\lambda} - 1)/\lambda = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik}(\lambda) + \epsilon_i \qquad (9)$$

and note that the original dependent variable can be expressed as

$$Y_i = F(Z_i) = (\lambda Z_i + 1)^{1/\lambda} \qquad (10)$$

where $F(Z_i)$ denotes the inverse of the Box-Cox transformation

Expressing Y_i as a second-order Taylor expansion around the expected value of the transformed dependent variable yields

$$\begin{aligned} Y_i \approx\ & F(\tilde{Z}) + (Z_i - \tilde{Z}_i)(\lambda \tilde{Z}_i + 1)^{(1-\lambda)/\lambda} \\ & + \tfrac{1}{2}(Z_i - \tilde{Z}_i)^2(1 - \lambda)(\lambda \tilde{Z}_i + 1)^{(1-\lambda)/\lambda} \end{aligned} \qquad (11)$$

where $\tilde{Z}_i = E(Z_i)$ One derives the expected value of the expression in equation (11)

$$E(Y_i) \approx F(\tilde{Z}_i) + \tfrac{1}{2}\sigma^2 (1 - \lambda)[F(\tilde{Z}_i)]^{1-2\lambda} \qquad (12)$$

by noting that the second term on the right-hand side (RHS) of equation 11 vanishes, that the expected value of $E(Z_t - \tilde{Z}_t)^2$ is the equation error variance, and that

$$[(\lambda \tilde{Z}_i + 1)^{(1-2\lambda/\lambda)}] = [(\lambda \tilde{Z}_i + 1)^{1/\lambda} \bullet (\lambda \tilde{Z}_i + 1)^{-2}]$$

$$= [F(\tilde{Z}_i)]^{1-2\lambda} \qquad (13)$$

The expression given in equation 12 differs from the simple formula of equation 8 by the second term on the RHS of equation 12 The sign of this term, which is uniquely determined by the value of λ, indicates the direction of bias involved by using equation 8 in lieu of the formula given in equation 12 A negative bias is generally present if $\lambda < 1$, and a positive bias is present if $\lambda > 1$

We examine the small sample performance of the model parameters using a variety of performance statistics to measure bias and variation. For each parameter estimated in equation (6), let $\tilde{\theta}_i$ be the i-th sample value of the parameter, let $\sigma(\theta_i)$ be the asymptotic standard deviation of θ_i, and define the following·

Mean bias = $\sum \tilde{\theta}_i/N - \theta$,

Mean absolute bias = $\sum |\tilde{\theta}_i - \theta| /N$,

Root mean square error = $[\sum (\tilde{\theta}_i - \theta)^2/N]^{1/2}$, and

Mean asymptotic standard deviation = $\sum \sigma(\tilde{\theta}_i)/N$

We use these statistics to evaluate the forecasting performance of the various models by assuming that $\tilde{\theta}$ represents the forecast value and θ represents the true value of the observation to be predicted

Estimation and Forecast Performance

Table 1 shows mean bias (MB) and mean absolute bias (MAB) statistics for the coefficient estimates of the alternative simulations The most striking result is perhaps the remarkable similarity between our results and those of Spitzer (*11*) Like Spitzer, we find that, except for β_0 in models with $\lambda > 0$, the MB's are relatively small and do not appear to indicate systematic under- or overestimation of para-

[4]Elasticity formulas frequently employed in studies using the transformation-of-variables technique are also in error because they are based on the same erroneous assumption about the expected value of the dependent variable

Table 1—Mean bias (MB) and mean absolute bias (MAB) of parameter estimates

λ	N	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\lambda}$	$\frac{MAB(\lambda)}{\lambda}$
	Sample size			*Estimates*		
MB						
−1 5	30	−0 057	0 024	0 042	0 046	NA
	60	− 245	012	035	044	NA
−1 0	30	− 085	009	040	025	NA
	60	− 095	016	020	017	NA
−0 15	30	− 021	− 014	025	−001	NA
	60	− 118	− 013	016	− 000	NA
1 0	30	3 667	− 154	003	− 017	NA
	60	2 026	− 117	− 005	020	NA
1 5	30	4 394	− 156	002	027	NA
	60	2 302	− 116	− 006	030	NA
MAB						
−1 5	30	952	061	092	124	0 083
	60	782	045	075	102	068
−1 0	30	914	064	106	098	068
	60	650	042	067	065	065
−0 15	30	1 538	075	137	029	193
	60	1 187	046	100	020	133
1 0	30	7 071	521	079	233	233
	60	4 203	345	045	146	146
1 5	30	7 884	517	076	334	223
	60	4 575	342	043	210	140

NA =Not applicable

meters However, the MAB's for β_0 and β_1 in models with positive λ's are several times larger than their counterparts in models with negative λ's The MAB's for all model parameter estimates decline with increased sample size Coupled with similar findings by Spitzer, this decline indicates the estimates are consistent The MAB of λ as a percentage of λ generally increases as λ increases, indicating that the variance of λ increases as λ increases Spitzer also noted this phenomenon Therefore, the problem seems not to be one of a small number of sample replications

The MB and MAB statistics show that (1) parameter estimates are unbiased and consistent, (2) models with positive λ's perform less well than other models in terms of MB and MAB statistics, (3) the variance of λ seems to increase as λ increases, and (4) our results are similar to those of Spitzer, which is comforting in terms of the reliability of the test statistics and because we obtained our results using different estimation techniques

Table 2 presents the root mean square error (RMSE) and mean asymptotic standard deviation (MASD) statistics for the parameter estimates If parameter estimation bias is small, the MASD's should be good approximations to the RMSE's (that is, the ratio of the MASD and RMSE for a parameter should approach 1 as sample size increases) The MASD's and RMSE's for all models decline as sample size increases, and their ratio is virtually equal to 1 for N = 60 in all cases This finding suggests that the estimated variances are consistent However, the statistics for positive λ are many times larger than those for the models with nega-

Table 2—Root mean square error (RMSE) and mean absolute standard deviation (MASD) of parameter estimates

λ	N	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\lambda}$
	Sample size	Estimates			
RMSE					
−1 5	30	1 207	0 075	0 116	0 154
	60	963	055	099	130
−1 0	30	1 152	080	156	129
	60	810	054	088	084
−0 15	30	1 978	091	177	036
	60	1 531	058	132	026
1 0	30	11 081	693	102	301
	60	6 257	467	061	191
1 5	30	12 805	690	098	431
	60	6 952	464	059	275
MASD					
−1 5	30	1 640	091	171	232
	60	1 037	061	102	142
−1 0	30	1 397	098	182	172
	60	892	061	106	104
−0 15	30	2 596	100	243	049
	60	1 562	061	145	029
1 0	30	13 378	882	119	370
	60	6 387	491	066	209
1 5	30	15 152	870	114	530
	60	6 990	485	064	300

tive λ, except for the statistics for β_2, which follow no obvious pattern The results indicate that estimates from models with $\lambda < 0$ tend to be more precise

One indicator of the concentration of the parameter estimates around the true parameter is the percent age of the estimates within ±20 percent of the true parameter (table 3) The results strongly indicate that models with a negative λ perform better An examination of table 3 reveals that a larger percentage of the parameter estimates fall within the 20-percent range as λ decreases and as sample size increases The exception is β_2, which shows no clear relationship with λ However, as sample size increases, all parameters become more highly concentrated around the true parameter values Thus, the parameter estimates obtained from larger samples and models with smaller λ are more precise than other models

One must be cautious when using standard t-tests for hypothesis testing To examine this issue further, we constructed two hypotheses The first is a true hypothesis

$$H_o \quad \beta_j = \beta_{jo}$$

where β_{jo} is the true value of β_j Using a two-tailed test at the 0 05 significance level, we would expect to reject 5 percent of these hypotheses Using the same procedures, we also tested the following false hypothesis

$$H_o \quad \beta_j = 0$$

This test shows the power of the t-test (table 4)

The true hypothesis was rejected in 5 percent or less of the replications for models with a negative λ The results were mixed for models with positive λ's,

Table 3—Percentage of estimates within 20 percent of true parameters

λ	N	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\lambda}$
	Sample size		*Percent*		
−1.5	30	87	100	60	95
	60	95	100	74	97
−1.0	30	88	100	59	91
	60	98	100	79	96
−0.15	30	66	100	45	62
	60	77	100	65	73
1.0	30	22	37	70	54
	60	33	54	89	74
1.5	30	21	38	70	55
	60	32	56	89	75

Table 4—Rejections (R) and acceptances (A) of hypotheses[1]

λ	N	$\tilde{\beta}_0$		$\tilde{\beta}_1$		$\tilde{\beta}_2$		$\tilde{\lambda}$	
		R	A	R	A	R	A	R	A
	Sample size				*Numbers*				
−1.5	30	5	0	2	0	1	6	2	0
	60	5	0	5	0	3	0	5	0
−1.0	30	2	0	3	0	1	8	4	1
	60	5	0	4	0	2	0	2	0
−0.15	30	3	0	5	0	4	37	2	11
	60	5	0	4	0	5	0	4	0
1.0	30	9	100	6	54	1	1	3	24
	60	8	70	5	0	8	0	6	0
1.5	30	10	100	6	54	2	1	3	20
	60	8	81	5	0	9	0	6	0

[1]R denotes the number of samples out of 100 in which the true hypothesis $\beta_k = \beta_{k_0}$ is rejected at the 0.05 level. A denotes the number of samples out of 100 in which the false hypothesis $\beta_k = 0$ is accepted as true.

although they are unambiguously worse than the statistics from the models with negative λ's.

Models with $\lambda < 0$ generally performed better than those with positive λ. For example, the false hypothesis, $\beta_0 = 0$, is never rejected for models with $\lambda = 1.0$, 1.5, and N = 30. In fact, even for the larger sample size (N = 60), a high percentage of the false hypotheses were accepted for the samples with $\lambda = 1.5$. However, the power of the test did tend to improve as sample size increased. The reason for the poor performance of the model with $\lambda > 0$ is, of course, the imprecision with which these model parameters are estimated. These test results are not encouraging for the application of t-tests to parameters estimated from Box-Cox models, especially for models with positive λ.

Table 5 reports test statistics for the equation error term. Contrary to Spitzer's assertion, our results

Table 5—Sample statistics for the error term, $\epsilon \sim N(0, 0.426)$

λ	N	MB of variance	MAB of variance	RMSE of variance	Percentage rejection of normality
	Sample size		*Estimates*		*Percent*
−1 5	30	−0 070	0 164	0 201	0
	60	− 058	134	164	3
−1 0	30	− 038	170	227	2
	60	− 028	121	144	0
−0 15	30	− 007	138	181	0
	60	− 013	108	133	0
1 0	30	1 413	1 653	4 394	35
	60	474	667	1 380	31
1 5	30	1 898	2 140	6 121	37
	60	588	787	1 541	34

Note MB = mean bias, MAB = mean absolute bias, and RMSE = root mean square error

clearly indicate that systematic bias occurs in the estimation of the error variance[5] The models underestimate the true variance for λ less than zero, and they overestimate the variance for positive λ This finding has serious implications for researchers because equation (12) expresses the expected value of the untransformed dependent variable as a function of the variance Thus, forecasts made with equation (12) using an estimate of the variance will be biased, and the bias will depend on the size of the bias in the variance and the value of λ

The MAB's and the RMSE's decline as sample size increases, indicating consistency However, the test statistics for the models with a positive λ are many times larger than their counterparts derived from the models with a negative λ In other words, the models with a positive λ once again performed far worse than their counterparts with a negative λ

The error distribution cannot be strictly normal because of the limited range of the dependent variable However, if the bounds implied on the error distribution occur in the extreme tails of the distribution as in the Monte Carlo experiment, departure from normality would not be expected to be significant. Table 5 shows the results of a test for normality of the estimated model error term using a two-tailed Kolmogorov-Smirnov goodness-of-fit at the 0 05 level Regardless of sample size, normality was rejected in approximately one-third of the replications for models with positive λ However, in the models with negative λ, normality was not rejected in the vast majority of cases

Model performance is frequently evaluated in terms of its overall fit to the sample data R^2, the coefficient of multiple determination, is probably the most often cited statistic for this purpose When applying R^2 to models with a transformed dependent variable, one must be careful to compute it in terms of the original untransformed dependent variable because the untransformed dependent variable is the variable of interest and represents a standard for comparisons across models Table 6 shows the average R^2's for the estimated models The R^2's are higher for the extreme positive and negative values of λ than for $\lambda = -0.15$ When the R^2 criterion was used, models with $\lambda > 0$ performed the best of all models R^2 decreased in larger size samples for $\lambda < 0$, but remained high and stable in models with positive λ The drop in R^2 as sample size increases suggests that randomness in smaller samples may have more influence on the parameter estimates than in larger samples, resulting in a better fit to the particular sample, but not necessarily to the population of interest This conjecture is consistent with the larger variances for the parameter estimates obtained in the smaller samples

[5]Except for N = 30 with $\lambda = -1.5$, Spitzer's reported results (*11*) indicate the same type of bias as we find We believe that Spitzer may have been too generous in stating that no error variance estimation bias appeared in his replications

Package LLRANDOM II " NPS55-81-005 Monterey, CA Naval Postgraduate School, 1981

(10) Liem, Tran C, M Dagenais, and Marc Gaudry "A Program for Estimating Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals " Publication No 301 University of Montreal, Centre de Récherche sur les Transports, 1983

(11) Spitzer, John J "A Monte Carlo Investigation of the Box-Cox Transformation in Small Samples," *Journal of the American Statistical Association*, Vol 73, 1978, pp 488-95

(12) __________ "A Primer on Box-Cox Estimation," *Review of Economics and Statistics*, Vol 64, 1982, pp 307-13

(13) Tukey, John W "On the Comparative Anatomy of Transformations," *Annals of Mathematical Statistics*, Vol 28, 1957, pp 602-32

(14) Zarembka, Paul "Transformation of Variables in Econometrics," in *Frontiers in Econometrics* (ed P Zarembka) New York Academic Press, 1974, pp 81-104

Appendix: Data Generation and Estimation

Some 100 successive data samples were generated for each model and sample size specification in the Monte Carlo study by use of a procedure set forth by Spitzer (*11*) Each generated data sample was divided into two subsamples one used for estimation and the other for forecast evaluation Without loss of generality, the first N observations were placed in the estimation sample and the remaining K observations were placed in the forecast sample, where N = (30, 60) denotes the estimation sample size and K = 10 denotes the forecast sample size

The transformed independent variables were obtained as follows first, N + K pairs of uniform pseudorandom numbers (W_{1t}, W_{2t}) were generated by use of the Lehmer multiplicative congruential method from the LLRANDOM II computer package (*9*) This generator has the form

$$U_{n+1} = A \cdot U_n \text{ (modulo } 2^{31} - 1)$$

where A = 397204092 When this value of A is used, the generator has very good statistical properties A starting seed value for the process was specified as $U_0 = 4312657$

Next, the N + K pairs were forced to orthogonality and standardized to zero mean and unit variance The transformed independent variables $X_1(\lambda)$ and $X_2(\lambda)$ were obtained from the W_t as

$$X_{1t}(\lambda) = 5 + (3)^{1/2} W_{1t},$$

$$X_{2t}(\lambda) = 45 + (0\ 4)(27)^{1/2} W_{1t}$$

$$+ [(0\ 84)(27)^{1/2}] W_{2t},\ t = 1, 2, \quad , N + K$$

For negative λ, each $X_{it}(\lambda)$ was multiplied by −1 0 to ensure that X_{it} was in the positive domain as required by the Box-Cox transformation This specification implies a correlation between $X_1(\lambda)$ and $X_2(\lambda)$ of 0 4 The inverse Box-Cox transformation was applied to the $X_i(\lambda)$ to obtain X_1 and X_2

We obtain the Y_t by untransforming the $Y_t(\lambda)$ computed from

$$Y_t(\lambda) = 9\ 0 - 1\ 5\ X_{1t}(\lambda) + 0\ 5\ X_{2t}(\lambda) + \epsilon_t$$

where ϵ_t is an independently, identically distributed normal random error term generated with mean zero and variance 0 426 If $Y_t(\lambda)$ fell outside the feasible range such that the untransformed Y_t could not be computed, then another error term was generated to compute a replacement This situation occurred only infrequently, suggesting that truncation of the error term (deviation from the assumption of normality) was not a significant problem for the specified models

Marsaglia's "rectangular-wedge-tail" procedure as implemented in LLRANDOM-II was used to generate the pseudorandom normal error term The error variance, σ^2, was chosen to make the residual variance approximately 5 percent of the total variance of $Y(\lambda)$.