10th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 1-2, 2007

A Bootstrap Procedure in a Spatial-Temporal Model for Corn Production Data

by

Charlotte F. Dumanjug

For additional information, please contact:

Author's name	:	Charlotte F. Dumanjug
Designation	:	Professor
Affiliation	:	Mindanao State University-IIT
Address	:	Iligan City
Telefax. no.	:	(0632) 9280881
E-mail	:	

A BOOTSTRAP PROCEDURE IN A SPATIAL-TEMPORAL MODEL FOR CORN PRODUCTION DATA

Charlotte Fernan Dumanjug

Bootstrap inference is proposed in estimating the parameters of a spatial-temporal model in which a hybrid of the backfitting algorithm and the Cochrane-Orcutt procedure is employed. Agricultural data is utilized to illustrate the proposed method.

Keywords: spatial-temporal, bootstrap, backfitting,

The estimation of a spatial-temporal model using a hybrid of the backfitting algorithm and Cochrane-Orcutt procedure from Landagan and Barrios (2007), is illustrated using Philippine Corn Production Data. The complex process inherent in crop growth, in both time and space, is aptly accounted for with such a model.

A bootstrap procedure to estimate the sampling distribution of statistics used to estimate the parameters in the model is essential to facilitate in making inferences.

1 A Spatial-Temporal Model

The model postulated in Landagan and Barrios (2007), which aptly accounts for spatial and temporal associations inherent in yield data, is defined as follows:

$$Y_{it} = X_{it}b + w_{it}g + e_{it} \quad i = 1,...,N \quad t = 1,...,T,$$

where Y_{it} is the response variable from location *i* at time *t*, X_{it} the set of covariates from location *i* at time *t*, w_{it} the set of neighborhood system of location *i* at time *t*, and e_{it} the error component that would take the disturbances. The error component is investigated for autoregressive behavior, postulated as $e_{it} = m_i + u_{it}$. The disturbances follow a one-way error component with individual effects $m_i \sim IID(0, s_m^2)$ and the remainder disturbances u_{it} following a stationary AR(*p*). The estimation procedure is a hybrid of the backfitting algorithm and the Cochrane-Orcutt procedure and proceeds in the following manner:

Step 1: The parameters b and r are simultaneously estimated using regression model with autocorrelated errors given by $Y_{it} = bX_{it} + re_{i(t-1)} + a_{it}$ using the Cochrane-Orcutt procedure. The error component a_{it} may contain the spatial component that is ignored in the initial estimation of b and r.

Define the resulting residuals as $e_{it} = Y_{it} - \hat{Y}_{it}$ where $\hat{Y}_{it} = \hat{b}X_{it} + \hat{r}e_{i(t-1)}$. In empirical modeling, this is executed per spatial unit.

Step 2: Another regression is performed on the residuals in Step 1 with the neighborhood variable to estimate the spatial component. This model is represented by $e_{it} = gW_{it} + a_{it}$ and estimated per time point.

Step 3: A new dependent variable is then recomputed adjusting for the spatial component, given by $Y_{it}^{1} = Y_{it} - \hat{g}W_{it}$. This sets aside the spatial effect to focus on the covariate-temporal effect when a regression with autocorrelated errors is again fitted in Step 1. The iteration converges when there are minimal changes in the values of the parameter estimates.

2 Bootstrap Procedure for the Spatial-Temporal Model

The step-by-step procedure for carrying out nonparametric bootstrap in the spatial-temporal model under consideration follows.

Step 1: Simultaneously estimate the covariate and temporal effects, b and r, respectively, for each of the 63 provinces, using the Cochrane-Orcutt procedure. Estimate the residuals from the 63 estimated regression equations.

Step 2: Regress the residuals with the neighborhood variable for each of the 52 time points. The spatial effect, *?*, for each time point is estimated here.

Step 3: Compute new data values with the *?* estimates, and use these in step 1, iterating three times. This yields nine sets of parameter estimates, three for each parameter.

Step 4: For each set of parameter estimates:

(i) Draw a simple random sample of size *n* with replacement, where *n* is the number of observations in each set.

(ii) Compute the mean. This gives the value of the statistic computed using the bootstrap sample in *(i)*.

(iii) Repeat *(i)* and *(ii)* *B* times, yielding *B* bootstrap means.

The nine sets of parameter estimates are tested for normality using the Shapiro-Wilk W test. Histograms that describe the shape of the distribution, is employed. The test will verify whether a normal distribution can be used to approximate the sampling distribution of the parameters estimated from an iterative method.

Bootstrap point estimates are determined using 1000 and 5000 replications. Although changes in the estimates are expected to be minimal for large replications, this is intended for the computation of the 95% Bias-Corrected confidence interval.

The estimated bias and standard error are used to evaluate the bootstrap estimators. In addition, the coefficient of variation is also computed to assess the reliability of the estimates.

3 Philippine Corn Production Data

Quarterly corn yield in all types (white and yellow), in 63 provinces in the Philippines from 1990-2002 (52 time points), is used as response variable. Provinces with a considerable amount of missing observations in some time points are excluded. The only covariate used in this study is area harvested as it is a good indicator of efficiency in crop yield. Often it is expected that with larger area, crop yield is maximized. Regional mean yield is used as a neighborhood variable in the estimation of the spatial parameter.

4 Results and Discussions

Point and interval estimates of the model parameters are presented in this section. Results of normality tests of the parameter estimates are also presented.

Table 3.1
Normality Test Results for Parameter Estimates

Variable	*W*	*Z*	Prob>*z*
$\hat{b}^{(1)}$	0.40861	7.586	0.00000
$\hat{b}^{(2)}$	0.34990	7.790	0.00000
$\hat{b}^{(3)}$	0.43850	7.474	0.00000
$\hat{r}^{(1)}$	0.96821	1.267	0.10257
$\hat{r}^{(2)}$	0.95739	1.900	0.02872
$\hat{r}^{(3)}$	0.97016	1.130	0.12915
$\hat{g}^{(1)}$	0.91927	2.918	0.00176
$\hat{g}^{(2)}$	0.95492	1.673	0.04720
$\hat{g}^{(3)}$	0.89943	3.388	0.00035

Table 3.1 shows that the b and g parameter estimates are not normally distributed. The r parameter estimates from the first and third iterations are apparently symmetric. From the histograms in Appendix A, the b estimates appear to be skewed. The r estimates from the second iteration are slightly skewed to the left (Appendix B). From Appendix C, observe that the g parameter estimates from both first and third iterations, are evidently skewed to the left, while estimates from the second iteration are skewed to the right. From this analysis, the asymmetric distribution of most of the parameter estimates necessitates the employment of the nonparametric bootstrap to estimate the empirical distribution of the parameters.

Table 3.2
Comparison of b Estimates from Three Iterations
With 1000 Replications and Resample Size of 63

	$Bias(\hat{q})$	$\hat{q}$	$s.e.(\hat{q})$	*CV, %*
$\hat{b}^{(1)}$	5.74e-07	.0000598	.0000654	109.3645
$\hat{b}^{(2)}$	-4.55e-07	.0001025	.000065	63.4146
$\hat{b}^{(3)}$	1.67e-07	.0000652	.0000551	84.5092

Significant differences in parameter estimates can be observed among three iterations, although the difference between estimates from the first and third iterations is minimal. The magnitude of the bias decreases with further iterations. The coefficient of variation does not provide valuable information on the characteristics of the sampling distribution of the b estimates for the reason that the b estimates are found to be skewed.

Table 3.3
Comparison of r Estimates from Three Iterations
With 1000 Replications and Resample Size of 63

	$Bias(\hat{q})$	$\hat{q}$	$s.e.(\hat{q})$	*CV, %*
$\hat{r}^{(1)}$	-.0017675	.3283336	.0423513	12.8989
$\hat{r}^{(2)}$	.000804	.3173595	.0418971	13.2018
$\hat{r}^{(3)}$	-.0007217	.2999271	.0421254	14.0452

Bias of estimates of r from the three iterations is tolerable, not exceeding 0.25 in ratio to the standard error. It can be observed, from Table 3.3, that bias decreases in magnitude with further iterations. Changes in parameter estimates among the three iterations are minimal. Notice the increase in the coefficient of variation with further iteration.

Table 3.4
Comparison of g Estimates from Three Iterations
With 1000 Replications and Resample Size of 52

	$Bias(\hat{q})$	$\hat{q}$	$s.e.(\hat{q})$	*CV, %*
$\hat{g}^{(1)}$	.0003451	-.0006152	.0179791	2922.4804
$\hat{g}^{(2)}$	.0002114	.0325786	.0065325	20.0515
$\hat{g}^{(3)}$	.0001278	.0088663	.0156059	176.6126

The estimates of bias appear to be tolerable in Table 3.4, decreasing with further iteration. Differences in the parameter estimates of g are significant with relatively large standard errors, except for the estimate from the second iteration. Because the g estimates are skewed, from the normality tests, the coefficient of variation does not give valuable information.

Table 3.5
Comparison of 95% Bias-Corrected Interval Estimates of b from Three Iterations
With 1000 and 5000 Replications and Resample Size of 63

	1000 Replications		5000 Replications	
	BC Interval	Width	BC Interval	width
$\hat{b}^{(1)}$	(-.0000515 , .0001976)	2.491e-04	(-.0000557, .0002081)	2.638e-04
$\hat{b}^{(2)}$	(7.52e-06, .0002712)	2.637e-04	(5.61e-06, .0002674)	2.618e-04
$\hat{b}^{(3)}$	(-.0000434, .0001762)	2.196e-04	(-.0000382, .0001901)	2.283e-04

Observe from Table 3.5 that the width of confidence intervals, of estimates from second iteration, decreased with larger replication size. On the estimates from the first and third iterations, however, width increases with larger replication size. Interval width decreases with further iterations.

Table 3.6
Comparison of 95% Bias-Corrected Interval Estimates of r from Three Iterations
With 1000 and 5000 Replications and Resample Size of 63

	1000 Replications		5000 Replications	
	BC Interval	Width	BC Interval	width
$\hat{r}^{(1)}$	(.243511, .4113668)	0.16786	(.2458784, .4134107)	0.16753
$\hat{r}^{(2)}$	(.2309529, .3944302)	0.15490	(.2321706, .3929713)	0.1608
$\hat{r}^{(3)}$	(.2144869, .3835097)	0.16911	(.2164173, .38383)	0.16741

With the 95% Bias-Corrected interval estimates for r, a decrease in width, although minimal, is observed with larger replication size, for estimates from the first and third iterations. It can also be observed that, for both replication sizes, the second iteration gives narrower intervals. With 1000 replications, the third iteration produced the widest interval for r, while with 5000 replications, the widest interval is observed from the first iteration.

Table 3.7 gives the 95% Bias-Corrected confidence intervals for g, from three iterations, in 1000 and 5000 bootstrap replications. Narrower intervals are observed from the second iteration, in both replication sizes. Wider intervals can be observed from the first iteration.

Table 3.7
Comparison of 95% Bias-Corrected Interval Estimates of g from Three Iterations
With 1000 and 5000 Replications and Resample Size of 52

	1000 Replications		5000 Replications	
	BC Interval	Width	BC Interval	width
$\hat{g}^{(1)}$	(-.0401332, .0299469)	0.0700	(-.0389858, .0333229)	0.0723
$\hat{g}^{(2)}$	(.0203886, .0457117)	0.02532	(.0200562, .0454229)	0.02537
$\hat{g}^{(3)}$	(-.0241908, .0362222)	0.06041	(-.0227039, .0351108)	0.05781

5 Conclusions and Recommendations

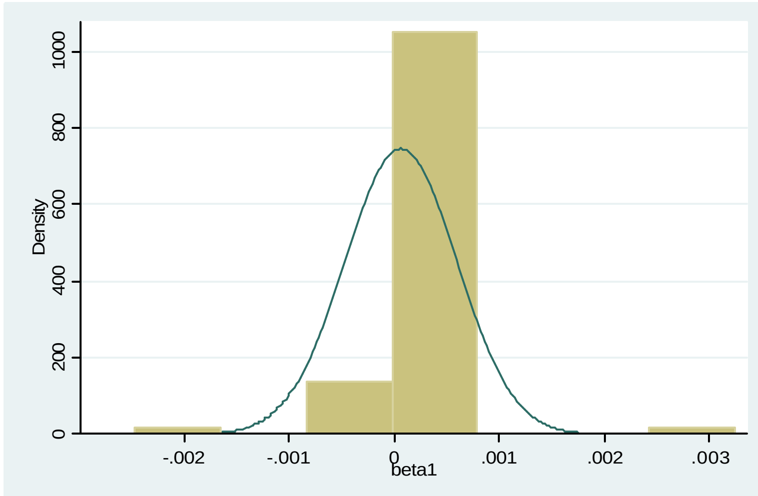
The estimated bias for all parameter estimates is generally contained, suggesting optimality of the iterative estimation procedure. Minimal changes in the empirical distributions, of estimates for both the covariate effect, b and temporal effect, r , from the three iterations, are observed. The unstable estimates of g can be explained by the fact that it accounts for a combination of a diverse set of spatial determinants.

Construction of blocks of consecutive observations could possibly result to more efficient and stable estimates of the parameters. The behavior of the estimates from further iterations may also be explored.

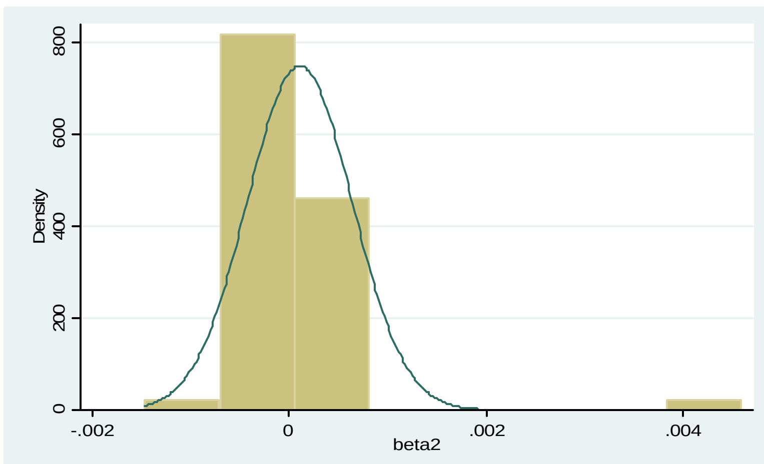
Reference:

Landagan, O.Z. and E.B. Barrios (2007), “An Estimation Procedure for a Spatial-Temporal Model,” *Statistics and Probability Letters*, 77:401– 406

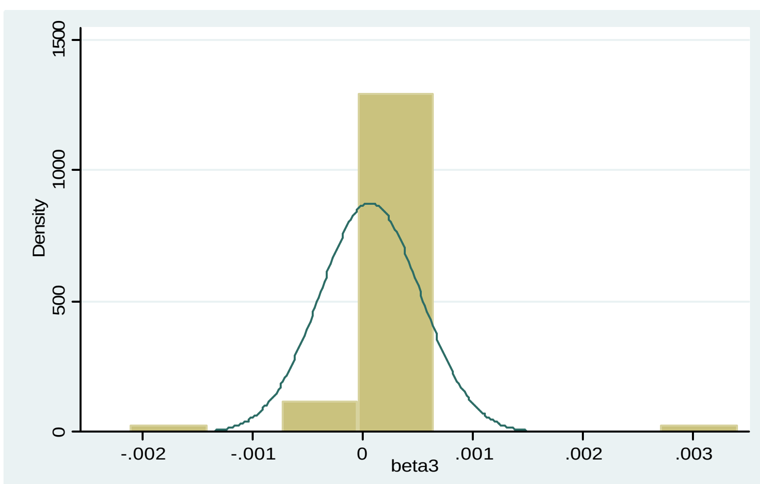
Appendix A *Histogram of b Parameter Estimates from Three Iterations*



$\hat{b}^{(1)}$

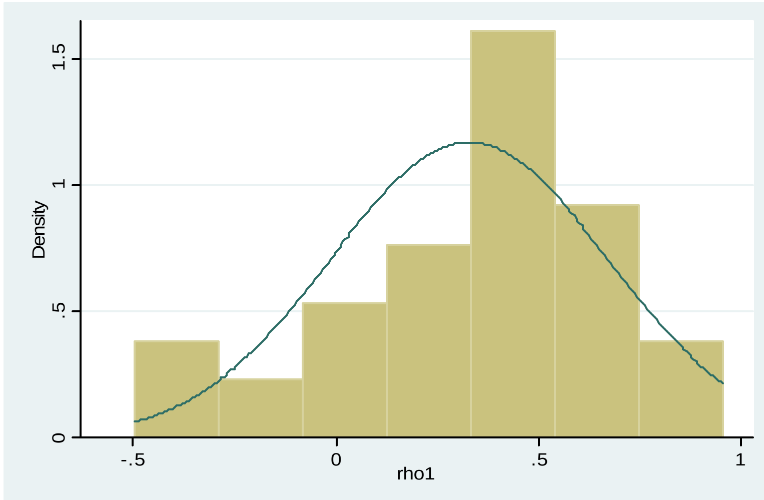


$\hat{b}^{(2)}$

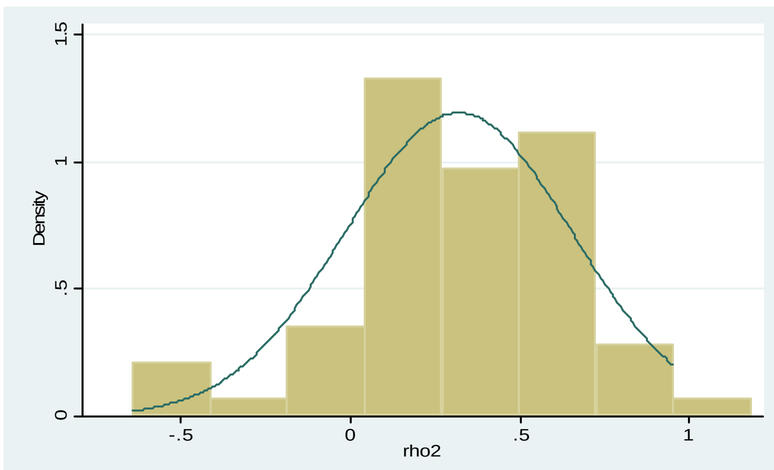


$\hat{b}^{(3)}$

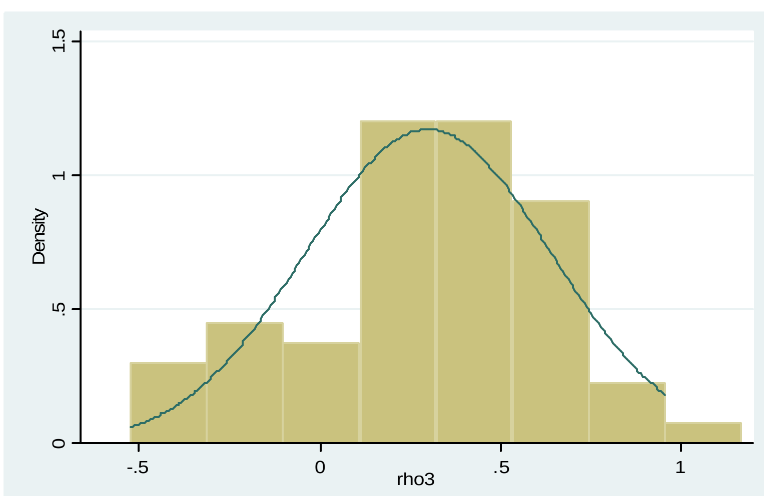
Appendix B *Histogram of r Parameter Estimates from Three Iterations*



$\hat{r}^{(1)}$

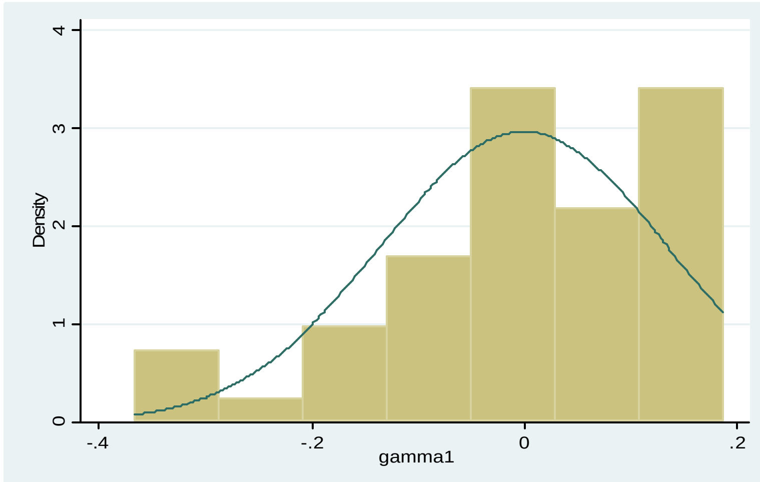


$\hat{r}^{(2)}$

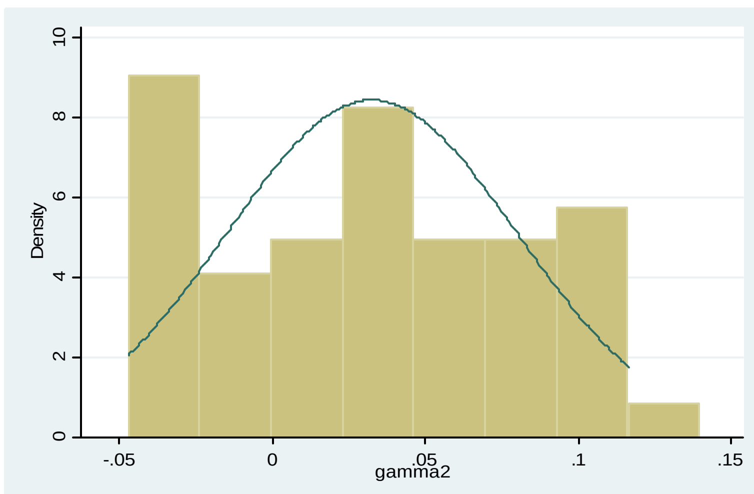


$\hat{r}^{(3)}$

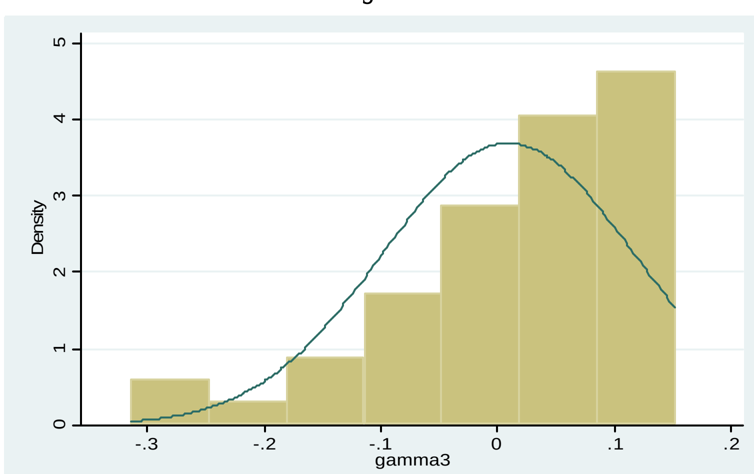
Appendix C *Histogram of g Parameter Estimates from Three Iterations*



$\hat{g}^{(1)}$



$\hat{g}^{(2)}$



$\hat{g}^{(3)}$