DOCUMENT RESUME

ED 458 265		TM 033 428
AUTHOR	Lane, Ginny G.	
TITLE	The Beginner's Guide to the Bootstrap Method of Resampling.	
PUB DATE	2000-01-00	
NOTE	17p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 27-29, 2000).	
PUB TYPE	Guides - Non-Classroom (055) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC01 Plus Postage.	
DESCRIPTORS	*Educational Research; *Estimation (Mathematics)	
IDENTIFIERS	*Bootstrap Methods; *Resampling Techniques; Research Replication	

ABSTRACT

The bootstrap method of resampling can be useful in estimating the replicability of study results. The bootstrap procedure creates a mock population from a given sample of data from which multiple samples are then drawn. The method extends the usefulness of the jackknife procedure as it allows for computation of a given statistic across a maximal number of fluctuations in the original sample from which the bootstrap data are based. A sample set of data is used to demonstrate the bootstrap procedure for a univariate multiple regression analysis. (Contains 1 figure, 2 tables, and 16 references.) (Author/SLD)

The Beginner's Guide to the Bootstrap Method of Resampling

Ginny G. Lane

University of North Texas

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX, January 27-29, 2000.

Abstract

The bootstrap method of resampling can be useful in estimating the replicability of study results. The bootstrap procedure creates a mock population from a given sample of data from which multiple samples are then drawn. The method extends the usefulness of the jackknife procedure as it allows for computation of a given statistic across a maximal number of fluctuations in the original sample from which the bootstrap data are based. A sample set of data is used to demonstrate the bootstrap procedure for a univariate multiple regression analysis.

The Beginner's Guide to the Bootstrap Method of Resampling

Result replicability, along with statistical significance and result importance, is one of the essential elements of the "research triumvirate" (Carver, 1987; Tukey, 1969). The whole purpose of taking the trouble and expense to collect samples and data is to make inferences about a particular population of interest. If study results will not replicate or generalize to the population of interest then they are of limited value (Thompson, 1992; Tukey, 1969). Unfortunately, it is often impractical, if not impossible, to replicate studies conducted in the social and behavioral sciences. It can take years to reproduce the conditions of the original study and, for many researchers, replication is cost prohibitive. One must also consider the timeliness of getting results to press. It is generally not in the interest of science to delay publication of important results for years until the entire study can be repeated with precision. Fortunately, growing accessibility to inexpensive computer power has encouraged statisticians to explore internal replication procedures as alternatives to complete study replication. One of the most useful internal replication methods gaining popularity among statisticians is the bootstrap method of resampling.

The bootstrap method of resampling, invented in the 1970s by Bradley Efron, is a computing-intensive procedure that simulates a population using the original sample set of data (Chernick, 1999). The simulated population is used to make judgements regarding the statistical analyses performed on the original sample set of data. Instead of relying of the theoretical sampling distributions for certain sample sizes, the bootstrap procedure creates an empirical distribution for a sample statistic through repeated sampling with replacement from the original sample (Chernick, 1999; Diaconis & Efron, 1983; Lunneborg, 1992; Thompson, 1992).

The bootstrap is conceptually quite simple and can be applied to a wide range of statistics. Even an inexperienced researcher can follow the logic: A random sample of data is drawn from a population of interest. Sample statistics are computed. The original sample is then copied many times to create a pseudo-population. Many random samples (of size equal to the original sample) are then drawn (with replacement) from this pseudo-population. Statistics are computed for each sample in order to create a distribution of each sample statistic. Statisticians (Chernick, 1999; Diaconis & Efron, 1983; Lunneborg, 1987) have empirically demonstrated through computer simulations that the sampling distribution created from the bootstrap samples (F^*) mirrors the true sampling distribution of the statistic (F). The number of replications required to create the "ideal" bootstrap sampling distribution (F^*) is n^n, where n is original sample size (Fox, 1997). However, Efron and Diaconis (1983) have demonstrated that one can approach the ideal F^* distribution with as few as 100 replications.[1]

The bootstrap distributions can be used for two purposes (Hinkle & Winstead, 1990; Lunneborg, 1997; Thompson, 1992). First, means and standard deviations of the bootstrap sample statistics may be computed and used to create confidence intervals around the original sample statistics. Thus, the researcher has some evidence of the stability of his or her results over many different configurations of samples. Second, the empirically created distributions can be used to make decisions regarding statistical significance when theoretical sampling distributions are not available.

One of the biggest advantages of the bootstrap method is that it does not require the assumption that the standard errors in the observed values be randomly and normally distributed in order to work effectively (Chernick, 1999; Hinkle & Winstead, 1990; Lunneborg, 1987; Reinhardt, 1992; Thompson, 1992). This assumption of normality is often required before classical statistical analysis can proceed (Lunneborg, 1987). Instead, the bootstrap method creates its own empirical distribution from the data at hand.

This is an important discovery considering that most classical statistical tests make normality assumptions about the sampling distribution. Often, this assumption is tenuous for data collected in the social and behavioral sciences (Bickel & Freedman, 1981). Until Efron invented the bootstrap, the researcher had two choices for making statistical decisions (Lunneborg, 1987). One was to assume he or she knew everything about the form of the distribution and use parametric techniques that utilize the theoretical sampling distributions. The other was to assume one knew nothing about the form of the distribution and use nonparametric techniques. The bootstrap finally offers a compromise between "everything and nothing".

Another advantage of the bootstrap procedure is that it avoids some of the problems associated with statistical significance and sample size (Fan, 1994; Thompson, 1992). It has been shown through simulated experiments that statistical significance is a function of sample size (Lane, 1999; Morrison & Hinkle, 1970; Thompson, 1998). If the sample size is large enough, statistical significance is assured, regardless of the effect size. The bootstrap can be useful in research areas where it is difficult to obtain large samples (e.g., special education). If a study's results are not statistically significant even

[1] As few as 100 replications may be sufficient for descriptive statistics such as means and standard deviations. However, for more complicated procedures or for creating confidence intervals, it has been

though the effect size is noteworthy, the bootstrap can give some estimate of result replicability. This could provide a basis for publication of important research that may be confounded by small sample size.

Given the importance of result replicability and the availability of internal replication procedures, it is surprising that many researchers fail to address the issue (Giroir, 1989; Reinhardt, 1992). Even though replication is one of the basic principles of research, surveys of professional journals in the behavioral sciences find that researchers pay little attention to this principle and fail to evaluate it in inappropriate ways. One reason for the lapse is the widespread misuse of statistical significance tests (SSTs). Many researchers mistakenly assume that statistically significant results indicate the probability that results will replicate in future samples (Carver, 1978; Daniel, 1998; Thompson, 1998). Thus, there is a large population of researchers who feel it unnecessary to bother with internal replication. Many have illuminated the folly of this assumption (Carver 1978; Daniel 1998; Lane 1999; Thompson 1998). SSTs are predicated on a true null hypothesis in the population of interest. The researcher cannot make further inferences about probabilities in the population based on calculated p results from SSTs because the population was "forcibly" set at zero (null). The calculated p value speaks only to the probability of the sample results.

Unfamiliarity with internal replication procedures is another reason why many researchers fail to address the issue. Many of those involved in educational research (author included) are neither expert statisticians nor computer prodigies. Although there are several software packages available for performing bootstrap procedures (e.g., S-Plus), most require user interface with the program syntax. Because these programs are

suggested that as many as 1000 replications may be needed (Diaconis & Efron, 1983; Fox, 1997).

relatively new, many educational researchers have not been exposed to the procedure. Many are unprepared for the challenge of learning a new program, assuming that they are even aware of the procedure. Given the pressures of daily life and the consuming nature of the researchers' primary interests, the reluctance to dedicate time to learning these new programs is understandable. However, this does not eliminate the need for the researcher to become more familiar with these procedures.

Data Example

The following example demonstrates the ease of the bootstrap procedure for those unfamiliar with the available computer programs. The entire exercise was created using SPSS—an extremely user-friendly statistical package with "point and click" commands. Although SPSS does not have a bootstrap function, the procedure can be roughly computed using the steps described below. This particular example also demonstrates the usefulness of the bootstrap procedure for small samples that yield notable effect sizes but fail to achieve statistical significance. A hypothetical population of 220 was created for heuristic purposes. The data were analyzed via multiple regression. There were two predictors and one dependent variable. The following steps were employed:

1. A random sample of 11 was selected from the population. Although the analysis yielded an R^2 of .423, the results were not statistically significant ($\underline{p}<.05$).
2. The sample was copied 20 times to create a pseudo population of 220.
3. 100 random samples (of size 11) were selected (within SPSS) from the pseudo-population. Effect size (R^2) and regression beta weights were computed for each of the 100 samples.

4. Statistics for the 100 bootstrap samples were plotted (see Figure 1). Means and standard deviations were computed and used to create confidence intervals around each statistic.
5. Results were compared to original sample estimates (see Table 1).

It should be noted that this procedure is a somewhat "crude" method of performing the bootstrap procedure. Rather than truly sampling from the original data with replacement of cases, the procedure as illustrated herein creates a "mega data file" in which each case is duplicated several (i.e., 10) times over. However, the method utilized here presents a defensible method for computing bootstrap estimates for the researcher whose technical expertise precludes use of sophisticated "programmable" statistical software.

The results of the bootstrap estimates are summarized in Table 1. The original sample estimates for all three statistics fall within the confidence intervals created by the bootstrap distributions. Because this is a hypothetical example for which we have "population" data, we have the luxury of also comparing results to the actual population parameters. All fall within the computed confidence intervals. The results of this bootstrap procedure indicate that the chances of replication of the original sample results are high even though the original result was are not statistically significant ($p<.05$). It is not unreasonable to assume that given a larger sample size for this experiment, we would have achieved statistical significance.

Table 2 illustrates the percentile method of computing confidence intervals. Instead of using normal theory confidence intervals (X + or – 1.96 * standard deviation), the intervals are determined by the actual cases that mark the 95% cut off. The cases for each statistic are first arranged in ascending order. The upper and lower limits are calculated by the following:

Lower limit cutoff = N (alpha/2)

100 (.05/2)

2.5

Upper limit cutoff = N(1-alpha/2)

100 (.05/2)

97.5

Therefore case numbers 2.5 and 97.5 mark the boundaries of the 95% confidence interval. The percentile confidence intervals and the normal theory confidence intervals were similar for this example although the percentile intervals were slightly smaller.

Conclusion

Critics of the bootstrap procedure argue that the bootstrap estimates are bound by the limitations of the original sample. This is quite true, but it is contradictory to be willing to use the original sample to estimate population parameters but unwilling to use that same sample to offer insight into the stability of those parameter estimates (Thompson, 1992). No analytic procedure can take a researcher beyond the limits of a given data set but at least the bootstrap analyses can give the researcher more confidence in results that replicate over numerous configurations of subjects. And while there may be some bias in the bootstrap estimates, biased estimates of replication are better than blind assumption. As Bruce Thompson explained (1992, p. 21):

> ...because (bootstrap) analyses capitalized during resampling on the commonalities inherent in a given sample in hand, such analyses always yield somewhat inflated evaluations of replicability. But inflated empirical evaluations of replicability are often superior to a mere presumption of replicability, especially when the researcher can take this capitalization into account during the interpretation.

According to Sir Ronald Fisher, famed statistician, replication is the basis of scientific truth (Tukey, 1969). Therefore, it is incumbent upon the researcher to become familiar with advances in statistical analysis that can address result replicability. It is incumbent upon journal editors and publishers to demand analysis of replicability and proper interpretation of statistical significance tests.

References

Bickel, P. & Freedman, D. (1981). Some asymptotic theory for the bootstrap. Annals of Statistics, 9(4), 1196-1217.

Carver, R.P. (1978) The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Chernick, M.R. (1999). Bootstrap methods: A practitioner's guide. New York: John Wiley & Sons, Inc.

Daniel, L.G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. Research in the Schools, 5(2), 23-32.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-130.

Fan, X. (1994). Does bootstrap procedure provide biased estimates? An empirical examination for a case of multiple regression. (Report No. TM-021-969). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction No. ED 372124)

Fox, J. (1997). Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage Publications.

Giroir, M.M. (1989). Ways of estimating the probability that results will replicate. (Report No. TM-014-264). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED 314449)

Hinkle, D.E., & Winstead, W.H. (1990). Bootstrap methods: A very leisurely look. (Report No. TM-014-767). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction No. ED 318751)

Lane, G. G. (1999). Show me the magnitude! The consequences of overreliance on null hypothesis significance testing. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL.

Lunneborg, C.E. (1987). Bootstrap applications for the behavioral sciences (vol.1). Seattle: University of Washington.

Morrison, D.E., & Henkel, R.E. (Eds.). (1970). The significance test controversy: A reader. Chicago: Alkine Publishing Company.

Reinhardt, B.M. (1992). Estimating result replicability using double cross-validation and bootstrap methods. (Report No. TM-018-581). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED 346164)

Thompson, B. (1992). Exploring the replicability of a study's results: Bootsrap statistics for the multivariate case. (Report No. TM-018-174). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction No. ED 344895)

Thompson, B. (1997). If statistical significance tests are broken/misused, what practices should supplement or replace them? (Report No. TM-027-613). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED413342)

Tukey, J.W. (1969). Analyzing data: Sanctification or detective work? American Psychologist, 24(1), 83-91.

Table 1

Bootstrap Results: Estimates of R^2 and Beta Weights (100 Replications)

Population Results	Statistic/ Parameter	Original Sample Results	Bootstrap Mean of the Statistic	95 % Confidence Intervals (normal theory)	95 % Confidence Intervals (Percentile)
.542	R^2	.423	.540	.187, .892	.164, .894
.656	X_1 Beta Weight	.462	.360	-.030, .950	-.114, .914
.179	X_2 Beta Weight	.293	.470	-.306, .986	-.363, .984

Table 2

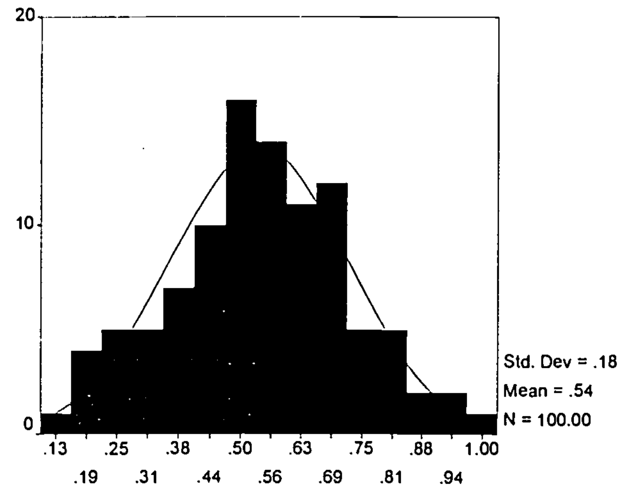
Percentile Confidence Intervals

Case #	RSQUARED	X1BETA	X2BETA	Case #	RSQUARED	X1BETA	X2BETA
1	.127	-.110	-.509	51	.552	.440	.327
2	.158	-.078	-.339	52	.554	.450	.328
3	.169	-.006	-.291	53	.555	.450	.330
4	.189	.016	-.219	54	.557	.462	.334
5	.216	.062	-.213	55	.559	.463	.335
6	.221	.087	-.200	56	.560	.470	.336
7	.229	.106	-.181	57	.560	.470	.344
8	.249	.113	-.089	58	.560	.486	.355
9	.268	.118	-.077	59	.565	.492	.362
10	.280	.125	-.021	60	.567	.501	.363
11	.311	.160	-.017	61	.577	.502	.368
12	.316	.179	-.006	62	.585	.510	.369
13	.320	.180	.008	63	.597	.510	.375
14	.322	.183	.033	64	.597	.518	.377
15	.333	.203	.045	65	.600	.527	.393
16	.370	.225	.053	66	.606	.532	.400
17	.375	.226	.070	67	.608	.534	.425
18	.390	.230	.077	68	.617	.547	.448
19	.402	.230	.093	69	.620	.557	.452
20	.403	.231	.111	70	.625	.558	.461
21	.405	.238	.118	71	.630	.561	.461
22	.406	.250	.120	72	.645	.612	.469
23	.409	.290	.122	73	.650	.615	.477
24	.409	.300	.146	74	.667	.621	.499
25	.411	.302	.156	75	.668	.629	.510
26	.416	.303	.163	76	.668	.636	.526
27	.419	.310	.167	77	.673	.656	.546
28	.427	.310	.175	78	.679	.660	.567
29	.432	.315	.181	79	.682	.665	.601
30	.433	.325	.186	80	.686	.671	.602
31	.443	.327	.198	81	.690	.673	.608
32	.452	.332	.206	82	.695	.673	.613
33	.490	.348	.211	83	.700	.730	.619
34	.498	.350	.214	84	.704	.752	.665
35	.499	.350	.215	85	.709	.755	.726
36	.502	.350	.217	86	.721	.762	.750
37	.502	.367	.221	87	.738	.763	.803
38	.503	.386	.226	88	.741	.809	.827
39	.503	.395	.230	89	.747	.813	.829
40	.504	.400	.264	90	.751	.816	.844
41	.505	.402	.264	91	.785	.819	.847
42	.505	.402	.265	92	.792	.840	.855
43	.506	.402	.267	93	.798	.845	.874
44	.508	.420	.276	94	.805	.863	.923
45	.509	.430	.279	95	.816	.902	.930
46	.510	.430	.289	96	.860	.903	.965
47	.511	.430	.291	97	.873	.914	.978
48	.523	.437	.293	98	.915	.915	.990
49	.534	.440	.303	99	.947	.924	1.060
50	.542	.440	.307	100	.994	1.024	1.149

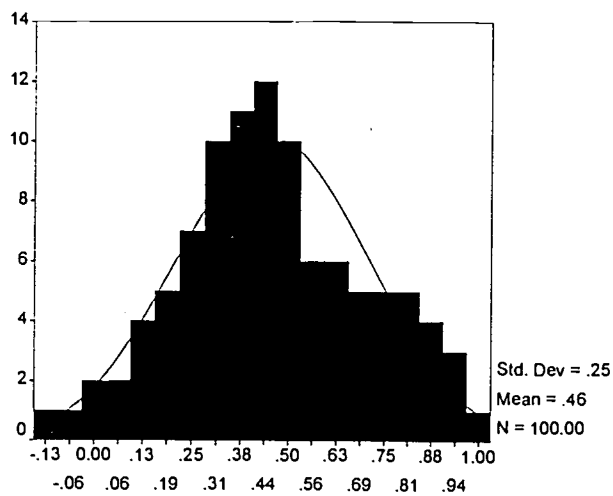
Figure 1

Bootstrap Distributions R^2, X1 Beta Weights, and X2 Beta Weights

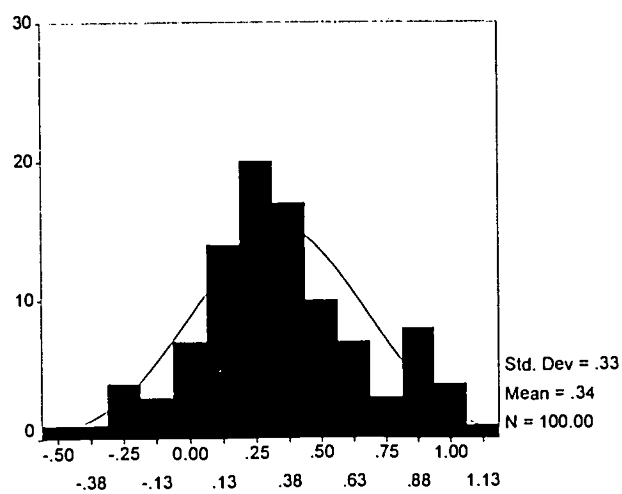
1. R^2 Distribution (100 Replications)



2. X1 Beta Weight Distribution (100 Replications)



3. X2 Beta Weight Distribution (100 Replications)



TM033428

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The Beginner's Guide to the Bootstrap Method of Resampling

Author(s): Ginny G. Lane

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Ginny G Lane	Printed Name/Position/Title: Ginny G. Lane	
Organization/Address: University of North Texas	Telephone: (817)431-5137	FAX:
	E-Mail Address: ginnylane@prodigy.net	Date: 1/28/00

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com