DOCUMENT RESUME

ED 466 648 TM 034 259

AUTHOR	Kromrey, Jeffery D.; Romano, Jeanine
TITLE	The Influence of Influence Diagnostics: An Empirical Investigation of the Effects of Removing Extreme Data Points.
PUB DATE	2001-04-00
NOTE	78p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2001).
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC04 Plus Postage.
DESCRIPTORS	Monte Carlo Methods; *Multivariate Analysis; Simulation; *Statistical Bias
IDENTIFIERS	*Data Elements; *Outliers

ABSTRACT

Monte Carlo methods were used to investigate the effects of removing extreme data points identified by five indices of influence. Multivariate normal data were simulated and observations were removed from samples if they exceeded the criteria suggested in the literature for each influence statistic. Factors included in the design of the Monte Carlo study were the number of regressor variables, population multiple correlation, degree of multicolinearity, and sample size. Conditions were simulated in which all sample observations were drawn from a single population and conditions in which a single observation in each sample was drawn from a different population (presenting either an extreme residual or an extreme value in the space of the regressor variables). Results were evaluated in terms of statistical bias in the regression parameter estimates and the sample R squared value. Results in general suggest that caution is needed in the screening of samples for outliers and influential observations. (Contains 4 figures, 12 tables, and 6 references.) (Author/SLD)

Running head: INFLUENCE DIAGNOSTICS

The Influence of Influence Diagnostics: An Empirical Investigation of the Effects of Removing Extreme Data Points

Jeffrey D. Kromrey
University of South Florida

Jeanine Romano
University of Tampa

Address Correspondence to:
Jeffrey D. Kromrey
Dept. of Educational Measurement and Research
University of South Florida
4202 E. Fowler Ave., EDU 162
Tampa, FL 33620
(813) 974-5739
kromrey@tempest.coedu.usf.edu

Paper presented at the annual meeting of the American Educational Research Association, April 10 – 14, 2001, Seattle.

Abstract

Monte Carlo methods were used to investigate the effects of removing extreme data points identified by five indices of influence. Multivariate normal data were simulated and observations were removed from samples if they exceeded the criteria suggested in the literature for each influence statistic. Factors included in the design of the Monte Carlo study were the number of regressor variables, population multiple correlation, degree of multicollinearity, and sample size. Conditions were simulated in which all sample observations were drawn from a single population and conditions in which a single observation in each sample was drawn from a different population (presenting either an extreme residual or an extreme value in the space of the regressor variables). Results were evaluated in terms of statistical bias in the regression parameter estimates and the sample R^2 value.

The Influence of Influence Diagnostics: An Empirical Investigation of the Effects of Removing Extreme Data Points

The purpose of this research was to examine outlier detection strategies and the effects of their use on the resulting regression equation parameter estimates. Linear regression analysis is used to make predictions about the behavior of data. By using the method of least squares, data are fit to a linear model and the resulting sample equation is used to draw inferences about the population from which the sample was obtained. Unusual data points known as outliers can have a critical impact on the sample regression equation. Outlier detection strategies are recommended for screening samples to determine which observations should be used to obtain the sample estimates of regression parameters (Bollen & Jackman, 1985; Mongomery & Peck, 1992; Fox, 1997).

Outlier Detection Strategies

Many statistics may be used for the detection of outlying observations, but this study focused on the five indices most commonly suggested: (1) leverage, (2) studentized residuals or RSTUDENT, (3) Cook's D, (4) DFITS and (5) DFBETAS. All of these indices begin with the general linear regression model:

$$y = Xb + \varepsilon$$

where y is an n x 1 vector of values for the dependent variable,
X is an n x k matrix of observations on the independent variables ,
b is a k x1 vector of regression coefficients, and
ε is an n x 1 vector of disturbances or residuals.

The least squares regression coefficients and the predicted values of the dependent variable are obtained as

$$b = (X^T X)^{-1} X^T y \text{ and}$$

$$\hat{y} = X(X^T X)^{-1} X^T y$$

The matrix **H**, known as the "hat matrix," is then defined to be

$$H = X(X^T X)^{-1} X^T$$

The diagonal elements of **H** denoted as h_{ii} are called leverage values. These values represent the extent to which each observation presents extreme values on the predictor variables.

The RSTUDENT or studentized residuals are transformations of the least squares residuals (i.e., $e_i = y_i - \hat{y}_i$), using the leverage value for the i^{th} observation and estimated variance of the i^{th} residual. Specifically, the RSTUDENT values are defined as:

$$RSTUDENT = \frac{e_i}{\sqrt{s_i^2(1-h_{ii})}}$$

DFITS values represent changes in all the regression coefficients that result when a single case is removed from the sample. DFITS are defined as:

$$DFITS = \frac{(\hat{y} - X_{(i)}b_{(i)})}{\sqrt{s_i^2 h_i}}$$

In contrast to DFITS, DFBETAS represent an index of the extent to which each regression coefficient changes when a case is omitted from the sample. These values, for regression weight b_j, are defined as:

$$DFBETA_j = \frac{b_j - b_{j(i)}}{\sqrt{s_i^2 (X^T X)_{jj}^{-1}}}$$

Cook's distance is a measure of squared distance between the least square estimate derived using all n points (**b**) and the estimate obtained by deleting the i^{th} observation (i.e., using $n - 1$ observations for the estimate, represented by $\mathbf{b}_i$). Cook's D is defined as:

$$D = \frac{(b_{(i)} - b)^T X^T X (b_{(i)} - b)}{(k+1)(MS_E)}$$

After values of these indices have been calculated for each observation in the sample, the obtained values are compared to criteria to determine if they are large enough to suggest that the observation is an outlier or an influential data point. The criteria suggested in the literature (e.g., by Bollen & Jackman, 1985) are functions of sample size and number of regressors (see Table 1).

A Paucity of Research on Outlier Detection

The impact of outliers and influential data points on the estimation of linear regression models is an area of research that has received very little attention (Chatterjee & Yilmaz, 1992). While several textbooks provide introductions to regression diagnostics (Pedhazur, 1997;

Mongomery & Peck, 1992; Fox, 1997), and current software including SPSS (SPSS, Inc., 1988) and SAS (SAS Institute, Inc., 1982) will provide information about outliers, very little systematic research has been conducted to examine the extent to which the removal of outliers influences the resulting regression model.

For example, in their expositions on regression diagnostics Bollen and Jackman (1985) examined data in two empirical settings drawn from cross-national comparative research (the relation between voting turnout and income inequality in industrial societies and an analysis of economic dependency and political democracy). They found that diagnostics were helpful in identifying problems of sample composition and measurement error. Similarly, in a review of regression diagnostics Chatterjee and Yilmaz (1992) looked at a small set of data to detect influential data points. After detecting influential data using the five diagnostics described above, they reported a large change in the estimated regression equations when a influential data point was removed. They concluded that as little as 1% of influential data points may affect multi-collinearity and may alter estimates of parameters and other statistics in an unpredictable way.

Although such anecdotal reports are useful to suggest the importance of screening data for outliers or influential observations, little evidence is available about the effects of data screening and the removal of influential data points on subsequent inferences about the population from which the sample was obtained. The purpose of this study was to examine the effects of outlier removal on the accuracy of such inferences. Specifically, we hypothesized that if all observations are sampled from a common multivariate normal population, the screening and removal of observations would result in biases in the regression estimates. In contrast, if actual aberrant observations are present in the samples, the removal of such observations would reduce bias in the estimates.

Method

The research was a Monte Carlo study in which random samples were simulated under known and controlled population conditions. In the Monte Carlo study, samples were generated from multivariate normal populations, regression equation parameters were estimated, and the samples were subsequently screened for outliers and influential data points. Such data points were removed from the sample and the regression equation parameters were re-estimated using the reduced sample.

We applied each index of "outlyingness" separately to each sample, using the "critical values" suggested in the literature. That is, each sample was first screened using only the leverage values, and the most extreme observation was removed if the leverage exceeded $2k/n$. If an observation was removed, the equation was re-estimated, using the remaining $n - 1$ observations and the leverage values were recomputed. This iterative process of estimating and screening for outliers continued until no observations were identified as presenting extreme values. The process was then repeated (using the entire original sample) applying each influence diagnostic statistic (e.g., DFITS, Cook's distance, etc.).

The Monte Carlo study included five factors in the design. These factors were (a) the true population multiple correlation (with $\rho^2 = 0.10, 0.30$, and 0.60), (b) number of regressor variables (with $k = 2$ and 5), (c) sample sizes (with $n = 5{*}k$, $10{*}k$, and $50{*}k$), (d) degree of multicollinearity (with average inter-regressor correlations of approximately 0.1, 0.3 and 0.5), and type of aberrant observation present in the sample (extreme residual, regressor outlier, and no aberrant observation). The correlation matrices used as the bases for the simulations were obtained from matrices reported in the educational research literature.

Samples were generated according to three conditions of aberrance. In two of these conditions, a single observation was produced that differed from the n – 1 remaining observations in the sample. Two types of such aberrant observations were investigated. Samples that included an observation with an extreme residual were produced by randomly selecting a single observation from each sample and computing that observation's residual from the population regression equation. The residual for the observation was then increased by $3\sigma_\varepsilon$ and the value of the criterion variable was recomputed using the larger residual. Samples that included a regressor outlier were produced by sampling one observation from a multivariate normal population with a mean of 3.0 on each regressor variable, and sampling the remaining n – 1 observations from a population with a mean of 0.0 on each regressor. Finally, samples were generated that included no aberrant observations (that is, all n observations were sampled from the same multivariate normal population).

The research was conducted using SAS/IML version 6.12 and 8.1. Conditions for the study were run under Windows 98. Normally distributed random variables were generated using the RANNOR random number generator in SAS. A different seed value for the random number generator was used in each execution of the program. The program code was verified by hand-checking results from benchmark datasets.

For each condition investigated, 5,000 samples were generated. The use of 5,000 samples provides adequate precision for the investigation of the bias in sample regression parameter estimates. For example, 5,000 samples provides a maximum 95% confidence interval width around an observed proportion that is $\pm .014$ (Robey & Barcikowski, 1992).

The effects of outlier screening were evaluated by calculating the bias in the sample estimates of individual regression parameters and the sample estimates of ρ^2. The regression equation obtained from each sample *before* outlier screening was compared to the known population regression equation (i.e., the equation used as the basis for the data generation). Similarly, the equations estimated (a) after the removal of the most extreme outlier (if at least one outlying data point was identified), and (b) after the removal of the two most extreme outliers were compared to the population parameters.

Regression parameter estimates were compared in terms of their statistical bias as estimates of the population parameters. The bias was estimated using

$$Bias(b_i) = \frac{\sum_{j=1}^{J}(b_{ij} - \beta_i)}{J}$$

where $Bias(b_i)$ = estimated bias in the i^{th} regression weight,
b_{ij} = i^{th} weight in the j^{th} sample
β_i = population value of the i^{th} regression weight,
J = number of samples simulated.

The bias among all of the regression weights was then computed as the mean absolute value of bias in the k weights. Absolute values were used to prevent positive and negative biases in the weights from canceling one another. An analogous bias equation was used for the estimation of bias in the sample R^2 as an estimate of ρ^2.

In addition to the estimation of statistical bias in the sample estimates, we evaluated the extent of agreement among the indices in individual observations being identified as outliers or influential cases. Finally, for conditions that included one of the two types of aberrant observations, we calculated the proportions of samples in which the aberrant observation was correctly flagged by each index.

Results

Conditions with All Observations from a Common Population

An initial consideration of the results of this research is the extent to which bias is induced by outlier detection methods if no "true" aberrant observations are present in the sample.

Bias in R^2. The sample estimates of statistical bias in the value of R^2 with a single observation removed, if flagged by each influence statistic, are presented in Table 2. The last column of the table provides, as a reference value, the estimated bias in sample R^2 prior to removing any observations. As anticipated, the bias in R^2 before removing observations (reflected in these reference values) increases with smaller samples and smaller values of ρ^2. That is, research designs with small samples drawn from populations with small values of ρ^2 are expected evidence more bias whether or not observations are removed as a result of applying the screening diagnostics. For conditions with large samples (n = 50*k), Cook's distance did not flag any observations for removal, therefore, bias estimates could not be calculated.

To provide an overview of the results, the bias estimates in Table 2 are graphed in Figure 1 as a series of box-and-whisker plots. As is evident in this figure, the indices differ in the amount of bias induced in R^2 when they are used to identify and remove a single observation. Further, the magnitude of induced bias is substantial for some conditions. Overall, the use of leverage to identify observations appears to induce the least amount of bias, while Cook's distance and R_Student appear to induce the greatest amount of bias.

The sample estimates of bias with two observations removed (if flagged) are presented in Table 3 and Figure 2. These data suggest an increase in the magnitude of bias with the removal of a second observation, but no substantive change in the pattern of results.

A closer inspection of the data in tables 2 and 3 suggests that the bias resulting from the use of influence diagnostics is related to several factors included in the experimental design used in this research. For example, the methods tend to converge as sample size increases. With k = 2, ρ^2 = .10, $\bar{r}_{12}$ = .30, and n = 10, the bias with one observation removed ranged from .40 (using R_Student to identify observations) to .20 (using leverage). Under this condition with large samples (n = 100), the bias ranged from only .02 to .01. These data are graphed in Figure 3 (for the removal of one observation) and Figure 4 (for the removal of two observations). The differences among methods are apparent with small samples, as is the large magnitude of bias resulting from the use of all methods examined except leverage. The use of DFBETAs and DFFITS result in nearly equal amounts of bias, while greater magnitudes of bias are observed for R_Student and Cook's distance. In both figures, the bias obtained from the use of leverage to screen observations closely approximates the bias observed before deleting observations.

The pattern of bias observed appears to be a function of absolute sample size rather than the ratio of sample size to number of regressors. Figure 5 provides bias estimates resulting from DFBETAs, R_Student, and leverage obtained in 2-regressor and 5-regressor models. The bias resulting from the application of the screening diagnostics in the 2-regressor models (having a smaller sample size for a given n-to-k ratio) exceeds that obtained in the 5-regressor models (having a larger sample size for a given ratio).

The magnitude of bias in R^2 appears consistent across levels of ρ^2. Figure 6 provides bias results from the removal of a single observation across levels of ρ^2 (for k = 5, n = 25, and $\bar{r}_{12}$ = .5). The lines in the graph are approximately parallel reflecting a consistent amount of bias induced by each method (with the exception of Cook's distance applied to small values of ρ^2, a condition that yielded a larger magnitude of bias).

Bias in Regression Weights. Tables 4 and 5 present the average bias in regression weights with one and two observations removed, respectively. As with the previous tables, the bias observed before removing observations is presented as a reference column. The bias values are all positive because bias was calculated as the average absolute value of bias across the k regression weights. These bias values are graphed in Figures 7 and 8 to provide an overview of the results. In contrast to the bias obtained with R^2, the bias in the regression weights appears very small in magnitude and the differences among methods is less obvious (with the exception of Cook's distance).

A closer inspection of Tables 4 and 5, however, suggests that the magnitude of bias in the regression weights is also a function of sample size. Figures 9 and 10 present bias as a function of sample size with one and two observations removed, respectively, for the condition k = 2, ρ^2 = .10, and $\bar{r}_{12}$ = .30. Differences among the influence diagnostics are evident for the small samples (n = 10), but the bias estimates converge for larger samples. The extreme bias resulting from the use of Cook's distance is apparent in these figures.

Agreement Among Indices

The agreement among the indices in the specific observations identified are presented in Table 6. This table presents the mean level of agreement across the conditions examined in this study, as well as the maximum and minimum agreement observed. The inter-index agreement on the first observation identified is presented above the main diagonal, while the agreement on the first two (disregarding order of identification) are below the diagonal.

As suggested by the bias results reported above, DFFITS and DFBETAs evidence a high level of agreement in the cases identified (mean agreement = .73 on the first case, with a range of .61 to .85). The R_Student index evidenced a moderate level of agreement with both DFFITS and DFBETAs (mean agreement = .40 and .37, respectively, for the first case identified). Leverage and Cook's D evidenced lower levels of agreement with the other indices. For leverage, the mean agreement on the first observation ranged from .03 (agreement with R_Student) to .19 (agreement with DFFITS). Similarly, for Cook's D, the mean agreement ranged from .06 (agreement with leverage and DFBETAs) to .19 (agreement with R_Student).

Conditions with Samples that Contain an Aberrant Observation

Successful Identification of Abberant Observations. Tables 7 and 8 present the proportion of samples in which each outlier detection method identified an aberrant observation as the first or second observation flagged. For samples generated with an observation having an excessively large residual (Table 7), the Rstudent index identified the observation as the first one flagged with rates ranging from 0.77 to 0.93. DFBETAs and DFFITS were nearly as successful with small samples, but their performance deteriorated with the larger samples examined. Leverage and Cook's D evidenced nearly complete inability to identify these aberrant observations.

For samples generated with an observation having extreme values among the regressor variables (Table 8), leverage was the most successful at small values of ρ^2. For example, with ρ^2 =.10 leverage identified the aberrant observation as the first one flagged at rates ranging from .62 to .85 with two regressors and ranging from .62 to .97 with five regressors. The method became less successful with larger values of ρ^2 and with higher degree of intercorrelation among the regressors. As ρ^2 increased, however, both the DFFITS and DFBETAs evidenced improved performance in detecting the aberrant observation. For example, with two regressors, $\rho^2 = .10$, and regressor intercorrelation of .30, DFFITS correctly identified the aberrant observation with rates between only .39 and .46, while the rates for leverage ranged from .73 to .82. In contrast, with two regressors, $\rho^2 = .60$, and regressor intercorrelation of .30, the rates for DFFITS ranged from .84 to .97, while the rates for leverage ranged from .72 to .80.

Bias in R^2.The sample estimates of statistical bias in the value of R^2 with a single observation removed, if flagged by each influence statistic, are presented in Tables 9 and 10, for the two types of aberrant observations simulated. As with previous tables, the last column of these tables provides, as a reference value, the estimated bias in sample R^2 prior to removing any observations. The overall distribution of the bias values in Tables 9 and 10 are presented as box-and-whisker plots in Figures 11 and 12, respectively. For samples that contain an observation with an exceptionally large residual (Table 9 and Figure 11), the removal of an observation flagged by the screening methods did not, in general, improve the estimation of R^2. The typical effect of such removal was to induce a small positive bias in the sample R^2 value. The effect was most pronounced with Cook's D method. For example when k = 2, $\rho^2 = .10$, and $\bar{r}_{12} = .30$ there was an extreme decrease in bias as sample size increased.(Table 10 and Figure 15). When k = 2, $\rho^2 = .60$, and $\bar{r}_{12} = .30$, the decrease in bias was even more extreme with Cook's D method(Table 10 and Figure 16) . Interestingly, the use of leverage for screening appeared to induce the least bias in the R^2 values. The leverage statistic was not effective in identifying the "correct" observations with this type of aberrance. Apparently the observations this statistic is flagging for removal have a relatively benign impact on the sample R^2.

For samples that contain an observation with unusual values in the regressor variables (Table 10 and Figure 12), more substantial impacts on the sample R^2 value are evident. Specifically, the large negative biases in R^2 that are present in small samples are effectively reduced by all of the screening methods(Figures 15 and 16). Unfortunately, some degree of positive bias is induced in many of the sample conditions. The most effective of the statistics examined, again, appears to be leverage. For the type of outlier simulated in these conditions, leverage was effective in identifying the correct observations and over the set of conditions examined was the most effective in terms of yielding the least biased estimated of the population squared multiple correlation.

Bias in Regression Weights. Tables 11 and 12 present the average bias in regression weights with a single observation removed for the two types of aberrance examined. As with previous presentation of bias in regression weights, the bias values are all positive because bias was calculated as the average absolute value of bias across the k regression weights. These bias values are graphed in Figures 13 and 14 to provide an overview of the results. As with the regression weight bias results obtained with all observations sampled from a single population, the effect of an observation with a large residual is very small (Table 11 and Figure 13). The screening and removal of observations provided little effect overall on the bias in regression weights with this type of aberrant observation. A notable difference is the use of Cook's D as a screening statistic, a statistic whose use induces substantial bias in the weights.

More notable effects are evident with samples that contain an observation with unusual values in the regressor variables (Table 12 and Figure 14). In these conditions, the presence of the aberrant observation has a substantial biasing effect on the sample regression weight, especially with small samples (Table 12 and Figures 17 and 18). All of the screening statistics were effective in reducing the bias in the regression weights with this type of aberrance. Although all of the statistics were effective when the resulting regression weights are compared to those obtained from the samples before outlier screening, the most effective among the methods appear to be DFFITS and leverage. For example when k = 2, ρ^2 = .10, and $\bar{r}_{12}$ = .30. the bias in regression weights is substantially smaller for DFFITS and leverage as sample size increases (Table 12 and Figure 17). When k = 2, ρ^2 = .60, and $\bar{r}_{12}$ = .30 a similar pattern is present (Table 12 and Figure 18). These screening statistics were the most accurate among those examined in identifying the aberrant observation in the samples and such accuracy is reflected in the reduced bias in sample regression weights.

Conclusions

The results of this research, in general, suggest that prudence and caution are needed in the screening of samples for outliers and influential observations. To some extent, apparent outliers in a sample reflect actual variability of the population and their elimination results in biased parameter estimates. When samples were generated from a single population and no unusual observations were deliberately included, the bias induced in the sample R^2 when these methods are applied to small samples was substantial. Although the bias induced in the regression weights was of a much lower magnitude, with small samples such bias is probably not ignorable. With equations estimated from large samples, the biases resulting from the use of these indices converge toward the bias expected without removal of any observations.

The only screening index that appeared to result in no bias increase in such samples (indeed, in some conditions bias was *reduced* after removal of observations) was leverage. The reader is reminded that the leverage index is calculated from the matrix of observations on the regressor variables only. That is, the criterion variable is irrelevant to this statistic. The other indices, by including in their calculation the criterion variable and its estimation, systematically identify observations whose removal biases the sample equation and the sample estimate of the population coefficient of determination.

The other side of the influence coin is the effect of unusual observations that actually results from processes that differ from those providing the majority of the sample. We attempted to simulate this with samples that included a single observation reflecting one of two types of aberrance. As expected, the effects of such aberrant observations are notably greater with small samples. Observations with large residuals had relatively little impact on the bias of either R^2 or the sample regression weights. With small samples, R^2 is already a biased estimator of ρ^2 and the presence of a larger than normal residual in some conditions reduced this small sample bias (at the most extreme these single observations resulted in small sample R^2 being negatively biased as an estimator). The screening and removal of a flagged observation evidenced little impact overall on the bias in R^2 or the regression weights. Observations with an unusual value among the regressor variables showed a greater effect, notably in the bias of sample regression weights. For this type of aberrant observation, screening and removal was effective in reducing bias, and both leverage and DFFITS were the most effective screening statistics.

Several limitations need to be considered in the interpretation of these results. First, the outlier detection strategies we employed were designed for detecting a single extreme observation. Multiple outliers in a sample may "mask" one another so that they cannot be detected by these techniques. Secondly, we simulated only multivariate normal data; nonnormal data may behave very differently. Finally, in this simulation study the regression models we employed represented the correct functional form of the relationship between the regressors and the criterion variable (i.e., linear, additive models) and either all observations were randomly sampled from a common multivariate population or a single aberrant observation was present in each sample. In actual field research, outlier removal from samples may improve the ability to identify the correct functional form of relationship (e.g., nonlinear or nonadditive models) and may improve estimates when samples are comprised of a mixture of distributions. Further research is needed to investigate the use of these diagnostics in such samples.

References

Bollen, K.A., & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. *Sociological Methods and Research, 13,* 510 -542.

Chatterjee, S. & Yilmaz, M. (1992) . A review of regression diagnostics for behavior research. *Applied Psychological Measurement. 16(3),* 209 - 227.

Fox, J (1997) *Applied regression analysis, linear models, and related methods.* California: Sage.

Montgomery, D. C. & Peck, E.A. (1992) *Introduction to linear regression analysis* (2nd ed.) New York: Wiley.

Pedhazur, E (1997). *Multiple regression in behavioral research* (3rd ed.) New York: Holt, Rinehart & Winston.

Robey, R. & Barciowski, R. (1992) Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematics and Statistical Psychology* 45 283 -288.

Table 1
Criterion Values for Influence Diagnostics.

Leverage (h_{ii})	$\frac{2k}{n}$
RSTUDENT	Critical t-value using $df = n - k - 1$
Cook's D	1.00
DFITS	$2\sqrt{\frac{k}{n}}$
DFBETAS	$\frac{2}{\sqrt{n}}$

Table 2
Estimated Bias in Sample R^2 with One Observation Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.3656	0.3262	0.2073	0.4035	0.3344	0.1931
			0.3	0.3561	0.3214	0.1993	0.4012	0.3291	0.1875
			0.5	0.3827	0.3340	0.2166	0.4172	0.3455	0.1982
		0.3	0.1	0.2467	0.2333	0.1111	0.3194	0.2384	0.1241
			0.3	0.2441	0.2202	0.1007	0.2995	0.2197	0.1174
			0.5	0.2303	0.2261	0.1044	0.3081	0.2247	0.1175
		0.6	0.1	0.0943	0.0885	0.0081	0.1705	0.0873	0.0330
			0.3	0.1276	0.1336	0.0239	0.2024	0.1333	0.0592
			0.5	0.1334	0.1236	0.0216	0.2037	0.1285	0.0513
	20	0.1	0.1	0.2018	0.1373	0.0849	0.1556	0.1389	0.0885
			0.3	0.1927	0.1353	0.0824	0.1516	0.1354	0.0864
			0.5	0.1929	0.1370	0.0842	0.1550	0.1381	0.0883
		0.3	0.1	0.1030	0.0939	0.0333	0.1231	0.0967	0.0522
			0.3	0.1305	0.0913	0.0262	0.1203	0.0916	0.0473
			0.5	0.1114	0.0992	0.0289	0.1294	0.0953	0.0536
		0.6	0.1	0.0450	0.0514	-0.0129	0.0832	0.0512	0.0197
			0.3	0.0548	0.0571	-0.0102	0.0927	0.0557	0.0225
			0.5	0.0475	0.0574	-0.0100	0.0924	0.0539	0.0232
	100	0.1	0.1	---	0.0218	0.0145	0.0250	0.0224	0.0175
			0.3	---	0.0206	0.0124	0.0240	0.0205	0.0157
			0.5	---	0.0216	0.0135	0.0250	0.0220	0.0164
		0.3	0.1	---	0.0164	0.0012	0.0252	0.0168	0.0096
			0.3	---	0.0158	-0.0009	0.0246	0.0155	0.0084
			0.5	---	0.0177	0.0004	0.0255	0.0170	0.0095
		0.6	0.1	---	0.0093	-0.0075	0.0194	0.0102	0.0027
			0.3	---	0.0119	-0.0062	0.0215	0.0116	0.0044
			0.5	---	0.0125	-0.0063	0.0217	0.0121	0.0044
5	25	0.1	0.1	0.3456	0.2479	0.1865	0.2701	0.2540	0.1809
			0.3	0.3299	0.2474	0.1824	0.2685	0.2521	0.1797
			0.5	0.3274	0.2499	0.1867	0.2688	0.2531	0.1772
		0.3	0.1	0.2528	0.1899	0.1274	0.2166	0.1965	0.1323
			0.3	0.2841	0.1994	0.1324	0.2216	0.2016	0.1359
			0.5	0.2228	0.1881	0.1242	0.2109	0.1918	0.1286
		0.6	0.1	0.1215	0.0977	0.0503	0.1196	0.1010	0.0596
			0.3	0.1114	0.1006	0.0522	0.1219	0.1032	0.0616
			0.5	0.1410	0.1042	0.0548	0.1229	0.1044	0.0631
	50	0.1	0.1	0.0933	0.1121	0.0854	0.1181	0.1131	0.0867
			0.3	___	0.1123	0.0850	0.1185	0.1126	0.0868
			0.5	---	0.1134	0.0867	0.1192	0.1143	0.0883
		0.3	0.1	0.0370	0.0868	0.0569	0.0985	0.0898	0.0626
			0.3	0.1572	0.0948	0.0615	0.1035	0.0953	0.0674
			0.5	0.1249	0.0913	0.0576	0.1007	0.0912	0.0636
		0.6	0.1	0.0635	0.0465	0.0199	0.0569	0.0478	0.0272
			0.3	0.0794	0.0516	0.0230	0.0611	0.0519	0.0304
			0.5	0.1866	0.0517	0.0215	0.0602	0.0506	0.0296

Table 2 (con't)
Estimated Bias in Sample R^2 with One Observation Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
	250	0.1	0.1	---	0.207	0.0163	0.0219	0.0207	0.0173
			0.3	---	0.0197	0.0151	0.0207	0.0193	0.0162
			0.5	---	0.0200	0.0158	0.0210	0.0198	0.0166
		0.3	0.1	---	0.0174	0.0103	0.0204	0.0176	0.0124
			0.3	---	0.0169	0.0094	0.0193	0.0167	0.0113
			0.5	---	0.0173	0.0092	0.0202	0.0173	0.0118
		0.6	0.1	---	0.0110	0.0036	0.0142	0.0111	0.0062
			0.3	---	0.0105	0.0028	0.0133	0.0103	0.0052
			0.5	__	0.0118	0.0035	0.0143	0.0113	0.0060

Table 3
Estimated Bias in Sample R^2 with Two Observations Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.4984	0.4540	0.2377	0.6177	0.4761	0.1931
			0.3	0.4934	0.4438	0.2259	0.6105	0.4646	0.1875
			0.5	0.5291	0.4634	0.2319	0.6300	0.4818	0.1982
		0.3	0.1	0.3405	0.3321	0.1192	0.4873	0.3437	0.1241
			0.3	0.3665	0.3044	0.0989	0.4465	0.3202	0.1174
			0.5	0.3270	0.3251	0.1058	0.4760	0.3377	0.1175
		0.6	0.1	0.2796	0.1432	-0.0150	0.3036	0.1556	0.0330
			0.3	0.1823	0.2011	0.0029	0.3162	0.2055	0.0592
			0.5	0.1730	0.1934	-0.0039	0.3112	0.1989	0.0513
	20	0.1	0.1	0.3150	0.1813	0.0855	0.2217	0.1870	0.0885
			0.3	0.2720	0.1795	0.0806	0.2240	0.1831	0.0864
			0.5	0.2816	0.1815	0.0819	0.2286	0.1849	0.0883
		0.3	0.1	0.1867	0.1349	0.0177	0.2001	0.1389	0.0522
			0.3	0.1536	0.1326	0.0092	0.1914	0.1314	0.0473
			0.5	0.1260	0.1386	0.0125	0.1997	0.1356	0.0536
		0.6	0.1	0.0120	0.0777	-0.0368	0.1350	0.0797	0.0197
			0.3	0.1818	0.0880	-0.0356	0.1523	0.0867	0.0225
			0.5	0.1091	0.0865	-0.0353	0.1499	0.0847	0.0232
	100	0.1	0.1	---	0.0257	0.0121	0.0315	0.0264	0.0175
			0.3	---	0.0245	0.0097	0.0307	0.0245	0.0157
			0.5	---	0.0261	0.0109	0.0319	0.0263	0.0164
		0.3	0.1	---	0.0219	-0.0055	0.0380	0.0233	0.0096
			0.3	---	0.0223	-0.0084	0.0382	0.0223	0.0084
			0.5	---	0.0248	-0.0066	0.0388	0.0236	0.0095
		0.6	0.1	---	0.0143	-0.0158	0.0326	0.0162	0.0027
			0.3	---	0.0178	-0.0144	0.0350	0.0178	0.0044
			0.5	---	0.0187	-0.0148	0.0357	0.0186	0.0044
5	25	0.1	0.1	0.5853	0.3098	0.1904	0.3576	0.3207	0.1809
			0.3	0.4386	0.3081	0.1890	0.3550	0.3181	0.1797
			0.5	0.4233	0.3119	0.1912	0.3562	0.3197	0.1772
		0.3	0.1	0.2274	0.2393	0.1270	0.2937	0.2509	0.1323
			0.3	---	0.2532	0.1317	0.3016	0.2592	0.1359
			0.5	0.1178	0.2411	0.1241	0.2879	0.2474	0.1286
		0.6	0.1	---	0.1287	0.0421	0.1666	0.1349	0.0596
			0.3	0.1186	0.1344	0.0443	0.1706	0.1368	0.0616
			0.5	0.1879	0.1380	0.0481	0.1735	0.1396	0.0631
	50	0.1	0.1	---	0.1349	0.0847	0.1484	0.1370	0.0867
			0.3	---	0.1354	0.0843	0.1475	0.1361	0.0868
			0.5	---	0.1354	0.0859	0.1472	0.1382	0.0883
		0.3	0.1	---	0.1081	0.0528	0.1299	0.1136	0.0626
			0.3	---	0.1176	0.0574	0.1349	0.1190	0.0674
			0.5	---	0.1141	0.0526	0.1333	0.1154	0.0636
		0.6	0.1	---	0.0623	0.0140	0.0807	0.0640	0.0272
			0.3	---	0.0689	0.0166	0.0854	0.0691	0.0304
			0.5	---	0.0693	0.0151	0.0850	0.0683	0.0296

Table 3 (con't)
Estimated Bias in Sample R^2 with Two Observations Removed from Samples Drawn from Single Population

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
	250	0.1	0.1	—	0.234	0.0156	0.0257	0.0235	0.0173
			0.3	—	0.0223	0.0143	0.0245	0.0221	0.0162
			0.5	---	0.0227	0.0149	0.0246	0.0226	0.0166
		0.3	0.1	—	0.0216	0.0085	0.0270	0.0221	0.0124
			0.3	—	0.0210	0.0079	0.0257	0.0209	0.0113
			0.5	---	0.0219	0.0073	0.0269	0.0218	0.0118
		0.6	0.1	—	0.0149	0.0015	0.0207	0.0154	0.0062
			0.3	---	0.0149	0.0008	0.0199	0.0147	0.0052
			0.5	—	0.0164	0.0014	0.0208	0.0208	0.0060

Table 4
Estimated Bias in Sample Regression Weights with One Observation Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.0166	0.0041	0.0036	0.0067	0.0020	0.0016
			0.3	0.0347	0.0086	0.0106	0.0134	0.0110	0.0096
			0.5	0.0151	0.0082	0.0077	0.0233	0.0125	0.0101
		0.3	0.1	0.0199	0.0050	0.0079	0.0072	0.0064	0.0028
			0.3	0.0261	0.0113	0.0146	0.0197	0.0103	0.0090
			0.5	0.0241	0.0090	0.0088	0.0137	0.0084	0.0088
		0.6	0.1	0.0737	0.0247	0.0086	0.0412	0.0381	0.0160
			0.3	0.0112	0.0041	0.0049	0.0091	0.0054	0.0027
			0.5	0.0205	0.0104	0.0036	0.0105	0.0090	0.0058
	20	0.1	0.1	0.0219	0.0029	0.0033	0.0054	0.0030	0.0030
			0.3	0.0157	0.0018	0.0020	0.0053	0.0022	0.0031
			0.5	0.0216	0.0042	0.0015	0.0040	0.0044	0.0017
		0.3	0.1	0.0136	0.0038	0.0030	0.0042	0.0023	0.0041
			0.3	0.0292	0.0036	0.0016	0.0035	0.0037	0.0023
			0.5	0.0292	0.0019	0.0016	0.0021	0.0024	0.0009
		0.6	0.1	0.0149	0.0026	0.0018	0.0020	0.0019	0.0019
			0.3	0.0096	0.0011	0.0025	0.0017	0.0007	0.0015
			0.5	0.0292	0.0025	0.0021	0.0043	0.0025	0.0018
	100	0.1	0.1	—	0.0006	0.0005	0.0005	0.0005	0.0005
			0.3	—	0.0011	0.0008	0.0011	0.0011	0.0009
			0.5	—	0.0008	0.0010	0.0016	0.0008	0.0011
		0.3	0.1	—	0.0014	0.0012	0.0010	0.0011	0.0009
			0.3	—	0.0021	0.0019	0.0014	0.0021	0.0016
			0.5	—	0.0007	0.0010	0.0007	0.0006	0.0009
		0.6	0.1	—	0.0008	0.0009	0.0010	0.0007	0.0010
			0.3	—	0.0006	0.0005	0.0003	0.0004	0.0004
			0.5	—	0.0014	0.0016	0.0014	0.0016	0.0016
5	25	0.1	0.1	0.0348	0.0019	0.0009	0.0020	0.0016	0.0012
			0.3	0.0264	0.0024	0.0024	0.0033	0.0023	0.0018
			0.5	0.0164	0.0071	0.0080	0.0055	0.0067	0.0069
		0.3	0.1	0.0126	0.0033	0.0035	0.0039	0.0040	0.0035
			0.3	0.0410	0.0046	0.0036	0.0024	0.0028	0.0026
			0.5	0.0137	0.0074	0.0051	0.0040	0.0072	0.0045
		0.6	0.1	0.0158	0.0020	0.0025	0.0020	0.0021	0.0017
			0.3	0.0332	0.0017	0.0019	0.0024	0.0020	0.0022
			0.5	0.0197	0.0033	0.0024	0.0031	0.0029	0.0025
	50	0.1	0.1	0.1074	0.0011	0.0014	0.0016	0.0013	0.0010
			0.3	—	0.0015	0.0018	0.0012	0.0017	0.0015
			0.5	—	0.0022	0.0025	0.0019	0.0023	0.0025
		0.3	0.1	0.1379	0.0015	0.0015	0.0009	0.0013	0.0014
			0.3	0.1141	0.0023	0.0017	0.0015	0.0023	0.0014
			0.5	0.1784	0.0017	0.0013	0.0020	0.0012	0.0021
		0.6	0.1	0.0312	0.0011	0.0009	0.0014	0.0009	0.0009
			0.3	0.1695	0.0010	0.0018	0.0014	0.0011	0.0015
			0.5	0.1181	0.0019	0.0023	0.0023	0.0022	0.0019

Table 4 (con't)
Estimated Bias in Sample Regression Weights with One Observation Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
	250	0.1	0.1	—	0.0004	0.0005	0.0006	0.0005	0.0005
			0.3	—	0.0008	0.0008	0.0007	0.0008	0.0011
			0.5	—	0.0014	0.0013	0.0015	0.0013	0.0014
		0.3	0.1	—	0.0009	0.0008	0.0008	0.0009	0.0009
			0.3	—	0.0025	0.0023	0.0023	0.0022	0.0024
			0.5	—	0.0014	0.0016	0.0016	0.0013	0.0013
		0.6	0.1	—	0.0004	0.0004	0.0004	0.0004	0.0004
			0.3	—	0.0004	0.0004	0.0005	0.0005	0.0005
			0.5	—	0.0009	0.0009	0.0010	0.0010	0.0009

Table 5
Estimated Bias in Sample Regression Weights with Two Observations Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.0150	0.0045	0.0031	0.0108	0.0016	0.0016
			0.3	0.1041	0.0131	0.0169	0.0119	0.0107	0.0096
			0.5	0.1326	0.0192	0.0176	0.0299	0.0238	0.0101
		0.3	0.1	0.0370	0.0051	0.0061	0.0100	0.0115	0.0028
			0.3	0.0409	0.0124	0.0094	0.0139	0.0073	0.0090
			0.5	0.0626	0.0134	0.0123	0.0327	0.0169	0.0088
		0.6	0.1	0.2378	0.0277	0.0289	0.0539	0.0475	0.0160
			0.3	0.0652	0.0061	0.0031	0.0077	0.0097	0.0027
			0.5	0.0728	0.0156	0.0068	0.0306	0.0166	0.0058
	20	0.1	0.1	0.0536	0.0025	0.0035	0.0057	0.0011	0.0030
			0.3	0.0722	0.0031	0.0029	0.0067	0.0027	0.0031
			0.5	0.2094	0.0035	0.0017	0.0032	0.0030	0.0017
		0.3	0.1	0.0930	0.0030	0.0023	0.0052	0.0029	0.0041
			0.3	0.0705	0.0038	0.0022	0.0042	0.0051	0.0023
			0.5	0.0853	0.0018	0.0026	0.0037	0.0035	0.0009
		0.6	0.1	0.1346	0.0025	0.0027	0.0031	0.0020	0.0019
			0.3	0.1835	0.0010	0.0017	0.0024	0.0006	0.0015
			0.5	0.0898	0.0032	0.0010	0.0051	0.0030	0.0018
	100	0.1	0.1	—	0.0006	0.0006	0.0004	0.0003	0.0005
			0.3	—	0.0012	0.0008	0.0012	0.0014	0.0009
			0.5	—	0.0015	0.0007	0.0016	0.0009	0.0011
		0.3	0.1	—	0.0014	0.0010	0.0010	0.0011	0.0009
			0.3	—	0.0022	0.0018	0.0011	0.0020	0.0016
			0.5	—	0.0008	0.0011	0.0005	0.0006	0.0009
		0.6	0.1	—	0.0009	0.0009	0.0010	0.0008	0.0010
			0.3	—	0.0008	0.0005	0.0005	0.0007	0.0004
			0.5	—	0.0015	0.0017	0.0019	0.0015	0.0016
5	25	0.1	0.1	0.1690	0.0024	0.0021	0.0039	0.0022	0.0012
			0.3	—	0.0017	0.0017	0.0020	0.0017	0.0015
			0.5	0.1730	0.0062	0.0067	0.0055	0.0064	0.0069
		0.3	0.1	0.1894	0.0031	0.0036	0.0051	0.0026	0.0035
			0.3	—	0.0058	0.0052	0.0045	0.0038	0.0026
			0.5	0.1653	0.0077	0.0037	0.0025	0.0061	0.0045
		0.6	0.1	—	0.0025	0.0028	0.0022	0.0021	0.0017
			0.3	0.1555	0.0018	0.0016	0.0038	0.0030	0.0022
			0.5	0.3374	0.0037	0.0021	0.0032	0.0038	0.0025
	50	0.1	0.1	—	0.0009	0.0017	0.0016	0.0014	0.0010
			0.3	—	0.0017	0.0020	0.0017	0.0016	0.0015
			0.5	—	0.0020	0.0014	0.0028	0.0019	0.0025
		0.3	0.1	—	0.0012	0.0017	0.0008	0.0014	0.0014
			0.3	—	0.0025	0.0017	0.0020	0.0025	0.0014
			0.5	—	0.0022	0.0015	0.0034	0.0018	0.0021
		0.6	0.1	—	0.0012	0.0009	0.0016	0.0011	0.0009
			0.3	—	0.0009	0.0017	0.0015	0.0013	0.0015
			0.5	—	0.0022	0.0018	0.0028	0.0024	0.0019

Table 5 (con't)
Estimated Bias in Sample Regression Weights with Two Observations Removed from Samples Drawn from Single Population.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
	250	0.1	0.1	—	0.0008	0.0008	0.0007	0.0008	0.0007
			0.3	—	0.0013	0.0012	0.0012	0.0012	0.0011
			0.5	—	0.0013	0.0013	0.0014	0.0012	0.0014
		0.3	0.1	—	0.0009	0.0009	0.0009	0.0009	0.0009
			0.3	—	0.0024	0.0022	0.0024	0.0024	0.0024
			0.5	—	0.0014	0.0016	0.0015	0.0012	0.0013
		0.6	0.1	—	0.0004	0.0004	0.0004	0.0005	0.0004
			0.3	—	0.0005	0.0005	0.0004	0.0004	0.0005
			0.5	—	0.0011	0.0008	0.0011	0.0010	0.0009

Table 6
Agreement Among Indices in Cases Identified.

Mean Agreement Among Methods

	DFFIT	Leverage	RStudent	DFBETAs	Cook D
DFFIT	----	0.19	0.4	0.73	0.08
Leverage	0.24	----	0.03	0.18	0.06
RStudent	0.43	0.05	----	0.37	0.19
DFBETAs	0.76	0.24	0.38	----	0.06
Cook D	0.06	0.06	0.27	0.04	----

Maximum Agreement Among Methods

	DFFIT	Leverage	RStudent	DFBETAs	Cook D
DFFIT	----	0.34	0.48	0.85	0.29
Leverage	0.37	----	0.08	0.34	0.24
RStudent	0.51	0.1	----	0.43	0.49
DFBETAs	0.86	0.38	0.47	----	0.26
Cook D	0.22	0.16	0.65	0.18	----

Minimum Agreement Among Methods

	DFFIT	Leverage	RStudent	DFBETAs	Cook D
DFFIT	----	0.12	0.3	0.61	0
Leverage	0.16	----	0.01	0.1	0
RStudent	0.35	0.02	----	0.28	0
DFBETAs	0.66	0.16	0.3	----	0
Cook D	0	0	0	0	----

Note. Values above the diagonal are agreement on the first observation. Those below the diagonal are agreement on the first two without regard to order of identification.

Table 7
Proportions of Samples in which an Obersvation with an Extreme Residual was Flagged.

				DFFITS		Leverage		RStudent		DFBETAs		Cook's D	
k	ρ^2	$\bar{r}_{12}$	N	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
2	0.10	0.10	10	0.73	0.14	0.10	0.09	0.77	0.02	0.71	0.16	0.18	0.05
			20	0.79	0.14	0.05	0.05	0.93	0.04	0.76	0.15	0.07	0.00
			100	0.58	0.18	0.01	0.01	0.93	0.06	0.55	0.19	0.00	0.00
		0.30	10	0.74	0.14	0.10	0.10	0.78	0.02	0.71	0.16	0.19	0.05
			20	0.79	0.14	0.05	0.05	0.93	0.04	0.75	0.16	0.07	0.00
			100	0.56	0.19	0.01	0.01	0.92	0.07	0.54	0.19	0.00	0.00
		0.50	10	0.74	0.15	0.10	0.11	0.79	0.02	0.70	0.17	0.18	0.05
			20	0.80	0.14	0.05	0.05	0.93	0.04	0.76	0.16	0.07	0.00
			100	0.56	0.19	0.01	0.01	0.93	0.06	0.52	0.20	0.00	0.00
	0.30	0.10	10	0.74	0.15	0.10	0.10	0.77	0.02	0.70	0.16	0.18	0.05
			20	0.80	0.14	0.05	0.05	0.93	0.03	0.76	0.16	0.07	0.00
			100	0.56	0.18	0.01	0.01	0.93	0.06	0.54	0.19	0.00	0.00
		0.30	10	0.74	0.15	0.10	0.10	0.78	0.02	0.70	0.17	0.18	0.04
			20	0.79	0.15	0.05	0.05	0.93	0.03	0.76	0.16	0.08	0.00
			100	0.57	0.18	0.01	0.01	0.93	0.07	0.54	0.19	0.00	0.00
		0.50	10	0.73	0.15	0.10	0.10	0.77	0.02	0.69	0.17	0.18	0.05
			20	0.79	0.14	0.05	0.05	0.92	0.04	0.74	0.17	0.07	0.00
			100	0.56	0.19	0.01	0.01	0.93	0.06	0.52	0.20	0.00	0.00
	0.60	0.10	10	0.73	0.15	0.10	0.10	0.77	0.03	0.70	0.16	0.17	0.05
			20	0.80	0.14	0.05	0.05	0.93	0.04	0.76	0.15	0.07	0.00
			100	0.56	0.19	0.01	0.01	0.93	0.06	0.54	0.19	0.00	0.00
		0.30	10	0.74	0.14	0.09	0.11	0.78	0.02	0.70	0.17	0.18	0.05
			20	0.79	0.15	0.05	0.05	0.93	0.04	0.75	0.16	0.08	0.00
			100	0.56	0.20	0.01	0.01	0.93	0.07	0.53	0.19	0.00	0.00
		0.50	10	0.73	0.15	0.11	0.10	0.79	0.02	0.69	0.17	0.19	0.04
			20	0.78	0.15	0.05	0.05	0.93	0.04	0.74	0.17	0.07	0.00
			100	0.57	0.18	0.01	0.01	0.93	0.06	0.53	0.20	0.00	0.00

Table 7 (Continued)
Proportions of Samples in which an Obersvation with an Extreme Residual was Flagged.

				DFFITS		Leverage		RStudent		DFBETAs		Cook's D	
k	ρ^2	$\bar{r}_{12}$	N	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
5	0.10	0.10	25	0.72	0.16	0.04	0.03	0.87	0.06	0.66	0.18	0.04	0.00
			50	0.71	0.16	0.02	0.02	0.91	0.06	0.65	0.18	0.00	0.00
			250	0.49	0.16	0.00	0.00	0.86	0.11	0.50	0.15	0.00	0.00
		0.30	25	0.72	0.15	0.04	0.03	0.86	0.06	0.64	0.18	0.04	0.00
			50	0.70	0.17	0.02	0.02	0.92	0.06	0.64	0.18	0.00	0.00
			250	0.50	0.17	0.00	0.00	0.85	0.11	0.48	0.15	0.00	0.00
		0.50	25	0.72	0.16	0.04	0.04	0.85	0.06	0.64	0.18	0.04	0.00
			50	0.71	0.16	0.02	0.02	0.92	0.06	0.62	0.17	0.00	0.00
			250	0.50	0.16	0.00	0.00	0.86	0.11	0.45	0.15	0.00	0.00
	0.30	0.10	25	0.71	0.16	0.04	0.03	0.85	0.06	0.62	0.18	0.04	0.00
			50	0.71	0.16	0.02	0.02	0.92	0.06	0.60	0.18	0.00	0.00
			250	0.50	0.16	0.00	0.00	0.85	0.11	0.41	0.15	0.00	0.00
		0.30	25	0.72	0.15	0.04	0.03	0.87	0.06	0.65	0.18	0.04	0.00
			50	0.72	0.15	0.02	0.02	0.91	0.07	0.66	0.17	0.00	0.00
			250	0.50	0.17	0.00	0.01	0.86	0.11	0.49	0.16	0.00	0.00
		0.50	25	0.73	0.15	0.04	0.04	0.86	0.06	0.64	0.17	0.04	0.00
			50	0.70	0.16	0.02	0.02	0.91	0.07	0.64	0.17	0.00	0.00
			250	0.49	0.16	0.00	0.00	0.86	0.11	0.46	0.15	0.00	0.00
	0.60	0.10	25	0.71	0.16	0.04	0.03	0.86	0.07	0.64	0.19	0.04	0.00
			50	0.71	0.17	0.01	0.02	0.91	0.07	0.66	0.17	0.00	0.00
			250	0.50	0.16	0.00	0.00	0.85	0.11	0.50	0.15	0.00	0.00
		0.30	25	0.71	0.17	0.04	0.04	0.87	0.06	0.64	0.19	0.04	0.00
			50	0.70	0.16	0.02	0.02	0.92	0.06	0.64	0.18	0.00	0.00
		0.50	25	0.72	0.16	0.04	0.04	0.86	0.06	0.65	0.17	0.04	0.00
			50	0.71	0.16	0.02	0.02	0.91	0.07	0.64	0.18	0.00	0.00
			250	0.48	0.17	0.01	0.00	0.85	0.12	0.46	0.15	0.00	0.00

Table 8

Proportions of Samples in which an Obersvation with Extreme Regressor Values was Flagged.

				DFFITS		Leverage		RStudent		DFBETAs		Cook's D	
k	ρ^2	$\bar{r}_{12}$	N	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
2	0.10	0.10	10	0.43	0.16	0.84	0.09	0.05	0.02	0.28	0.14	0.35	0.02
			20	0.40	0.15	0.85	0.08	0.05	0.03	0.28	0.14	0.16	0.00
			100	0.30	0.09	0.81	0.08	0.03	0.02	0.22	0.10	0.00	0.00
		0.30	10	0.44	0.16	0.78	0.11	0.08	0.02	0.27	0.14	0.35	0.03
			20	0.46	0.12	0.82	0.09	0.10	0.04	0.30	0.14	0.20	0.00
			100	0.39	0.09	0.73	0.10	0.06	0.04	0.28	0.11	0.00	0.00
		0.50	10	0.40	0.17	0.73	0.12	0.07	0.02	0.21	0.14	0.31	0.03
			20	0.37	0.12	0.74	0.11	0.07	0.04	0.20	0.11	0.13	0.00
			100	0.28	0.09	0.62	0.11	0.04	0.03	0.15	0.09	0.00	0.00
	0.30	0.10	20	0.74	0.08	0.86	0.07	0.30	0.07	0.59	0.13	0.45	0.00
			100	0.80	0.07	0.81	0.08	0.34	0.11	0.71	0.11	0.06	0.00
		0.30	10	0.65	0.12	0.79	0.10	0.21	0.03	0.44	0.14	0.56	0.03
			20	0.75	0.08	0.81	0.10	0.33	0.07	0.56	0.15	0.45	0.00
			100	0.79	0.07	0.71	0.11	0.36	0.11	0.64	0.14	0.05	0.00
		0.50	10	0.59	0.13	0.74	0.11	0.19	0.03	0.35	0.14	0.48	0.03
			20	0.68	0.10	0.75	0.11	0.30	0.07	0.47	0.16	0.37	0.00
			100	0.72	0.09	0.65	0.11	0.31	0.11	0.52	0.15	0.03	0.00
	0.60	0.10	10	0.84	0.07	0.83	0.09	0.41	0.03	0.65	0.13	0.80	0.03
			20	0.95	0.03	0.85	0.08	0.67	0.05	0.84	0.09	0.80	0.00
			100	0.98	0.01	0.80	0.08	0.83	0.07	0.95	0.03	0.31	0.00
		0.30	10	0.84	0.07	0.78	0.11	0.44	0.03	0.63	0.13	0.78	0.03
			20	0.94	0.03	0.80	0.10	0.68	0.06	0.81	0.10	0.76	0.00
			100	0.97	0.02	0.72	0.10	0.82	0.07	0.92	0.05	0.25	0.00
		0.50	10	0.81	0.08	0.74	0.12	0.42	0.03	0.56	0.15	0.71	0.03
			20	0.91	0.04	0.74	0.11	0.63	0.07	0.73	0.13	0.65	0.00
			100	0.94	0.03	0.65	0.11	0.77	0.09	0.84	0.09	0.15	0.00

Table 8 (Continued)
Proportions of Samples in which an Obersvation with Extreme Regressor Values was Flagged.

				DFFITS		Leverage		RStudent		DFBETAs		Cook's D	
k	ρ^2	$\bar{r}_{12}$	N	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
5	0.10	0.10	25	0.48	0.12	0.95	0.04	0.09	0.04	0.25	0.11	0.23	0.00
			50	0.53	0.10	0.97	0.02	0.10	0.05	0.34	0.10	0.08	0.00
			250	0.56	0.07	0.97	0.02	0.08	0.04	0.40	0.10	0.00	0.00
		0.30	25	0.40	0.12	0.76	0.11	0.11	0.05	0.16	0.09	0.13	0.00
			50	0.42	0.10	0.79	0.09	0.11	0.06	0.18	0.09	0.03	0.00
			250	0.39	0.09	0.72	0.09	0.07	0.04	0.15	0.08	0.00	0.00
		0.50	25	0.29	0.10	0.62	0.13	0.08	0.04	0.11	0.07	0.06	0.00
			50	0.27	0.09	0.62	0.13	0.06	0.04	0.10	0.07	0.01	0.00
			250	0.21	0.06	0.52	0.11	0.03	0.02	0.07	0.04	0.00	0.00
	0.30	0.10	25	0.63	0.12	1.00	0.00	0.08	0.03	0.50	0.13	0.43	0.00
			50	0.68	0.08	1.00	0.00	0.08	0.05	0.64	0.11	0.27	0.00
			250	0.78	0.04	1.00	0.00	0.10	0.05	0.82	0.05	0.02	0.00
		0.30	25	0.51	0.11	0.78	0.10	0.17	0.05	0.22	0.11	0.18	0.00
			50	0.57	0.10	0.80	0.09	0.20	0.08	0.27	0.11	0.05	0.00
			250	0.56	0.08	0.74	0.08	0.16	0.07	0.28	0.11	0.00	0.00
		0.50	25	0.50	0.11	0.64	0.13	0.21	0.05	0.23	0.11	0.15	0.00
			50	0.54	0.10	0.63	0.13	0.23	0.09	0.27	0.12	0.03	0.00
			250	0.50	0.10	0.53	0.11	0.16	0.08	0.23	0.09	0.00	0.00
	0.60	0.10	25	0.75	0.07	0.95	0.03	0.28	0.06	0.49	0.12	0.48	0.00
			50	0.85	0.04	0.97	0.02	0.39	0.09	0.68	0.09	0.33	0.00
			250	0.92	0.03	0.97	0.02	0.45	0.11	0.83	0.06	0.01	0.00
		0.30	25	0.93	0.03	0.79	0.10	0.64	0.07	0.61	0.15	0.67	0.00
			50	0.98	0.01	0.82	0.09	0.79	0.07	0.79	0.10	0.48	0.00
			250	0.99	0.01	0.75	0.08	0.87	0.05	0.87	0.07	0.01	0.00
		0.50	25	0.84	0.06	0.59	0.13	0.58	0.07	0.52	0.13	0.42	0.00
			50	0.91	0.04	0.59	0.13	0.71	0.09	0.62	0.15	0.18	0.00
			250	0.92	0.04	0.47	0.11	0.72	0.09	0.65	0.12	0.00	0.00

Table 9

Estimated Bias in Sample R^2 with One Observation Removed in which an Obersvation with an Extreme Residual was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.2622	0.2282	0.1245	0.2422	0.2311	0.1045
			0.3	0.3109	0.2684	0.1749	0.2711	0.2683	0.1531
			0.5	0.306	0.2662	0.177	0.2663	0.2658	0.1532
		0.3	0.1	0.1798	0.1678	0.0277	0.1922	0.1726	0.0144
			0.3	0.155	0.1482	0.0006	0.1759	0.1513	-0.0101
			0.5	0.1713	0.1613	0.0136	0.187	0.1648	0.0053
		0.6	0.1	0.0123	0.0385	-0.154	0.0875	0.0451	-0.1335
			0.3	0.0332	0.0489	-0.136	0.0965	0.0575	-0.1245
			0.5	0.0335	0.0493	-0.1364	0.0944	0.0565	-0.1236
	20	0.1	0.1	0.1083	0.0835	0.018	0.0877	0.0854	0.0224
			0.3	0.1307	0.1028	0.0572	0.1036	0.1031	0.05675
			0.5	0.1217	0.099	0.0546	0.1002	0.0978	0.0543
		0.3	0.1	0.0746	0.0523	-0.0466	0.0661	0.055	-0.0326
			0.3	0.0857	0.0475	-0.0585	0.0609	0.0503	-0.045
			0.5	0.0683	0.0448	-0.0556	0.058	0.0464	-0.0416
		0.6	0.1	0.0238	-0.0011	-0.1337	0.0216	0.0029	-0.1043
			0.3	0.0243	0.0064	-0.1253	0.0282	0.0103	-0.0992
			0.5	0.0001	0.0036	-0.1299	0.0257	0.0086	-0.1017
	100	0.1	0.1	0.0426	0.009	-0.0085	0.0141	0.0097	-0.0033
			0.3	0.0074	0.014	0.0028	0.0171	0.0143	0.006
			0.5	---	0.0148	0.0036	0.0174	0.0147	0.0066
		0.3	0.1	0.0204	0.0027	-0.0231	0.0111	0.0027	-0.015
			0.3	0.0843	0.0016	-0.0268	0.0101	0.0018	-0.0181
			0.5	0.1534	0.0015	-0.0255	0.0103	0.0016	-0.0176
		0.6	0.1	---	-0.0074	-0.0379	0.0033	-0.0068	-0.0275
			0.3	0.0109	-0.0065	-0.0386	0.0041	-0.006	-0.0283
			0.5	-0.1265	-0.0065	-0.0381	0.0041	-0.0059	-0.0279

Table 9 (Continued)
Estimated Bias in Sample R^2 with One Observation Removed in which an Obersvation with an Extreme Residual was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
5	25	0.1	0.1	0.2630	0.2075	0.1542	0.2071	0.2102	0.1484
			0.3	0.2611	0.2101	0.1540	0.2087	0.2088	0.1480
			0.5	0.2537	0.2079	0.1544	0.2095	0.2089	0.1492
		0.3	0.1	0.2041	0.1434	0.0641	0.1549	0.1484	0.0629
			0.3	0.1906	0.1546	0.0743	0.1627	0.1583	0.0744
			0.5	0.176	0.1385	0.0592	0.1495	0.1414	0.0587
		0.6	0.1	0.0900	0.0492	-0.0393	0.0671	0.0549	-0.0314
			0.3	0.1011	0.0568	-0.0328	0.0745	0.0607	-0.0249
			0.5	0.0926	0.0539	-0.0359	0.0699	0.0574	-0.0280
	50	0.1	0.1	0.1753	0.092	0.0646	0.0931	0.0924	0.0652
			0.3	0.1231	0.0915	0.0644	0.093	0.092	0.0654
			0.5	0.1012	0.0926	0.066	0.0934	0.0936	0.0663
		0.3	0.1	0.0948	0.0582	0.0119	0.0662	0.0607	0.0168
			0.3	0.1522	0.0641	0.0183	0.0693	0.0652	0.0226
			0.5	0.0881	0.0601	0.0116	0.0667	0.0608	0.0165
		0.6	0.1	0.0518	0.0211	-0.0317	0.031	0.0236	-0.0241
			0.3	0.0593	0.0197	-0.0344	0.0298	0.0214	-0.0268
			0.5	0.0509	0.0213	-0.0329	0.0313	0.0238	-0.0254
	250	0.1	0.1	---	0.0166	0.0109	0.0179	0.0167	0.012
			0.3	---	0.0165	0.0109	0.0177	0.0166	0.0118
			0.5	---	0.0168	0.0114	0.0179	0.0168	0.0123
		0.3	0.1	---	0.0077	-0.0027	0.0112	0.0083	-0.0004
			0.3	---	0.0107	0.0003	0.0132	0.0104	0.0024
			0.5	---	0.009	-0.0019	0.012	0.0089	0.0003
		0.6	0.1	---	0.0027	-0.0088	0.0063	0.0029	-0.0063
			0.3	---	0.0022	-0.0099	0.0055	0.0021	-0.0074
			0.5	___	0.0034	-.0085	0.0068	0.0033	-0.0060

Table 10

Estimated Bias in Sample R^2 with One Observation Removed in which an Obersvation with Extreme Regressor Values was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.2892	0.2784	0.1851	0.3681	0.2811	0.1578
			0.3	0.3323	0.3054	0.1984	0.3781	0.3114	0.151
			0.5	0.3416	0.3078	0.2036	0.3913	0.3157	0.1653
		0.3	0.1	0.2200	0.1892	0.1130	0.2557	0.1994	0.0030
			0.3	0.1997	0.1597	0.0847	0.2348	0.1749	-0.0231
			0.5	0.2156	0.1616	0.0792	0.2407	0.1801	-0.0069
		0.6	0.1	0.0721	0.0492	-0.0058	0.1125	0.0672	-0.156
			0.3	0.0883	0.0507	-0.0171	0.118	0.0756	-0.1861
			0.5	0.0929	0.0491	-0.0222	0.1198	0.0787	-0.1691
	20	0.1	0.1	0.1668	0.1121	0.0743	0.1341	0.1134	0.064
			0.3	0.2011	0.1186	0.0839	0.1259	0.1239	0.0525
			0.5	0.2027	0.1207	0.083	0.1323	0.1254	0.067
		0.3	0.1	0.1336	0.0609	0.0348	0.0631	0.0693	-0.0535
			0.3	0.1185	0.0493	0.02	0.0591	0.0614	-0.0661
			0.5	0.1339	0.0523	0.0192	0.0589	0.0664	-0.0521
		0.6	0.1	0.0362	0.0107	-0.0223	0.0243	0.0202	-0.1522
			0.3	0.046	0.0118	-0.0319	0.0244	0.0256	-0.173
			0.5	0.0505	0.0061	-0.0386	0.0248	0.0237	-0.1506
	100	0.1	0.1	0.0462	0.0178	0.0119	0.0221	0.0186	0.0102
			0.3	0.0672	0.0165	0.0123	0.0142	0.0185	0.0048
			0.5	0.1402	0.0184	0.0139	0.019	0.0209	0.0105
		0.3	0.1	0.057	0.0093	0.0013	0.0042	0.0111	-0.0272
			0.3	0.0447	0.0067	-0.0055	0.0047	0.0095	-0.0288
			0.5	0.0621	0.0056	-0.0055	0.0056	0.0103	-0.0234
		0.6	0.1	0.0184	0.0014	-0.0124	0.0018	0.0021	-0.0533
			0.3	0.0191	0.0016	-0.0181	0.0028	0.0033	-0.0591
			0.5	0.028	0.003	-0.0179	0.0053	0.0054	-0.0482

Table 10 (Continued)
Estimated Bias in Sample R^2 with One Observation Removed in which an Obersvation with Extreme Regressor Values was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
5	25	0.1	0.1	0.2777	0.2292	0.1839	0.2428	0.2303	0.1491
			0.3	0.3018	0.2184	0.1727	0.2366	0.2274	0.1406
			0.5	0.2914	0.229	0.1738	0.2498	0.238	0.1573
		0.3	0.1	0.1736	0.1692	0.134	0.1996	0.1656	0.1097
			0.3	0.2331	0.1641	0.1248	0.1857	0.1736	0.0843
			0.5	0.2272	0.1524	0.1067	0.1745	0.1643	0.0714
		0.6	0.1	0.0895	0.0741	0.0578	0.0892	0.0743	0.0116
			0.3	0.0894	0.0512	0.0282	0.0731	0.0689	-0.0706
			0.5	0.0987	0.053	0.0088	0.0758	0.0713	-0.0468
	50	0.1	0.1	0.1628	0.0984	0.0854	0.0963	0.1009	0.0605
			0.3	0.1984	0.0942	0.0815	0.0961	0.1023	0.0612
			0.5	0.2162	0.1032	0.0834	0.1071	0.108	0.0754
		0.3	0.1	0.1017	0.0758	0.0644	0.0854	0.076	0.0473
			0.3	0.1563	0.0712	0.0578	0.0767	0.0795	0.0292
			0.5	0.1446	0.064	0.0447	0.0714	0.073	0.0222
		0.6	0.1	0.0519	0.0304	0.0248	0.035	0.0303	-0.0115
			0.3	0.0524	0.025	0.0081	0.031	0.0317	-0.0684
			0.5	0.063	0.0199	-0.0092	0.0321	0.0312	-0.0469
	250	0.1	0.1	---	0.0179	0.0172	0.0147	0.019	0.0089
			0.3	---	0.0161	0.0153	0.0159	0.0187	0.0102
			0.5	---	0.0171	0.0147	0.0177	0.0185	0.0129
		0.3	0.1	0.0425	0.0147	0.013	0.0147	0.0148	0.0056
			0.3	---	0.0139	0.0111	0.0143	0.0161	0.0038
			0.5	---	-.03092	-0.3138	-0.3076	-0.3067	-0.3183
		0.6	0.1	0.0236	0.0064	0.0054	0.0068	0.0063	-0.0064
			0.3	0.0163	0.0045	-0.0022	0.0055	0.0056	-0.0225
			0.5	0.0264	0.0031	-0.0068	0.0056	0.0055	-0.0147

Table 11
Estimated Bias in Regression Weights with One Observation Removed in which an Obersvation with an Extreme Residual was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.0098	0.0113	0.0083	0.0064	0.0079	0.0062
			0.3	0.0087	0.0049	0.0082	0.0037	0.0054	0.0060
			0.5	0.0171	0.0034	0.0028	0.0056	0.0040	0.0050
		0.3	0.1	0.0100	0.0027	0.0092	0.0032	0.0024	0.0096
			0.3	0.0099	0.0057	0.0051	0.0068	0.0058	0.0070
			0.5	0.0053	0.0081	0.0043	0.0070	0.0076	0.0072
		0.6	0.1	0.0065	0.0038	0.0052	0.0045	0.0046	0.0058
			0.3	0.0127	0.0055	0.0077	0.0030	0.0054	0.0055
			0.5	0.0071	0.0017	0.0049	0.0038	0.0025	0.0028
	20	0.1	0.1	0.0256	0.0037	0.0015	0.0037	0.0041	0.0016
			0.3	0.0139	0.0035	0.0033	0.0029	0.0035	0.0026
			0.5	0.0089	0.0068	0.0107	0.0082	0.0090	0.0116
		0.3	0.1	0.0023	0.0046	0.0041	0.0042	0.0036	0.0027
			0.3	0.0206	0.0025	0.0027	0.0018	0.0040	0.0021
			0.5	0.0080	0.0052	0.0037	0.0045	0.0052	0.0037
		0.6	0.1	0.0086	0.0014	0.0020	0.0016	0.0017	0.0019
			0.3	0.0047	0.0047	0.0045	0.0029	0.0037	0.0023
			0.5	0.0053	0.0031	0.0036	0.0025	0.0024	0.0033
	100	0.1	0.1	0.0383	0.0009	0.0015	0.0010	0.0009	0.0013
			0.3	0.1272	0.0009	0.0015	0.0011	0.0012	0.0013
			0.5	---	0.0018	0.0024	0.0020	0.0019	0.0022
		0.3	0.1	0.0927	0.0011	0.0010	0.0010	0.0011	0.0011
			0.3	0.0920	0.0011	0.0011	0.0008	0.0012	0.0012
			0.5	0.2340	0.0006	0.0003	0.0003	0.0007	0.0005
		0.6	0.1	---	0.0008	0.0009	0.0010	0.0009	0.0010
			0.3	0.0553	0.0005	0.0007	0.0007	0.0006	0.0004
			0.5	0.0766	0.0005	0.0004	0.0006	0.0007	0.0006

Table 11 (Continued)
Estimated Bias in Regression Weights with One Observation Removed in which an Obersvation with an Extreme Residual was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
5	25	0.1	0.1	0.0107	0.0034	0.0050	0.0025	0.0033	0.0036
			0.3	0.0245	0.0031	0.0040	0.0031	0.0030	0.0033
			0.5	0.0121	0.0053	0.0033	0.0050	0.0038	0.0035
		0.3	0.1	0.0120	0.0021	0.0026	0.0031	0.0036	0.0027
			0.3	0.0175	0.0042	0.0034	0.0035	0.0048	0.0033
			0.5	0.0226	0.0038	0.0040	0.0027	0.0040	0.0012
		0.6	0.1	0.0094	0.0025	0.0016	0.0024	0.0020	0.0019
			0.3	0.0154	0.0017	0.0013	0.0009	0.0012	0.0013
			0.5	0.0146	0.0017	0.0030	0.0020	0.0014	0.0025
	50	0.1	0.1	0.0561	0.0021	0.0021	0.0016	0.0016	0.0020
			0.3	0.0236	0.0037	0.0025	0.0029	0.0037	0.0023
			0.5	0.0379	0.0027	0.0029	0.0019	0.0019	0.0032
		0.3	0.1	0.0755	0.0021	0.0019	0.0015	0.0014	0.0023
			0.3	0.0344	0.0017	0.0014	0.0018	0.0013	0.0016
			0.5	0.0360	0.0017	0.0022	0.0014	0.0017	0.0019
		0.6	0.1	0.0331	0.0014	0.0021	0.0010	0.0014	0.0014
			0.3	0.0198	0.0022	0.0024	0.0023	0.0025	0.0024
			0.5	0.0282	0.0017	0.0016	0.0016	0.0016	0.0018
	250	0.1	0.1	---	0.0011	0.0010	0.0010	0.0011	0.0010
			0.3	---	0.0006	0.0007	0.0007	0.0005	0.0007
			0.5	---	0.0007	0.0009	0.0008	0.0008	0.0009
		0.3	0.1	---	0.0010	0.0009	0.0010	0.0010	0.0009
			0.3	---	0.0007	0.0008	0.0008	0.0007	0.0008
			0.5	---	0.0006	0.0006	0.0004	0.0006	0.0006
		0.6	0.1	---	0.0005	0.0005	0.0005	0.0005	0.0005
			0.3	---	0.0003	0.0003	0.0003	0.0003	0.0003
			0.5	---	0.0011	0.0010	0.0010	0.0010	0.0010

Table 12
Estimated Bias in Regression Weights with One Observation Removed in which an Obersvation with Extreme Regressor Values was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
2	10	0.1	0.1	0.0430	0.0635	0.0288	0.0860	0.0263	0.1773
			0.3	0.0321	0.0562	0.0264	0.0743	0.0311	0.0973
			0.5	0.0333	0.0418	0.0193	0.0502	0.0215	0.0718
		0.3	0.1	0.0430	0.0635	0.0288	0.0860	0.0263	0.1773
			0.3	0.0370	0.0691	0.0332	0.0825	0.0266	0.1710
			0.5	0.0305	0.0769	0.0434	0.0791	0.0349	0.1509
		0.6	0.1	0.0086	0.0571	0.0411	0.0667	0.0187	0.2172
			0.3	0.0111	0.0597	0.0493	0.0554	0.0161	0.2197
			0.5	0.0134	0.0696	0.0513	0.0560	0.0195	0.1965
	20	0.1	0.1	0.0776	0.0177	0.0052	0.0344	0.0100	0.0414
			0.3	0.0811	0.0315	0.0109	0.0517	0.0162	0.0668
			0.5	0.0679	0.0305	0.0128	0.0405	0.0167	0.0485
		0.3	0.1	0.0530	0.0341	0.0186	0.0681	0.0157	0.1275
			0.3	0.0424	0.0372	0.0221	0.0603	0.0152	0.1199
			0.5	0.0496	0.0412	0.0248	0.0584	0.0175	0.1062
		0.6	0.1	0.0142	0.0172	0.0223	0.0330	0.0032	0.1544
			0.3	0.0177	0.0201	0.0279	0.0318	0.0025	0.1525
			0.5	0.0179	0.0284	0.0325	0.0330	0.0072	0.1332
	100	0.1	0.1	0.0318	0.0081	0.0036	0.0134	0.0064	0.0143
			0.3	0.0453	0.0096	0.0037	0.0174	0.0062	0.0192
			0.5	0.0818	0.0088	0.0038	0.0120	0.0051	0.0132
		0.3	0.1	0.0335	0.0060	0.0051	0.0209	0.0029	0.0382
			0.3	0.0270	0.0087	0.0091	0.0191	0.0043	0.0353
			0.5	0.0351	0.0102	0.0086	0.0178	0.0040	0.0298
		0.6	0.1	0.0104	0.0019	0.0082	0.0065	0.0008	0.0475
			0.3	0.0122	0.0024	0.0102	0.0057	0.0010	0.0445
			0.5	0.0162	0.0034	0.0105	0.0059	0.0006	0.0372

Table 12 (Continued)

Estimated Bias in Regression Weights with One Observation Removed in which an Obersvation with Extreme Regressor Values was Flagged.

k	N	ρ^2	$\bar{r}_{12}$	Cook's D	DFBETAs	Leverage	R_Student	DFFITS	All Obs
5	25	0.1	0.1	0.0591	0.0293	0.0036	0.0449	0.0121	0.0560
			0.3	0.0519	0.0301	0.0091	0.0332	0.0127	0.0422
			0.5	0.0643	0.0294	0.0137	0.0357	0.0177	0.0412
		0.3	0.1	0.0830	0.0142	0.0031	0.0677	0.0064	0.0911
			0.3	0.0471	0.0331	0.0115	0.0357	0.0137	0.0512
			0.5	0.0527	0.0422	0.0222	0.0426	0.0214	0.0642
		0.6	0.1	0.0293	0.0231	0.0034	0.0379	0.0059	0.0677
			0.3	0.0131	0.0268	0.0157	0.0205	0.0026	0.0832
			0.5	0.0164	0.0248	0.0240	0.0231	0.0074	0.0636
	50	0.1	0.1	0.0584	0.0175	0.0026	0.0334	0.0088	0.0403
			0.3	0.0573	0.0183	0.0043	0.0222	0.0084	0.0264
			0.5	0.0929	0.0197	0.0080	0.0215	0.0114	0.0244
		0.3	0.1	0.0787	0.0085	0.0019	0.0628	0.0055	0.0757
			0.3	0.0413	0.0192	0.0062	0.0232	0.0074	0.0326
			0.5	0.0542	0.0222	0.0133	0.0270	0.0109	0.0389
		0.6	0.1	0.0229	0.0094	0.0022	0.0240	0.0029	0.0482
			0.3	0.0104	0.0083	0.0085	0.0083	0.0013	0.0539
			0.5	0.0150	0.0118	0.0148	0.0090	0.0023	0.0378
	250	0.1	0.1	---	0.0044	0.0006	0.0104	0.0024	0.0119
			0.3	---	0.0048	0.0015	0.0058	0.0032	0.0067
			0.5	---	0.0049	0.0025	0.0057	0.0032	0.0061
		0.3	0.1	0.0375	0.0021	0.0009	0.0263	0.0024	0.0312
			0.3	---	0.0046	0.0018	0.0062	0.0020	0.0083
			0.5	---	0.1886	0.1878	0.1889	0.1874	0.1899
		0.6	0.1	0.0156	0.0014	0.0005	0.0065	0.0006	0.0145
			0.3	0.0105	0.0012	0.0027	0.0012	0.0005	0.0140
			0.5	0.0248	0.0025	0.0042	0.0020	0.0007	0.0089

Figure 1

Bias in R-Square with One Observation Removed

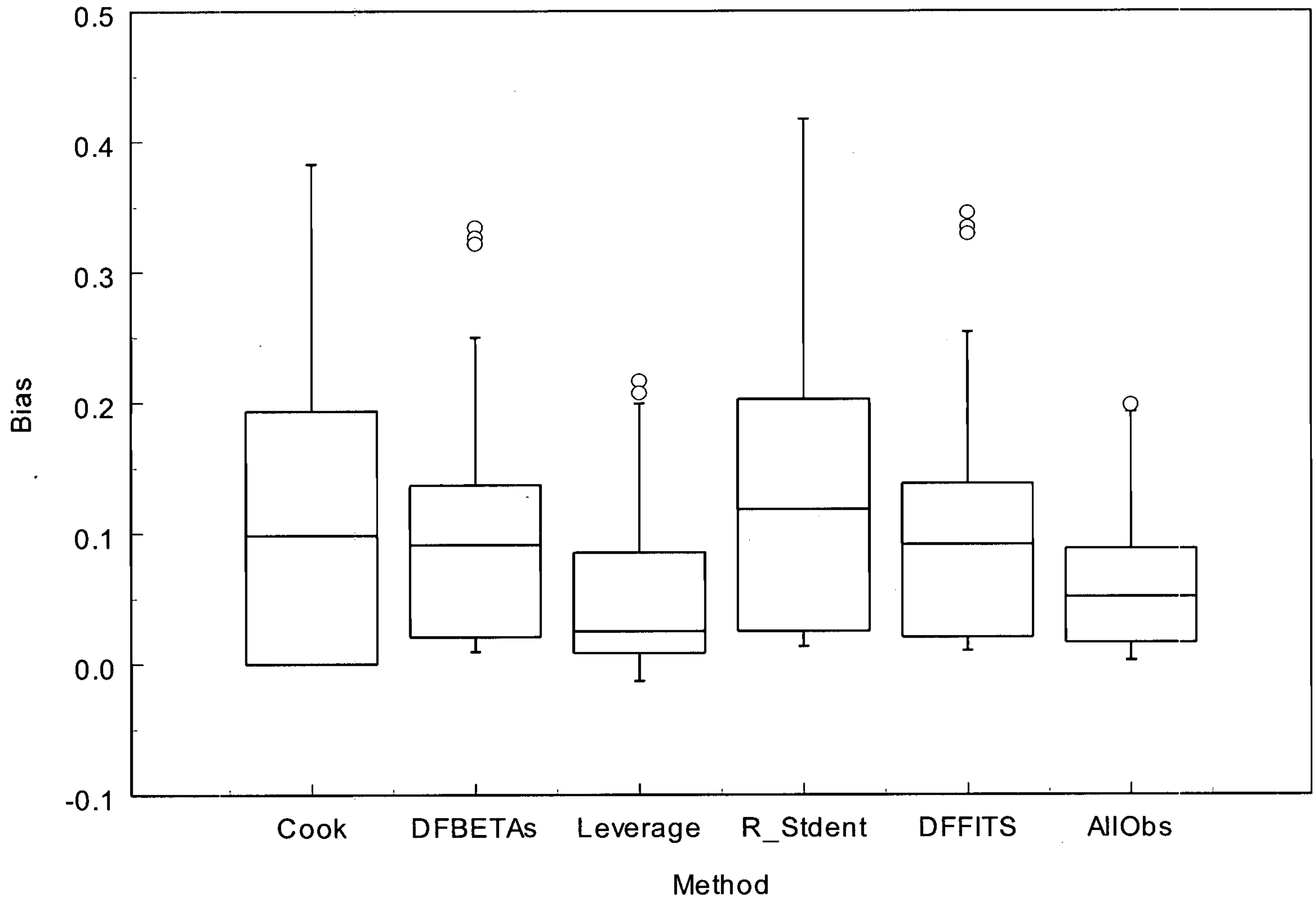


Figure 2

Bias in R-Square with Two Observations Removed

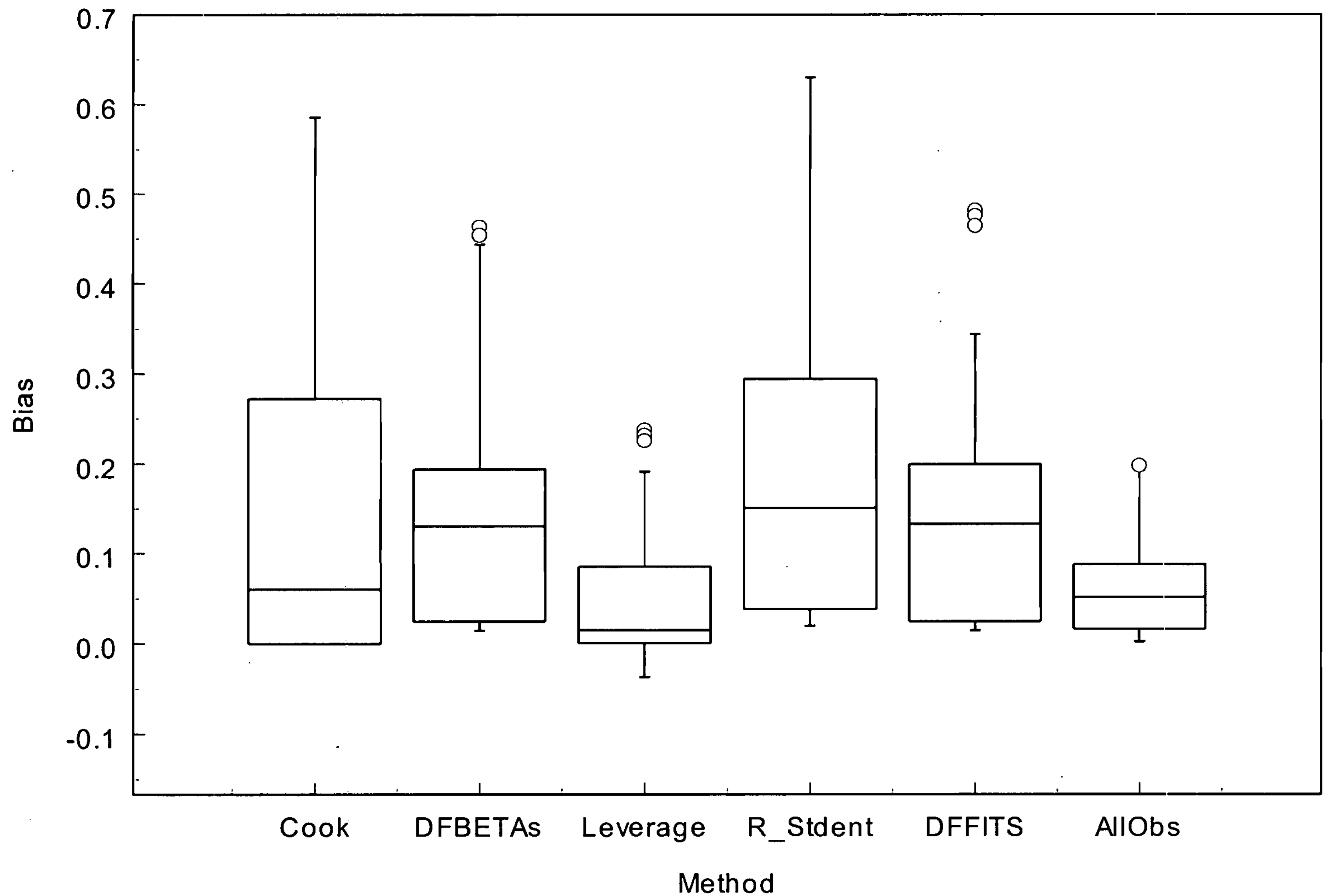


Figure 3
Estimated Bias in R^2 with One Observation Removed
k = 2, Population R^2 = .10, mean r_{12} = .30

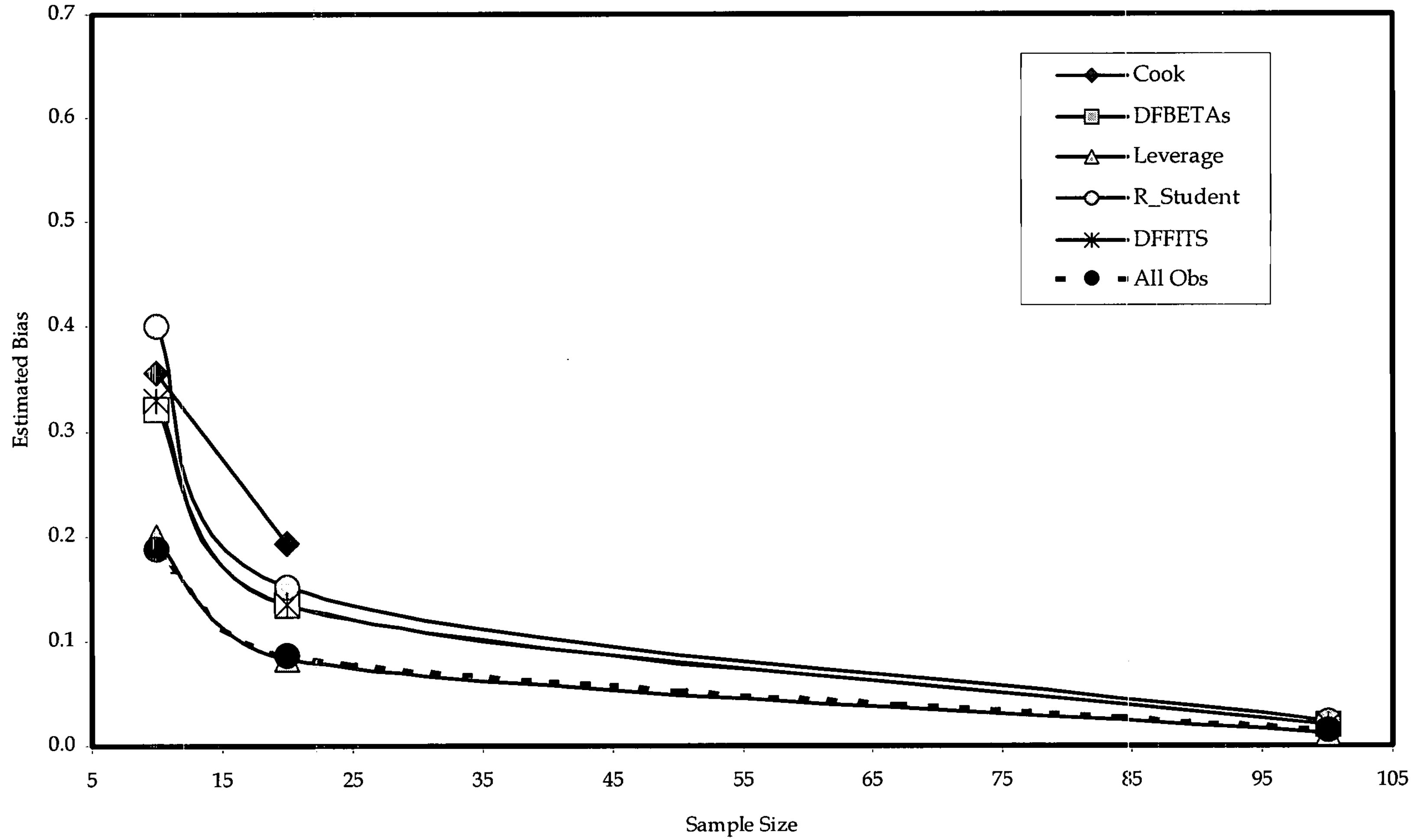


Figure 4
Estimated Bias in R^2 with Two Observations Removed
k = 2, Population R^2 = .10, mean r_{12} = .30

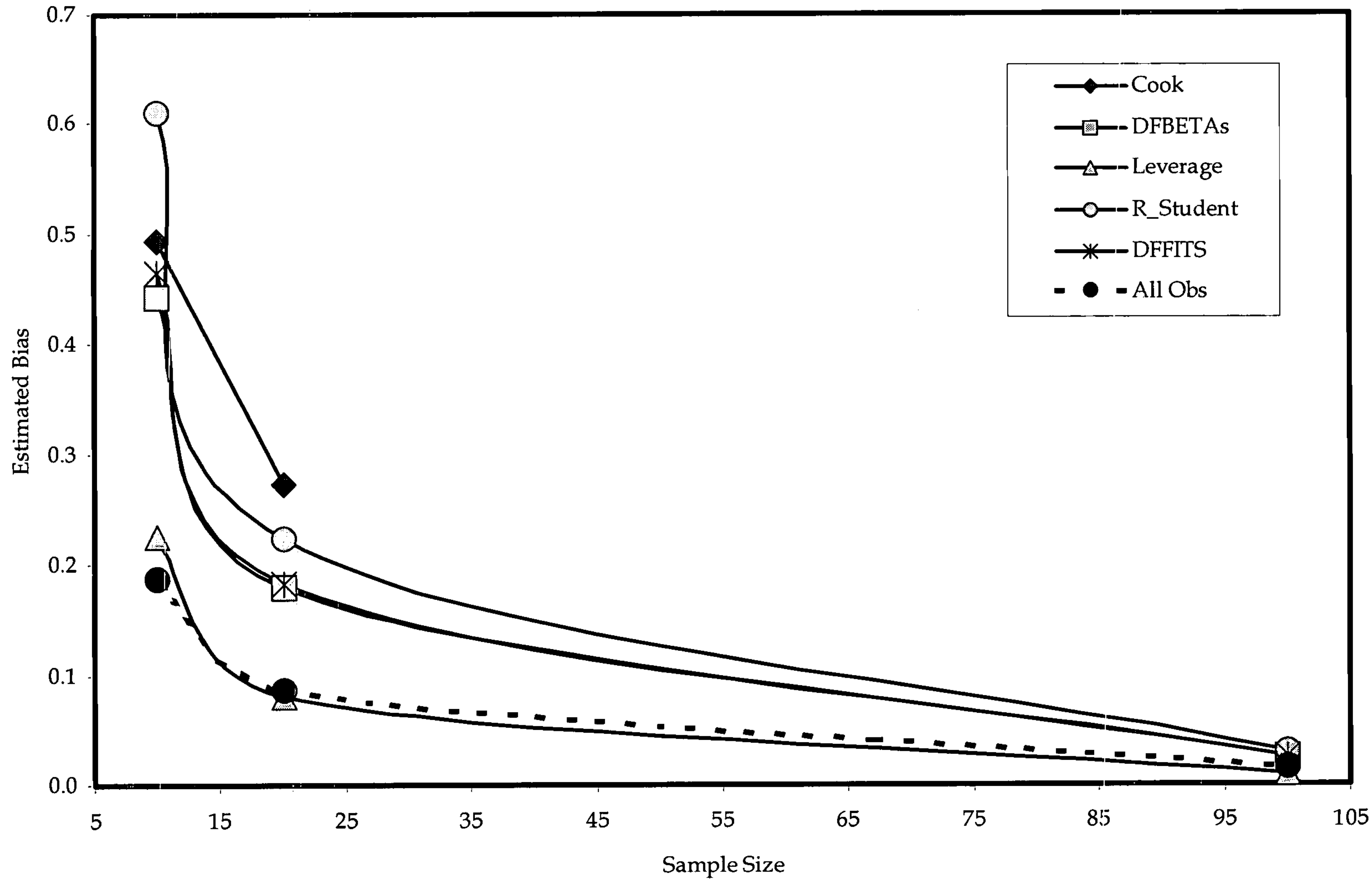


Figure 5
Estimated Bias in R^2 with One Observation Removed
Population R^2 = .60, Mean r_{12} = .30

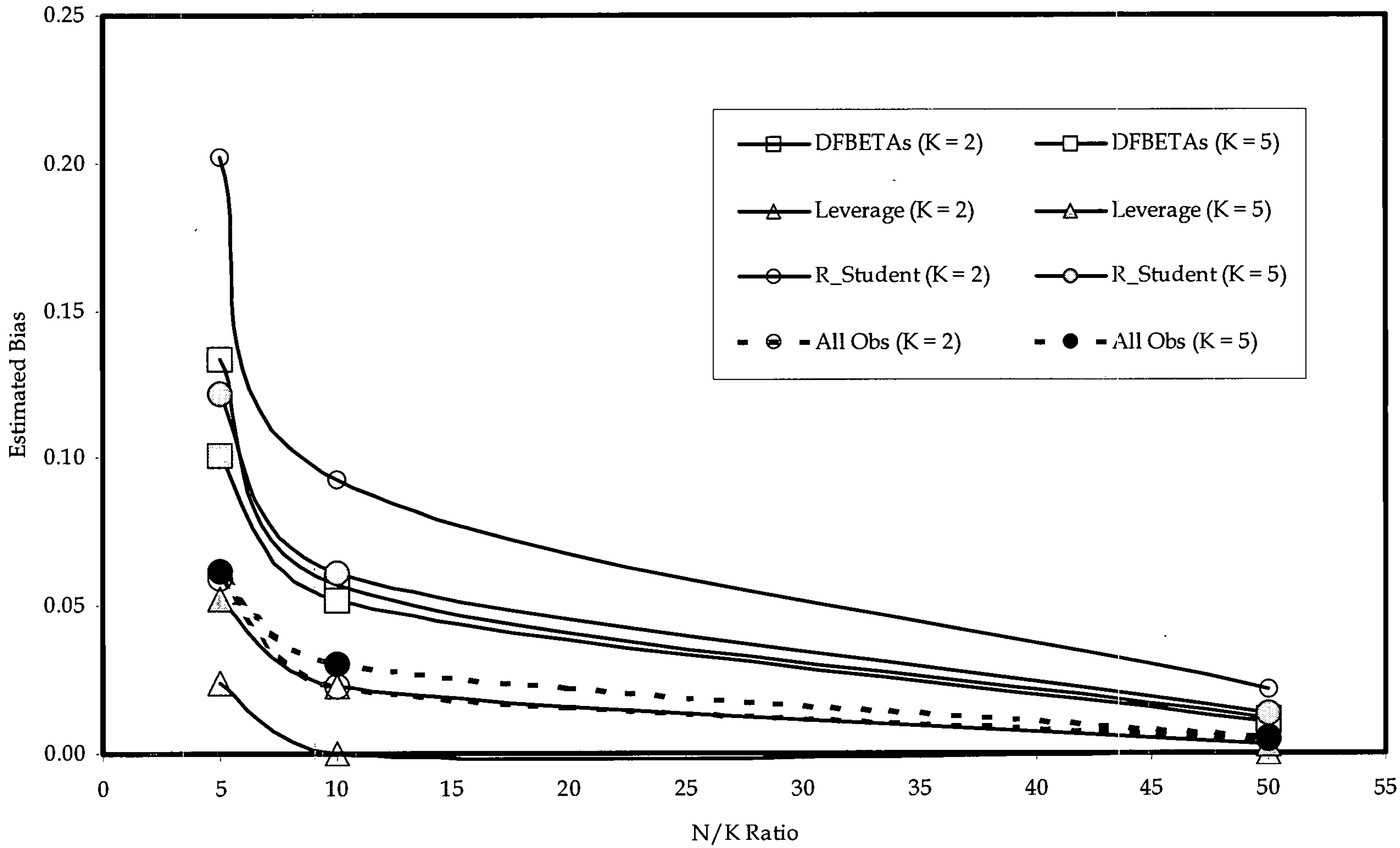


Figure 6
Estimated Bias in R^2 by Population R^2 with One Observation Removed
$k = 5$, $n = 25$, Mean $r_{12} = .5$

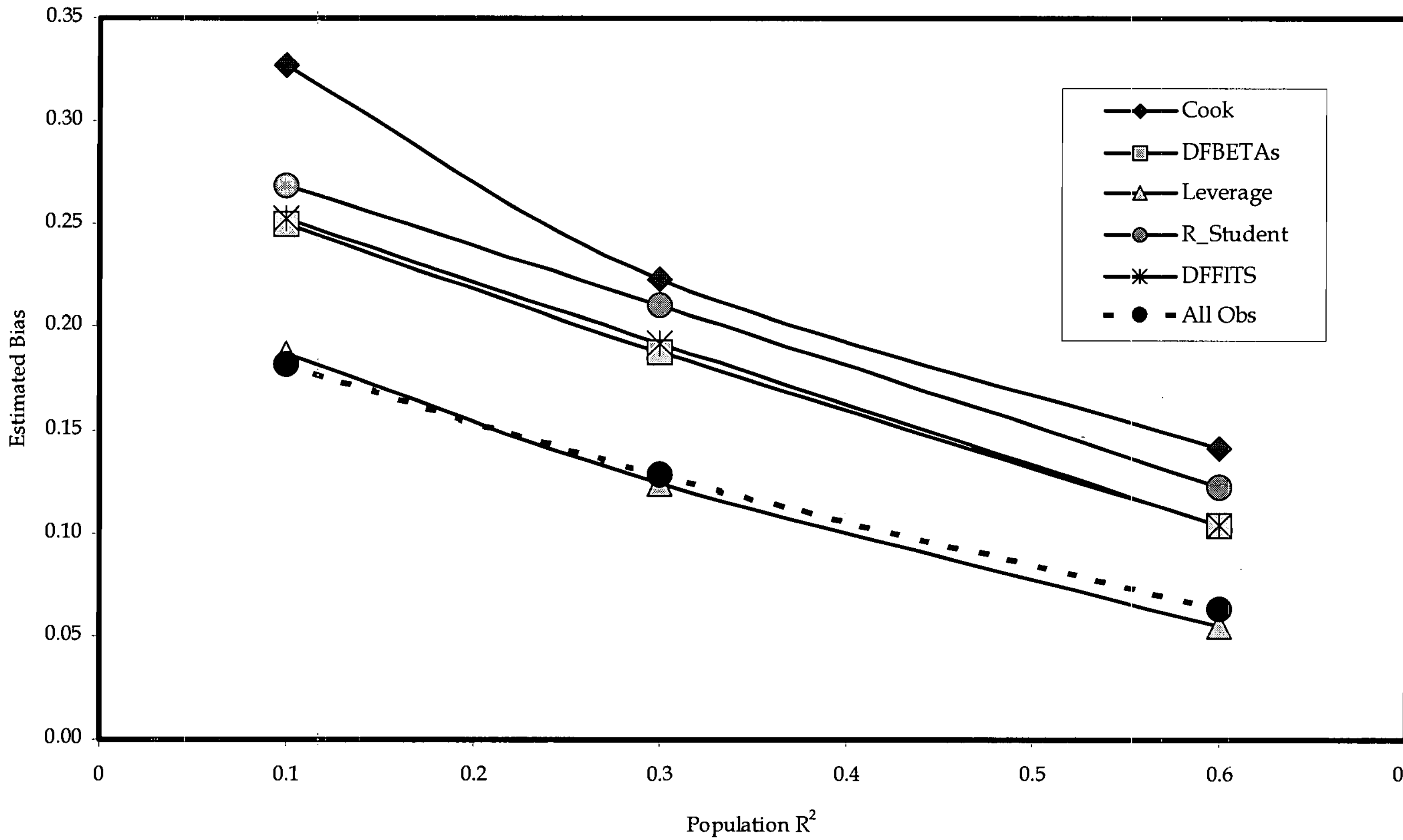


Figure 7

Bias in Regression Weights with One Observation Removed

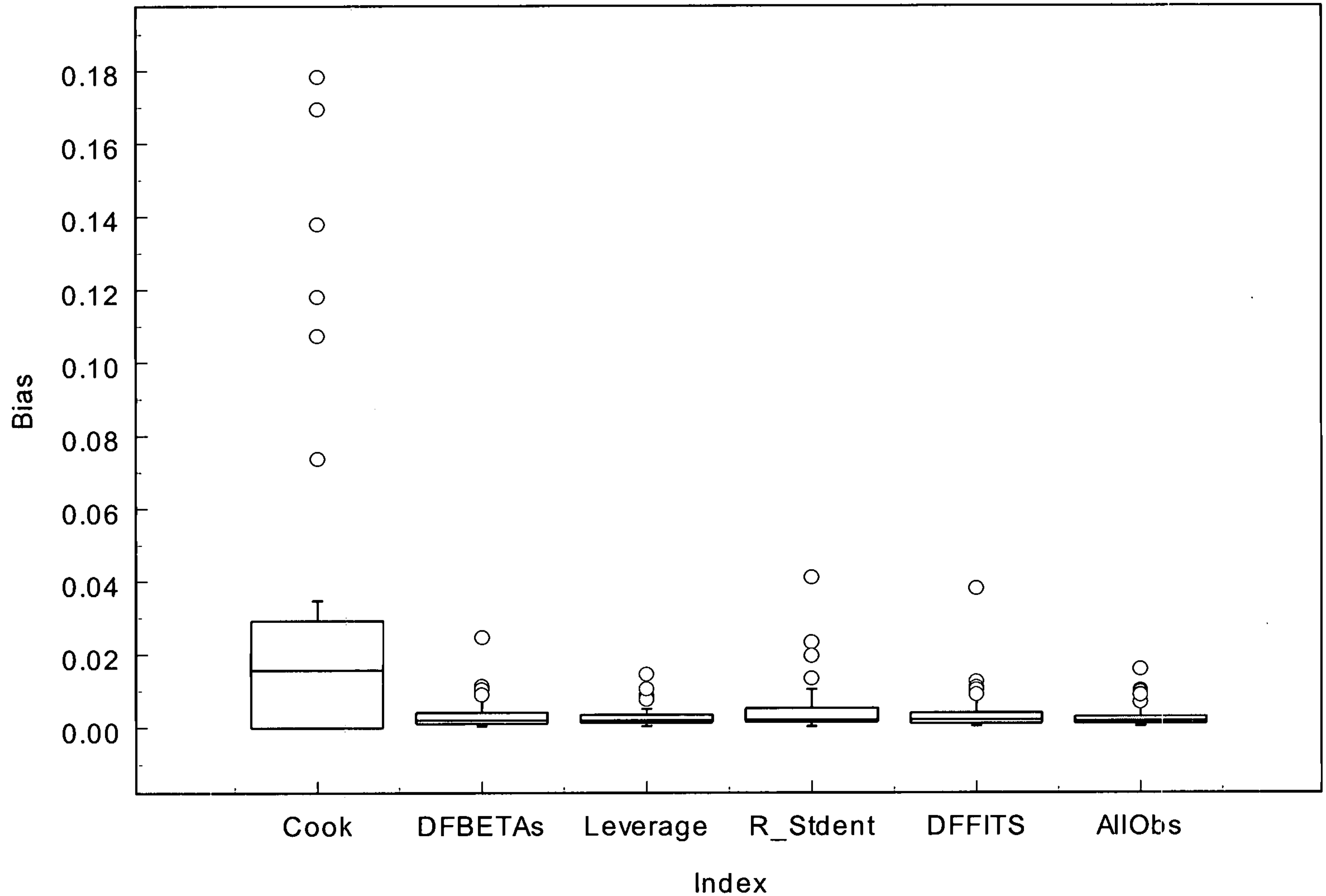


Figure 8

Bias in Regression Weights with Two Observations Removed

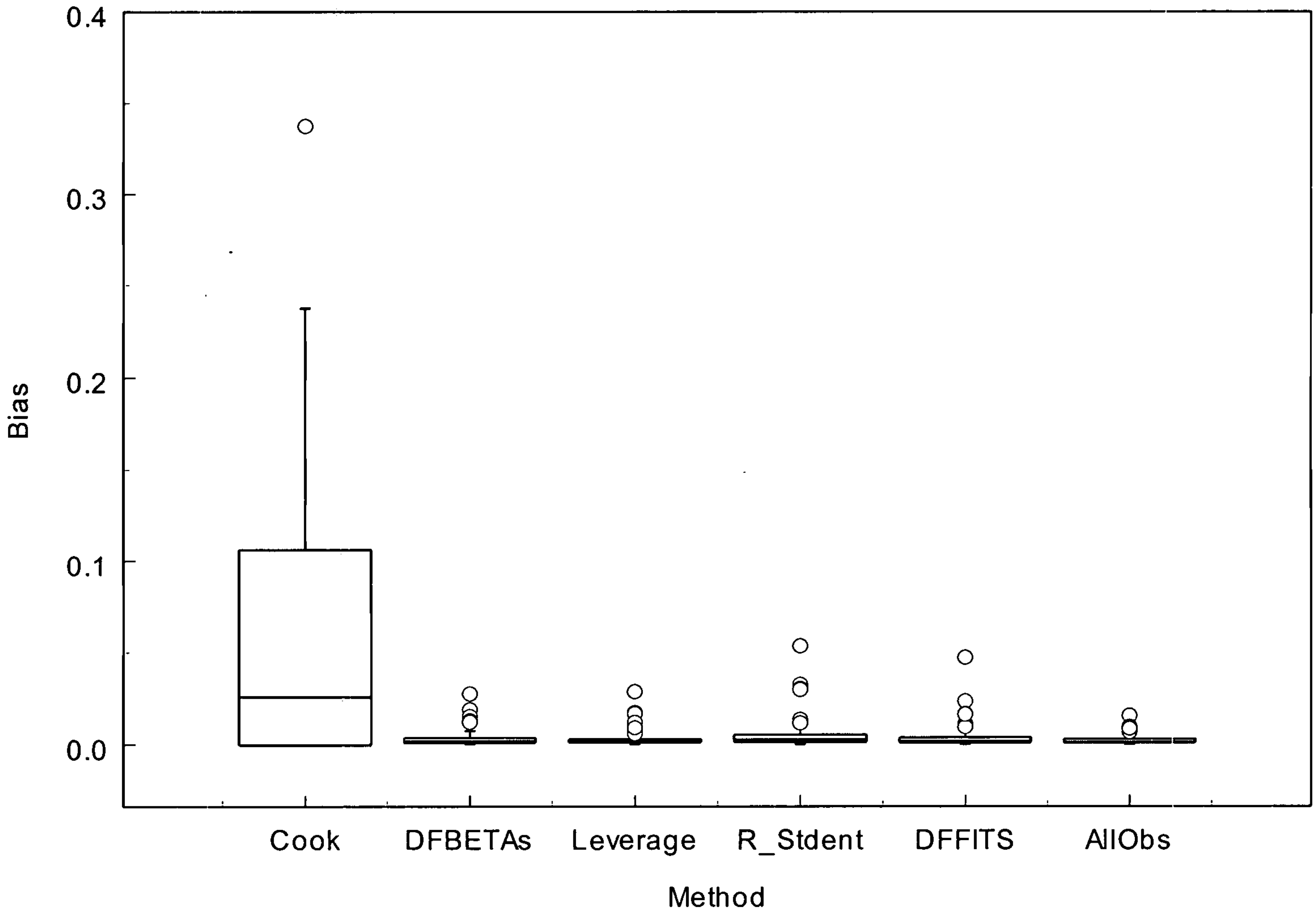


Figure 9
Estimated Bias in Regression Weights with One Observation Removed
k = 2, Population R^2 = .10, r_{12} = .30

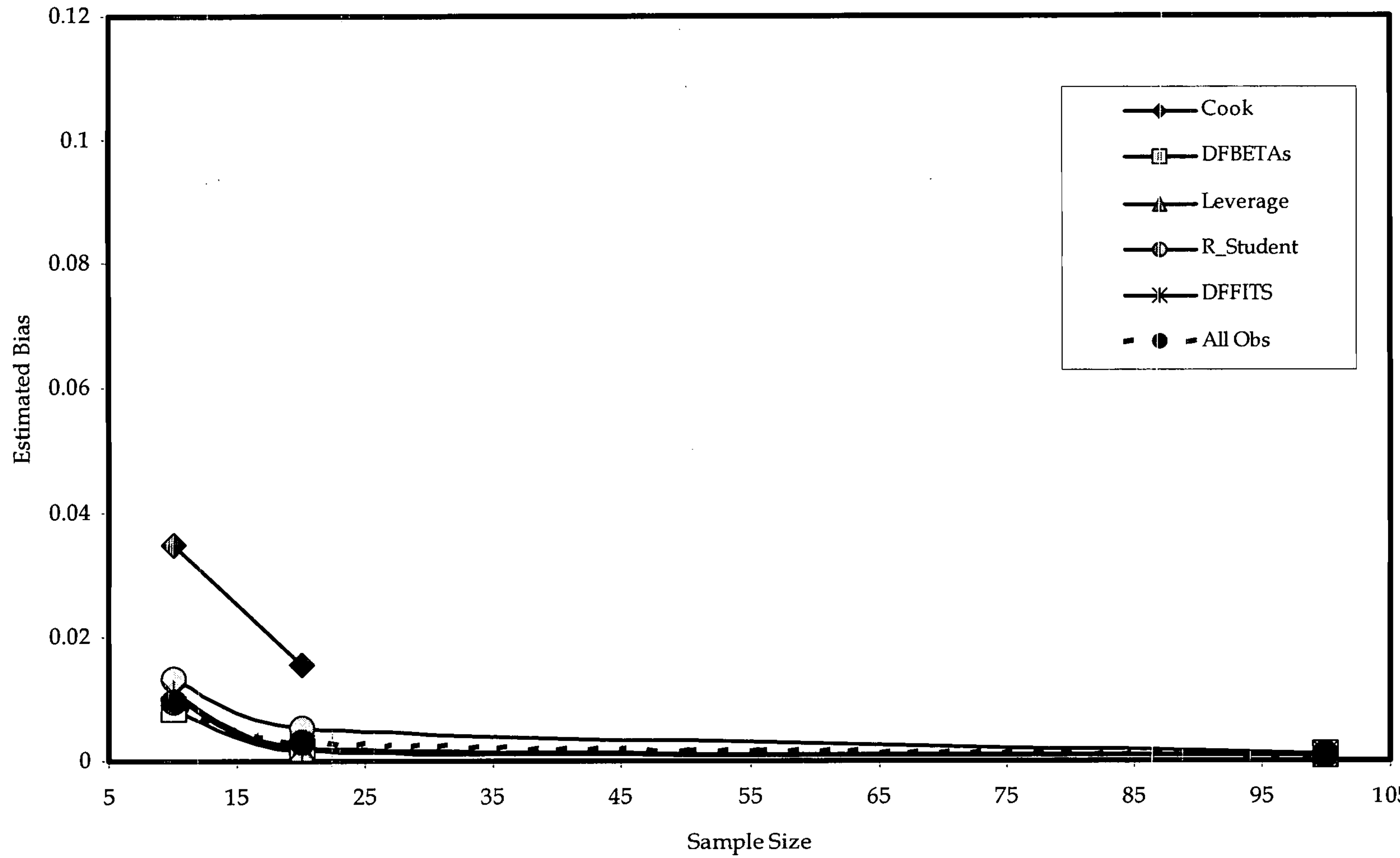


Figure 10
Estimated Bias in Regression Weights with Two Observations Removed
k = 2, Population R^2 = .10, r_{12} = .30

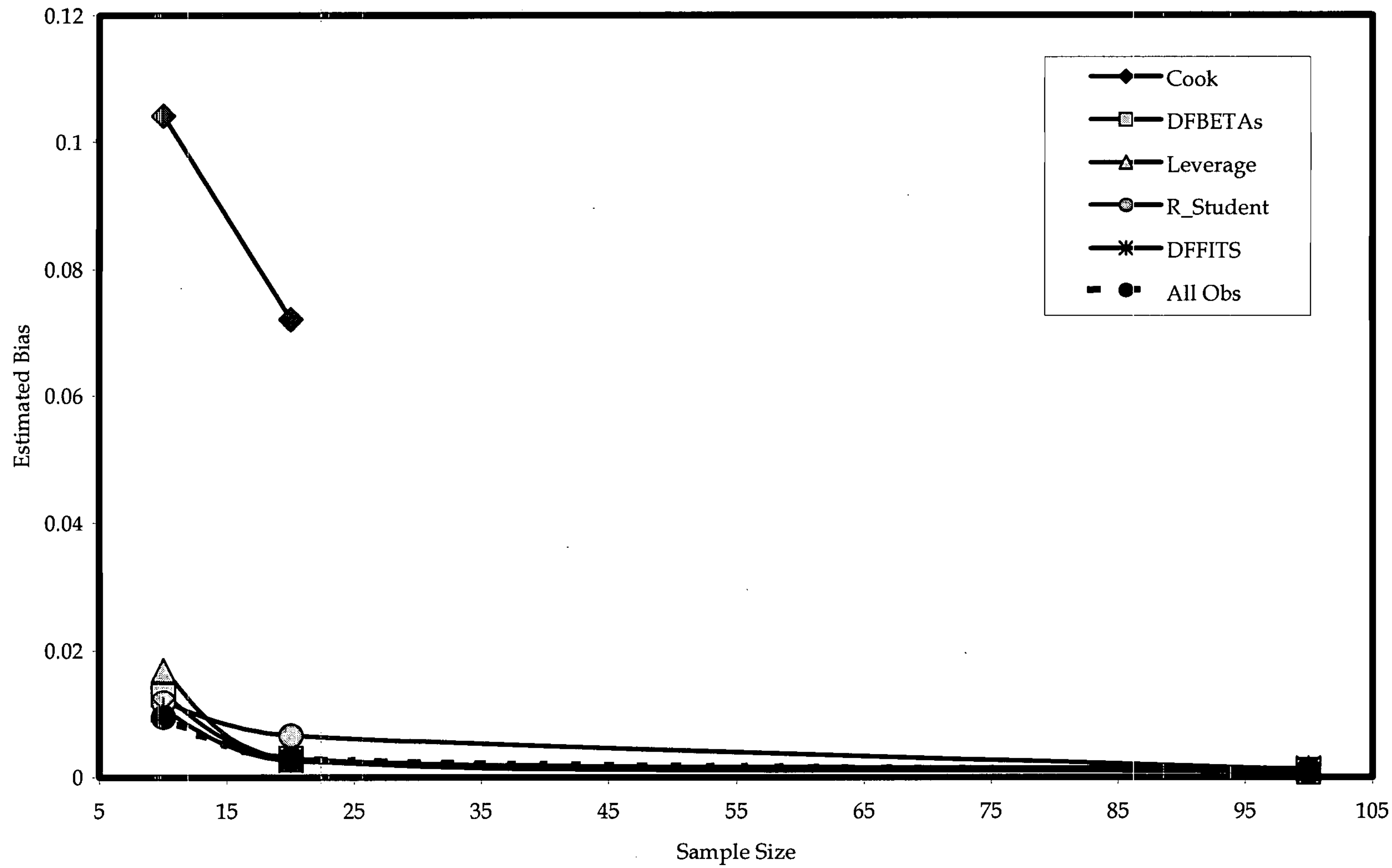


Figure 11

Bias in R-Square with One Observation Removed (Large Residual)

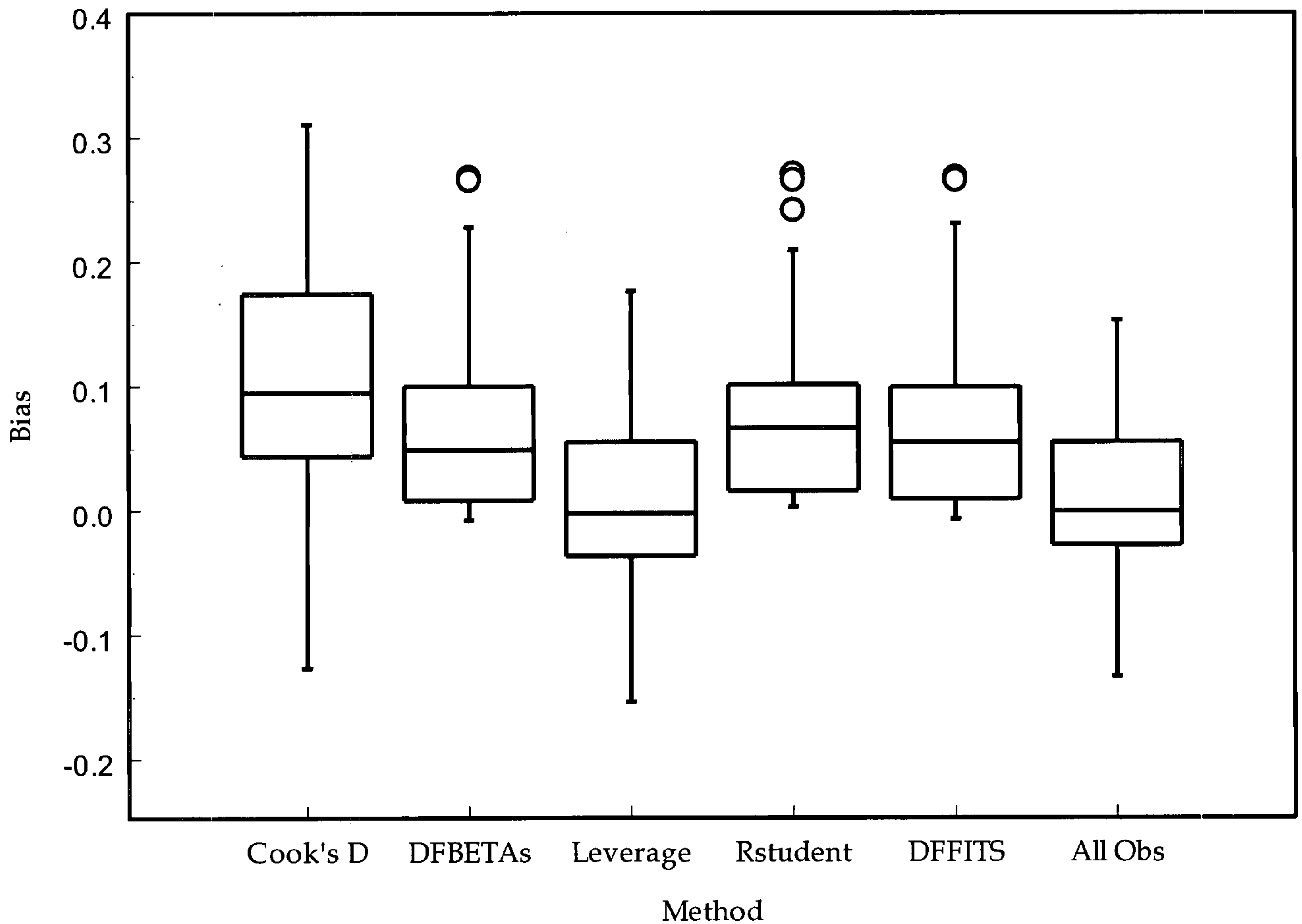


Figure 12

Bias in R-Square with One Observation Removed (Regressor Outlier)

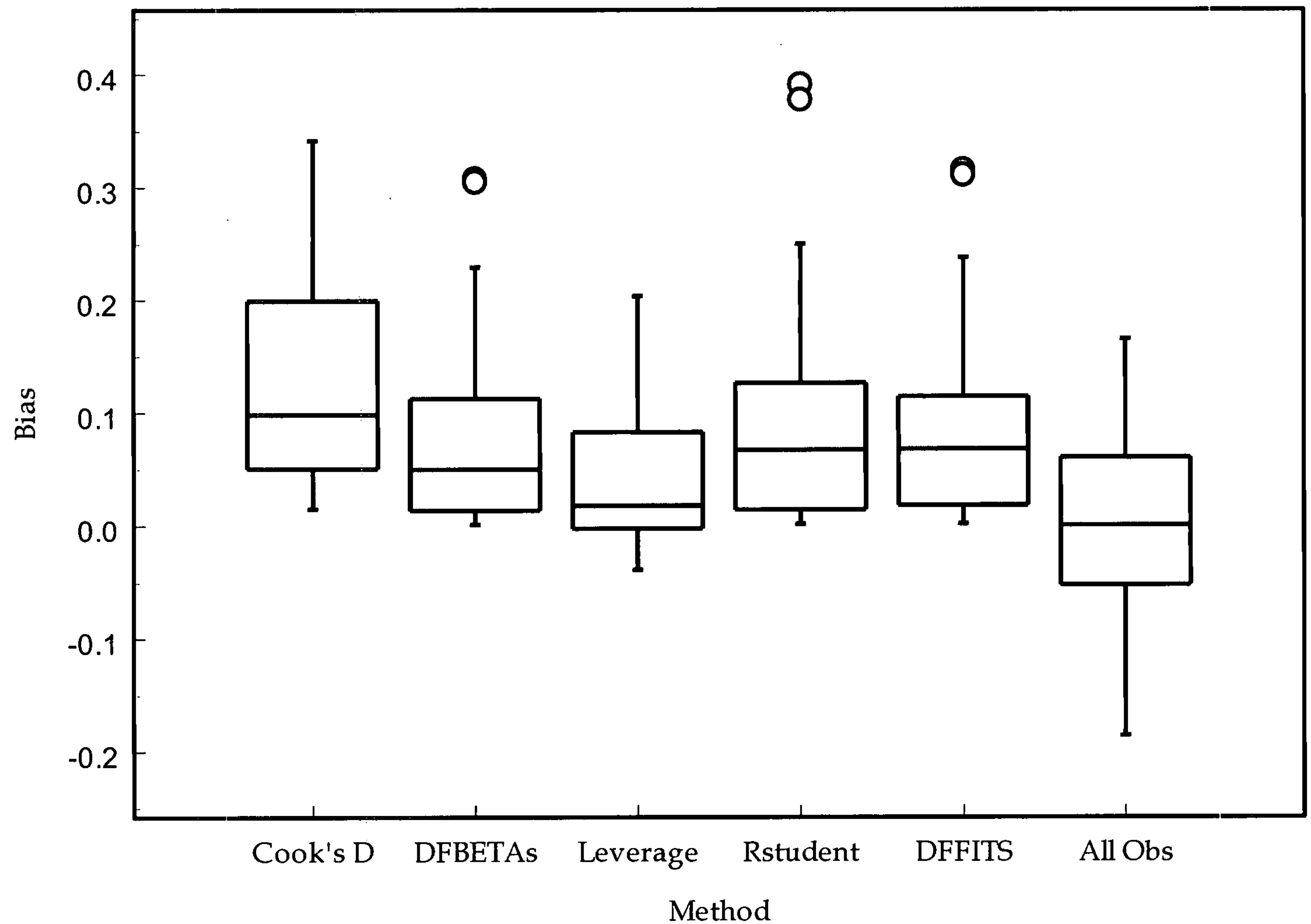


Figure 13

Bias in Regression Weights with One Observation Removed (Large Residual)

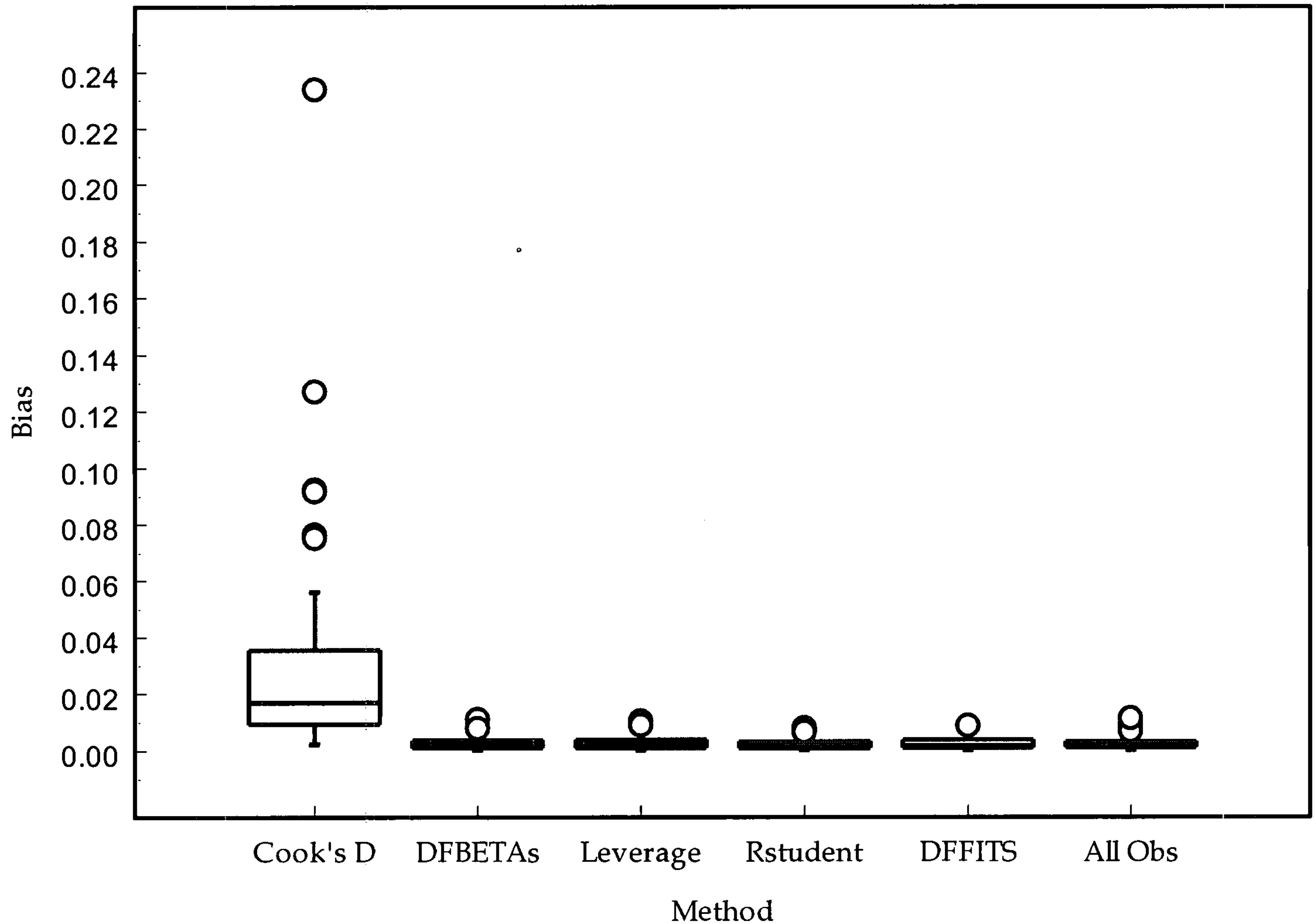
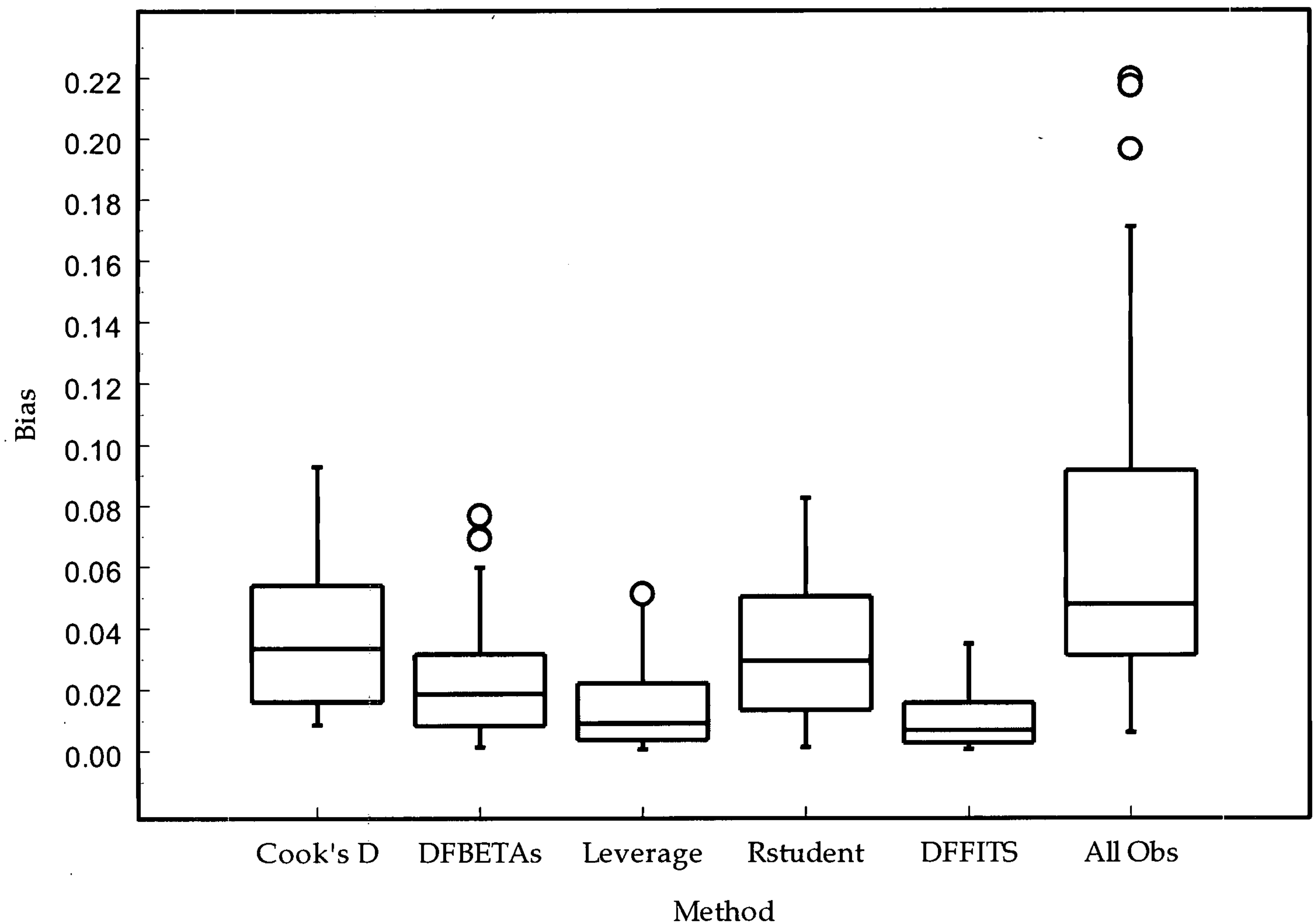


Figure 14

Bias in Regression Weights with One Observation Removed (Regressor Outlier)



TM034259

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 Influence of Influence Diagnostics:An Empirical Investigation of the Effects of Removing Extreme Data Points

Author(s): Jeffrey D. Kromrey & Jeanine Romano

Corporate Source: University of South Florida

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 [X]	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: [signature]	Printed Name/Position/Title: Jeffrey D. Kromrey / Professor	
Organization/Address: Univ. of South Florida	Telephone: 813 974-5739	FAX: 813 974-4495
	E-Mail Address: kromrey@tempest.coedu.usf.edu	Date: 13 May 2002

(over)

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:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

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
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)