# Should High-Low Go: An Analysis Using the Bootstrap

**Jon Woodroof**
**University of Tennessee**
**Terry Ward**
**Middle Tennessee State University**
**Bill Burg**
**University of Alabama - Birmingham**

**Abstract**

In order for managers to be able to estimate break-even numbers for budgeting purposes, historical total costs must be able to be separated into their fixed and variable cost components. There are generally two methods for teaching this: the high-low method, and the method of least squares regression. The high-low method is considered theoretically inferior to the method of least squares, yet it continues to be taught in accounting courses. The argument for its continuation has been that it is "quick and easy". However, with the proliferation of electronic spreadsheets, this advantage can also be attributed to the method of least squares. The high-low method's continued coverage in accounting textbooks would seem to indicate that educators feel that the results generated by each method are not significantly different.

This paper compares these methods by using a bootstrapping technique. Bootstrapping facilitates the simulated generation of entire distributions from a sample and allows statistical comparisons to be made between the distributions. The results of this study indicate that the high-low method, while easy to use, may be giving results that are significantly different from results obtained from regression. Because students now have the ability to do regression easily and inexpensively using a spreadsheet, and because of the theoretical shortcomings of the high-low method, it may be that educators should discontinue using and teaching the high-low method altogether.

## Introduction

In order for managers to be able to estimate break-even numbers for budgeting purposes, historical total costs (which are mixed in nature) must be able to be separated into their fixed and variable cost components. There are generally two methods for teaching this: the high-low method, and the method of least squares regression. Using the high-low method, one determines the slope of the variable cost line (and ultimately the amount of fixed costs) by identifying high and low total cost data points and high and low activity data points and dividing the change in total cost at these high and low levels by the change in activity. Alternatively using the method of least squares regression, the

slope line is mathematically fitted to the data. The point where this slope line crosses the Y-axis indicates the amount of the total costs that are fixed.

Advocates of the high-low method have historically argued that it is "quick and easy" and does not require sophisticated analysis tools that may not be readily available to accounting practitioners and students. Advocates of the method of least squares regression insist that relying on only two data points to estimate the fixed and variable components of total costs (and discarding the rest of the data) is too simplistic an approach and may lead managers to make bad decisions.

## Problem

The high-low method has been known to have serious theoretical and practical flaws for several decades (Johnson and Harrell, 1999). Conversely, the method of least squares regression has long been recognized as the superior technique (Nurnberg, 1977). Regression is a much more sophisticated mathematical technique, and accounting researchers more than 30 years ago showed that regression can be a valuable tool in accounting (Comiskey, 1966; Benston, 1966). Today, regression can be performed easily using a common electronic spreadsheet.

Because electronic spreadsheets have been readily available to both accounting practitioners and students for several years now, utilization of the least squares regression method should be on the increase, and teaching and promoting the simplistic (and perhaps flawed) high-low method should be on the decrease. However, there is no evidence that this is the trend. On the contrary, both methods continue to be taught in accounting principles, cost, and managerial accounting textbooks. In fact, texts published within the past few years continue to give a significant amount of coverage to the high-low method and include several high-low problems and exercises (Johnson and Harrell, 1999).

There can be only one good reason for this (reluctance to change is not considered here as a good reason). Evidently, there is a perception that the two methods do not produce results that materially differ from each other. The source of this perception is not entirely clear -- accounting research literature describing empirical studies on this topic is unusually sparse. One reason for this scarcity is that it is very difficult to get companies to release real cost data. A second reason is that the very exercise of comparing methods for separating total costs into their fixed and variable components is innately problematic. How do you statistically determine which method is "better"? And how can this determination for a given set of data in a particular industry be made easily and inexpensively? One approach for doing this is bootstrapping.

## Bootstrapping

Bootstrapping is a powerful technique for making statistical inferences about a population characteristic (Efron, 1982). However, it differs from the traditional approach to statistical inferencing in that,

> *Bootstrapping relies on an analogy between the sample and the population from which the sample was drawn. The central idea is that it may sometimes be better to draw conclusions about the characteristics of a population strictly from the sample at hand, rather than by making perhaps unrealistic assumptions about that population* (Mooney and Duval, 1993, p.1).

Bootstrapping, through a resampling simulation, uses large numbers of repetitive computations and empirically generates an estimate of a statistic's entire distribution. It does this by examining the variation of the statistic of interest within the sample. Bootstrapping is a nonparametric technique where normality is not assumed. Thus, it can be employed to estimate the underlying sampling distributions of statistics that cannot be assumed to be normal (Mooney and Duval, 1993). What is more, bootstrapping enables confidence intervals to be derived, statistical comparisons to be made between distributions, and the accuracy of a particular estimator of an unknown parameter to be measured (Efron and Tibshirani, 1986).

The methodology is as follows: The original data is randomly sampled with replacement (so that each resample has the same number of elements as did the original data). Using each new resample, the statistic of interest is recalculated for each of the methods being compared. Next, the statistic calculated using one method is subtracted from the static calculated using the comparison method, and the results are stored in a vector. This is done at least 1000 times (Efron, 1986). The resulting vector of differences is then sorted in ascending order, confidence interval endpoints of a particular alpha level of confidence are identified, and inferences about the methods being compared are made. This methodology can be implemented easily and inexpensively by designing a simple spreadsheet template (for a more thorough discussions of bootstrapping in a spreadsheet, see Woodroof, 1997).

## The Nurnberg Study

This paper references and builds on one of the few efforts (outside of the coverage these topics get in cost accounting textbooks) that has investigated these methods -- the Nurnberg study (Nurnberg, 1977). In the first part of the Nurnberg study, an analysis was made comparing three implementation of the high-low method: 1) actual high-low pairs based upon highest and lowest costs levels; 2) actual high-low pairs based upon highest and lowest activity levels; and 3) hypothetical high-low pairs based on absolute highs and lows regardless of actual pairing. Nurnberg used a data set borrowed from Horngren (1972) consisting of *Total Trucking Labor Cost* (the total costs to be separated into their fixed and variable components) and *Direct Labor Cost* (the activity that is assumed to be correlated with the trucking labor), and he argued that, unless the highest cost is incurred at the highest activity level and the lowest cost is incurred at the lowest activity level, the three implementations of the high-low method can result in different variable cost slopes and fixed cost amounts for the same data set (Nurnberg, 1977). Table 1 shows these results.

**Table 1: Nurnberg's Comparison of Three High-Low Methods**

| | Actual Pairing Based on Total Cost | Actual Pairing Based on Activity (Direct Labor) | Hypothetical pairing |
|---|---|---|---|
| **High Total Cost** | 11,200 | 11,000 | 11,200 |
| **Low Total Cost** | 6,400 | 6,400 | 6,400 |
| **High Activity** | 378,000 | 384,000 | 384,000 |
| **Low Activity** | 180,000 | 180,000 | 180,000 |
| **Variable cost per unit** | 0.02424 | 0.02255 | 0.02353 |
| **Fixed cost component** | 2,036 | 2,341 | 2,165 |

However, Nurnberg did not test whether these differences for this data set were statistically significant.

## The Research Question

For the current study, the high-low implementation based on total cost was dropped from the analysis due to Nurnberg's (as well as others) *apriori* conclusions concerning the nature of actual high-low pairs based upon highest and lowest *cost* levels. He states:

> *Implicit throughout most discussions of cost behaviour in the managerial accounting literature is the assumption of fewer errors in the measurement of activity levels than cost levels. This in turn is due to the fact that activity levels are often measured in physical terms and, as such, are devoid of the ambiguities inherent in accounting accruals, deferrals, and allocations, to which measurements of cost levels are subject. Accordingly, extreme activity levels are less likely to reflect the abnormal than extreme cost levels* (Nurnberg, 1977, p. 433).

Therefore, this paper uses the same Horngren data set (Appendix A) and uses bootstrapping to empirically compare the following three methods of separating total costs into their fixed and variable components: 1) least squares regression; 2) actual high-low pairs based upon highest and lowest activity levels; and 3) hypothetical high-low pairs based on absolute highs and lows regardless of actual pairing. The calculations for the three methods are shown in Table 2.

**Table 2: Comparison of Methods to Separate Total Costs**

| | Least Squares Regression | Actual Pairing Based on Activity | Hypothetical Pairing |
|---|---|---|---|
| **High Total Cost** | | 11,000 | 11,200 |
| **Low Total Cost** | | 6,400 | 6,400 |
| **High Activity** | | 384,000 | 384,000 |
| **Low Activity** | | 180,000 | 180,000 |
| **Variable cost per unit** | 0.02155 | 0.02255 | 0.02353 |
| **Fixed cost component** | 2,918 | 2,341 | 2,165 |

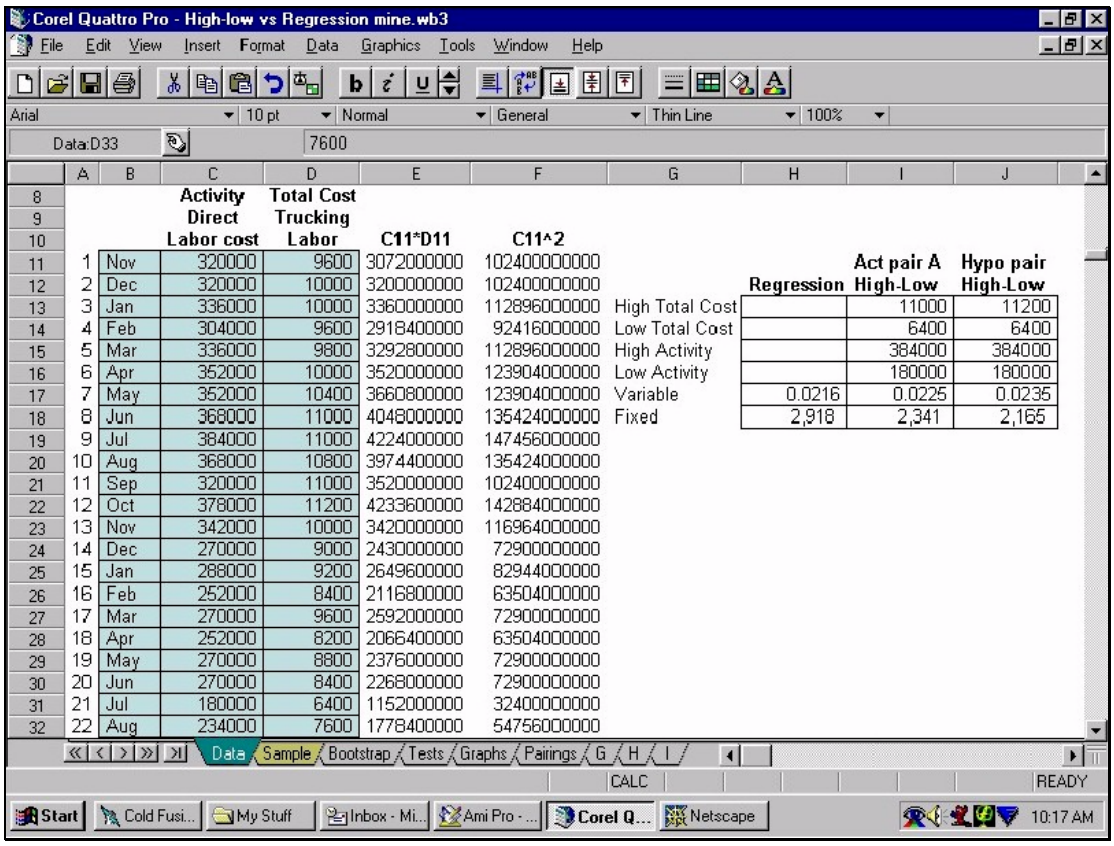
Thus, the research question to be investigated is, "Is there a significant difference among these three methods?" Stated in the null,

> *There is no difference among least squares regression, actual pairing high-low, and hypothetical pairing high-low in their abilities to separate total costs into their fixed and variable components.*

Bootstrapping is used to address this question. The following section provides a discussion of a spreadsheet template that performs this powerful statistical technique.

## The Spreadsheet Template

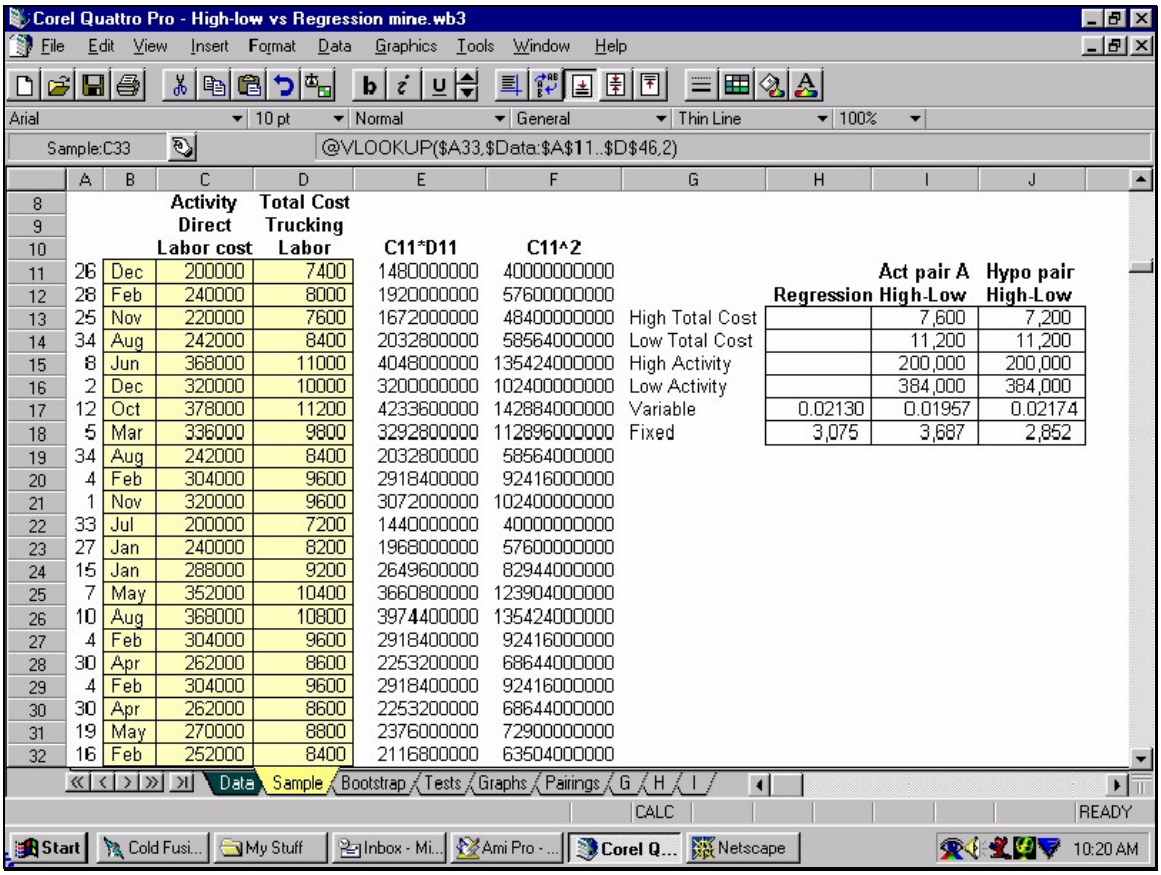
This template can be build using any electronic spreadsheet. For our example, we chose Corel's Quattro Pro spreadsheet. First, the data (Appendix A) is entered into the Data page of the spreadsheet - shown in Figure 1. In column A is an index (1 through 36) indicating the number of the particular data point. This will be used later to randomly select a data point. Of course, columns C and D are needed to calculate least squares regression (details of this calculation will not be presented here, since that is not the purpose of this paper). This page calculates and displays the variable and fixed components of total cost for the *original* data using the three methods.




| | | | Activity Direct Labor cost | Total Cost Trucking Labor | C11*D11 | C11^2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Nov | | 320000 | 9600 | 3072000000 | 102400000000 | | | | |
| 2 | Dec | | 320000 | 10000 | 3200000000 | 102400000000 | | Regression | Act pair A High-Low | Hypo pair High-Low |
| 3 | Jan | | 336000 | 10000 | 3360000000 | 112896000000 | High Total Cost | | 11000 | 11200 |
| 4 | Feb | | 304000 | 9600 | 2918400000 | 92416000000 | Low Total Cost | | 6400 | 6400 |
| 5 | Mar | | 336000 | 9800 | 3292800000 | 112896000000 | High Activity | | 384000 | 384000 |
| 6 | Apr | | 352000 | 10000 | 3520000000 | 123904000000 | Low Activity | | 180000 | 180000 |
| 7 | May | | 352000 | 10400 | 3660800000 | 123904000000 | Variable | 0.0216 | 0.0225 | 0.0235 |
| 8 | Jun | | 368000 | 11000 | 4048000000 | 135424000000 | Fixed | 2,918 | 2,341 | 2,165 |
| 9 | Jul | | 384000 | 11000 | 4224000000 | 147456000000 | | | | |
| 10 | Aug | | 368000 | 10800 | 3974400000 | 135424000000 | | | | |
| 11 | Sep | | 320000 | 11000 | 3520000000 | 102400000000 | | | | |
| 12 | Oct | | 378000 | 11200 | 4233600000 | 142884000000 | | | | |
| 13 | Nov | | 342000 | 10000 | 3420000000 | 116964000000 | | | | |
| 14 | Dec | | 270000 | 9000 | 2430000000 | 72900000000 | | | | |
| 15 | Jan | | 288000 | 9200 | 2649600000 | 82944000000 | | | | |
| 16 | Feb | | 252000 | 8400 | 2116800000 | 63504000000 | | | | |
| 17 | Mar | | 270000 | 9600 | 2592000000 | 72900000000 | | | | |
| 18 | Apr | | 252000 | 8200 | 2066400000 | 63504000000 | | | | |
| 19 | May | | 270000 | 8800 | 2376000000 | 72900000000 | | | | |
| 20 | Jun | | 270000 | 8400 | 2268000000 | 72900000000 | | | | |
| 21 | Jul | | 180000 | 6400 | 1152000000 | 32400000000 | | | | |
| 22 | Aug | | 234000 | 7600 | 1778400000 | 54756000000 | | | | |

**Figure 1: Data Page**

The next spreadsheet page is the Sample page, shown in Figure 2. Each time the spreadsheet is recalculated, formulas on this page randomly sample the data on the Data page, and then calculate and display the variable and fixed components of total cost for the *sampled* data using the three methods.




| | | | Activity Direct Labor cost | Total Cost Trucking Labor | C11*D11 | C11^2 | | Regression | Act pair A High-Low | Hypo pair High-Low |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 26 | Dec | 200000 | 7400 | 1480000000 | 40000000000 | | | | |
| 12 | 28 | Feb | 240000 | 8000 | 1920000000 | 57600000000 | | | | |
| 13 | 25 | Nov | 220000 | 7600 | 1672000000 | 48400000000 | High Total Cost | | 7,600 | 7,200 |
| 14 | 34 | Aug | 242000 | 8400 | 2032800000 | 58564000000 | Low Total Cost | | 11,200 | 11,200 |
| 15 | 8 | Jun | 368000 | 11000 | 4048000000 | 135424000000 | High Activity | | 200,000 | 200,000 |
| 16 | 2 | Dec | 320000 | 10000 | 3200000000 | 102400000000 | Low Activity | | 384,000 | 384,000 |
| 17 | 12 | Oct | 378000 | 11200 | 4233600000 | 142884000000 | Variable | 0.02130 | 0.01957 | 0.02174 |
| 18 | 5 | Mar | 336000 | 9800 | 3292800000 | 112896000000 | Fixed | 3,075 | 3,687 | 2,852 |
| 19 | 34 | Aug | 242000 | 8400 | 2032800000 | 58564000000 | | | | |
| 20 | 4 | Feb | 304000 | 9600 | 2918400000 | 92416000000 | | | | |
| 21 | 1 | Nov | 320000 | 9600 | 3072000000 | 102400000000 | | | | |
| 22 | 33 | Jul | 200000 | 7200 | 1440000000 | 40000000000 | | | | |
| 23 | 27 | Jan | 240000 | 8200 | 1968000000 | 57600000000 | | | | |
| 24 | 15 | Jan | 288000 | 9200 | 2649600000 | 82944000000 | | | | |
| 25 | 7 | May | 352000 | 10400 | 3660800000 | 123904000000 | | | | |
| 26 | 10 | Aug | 368000 | 10800 | 3974400000 | 135424000000 | | | | |
| 27 | 4 | Feb | 304000 | 9600 | 2918400000 | 92416000000 | | | | |
| 28 | 30 | Apr | 262000 | 8600 | 2253200000 | 68644000000 | | | | |
| 29 | 4 | Feb | 304000 | 9600 | 2918400000 | 92416000000 | | | | |
| 30 | 30 | Apr | 262000 | 8600 | 2253200000 | 68644000000 | | | | |
| 31 | 19 | May | 270000 | 8800 | 2376000000 | 72900000000 | | | | |
| 32 | 16 | Feb | 252000 | 8400 | 2116800000 | 63504000000 | | | | |

**Figure 2: Sample Page**

The formulas in row 11, columns A through D are presented below:

***Formula 1* (A11):** @INT(@RAND*@COUNT($Data:$A$11..$A$46))+1
***Formula 2* (B11):** @VLOOKUP($A11,$Data:$A$11..$D$46,1)
***Formula 3* (C11):** @VLOOKUP($A11,$Data:$A$11..$D$46,2)
***Formula 4* (D11):** @VLOOKUP($A11,$Data:$A$11..$D$46,3)

Formula 1 uses the @RAND function to randomly choose a number between 1 and the total number of data points in the original data (36). Formulas 2, 3 and 4 use the table LOOKUP function to retrieve data associated with the randomly chosen data point. These formulas can be copied down to the remaining 35 rows.

The next spreadsheet page shows the first of two spreadsheet macros. This macro, shown in Figure 3, allows the bootstrap to iterate automatically.

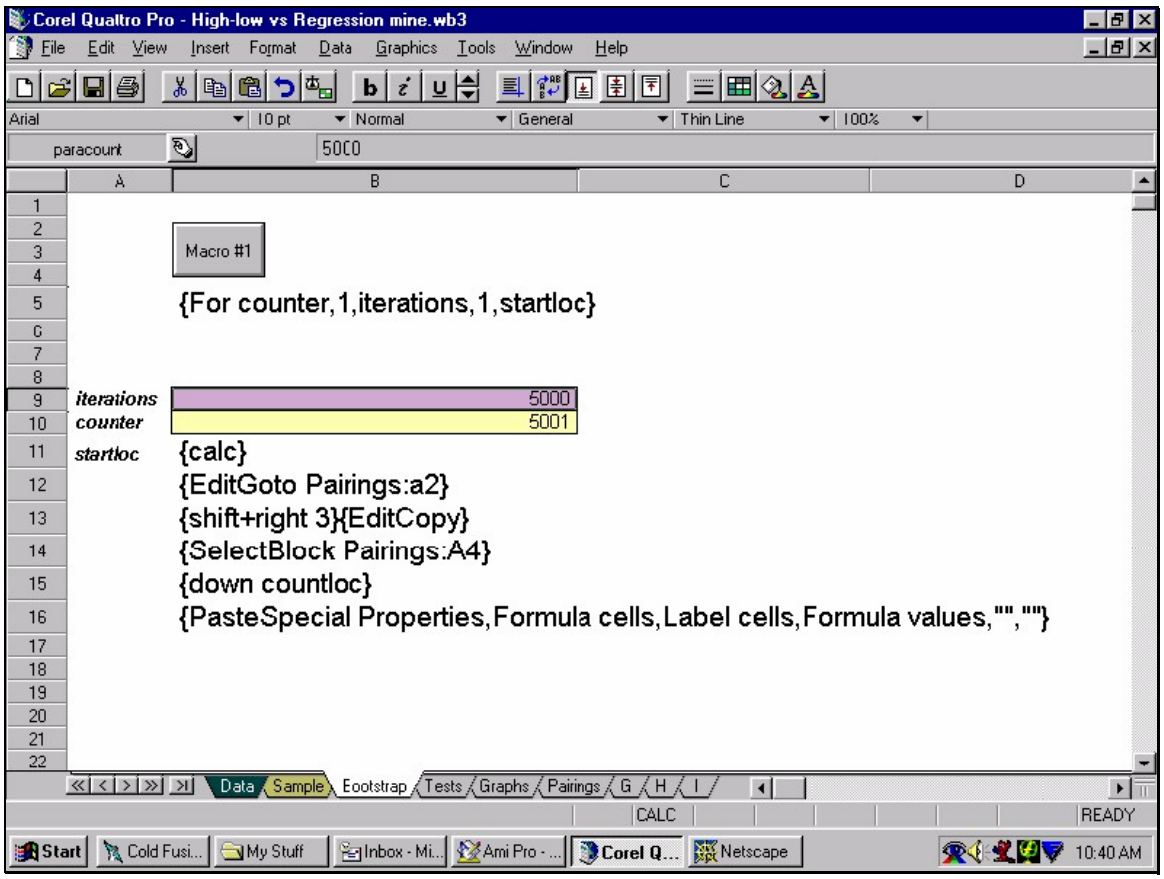



**Figure 3: Bootstrap Macro**

Here is what this macro is doing. First, a FOR loop is defined. The logic is: "For as long as the counter is less than or equal to the number of iterations, execute the code beginning in the start location". B11 is the start location for the code that gets executed. So, for each increment, the spreadsheet is CALCulated (a new sample is drawn and the calculations using each of the three methods are performed). The results of these calculations (only the fixed component was used in the comparison) are displayed on the Pairings page in row 2, shown in Figure 4. This row is selected and copied to the row associated with the current iteration number. Figure 4 shows the results after six iterations. Five thousand iterations were run in order to give more data points for the resulting distributions.

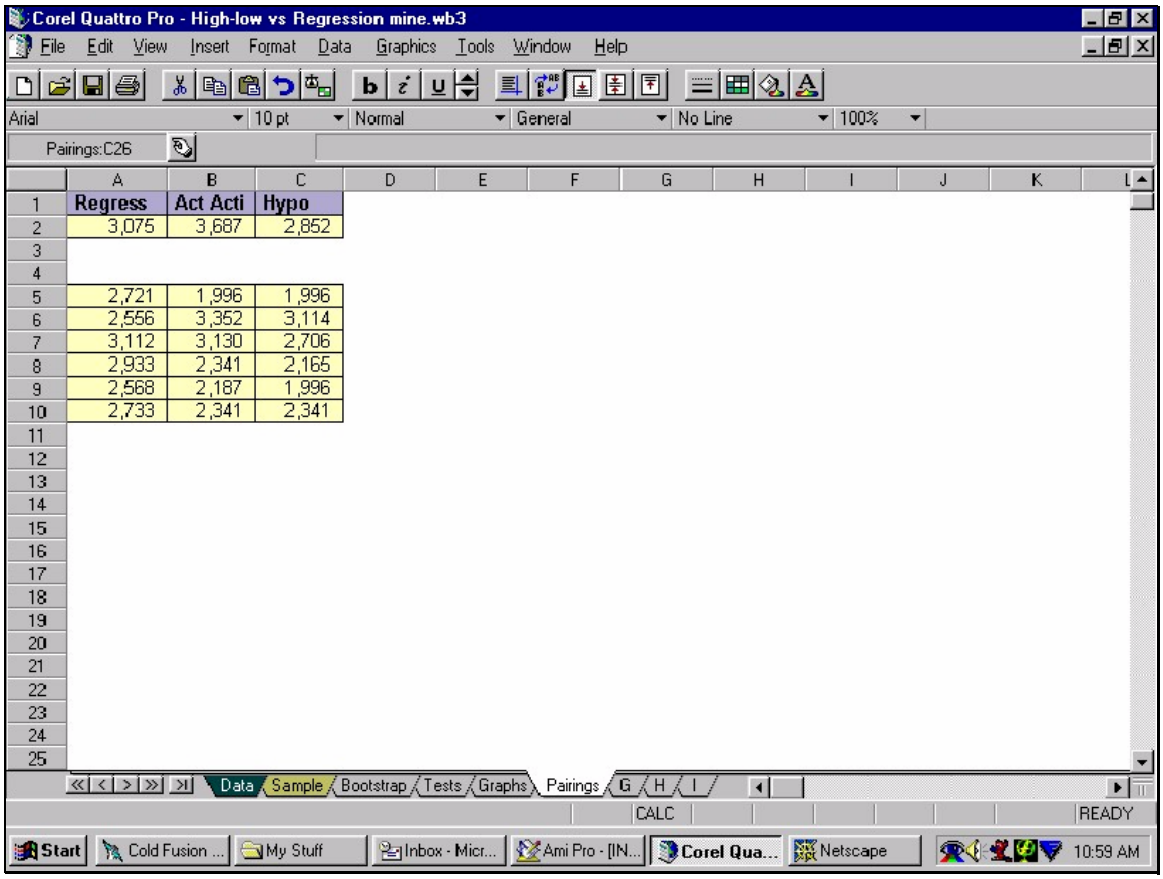



**Figure 4: Pairings Page**

The second macro is found on the Test Page shown in Figure 5. This macro uses the output of the 5000 iterations that were previously written to the Pairings Page. It takes the fixed cost components for the two models being compared and copies them into the range on the Test page beginning with B10. The vector of fixed cost points from one model is subtracted from the vector of fixed cost points from the second model, creating a vector of differences beginning in D10. The macro then copies the values found in the differences vector into column G and sorts this column in ascending order.

Column F contains percentage bins increments of .02%, from 0.00% to 100.00%. This assigns each of the 5000 ranked data points in the differences vector to a percentile place in the distribution. Then, this column is used to look up the difference associated with the particular level of confidence entered into cell G5. The respective lower and upper limits of the confidence interval are found by using the following two formulas:

***Formula 5* (F3):** @VLOOKUP(@VLOOKUP(J6,F11 through G5010,1))
***Formula 6* (G3):** @VLOOKUP(@VLOOKUP(J5,F11 through G5010,1)).

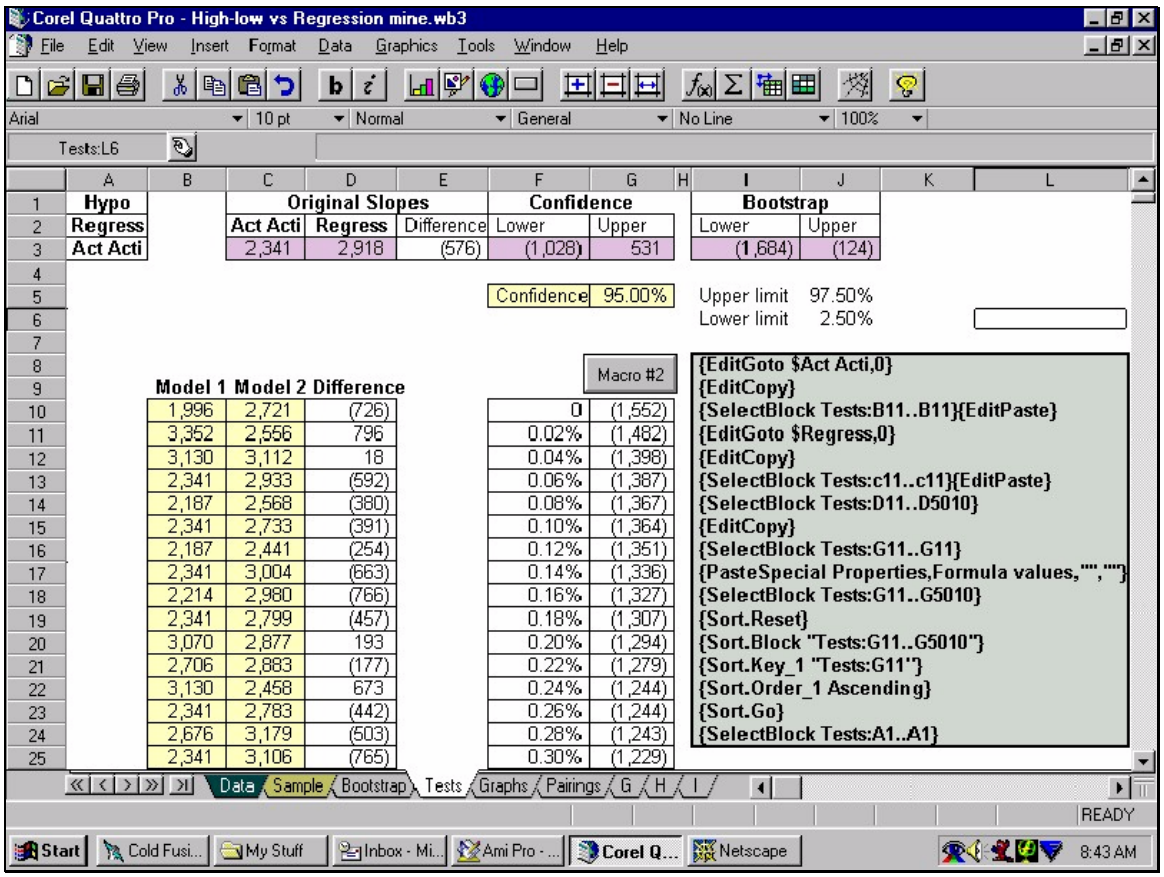



**Figure 5: Test Page**

Once the confidence interval has been determined, the lower and upper end of the bootstrap range can be calculated using Formulas 7 and 8.

***Formula 7 (I3):*** *-$Tests:$G$3+(2*$E$3)*
***Formula 8 (J3):*** *-$Tests:$F$3+(2*$E$3)*

For the *lower* end of the bootstrap range, the upper confidence limit is subtracted from two times the original difference in the fixed cost components of the models being compared. For the *upper* end of the bootstrap range, the lower confidence limit is subtracted from two times the original difference in the fixed cost components of the models being compared. If zero is *not* within the bootstrapped range, the two distributions are significantly different at the confidence level used.

## Results

Three comparisons were made among Least Squares Regression (LSR), Actual High-Low Pairing based on Activity (APA), and Hypothetical High-Low Pairing (HP): 1) LSR vs. APA; 2) LSR vs. HP; and 3) APA vs. HP. The results are shown in Table 3.

**Table 3: Results of Bootstrapping**

| Comparison | Significance |
|---|---|
| LSR vs. APA | 95 % |
| LSR vs. HP | 99 % |
| APA vs. HP | not significant |

In the case of the data set used in this study, the results indicate that using the least squares regression method of separating total cost into its fixed and variable components does produce results that are significantly different from the results produced by both implementations of the high-low method (while the two implementations of the high-low method were not found to produce significantly different results). Therefore, the null is rejected. Graphs showing the generated estimated distributions for the significant comparisons are shown in Figures 6 and 7. As can be readily observed, both graphs show distributions that are not normal. Bootstrapping is one of the few techniques that can be used to compare such distributions, and, as shown, it can be performed easily in a spreadsheet.

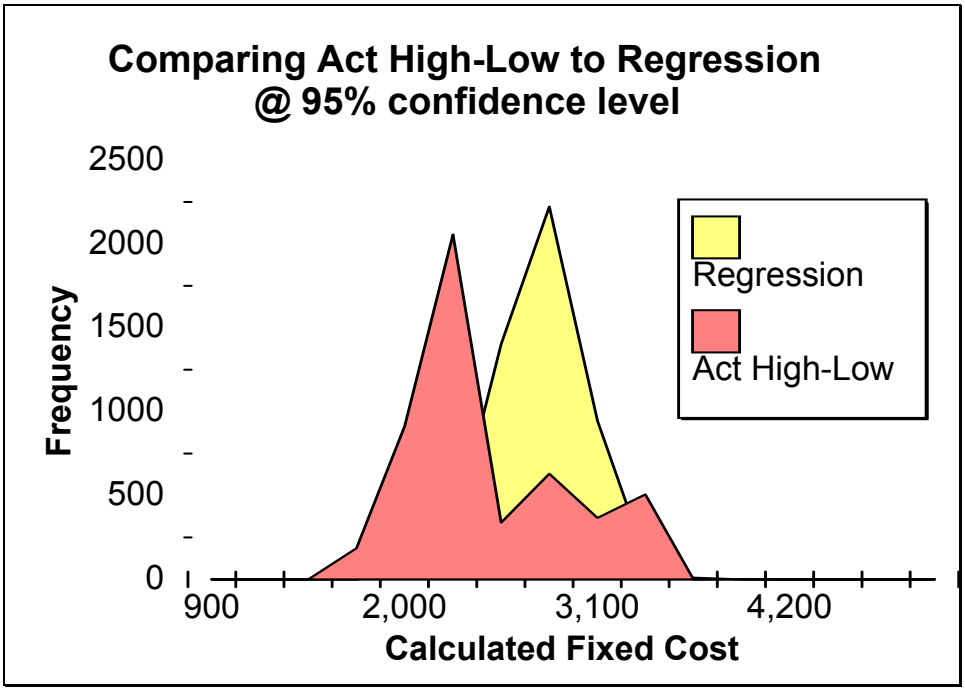



**Figure 6: Regression vs. Actual High-Low Pairing Based on Activity**

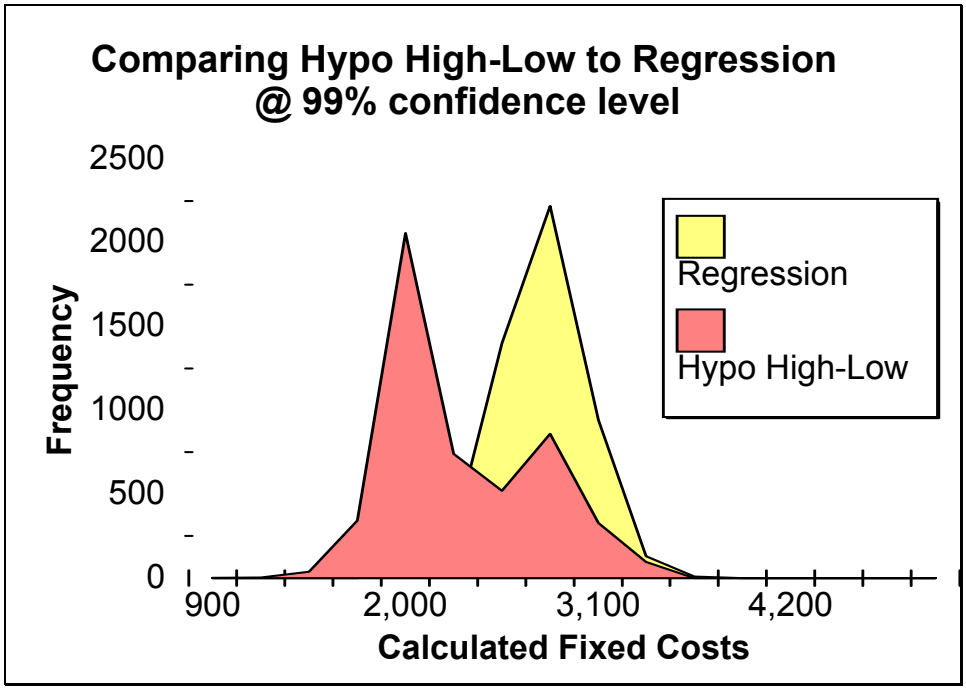



**Figure 7: Regression vs. Hypothetical High-Low Pairing**

## Conclusion and Future Direction

The technique of bootstrapping in a spreadsheet provides practitioners and educators with a powerful, yet simple to use, tool to compare methods for separating total costs into their fixed and variable components. The results of this study indicate that the high-low method, while easy to use, may be giving results that are significantly different from results obtained from regression. Because students now have the ability to do regression easily and inexpensively in a spreadsheet, and because of the theoretical shortcomings of the high-low method, it may be that educators should discontinue using and teaching the high-low method altogether.

Much more analysis needs to be done using real cost data from various industries. Researchers and accounting educators are encouraged to use this bootstrapping technique on various total cost data sets from various industries to compare the high-low method with the regression method in order to determine which method(s) of separating total costs into their fixed and variable components should continue to be used and taught. By doing this, a body of empirical evidence about these methods can be collected and analyzed.

## References:

Benston, G.J. (1966) Multiple Regression Analysis of Cost Behavior, *The Accounting Review*, October, 657-672.

Comiskey, E.E. (1966) Cost Control by Regression Analysis, *The Accounting Review*, April, 235-238.

Efron, B., (1982). The Jackknife, the Bootstrap, and Other Resampling Plans. CBMS-NSF Monograph 38, Society of Industrial and Applied Mathematics, Philadelphia.

Efron, B., and Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* 1:54-77.

Horngren, C.T. (1972) Cost Accounting: A Managerial Emphasis, 3rd ed.; Englewood Cliffs, J.J.: Prentice-Hall, Inc.

Johnson, K.H. and Harrell, H.W. (1999) Comparison of the High-Low and Regression Methods of Analyzing Mixed Costs Using a Spreadsheet Simulation, working paper.

Mooney, C.Z., and Duval, R.D. (1993). Bootstrapping: A Nonparametric Approach to Statistical Inference (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-095). Newbury Park, CA: Sage, 1993.

Nurnberg, H. (1977) An Unrecognized Ambiguity of the High-Low Method, *Journal of Business Finance and Accounting*, 4:4,.

Woodroof, J. (1997) Making Statistical Comparisons: An Application of the Bootstrap Using a Spreadsheet," *Review of Accounting Information Systems*, Summer, Volume 1, Number 3, pp 73-86.

# Appendix A: Horngren Data

| | | Total Cost Trucking Labor | Activity Direct Labor cost |
|---|---|---|---|
| 1 | Nov | 320000 | 9600 |
| 2 | Dec | 320000 | 10000 |
| 3 | Jan | 336000 | 10000 |
| 4 | Feb | 304000 | 9600 |
| 5 | Mar | 336000 | 9800 |
| 6 | Apr | 352000 | 10000 |
| 7 | May | 352000 | 10400 |
| 8 | Jun | 368000 | 11000 |
| 9 | Jul | 384000 | 11000 |
| 10 | Aug | 368000 | 10800 |
| 11 | Sep | 320000 | 11000 |
| 12 | Oct | 378000 | 11200 |
| 13 | Nov | 342000 | 10000 |
| 14 | Dec | 270000 | 9000 |
| 15 | Jan | 288000 | 9200 |
| 16 | Feb | 252000 | 8400 |
| 17 | Mar | 270000 | 9600 |
| 18 | Apr | 252000 | 8200 |
| 19 | May | 270000 | 8800 |
| 20 | Jun | 270000 | 8400 |
| 21 | Jul | 180000 | 6400 |
| 22 | Aug | 234000 | 7600 |
| 23 | Sep | 216000 | 7600 |
| 24 | Oct | 240000 | 7800 |
| 25 | Nov | 220000 | 7600 |
| 26 | Dec | 200000 | 7400 |
| 27 | Jan | 240000 | 8200 |
| 28 | Feb | 240000 | 8000 |
| 29 | Mar | 262000 | 8400 |
| 30 | Apr | 262000 | 8600 |
| 31 | May | 282000 | 9000 |
| 32 | Jun | 282000 | 9200 |
| 33 | Jul | 200000 | 7200 |
| 34 | Aug | 242000 | 8400 |
| 35 | Sep | 262000 | 8800 |
| 36 | Oct | 284000 | 8600 |