DEFENSE CONTRACT AUDIT AGENCY
DEPARTMENT OF DEFENSE
8725 JOHN J. KINGMAN ROAD, SUITE 2135
FORT BELVOIR, VA 22060-6219

IN REPLY REFER TO

OTS 730.5

June 10, 2011
11-OTS-027(R)

MEMORANDUM FOR REGIONAL DIRECTORS, DCAA
DIRECTOR, FIELD DETACHMENT, DCAA
HEADS OF PRINCIPAL STAFF ELEMENTS, HQ, DCAA

SUBJECT: Guidance on Curvilinear Regression and Adjusted R^2 in EZ-Quant Regression Analysis Version 1.2.0.

This memorandum provides guidance on two features included in the new EZ-Quant Regression module, version 1.2.0:

- Use of curvilinear (non-linear) regression analysis, and
- Adjusted R^2 (adjusted correlation of determination)

Curvilinear Regression:

Besides the linear regression equation, EZ-Quant Regression Analysis version 1.2.0 also includes two curvilinear equations using a single independent variable (simple regression analysis). These curvilinear equations are the exponential and power curves which are occasionally useful in contract auditing.

A historical relationship between a dependent and independent variable (such as factory overhead and direct manufacturing labor) may exhibit either a linear or non-linear relationship. If there is correlation between the two variables being plotted, the general path of the scatter of points might also indicate whether the correlation is linear or curved (non-linear). If the central path through the pattern of points from the lowest to the highest *x* value is approximately a straight line, the correlation is linear and there is said to be a constant or straight-line relationship between the variables. If the central path through the pattern is curved, the correlation is curvilinear and a non-linear relationship is said to exist between the variables.

It is important to distinguish between linear and nonlinear relationships because different computational analysis techniques are required. Substantial errors may result from assuming a straight-line or linear relationship when the true relationship is curvilinear. An important reason for constructing and evaluating a scatter diagram prior to examining the computational analysis is to determine, from the pattern of the points, the type of analysis that should be used. If a curvilinear relationship exists, it should be evident in the scatter diagram. However, the determination of a functional relationship, either linear or non-linear, should not be made exclusively from a scatter diagram, especially if only a few plot points are available. Visual

evidence in this respect is not exclusively compelling by itself. Auditors should also consider what type of functional relationship between the dependent and independent variables exists and whether it makes sense and is logically supportable.

There are a number of curvilinear equations, but only a few have been found to have application in contract cost audits. EZ-Quant Regression Analysis version 1.2.0 includes two curvilinear equations that have been useful in contract auditing which are the power and exponential curves. Either of these curvilinear equations can be used if a curvilinear relationship between the dependent and independent variables is evidenced by the historical data and it is logical. The power curve equation is expressed as $y = ax^b$, and the exponential curve equation is expressed as $y = ab^x$. Auditors should use the equation which has been judgmentally determined to have a logical relationship between the variables.

Auditors should be able to support the use of a non-linear equation for estimating the relationship between variables. The auditor should be able to explain why the specific type of estimated relationship is a reasonable representation of the actual interaction between the dependent and independent variables. This logical support is particularly important when the range of independent variable values is narrow, in which case most curves will themselves appear to be nearly linear within that range and more clearly non-linear only outside the range. Caution is advised concerning predictions from the curve or line using independent variable values significantly outside the data's relevant range. As values move increasingly further outside the relevant range, less reliable predictions are generated.

Adjusted R^2:

The coefficient of determination or R^2 represents the proportion of variation in the dependent variable *(y)* that can be explained by the independent variables *(x_1, x_2, etc.)* in a regression model. The coefficient of determination or R^2 can also be used to determine which of two different equations, each having the same dependent variable but a different independent variable, provides the better fit to historical data. The two computed R^2 amounts can be compared to each other to determine which independent variable provides a better predictor of the dependent variable. A direct comparison between the values of R^2 obtained in two different regression analyses is valid only if both analyses use (1) the same number of observations, and (2) the same number of independent variables. Otherwise, comparisons need to be made using the Adjusted R^2, since it is not limited to these specific circumstances.

An analysis can be made using the Adjusted R^2, a measure similar to R^2 that accounts for the number of independent variables and the number of observations. The Adjusted R^2 is a modification of R^2 that adjusts for the number of independent variables included in a regression analysis. As independent variables *(x)* are added to the model, each independent variable will explain some of the variance in the dependent variable *(y)* simply due to chance. One could continue to add independent variables to the model which would continue to improve the ability of the independent variables to explain the dependent variable, although some of this increase in the R^2 would be simply due to chance. The R^2 value will always increase when a new independent variable is added, but the Adjusted R^2 increases only if the new independent variable

improves the analysis more than would be expected by chance. The Adjusted R^2 attempts to yield a more true value to estimate R^2.

Adjusted R^2 values provide an objective means of choosing between two equally logical equations, whether or not they use different numbers of independent variables or different numbers of observations. The greater the number of observations used and the more substantial the differences are between the Adjusted R^2 values, the more reliable the comparison.

Field audit office personnel should direct questions regarding this memorandum to their regional offices. Regional offices should direct their questions to the Technical Audit Services Division, at (703) 767-2238, or email DCAA-OTS@dcaa.mil.

/s/ Thomas F. Trefry
/for/
Nina I.S. Kissinger
Deputy Assistant Director
Operations

DISTRIBUTION: C