Presented at the 2016 ICEAA Professional Development & Training Workshop - www.iceaaonline.com/atlanta2016

PRT-204

Develop PRESS for Nonlinear Equations

Presented at the
International Cost Estimating and Analysis Association (ICEAA)
Professional and Training Workshop
Atlanta Georgia
7-10 June 2016

Dr. Shu-Ping Hu

21 June 2016

TECOLOTE RESEARCH, INC.
420 S. Fairview Avenue, Suite 201
Goleta, CA 93117-3626
(805) 571-6366

TECOLOTE RESEARCH, INC.
5266 Hollister Avenue, Suite 301
Santa Barbara, CA 93111-2089
(805) 964-6963

ABSTRACT

This paper develops the algorithm to calculate the predicted residual sum of squares (PRESS) and predicted R^2 statistics when deriving nonlinear equations, including the Minimum-Unbiased-Percent Error (MUPE) cost estimating relationships (CER).

PRESS is a well-known "leave-one-out" cross-validation statistic. Another related statistic is the Predicted R^2 that is defined as the fraction of the variation in the dependent variable explained by the "leave-one-out" model. These two statistics are commonly used in regression analysis to determine (1) how well the model predicts for new observations and (2) if the model's R^2 is inflated by including more independent variables than necessary. Predicted R^2 is more useful than Adjusted R^2 when measuring the model's predictive power because it is calculated using observations not included in model building. However, many statistical packages only provide the PRESS and Predicted R^2 statistics for ordinary least squares (OLS). This paper develops a method to approximate these two cross-validation statistics for nonlinear regression such as MUPE and demonstrates their value in practical applications.

PURPOSE

The objectives of this paper are twofold. First, explain the importance of using the Predicted residual sum of squares (PRESS) and Predicted R^2 statistics in regression analysis. Although these two statistics are commonly used in regression analysis, they are seldom applied in the cost analysis community when evaluating the usefulness of a CER. This paper advocates the concept of cross-validation and recommends using PRESS and Predicted R^2 for cost analysis.

Second, many commercial statistical packages only provide the PRESS and Predicted R^2 statistics for linear and log-linear models. These two statistics are not readily available for nonlinear equations, such as MUPE or Minimum-Percentage Error Regression under Zero-Percentage Bias (ZMPE) CER. It would be very helpful to model builders if these two cross-validation statistics are available for nonlinear regression, which can be used to evaluate various fitted equations, including MUPE.

This paper investigates the possibilities to approximate the PRESS statistic for nonlinear equations using a **single** regression method for three different error terms, additive, MUPE, and log error terms. Although we can calculate PRESS directly by definition for a nonlinear equation, we should avoid running nonlinear regression **multiple** times. This is because (1) the computation process is extremely tedious, (2) one or more of the nonlinear regression runs may not converge, and (3) each run can be easily trapped in local minima. This paper proposes a single regression approach to derive the solution efficiently. The following topics will be discussed:

- Introduction
- Additive and Multiplicative Error Models
- Calculate PRESS for Linear Models
- Use of PRESS and Predicted R^2
- Algorithm

- Example Section for Validation
- Conclusions

INTRODUCTION

PRESS, a well-known "leave-one-out" (LOO) statistic, is commonly used in regression analysis for cross-validation. PRESS evaluates the fit of a specific model to a given data set by (1) systematically removing each observation from the data set, (2) refitting the equation, and (3) computing the square of the residual for the removed data point. The smaller the PRESS statistic, the better the model's predictive capability.

Definition of PRESS

The PRESS statistic is defined as the sum of the squares of all the residuals such that the predicted value is calculated for the omitted observation in each refitted regression model (see Allen, 1974). In mathematical terms, it is given by

$$PRESS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,-i})^2 \tag{1}$$

where:

n = the sample size
w_i = the weighting factor for the ith observation
y_i = the ith observation
$\hat{y}_{i,-i}$ = the ith predicted value from the equation fitted without the ith observation

Each residual in the PRESS statistic is commonly named the PRESS residual or LOO residual:

$$ith\ PRESS\ residual = y_i - \hat{y}_{i,-i} \quad for\ i = 1,...,n \tag{2}$$

The PRESS statistic can be used to evaluate a number of fitted equations for the same data set, with the lowest value of PRESS indicating the best predicative power. CERs that are over-fitted may tend to produce small residuals for observations included in the curve-fitting process but large residuals for observations that are excluded.

Definition of Predicted R^2

PRESS, calculated in unit space, is an absolute measure. Using PRESS alone, however, analysts do not really know whether this PRESS measure calculated for a given model is good or bad (or how bad it is). The Predicted R^2 statistic captures the predictive power of the model and puts PRESS in perspective:

$$\text{Predicted } R^2 = 1 - PRESS/SST \tag{3}$$

where SST is the total sum of squares of the dependent variable. Based upon Equations 1 and 3, the larger the Predicted R^2, the more favorable the model becomes. In addition, Predicted R^2 is more useful than Adjusted R^2 to test the robustness of a model because this statistic is calculated using observations not included in the regression.

ADDITIVE AND MULTIPLICATIVE ERROR MODELS

Both additive and multiplicative error models are briefly described below before explaining the math formula to compute the PRESS statistic.

Additive Error Model

An additive error model is generally stated as follows:

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \epsilon_i = f_i + \epsilon_i \quad (for\ i = 1, \ldots, n) \tag{4}$$

where:

- y_i = the observed dependent variable of the ith data point, i = 1 to n
- $f(\mathbf{x}_i, \boldsymbol{\beta})$ = f_i = the value of the hypothesized equation at the ith data point
- $\mathbf{x}_i$ = the ith data vector of the independent variables
- $\boldsymbol{\beta}$ = the vector of unknown parameters to be estimated by the regression equation
- ε_i = the error term with a mean of 0 and variance σ^2 (assumed to be independent of the cost drivers)
- n = the sample size

The corresponding standard error of estimate for an additive CER is commonly termed the standard error of estimate (SEE) or CER's standard error:

$$SEE = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / (n - p)} \tag{5}$$

where $\hat{y}_i$ is used to denote the predicted value of the ith data point and p is the total number of estimated parameters. Based upon Equation 4, the error distribution of the dependent variable is assumed to be the same across the entire data range regardless of the size of the dependent variable (e.g., cost). This is not a realistic assumption in cost estimating, especially when the cost elements are approaching the upper or the lower end of the data range.

Multiplicative Error Model

Multiplicative error terms are preferred in the cost analysis field because the error of an individual cost observation is generally proportional to the magnitude of the hypothetical equation rather than some fixed amount. In such cases, it is appropriate to hypothesize a multiplicative error term for a CER. A multiplicative error model is generally specified as

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) * \varepsilon_i = f_i * \varepsilon_i \quad \text{for i = 1, …, n} \tag{6}$$

(The definitions of y_i, $f(\mathbf{x}_i, \boldsymbol{\beta})$, $\mathbf{x}_i$, $\boldsymbol{\beta}$ and ε_i are the same as given in Equation 4). Unlike the additive error model, the standard deviation of the <u>dependent variable</u> in Equation 6 is proportional to the level of the hypothetical equation rather than some fixed amount across the entire data range. There are three popular methods to fit multiplicative error models: the Log-Error Model, the Minimum-Unbiased-Percentage-Error (MUPE) and Minimum-Percentage Error Regression under Zero-Percentage Bias (ZMPE) methods.

Log-Error Model. If the multiplicative error term (ε_i) in Equation 6 is further assumed to follow a log-normal distribution with a mean of zero and a variance of σ^2 (in log space), then the error can be measured by the following:

$$e_i = \ln(\varepsilon_i) = \ln(y_i) - \ln(f(\mathbf{x}_i, \boldsymbol{\beta})) \tag{7}$$

where "ln" stands for the natural logarithm function. The objective is then to minimize the sum of squared e_is (i.e., $(\Sigma(\ln(\varepsilon_i))^2$). The corresponding standard error of estimate for a log-error model is then given by

$$SEE = \sqrt{\sum_{i=1}^{n} (\ln(y_i) - \ln(\hat{y}_i))^2 / (n-p)} \tag{8}$$

If the transformed function is linear in log space, then OLS can be applied in log space to derive a solution for $\boldsymbol{\beta}$. In this situation, the CER is termed a log space OLS equation (LOLS) or a log-linear CER. If not, we need to apply a nonlinear regression technique to derive a solution.

MUPE Method. The general specifications of the MUPE and ZMPE methods are the same as given above (Equation 6), except that the error term is assumed to have a mean of **one** and variance, σ^2. Based upon this assumption of a multiplicative model, a generalized error term is defined by

$$e_i = \frac{y_i - f(\mathbf{x}_i, \boldsymbol{\beta})}{f(\mathbf{x}_i, \boldsymbol{\beta})} \tag{9}$$

where e_i now has a mean of **0** and variance σ^2 (since e_i given above becomes "$\varepsilon_i - 1$").

The difference between this percentage error (Equation 9) and the traditional percentage error is in the denominator, where <u>predicted</u> value instead of actual value is used as the baseline. The objective of the MUPE method is to find the values of the parameter vector $\boldsymbol{\beta}$ that minimize the sum of squares due to error (SSE). Instead of minimizing SSE in a single pass, the MUPE method solves for the hypothetical function, $f(\boldsymbol{x},\boldsymbol{\beta})$, in the numerator separately from the function in the denominator through an iterative process:

$$\text{Minimize} \quad \sum_{i=1}^{n} \left(\frac{y_i - f(\mathbf{x}_i, \boldsymbol{\beta}_k)}{f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_{k-1})} \right)^2 \tag{10}$$

where k is the iteration number and $f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_{k-1})$ is the predicted value from the previous iteration.

Based upon Equation 10, the weighting factor of each residual in the current iteration is equal to the reciprocal of the predicted value from the previous iteration. Since the denominator in Equation 10 is kept fixed throughout each iteration, the MUPE technique turns out to be a weighted least squares (WLS) process with an additive error. The final solution is derived when the change in the estimated parameters ($\boldsymbol{\beta}$ vector) between the current iteration and the previous iteration is within the analyst-specified tolerance limit. This optimization technique (Equation 10) is commonly referred to as Iteratively Reweighted Least Squares (IRLS; see Seber and Wild, 1989; Weisberg 1985; Wedderburn 1974). The corresponding standard error of estimate for the MUPE CER is commonly termed multiplicative error or standard percent error (SPE):

$$SPE = \sqrt{\sum_{i=1}^{n}((y_i - \hat{y}_i)/\hat{y}_i)^2/(n-p)} \tag{11}$$

Again, $\hat{y}_i$ is the predicted value in unit space for the ith data point and p is the total number of estimated parameters. The MUPE CER provides consistent estimates of the parameters and has zero proportional bias for all points in the data set. See Hu (2001) or Hu and Sjovold (1994) for detailed descriptions of the MUPE method.

ZMPE Method. The ZMPE method is an alternative approach to minimize the sum of squared proportional errors directly in unit space. Mathematically, it is stated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{n}\left(\frac{y_i - f(\mathbf{x}_i,\boldsymbol{\beta})}{f(\mathbf{x}_i,\boldsymbol{\beta})}\right)^2$$

$$\text{Subject to} \quad \sum_{i=1}^{n}\frac{y_i - f(\mathbf{x}_i,\boldsymbol{\beta})}{f(\mathbf{x}_i,\boldsymbol{\beta})} = 0 \tag{12}$$

This alternative method is a "constrained" minimization process (see Book and Lao, 1999; Hu and Smith, 2007). Both MUPE and ZMPE CERs have zero proportional bias for all the points in the data set, i.e., zero sample bias.

CALCULATE PRESS FOR LINEAR MODELS

The PRESS statistic can be calculated directly by its definition using Equation 1. However, the process can be very cumbersome as it requires refitting the regression equation many times. For example, given a data set of 100 observations, analysts have to refit the regression equation 100 times to compute the PRESS statistic using Equation 1. For linear models, however, there is an easy way to compute PRESS using a **single** regression instead of multiple regressions as given by the definition. The leverage value will be discussed first before explaining the simple formula to compute PRESS.

Leverage Value

In outlier analysis, the test statistic for an extreme value of the <u>predictors</u> is the leverage value (LV). The ith LV is, in fact, the <u>i[th] diagonal element of the hat matrix **H**</u>, and **H** is defined as

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W} \tag{13}$$

where

X = the design matrix of the predictors

V = the variance/covariance matrix of the errors of the observations

W = the weighting matrix (= $\mathbf{V}^{-1}$)

This definition is generalized to the cases where the weights are not identical and/or the errors are correlated among the observations. If all the observations are not correlated, then the off-diagonal elements of **V** will be zero and the weighting factor for the ith observation is simply the

reciprocal of the ith diagonal element of **V**. In this case, the LV of the ith observation (denoted by h_{ii}) is computed by

$$h_{ii} = w_i \mathbf{x}_i (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{x}_i' \quad \text{for the i}^{\text{th}} \text{ observation (i = 1, …, n)} \tag{14}$$

where

w_i = the weighting factor for the ith observation

$\mathbf{x}_i$ = the ith driver vector, i.e., $\mathbf{x}_i = (1, x_{1i}, x_{2i}, ..., x_{ki})$ if intercept is to be estimated

n = the total number of observations (i.e., the sample size)

The purpose of using a weighting variable in the process is when some of the observations in the data set are less reliable than others. This means the variances of the data points are not all equal. When this occurs, a weighting variable is added to reflect the relative quality of each data point. Statistically, the weighting factors should be chosen inversely proportional to the magnitude of the relative variances of the observations. The larger the variance of the data point, the less reliable the data point becomes. (Note: the weighting variable is included in the computation of leverage values to cover the topic of WLS.)

The hat matrix is very useful for outlier analysis. It can be used to construct the predictive values, standardized residual, Cook's D statistic, and leverage value. The predicted value of Y (i.e., $\hat{Y}$) can be expressed using the hat matrix **H**:

$$\hat{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y = \mathbf{H}Y \tag{15}$$

Therefore, the predicted value for the ith data point can be written as a linear combination of the n observations. Note that the LV depends only on the independent variables; it does not involve the dependent variable Y. If the ith LV is large, the observation y_i has driver values that are far from the center of the database. This measure is used to identify if an observation has an unusual predictor value. For a simple linear model (p = 2, **V** = **I**), the ith leverage value is given by $h_{ii} = 1/n + (x_i - \bar{x})^2 / \sum_k (x_k - \bar{x})^2$.

Here is an important property of the leverage values. The sum of all the leverage values equals the total number of estimated parameters.

$$\sum_{i=1}^{n} h_{ii} = Trace\,(H) = p \tag{16}$$

where the term "Trace" stands for the sum of the diagonal elements of a square matrix and the letter p is used to denote the total number of estimated parameters, including the intercept. This implies that the average of the LVs is p/n for an OLS equation. Therefore, an observation is flagged to have an unusual value of the predictors if its leverage value is greater than 2p/n or 3p/n.

Leverage points might impact the goodness-of-fit measures for the regression equation. For instance, the value of R^2 may drop substantially when removing the leverage point. For further information, see Belsley, Kuh, & Welsch (1981) or Cook and Weisberg (1982).

Simple Formula Using Leverage Value

For linear models, there is a much simpler way than Equation 1 to compute PRESS by a single regression equation using all the observations. It turns out that the PRESS residual for each observation is, in fact, the ordinary residual weighted by a function of the leverage value:

$$ith\ PRESS\ residual = (y_i - \hat{y}_{i,-i}) = \frac{(y_i - \hat{y}_i)}{(1 - w_i \mathbf{x_i}(\mathbf{X'WX})^{-1}\mathbf{x_i}')} = \frac{y_i - \hat{y}_i}{1 - h_{ii}} \quad (i = 1,\ldots, n) \quad (17)$$

where:

y_i = the ith observation

$\hat{y}_{i,-i}$ = the ith predicted value from the equation fitted without the ith observation

$\hat{y}_i$ = the ith predicted value using the equation fitted by **all** the observations

The definitions of w_i, h_{ii}, **W**, **X**, and $\mathbf{x}_i$ are described above. For a proof of Equation 17, see Montgomery and Peck, 1992.

According to Equation 17, the larger the leverage value (h_{ii}), the larger (in absolute term) the ith PRESS residual becomes. The observations with large PRESS residuals are generally high influential points. Furthermore, a large difference between the ordinary and PRESS residuals usually indicates a data point where the model may fit it well, but a model built without that data point may predict it poorly. The proof of Equation 17 is based upon a useful identity found in matrix inverse operation:

$$(\mathbf{Z} + c\mathbf{bb'})^{-1} = \mathbf{Z}^{-1} - \frac{c}{1 + c\mathbf{b'Z}^{-1}\mathbf{b}}\mathbf{Z}^{-1}\mathbf{bb'Z}^{-1} \quad (18)$$

where **Z** is a p-by-p nonsingular matrix, **b** is a p-by-1 vector, and c is a scalar. See Morrison (1976) or Bartlett (1951) for a detailed derivation of Equation 18.

Using Equation 17, the PRESS statistic can be computed by

$$PRESS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,-i})^2 = \sum_{i=1}^{n} w_i \frac{(y_i - \hat{y}_i)^2}{(1 - h_{ii})^2} \quad (19)$$

Therefore, the PRESS statistic is derived by a **single** regression, which saves a lot of computing time and effort. Note that the weighting factors (w_i's) are set to be one if the variances of the errors are the same.

USE OF PRESS AND PREDICTED R^2

Weight-Based CER

An illustrative example is given below to explain the usefulness of the PRESS and Predicted R^2 measures in regression analysis. Table 1 is a hypothetical data set where the weight of a black box is used to predict its cost. The weighting factors are assigned arbitrarily for demonstration purposes.

Table 1: **Black Box Cost and Weight with Weighting Variable**

Data Point	Cost $K	Weight in lbs	WF (*w*)
Obs 1	538.84	6.83	1.00
Obs 2	363.77	5.16	1.00
Obs 3	405.83	9.17	1.00
Obs 4	549.91	8.53	1.00
Obs 5	619.03	10.06	1.00
Obs 6	660.29	11.15	0.85
Obs 7	470.80	5.92	1.00
Obs 8	668.95	11.10	1.00
Obs 9	385.21	4.33	1.00
Obs 10	583.21	7.62	1.00
Obs 11	337.02	4.37	1.00
Obs 12	555.82	13.94	0.75
Obs 13	542.05	9.92	1.00
Obs 14	707.17	14.02	1.00
Obs 15	660.15	9.18	1.00
Obs 16	315.45	2.32	1.00
Obs 17	656.81	10.59	0.62
Obs 18	1701.28	25.12	1.00

(Note: WF, denoted by *w* in the formulas, is the weighting factor for each observation.)

Two weight-based CERs are used to illustrate the usefulness of the PRESS and predicted R^2 statistics in regression analysis. One CER is a weighted **linear** model; the other is a weighted **log-linear** model. Both examples use the same data set as given in Table 1.

Example 1: A linear CER is derived using an OLS regression model with the weighting variable, WF: Cost = 78.06 + 55.51(Weight). Its adjusted R^2 is 84.82%. Table 2 is the respective outlier analysis table:

Table 2: Outlier Analysis Table (for a weighted linear CER)

Obs #	Cost	Predicted Y Value	Residual	Std. Residual	Leverage	Cook's Distance
1	538.8400	457.1922	81.647777	0.719363	0.072174	0.020127
2	363.7700	364.4906	-0.720644	-0.006438	0.097687	0.000002
3	405.8300	587.0855	-181.255453	-1.584996	0.058111	0.077497
4	549.9100	551.5591	-1.649100	-0.014431	0.059443	0.000007
5	619.0300	636.4893	-17.459289	-0.152779	0.059406	0.000737
6	660.2900	696.9951	-36.705110	-0.295601	0.056084	0.002596
7	470.8000	406.6782	64.121811	0.568733	0.084478	0.014923
8	668.9500	694.2196	-25.269613	-0.221850	0.065559	0.001727

9	385.2100	318.4174	66.792596	0.602606	0.115163	0.023631
10	583.2100	501.0451	82.164934	0.720981	0.064596	0.017948
11	337.0200	320.6378	16.382199	0.147725	0.114248	0.001407
12	555.8200	851.8678	-296.047807	-2.269557	0.080871	0.226605
13	542.0500	628.7179	-86.667899	-0.758213	0.058960	0.018010
14	707.1700	856.3086	-149.138601	-1.341295	0.109559	0.110678
15	660.1500	587.6406	72.509448	0.634060	0.058105	0.012401
16	315.4500	206.8425	108.607550	1.012128	0.170678	0.105413
17	656.8100	665.9096	-9.099550	-0.062009	0.038389	0.000077
18	1701.2800	1472.4688	228.811204	3.220733	0.636488	9.081357

Based upon Equation 19, the PRESS statistic is given by

$$PRESS = \sum_{i=1}^{n} w_i \frac{(\text{Residual}_i)^2}{(1-\text{Leverage}_i)^2} = \sum_{i=1}^{n} w_i \frac{(y_i - \hat{y}_i)^2}{(1-h_{ii})^2} = 599{,}480.81$$

The corresponding predicted R^2 is then

Predicted R^2 = 1 – PRESS/SST = 1 – 599,480.81/1,555,385 = 61.46%

where $SST = \sum_i (w_i)(y_i - \bar{y}_w)^2 = \sum_i (w_i)(y_i - 594.31)^2$ = 1,555,385 and $\bar{y}_w$ (=594.31) is the weighted mean of the dependent variable. Note that the predicted R^2 (**61.46**%) is much smaller than the adjusted R^2 (**84.82**%) for this equation, which indicates that the model does not predict new observations as well as it fits the existing data.

Listed below is a Goodness-of-Fit Statistics table, including PRESS and Predicted R^2:

Table 3: Goodness-of-Fit Statistics (for a weighted linear CER)

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef	PRESS	R-Squared (Predicted)
117.8320	85.72%	84.82%	0.9258	599,480.8	61.46%

Example 2: A log-linear CER is generated using an OLS regression model in log space with the weighting variable, WF: Cost = 146.99(Weight$^{0.6268}$). Its adjusted R^2 is 76.35%. Table 4 is the respective outlier analysis table:

Table 4: Outlier Analysis Table (for a weighted log-linear CER)

Obs #	Ln(Cost)	Predicted Y Value	Residual $(y_i - \hat{y}_i)$	Std. Residual	Leverage (h_{ii})	Cook's Distance
1	6.2894	6.1946	0.0948	0.5323	0.0644	0.0097
2	5.8965	6.0188	-0.1223	-0.7004	0.1010	0.0276
3	6.0059	6.3792	-0.3733	-2.0914	0.0611	0.1422
4	6.3098	6.3339	-0.0241	-0.1351	0.0585	0.0006
5	6.4282	6.4373	-0.0091	-0.0514	0.0674	0.0001
6	6.4927	6.5018	-0.0091	-0.0471	0.0669	0.0001
7	6.1544	6.1050	0.0495	0.2798	0.0790	0.0034

8	6.5057	6.4990	0.0068	0.0382	0.0781	0.0001
9	5.9538	5.9089	0.0449	0.2627	0.1404	0.0056
10	6.3685	6.2632	0.1054	0.5896	0.0590	0.0109
11	5.8201	5.9147	-0.0946	-0.5529	0.1380	0.0245
12	6.3204	6.6417	-0.3213	-1.5824	0.0887	0.1218
13	6.2954	6.4285	-0.1332	-0.7481	0.0662	0.0199
14	6.5613	6.6453	-0.0841	-0.4863	0.1195	0.0161
15	6.4925	6.3799	0.1125	0.6306	0.0611	0.0129
16	5.7540	5.5178	0.2362	1.6338	0.3841	0.8325
17	6.4874	6.4695	0.0179	0.0784	0.0450	0.0001
18	7.4391	7.0108	0.4283	2.8229	0.3216	1.8886

Based upon Equation 19, the PRESS statistic is given by

$$PRESS = \sum_{i=1}^{n} w_i \frac{(\text{Residual}_i)^2}{(1-\text{Leverage}_i)^2} = \sum_{i=1}^{n} w_i \frac{(y_i - \hat{y}_i)^2}{(1-h_{ii})^2} = 0.9008$$

The corresponding predicted R^2 is then

Predicted R^2 = 1 – PRESS/SST = 1 – 0.9008/2.4387 = 63.06%

where $SST = \sum_i (w_i)(\ln(y_i) - \bar{y}_w)^2 = 2.4387$ and $\bar{y}_w = 6.3041$ are both evaluated in log space. Note that the predicted R^2 (**63.06**%) is 13.3% less than the adjusted R^2 (**76.35**%) for this equation, which indicates that the model does not predict new observations as well as it fits the existing data.

Table 5 lists Goodness-of-Fit Statistics, including PRESS and Predicted R^2:

Table 5: Goodness-of-Fit Statistics (for a weighted log-linear CER)

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef	PRESS	R-Squared (Predicted)
0.1842	77.74%	76.35%	0.8817	0.9008	63.06%

ALGORITHM

The goal of this paper is to approximate the PRESS statistic for nonlinear CERs using a single regression instead of running multiple regression equations. Three error terms, additive, log-error, and MUPE, will be validated for this approach.

A nonlinear model with an additive error term is generally stated as follows (see Draper and Smith, 1981 or Morrison, 1983):

$$\mathbf{y} = f(\mathbf{X},\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{20}$$

where:

$\mathbf{y}$ is a vector of the dependent variable (e.g., observed cost); i.e., $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and n is the sample size

$f(\mathbf{X},\boldsymbol{\theta})$ is a nonlinear function of $\mathbf{X}$ and $\boldsymbol{\theta}$

$\mathbf{X}$ is a set of n vectors of the independent variables; i.e., $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$

$\boldsymbol{\theta}$ is a set of p unknown parameters, i.e., $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$

$\boldsymbol{\varepsilon}$ is the error term vector with mean of $\mathbf{0}$ and variance $\mathbf{V}\sigma^2$ ($\boldsymbol{\varepsilon}$ is assumed to be independent of the cost drivers)

For each data point, it can be written as follows:

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \varepsilon_i = f_i + \varepsilon_i \quad \text{for i = 1, 2, …, n} \tag{21}$$

where:

y_i = the dependent variable of the ith data point (i = 1, 2,…, n); n is the sample size

$f_i = f(\mathbf{x}_i, \boldsymbol{\theta})$ = the value of the hypothesized equation for the ith data point

$\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})$, which is a set of k observed independent variables for data point i

ε_i is the error term for the ith data point with a mean of 0 and a variance σ_i^2

The Taylor series expansion of the nonlinear function f at a given point $\boldsymbol{\theta}^o$ is given by

$$f(\mathbf{x}_i, \boldsymbol{\theta}) \cong f(\mathbf{x}_i, \boldsymbol{\theta}^0) + \left.\frac{\partial f_i}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) = f_i^o + \left.\frac{\partial f_i}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \qquad \text{for i = 1,…,n} \tag{22}$$

where $\left.\frac{\partial f_i}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$ denotes the partial derivatives of $f(\mathbf{x}_i, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$**,** evaluated at the point $\boldsymbol{\theta}^o$. Note that the above approximation ignores the second and higher order terms of partial derivatives.

The above hypothesized model can be further simplified using matrix notations:

$$\mathbf{y} \cong f^o + \mathbf{Z}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + \boldsymbol{\varepsilon} \quad \Rightarrow \mathbf{y} - f^0 \cong \mathbf{Z}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + \boldsymbol{\varepsilon} \tag{23}$$

where:

$\boldsymbol{f}^o$ is a n-by-1 vector of the hypothesized equation evaluated at the point $\boldsymbol{\theta}^0$ and each data point; i.e., $\boldsymbol{f}^o = f(\mathbf{X}, \boldsymbol{\theta}^0) = (f_1^0, f_2^0, ..., f_n^0)'$

$\mathbf{Z}_0$ is an n-by-p matrix of the partial derivatives of $f(X,\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, evaluated at the point $\boldsymbol{\theta}^o$ and each observation; namely,

$$\mathbf{Z}_0 = \left.\frac{\partial f}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \left.\begin{pmatrix} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} & \cdots & \frac{\partial f_1}{\partial \theta_p} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_i}{\partial \theta_1} & \frac{\partial f_i}{\partial \theta_2} & \cdots & \frac{\partial f_i}{\partial \theta_p} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial \theta_1} & \frac{\partial f_n}{\partial \theta_2} & \cdots & \frac{\partial f_n}{\partial \theta_p} \end{pmatrix}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \left.\begin{pmatrix} \frac{\partial f_1}{\partial \boldsymbol{\theta}} \\ \cdots \\ \frac{\partial f_i}{\partial \boldsymbol{\theta}} \\ \cdots \\ \frac{\partial f_n}{\partial \boldsymbol{\theta}} \end{pmatrix}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \begin{pmatrix} \mathbf{z}_1 \\ \cdots \\ \mathbf{z}_i \\ \cdots \\ \mathbf{z}_n \end{pmatrix} \tag{24}$$

where $\mathbf{z}_i$ is the ith row of $\mathbf{Z}_o$, i.e., $\left.\frac{\partial f_i}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \mathbf{z}_i$. Note that the design matrix does not exist for nonlinear models. To approximate the denominator of the PRESS residual, the partial derivative

matrix $\mathbf{Z}_o$ is used to mimic OLS's design matrix **X**. Therefore, the PRESS residual for the ith data point can be estimated by

$$ith\ PRESS\ residual = (y_i - \hat{y}_{i,-i}) \cong \frac{(y_i - \hat{y}_i)}{(1 - w_i \mathbf{z}_i (\mathbf{Z_0'WZ_0})^{-1} \mathbf{z}_i')} = \frac{(y_i - \hat{y}_i)}{(1 - H_i)} \quad (i = 1,.., n) \quad (25)$$

where **W** and w_i are defined above (see Equations 13 and 14) and H_i is given by

$$H_i = w_i \mathbf{z}_i (\mathbf{Z_0'WZ_0})^{-1} \mathbf{z}_i' \quad \text{(for } i = 1,\ldots, n) \quad (26)$$

The statistic H_i (Equation 26) is used to estimate the leverage value for the ith data point, while the matrix $\mathbf{Z}_o$ is used as a proxy for OLS's design matrix **X**.

Given Equation 25, the PRESS statistic for a nonlinear model is then approximated as

$$Nonlinear\ PRESS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,-i})^2 \cong \sum_{i=1}^{n} w_i \frac{(y_i - \hat{y}_i)^2}{(1 - H_i)^2} \quad (27)$$

where H_i is used to estimate the leverage value of the ith data point for a nonlinear model.

The partial derivative matrix $\mathbf{Z}_o$ is widely used for nonlinear analysis—it is the basis for deriving the standard errors of the estimated parameters, the confidence intervals of the mean values, etc. However, the validity of using Equation 25 to approximate the PRESS residual depends upon whether the linearized form is a good approximation of the true model. Nonetheless, Equation 25, as well as Equation 27, is very useful for cross-validation.

EXAMPLE SECTION FOR VALIDATION

This paper developed a single regression method to approximate the PRESS residual for nonlinear equations. Three error terms, additive, log-error, and MUPE, will be validated for developing PRESS using a single regression. A cost vs. weight sample data set is also used to validate the approximated PRESS residual for a nonlinear CER. Table 6 is a hypothetical data set of generic satellite electronic units. Again, the weighting factors are assigned arbitrarily for demonstration purposes.

Table 6: **Cost and Weight of Generic Satellite Electronic Units**

Data Point	Cost $K	Weight in lbs	WF (w)
Obs 1	3106.64	77.05	1.00
Obs 2	29166.32	1236.77	0.62
Obs 3	4820.48	232.14	1.00
Obs 4	34111.22	863.36	1.00
Obs 5	6387.04	224.40	1.00
Obs 6	20871.60	720.44	0.75
Obs 7	28621.92	959.33	1.00
Obs 8	19796.80	332.50	1.00
Obs 9	7526.40	269.42	0.50
Obs 10	6002.24	123.84	1.00

Data Point	Cost $K	Weight in lbs	WF (w)
Obs 11	11668.48	316.15	1.00
Obs 12	6329.12	59.77	1.00
Obs 13	4683.20	59.17	1.00
Obs 14	21068.72	369.12	1.00

(Note: WF, denoted by w in the formulas, is the weighting factor for each observation.)

The SST for the dependent variable y (cost) is given by

$$SST = \sum_{i=1}^{n}(w_i)(y_i - \bar{y}_w)^2 = 1{,}375{,}549{,}914.28 \tag{28}$$

where w_i's are the weighting factors and $\bar{y}_w$ is a weighted mean of the cost variable:

$$\bar{y}_w = \sum_i (w_i \times y_i) / \sum_i w_i = 14{,}304.26 \tag{29}$$

Additive Error CER: y = a(x^b) + ε (E(**ε**) = **0**, Var(**ε**) = **V**σ^2)

A nonlinear weight-based CER is generated from this data set. Its error term is assumed to be additive; a weighting variable (WF) is also applied. The fitted equation is given below:

Cost = 225.5949*Weight^0.70886 Adj. R^2 = 82.4% (30)

Table 7 compares the PRESS residuals approximated by a single regression (Equation 25) with those calculated by the direct approach using the LOO definition (Equation 2).

Table 7: Comparison of PRESS Residual between Single Regression and its Definition

Data Point	WF (w)	Residual ($y_i - \hat{y}_i$)	H_i	$y_i - \hat{y}_{i,-i}$ Single Regr	$y_i - \hat{y}_{i,-i}$ LOO Def	% Difference
Obs 1	1.00	-1,800.10	0.08381	-1,964.77	-1,956.66	0.41%
Obs 2	0.62	-5,936.22	0.38303	-9,621.64	-9,688.84	-0.69%
Obs 3	1.00	-5,902.66	0.12094	-6,714.77	-6,660.69	0.81%
Obs 4	1.00	6,903.81	0.22278	8,882.69	8,868.27	0.16%
Obs 5	1.00	-4,081.41	0.12077	-4,642.04	-4,606.37	0.77%
Obs 6	0.75	-3,060.16	0.11428	-3,455.01	-3,454.69	0.01%
Obs 7	1.00	-696.20	0.29367	-985.66	-979.13	0.67%
Obs 8	1.00	5,963.23	0.11769	6,758.65	6,742.03	0.25%
Obs 9	0.50	-4,390.71	0.06033	-4,672.61	-4,656.59	0.34%
Obs 10	1.00	-866.56	0.10543	-968.69	-962.24	0.67%
Obs 11	1.00	-1,679.36	0.11862	-1,905.38	-1,893.06	0.65%
Obs 12	1.00	2,230.74	0.07178	2,403.23	2,385.89	0.73%
Obs 13	1.00	614.02	0.07131	661.17	657.43	0.57%
Obs 14	1.00	6,171.70	0.11555	6,977.99	6,968.17	0.14%
PRESS				332,790,557	331,656,343	0.34%
Pred R^2				75.81%	75.89%	0.1%

Here, the H column contains the diagonal elements of the nonlinear "hat" matrix and its ith element is given by Equation 26. Hence, the PRESS residual of the ith data point is approximated by $(y_i - \hat{y}_i)/(1 - H_i)$ using a single regression (Column 5). The last column of Table 7 illustrates the percentage differences between the PRESS residuals calculated by the single regression method and direct approach. The percentage differences are all within **1%**.

The PRESS statistic calculated using the definition (Column 6) is given by

$$PRESS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,-i})^2 = 331{,}656{,}343 \tag{31}$$

The PRESS statistic calculated using a single regression approximation (Column 5) is given by

$$PRESS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,-i})^2 \cong \sum_{i=1}^{n} w_i \left((y_i - \hat{y}_i)/(1 - H_i)\right)^2 = 332{,}790{,}557 \tag{32}$$

The percent difference between these two PRESS statistics is 0.34%, which is well within **one** percent. The corresponding Predicted R^2 measures are 75.81% and 75.89%, respectively. Below is a summary table:

Table 8: **Comparisons of PRESS and Predicted R^2 for Cost = aWeight^b + ε**

Statistics	Single Regr	LOO Definition	% Difference
PRESS	332,790,557	331,656,343	0.34%
Predicted R^2	75.81%	75.89%	0.1%

where Predicted R^2 = 1 – PRESS/SST (see Equation 3) and SST = 1,375,549,914. Table 8 shows that the single regression approach provides very good approximations of the PRESS and Predicted R^2 statistics for this nonlinear power form equation.

MUPE CER: y = a(x^b)*ε

With the same data set, the MUPE assumption is now applied to the power-form equation. Given below is the respective MUPE CER:

Cost = 241.06*Weight^0.69115 Adj. R^2 = 68.7% (33)

As explained above, the MUPE technique turns out to be a WLS process with an additive error and the weighting factor for each data point is the square of the reciprocal of its predicted value. Therefore, we proceed using the same single regression process as the additive error case. The challenge for the MUPE case is that the weighting factors are **not** fixed in the individual leave-one-out regression models. In other words, the weighting factors vary from one leave-one-out model to another. Consequently, using Equation 27 to derive PRESS for a MUPE CER may not be as good as the additive error case with fixed weighting factors.

Table 9 compares MUPE's PRESS residuals derived by a single regression with those calculated using the direct approach. Again, the ith PRESS residual in Column 4 is derived by a single regression approximation: $(y_i - \hat{y}_i)/(1 - H_i)$.

Table 9: Comparison of MUPE's PRESS Residuals and Percentage Errors between Single Regression Method and Direct Approach

Data Point	Residual $(y_i - \hat{y}_i)$	H_i	$y_i - \hat{y}_{i,-i}$ Single Regr	$y_i - \hat{y}_{i,-i}$ LOO Def	% Error Single Regr	% Error LOO Def
Obs 1	-1,748.13	0.19334	-2,167.11	-2,139.71	41%	41%
Obs 2	-3,898.10	0.24316	-5,150.50	-5,049.08	15%	15%
Obs 3	-5,583.84	0.07352	-6,026.96	-6,026.13	56%	56%
Obs 4	8,319.85	0.17099	10,035.90	9,947.85	-42%	-41%
Obs 5	-3,776.26	0.07447	-4,080.09	-4,079.19	39%	39%
Obs 6	-1,887.45	0.14205	-2,199.96	-2,185.35	10%	9%
Obs 7	881.53	0.19013	1,088.48	1,075.85	-4%	-4%
Obs 8	6,459.69	0.07424	6,977.74	6,975.68	-54%	-54%
Obs 9	-4,005.93	0.07145	-4,314.18	-4,314.19	36%	36%
Obs 10	-736.93	0.11922	-836.68	-833.13	12%	12%
Obs 11	-1,211.84	0.07296	-1,307.22	-1,307.06	10%	10%
Obs 12	2,255.86	0.24700	2,995.83	2,956.66	-90%	-88%
Obs 13	638.25	0.24933	850.24	835.49	-22%	-22%
Obs 14	6,732.88	0.07812	7,303.42	7,298.36	-53%	-53%
PRESS					242.9%	237.9%
Pred R^2					56.4%	57.3%

As described above, the weighting factor (w_i) for each data point is 1/(predicted value)^2 for the MUPE CER. Therefore, MUPE's PRESS statistic is the sum of all percentage errors (see Equation 34 below).

As shown in Table 9, the individual PRESS percent error approximated by the single regression method closely follows its definition except for observation 12, which is off by two percent (columns in light blue). The individual PRESS residuals derived by the single regression are also very close to those calculated directly (columns in orange); the differences are all within **2%**.

The PRESS statistic for MUPE using the LOO definition (Column 7) is given by

$$MUPE's\ PRESS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,-i})^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{i,-i})^2}{(\hat{y}_{i,-i})^2} = \sum_{i=1}^{n} (\%\ Error_{(i)})^2 = 237.9\% \quad (34)$$

where the predicted values (circled in Equation 34) are derived from the leave-one-out model to match the LOO definition.

The PRESS statistic using the single regression approximation (Column 6) is given by

$$MUPE's\ PRESS = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{i,-i})^2}{(\hat{y}_{i,-i})^2} \cong \sum_{i=1}^{n} \frac{((y_i - \hat{y}_i)/(1 - H_i))^2}{(\hat{y}^*_{i,-i})^2} = 242.9\% \quad (35)$$

where the predicted value, $\hat{y}^*_{i,-i}$, is calculated as the actual cost minus the PRESS residual, to match the LOO definition.

The percent difference between these two MUPE's PRESS statistics is 2.1%, which is still reasonably good. The respective Predicted R^2 measures are calculated to be 56.4% and 57.3%. Below is a summary table:

Table 10: Comparisons of MUPE's PRESS and Predicted R^2 for Cost = (aWeight^b)ε

Statistics	Single Regr	LOO Definition	Difference	% Difference
PRESS	2.429	2.379	0.005	2.10%
Predicted R^2	56.4%	57.3%	-0.9%	-1.57%

where Predicted R^2 = 1 – PRESS/SST and SST = 5.56962. Table 10 demonstrates that the single regression approach provides very good approximations of the PRESS and Predicted R^2 statistics for this power form MUPE equation.

Log-Error CER: $y = a(x^b)*\varepsilon$ ($\varepsilon \sim LN(\mathbf{0}, \mathbf{I}\sigma^2)$)

A power form weight-based CER is generated from this data set. Its error term is assumed to be multiplicative; a weighting variable (WF) is applied as in the additive case. The regression is done in log space and the fitted equation is given below:

Cost = 200.1*Weight^0.7167 Adj. R^2 = 76.7% (36)

Table 11 compares the PRESS residuals approximated by a single regression (Equation 25) with those calculated by the direct approach (Equation 2).

Table 11: Comparison of PRESS Residual between Single Regression and Direct Approach

Data Point	WF (w)	Residual $\ln(y_i) - \ln(\hat{y}_i)$	H_i	$\ln(y_i) - \ln(\hat{y}_{i,-i})$ Single Regr	$\ln(y_i) - \ln(\hat{y}_{i,-i})$ LOO Def	% Diff
Obs 1	1.00	-0.3711	0.19835	-0.4629	-0.4629	0%
Obs 2	0.62	-0.1210	0.17465	-0.1466	-0.1466	0%
Obs 3	1.00	-0.7222	0.07859	-0.7838	-0.7838	0%
Obs 4	1.00	0.2932	0.19893	0.3660	0.3660	0%
Obs 5	1.00	-0.4165	0.07927	-0.4523	-0.4523	0%
Obs 6	0.75	-0.0683	0.12403	-0.0780	-0.0780	0%
Obs 7	1.00	0.0422	0.22098	0.0542	0.0542	0%
Obs 8	1.00	0.4330	0.08311	0.4722	0.4722	0%
Obs 9	0.50	-0.3834	0.03893	-0.3989	-0.3989	0%
Obs 10	1.00	-0.0526	0.12212	-0.0599	-0.0599	0%
Obs 11	1.00	-0.0595	0.08118	-0.0648	-0.0648	0%
Obs 12	1.00	0.5225	0.25448	0.7009	0.7009	0%
Obs 13	1.00	0.2286	0.25693	0.3076	0.3076	0%
Obs 14	1.00	0.4203	0.08843	0.4611	0.4611	0%
PRESS				2.2969	2.2969	0%
Pred R^2				70.93%	70.93%	0%

(Note: SST = 7.9017 in this case.)

The PRESS statistic derived by the LOO definition (Column 6) is given by

$$PRESS = \sum_{i=1}^{n} w_i (\ln(y_i) - \ln(\hat{y}_{i,-i}))^2 = 2.2969 \tag{37}$$

The PRESS statistic calculated using the approximation (Column 5) is given by

$$PRESS = \sum_{i=1}^{n} w_i (\ln(y_i) - \ln(\hat{y}_{i,-i}))^2 \cong \sum_{i=1}^{n} w_i \left((\ln(y_i) - \ln(\hat{y}_i)) / (1 - H_i)\right)^2 = 2.2969 \tag{38}$$

This case shows there is no difference between these two PRESS statistics or their respective predicted R^2 measures. This is because the power-form equation becomes linear in log space and Equation 26 (based upon the partial derivative matrix $\mathbf{Z}_0$) precisely estimates the leverage values for the log-linear CER. A **linear** equation regressed in log space is also tested as an excursion; the differences between the single regression and direct method are also within 2%.

CONCLUSIONS

PRESS and Predicted R^2 are useful cross-validation statistics. Use PRESS to evaluate the model's predictive power. The smaller the PRESS statistic, the better the model's predictive capability.

PRESS is an absolute measure; predicted R^2 puts PRESS in a relative scale. PRESS alone does not provide useful insight into the robustness of the regression model. Analysts do not really know whether the PRESS measure calculated for a particular regression model is good or bad because PRESS is an absolute measure. The Predicted R^2 statistic captures the predictive power of the model and puts PRESS in perspective.

Use predicted R^2, not adjusted R^2, to judge the predictive power of the regression model. The adjusted R^2 measure is very popular to measure the quality of the fit in regression analysis. However, predicted R^2 is more useful than adjusted R^2 to access the ability to predict **future** observations because predicted R^2 is calculated using observations not included in the regression.

Recommend using PRESS and Predicted R^2 for all equations. Many commercial statistical packages only provide these two statistics for linear and log-linear models. They are not readily available for nonlinear equations, such as MUPE or ZMPE. These two cross-validation statistics should be very valuable to model developers when deriving nonlinear CERs**.**

Recommend using a single regression algorithm developed in this paper to approximate PRESS and Predicted R^2 for nonlinear equations, including MUPE. Although PRESS can be calculated directly by definition for a nonlinear equation, avoid running nonlinear regression multiple times. This is because (1) the computation process is extremely tedious, (2) one or more of the nonlinear regression runs may not converge, and (3) each run can be easily trapped in local minima. This paper proposes a single regression approach to derive the solution efficiently. The approximated PRESS and Predicted R^2 measures are validated against their definitions for three error terms: additive, MUPE, and log-error terms. The partial derivative matrix $\mathbf{Z}_0$ (as a proxy for OLS's design matrix $\mathbf{X}$) works really well. A data set of generic satellite electronic units is used in the validation process. We should also generate the PRESS and Predicted R^2 measures for the **ZMPE** CERs in the future.

Presented at the 2016 ICEAA Professional Development & Training Workshop - www.iceaaonline.com/atlanta2016

REFERENCES

1. Allen, D. M., "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction," Technometrics, 16, (1974), 125–127.
2. Bartlett, M. S., "An inverse adjustment arising in discriminant analysis," Annals of Mathematical Statistics, Volume 22 (1951), pp. 107-111.
3. Belsley, D. A.; E. Kuh; R. E. Welsch, "Regression Diagnostics," John Wiley & Sons, Inc., 1981.
4. Book, S. A. and N. Y. Lao, "Minimum-Percentage-Error Regression under Zero-Bias Constraints," Proceedings of the 4th Annual U.S. Army Conference on Applied Statistics, 21-23 Oct 1998, U.S. Army Research Laboratory, Report No. ARL-SR-84, November 1999, pages 47-56.
5. Cook, R. D. and S. Weisberg, "Residuals and Influence in Regression," Chapman and Hall, 1982.
6. Draper, N. R. and H. Smith, "Applied Regression Analysis (2nd edition)," New York: John Wiley & Sons, Inc., 1981.
7. Hu, S., "The Minimum-Unbiased-Percentage-Error (MUPE) Method in CER Development," 3rd Joint Annual ISPA/SCEA International Conference, Vienna, VA, 12-15 June 2001.
8. Hu, S. and A. R. Sjovold, "Multiplicative Error Regression Techniques," 62nd MORS Symposium, Colorado Springs, Colorado, 7-9 June 1994.
9. Hu, S. and A. Smith, "Why ZMPE When You Can MUPE," 6th Joint Annual ISPA/SCEA International Conference, New Orleans, LA, 12-15 June 2007.
10. Montgomery D. C. and E. A. Peck, "Introduction to Linear Regression Analysis (2nd edition)," Wiley-Interscience, 1992.
11. Morrison, D. F., "Multivariate Statistical Methods, 2nd ed.", New York: McGraw-Hill Book Company, 1976.
12. Morrison, D. F., "Applied Linear Statistical Methods", New Jersey: Prentice-Hall, Inc. 1983.
13. Seber, G. A. F. and C. J. Wild, "Nonlinear Regression," New York: John Wiley & Sons, 1989, pages 37, 46, 86-88.
14. Wedderburn, R. W. M., "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," Biometrika, Vol. 61, Issue 3 (December 1974), pages 439-447.
15. Weisberg, S., "Applied Linear Regression, 2nd Edition," New York: John Wiley & Sons, 1985, pages 87-88.