PRT-191

Generalized Degrees of Freedom (GDF)

Dr. Shu-Ping Hu

22 June 2015

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

Generalized Degrees of Freedom (GDF)

Dr. Shu-Ping Hu

ABSTRACT

Two popular regression methods for the multiplicative-error model are the Minimum-Unbiased-Percent Error (MUPE) and Minimum-Percentage Error under the Zero-Percentage Bias (ZMPE) methods. The MUPE method, an Iteratively Reweighted Least Squares (IRLS) regression, does not use any constraints, while the ZMPE method requires a constraint as part of the curve-fitting process.

However, ZMPE users do not adjust the degrees of freedom (DF) to account for constraints included in the regression process. As a result, fit statistics for the ZMPE equations, e.g., the standard percent error (SPE) and generalized R^2 (GRSQ), can be incorrect and misleading. This results in incomparable fit statistics between ZMPE and MUPE. This paper details why DF should be adjusted and recommends a Generalized Degrees of Freedom (GDF) measure to calculate fit statistics for constraint-driven cost estimating relationships (CER). It also explains why ZMPE's standard error underestimates the spread of the CER error distribution. Illustrative examples are provided.

OUTLINE

This paper introduces a new concept to the cost community. We explain in detail why degrees of freedom (DF) should be adjusted for any constraint-driven regression equations and we use ZMPE as an example. The following topics will be discussed:

- Purpose
- Using constraints in CER development
- Additive and Multiplicative-Error Models
- ZMPE vs. MUPE:
  - Is ZMPE CER Unbiased?
  - Is ZMPE's Standard Percent Error (SPE) Unbiased?
- Definition of Generalized Degrees of Freedom (GDF)
- Calculate Fit Statistics Using GDF
- Example Section
- Conclusions
- Future Study Items for Constraint Driven Regression

PURPOSE

The main objectives of this paper are threefold. First, we address a potential shortcoming when deriving constraint-driven equations, e.g., ZMPE CERs. Second, we suggest an appropriate adjustment (GDF measure) to calculate the CER fit statistics when constraints are included in the

regression process. Without a proper adjustment for DF, the fit measures for the ZMPE equations, e.g., the standard percent error (SPE) and generalized R^2 (GRSQ), can be incorrect and misleading. This results in an unfair comparison with the statistical measures derived by the MUPE method.

Lastly, we explain why using ZPME CERs without adjustment in cost uncertainty analysis may unduly tighten the S-curve. The unadjusted SPE for the ZMPE equation is smaller than the SPE for the MUPE equation. Since the SPE measures are used as the basis for cost uncertainty analysis, using unadjusted SPE measures for ZMPE equations may have a direct impact on the cost uncertainty results. A math proof is provided to support this conclusion.

USING CONSTRAINTS IN CER DEVELOPMENT

Solver (an Excel add-in program) is a popular tool used to generate nonlinear CERs, especially when constraints are specified. Many analysts have been using the following methodologies in Solver to develop CERs:

- Minimizing the sum of squared residuals under the Zero-Percentage Bias constraint (i.e., the mean of the percentage errors is zero)
- Minimizing the sum of squared percentage errors under the Zero-Percentage Bias constraint (i.e., the ZMPE CER)
- Minimizing the sum of squared percentage errors or residuals in log space under the Zero-Bias constraint (i.e., the mean of the residuals is zero) using the Balance-Adjustment Factor (BAF) (see Book, 2006)

Considerations. These methodologies are very different from traditional methods, such as ordinary least squares (OLS) or log-linear models (OLS performed in log space). When various constraints are specified, we may not have the DF as given by the traditional definition when constraints are **not** part of the curve fitting process. In fact, the DF should be reduced if constraints are specified when deriving CERs. (See the explanation in the definition section.) In addition, constraints should be made consistent with the underlying hypothesis. Mixing a multiplicative constraint with an additive error model or vice versa is neither logical nor necessary.

Suggestions. Here are a few suggestions when using constraints in Solver:

- Do not abuse Solver. Although Solver can easily handle constraints in the minimization process, we should not specify excessive constraints.
- Explore different starting points to see if the solution stabilizes when using Solver. As we know, Solver is sensitive to the starting points and can be easily trapped in local minima, especially when fitting complicated equations or the sample size is small. We also know that the ZMPE CERs seem to be less stable than the MUPE ones (see Hu and Smith, 2007). Be cautious when fitting ZMPE equations to small samples. As a precaution, explore different starting points to see if they lead to the same solution to ensure the resultant solution is stable.
- Specify "meaningful" constraints. Make sure constraints included in the process are necessary, logical, and statistically sound as DF can be reduced by additional constraints.

If we lose one DF for each constraint, we want to ensure it is worthwhile. Since we generally deal with small samples, we should preserve as many DF as we can.

- Calculate the DF and fit statistics properly if constraints are specified when deriving CERs, as well as any estimating relationships.

ADDITIVE AND MULTIPLICATIVE ERROR MODELS

We first introduce additive and multiplicative error models and then explain three different methodologies for fitting multiplicative error models.

Additive Error Model. An additive error model is generally stated as follows:

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

where:

Y_i = observed cost 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
$\mathbf{x}_i$ = vector of the cost driver variables at the ith data point
$\boldsymbol{\beta}$ = vector of coefficients (unknown parameters) to be estimated by the regression equation
ε_i = error term with a mean of 0 and variance σ^2 (assumed to be independent of the cost drivers)
n = 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:

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

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 1, 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 experience tells us that 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

$$Y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) * \varepsilon_i = f_i * \varepsilon_i \quad \text{for i = 1, …, n} \tag{3}$$

(Most of the definitions of Y_i, $f(\mathbf{x}_i,\boldsymbol{\beta})$, etc. are the same as given in Equation 1). Unlike the additive error model, the standard deviation of the dependent variable in Equation 3 is proportional to the level of the hypothetical equation rather than some fixed amount across the entire data range. Before explaining the MUPE and ZMPE methods, we will start with the minimum-percentage-error (MPE) method.

A. MPE Method. The general specification for MPE, MUPE, and ZMPE models is the same as given above (Equation 3), except that the error term is assumed to have a mean of 1 and variance, σ^2. Based upon this assumption of a multiplicative model, a generalized error term is defined by

$$e_i = \frac{y_i - f(\mathbf{x}_i, \boldsymbol{\beta})}{f(\mathbf{x}_i, \boldsymbol{\beta})} \tag{4}$$

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 4) and the traditional percentage error is in the denominator, where predicted value instead of actual value is used as the baseline. The objective of the MPE method is to find the values of the parameter vector $\boldsymbol{\beta}$ that minimize the sum of squares due to error (SSE); namely, the sum of squared e_is (see Young 1991):

$$\text{Minimize} \quad \sum_{i=1}^{n} \left(\frac{y_i - f(\mathbf{x}_i, \boldsymbol{\beta})}{f(\mathbf{x}_i, \boldsymbol{\beta})} \right)^2 = SSE = \sum_{i=1}^{n} e_i^{\,2} \tag{5}$$

However, the MPE solution is biased high because it is derived directly in a single pass, rather than by an iterative process like MUPE. As the hypothetical function $f(\boldsymbol{x},\boldsymbol{\beta})$ appears in both numerator and denominator, this process has a tendency to make $f(\boldsymbol{x},\boldsymbol{\beta})$ higher than it should be. See Hu and Sjovold (1994) for details.

B. MUPE Method. To eliminate this bias, 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,

$$\text{Minimize} \quad \sum_{i=1}^{n} \left(\frac{y_i - f(\mathbf{x}_i, \boldsymbol{\beta}_k)}{f(\mathbf{x}_i, \boldsymbol{\beta}_{k-1})} \right)^2 = \sum_{i=1}^{n} \left(\frac{y_i - f_k(\mathbf{x}_i)}{f_{k-1}(\mathbf{x}_i)} \right)^2 \tag{6}$$

where k is the iteration number and the other terms are as defined previously.

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 6 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 6) 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):

$$SPE = \sqrt{\sum_{i=1}^{n} ((y_i - \hat{y}_i)/\hat{y}_i)^2 / (n-p)} \tag{7}$$

Again, $\hat{y}_i$ is the predicted value in unit space for the i^{th} data point and p is the total number of estimated parameters. Note the SPE measure is the square root of SSE adjusted for its degrees of freedom. The MUPE CER provides consistent estimates of the parameters and has zero

proportional error for all points in the data set. See Hu (2001) or Hu and Sjovold (1994) for detailed descriptions of the MUPE method.

C. ZMPE Method. This is another common method used to reduce the positive proportional error when minimizing Equation 5 directly. Mathematically, it is stated as follows:

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

$$\text{Subject to} \quad \sum_{i=1}^{n} \frac{y_i - f(\mathbf{x}_i, \boldsymbol{\beta})}{f(\mathbf{x}_i, \boldsymbol{\beta})} = 0 \tag{8}$$

This method is a "constrained" minimization process. It is commonly referred to as the zero percentage bias method under MPE, i.e., the ZPB/MPE or ZMPE method (see Book and Lao, 1999). The objectives of the MPE and ZMPE methods are the same except that a constraint is applied to the ZMPE method. Both MUPE and ZMPE CERs have zero percentage error for all the points in the data set, i.e., zero sample bias. Unlike ZMPE, however, MUPE does not include any constraints as part of the regression process (see Equation 6).

ZMPE VS. MUPE

Is ZMPE CER unbiased? As discussed above, both MUPE and ZMPE CERs have zero sample bias; namely, the average percentage error is zero for the data set used to generate the CER:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{\hat{y}_i} = 0 \tag{9}$$

Note that Equation 9 is commonly termed zero percentage bias (ZPB) or zero sample bias. For MUPE CERs, the ZPB condition is achieved through the iterative minimization process. For ZMPE CERs, however, this condition is obtained by using a constraint (see Equation 8).

Here is a common question: does the ZPB property imply that the fitted CER is unbiased? The answer is no—this is **not** necessarily true because the ZPB constraint can be applied to any minimization process. There is no guarantee that the CER result will be unbiased; namely, the condition "$E(\hat{Y}) = f(X,\beta)$" may not be satisfied.

On the other hand, MUPE is the best linear **unbiased** estimator (BLUE) for linear models. For linear CERs, e.g., $Y = (a + bX_1 + cX_2)*\varepsilon$, the MUPE method produces unbiased estimates of the parameters and the function mean; it also provides smaller variances for the parameters and for linear functions of the parameters. (See Draper and Smith, 1981)

Furthermore, the MUPE CER provides consistent estimates of the parameters and the mean of the equation; MUPE's parameter estimators are also the quasi maximum likelihood estimators (QMLE) of the parameters. However, there is **no objective interpretation** of the ZMPE CER—we do not know whether the estimated costs derived from the ZMPE CERs are the mean, median, or mode of the distribution.

Is a smaller SPE better? Should we select a CER solely based upon its SPE value? The answer is **no** for both questions. SPE is the CER's standard error of estimate, which is used

to measure the model's overall error of estimation; it is the one-sigma (percent) spread of the MUPE or ZMPE CER errors. In mathematical terms, we use **SPE^2 to estimate σ^2**, which is the variance of the error term, ε. Based upon many empirical examples, the SPE measure generated by the ZMPE method is always smaller than or equal to the SPE derived by the MUPE method (see Book, 2006):

$$\mathbf{SPE_{(ZMPE)} \leq SPE_{(MUPE)}} \quad (10)$$

Therefore, the estimated variance of the error term derived by the ZMPE method is also smaller than or equal to that derived by the MUPE method.

$$\mathbf{SPE^2_{(ZMPE)} \leq SPE^2_{(MUPE)}} \quad (11)$$

In Equations 10 and 11, the equal sign holds only for simple factor equations. Since ZMPE and MUPE methods generate the same simple factor equations, their SPE measures are also the same. See Hu (June 2010) for details.

Given Equation 10, should we select a ZMPE CER because its SPE measure is smaller than its MUPE counter-part? The answer is no—we should consider all the relevant fit and predictive measures when deriving CERs (Hu, 2008). Although a smaller SPE usually indicates a tighter fit, we should not rely solely on this measure when selecting CERs. If so, we would use the MPE method for CER development, which is proven to be wrong (see Hu 2001 and Hu and Sjovold, 1994).

Is ZMPE's SPE^2 an unbiased estimator of σ^2? The answer is **no**; in fact, ZMPE's SPE measure underestimate the spread of the CER error distribution. We will use a linear model as an illustrative example and will apply an expectation formula to a WLS to prove this conclusion.

The purpose of using a WLS is when some of the observations in the data set are less reliable than others (i.e., the data points are not of the same quality). 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. For a general linear model, we can hypothesize the WLS as follows:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \quad (12)$$

where:

$\mathbf{Y}$ = the vector of the dependent variable
$\mathbf{X}$ = the design matrix
β = the vector of the unknown parameters, i.e., $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$
p = the total number of estimated parameters
ε = the error term with mean of $\mathbf{0}$ and variance $\mathbf{V}\sigma^2$
$\mathbf{V}$ = the variance/covariance matrix without the constant factor σ^2

Note that Equation 12 also applies to a linear MUPE equation because the MUPE technique is also a WLS process.

It can be shown that there exists a nonsingular symmetric matrix $\mathbf{P}$ such that $\mathbf{P'P} = \mathbf{P^2} = \mathbf{V}$. If we multiply both sides of Equation 12 by the inverse matrix of $\mathbf{P}$ (denoted by $\mathbf{P^{-1}}$), a modified model for the new variable $\mathbf{Z}$ is given by

$$\mathbf{Z} = (\mathbf{P}^{-1}\mathbf{X})\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon} = (\mathbf{Q})\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon} \tag{13}$$

where $\mathbf{Z} = \mathbf{P}^{-1}\mathbf{Y}$ (a n-by-1 vector) and $\mathbf{Q} = \mathbf{P}^{-1}\mathbf{X}$ (a n-by-p matrix). The new error term, $\mathbf{P}^{-1}\boldsymbol{\varepsilon}$, for the variable $\mathbf{Z}$ is now distributed with the **same** variance for all observations, i.e., $V(\mathbf{Z}) = V(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) = \mathbf{I}\sigma^2$, where $\mathbf{I}$ is an identity matrix. (Note: variance of $\mathbf{Z} = \text{Var}(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) = \mathbf{P}^{-1}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{P}\mathbf{P}^{-1}\sigma^2 = \mathbf{I}\sigma^2$.) Hence, we can apply **OLS** to derive a solution for Equation 13. We can also claim that the expected value of SSE for Equation 13 is σ^2 times n minus p:

$$E(SSE_{(Z)}) = E(\sum_{i=1}^{n}(z_i - \hat{z}_i)^2) = E(\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}}) = \sigma^2(n - p) \tag{14}$$

where n is the sample size and p is the number of estimated parameters as given above. The proof of Equation 14 is based upon Equation 15 given below:

If the random vector $\mathbf{y}$ has an expected value $\boldsymbol{\mu}$ ($E(\mathbf{y}) = \boldsymbol{\mu}$) and a variance-covariance matrix $\boldsymbol{\Sigma}$, the quadratic form $\mathbf{y}'\mathbf{A}\mathbf{y}$ has the following expected value:

$$E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \text{trace}(\boldsymbol{\Sigma}\mathbf{A}) \tag{15}$$

where $\mathbf{A}$ is a square matrix and trace is the sum of the diagonal elements of a square matrix. Note: Equation 15 holds whether or not the random vector $\mathbf{y}$ is multi-normally distributed. The SSE of the new dependent variable $\mathbf{Z}$ under OLS is given by the quadratic form:

$$\begin{aligned} SSE &= (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}}) = (\mathbf{Z} - \mathbf{Q}\hat{\boldsymbol{\beta}})'(\mathbf{Z} - \mathbf{Q}\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{Z} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Z})'(\mathbf{Z} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Z}) \\ &= (\mathbf{Z} - \mathbf{H}\mathbf{Z})'(\mathbf{Z} - \mathbf{H}\mathbf{Z}) = \mathbf{Z}'(\mathbf{I} - \mathbf{H})\mathbf{Z} \end{aligned} \tag{16}$$

where $\mathbf{H} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ (the hat matrix) and $\hat{\boldsymbol{\beta}}$ is the LS solution for the unknown parameters. Based upon Equation 15, we can then derive the expected value of SSE for the dependent variable $\mathbf{Z}$:

$$\begin{aligned} E(SSE) &= E(\mathbf{Z}'(\mathbf{I} - \mathbf{H})\mathbf{Z}) = E(\mathbf{Z})'(\mathbf{I} - \mathbf{H})E(\mathbf{Z}) + trace[(\mathbf{I} - \mathbf{H})\text{V}(\mathbf{Z})] \\ &= (\mathbf{Q}\boldsymbol{\beta})'(\mathbf{I} - \mathbf{H})(\mathbf{Q}\boldsymbol{\beta}) + trace[(\mathbf{I} - \mathbf{H})\mathbf{I}\sigma^2] \\ &= \boldsymbol{\beta}'\mathbf{Q}'(\mathbf{I} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}')\mathbf{Q}\boldsymbol{\beta} + \sigma^2(n - trace[\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}']) \\ &= 0 + \sigma^2(n - trace[(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Q}]) \\ &= 0 + \sigma^2(n - trace[\mathbf{I}_\mathbf{p}]) = \sigma^2(n - p) \end{aligned} \tag{17}$$

Note that Equation 14 does **not** require the random vector $\mathbf{Z}$ to be multi-normally distributed (Morrison, 1983). We may also use Equation 14 as an approximation for a nonlinear model where the design matrix is approximated by the partial derivatives of the hypothetical CER with respect to each parameter evaluated at the parameter estimates. For a nonlinear model, however, the validity of Equation 14 depends upon whether the linearized form is a good approximation of the true model. (If the normality assumption holds (i.e., $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$), then SSE/σ^2 follows a chi-square distribution with (n – p) degrees of freedom. See Draper and Smith, 1981).

Equation 16 can also be expressed using a quadratic form of Y:

$$
\begin{aligned}
SSE &= (\mathbf{Z}-\mathbf{Q}\hat{\boldsymbol{\beta}})'(\mathbf{Z}-\mathbf{Q}\hat{\boldsymbol{\beta}}) = \mathbf{Z}'(\mathbf{I}-\mathbf{H})\mathbf{Z} \\
&= (\mathbf{P}^{-1}\mathbf{Y}-\mathbf{P}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{P}^{-1}\mathbf{Y}-\mathbf{P}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{P}^{-1}\mathbf{P}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \qquad (18) \\
&= \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}-\mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \mathbf{Y}'(\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{R}\mathbf{V}^{-1})\mathbf{Y}
\end{aligned}
$$

where $\mathbf{R} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'$ is the hat matrix and $\mathbf{V}^{-1}$ (= $\mathbf{P}^{-1}\mathbf{P}^{-1}$) is viewed as the weighting matrix for WLS. Just as Equation 17, we can also verify that the mean of SEE using Equation 18 is also $(n-p)*\sigma^2$. This SSE can be further simplified when all the observations are uncorrelated:

$$SSE = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \sum w_i (y_i - \hat{y}_i)^2 \qquad (19)$$

Note that for MUPE CERs, the diagonal element of $\mathbf{V}^{-1}$ is the reciprocal of the squared predicted value and the rest elements of $\mathbf{V}^{-1}$ are all zeroes.

Based upon Equation 17, we can drive the following formula for linear MUPE CERs:

$$E(SSE/(n-p)) = E(MSE) = E(SPE^2) = \boldsymbol{\sigma^2} \qquad (20)$$

where MSE stands for mean squared error, which is the sum of squares due to error, adjusted by its DF. Given Equations 11 and 20, we conclude that ZMPE's SPE measure is **not** an unbiased estimator of the error term; in fact, it underestimates the standard deviation (σ) of the error distribution.

$$\mathbf{E(SPE^2_{(ZMPE)}) \le E(SPE^2_{(MUPE)}) = \sigma^2} \qquad (21)$$

Just as Equation 11, the equal sign in Equation 21 holds only for simple factor CERs. Therefore, here is an important message for analysts using ZMPE's SPE for uncertainty analysis:

> Using ZMPE CERs in cost uncertainty analysis may unduly **tighten the S-curve** because their SPEs underestimate the CER error distribution.

We use prediction intervals (PI) to model CER errors for cost uncertainty analysis. Since a PI is a function of the standard error of the regression (e.g., SPE), along with other factors, the smaller the standard error, the tighter the PI becomes. Consequently, the impact on the risk session can be substantial when using underestimated ZMPE's SPEs in numerous WBS elements.

Analysts may ask why ZMPE's SPE is biased low. We believe this is because its degrees of freedom did not get adjusted properly. In the section below, we will introduce the concept of GDF to explain why ZMPE CERs underestimate CER errors.

DEFINITION OF GDF

Degrees of Freedom (DF). Statisticians use the term *degrees of freedom* to characterize the number of **independent** pieces of information contained in a statistic. (Note that the term "independent" means they are free to vary.) Any sum of squares, therefore, has a DF associated with it. This number indicates how many pieces of independent information from the n independent observations, Y_1, Y_2, …,, Y_n, are needed to compile the sum of squares. For example, the total sum of squares (SST) needs (n – 1) independent pieces because there is one constraint such that the sum of $(Y_1 - \overline{Y})$, $(Y_2 - \overline{Y})$, …, and $(Y_n - \overline{Y})$ is zero by definition.

For a simple regression model where $Y = \alpha + \beta X + \varepsilon$, we can compute the regression sum of squares (SSR) from one function of $Y_1, Y_2, \ldots, Y_n$; namely, $\hat{\beta}^2$ as $SSR = \hat{\beta}^2 \Sigma_i (x_i - \bar{x})^2$. Therefore, SSR has one DF, which is the number of the independent variables. As for the sum of squares due to error (SSE), it has (n – 2) DF to measure the variability of the fitted line since both the intercept and slope are estimated by the regression equation. (Note: $SSE = \Sigma_i (y_i - \hat{y}_i)^2$ for equations with an additive error term.)

For a regression with multiple independent variables, the DF for SSE is defined to be the number of observations in the equation (sample size) minus the number of parameters estimated from the data. Intuitively, the DF for SSE is the excess points that can be used to judge the quality of the fit. The DF for SSR is the number of independent variables in the model. Table 1 below lists the DFs for the various sums of squares.

Table 1: Degrees of Freedom for Sum of Squares

Source	DF (w/ Intercept)	DF (No Intercept)
SSR	k	k
SSE	n – k – 1	n – k
SST	n – 1	n

where k stands for the number independent variables and n is the sample size.

Generalized Degrees of Freedom (GDF). When constraints are introduced into regression analysis, the number of **independent** pieces of information contained in SSE decreases. In other words, when additional constraints are present, we cannot search as freely as we can in an unconstrained domain to find a solution. Consequently, this results in a loss of DF. Generally, analysts count DF as the DF for the unconstrained process minus the number of constraints included in the process.

To solve the unknowns in regression analysis, each "normal" equation introduced into the system can be viewed as a constraint that restricts one DF. For example, in a simple OLS regression model where $Y = \alpha + \beta X + \varepsilon$, the objective function is given by

$$F = \sum_{i=1}^{n} (y_i - \alpha - \beta * x_i)^2 \tag{22}$$

The least square solutions for the intercept and slope parameters are derived by taking the partial derivatives of F with respect to α and β, respectively, and then setting them to zero:

$$\frac{\partial F}{\partial \alpha} = 0 \quad \rightarrow \quad \sum_{i=1}^{n} (y_i - \alpha - \beta * x_i) = 0 \tag{23}$$

$$\frac{\partial F}{\partial \beta} = 0 \quad \rightarrow \quad \sum_{i=1}^{n} x_i (y_i - \alpha - \beta * x_i) = 0 \tag{24}$$

Since every normal equation (e.g., Equations 23 and 24) can be viewed as a constraint (which results in a loss of a DF), the DF for SSE in this simple example is n – 2 (as we explained in the DF section). If we specify an additional constraint in the minimization process such that the sum of all proportional errors is zero:

$$\sum_{i=1}^{n}(y_i - \alpha - \beta * x_i)/(\alpha + \beta * x_i) = 0 \quad (25)$$

Then the total number of constraints is increased to three from two. Consequently, the resultant DF becomes n – 2 – 1 as we are restricted by this additional constraint when searching for the minimum.

By definition, each constraint introduced into the system will restrict one DF. Therefore, just like each unknown in the regression equation, each constraint specified in the curve-fitting process should be counted as a loss of one DF when calculating the SE or SPE measure.

Furthermore, we should take redundancy into account. For example, if two additional constraints are specified in a regression model but one constraint can be derived from the other, then we should only count a loss of one DF rather than two. Another example, if a constraint is directly related to the normal equations, then it does not count towards a loss of a DF. Based upon this concept, GDF is defined as follows:

GDF = n – p – (Number of Constraints) + (Number of Redundancies) (26)

where n is the sample size and p is the total number of estimated parameters. Note that the constraints in Equation 26 are referring to the equality constraints; the GDF for the inequality constraints is another topic. Based upon Equation 26, GDF can be summarized as:

$$GDF = \begin{cases} n - p & \textit{for MUPE} \\ n - p - 1 & \textit{for ZMPE (except for simple factor CERs)} \end{cases} \quad (27)$$

CALCULATE FIT STATISTICS USING GDF

Why is GDF important and why do we care about reporting the correct DF? The answer is straightforward— the DF measure is essential because all the fit statistics are based upon this number. Therefore, we should calculate the fit statistics (SEE, SPE, Adjusted R^2, and Generalized R^2) using the proper DF, namely, GDF when constraints are specified as part of the regression process. When GDF is applied, the fit measures derived by the ZMPE method should be comparable to those derived by the MUPE method.

Modifying SEE and SPE Using GDF. Both the SEE and SPE measures should be calculated using GDF when constraints are specified in the process:

$$SEE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / GDF} \quad (28)$$

$$SPE = \sqrt{\sum_{i=1}^{n}((y_i - \hat{y}_i)/\hat{y}_i)^2 / GDF} \quad (29)$$

This SPE measure (Equation 29) is identical to Equation 7 for MUPE CERs. However, it is different for ZMPE and should be modified as follows:

$$SPE_{(ZMPE)} = SPE_{(c)}\sqrt{\frac{n-p}{n-p-1}} \quad (30)$$

where $SPE_{(c)}$ stands for the current calculation, not applying GDF. The updated SPE measure for ZMPE is larger than the current value. Once GDF is applied, we can compare MUPE's SPE with ZMPE's SPE using Equation 29. Note that the SPE measure is regarded as the CER's standard error of estimate; it is the one-sigma spread of the MUPE/ZMPE CER.

Modifying Adjusted R^2 using GDF for Additive-Error CERs. *Adjusted R^2* calculated in unit space is used to evaluate the predictive capability of the regression equation for **additive-error** models. Both R^2 and *Adjusted R^2* can be evaluated in either fit space or unit space. "Fit space" refers to the domain where the regression is derived by the optimization technique, while "unit space" denotes the domain of the dependent variable. We recommend updating *Adjusted R^2 in unit space* by GDF when constraints are specified in the curve fitting process:

$$Adj.\,R^2\ in\ unit\ space = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / GDF}{\sum (y_i - \bar{y})^2 / (n-1)} = \frac{MSE_{\bar{Y}} - MSE_f}{MSE_{\bar{Y}}} = \frac{SEE_{\bar{Y}}^2 - SEE_f^2}{SEE_{\bar{Y}}^2} \quad (31)$$

where MSE_f is the CER's mean squared error (MSE) and SEE_f is the CER's standard error. We can use *Adjusted R^2 in unit space* to compare the CER's performance to the starting point, the MSE of an average cost CER (i.e., $MSE_{\bar{Y}}$) when the driver variables are not present. Note that GDF equals n – p when no constraints are specified in the curve-fitting process.

Modifying the Adjusted R^2 using GDF for MUPE and ZMPE CERs. Equation 32 is a modified Adjusted R^2 for MUPE and ZMPE CERs. To be consistent with the error definition and the fitting methodology (see Equations 4 and 6), we recommend using this measure to evaluate the predictive capability of MUPE/ZMPE CERs:

$$Adj.\,R^2\ for\ MUPE/ZMPE = 1 - \frac{\sum ((y_i - \hat{y}_i)/\hat{y}_i)^2 / GDF}{\sum ((y_i - \bar{y})/\bar{y})^2 / (n-1)} = \frac{SPE_{\bar{Y}}{}^2 - SPE_f{}^2}{SPE_{\bar{Y}}{}^2} \quad (32)$$

This modified Adjusted R^2 compares the estimated variances of MUPE or ZMPE with its baseline (i.e., an average CER when driver variables are not available). For example, if the SPE for a MUPE CER is 0.35, we don't really know whether this number is good or bad (or how good it is). However, if *Adjusted R^2 for MUPE* is calculated to be 0.75 (given the sample variance of an average CER is 1.4), then we know the reduction of variance is 75% when applying this CER. In short, *Adjusted R^2 for MUPE/ZMPE* is a relative measure, which puts MUPE and ZMPE's SPE^2 (MSE) in perspective. See Hu (2010) for details.

We first introduced *Adjusted R^2 for MUPE/ZMPE* in 2008 at the SCEA/ISPA Joint Annual Conference. We now suggest using GDF in the computation as an additional adjustment for ZMPE. Since Equation 32 takes constraints into account, we can use it to compare the *Adjusted R^2* measure between MUPE and ZMPE CERs. We can also use Equation 32 to compare across different MUPE or ZMPE CERs, but not to compare across different methods, e.g., a MUPE CER versus an additive error CER. As noted above, GDF equals n – p when no constraints are specified in the curve-fitting process.

Modifying the Pearson's Correlation Coefficient (r) for Small Samples. Pearson's correlation coefficient (Pearson's r) measures the linear association between two sets of numbers $\{x_i\}$ and $\{y_i\}$ and is defined as follows:

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (33)$$

where $\bar{x}$ and $\bar{y}$ are the means of $\{x_i\}$ and $\{y_i\}$, respectively, and n is the sample size. Pearson's r measures the degree to which two sets of data move together in a linear manner. It ranges between -1.0 to +1.0, where -1 indicates a perfect inverse relationship, 0 indicates no correlation, and +1 indicates a perfect positive relationship.

We should consider the sample size when applying correlations to our cost uncertainty models. For example, a correlation of 0.8 derived from 30 data points is much more significant than the same correlation computed from just five data points. Unfortunately, the sample size adjustment is not accounted for in the definition of Pearson's correlation coefficient. The following formula relates the adjusted R^2 to R^2:

$$\text{Adjusted } R^2 = R^2 - (1 - R^2)*(p - 1)/(n - p) = 1 - (1 - R^2)*(n - 1)/(n - p) \quad (34)$$

where p is the total number of estimated coefficients and n is the sample size.

As inspired by Equation 34, we now establish a heuristic *Pearson's r squared* (denoted by $\bar{r}^2$) to correct for small samples, assuming p = 2 for a one-independent model:

$$\bar{r}^2 = r^2 - (1 - r^2)/(n - 2) = 1 - (1 - r^2)*(n - 1)/(n - 2) \quad (35)$$

Pearson's adjusted correlation coefficient (Pearson's adjusted r) is then given by

$$Pearson's\ adjusted\ r = \begin{cases} \text{sign}(r)*\bar{r} & if\ \bar{r}^2 > 0 \\ 0 & if\ \bar{r}^2 \le 0 \end{cases} \quad (36)$$

where r is used to denote Pearson's correlation coefficient.

Instead of using Equation 33, we suggest using *Pearson's adjusted r* (Equation 36) for correlation analysis, such as analyzing cost versus cost or residual versus residual correlation between two separate CERs. Here, the value of p is assumed to be **two** as we do **not** use Equation 36 to analyze correlation between the actual observation (y) and its CER's predicted value ($\hat{y}$); namely, Equation 36 is not used within a CER. When evaluating a CER, use GRSQ (see below) to measure the linear association between a CER's actual value and its predicted value. We can also take the square root of GRSQ (like Equation 36) to calculate *Pearson's adjusted r* when evaluating CERs.

Definition of GRSQ (r^2). Generalized R-squared (GRSQ, also denoted by the symbol r^2) is commonly used for evaluating the quality of a nonlinear CER. By definition, GRSQ is the square of Pearson's correlation coefficient between the actual observations, y_is and CER predicted values, $\hat{y}_i$s (see Young 1992):

$$GRSQ = r^2(y,\hat{y}) = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})\left(\hat{y}_i - \bar{\hat{y}}\right)\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2} \quad (37)$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the averages of the actual (y) and CER-predicted values ($\hat{y}$), respectively.

Modifying GRSQ (r^2) Using GDF for Correcting Degrees of Freedom. As shown by the definition above, GRSQ does not take the DF or even the sample size into consideration. For example, a GRSQ of 0.8 derived from 30 observations should be more significant than the same GRSQ simply based upon five data points. This shortcoming is the same as we noted above for the Pearson's correlation coefficient. We introduced a modified GRSQ to correct for DF in 2008; we now suggest making further adjustment for ZMPE by including GDF in the computation:

$$GRSQ_{(GDF)} = \begin{cases} r^2 - (1-r^2) * \dfrac{p-1}{GDF} & if \;\; p > 1 \\ r^2 - (1-r^2) * \dfrac{1}{n-1} & if \;\; p = 1 \end{cases} \quad (38)$$

where $r^2 = r^2(y,\hat{y})$ = GRSQ; n and p are defined above. Again, GDF = n – p if no constraints are specified in the curve-fitting process.

When dealing with a simple factor equation, e.g., p = 1, the term "(p – 1)/GDF" in Equation 38 vanishes. However, the sample size should still be accounted for when it happens, so a modified term "1/(n – 1)" is used in this case. See Hu (2010) for details.

Note that Equation 38 only adjusts DF; GRSQ (as well as *Adjusted Pearson's r*) is still insensitive to different fitting methods and equation forms because it only measures the linear association between two sets of numbers, not the actual deviations between them (see Hu, 2010 for details).

EXAMPLE SECTION

In this section, we use examples to demonstrate that the SPE measures for the ZMPE equations are no longer smaller than their MUPE counter-part when GDF is used in the computation of SPE.

Weight-Based CER. Listed below is a hypothetical data set where the weight variable is used to predict the cost of a black box:

Table 2: **Cost vs. Weight Database**

Data Point	Cost $K	Weight (lbs)
Obs 1	135.0	4.18
Obs 2	6.5	0.32
Obs 3	8.0	0.57
Obs 4	64.6	2.34
Obs 5	32.9	0.50

Data Point	Cost $K	Weight (lbs)
Obs 6	95.4	2.70
Obs 7	67.0	4.54
Obs 8	112.2	4.42
Obs 9	29.2	0.55
Obs 10	22.7	0.20
Obs 11	16.9	0.80
Obs 12	35.0	1.75

Four different CERs are generated using both the MUPE and ZMPE methods. See Tables 3 and 4 for the fitted equations and their corresponding fit and predictive measures:

Table 3: ZMPE CERs and Statistics

ZMPE	CERs	SPE	$SPE_{(GDF)}$	$\underline{R}^2$	$\underline{R}^2_{(GDF)}$	GRSQ	$GRSQ_{(GDF)}$
Linear	12.794 + 19.16*Weight	45.7%	48.2%	68.9%	65.4%	77.5%	75.0%
LogLinear	36.889*Weight^(0.5882)	48.9%	51.5%	66.9%	60.4%	77.4%	74.9%
Semi-Log	17.881*(1.5314)^Weight	47.1%	49.7%	64.4%	63.2%	68.9%	65.4%
Triad	16.785+12.034*Weight^(1.349)	47.5%	50.4%	66.4%	62.2%	75.5%	69.4%

Note: $\underline{R}^2$ is used to denote the adjusted R^2 for MUPE and ZMPE CERs.

Table 4: MUPE CERs and Statistics

MUPE	CERs	SPE	$\underline{R}^2$	GRSQ
Linear	10.528 + 21.2975*Weight	46.5%	67.8%	77.5%
LogLinear	36.2953*Weight^(0.6635)	50.2%	65.5%	77.7%
Semi-Log	16.756*(1.5835)^Weight	47.4%	66.5%	67.0%
Triad	15.026+12.9863*Weight^(1.386)	48.4%	65.1%	75.3%

As shown by Tables 3 and 4, the unadjusted SPE measures for these ZMPE CERs are all smaller than their respective SPE measures generated by the MUPE method. However, ZMPE's updated SPE measures using GDF become larger than their MUPE counter-part (see the numbers in red). Furthermore, ZMPE's updated adjusted R^2 and GRSQ measures are all smaller than those generated by the MUPE method when GDF is applied in the computation (see the numbers in purple and green). Consequently, MUPE outperforms ZMPE based upon all three statistics.

Note that the semi-log equation has the worst GRSQ among the four CERs. Although the semi-log equation is tighter than the log-linear and triad equations, its GRSQ is much worse than those of the log-linear and triad CERs (see Tables 3 and 4). This example demonstrates that we cannot use GRSQ alone to select a best CER because GRSQ only measures the linear association between the actual observations and the predicted values, not the difference between them (Hu, 2010). Here is a cautionary note for semi-log CERs: despite the fit measures, use the semi-log equation with caution as it goes up exponentially beyond the data range.

Graph 1 is a pictorial representation of the four MUPE CERs listed in Table 4. Based upon Table 4 and Graph 1, the linear CER is the best choice for predicting the cost of this black box. Both the intercept and slope coefficient of the linear CER are also significant. For details, see CO$TAT's regression reports of the fit measures in Table 5.

Graph 1: MUPE CERs

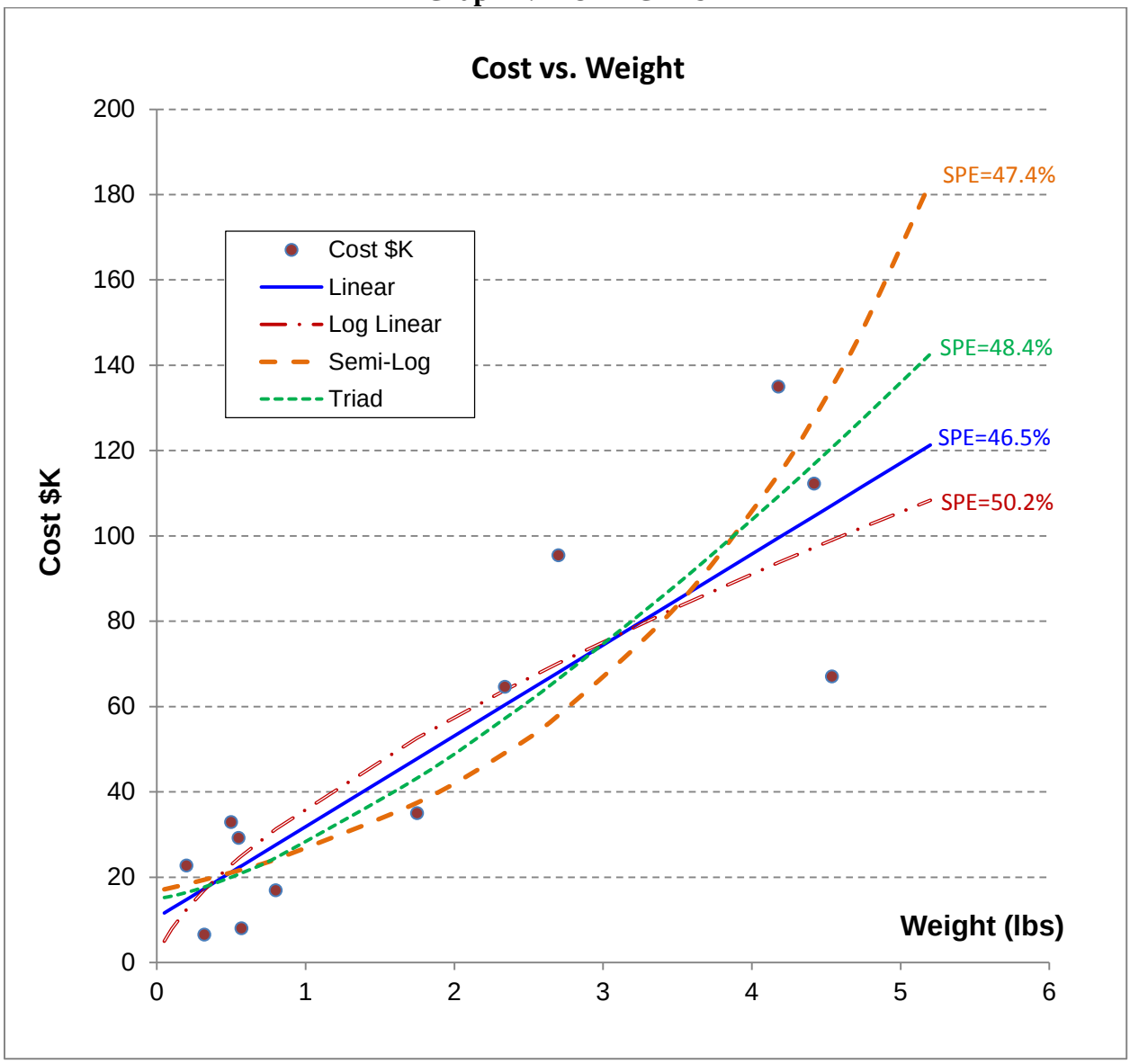

Table 5: CO$TAT Fit Measure Reports for the Linear CER: Y = 10.528 + 21.2975*Weight

Coefficients Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	10.5281	4.7648		2.2096	0.0515	0.949
X	21.2975	5.4030	0.7800	3.9418	0.0028	0.997

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.4649	60.84%	56.93%	0.7800

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	P-Value	Prob Not Zero
Regression	1	3.3579	3.3579	15.5379	0.0028	0.9972
Residual (Error)	10	2.1611	0.2161			
Total	11	5.5190				

CONCLUSIONS

Make sure the constraints (if any) are meaningful, logical, and statistically sound when adding them to the curve fitting process. Although Solver can easily handle constraints when deriving CERs, we should not use the constraint feature excessively. Make sure every constraint is required and satisfies engineers' reasoning. Also, explore different starting points when using Solver. Since Solver can be easily trapped in local minima, especially when fitting complicated equations or ZMPE equations, we suggest exploring different starting points to ensure the solution stabilizes.

Adjust DF for constraint-driven equations (CERs and PERs). When constraints are specified in regression analysis, the number of independent pieces of information contained in SSE decreases. In other words, when additional constraints are present, we cannot search as freely as we can in an unconstrained domain to find a solution. We should adjust DF accordingly to capture the impact.

GDF = n – p – (# of Constraints) + (# of Redundancies), where p denotes the total number of estimated parameters (coefficients) in the equation and n is the sample size. We should take both the constraints and redundancies into account when counting GDF. One restriction is equivalent to a loss of one DF. However, if two constraints are specified in a regression model but one constraint can be derived from the other, then we should only count a loss of one DF, rather than two. Additionally, if a parameter is known, e.g., the startup cost is known or the rate slope is given, then this amounts to a gain of one DF.

Adjust DF for ZMPE CERs. DF should be subtracted by one for ZMPE CERs except for simple factor CERs. This is because the solution for this case is achieved using the constraint alone.

Do not adjust DF and goodness-of-fit measures for MUPE CERs. Since no constraints are specified when deriving MUPE CERs, we do not need to adjust DF and goodness-of-fit measures for MUPE equations.

Calculate the fit measures (SEE, SPE, Adjusted R2, and GRSQ) using GDF. All the fit statistics are driven by the DF measure. Therefore, calculate the fit statistics (e.g., SEE, SPE, Adjusted R^2 and Generalized R^2) using the proper measure of DF, namely, GDF when constraints are specified in the process. Using GDF, the ZMPE fit statistics should be comparable with those derived by the MUPE equation.

ZMPE's standard error underestimates the spread of the CER error distribution. We applied an expectation formula (Morrison, 1983) to prove the above conclusion. Since the SPE measure is the basis for cost uncertainty analysis, we should use an unbiased estimator for the error distribution. Using ZMPE's SPE without adjustment in cost uncertainty analysis may unduly tighten the S-curve.

FUTURE STUDY ITEMS FOR CONSTRAINT DRIVEN REGRESSION

Should we worry about specifying excessive constraints in a regression model? Adding excessive constraints into the curve-fitting process may cause the unknown parameters in the CER to be determined completely by the constraints. When this happens, there is **no** need to run regression analysis. How do we define the DF properly for this extreme case? In other words, if the number of constraints is equal to the number of estimated parameters, we do not use any curve-fitting methods to derive a solution. Consequently, we may have no degrees of freedom left to judge the quality of the fit due to lack of regression.

Below is a simple example using the constraint alone to generate a solution. Let us analyze a simple factor equation with a multiplicative error term:

$$Y_i = \beta * X_i * \varepsilon_i \qquad \text{for } i = 1, \ldots, n \tag{39}$$

where n is the sample size and ε_i is the error term (with a mean of 1 and a variance σ^2).

It can be shown that both the MUPE and ZMPE solutions are the same:

$$\hat{\beta}_{(ZMPE)} = \hat{\beta}_{(MUPE)} = \frac{\sum_{i=1}^{n} (Y_i / X_i)}{n} \tag{40}$$

Note that the ZMPE solution in this case is uniquely determined by the constraint instead of the minimization process (see Hu 2010):

$$\sum_{i=1}^{n} \frac{Y_i - \beta X_i}{\beta X_i} = \left(\sum_{i=1}^{n} \frac{Y_i}{\beta X_i} \right) - n = 0 \tag{41}$$

Also, an inequality constraint may not be treated the same as an equality constraint. How to treat inequality constraints in the curve-fitting or distribution-finding process is another topic.

REFERENCES

1. Book, S., "IRLS/MUPE CERs are not MPE-ZPB CERs," 28th Annual ISPA International Conference, Seattle, WA, 23-26 May 2006.
2. Book, S., "Significant Reasons to Eschew Log-Log OLS Regression when Deriving Estimating Relationships," 2012 ISPA/SCEA Joint Annual Conference, Orlando, FL, 26-29 June.
3. 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.
4. Cochran, W. G., "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance, Proceedings of the Cambridge Philosophical Society, vol. 30, 1934, pages 178-191.

5. Draper, N. R. and H. Smith, "Applied Regression Analysis (2nd edition)", New York: John Wiley & Sons, Inc., 1981.
6. 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.
7. Hu, S., "R^2 versus r^2," Journal of Cost Analysis and Parametrics, Volume 3, Issue 2, Fall 2010.
8. Hu, S., "Simple Mean, Weighted Mean, or Geometric Mean?" 2010 ISPA/SCEA Joint Annual Conference, San Diego, CA, 8-11 June 2010.
9. Hu, S. and A. R. Sjovold, "Multiplicative Error Regression Techniques," 62nd MORS Symposium, Colorado Springs, Colorado, 7-9 June 1994.
10. 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.
11. Morrison, D. F., "Applied Linear Statistical Methods", New Jersey: Prentice-Hall, Inc. 1983.
12. Searle, S. R. "Linear Models," New York: John Wiley & Sons, Inc., 1981.
13. Seber, G. A. F. and C. J. Wild, "Nonlinear Regression," New York: John Wiley & Sons (1989), pages 37, 46, 86-88.
14. Weisberg, S., "Applied Linear Regression, 2nd Edition," New York: John Wiley & Sons, 1985, pages 87-88.
15. 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.
16. Young, P. H., "GERM: Generalized Error Regression Model," 25th Annual DoD Cost Analysis Symposium, Leesburg, VA, September 1991.
17. Young, P. H., "Generalized Coefficient of Determination," 26th Annual DoD Cost Analysis Symposium, Leesburg, VA, September 1992.