TECOLOTE RESEARCH, INC.
Bridging Engineering and Economics
Since 1973

R^2 vs. r^2

2008 Joint SCEA/ISPA Conference

Dr. Shu-Ping Hu

24 - 27 June 2008

■ Los Angeles ■ Washington, D.C. ■ Boston ■ Chantilly ■ Huntsville ■ Dayton ■ Santa Barbara
■ Albuquerque ■ Colorado Springs ■ Columbus ■ Ft. Meade ■ Ft. Monmouth ■ Montgomery ■ Ogden ■ Patuxent River ■ Pensacola ■ San Diego
■ Charleston ■ Cleveland ■ Denver ■ New Orleans ■ Oklahoma City ■ Silver Spring ■ Warner Robins AFB ■ Vandenberg AFB

Presented at the 2008 SCEA-ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

TECOLOTE RESEARCH, INC.

Objectives

- Address the concerns/criticisms about Adjusted R^2 and Pearson's r^2 (GRSQ)
- Analyze GRSQ and provide a statistical rationale to explain the behaviors of GRSQ
- Propose modified Adjusted R^2 (for MUPE CERs) and GRSQ (to correct for degrees of freedom)

Outline

- **Multiplicative Error Models**
- **Definitions of GRSQ (r^2), R^2, Adj. R^2, and SPE**
  - Interpretations of Adj. R^2 and SPE
- **Properties of R^2/Adj. R^2 and GRSQ (r^2)**
- **Concerns about GRSQ (r^2) and R^2/Adj. R^2**
- **Analyze R^2/Adj. R^2 and GRSQ Using Examples**
  - GRSQ insensitive to different fitting methods and CER forms
  - Why is GRSQ insensitive?
- **Propose Modified Adj. R^2 (for MUPE) and GRSQ (for DF)**
- **Conclusions**

Multiplicative Error Models – Log-Error, MUPE, & ZMPE

Definition of error term for $Y = f(x)*\varepsilon$

- **Log-Error: $\varepsilon \sim LN(0, \sigma^2)$ ⇨ Least squares in log space**
  - Error = Log (Y) - Log f(X)
  - Minimize the sum of squared errors; process is done in log space
- **MUPE: $E(\varepsilon) = 1$, $V(\varepsilon) = \sigma^2$ ⇨ Least squares in weighted space**
  - Error = (Y-f(X))/f(X)
  - Minimize the sum of squared (percentage) errors iteratively

Note: $E((Y-f(X))/f(X)) = 0$

$V((Y-f(X))/f(X)) = \sigma^2$

- **ZMPE: $E(\varepsilon) = 1$, $V(\varepsilon) = \sigma^2$ ⇨ Least squares in weighted space**
  - Error = (Y-f(X))/f(X)
  - Minimize the sum of squared (percentage) errors with a constraint

$\Sigma_i (Error_i) = 0$

- **(MPE: Same as ZMPE but with no constraint)**

Outline

- **Multiplicative Error Models**
- ***Definitions of GRSQ (r^2), R^2, Adj. R^2, and SPE***
  - Interpretations of Adj. R^2 and SPE
- **Properties of R^2/Adj. R^2 and GRSQ (r^2)**
- **Concerns about GRSQ (r^2) and R^2/Adj. R^2**
- **Analyze R^2/Adj. R^2 and GRSQ Using Examples**
  - GRSQ insensitive to different fitting methods and CER forms
  - Why is GRSQ insensitive?
- **Propose Modified Adj. R^2 (for MUPE) and GRSQ (for DF)**
- **Conclusions**

Definitions – Pearson's r and GRSQ

- Pearson's correlation coefficient between two sets of numbers $\{x_i\}$ and $\{y_i\}$:

$$r_{xy} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

- Generalized R^2 (GRSQ): GRSQ is Pearson's r^2 between the actual $\{y_i\}$ and predicted $\{\hat{y}_i\}$ in unit space, i.e., GRSQ = $r^2(y, \hat{y})$

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

- GRSQ was first listed in CO$TAT's predictive measures circa 1991

TECOLOTE RESEARCH, INC.

Definition – R^2

- **(1) R^2 = (SST – SSE) / SST:**

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

- **(2) R^2 = SSR / SST:**

$$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

- **Definition 1 (not 2) is used to compute R^2 for nonlinear CERs**
  - Definitions 1 and 2 are not the same except for OLS
  - The fitted equation does not necessarily go through the mean
  - The cross-product term $(\Sigma (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}))$ is not necessarily equal to zero
- **R^2 (by Definition 1) is well-defined and applicable**
  - The traditional SSR (Regression Sum of Squares) is not used in the R^2 definition for nonlinear CERs
  - We do not use R^2 (as given in definition 1) to indicate the proportion of the explained variation for cases other than OLS
- **In OLS: $R^2 = r^2(y, \hat{y})$ = GRSQ** (true for OLS only)

TECOLOTE RESEARCH, INC.

Definition – Adj. R^2

- **Adjusted R^2 in unit space:**

$$Adj.\,R^2 = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{\sum (y_i - \hat{y}_i)^2/(n-p)}{\sum (y_i - \bar{y})^2/(n-1)}$$

(n = sample size and p = total number of estimated coefficients)

Note: Adj. R^2 can be evaluated in both the fit and unit spaces

- ***Adjusted R^2 in unit space* translates SSE from the absolute scale to the relative scale by**
  - Comparing SSE to SST
  - Adjusting degrees of freedom for small samples
- **Adj. $R^2 = 1 - (1 - R^2)*(n-1)/(n-p)$**

  $= R^2 - (1 - R^2)*(p-1)/(n-p)$
- **Note: We compare SSE to SST because $\bar{y}$ is the unbiased estimate for the univariate models ($y = a + \varepsilon$ or $y = a*\varepsilon$)**

Interpretation of Adj. R^2

- **What is Adjusted R^2 in unit space?**

$$Adj.\,R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2/(n-p)}{\sum (y_i - \bar{y})^2/(n-1)} = \frac{MSE_{\bar{Y}} - MSE_f}{MSE_{\bar{Y}}}$$

for additive models

- **This statistic is well-defined and applicable. It measures the percent difference between the CER's estimated variance and the sample variance of Y**
  - For example, if a CER's estimated variance is 0.1 while the sample variance of y is 0.5, then the CER's variance is only 20% of the sample variance. This reduction of variance, 80%, is the Adjusted R^2.
  - The reduction of variance is considered to be an "improvement" when applying the CER
- **We can use *Adjusted R^2 in unit space* to compare a CER's performance to the starting point, i.e., MSE of an average CER (when the driver variables are not available)**

TECOLOTE RESEARCH, INC.

Definition – SPE

- **Standard Percent Error (SPE) or Multiplicative Error:**

$$\text{SPE} = \text{SEE} = \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2}$$

(n = sample size and p = total number of estimated coefficients)

Note: SPE is CER's standard error of estimate (SEE)

- **SPE is used to measure the model's overall error of estimation; it is the one-sigma spread of the MUPE CER**
- **SPE is based upon the objective function; the smaller the value of SPE, the tighter the fit becomes**

Interpretation of SPE

$$SPE = \sqrt{\sum_{i=1}^{n}((Y_i - \hat{Y}_i)/\hat{Y}_i)^2 /(n-p)}$$

$Y = f(x)^*\varepsilon$, $\varepsilon \sim$ Distrn$(1, \sigma^2)$

- **SPE^2 is an estimate of σ^2**
- **SPE cannot be used to determine the significance of the regressed coefficients**
  - For example, if $f(X) = aX^b$, SPE is not used to test whether the regressed coefficient b is significant
  - Compare two regression models, where the exponent of one equation is 10 times larger than the other:

    $Y = a_1X_1^{0.5}\varepsilon_1$ vs. $Y = a_2X_2^{0.05}\varepsilon_2$ ($\varepsilon_1 \sim$ Distrn$(1,\sigma_1^2)$, $\varepsilon_2 \sim$ Distrn$(1,\sigma_2^2)$)

    Assume also that σ_2 in the second equation is much tighter than σ_1 (i.e., $SPE_1 >> SPE_2$). However, the exponent coefficient in the first CER should be more significant than the second one.
- **Beware of using SPE alone for selecting CERs**

Outline

- **Multiplicative Error Models**
- **Definitions of GRSQ (r^2), R^2, Adj. R^2, and SPE**
  - Interpretations of Adj. R^2 and SPE
- ***Properties of R^2/Adj. R^2 and GRSQ (r^2)***
- **Concerns about GRSQ (r^2) and R^2/Adj. R^2**
- **Analyze R^2/Adj. R^2 and GRSQ Using Examples**
  - GRSQ insensitive to different fitting methods and CER forms
  - Why is GRSQ insensitive?
- **Propose Modified Adj. R^2 (for MUPE) and GRSQ (for DF)**
- **Conclusions**

Properties of R^2 and r^2

- **R^2/Adj. R^2 (in unit space) and GRSQ (r^2) are <u>predictive</u> measures**
  - They are <u>not</u> used to evaluate the significance of the regressed coefficient
- **They can be easily influenced by outliers (i.e., leverage points)**
- **R^2, as well as Adj. R^2, measures how well the estimates <u>match</u> the database actuals**
  - The closer the R^2 measure (or Adjusted R^2) is to one, the closer the estimates match the actual observations
- **GRSQ (r^2) measures the <u>linear association</u> between the estimates and actuals, not how well the estimates track to the actual observations if the fit is not an OLS**
  - GRSQ is <u>not</u> an analog of "coefficient of determination" for non-OLS CERs

Outline

- **Multiplicative Error Models**
- **Definitions of GRSQ (r^2), R^2, Adj. R^2, and SPE**
  - Interpretations of Adj. R^2 and SPE
- **Properties of R^2/Adj. R^2 and GRSQ (r^2)**
- ***Concerns about GRSQ (r^2) and R^2/Adj. R^2***
- **Analyze R^2/Adj. R^2 and GRSQ Using Examples**
  - GRSQ insensitive to different fitting methods and CER forms
  - Why is GRSQ insensitive?
- **Propose Modified Adj. R^2 (for MUPE) and GRSQ (for DF)**
- **Conclusions**

Concerns about GRSQ (r^2)

- **Too much credence and emphasis have been placed on GRSQ (r^2) over the years**
  - Besides SPE, GRSQ has become the **only** statistical measure reported for non-OLS CERs
  - Many believe that "appropriateness of shape of a CER (whether it be linear or one of the various nonlinear forms such as Triad or power form) is measured by Pearson's r^2." ➔ **Not** true!
- **Many analysts mistakenly take GRSQ (r^2) for R^2**
  - Does a 90% GRSQ imply the CER has explained 90% of the variation in the data set if it is not an OLS? The answer is no, but many believe it is so.
  - As shown in USCM7, "Pearson's r^2 is the R^2 value that measures how well estimates match the database actuals to which they correspond." In fact, the above quoted statement is incorrect, namely, $r^2 \neq R^2$ if not an OLS.
- **Suggest using the symbol r^2 (not R^2) for GRSQ to avoid possible confusion**
  - Pearson's correlation coefficient is commonly denoted by the letter r (not R)

TECOLOTE RESEARCH, INC.

Concerns about R^2 & Adj R^2

- ***R^2*, as well as *Adjusted R^2*, has no value as a metric in cases other than OLS**
- **The formulas of *R^2* and *Adjusted R^2* are inapplicable**
- **Many good CERs may be dismissed when using *Adjusted R^2* because they might have a negative *Adjusted R^2***
  - As noted in the USCM8 document, a negative Adj. R^2 is a warning flag
  - This warning flag has probably led to the rejection of a number of **good** USCM8 CERs

TECOLOTE RESEARCH, INC.

No Worries about Adj R^2

- **No <u>single</u> measure is relied on to select the best CER**
  - <u>Not</u> possible to reject “logical” CERs just because of negative Adj. R^2
- **Several fit, as well as predictive, measures will be examined for USCM9; they were also examined for USCM7 & USCM8**
  - SPE, Approx T-stats, Pearson’s r, Adjusted R^2, MAD of % Errors, etc.

	Red	Yellow	Green
SPE (Multiplicative Error)	> 0.5	0.25 ~ 0.5	< 0.25
T-Stats (absolute value)	< 1.5	1.5 ~ 2.3	> 2.3
Pearson's r	< 0.6	0.6 ~ 0.8	> 0.8
Adj. R^2	< 0.45	0.45 ~ 0.65	> 0.65
MAD of % Errors	> 50%	25% ~ 50%	< 25%
RMS OF % Errors	> 60%	30% ~ 60%	< 30%

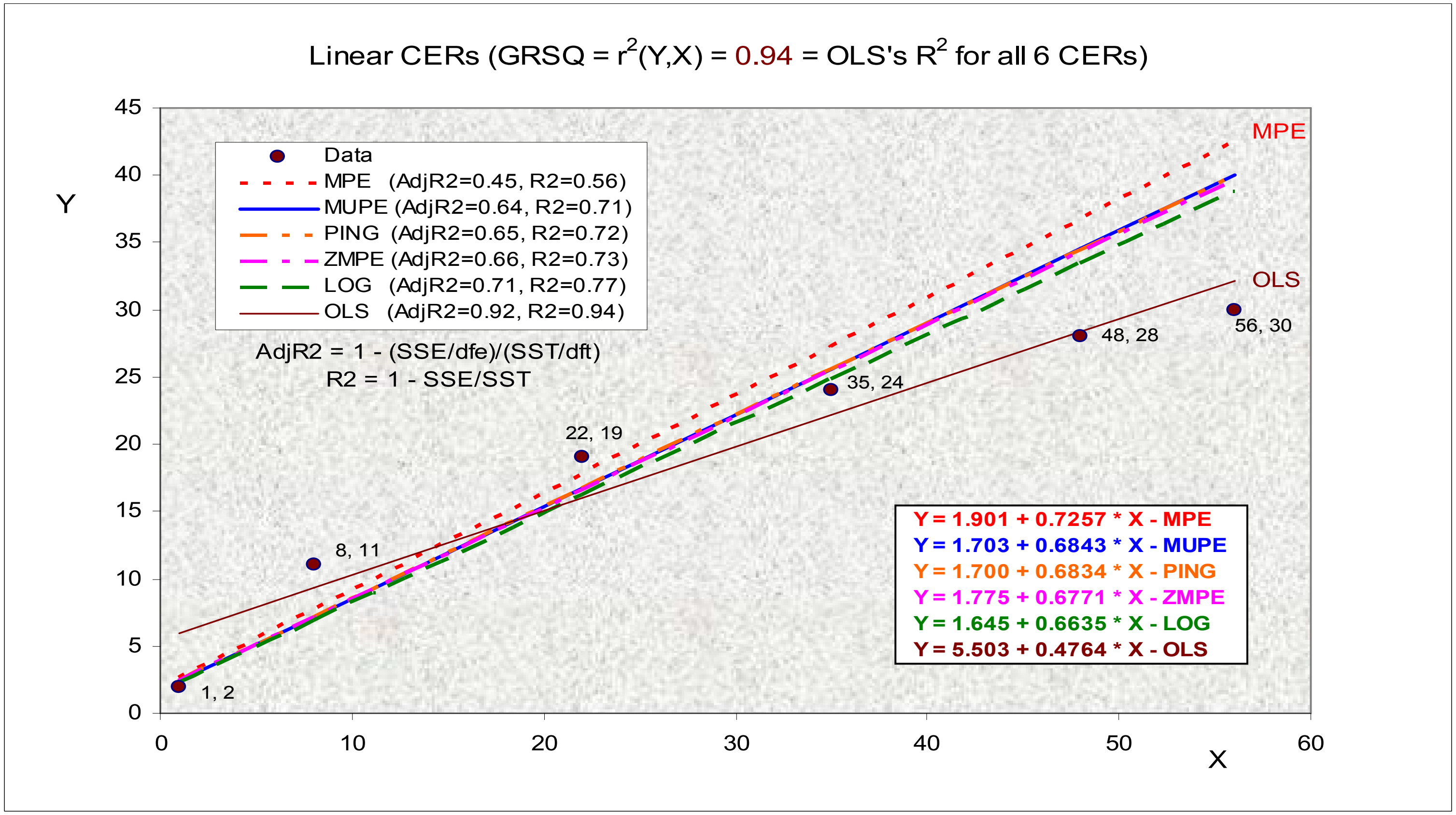
  - The first two are the fit measures (t-stats evaluate the significance of the coefficients); the remainder are predictive measures
- **Neither R^2 nor Adj. R^2 was used to indicate the proportion of the variation explained by the MUPE CER**
- **Counter examples (good SPE with negative Adj. R^2) can be found**

Outline

- **Multiplicative Error Models**
- **Definitions of GRSQ (r^2), R^2, Adj. R^2, and SPE**
  - Interpretations of Adj. R^2 and SPE
- **Properties of R^2/Adj. R^2 and GRSQ (r^2)**
- **Concerns about GRSQ (r^2) and R^2/Adj. R^2**
- ***Analyze R^2/Adj. R^2 and GRSQ Using Examples***
  - GRSQ insensitive to different fitting methods and CER forms
  - Why is GRSQ insensitive?
- **Propose Modified Adj. R^2 (for MUPE) and GRSQ (for DF)**
- **Conclusions**

r^2 Insensitive to Different Fitting Methods (1/5)

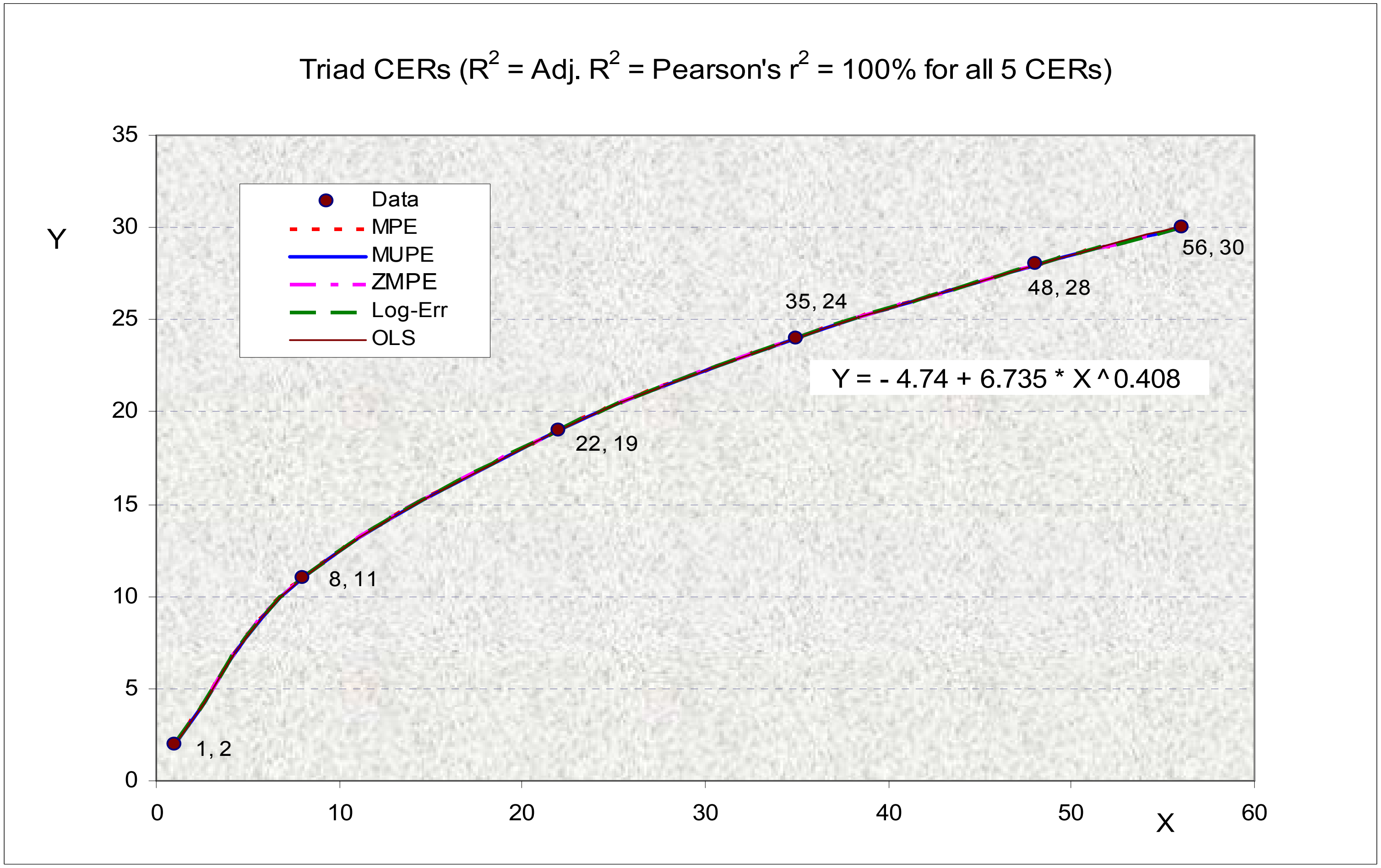
- This data set is obtained from Figure 3 of Reference 1. MPE, MUPE, ZMPE, Log-Error, and OLS CERs were fitted using this data set; the PING Factor was also applied to the log-error CER as an excursion.
- R^2/Adj. R^2 vary in these CERs while GRSQ (i.e., r^2) is the same for all six of them



1. Book, S. and Young, P.H., "The Trouble with R2," ISPA, Journal of Parametrics, Vol. XXV, No 1 (Summer 2006)

r^2 Insensitive to Different Fitting Methods (2/5)

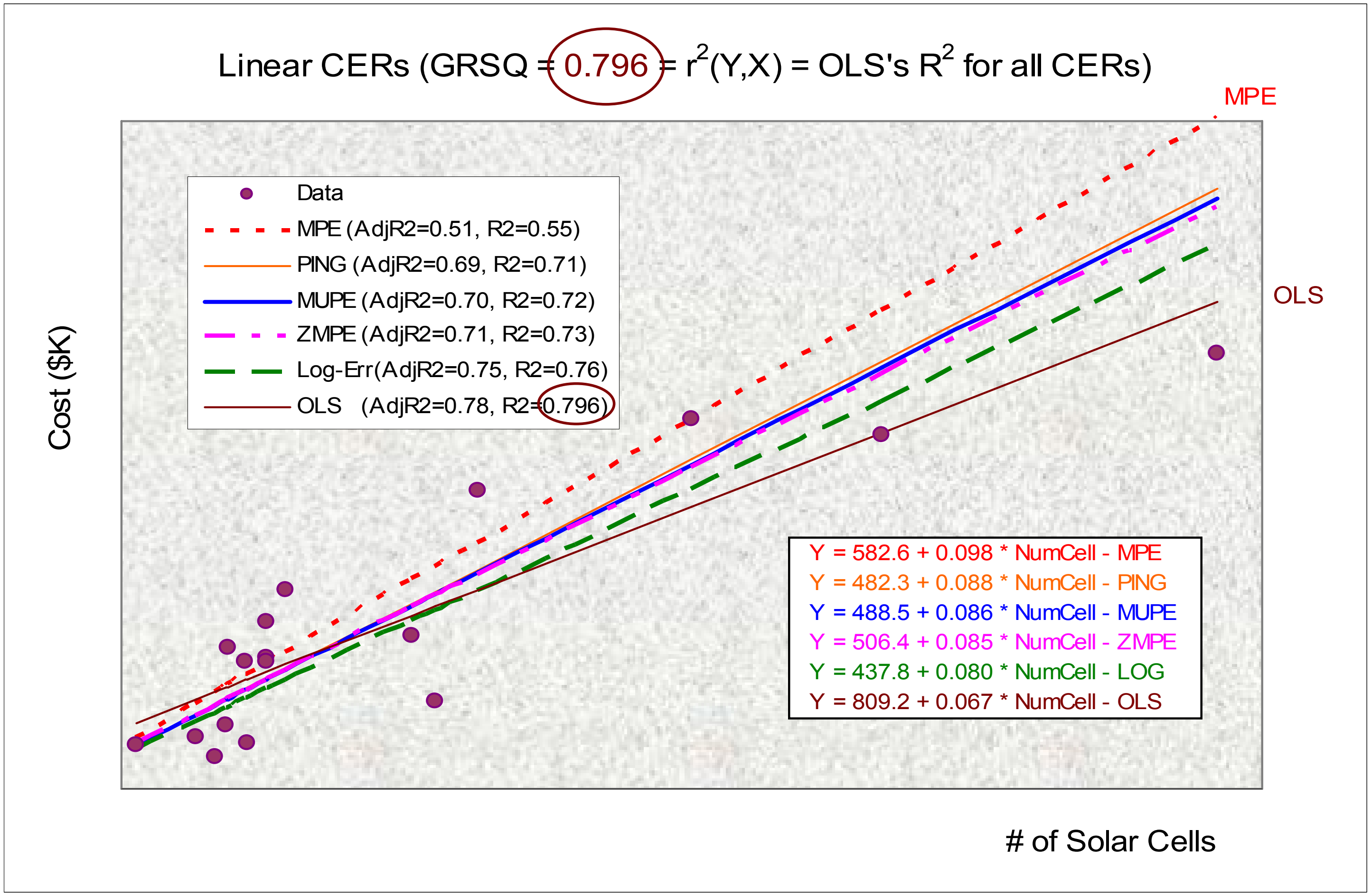
- Note that the TRIAD equation form fits this data set almost perfectly—all five CERs become one: Y = - 4.74 + 6.735*X $^{0.408}$



1. Book, S. and Young, P.H., "The Trouble with R2," ISPA, Journal of Parametrics, Vol. XXV, No 1 (Summer 2006)

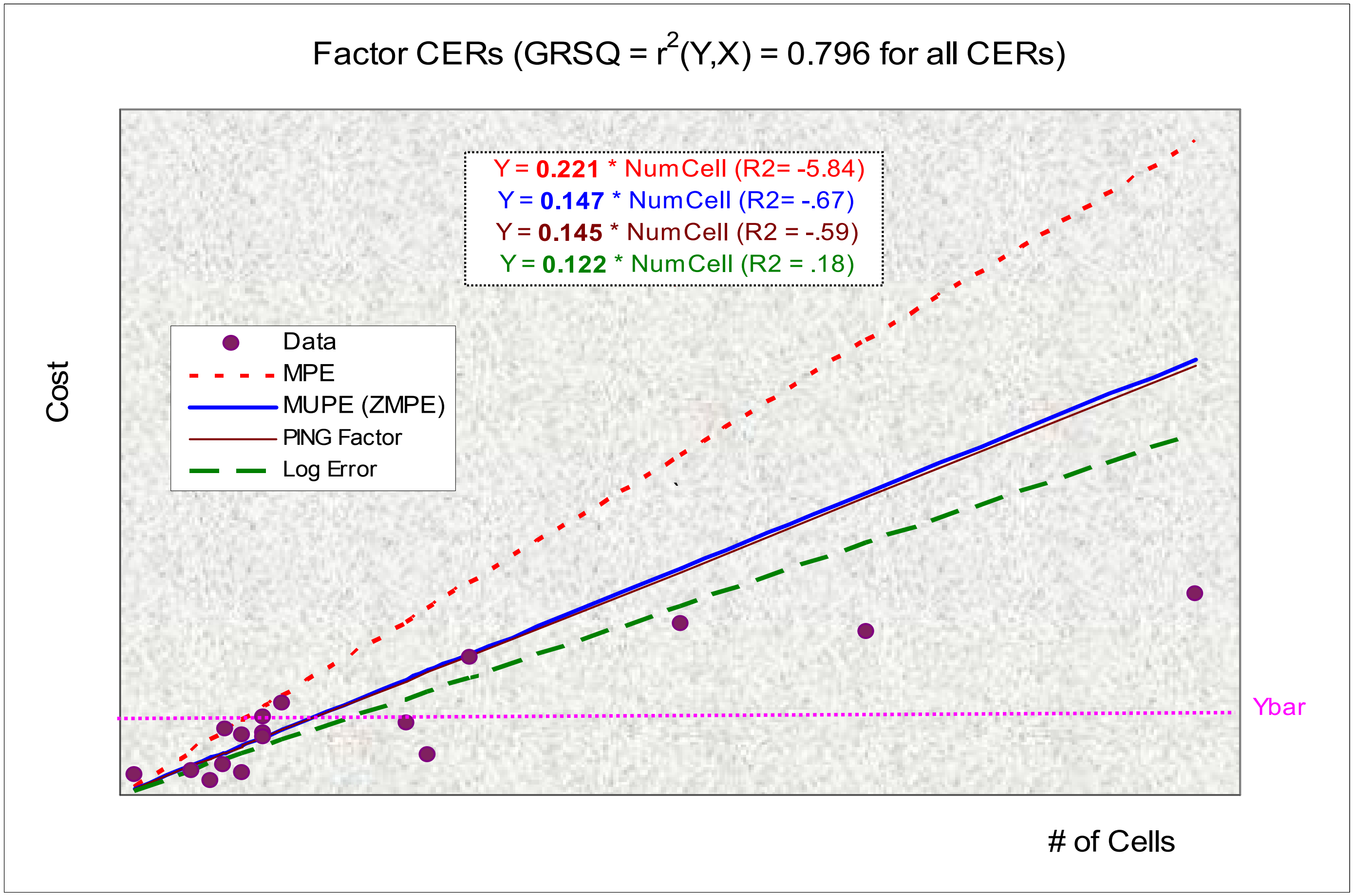
r^2 Insensitive to Different Fitting Methods (3/5)

- These CERs are derived from the USCM7 EPS generation data set
- GRSQ (r^2) is the same for all **linear** equations regardless of the fitting methods; GRSQ = r^2(cost, NumCells) for all 6 CERs; R^2/Adj. R^2 varies in all six



r^2 Insensitive to Different Fitting Methods (4/5)

- These factor CERs are derived from the USCM7 EPS generation data set
- GRSQ (r^2) is the same for all **factor** equations regardless of the fitting methods; GRSQ = r^2(cost, NumCells) for all CERs; R^2 varies in all



r^2 Insensitive to Different Fitting Methods (5/5)

TECOLOTE RESEARCH, INC.

- For simple linear and factor CERs, GRSQ is equal to the square of Pearson's correlation coefficient between the dependent and independent variables <u>regardless of the fitting methods</u>
  - **GRSQ = $r^2(Y,X)$** = OLS's **R^2** for simple linear or factor CERs, including MPE, MUPE, ZMPE, OLS, Log-Error, and PING-Factor (PF) CERs
- GRSQ (= $r^2(y, \hat{y})$) measures the <u>linear association</u> between y (actual cost) and $\hat{y}$ (predicted cost); it cannot detect the actual deviation between them
- Note: For CERs with multiple drivers, GRSQ cannot be measured by the pairwise correlation between the dependent and any individual independent variables

r^2 Insensitive to CER Forms

- As shown in Ref 1, most of the GRSQ numbers are about the same regardless of the CER forms (see Tables 1 ~ 5 in Ref 1)
- Here are two examples from Ref 1 (Table 3 and Table 5):

Parameter or Measure	OLS Linear $y = a + bx$	Multiplicative Error Linear $y = a + bx$	Log-Log Fit $y = ax^b$	Direct Nonlinear Fit $y = ax^b$
a	60.717	-89.829	62.889	48.267
b	47.018	64.842	0.905	1.063
GRSQ	0.615	0.615	0.616	0.614
R^2 =1 - *SSE/SST*	0.615	0.478	0.603	0.527
CrossTerm/SST	0.000	-0.741	0.043	-0.528

Parameter or Measure	OLS Linear $y = a + bx$	Multiplicative Error Linear $y = a + bx$	Log-Log Fit $y = ax^b$	Direct Nonlinear Fit $y = ax^b$
a	31.408	41.282	35.411	84.896
b	21.454	39.107	0.872	0.517
GRSQ	0.317	0.317	0.334	0.378
R^2 =1 - *SSE/SST*	0.317	-0.398	0.261	-0.003
CrossTerm/SST	0.000	-1.952	-0.338	-1.074

- GRSQ (r^2) is not sensitive to the exponent change
  - In Table 3, the exponent ranges from 0.9 to 1.06 while GRSQ is about the same
  - In Table 5, the exponent is between .5 and 1 while GRSQ ranges from .32 to .38
    Note: when the exponent is almost doubled, there is only a 16% change in GRSQ
  - In Table 5, the last CER doesn't seem to be the best, but its GRSQ is the largest

Why Is r^2 Insensitive?

Consider a Triad equation form: $Y = a + b*X^c$

- **GRSQ = $r^2(y, \hat{y}) = r^2(y, \hat{a} + \hat{b} * x^{\hat{c}}) = r^2(y, x^{\hat{c}})$**
- **Pearson's r is invariant under any linear transformations**
  - The fixed-cost term (a), slope parameter (b), and error term (ε) are not used in the computation of GRSQ
- **If several different methods generate a similar sensitivity parameter, i.e., the exponent, GRSQ should be similar among these CERs regardless of the following:**
  - whether or not the error term is additive or multiplicative
  - the size and sign of a and b
  - the fitting methods
- **Note: GRSQ is a constant for all CERs when c is set to be a fixed number (by engineering logic)**

Invariance <u>Not</u> a Desirable Property!!!

- **GRSQ, i.e., $r^2(y, \hat{y})$, remains the same when you multiply, divide, add, and/or subtract your estimate ($\hat{y}$) by <u>any</u> amount, which is <u>not</u> a desirable property**
  - $r^2(y, \hat{y}) = r^2(y, a + b^* \hat{y})$

$$r(a+bx, c+dy) = \frac{\sum_{i=1}^{n}(a+bx_i - a - b\bar{x})(c+dy_i - c - d\bar{y})}{\sqrt{\sum_{i=1}^{n}(a+bx_i - a - b\bar{x})^2}\sqrt{\sum_{i=1}^{n}(c+dy_i - c - d\bar{y})^2}} = \frac{bd\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{bd\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = r(x, y)$$

- **Invariance may seem to be a valuable characteristic, but is in fact a detrimental property for GRSQ**
  - Note: Neither R^2 nor Adjusted R^2 is invariant under linear transformations whether or not the CER is linear or nonlinear
- **A well-defined stat may not necessarily be a helpful one**
  - A “pro” of GRSQ: GRSQ always has the same meaning and value as a metric regardless of how the estimates are determined

TECOLOTE RESEARCH, INC.

$r^2 = R^2$ for Linear Models

- **GRSQ is insensitive to and does not have explanatory power for the multiplicative linear and factor CERs**
  - GRSQ = $\mathbf{r^2}(y, \hat{y}) = r^2(y, x)$ = OLS's $\mathbf{R^2}$ for simple linear and factor CERs whether or not the error is additive or multiplicative
- **Why should we use an OLS metric to describe the quality of multiplicative error models?**

Example / Equation	Method	GRSQ (r^2)	R^2	Adj. R^2
#2: Y = 582.6 + 0.098 * X	MPE	0.796	0.545	0.515
#2: Y = 482.3 + 0.088 * X	PING Factor	0.796	0.706	0.687
#2: Y = 488.5 + 0.086 * X	MUPE	0.796	0.721	0.702
#2: Y = 506.4 + 0.085 * X	ZMPE	0.796	0.731	0.714
#2: Y = 437.8 + 0.080 * X	Log Error	0.796	0.761	0.745
#2: Y = 809.2 + 0.067 * X	**OLS**	**0.796**	**0.796**	0.782
#3: Y = 0.221 * X	MPE	0.796	-5.84	-5.84
#3: Y = 0.147 * X (SPE=73%)	MUPE	0.796	-0.67	-0.67
#3: Y = 0.147 * X (SPE=73%)	ZMPE	0.796	-0.67	-0.67
#3: Y = 0.145 * X	PING Factor	0.796	-0.59	-0.59
#3: Y = 0.122 * X	Log Error	0.796	0.18	0.18

R^2 and Adj. R^2 vary under different fitting methods

GRSQ = 0.796 for all !?

Outline

- **Multiplicative Error Models**
- **Definitions of GRSQ (r^2), R^2, Adj. R^2, and SPE**
  - Interpretations of Adj. R^2 and SPE
- **Properties of R^2/Adj. R^2 and GRSQ (r^2)**
- **Concerns about GRSQ (r^2) and R^2/Adj. R^2**
- **Analyze R^2/Adj. R^2 and GRSQ Using Examples**
  - GRSQ insensitive to different fitting methods and CER forms
  - Why is GRSQ insensitive?
- ***Propose Modified Adj. R^2 (for MUPE) and GRSQ (for DF)***
- **Conclusions**

Modify Adj. R^2 for MUPE

- **Adjusted R^2 in unit space:**

$$Adj.\,R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n-p)}{\sum (y_i - \bar{y})^2 / (n-1)} = \frac{MSE_{\bar{Y}} - MSE_f}{MSE_{\bar{Y}}}$$

- **Adjusted R^2 for MUPE:**

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

- **This modified adjusted R^2 compares MUPE's SPE^2 to its baseline (i. e., SPE^2 of an average CER); it is more pertinent to the fitting methodology**
- ***Adjusted R^2 for MUPE* puts SPE in perspective**

Modify GRSQ for Degrees of Freedom (DF)

- **Generalized R^2 (GRSQ): GRSQ is Pearson's r^2 between the actual {y} and predicted {ŷ} values in unit space, i.e., GRSQ = $r^2(y, \hat{y})$**
  - GRSQ does not take the sample size or degrees of freedom into account
- **Modified GRSQ for DF:**

$$\begin{aligned} \mathbf{GRSQ_{DF}} &= \mathbf{r^2 - (1 - r^2)*(p-1)/(n-p)} && \text{if } p > 1 \\ &= \mathbf{r^2 - (1 - r^2)*1/(n-1)} && \text{if } p = 1 \end{aligned}$$

(r^2 = GRSQ, p = number of estimated coefficients, n = sample size)

- **This modified GRSQ takes the degrees of freedom and number of parameters into consideration**

Conclusions

- **Use GRSQ with caution for non-OLS CERs**
  - GRSQ is a constant for all CERs regardless of the fitting methods when developing simple linear and factor CERs
  - GRSQ is also insensitive to different CER forms (e.g., exponent change)
  - GRSQ only measures linear association between actual and predicted
- **Neither SPE nor GRSQ can determine whether the regression coefficients are significant; they can't detect model flaws either**
  - GRSQ does not have the same statistical meaning and value of the traditional R^2 if the regression is not an OLS (GRSQ = R^2 for OLS)
- **Adj. R^2 is a complement to GRSQ; use them together**
  - Both are predictive measures and can be influenced by outliers
  - Suggest using the symbol r^2 (not R^2) for GRSQ to avoid confusion, as Pearson's correlation is commonly denoted by r, not R
- **Knowing your data set is more important than comparing statistical measures**

Recommendations

- **Use "Adjusted R^2 for MUPE" to evaluate MUPE CERs, as it is more relevant to the fitting method**

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

- **Modify GRSQ to correct for degrees of freedom**

$$GRSQ_{DF} = r^2 - (1 - r^2)*(p-1)/(n-p) \quad \text{if } p > 1$$

$$= r^2 - (1 - r^2)*1/(n-1) \quad \text{if } p = 1$$

(r^2 = GRSQ, p = number of estimated coefficients, n = sample size)

- **Use $GRSQ_{DF}$ to update the correlation analyses for the USCM CERs and database**
- **Do not rely on a single measure for selecting the best CER**
  - Use relevant fit measures when evaluating CERs; use predictive measures (Adj. R^2 & GRSQ) for supplemental evaluations

References

1. Book, S. A. and Young, P. H., "The Trouble with R2," International Society of Parametric Analysts, Journal of Parametrics, Vol. XXV, No 1 (Summer 2006), pages 87-112
2. Hu, S. and Smith, A., "Why ZMPE When You Can MUPE," 6th Joint Annual ISPA/SCEA International Conference, New Orleans, LA, 12-15 June 2007
3. Hu, S., "The Impact of Using Log-Error CERs Outside the Data Range and PING Factor," 5th Joint Annual ISPA/SCEA Conference, Broomfield, CO, 14-17 June 2005
4. Nguyen, P., N. Lozzi, et al., "Unmanned Space Vehicle Cost Model, Eighth Edition," U. S. Air Force Space and Missile Systems Center (SMC/FMC), Los Angeles AFB, CA, November 2001
5. 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
6. Book, S. A. and Lao, N. Y., "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
7. Nguyen, P., N. Lozzi, et al., "Unmanned Spacecraft Cost Model, Seventh Edition," U.S. Air Force Space and Missile Systems Center (SMC/FMC), Los Angeles AFB, CA, August 1994
8. Hu, S. and Sjovold, A. R., "Multiplicative Error Regression Techniques," 62nd MORS Symposium, Colorado Springs, Colorado, 7-9 June 1994
9. Young, P. H., "Generalized Coefficient of Determination," 26th Annual DoD Cost Analysis Symposium, Leesburg, VA, September 1992
10. Young, P. H., "GERM: Generalized Error Regression Model," 25th Annual DoD Cost Analysis Symposium, Leesburg, VA, September 1991
11. Hu, S. and Sjovold, A. R., "Error Corrections for Unbiased Log-Linear Least Square Estimates," TR-006/2, March 1989
12. Seber, G. A. F., and C. J. Wild, "Nonlinear Regression," New York: John Wiley & Sons, 1989, pages 37, 46, 86-88
13. Weisberg, S., Applied Linear Regression, 2nd Edition," New York: John Wiley & Sons, 1985, pages 87-88
14. Goldberger, A. S., "The Interpretation and Estimation of Cobb-Douglas Functions," Econometrica, Vol. 35, July-Oct 1968, pp. 464-472

TECOLOTE RESEARCH, INC.
Bridging Engineering and Economics
Since 1973

Backup Slides

Measures of Fit: T1 vs. Wt & GEO Orbit Dummy Variable

TECOLOTE RESEARCH, INC.

MULTIPLICATIVE			N= 17		TT&C SUBSYSTEM RECURRING COST CER					
RUN	DEPENDENT VARIABLE	SE (MULT)	PEARSON'S CORR COEF	MAD of % ERROR	RMS OF % ERRORS	ADJ R^2	X1 COEF T SCORE***	X2 COEF T SCORE***	X3 COEF T SCORE***	CONST T SCORE***
FACTOR	T1K	0.2497	0.7130	18.0780	23.4464	-0.1158	13.8208	2.4570	0.0000	0.0000
	EQUATION	T1K = 41.794 * WEIGHTLB + 1064.627 * ORBIT								
	REMARKS									
LINEAR	T1K	0.1832	0.7110	11.3870	16.6326	0.4201	5.1526	1.1167	0.0000	3.5334
	EQUATION	T1K = 1737.572 + 24.408 * WEIGHTLB + 434.634 * ORBIT								
	REMARKS									
CURVE	T1K	0.1757	0.7480	10.6460	15.9338	0.4946	2.2980	5.2796	10.8654	0.0000
1	EQUATION	**T1K = 441.546 * WEIGHTLB ^ 0.491 * 1.13 ^ ORBIT**								
	REMARKS									
TRIAD	T1K	0.1825	0.7460	10.6370	15.9661	0.4528	0.3270	1.1012	6.8263	0.0904
	EQUATION	T1K = 292.978 + 339.597 * WEIGHTLB ^ 0.533 * 1.14 ^ ORBIT								
	REMARKS									

1. A significant intercept is omitted from the factor equation, which cannot be detected by examining the CER's SE or Pearson's r
2. Neither SPE nor Pearson's r^2 can detect model flaws; **Adj. R^2 may provide a clue**