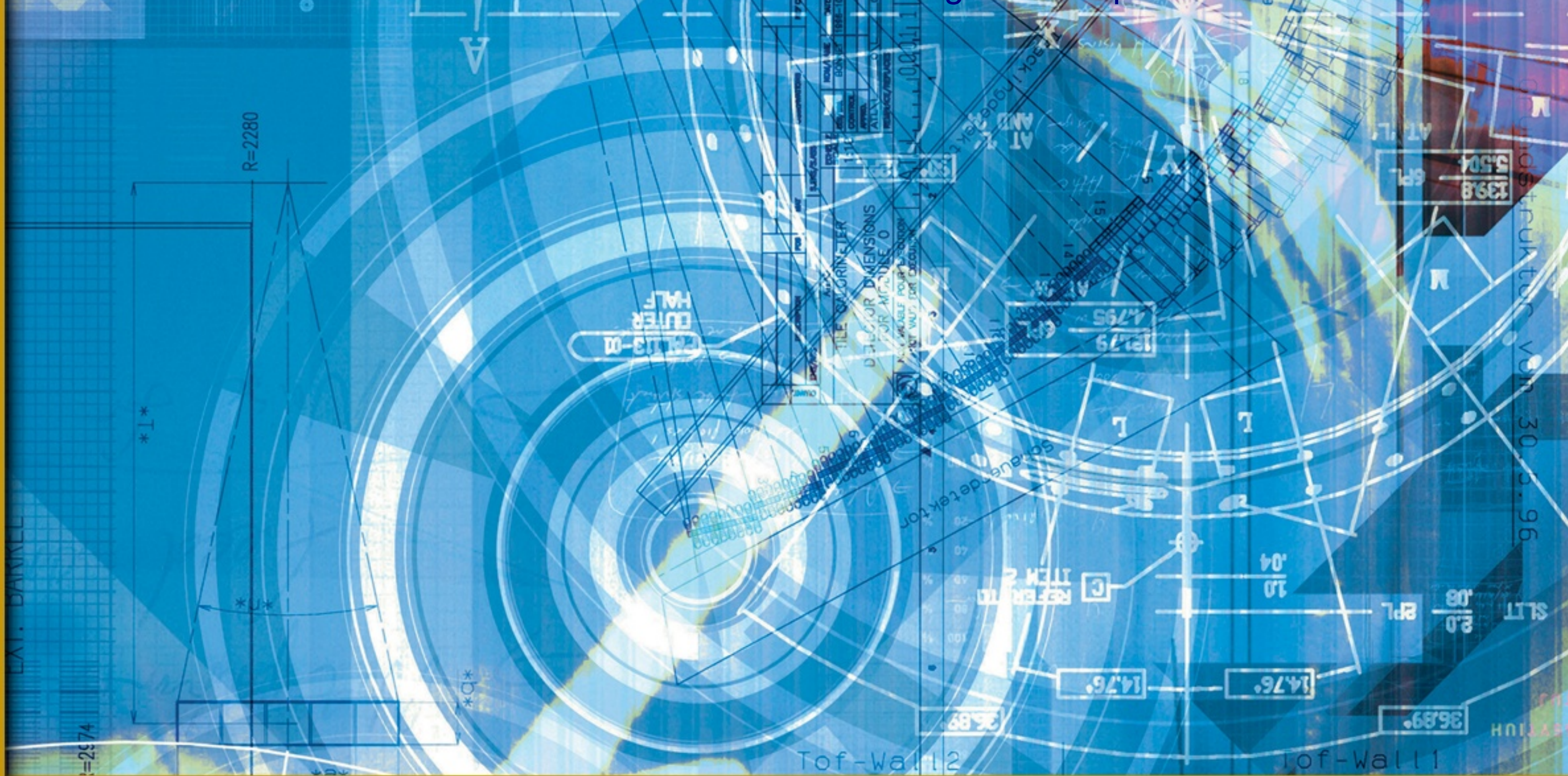


Significant Reasons to Eschew Log-Log OLS Regression when Deriving Estimating Relationships

Presented at 2012 SCEA/ISPA Joint Annual Conference and Training Workshop, June 26-29, 2012

Written by Stephen A. Book, MCR LLC, El Segundo CA
Presented by Raymond P. Covert, MCR LLC

MCR
CRITICAL THINKING.
SOLUTIONS DELIVERED.

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Abstract

Log-Log Ordinary Least Squares (LLOLS) regression, considered in the 18th and early 19th Centuries as the best (and, in fact, the only) method for fitting nonlinear algebraic relationships of the form $y = ax^b$ to data sets of (x,y) pairs, has a number of serious defects that make it far from adequate for CER development in the 21st. No other option was available 200 years ago, but the advances in computing power and techniques of statistical optimization available to us today leave no reason to stick with an obsolete method. In the 21st Century, we insist that 21st Century engineering technologies be applied, so why would we continue to develop CERs to estimate them using 18th Century statistical methods?

Continuing to derive CERs via LLOLS imposes a number of unfortunate burdens on the analyst that require several special adjustments to counteract them. Among these are the following: (1) LLOLS CERs do not minimize the error of estimating cost; (2) they are almost always biased low; (3) when a bias "correction" is made to them, quality metrics such as standard error and R^2 must be recalculated; and (4) the logarithmic space standard error that LLOLS reports is not related in any simple way to the CER's actual standard error.

Furthermore, restricting one's CER-derivation techniques to OLS and LLOLS involves one in a web of contradictions, among them: (1) Nonlinear CERs whose coefficients are derived by LLOLS must have fixed cost = zero, while linear CERs whose coefficients are derived by OLS are permitted to have nonzero fixed-cost terms; and (2) Nonlinear CERs derived by LLOLS must have standard errors expressible as a percentage of the estimate, while linear CERs derived by OLS must have standard errors expressed as plus/minus dollar values. Finally, in what may be the most significant issue that makes use of LLOLS impractical, there are only a very few nonlinear algebraic forms that can be treated using LLOLS, namely those that are amenable to the algebraic properties of logarithms.

The objective of this presentation is to encourage cost analysts to wean themselves off the 18th Century LLOLS technique and move on to 21st Century methods that have optimal estimating and statistical properties and a wider range of applicability.

MCR
CRITICAL THINKING.
SOLUTIONS DELIVERED.

Acknowledgments

The presenter would like to express his appreciation for the vital efforts of David L. Hansen, who as USAF SMC/FMC Cost Chief in the middle 1990s recognized the serious necessity of fixing a broken CER-development process and took strong action to rectify the problem. Rather than deciding to "keep doing it the way we've always done it," Mr. Hansen's attitude was "I understand the problem – let's fix it."

He would also like to express his gratitude to MCR's internal quality-review team members, Ray Covert and and Neal Hulkower for their valuable comments and suggestions that led to a substantial improvement of the exposition.

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Nonlinear Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is “Corrected” to Zero, the CER’s Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Nonlinear Cost-Estimating Relationships (CERs)
- Summary

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Algebraic Expressions for CERs

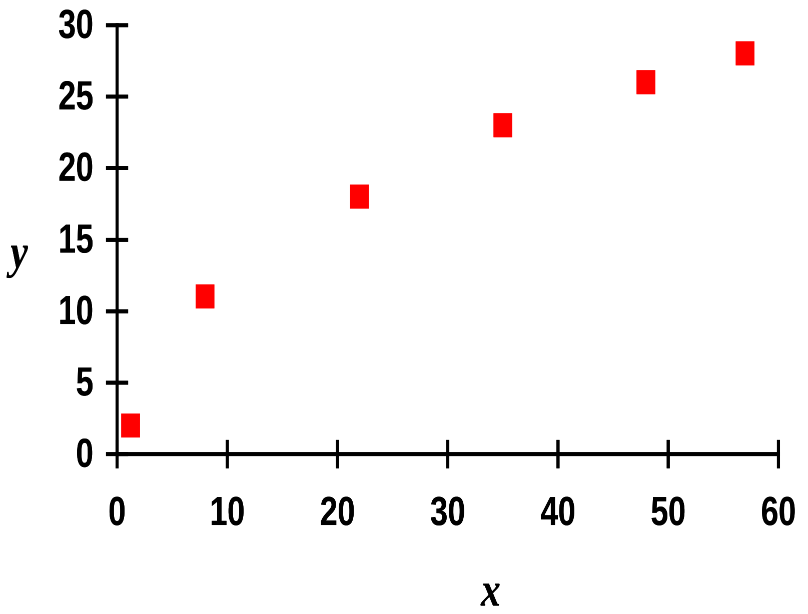
- y = Cost
  x = Technical Parameter (Cost Driver)
- Factor CER: $y = ax$
  Linear CER: $y = a + bx$
  "Nonlinear" CERs: Power CER: $y = ax^b$
  Exponential CER: $y = ab^x$
  Triad CER: $y = a + bx^c$
- a, b, c are Constant Coefficients or Exponents Derived from the Historical Data Base
- This Discussion Will Concentrate on the Case of Only One Cost Driver per CER – For Multiple Cost Drivers, the Concepts are the Same, but the Statistics is a Little More Complicated

CRITICAL THINKING. SOLUTIONS DELIVERED.

Ideal Historical Data (after Normalization)

COST	TECHNICAL PARAMETER
y_1	x_1
y_2	x_2
y_3	x_3
.	.
.	.
.	.
y_n	x_n

SCATTERGRAM



Note: Everything Said Regarding Statistical Procedures in this Presentation Applies not only to Cost, but also to Schedule, Weight, and Other Estimating Relationships

Contents

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

OLS Linear Regression

- **Linear CER Additive-Error Model:** $y = a + bx + \varepsilon$
  (i.e., Actual Cost = Estimated Cost + Error of Estimation)
- **OLS Regression Minimizes Sum of Squared Errors**
  - **Actual cost for data point** i **is** y_i
  - **Estimated cost for data point** i **is** $a + bx_i$
  - **Error of estimation for data point** i **is** $\varepsilon_i = y_i - (a + bx_i)$
  - **Choose values for** a **and** b **that minimize**
    $\sum (y_i - a - bx_i)^2 = \sum \varepsilon_i^2$
  - **Resulting estimates are (sample) unbiased, as well as unbiased in the formal statistical sense**
- **OLS Solution:**

$$b = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a = \frac{\sum y_i - b\sum x_i}{n}$$

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

OLS Standard Error of the Estimate

- A "One-Sigma"-type Error Bound on Error Implicit in OLS CERs
- Sample Standard Error of Estimate

$$\text{SEE} = \sqrt{\frac{1}{n-k}\sum_{i=1}^{n}\left(y_i - a - bx_i\right)^2}\ \text{Dollars}$$

where

– $y = a+bx$ is the OLS CER that Expresses Cost (y) in Terms of a Cost-Driving Technical or Programmatic Parameter (x)
– n is the Number of Data Points Used to Derive the CER
– k is the Number of Coefficients in the Algebraic Expression for the CER, e.g., $k = 2$ for the CER $y = a+bx$

- Sample Standard Error is an OLS CER Quality Metric

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Bias and Its Role in Estimating

- There is a Precise Quantitative Mathematical Definition of CER Bias*
- Qualitatively, though, Bias is the Tendency of a CER to, on the Average, Overestimate ("has positive bias" or "is biased high") or Underestimate ("has negative bias" or "is biased low") the Cost of Projects in the Data Base from which the CER was Derived
- We then Assume that the CER Will Estimate a New Project with the Same Bias Tendency as it Estimates the Projects in its Supporting Data Base
- We are Obligated to Believe This if we Assume that the CER is a Valid Tool to Use in Estimation

* For details, refer to any advanced-level statistics textbook such as R.J. Larsen and M.L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, 3rd Ed., Prentice-Hall, 2001, pages 338-344.

CRITICAL THINKING.
SOLUTIONS DELIVERED.

Footnote on Bias

- "Bias" is a Theoretical Statistical Concept, but "Sample Bias" is the Actual Value in Any Particular Case of Averaging a CER's Overestimates and Underestimates of the Costs of Projects in the Data Base from which the CER was Derived
  - Mechanically, We Calculate the Sample Bias by Summing the Overestimates (estimate minus actual = a positive number) and Underestimates (estimate minus actual = a negative number)
  - Then we Divide the Sum by the Number of Points in the Data Base to Obtain the Sample Bias
- Sample Bias is an Estimate of the Theoretical Bias
- In this Presentation, we will use the Word "Bias" to Mean "Sample Bias"

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

OLS Bias = Zero

- **Bias is a CER's Tendency to Produce Estimates that are Higher (biased high) or Lower (biased low) than the Actual Costs, on the "Average"**
- **Bias = Zero (dollars) Means that the CER Estimates the "Average" (Mean) Cost at Each Cost-Driver Value**
- **Sample Bias** $= \frac{1}{n}\sum_{i=1}^{n}\left(a + bx_i - y_i\right)$ **= 0 Dollars because**

$$a = n^{-1}\left(\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i\right)$$

  - $y = a+bx$ **is the OLS CER that Expresses Cost (*y*) in Terms of a Cost-Driving Technical or Programmatic Parameter (*x*)**
  - $y_i - a - bx_i$ **is the i^{th} "Residual," the Actual i^{th} Cost Value, Minus its Estimate**
  - n **is the Number of Data Points Used to Derive the CER**
- **Sample Bias is an OLS CER Quality Metric**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

R^2

- R^2 Measures Linearity of the Relationship between Actual (*ACT*) Costs and their Estimates (*EST*)
- Ideally this Relationship Should be Perfectly Linear: Estimates Should Equal Actuals Exactly, i.e., if Estimates are Graphed against Actuals, the Graph Should be the Straight Line $y = x$
- Sample R^2 = $$\frac{\left[n\sum_{k=1}^{n} ACT_k EST_k - \sum_{k=1}^{n} ACT_k \sum_{k=1}^{n} EST_k\right]^2}{\left[n\sum_{k=1}^{n} ACT_k^{\,2} - \left(\sum_{k=1}^{n} ACT_k\right)^2\right]\left[n\sum_{k=1}^{n} EST_k^{\,2} - \left(\sum_{k=1}^{n} EST_k\right)^2\right]}$$
  - R^2 is Usually Expressed as a Percentage, 0% to 100%, with 100% Signifying a Perfect Linear Predictive Relationship
  - n is the Number of Data Points Used to Derive the CER
- Sample R^2 is an OLS CER Quality Metric

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Example of a Nonlinear Concave Data Set

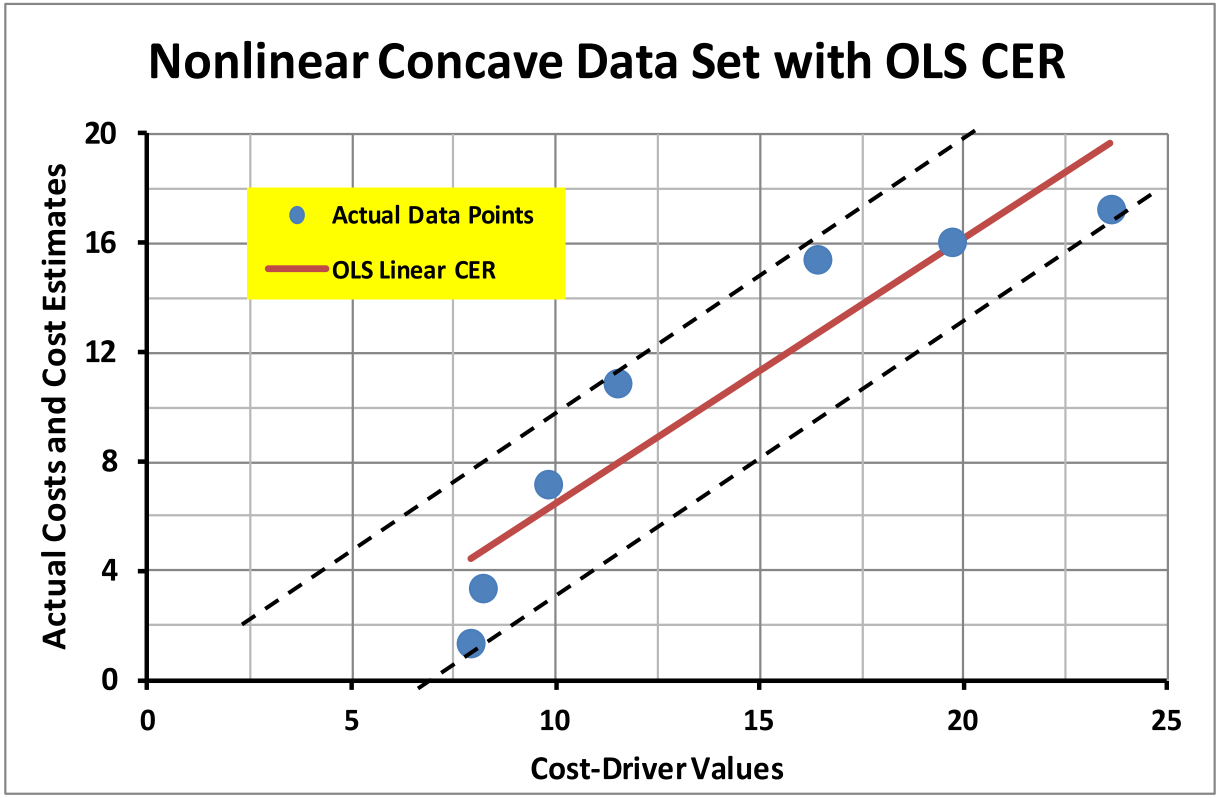
x Values (Cost Driver)	y Values (Actual Costs)	Predicted y Values (Cost Estimates)	Residuals = Actuals-Estimates
7.9	1.380	4.449	-3.069
8.2	3.395	4.740	-1.345
9.8	7.201	6.291	0.910
11.5	10.900	7.938	2.962
16.4	15.434	12.688	2.746
19.7	16.074	15.886	0.188
23.6	17.274	19.666	-2.392
Sums =	71.658	71.658	0.000

Note: OLS CER derived from x and actual y values is $y = a + bx$, where a = -3.207 and b = 0.969.

Note: A "concave" data set is one whose y trend bends downward as its x value moves to the right (see the next chart for an illustration). Obviously, all concave data sets are nonlinear.

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

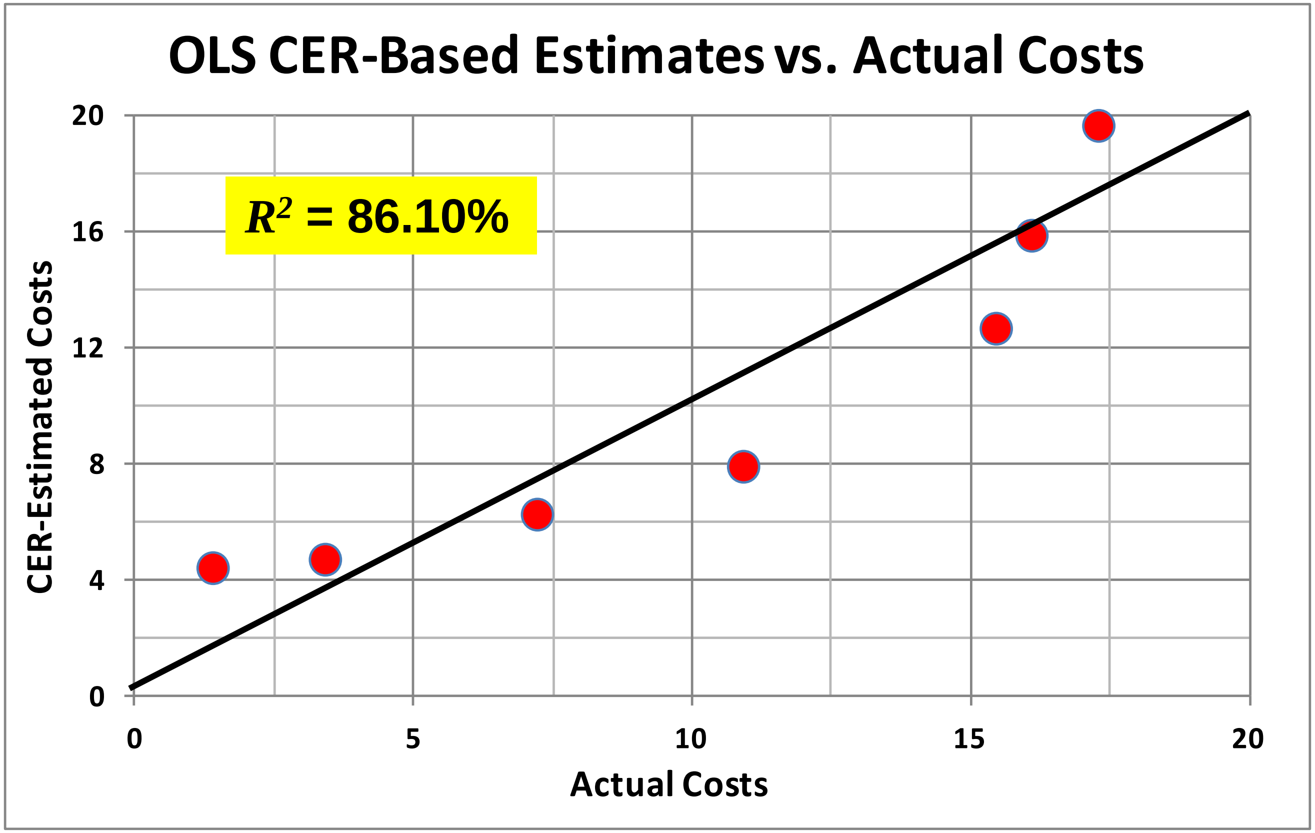
OLS CER and Its Quality Metrics



- **SEE = 2.613 Dollars (one-sigma bounds are illustrated)**
- **Bias = 0.000 Dollars (sample bias)**
- R^2 **= 86.10%**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Why R^2 is not 100%



- If All (x,y) = (Actual,Estimate) Pairs had Fallen on the 45° Line $y = x$ (i.e., All Estimates = Corresponding Actuals) , then R^2 Would have been 100%

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)
- Summary

CRITICAL THINKING. SOLUTIONS DELIVERED.

Using Logarithms

- Consider the Nonlinear Power CER Model $y = ax^b$
- Take Logarithms of Both Sides:

$$\log y = \log a + b \log x$$

- We Can Use OLS Mathematics to Determine the Values of a and b that will "Best" Estimate $\log y$:
  - Assume Additive-Error Model for $\log y$, i.e.

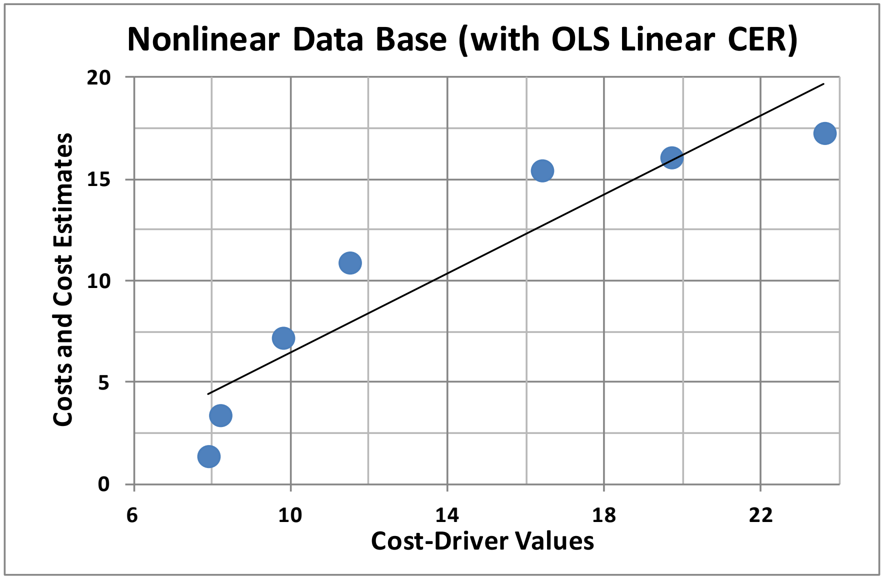
$$\log y = \log a + b \log x + E$$

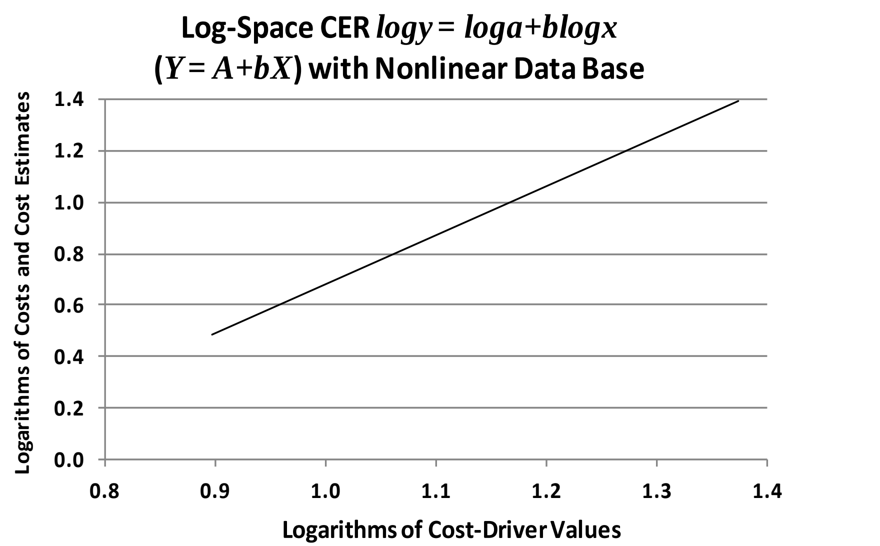
  - Here $E = \log y - (\log a + b \log x)$ is the Error of Estimation in Predicting Logarithm of Cost
  - Choose Values for a and b that Minimize

$$\sum(\log y_i - \log a - b \log x_i)^2 = \sum E_i^2$$

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

What Happens When a CER is Logarithmically Transformed?





- **Logarithmic Transformation of the Nonlinear Power CER $y = ax^b$ into a Linear Form Permits Use of OLS Formulas to Solve the Nonlinear Problem**
- **One Would Think that a Straight Line Would Fit the Logarithmic Data Somewhat Better than One Would Fit the Arithmetic Data**
- **Excel's "Trend Line" Function and Other Common "Quickie" Approaches Use This Technique**

CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)
- Summary

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

The LOLS Solution

- **In Log-Log Space the Logarithmic OLS (LOLS) CER has the form** $\log y = \log a + b \log x = A + b \log x$ **(where** $a = 10^{\log a} = 10^{A}$**)**
- **Use the OLS-Derived Formulas for the Values of** A **and** b

$$b = \frac{n\sum_{i=1}^{n}(\log y_i)(\log x_i) - \sum_{i=1}^{n}(\log y_i)\sum_{i=1}^{n}(\log x_i)}{n\sum_{i=1}^{n}(\log x_i)^2 - \left(\sum_{i=1}^{n}\log x_i\right)^2}$$

$$A = \log a = \frac{\sum_{i=1}^{n}\log y_i - b\left(\sum_{i=1}^{n}\log x_i\right)}{n}$$

- **The LOLS CER's Standard Error of Estimate is**

$$LSEE = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\log y_i - \log a - b\log x_i\right)^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}E_i^2}$$

MCR CRITICAL THINKING. SOLUTIONS DELIVERED.

Bias of the LOLS Solution

- **We Noted on the Previous Chart that the LOLS CER in Log-Log Space is Expressed as $\log y = A + b \log x$, where $A = \log a$ and $a = 10^A$**
- **The Output of Most Common Statistical Software Reports the LOLS CER's Bias (LB) as**

$$LB = \frac{1}{n}\sum_{i=1}^{n}\left(\log y_i - \log a - b \log x_i\right)$$

- **An LOLS CER is Really an OLS CER in Disguise, Based on Logarithms ($\log x_i$, $\log y_i$) of Data Points instead of Data Points Themselves (x_i, y_i), so LB is Zero in all Cases**
- **This Can be Seen Directly from the Expression for $A = \log a$ on the Previous Chart**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

R^2 of the LOLS CER Solution

- **Standard Output of Commercial Statistical Software Reports LR^2 as**

$$\frac{\left[n\sum_{k=1}^{n} \log ACT_k \log EST_k - \sum_{k=1}^{n} \log ACT_k \sum_{k=1}^{n} \log EST_k \right]^2}{\left[n\sum_{k=1}^{n} (\log ACT_k)^2 - \left(\sum_{k=1}^{n} \log ACT_k \right)^2 \right] \left[n\sum_{k=1}^{n} (\log EST_k)^2 - \left(\sum_{k=1}^{n} \log EST_k \right)^2 \right]}$$

- **We will See Later that the Result of this Calculation Provides Absolutely No Information about the R^2 Value of Original Nonlinear Power CER $y = ax^b$**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

LOLS Computations on the Nonlinear Concave Data Set

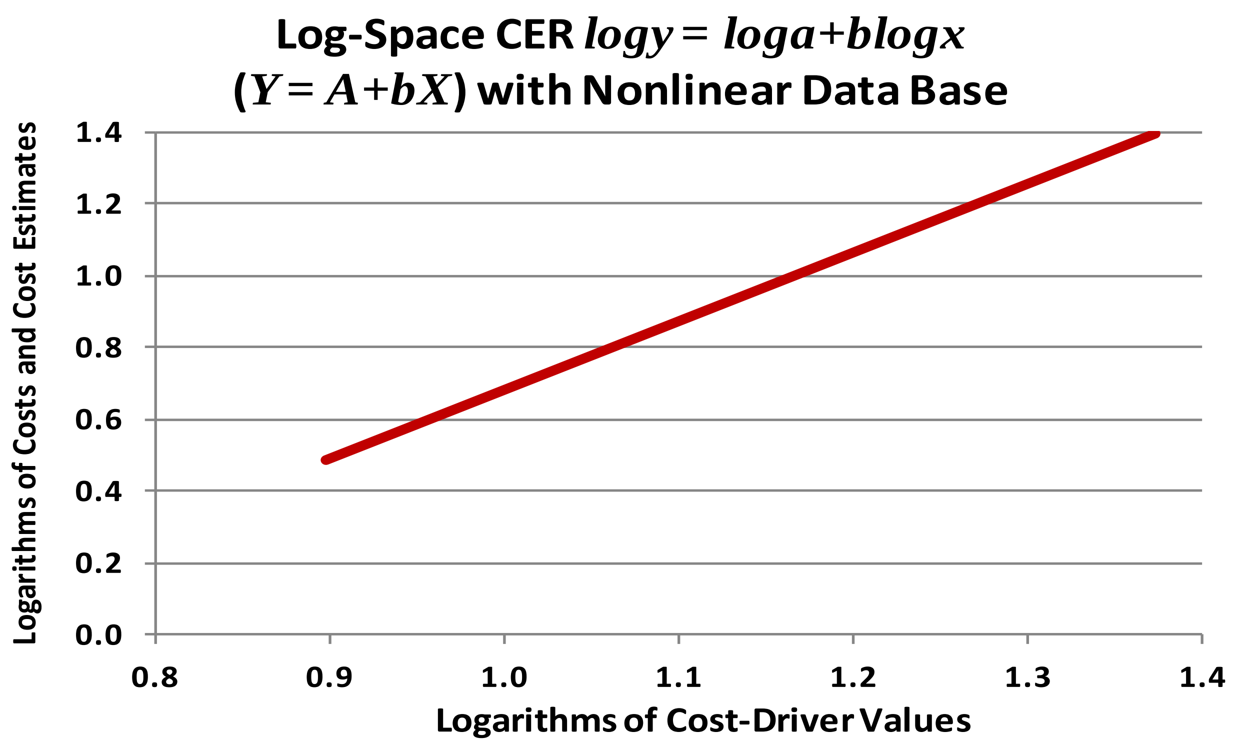
n	X Values (X = log x)	Y Values Y = log y)	X^2	Y^2	XY	Estimated Y Values	EstY-Y	$(EstY-Y)^2$
7	0.898	0.140	0.806	0.020	0.126	0.488	0.348	0.121
	0.914	0.531	0.835	0.282	0.485	0.519	-0.012	0.000
	0.991	0.857	0.983	0.735	0.850	0.666	-0.191	0.037
	1.061	1.037	1.125	1.076	1.100	0.798	-0.239	0.057
	1.215	1.188	1.476	1.412	1.444	1.091	-0.097	0.009
	1.294	1.206	1.676	1.455	1.561	1.243	0.037	0.001
	1.373	1.237	1.885	1.531	1.699	1.392	0.154	0.024

Denominator of b	1.499
Numerator of b	2.850
b	1.901
A	-1.218
a	0.060

Std Error (Log Space) LSEE =	0.2236
Bias (Log Space) LB =	0.0000
R^2 (Log Space) LR^2 =	75.60%

MCR
Critical Thinking. Solutions Delivered.

Log-Space CER (LOLS CER)



- **The LOLS CER Does Not Estimate Cost – It Estimates the Logarithm of the Cost**
- **The LOLS Process Minimizes the Error You Would Make if Your Intention Were to Estimate Log Cost**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Estimating Cost, Not its Logarithm

- **The Log-Log OLS (LLOLS) Power CER $y = ax^b$ is derived by Exponentiating the LOLS CER**

$$log\ y = log\ a + b\ log\ x$$

- **For Logarithms to Work, however, the Error Model of a Nonlinear Power CER Must be "Multiplicative," i.e.,**

$$y = ax^b\,\varepsilon$$

because Applying Logarithms Must Get Us to the OLS "Additive-Error" Model for Predicting the <u>Logarithm</u> of Cost (because OLS assumes additive error), namely $log\ y = log\ a + b\ log\ x + log\ (1+\varepsilon)$

- **But, with** $E_i = log\ \varepsilon_i$ **, LOLS Actually Minimizes**

$$\sum(log\,\varepsilon_i)^2 = \sum E_i^2 = \sum (log\,y_i - log\,a - b\ log\,x_i)^2$$

Instead of $\sum \varepsilon_i^2$

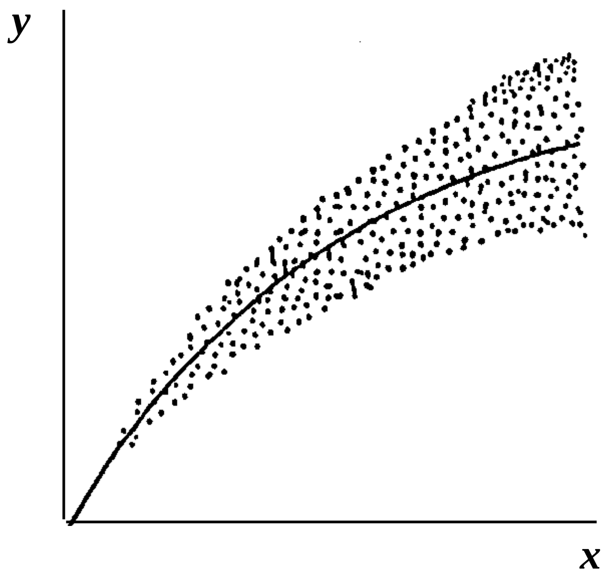
MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

A Word on Error of Estimation

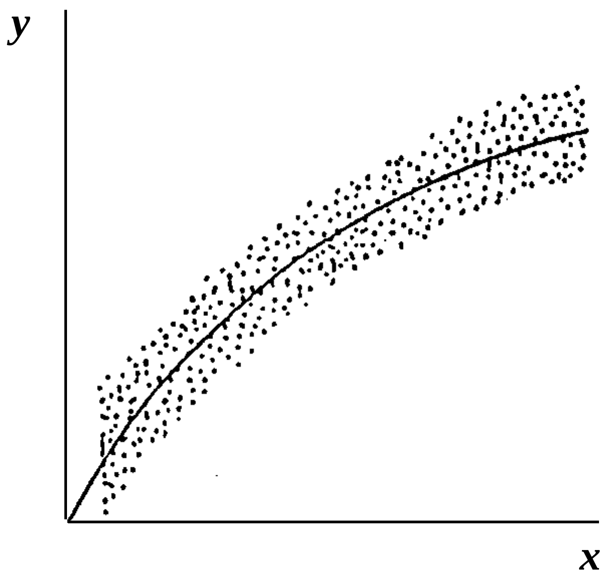
- **An OLS Linear CER is an Additive-Error CER having the form $y = a + bx + \varepsilon$**
- **However, Logarithms Cannot be Applied to a CER of the Form $y = ax^b + \varepsilon$ because Logarithms of Sums Cannot be Calculated**
- **So a Nonlinear CER Model Must be a "Multiplicative-Error" Model like the Form $y = ax^b\varepsilon$**
- **The Standard Error of a Multiplicative-Error CER is Expressed as a Multiple, Typically as a Percentage of the Estimate, rather than as a Constant Dollar Value Across the Estimating Range**
  - **"Error is ±30% of the Estimate"**
  - **Not "Error is ±$1,500"**

A Picture of the Two Error Models

Multiplicative Error



Additive Error



Reference: H.L. Eskew and K.S. Lawler, "Correct and Incorrect Error Specifications in Statistical Cost Models," *Journal of Cost Analysis*, Spring 1994, page 107.

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

The LLOLS Power CER $y = ax^b$ and Its Quality Metrics

- **The LLOLS Power CER:** $y = 0.060x^{1.901}$

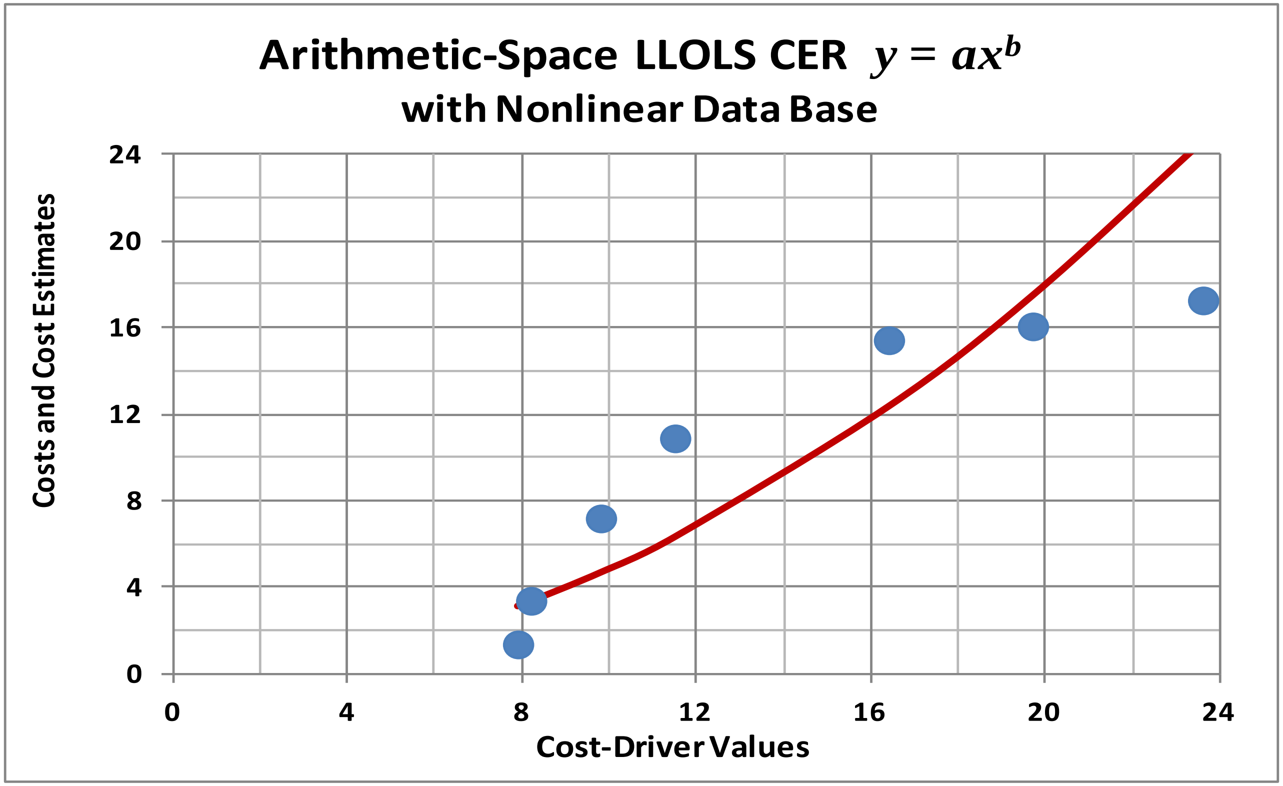
n	*x Values (Cost Driver)*	*y Values (Actual Costs)*	*Estimated y Values*	*Esty-y*	*(Esty-y)²*	*(Esty-y)/Esty*	*[(Esty-y)/Esty]²*
7	7.9	1.380	3.078	1.698	2.883	0.552	0.304
	8.2	3.395	3.304	-0.091	0.008	-0.028	0.001
	9.8	7.201	4.636	-2.565	6.577	-0.553	0.306
	11.5	10.900	6.284	-4.616	21.304	-0.734	0.539
	16.4	15.434	12.340	-3.094	9.573	-0.251	0.063
	19.7	16.074	17.486	1.412	1.994	0.081	0.007
	23.6	17.274	24.651	7.377	54.413	0.299	0.090
	Sums =	71.658	71.779	0.121	96.752	-0.634	1.309

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	4.3989
Bias (Arith Space) B =	0.0173
R^2 (Arith Space) R^2 =	77.63%

Mutliplicative-Error Quality Metrics	
Percentage Std Error =	51.17%
Percentage Bias =	-9.06%
R^2 =	77.63%

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Arithmetic Space LLOLS CER



- Exponentiating Linear LOLS CERs $logy = loga + blogx$ back to Arithmetic Space Reveals Nonlinearity of LLOLS CERs
- Disconnect between Concave Shape of the Data vs. CER's Convex Shape is due to Lack of a Fixed-Cost Term – More about this Later

CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)
- Summary

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS Does Not Minimize the Relevant Estimating Error

- The First Step in Deriving LLOLS Power CERs is Calculating the Values of *a* and *b* that Minimize the Sum of Squared Differences between the Logarithms of the Actuals and the Estimates, namely

$$\sum(\log y_i - \log a - b \log x_i)^2 = \sum (\log \varepsilon_i)^2$$

- But Minimizing $\sum (\log \varepsilon_i)^2$ not Same as Minimizing $\sum \varepsilon_i^2$ as in OLS Regression, so the *a* and *b* Values Turn out to be Different
- In Summary, the LOLS CER (in logarithmic space) …
  - Minimizes the Error of Estimating the Logarithm of Cost
  - … and its Standard Error is Expressed in Meaningless Units ("log dollars")
  - … so it Cannot Legitimately Claim to be an Optimal Way of Estimating Cost

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Even After the LOLS is Transformed to Arithmetic Space, Issues Remain

- **The Second Step in Deriving the LLOLS CER is to Exponentiate the Unusable Logarithmic-Space LOLS CER $\log y = \log a + b \log x$ to obtain a Usable Arithmetic-Space LLOLS CER $y = ax^b$**
- **LOLS Standard Error, Bias, and R^2 (which often are provided by an impersonal statistical software product) Cannot be Directly Converted to LLOLS' Standard Error, Bias, and R^2, so the Latter Numbers Must be Completely Recalculated**
- **Typically for LLOLS CERs, Standard Error Increases, Bias becomes Nonzero (it is 0 for LOLS CERs), and R^2 Decreases from their LOLS Values, Better Reflecting the Estimating Reality**

CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS Does Not Minimize CER Least-Squares Estimating Error

- **One Cause of Suboptimal LLOLS Quality Metrics is that Logarithms of Numbers are Usually Much Smaller than the Numbers Themselves and Therefore There is Lesser Distance between Them**
- **It follows that Different Values of *a* and *b* Will Result from Minimizing**

$$\sum(\log y_i - \log a - b \log x_i)^2 = \sum (\log \varepsilon_i)^2$$

than Would Result from Minimizing

$$\sum(y_i - a x_i^b)^2 = \sum (e_i)^2$$

where $e_i = y_i - a x_i^b$

- **But** $\sum(y_i - a x_i^b)^2 = \sum (e_i)^2$ **is Really What we Have to Minimize if we Want a CER that is as Close as Possible (in the sum-of-squared error additive-error sense) to the Data Points**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Footnote on Error Forms

- **Another Complicating Issue is that $\log e_i$ is not the same as $\log \varepsilon_i$**
- **What Causes this Situation is that the Basic LLOLS CER itself must be of the Multiplicative-Error Form, namely $y_i = ax_i^b\varepsilon_i$, in Order for the Logarithmic Transformation to Produce the LOLS CER Form**

$$\log y_i = \log a + b \log x_i + \log \varepsilon_i$$

- **But if we Want to Minimize Least-Squares Estimating Error for the Nonlinear CER $y_i = ax_i^b$ Just as we Do for Straight-Line OLS CERs, we Would Have to Minimize the Sum of Squared Errors $\sum(y_i - ax_i^b)^2$ Beginning with the Basic Additive-Error Form $y_i = ax_i^b + e_i$**
- **But That Algebraic Form Cannot be Handled using Logarithms because of the Algebraic Fact that $\log(U+V)$ Cannot be Simplified**

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- **Ordinary Least Squares (OLS) Regression**
- **The Logarithmic Transformation**
- **Log-Log OLS (LLOLS) Regression**
- **Five Unfortunate Characteristics of LLOLS**
  - **Error of Estimating Cost is Not Minimized**
  - **LLOLS CERs are Biased (usually low, but sometimes high)**
  - **When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated**
  - **CERs Derivable by LLOLS Must Have Fixed Cost = Zero**
  - **Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning**
- **A Few More Negatives**
- **A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)**
- **Summary**

MCR

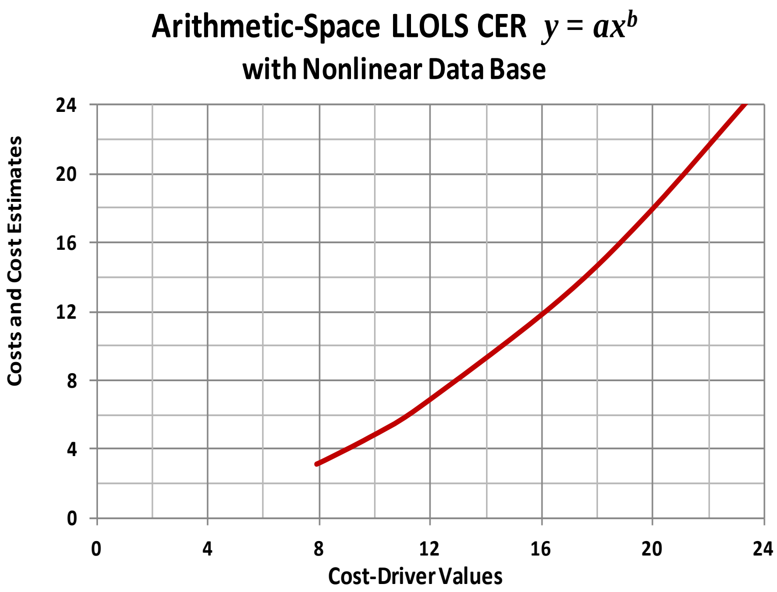
CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS CER for Our Concave Data Set is Biased High

- **Recall our Results for the LLOLS CER Based on Our Nonlinear Data Set**

CER: $y = 0.060x^{1.901}$

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	4.3989
Bias (Arith Space) B =	0.0173
R^2 (Arith Space) R^2 =	77.63%



- **Note that the Data Points Form a Concave Pattern, but the LLOLS CER is Convex (more about this later)**
- **Note also that the Bias in Arithmetic Space is Positive (= 0.0173 of the appropriate unit, e.g., $M)**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

What Happens in One Case of One Convex Data Set

- **Consider the Following Nonlinear Convex Data Set in Arithmetic Space:**

n	*x* Values (Cost Driver)	*y* Values (Actual Costs)	Estimated *y* Values
7	7.9	127.200	114.462
	8.2	138.300	121.288
	9.8	142.800	160.002
	11.5	177.400	205.159
	16.4	307.600	356.161
	19.7	483.100	473.573
	23.6	727.800	627.044
Sums =	97.100	2104.200	2057.689

- **The LLOLS Power CER Based on this Data Set has Coefficient and Exponent:**

a =	4.610
b =	1.554

Note: A "convex" data set is one whose y trend bends upward as its x value moves to the right (see the next chart for an illustration). Obviously, all convex data sets are nonlinear.

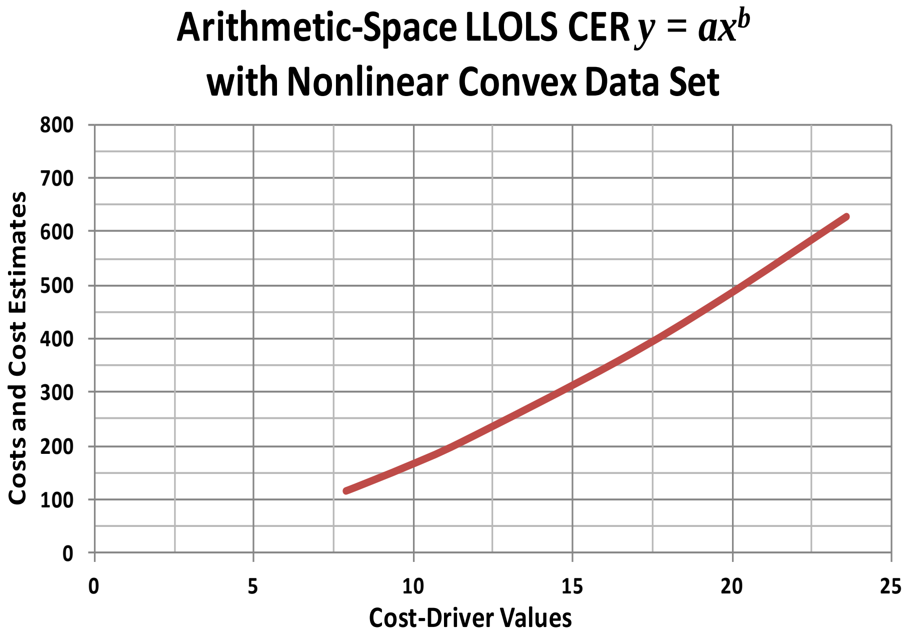
MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS CERs for a Convex Data Set Can be Biased Low

- **The LLOLS CER's Quality Metrics and Graphics are**

CER: $y = 4.601x^{1.554}$

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	53.1389
Bias (Arith Space) B =	-6.6445
R^2 (Arith Space) R^2 =	96.96%



- **Here the Data Points Form a Convex Pattern, and the LLOLS CER tracks that Pattern**
- **Note Here that Arithmetic-Space Bias is Negative (= -6.6445 of the appropriate unit, e.g., $M)**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)
- Summary

MCR
Critical Thinking. Solutions Delivered.

Bias and the LLOLS Process

- **The LOLS CER, Namely the OLS CER Relating $\log y$ and $\log x$, Having the Form $\log y = \log a + b \log x$, has Zero Bias, i.e., the Average of the Differences between the $\log a + b \log x_i$ and $\log y_i$ Values is Zero**
- **However, When the LOLS CER is Exponentiated to Become the LLOLS Nonlinear Power CER $y = ax^b$, the Latter CER no Longer has Zero Bias, i.e., the Average Difference between ax_i^b and y_i Values is NOT Zero**
- **Meyer(1941) Suggested that LLOLS CER Bias be Eliminated by Multiplying the CER by an Simple Exponential "Correction" Factor**
- **Since 1941, Several Factors Have Been Proposed to Support Various Bias-Correction Objectives**

Bias-Correction Factor Example

- The Professional Literature (see reference list at the end) Contains Several Additional Post-1941 Proposed Methods of Correcting Bias in LLOLS CERs
  - "Truncated Infinite Series" (Neyman and Scott, 1960)
  - "Reparametrization" (Heien, 1968)
  - "Tabulation" (Bradu and Mundlak, 1970)
  - "Smearing" (Duan, 1983)
  - "Inverse Linearization" (Miller, 1984)
  - "Ping Factor" (Hu and Sjovold, 1987)
- A Clear Summary of the Problem and Some of its Solutions was Presented to DoDCAS in 2003 by Jarvis and Rozzo of the OSD CAIG (now CAPE)
- To Keep Things as Simple as Possible in Examples, Let's Work with Simplest Possible Correction Factor, the "Balance-Adjustment Factor" (Book and Young, 1990) that Forces the Sample Bias to Zero

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Balance-Adjustment Factor (BAF)

- For a Set of Data Points (x_i, y_i) and an LLOLS CER $y = ax^b$ Derived from that Data Set, the BAF* is Defined and Calculated as

$$\beta = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} a x_i^b}$$

- To Prove that Multiplication of the LLOLS CER by the BAF "works" (i.e., produces a CER $y = \beta a x^b$ that has zero sample bias), Watch This: Sample Bias =

$$\sum_{i=1}^{n} (\beta a x_i^b - y_i) = \sum_{i=1}^{n} \beta a x_i^b - \sum_{i=1}^{n} y_i = \beta \sum_{i=1}^{n} a x_i^b - \sum_{i=1}^{n} y_i$$

$$= \left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} a x_i^b} \right) \sum_{i=1}^{n} a x_i^b - \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i = 0$$

* It is important to note a distinction between the BAF and the earlier factors: while the BAF is indeed the simplest proposed correction factor, its goal was to solve a simpler problem than the earlier factors do. They attempt to correct the theoretical bias, but the BAF corrects only the sample bias.

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

BAF Value for the LLOLS CER Derived from the Concave Data Set

- **The LLOLS CER Derived from the Nonlinear Concave Data Set is $y = 0.060x^{1.901}$ and its Quality Metrics are …**

x Values (Cost Driver)	*y Values (Actual Costs)*	*Estimated y Values*
7.9	1.380	3.078
8.2	3.395	3.304
9.8	7.201	4.636
11.5	10.900	6.284
16.4	15.434	12.340
19.7	16.074	17.486
23.6	17.274	24.651
Sums =	71.658	71.779

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	4.3989
Bias (Arith Space) B =	0.0173
R^2 (Arith Space) R^2 =	77.63%

- **The BAF Bias-Correction Factor is Therefore**

$$\beta = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} a x_i^b} = \frac{71.658}{71.779} = 0.9983$$

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Bias-Corrected CER of LLOLS Origin for Concave Data Set

- **The BAF-Corrected CER (no longer an LLOLS CER, of Course, but Still a Power CER) is**

$$y = 0.9983 \times 0.060x^{1.901} = 0.0599x^{1.901}$$

- **The New Estimates and Quality Metrics are …**

x Values (Cost Driver)	*y Values (Actual Costs)*	*Estimated y Values*
7.9	1.380	3.073
8.2	3.395	3.298
9.8	7.201	4.629
11.5	10.900	6.274
16.4	15.434	12.319
19.7	16.074	17.457
23.6	17.274	24.609
Sums =	71.658	71.658

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	4.3888
Bias (Arith Space) B =	0.0000
R^2 (Arith Space) R^2 =	77.63%

- **In this Case, as Well as the Bias Becoming Zero, the Standard Error Decreases Slightly (not guaranteed in all cases)**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

BAF Value for the LLOLS CER Derived from the Convex Data Set

- **The LLOLS CER Derived from the Convex Data Set is $y = 4.601x^{1.554}$ and its Quality Metrics are …**

x Values (Cost Driver)	*y Values (Actual Costs)*	*Estimated y Values*
7.9	127.200	114.462
8.2	138.300	121.288
9.8	142.800	160.002
11.5	177.400	205.159
16.4	307.600	356.161
19.7	483.100	473.573
23.6	727.800	627.044
97.100	2,104.200	2,057.689

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	53.1389
Bias (Arith Space) B =	-6.6445
R^2 (Arith Space) R^2 =	96.96%

- **The BAF Bias-Correction Factor is Therefore**

$$\beta = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} a x_i^b} = \frac{2,104.200}{2,057.689} = 1.0226$$

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Bias-Corrected CER of LLOLS Origin for Convex Data Set

- **The BAF-Corrected CER (no longer an LLOLS CER, but Again Still a Power CER) is**

$$y = 1.0226 \times 4.601 x^{1.554} = 4.705\ x^{1.554}$$

- **The New Estimates and Quality Metrics are …**

x Values (Cost Driver)	y Values (Actual Costs)	Estimated y Values
7.9	127.200	117.050
8.2	138.300	124.030
9.8	142.800	163.618
11.5	177.400	209.796
16.4	307.600	364.211
19.7	483.100	484.277
23.6	727.800	641.218
97.100	2,104.200	2,104.200

Additive-Error Quality Metrics	
Std Error (Arith Space) SEE=	49.9845
Bias (Arith Space) B =	0.0000
R^2 (Arith Space) R^2 =	96.96%

- **Notice that here also, along with the Bias Becoming Zero, the Standard Error Decreases Noticeably**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)
- Summary

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS CERs Cannot Have a Nonzero Fixed-Cost Term

- **Recall the Form of the OLS Linear Regression Equation: $y = a+bx$**
  - **a is a Fixed-Cost Term, a Cost Incurred Regardless of the Numerical Value of the Cost Driver x**
  - **a is the Estimate of Project Cost when $x = 0$, Typically Representing Start-Up Costs of Development or Production**
  - **b is a Factor that Estimates Cost per Cost-Driver Unit, e.g., a Pound, a Lines of Code, a Watt**
- **The LLOLS Nonlinear Power Regression Form $y = ax_i^b$ Does Not Accommodate a Fixed-Cost Term**
  - **It Cannot, because the Logarithm of $y = a+bx_i^c$ is Not Compatible with Algebraic Calculations**
  - **Therefore CER Development by OLS Regression Puts Us in a Contradictory Position: Linear CERs Can Accommodate a Fixed-Cost Term, but Power CERs Cannot**

MCR
Critical Thinking. Solutions Delivered.

"Fixed-Cost" Term Can be Negative

- **a Can be Negative in Situations where There is a Natural Limit on How Low the Cost-Driver Value Can Go, e.g., Telescope Focal Length that Must be Above a Certain Numerical Value to be Effective**
- **Of Course, in such Cases, a does not Actually Represent a "Cost," but is Merely an Unintended Consequence of Data Base Relationships that Result from Practical Limits on Cost Driver Values**
  - **But Even a Negative Value of a is a Necessary Part of a CER that Models Cost Behavior above the Cost Driver's Minimum Practical Limit**
  - **An Example of This Phenomenon is our Concave Data Set, in which the Cost-Driver Value Apparently Cannot Go Much Below $x = 8$, so a Fixed Cost a (were one to exist) would be Negative**

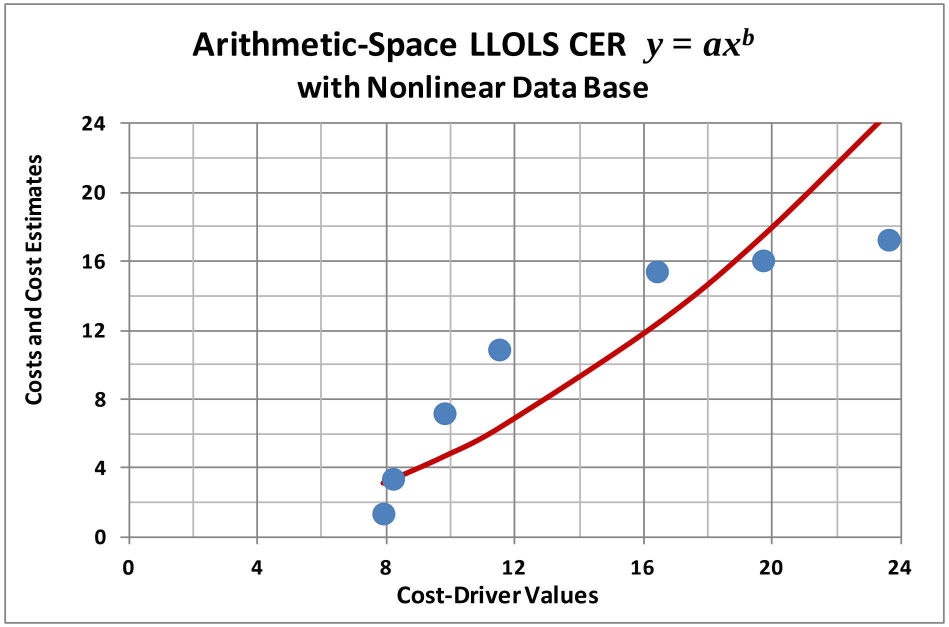
MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Footnote on CER Fixed-Cost Terms

- We Know from Basic Statistics that Regression Lines Derived from Data Need Not Pass through Any Actual Points in the Data Set
- However, if a CER has a Fixed-Cost Term a that is Calculated from the Data, the CER then Must Pass through the Point $(0,a)$ on the Vertical Axis
- Similarly, if a CER has no Fixed-Cost Term (which is vacuously not calculated from the data), then the Fixed Cost is *a fortiori* Zero
  - Therefore the CER Must Pass through the Point (0,0) on the Vertical Axis, namely the Point of Intersection of the Vertical and Horizontal Axes
  - To Put it More Starkly, an LLOLS CER Need Not Pass through Any Actual Data Points, but Must Pass through a Point that is not an Actual Data Point, namely (0,0)

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS CER Based on Nonlinear Concave Data Set



- **Notice the Disconnect between Concave Shape of the Data vs. Convex Shape of the CER due to Lack of a Fixed-Cost Term**
  - **Following the Data Pattern as it Moves toward the Vertical Axis (where the cost-driver value would be zero), we Note that a Fixed-Cost Term a Should Really be Negative**
  - **The LLOLS CER, however, is on Track to Intersect the Vertical Axis at the Zero Point as Log-Log OLS Requires it to Do**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

MCR

Critical Thinking. Solutions Delivered.

The Starting Point: Standard Error of an LOLS CER

- In Log-Log Space the Logarithmic OLS (LOLS) CER has the form $\log y = \log a + b \log x$
- The LOLS CER's Standard Error of Estimate is

$$LSEE = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\log y_i - \log a - b\log x_i\right)^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} E_i^2}$$

- Recall the Nonlinear Concave Data Set's LOLS CER:

$$\log y = -1.218 + 1.901 \log x$$

- … and its Quality Metrics:

Std Error (Log Space) LSEE =	0.2236
Bias (Log Space) LB =	0.0000
R^2 (Log Space) LR^2 =	75.60%

- To Get Us Started, We Know that the Standard Error of the LLOLS Power CER (the exponentiation of the LOLS CER) is Multiplicative, i.e., a Percentage of the Estimate, rather than a Constant Dollar Value

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

One Rumor about Standard Errors of LLOLS Power CERs

- **There Have Been Persistent Hard-to-Kill Oddball Rumors Regarding the Standard Error of the Derived LLOLS Power CER $y = ax^b$, which is $y = 0.60x^{1.901}$**
- **The First Rumor Instructs Us to Note the Log-Space LOLS *LSEE*, which in this Case is *0.2236*, and First Calculate $exp(0.2236) = 1.2505$ and $exp(-0.2236) = 0.7966$**
  - **Then the "Upper Standard Error" of the LLOLS Power CER is Said to be $exp(0.2236) -1= 0.2505 = 25.05\%$**
  - **And the "Lower Standard Error" of the CER is Said to be $exp(-0.2236)-1 = 0.7966-1 = -20.34\%$**
- **Although This Method is Occasionally Used, No Justification of its Validity Appears to be Available and Not a Lot is Known about its Origin**
- **However, it is Surely Incorrect, because We Know that the Percentage Standard Error is Really 51.17%**

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

A Currently Being Used Example of the First Rumor

The cost estimating relationship for xxxxxxxx is shown below with its statistics.

CER: Lot \$ = 61.789 (xxxxxxxxxx)$^{0.750}$ (xxxx)$^{-0.397}$ Qty$^{0.930}$ e$^{-0.223(\text{Follow-on/Prod})}$ e$^{0.832(\text{xxxxxxxxxxxx})}$ e$^{0.509(\text{xxxxxxxxxxxxxx})}$

T Statistics: (1.11) (6.22) (-3.12) (fixed) (fixed) (2.12) (4.43)

Statistics: R^2 = 98.6%
s = 0.147 (+15.8%, -13.7%)
(12 Data Points) (7 Degrees of Freedom)

Where:

Lot Cost = The manufacturing (recurring) cost of the units built in FY08\$K.
xxxxxxxxxx = The number of radiating elements in the xxxxxxx.
xxxx = The xxxx average xxxxxxx in xxxxxxxxxx.
Qty = The quantity of units manufactured in development or production
Follow-on/Prod =1 for follow-on developments, a xxxxxx development that utilizes man the same components/assemblies developed in a prior program (i.e. where a significant portion of the xxxxxx design is from a prior program or hardware developed in a prior program is reused in ne program) or for estimating the first lot of production, and
= 0 for a new development program.
xxxxxxxxxxxx = 1 for xxxxxxxxxxxxxxxx, and
= 0 for xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxx.
xxxxxxxxxxxxxx = 1 for a xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxx
= 0 for a xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

MCR CRITICAL THINKING. SOLUTIONS DELIVERED.

Another, Perhaps More Popular, Rumor about the Standard Error

- **The Second Rumor Advises Us to Interpret *LSEE* of the Log-Space LOLS $\log y = \log a + b \log x$, which in our Concave Data-Set Example is *0.2236*, as an Approximation to the Percentage Error**
- **If We Were to Follow this Advice, We Would Approximate the Percentage Standard Error of the LLOLS Power CER $y = ax^b$ as *22.36%***
- **Unfortunately, the Method Recommended by this Rumor is also Incorrect, the Mathematics of which was Carefully Explained by P. Young in 1999 (of course, we already know it's incorrect because 22.36% is not a "good" approximation to 51.17%)**

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Young's Specific Problem with the Second Rumor

- **Full Details are Available in Young's 1999 Article*, but We Will Sketch the Main Points Here**
- **If the Second Rumor were Correct, the LLOLS CER's Standard Error would be the Quantity, Expressed as a Percentage,**

$$LSEE = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\log y_i - \log a - b\log x_i\right)^2}$$

- **However, the Number we Really Want to Know, Expressed as a Percentage, is**

$$\sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\frac{y_i - ax_i^b}{ax_i^b}\right)^2}$$

- **Young Shows Mathematically that**
  - **If $LSEE$ is "small," then the Two Percentages will be Fairly Close to Each Other, so the Error Made by the Approximation would be "acceptable"**
  - **However, if $LSEE$ is "large," then "the overall approximating error can be very bad, and the use of [$LSEE$] will be effectively invalid and *meaningless*." (Young, page 64)**

***P.H. Young, P. , "The Meaning of the Standard Error of the Estimate in Logarithmic Space," *The Journal of Cost Analysis & Management*, Winter 1999, pages 59-65.**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

The Solution to the Rumor Problem

- **Now that Estimating the Percentage Standard Error of an LLOLS Power CER $y = ax^b$ as Proposed by the Rumors (rumor-mongers?) Have Been Demonstrated to be Incorrect, What Can We Do?**
- **The Only Way to Calculate the Percentage Standard Error Correctly is to Move the Problem Back into Arithmetic Space and Redo All the Calculations**
  - **Ignore the Logarithmic Space Quality Metrics (i.e., those associated with the LOLS CER $\log y = \log a + b \log x$)**
  - **Calculate the $ESTy$ (not $ESTY$) Values, this Time Using the LLOLS Power CER $y = ax^b$**
  - **Set Up the Computational Table and Calculate the Percentage Standard Error Using the Correct Formula**

$$\sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\frac{y_i - ax_i^b}{ax_i^b}\right)^2}$$

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

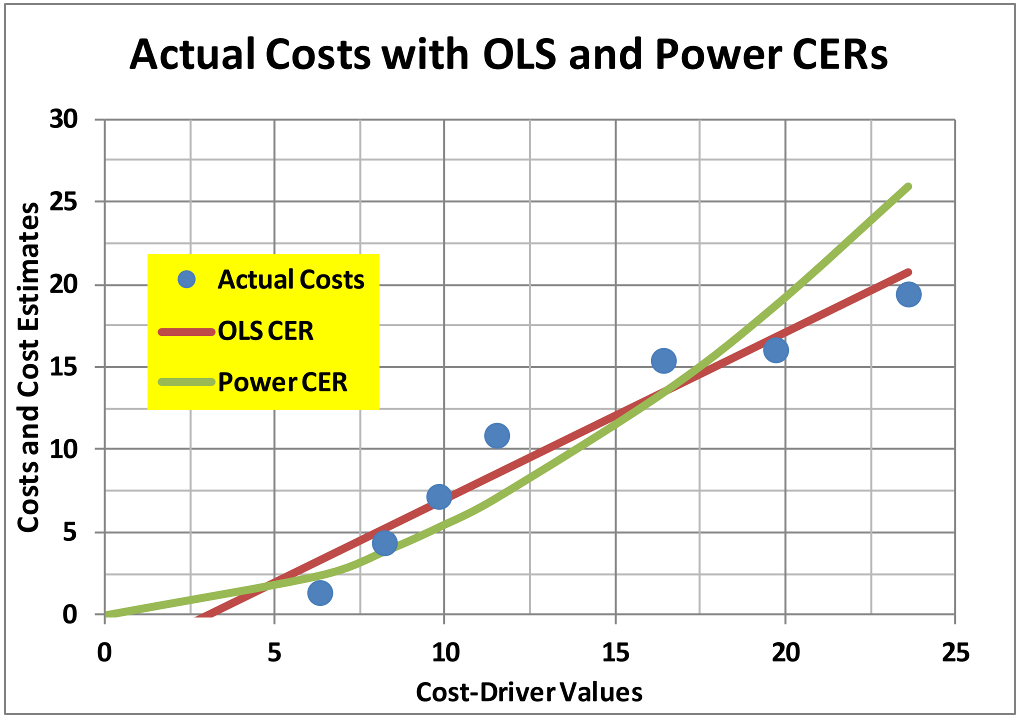
Identifying the "Best" CER

- There is no Universally Accepted Definition of How to Determine which is the "Best" of the Many Possible CERs that May be Calculated for Any Data Set You Happen to be Working with
- Most Developers First Compare the Quality Metrics of Several Candidate CERs and then Assess the Result One Judged "Best" for "Reasonableness"
- Quality Metrics Usually Compared are …
  - Standard Error
  - Bias
  - R^2
- Metrics such as t and F Scores as Applied to LOLS CERs Do Not Offer Relevant Information Regarding the Estimating Capability of LLOLS Power CERs

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Example: A "Nearly Linear" Data Set

n	x Values (Cost Driver)	y Values (Actual Costs)
7	6.3	1.380
	8.2	4.395
	9.8	7.201
	11.5	10.900
	16.4	15.434
	19.7	16.074
	23.6	19.453
Sums =	95.500	74.837



Without Converting the LOLS CER into an LLOLS Power CER in Arithmetic Space, An Analyst Might be Tempted (and many have) to Choose the LOLS CER as a Starting Point, Being Impressed by the Very Low Standard Error

Comparison of Reported Quality Metrics		
Quality Metric	OLS CER	LOLS CER
Standard Error	1.792	0.165
Bias	0.000	0.000
R^2	93.98%	86.25%

But Back in Arithmetic Space …

- … the Tale is Different

Quality Metric	OLS CER Dollar Values	LOLS CER Log-Dollar Values	LLOLS Power CER Dollar Values	LLOLS Power CER Percentages
Standard Error	1.792	0.165	3.806	37.79%
Bias	0.000	0.000	0.287	-5.04%
R^2	93.98%	86.25%	87.51%	

- It Turns out that the Very Impressive 0.165 is Denominated in “Log-Dollars,” not “Dollars,” because the LOLS CER Exists in Logarithmic Space
- When it is Transformed into the LLOLS Power CER in Arithmetic Space, the Standard Error Becomes 3.806 Dollars, which Compares Unfavorably with the OLS CER’s 1.792 Dollars
- So, after all, the OLS CER is Preferred if the Criterion for “Best” is Standard Error (by itself or even with R^2)

CRITICAL THINKING. SOLUTIONS DELIVERED.

As for Percentage Standard Errors …

- **An Analyst Might Apply the Incorrect Assumption that LOLS CER *LSEE = 0.165* Implies a Percentage Standard Error of 16.5% in the Resulting LLOLS Power CER**
- **Such a Person Would then Mistakenly Assume that the LLOLS Power CER Had a Smaller Percentage Standard Error than it Actually Does – its Correct Percentage Standard Error is 37.79%, not 16.5%**
- **To Summarize: An Analyst Using the Logarithmic OLS Method of CER Derivation, along with all the Faulty "Folklore" that Goes with it, Would Conclude that the LLOLS Power CER Better Fits the Data Set than Does the OLS Linear CER, Even though both the Graphics and a Correct Calculation of the Standard-Error Statistics Show Otherwise**

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Contradictory Constraints on OLS/LOLS/LLOLS Error Models

- **OLS CERs Have Additive-Error Models**
  - $y = a + bx + \varepsilon$
  - **Errors are Expressed in Terms of Dollars**
- **LOLS CERs Have Additive-Error Models**
  - $\log y = \log a + b \log x + \log \varepsilon$
  - **Errors are Expressed in Terms of Log-Dollars**
  - **Log-Dollars are Much Smaller in Magnitude than Dollars**
- **LLOLS CERs Have Multiplicative-Error Models**
  - $y = ax^b\varepsilon$
  - **Errors are Expressed as a Percentage of the Estimate**
- **These Three Kinds of Errors are Incommensurable, i.e., their Magnitudes Cannot be Directly Compared with Each Other**
- **This Circumstance Has Had Some Unfortunate and Embarrassing Consequences**

MCR
CRITICAL THINKING.
SOLUTIONS DELIVERED.

A Page from USAF's Unmanned Space Vehicle Cost Model, Version 6 (1988)

- **An USCM6 Linear OLS CER with its Quality Metrics:**

NONRECURRING CER
USCM6

SUBSYSTEM: THERMAL

COMPONENT: SUBSYSTEM

FUNCTION: Y = 297.69 + (.0011 * X1)

WHERE: Y = TOTAL NONRECURRING, FY86$ (K)

OBSERVATIONS: 14 R-SQUARE: .89

STANDARD ERROR: 776 ADJ R-SQUARE: .88

AVG Y VALUE: 1578

CORRELATION MATRIX

	NONREC	THERWT	SCBOL	SCWT	BOL
NONREC	1.0000	-.0620	.9130	.8600	.7720
THERWT	-.0620	1.0000	-.0220	.0210	-.0000

- **Note the Standard Error Value of 776 for This CER**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

The Very Next Page from that Very Same Cost Model Document

- **An USCM6 LLOLS Power CER with its Quality Metrics:**

RECURRING CER
USCM6

SUBSYSTEM: THERMAL

COMPONENT: SUBSYSTEM

FUNCTION: $Y = 13.07 \cdot (X1^{.77}) \cdot (1.99^{X2}) \cdot [1.21]$

where: Y = FIRST UNIT COST, FY86$ (K)

\# OBSERVATIONS: 28 R-SQUARE: .78

STANDARD ERROR: .65 ADJ R-SQUARE: .76

CORRELATION MATRIX

	LNT1L95	LNSCWT	LNTHERWT
LNT1L95	1.0000	.6480	8510
LNSCWT	.6480	1.0000	.5120
LNTHERWT	.8510	.5120	1.0000

- **Note the Standard Error Value of .65 for This CER**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

What About these Two CERs?

- Those Two Pages would lead a Reasonable Reader Inexorably to the Conclusion that Estimating Using the OLS CER is Much More Error-Prone than Using the LLOLS CER
  - After All, the OLS CER Has Standard Error 776, while the LLOLS CER Has Standard Error 0.65
  - The Remaining Items on Both Pages are Basically Similar
- In Fact, an Analyst May Ask, "Why Didn't the AF Derive an LLOLS CER for the Nonrecurring Case, rather than Stick with that Lousy OLS CER?"
- The Pages do not Make Clear that the OLS Error is Expressed in Dollars, while the LLOLS Error is Expressed in Log-Dollars
  - That's the Source of the Entire Apparent Contradiction
  - Of Course, We are not Even Mentioning the Fact that the Real Standard Error of the LLOLS CER is Actually not 0.65 and is Most Likely Nowhere Near it

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

LLOLS Denies Freedom of Choice to the CER Developer

- **The USCM6 (1988) Experience Occurred because the CER Developers at the Time were Painted into a Corner by the LLOLS CER-Development Process**
  - **If a CER Developer Wants to Derive a Linear CER of the Form $y = a + bx$, then he or she Must Use the Additive-Error Model**
  - **If a CER Developer Wants to Derive a Nonlinear Power CER of the Form $y = ax^b$, then he or she Must Use the Multiplicative-Error Model**
- **What if the Analyst Needs a Linear CER with a Multiplicative-Error Term or a Power CER with an Additive-Error Term?**
  - **Tough Luck! Log-Log OLS Can't Handle those Issues**
  - **Don't Feel Bad – The Problem is Log-Log OLS, not You**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

The Good News About USCM

- In the Early 1990s, under the Direction of SMC/FMC Cost Chief David L. Hansen, the USAF Ordered the OLS/LLOLS Inconsistency Corrected by Changing the Method by which USCM CERs were Developed
- The Method that Replaced Log-Log OLS, Applied in USCM7 (1994) and Later Versions of USCM, is “Iteratively Reweighted Least Squares” (IRLS)
- IRLS (called “MUPE” by the USAF) is One of the CER-Development Methods Available that Eliminate the Three Undesirable Aspects of Log-Log OLS
  - Power CERs are now Permitted to Include Fixed-Cost Terms
  - Standard Errors Can be Expressed in either Additive-Error or Multiplicative-Error Forms, at the Analyst’s Option
  - CERs of All Algebraic Forms are Permitted, Even if they are not Compatible with Logarithms

MCR

CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Curvilinear Cost-Estimating Relationships (CERs)
- Summary

General Multiplicative-Error Model Eliminates All These Problems

- No Logarithms
  - Functional Forms Predict Cost, Not Logarithm of Cost
  - Standard Errors Can Be Compared and Ranked in Magnitude for All Functional Forms (no 776 vs. 0.65 problem)
  - Error Model (Additive or Multiplicative) Can Be Chosen Independently of Functional Form
  - Unfortunately, there are no Formulas for the Coefficients as there are for a and b as in the OLS Case
- Make Use of Modern Computing Capability
  - Fortunately, in the 21st Century, the Least-Squares Minimization Problem Does Not Have to Be Solved Explicitly
  - Sequential-search Techniques Based on Newton's Method or Simplex Method Are Used to Find Error-minimizing Values of a and b (*Excel Solver* works OK for me, although some theorists prefer more high-class mathematics software)
  - All Functional Forms Can Be Considered, Even $y = a + bx^c$

Scenario: A Simple First Step to a Multiplicative-Error Model

- **Actual Cost Equals Estimate times Error:**

$$y = f(x) \times \varepsilon$$

- **Error is Ratio of Actual to Estimate, namely**

$$\varepsilon = \frac{y}{f(x)} = \frac{\text{Actual}}{\text{Estimate}}$$

- **Minimum Percentage Error (MPE) CERs: Choose *f(x)*'s Coefficients so that Sum of Squared Percentage Errors**

$$\sum(\varepsilon_i - 1)^2 = \sum\left[\frac{y_i - f(x_i)}{f(x_i)}\right]^2$$

  is as Small as Possible

- **Actual Cost = Estimate ± Percentage of Estimate**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

1st Quality Metric: Percentage Standard Error

- **"One-Sigma"-type Error Bound that Models the Error Term of Multiplicative-Error CERs of any Algebraic Form $y = f(x)$**
- **%SEE =** $\sqrt{\frac{1}{n-k}\sum_{i=1}^{n}\left[\frac{y_i - f(x_i)}{f(x_i)}\right]^2}$ **x 100%**
  - **$y = f(x)$ is the CER that Expresses Cost (y) in Terms of a Technical or Programmatic Cost Driver (x)**
  - **n is the Number of Data Points Used to Derive the CER**
  - **k is the Number of Coefficients in the CER's Algebraic Expression, e.g., $k = 2$ for the CER $y = ax^b$ and $k = 3$ for the CER $y = a + bx^c$**
- **Percentage Standard Error is a CER Quality Metric**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

2nd Quality Metric: Percentage Bias

- **Tecolote Analysts Found that the MPE Procedure Tends to Produce a Positive Sample Percentage Bias**

$$\frac{1}{n}\sum\left[\frac{f(x_i)-y_i}{f(x_i)}\right]$$

  - **This is Bad?**
  - **Bias Seems to be Attributable to the Fact that** $\sum\left(\frac{y_i - f(x_i)}{f(x_i)}\right)^2$ **will be Smaller if the** $f(x)$ **Values are Larger**
- **Percentage Bias is a CER Quality Metric**
  - **USCM's IRLS is a 1974-Vintage Statistical Regression Method that Reduces Percentage Bias "Essentially" to Zero, while "Almost" Minimizing Percentage Error**
  - **A Method of 1998 Vintage, Somewhat Easier to Understand, that Does a Better Job on Both those Issues is Called "Zero Percentage Bias, Minimum Percentage Error" (usually denoted "ZMPE" and pronounced "zimpy")**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

3rd Quality Metric: R^2 Between Estimates and Actuals

- R Denotes the Correlation Between Actuals (x values) and Estimates (y values), Measuring the Extent to which Relationship between x and y is Linear
  - R Does not Depend on Specific Coefficients or Form of Relationship
  - R^2 = Proportion of Variation in Estimates (y) that is Attributable, through a OLS Linear Relationship, to Variations in Actuals (x)
  - Larger (closer to 1.00) Values of R^2 Indicate Better Linear Fit
- If the CER is "Good", Estimates Should be Pretty Close to Actuals, i.e., the (Actual, Estimate) = (x,y) Points Should Lie Along Straight Line $y = x$
- R^2 ("Pearson's Correlation Squared") is a CER Quality Metric

$$R^2 = \frac{\left[n\sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k \sum_{k=1}^{n} y_k \right]^2}{\left[n\sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k \right)^2 \right]\left[n\sum_{k=1}^{n} y_k^2 - \left(\sum_{k=1}^{n} y_k \right)^2 \right]}$$

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Footnote: R^2 Can be Deceptive

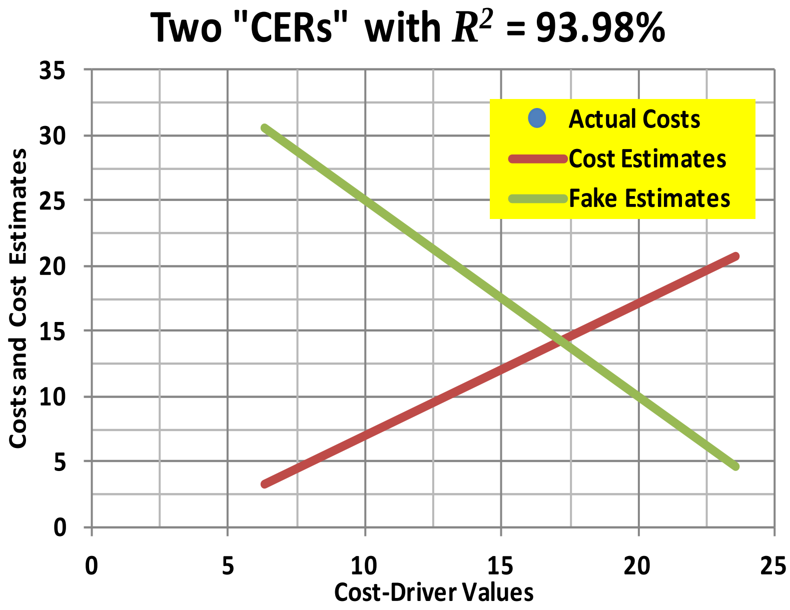
- R^2 Measures Only the Extent of Linearity in the (x,y) Data Set – No More and No Less
- This Means that R^2 Will Have Exactly the Same Numerical Value for Every Linear Relationship between x and y – whether it's a "Best-Fit" Regression Line or Just Any Old Line
- For Example, the Regression Line for the Set of Linear Data is $y = -3.017 + 1.005x$
- The R^2 Value for this Regression Line is 93.98%
- However R^2 also Equals 93.98% for the Straight Line $y = 40 – 1.5x$, which is Located Nowhere Near the Data Set, Graphical Details of which are Coming Up Next

Cost Driver	*Actual Costs*
6.3	1.380
8.2	4.395
9.8	7.201
11.5	10.900
16.4	15.434
19.7	16.074
23.6	19.453

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Footnote Continued: R^2 is the Same for all Straight-Line Relationships

Cost Driver	*Actual Costs*	*Cost Estimates*	*Fake Estimates*
6.3	1.380	3.283	30.550
8.2	4.395	5.200	27.700
9.8	7.201	6.814	25.300
11.5	10.900	8.529	22.750
16.4	15.434	13.473	15.400
19.7	16.074	16.802	10.450
23.6	19.453	20.737	4.600



- R^2 is the Same, namely 93.98%, for both “CER” Lines – Try it!
- Lesson to be Learned: Although R^2 is a Useful Metric, its Report is Very Specific – R^2 Cannot be Used by Itself to Measure CER “Goodness,” but Must be Used Only in Concert with Standard Error or a Similar Metric

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

ZMPE CER-Regression Technique

- **The ZMPE Technique was Developed* to Yield CERs Guaranteed to Have Minimum Possible Percentage Error among all Percentage-Unbiased CERs for a Given Data Set that Have the Algebraic Form being Considered**
- **ZMPE Pursues the Minimum-Percentage-Error Goal Directly (IRLS works in a different, slightly more complicated, way)**
  - **ZMPE Computes Minimum-Percentage-Error CER, Subject to Constraint that Percentage Bias be Exactly Zero**
  - **CERs are Derived Using "Constrained Optimization" – *Excel Solver* Provides the Capability to Carry this Process Out**

* **Book, S. and Lao, N, "Minimum-Percentage-Error Regression under Zero-Bias Constraints", *Proceedings of the Fourth Annual U.S. Army Conference on Applied Statistics, 21-23 October 1998*, U.S. Army Research Laboratory, Report No. ARL-SR-84, November 1999, pages 47-56.**

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Calculating CERs of the Algebraic Form $y = a + bx^c$ by ZMPE

- **USCM Refers to $y = a + bx^c$ as the "Triad" Form, an Algebraic Form that LLOLS Cannot Resolve**
- **ZMPE is Tasked to Minimize the Sum of Squared Percentage Errors, namely**

$$F(a,b,c) = \sum_{k=1}^{n} \left(\frac{y_k - a - bx_k^c}{a + bx_k^c} \right)^2 ,$$

- **… subject to the Constraint that the Summed Percent Bias Equals Zero, namely**

$$Summed\,\%Bias(a,b,c) = \sum_{k=1}^{n} \left(\frac{a + bx_k^c - y_k}{a + bx_k^c} \right) = 0$$

- **All This Automatically Minimizes the Percentage Standard Error and Bias, which are, Respectively,**

$$\%SEE = \sqrt{\frac{1}{n-3} \sum_{i=1}^{n} \left(\frac{y_i - a - bx_i^c}{a + bx_i^c} \right)^2} \quad \textbf{and} \quad \%Bias = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - a - bx_i^c}{a + bx_i^c} \right)$$

ZMPE CER for Nonlinear Concave Data Set

- **Based on Computations on the Historical Data …**

$$a = -660.62\ ; \quad b = 626.78\ ; \quad c = 0.027$$

- **Multiplicative-Error CER**

$$y = -660.62 + 626.78x^{0.027}$$

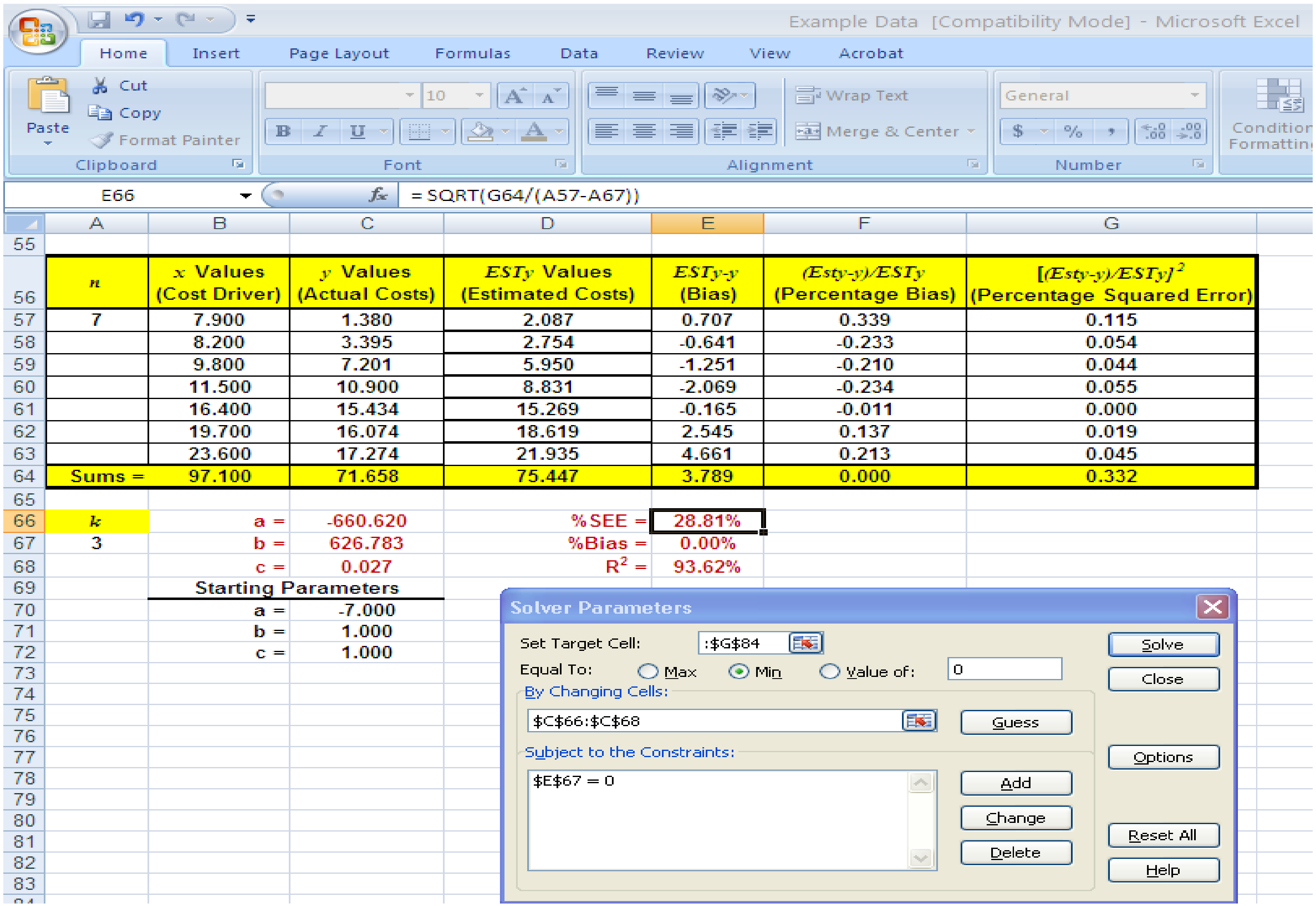
- **Standard Error of the Estimate *(%SEE)***

$$Standard\ Error = \sqrt{\frac{1}{n-3}\sum_{i=1}^{n}\left(\frac{y_i - a - bx_i^c}{a + bx_i^c}\right)^2} = \sqrt{\frac{1}{7-3}(0.3321)} = 0.2881$$

(Average 28.81% Across the Data Range)

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Screen Shot of Excel Spreadsheet that Implements ZMPE



Example Data [Compatibility Mode] - Microsoft Excel

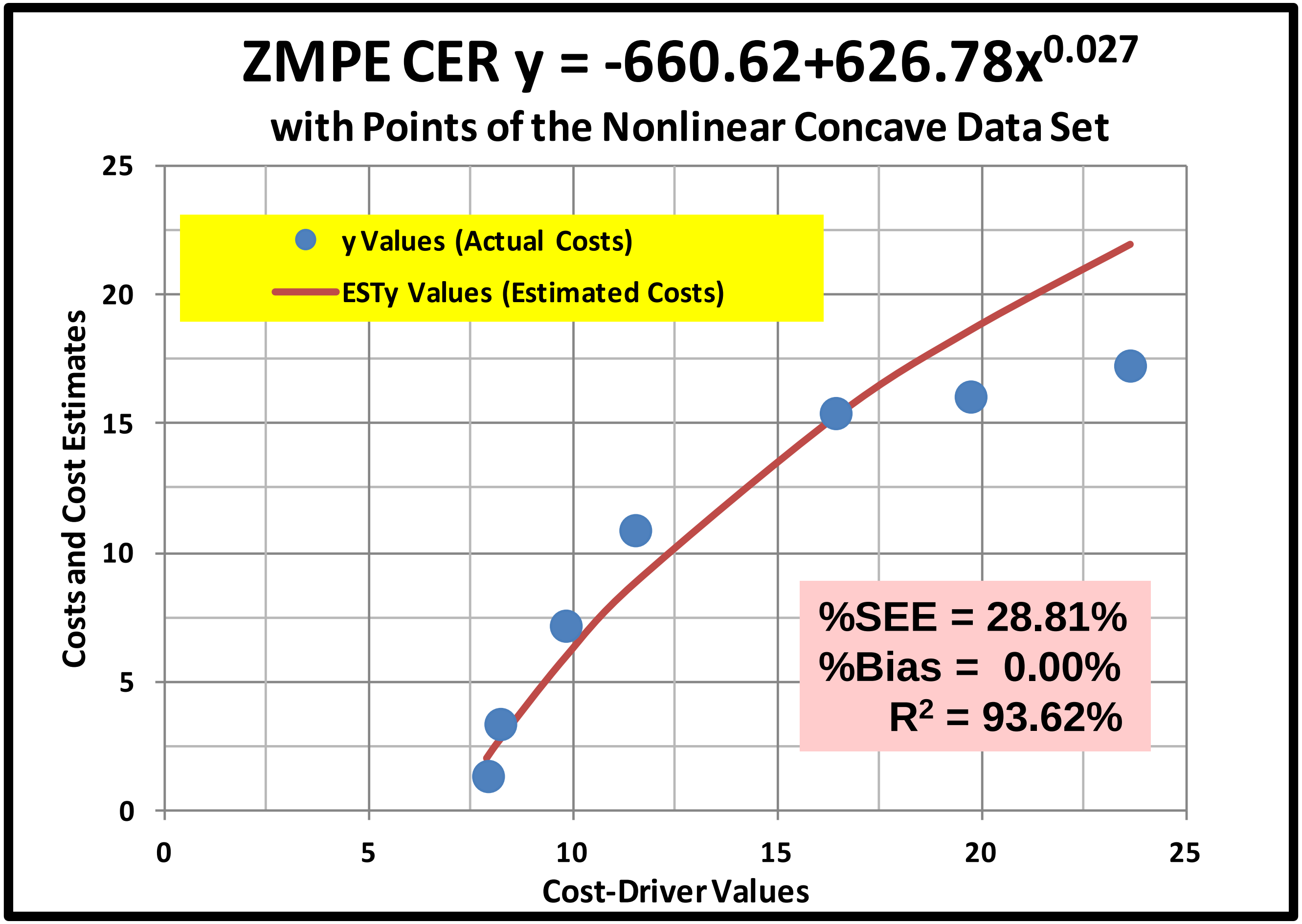
E66 | = SQRT(G64/(A57-A67))

n	x Values (Cost Driver)	y Values (Actual Costs)	ESTy Values (Estimated Costs)	ESTy-y (Bias)	(Esty-y)/ESTy (Percentage Bias)	$[(Esty-y)/ESTy]^2$ (Percentage Squared Error)
7	7.900	1.380	2.087	0.707	0.339	0.115
	8.200	3.395	2.754	-0.641	-0.233	0.054
	9.800	7.201	5.950	-1.251	-0.210	0.044
	11.500	10.900	8.831	-2.069	-0.234	0.055
	16.400	15.434	15.269	-0.165	-0.011	0.000
	19.700	16.074	18.619	2.545	0.137	0.019
	23.600	17.274	21.935	4.661	0.213	0.045
Sums =	97.100	71.658	75.447	3.789	0.000	0.332
k	a =	-660.620	%SEE =	28.81%		
3	b =	626.783	%Bias =	0.00%		
	c =	0.027	R^2 =	93.62%		

Starting Parameters

a =	-7.000
b =	1.000
c =	1.000

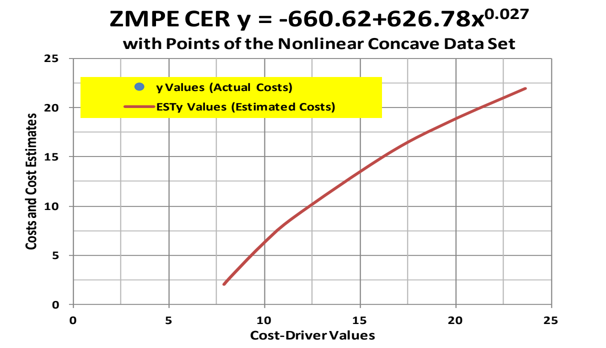
ZMPE CER and Quality Metrics for the Nonlinear Concave Data Set

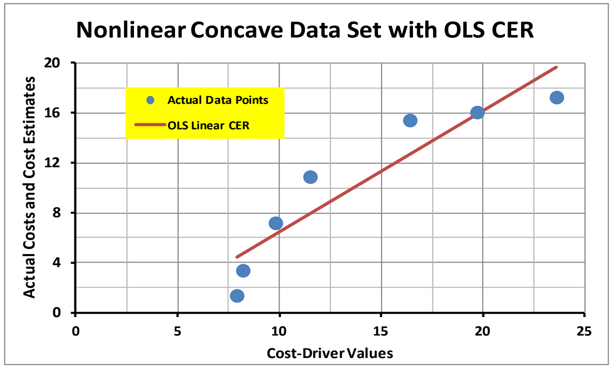


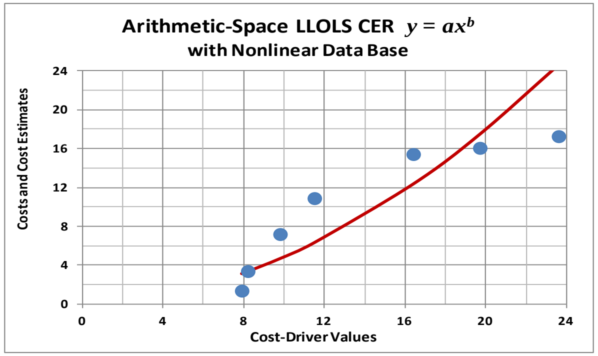
MCR
CRITICAL THINKING.
SOLUTIONS DELIVERED.

Comparison of Candidate CERs

- **ZMPE CER:** $y = -660.62 + 626.78x^{0.027}$
  - ***%SEE = 28.21% of the estimate***
  - ***%Bias = 0.00% of the estimate***
  - ***R^2 = 93.62%***
- **Linear CER:** $y = -3.207 + 0.969x$
  - ***SEE = 2.613* Dollars**
  - ***Bias = 0.000* Dollars**
  - ***R^2 = 86.10%***
- **LLOLS CER:** $y = 0.060x^{1.901}$
  - ***%SEE = 51.17% of the estimate***
  - ***%Bias = -9.06% of the estimate***
  - ***R^2 = 77.63%***



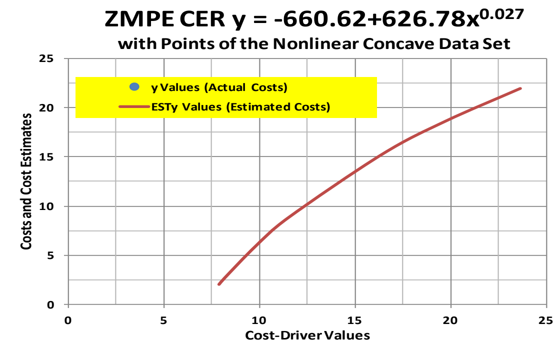


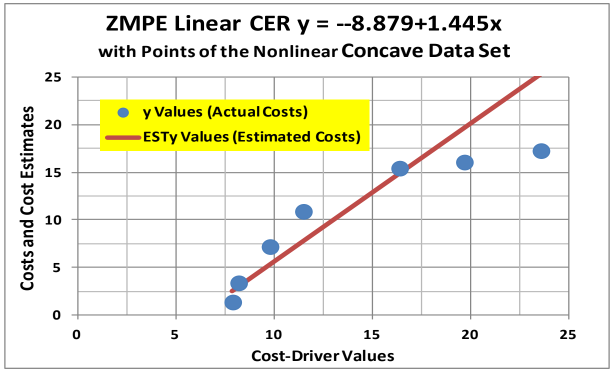


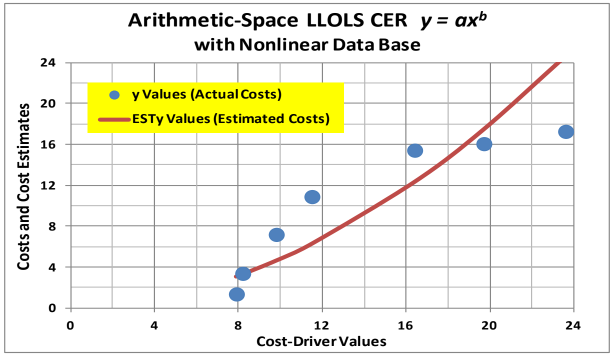
MCR
CRITICAL THINKING.
SOLUTIONS DELIVERED.

To be Fair, a Comparison of Only Multiplicative-Error CERs

- **ZMPE CER:** $y = -660.62 + 626.78x^{0.027}$
  - *%SEE = 28.21% of the estimate*
  - *%Bias = 0.00% of the estimate*
  - *R^2 = 93.62%*
- **ZMPE Linear CER:** $y = -8.879 + 1.445x$
  - *%SEE = 36.31% of the estimate*
  - *%Bias = 0.00% of the estimate*
  - *R^2 = 86.10%*
- **LLOLS CER:** $y = 0.060x^{1.901}$
  - *%SEE = 51.17% of the estimate*
  - *%Bias = -9.06% of the estimate*
  - *R^2 = 77.63%*



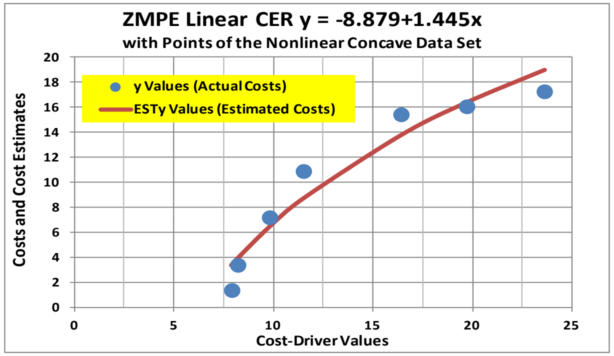


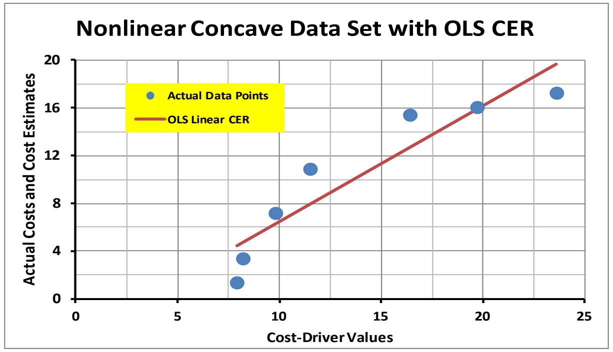


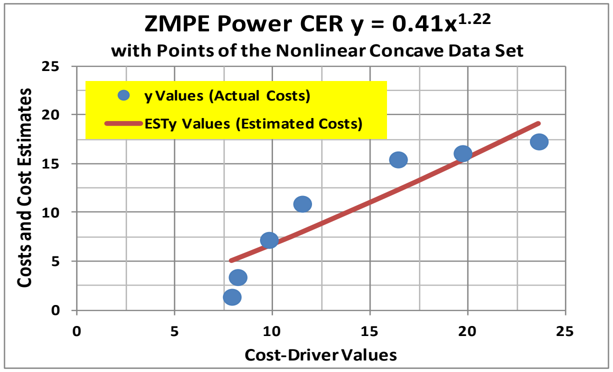
CRITICAL THINKING. SOLUTIONS DELIVERED.

To be Fair, a Comparison of Only Additive-Error CERs

- **ZMPE CER:** $y = -988.54 + 963.11x^{0.014}$
  - SEE = ***1.966*** **Dollars**
  - $Bias$ = ***0.000*** **Dollars**
  - $R^2 = 93.70\%$
- **OLS Linear CER:** $y = -3.207 + 0.969x$
  - SEE = ***2.613*** **Dollars**
  - $Bias$ = ***0.000*** **Dollars**
  - $R^2 = 86.10\%$
- **ZMPE Power CER:** $y = 0.410x^{1.215}$
  - SEE = ***2.284*** **Dollars**
  - $Bias$ = ***0.000*** **Dollars**
  - $R^2 = 84.16\%$







MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Contents

- Ordinary Least Squares (OLS) Regression
- The Logarithmic Transformation
- Log-Log OLS (LLOLS) Nonlinear Regression
- Five Unfortunate Characteristics of LLOLS
  - Error of Estimating Cost is Not Minimized
  - LLOLS CERs are Biased (usually low, but sometimes high)
  - When Bias is "Corrected" to Zero, the CER's Standard Error and R^2 Must be Recalculated
  - CERs Derivable by LLOLS Must Have Fixed Cost = Zero
  - Although the LLOLS Error Model is Multiplicative, the Reported Standard Error Has No Meaning
- A Few More Negatives
- A Better Way to Derive Nonlinear Cost-Estimating Relationships (CERs)
- Summary

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Summary: The Bad News

- CERs are Derived by Applying Statistical Analysis to a Cost Data Base that Reflects Historical Cost Experience
  - Objective is to Minimize the Error One Makes in Trying to Re-Estimate the Actual Costs in the Data Base Logarithmic OLS Regression
  - Unfortunately Logarithmic OLS Regression Minimizes the Wrong Error – the Error of Estimating the Logarithms of the Actual Costs
- The Path One Needs to Take to Calculate the Correct Magnitudes of the Right Error is not a Direct One
  - The Errors that Have Allegedly been Minimized in the Log-Log Step Have to be Recalculated in order to Apply to the Error of Estimating the Costs, not their Logarithms
  - There is no "Quickie" Way to Calculate the True Metrics

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

Summary: The Good News

- **General-Error Regression Eliminates All Problems Caused by Log-Log OLS Regression**
  - **Multiplicative-Error CERs May be More Appropriate than Additive-Error CERs, but Analysts Should Have a Choice**
  - **IRLS (aka "MUPE") and ZMPE Allow CERs of Any Algebraic Forms and Either Error Model to be Derived**
- **CER Quality Metrics Support Credibility of Estimates**
  - **Percentage or Dollar-Valued Standard Error of the Estimate**
  - **Percentage or Dollar-Valued Bias of the Estimate**
  - **Pearson's Correlation Squared between Estimates and Actuals**
- **Apparent (but not real) Downside: There are no Explicit Formulas that Can be Used to Calculate the Coefficients and Exponents – Use Your Computer**

MCR
Critical Thinking. Solutions Delivered.

Acronyms

BAF	Balance-Adjustment Factor
CAIG	Cost Analysis Improvement
CAPE	Cost Assessment and Program Evaluation
CER	Cost Estimating Relationship
DoDCAS	Department of Defense Cost Analysis Symposium
FMC	Financial Management, Cost
IRLS	Iteratively Reweighted Least Squares
LLOLS	Logarithm-Logarithm (Log-Log) Ordinary Least Squares
LOLS	Logarithmic Ordinary Least Squares
LSEE	Logarithmic Standard Error of the Estimate
M	Millions (usually of dollars)
MPE	Minimum Percentage Error
MUPE	Minimum Unbiased Percentage Error
OLS	Ordinary Least Squares
OSD	Office of the Secretary of Defense
SEE	Standard Error of the Estimate
SMC	(USAF) Space and Missile Systems Center
USAF	United States Air Force
USCM	Unmanned Space Vehicle Cost Model
ZMPE	Zero Percentage Bias, Minimum Percentage Error

MCR
CRITICAL THINKING. SOLUTIONS DELIVERED.

References, Chart 1 of 2

- **Book, S. and N. Lao (1999), "Minimum-Percentage-Error Regression under Zero-Bias Constraints",** ***Proceedings of the Fourth Annual U.S. Army Conference on Applied Statistics, 21-23 October 1998*****, U.S. Army Research Laboratory, Report No. ARL-SR-84, November 1999, pages 47-56.**
- **Book, S. and P. Young (1997), "General-Error Regression for Deriving Cost-Estimating Relationships",** ***The Journal of Cost Analysi*****s, Fall 1997, pages 1-28.**
- **Book, S. and P. Young (1990), "Optimality Considerations Related to the USCM-6 'Ping Factor'," ICA/NES National Conference, June 1990, 40 charts.**
- **Bradu, D. and Y. Mundlak (1970), "Estimation in Lognormal Linear Models,"** ***Journal of the American Statistical Association*****, Vol. 65, March 1970, pages 198-211.**
- **Duan, N. (1983), "Smearing Estimate: A Nonparametric Transformation Method,"** ***Journal of the American Statistical Association*****, Vol. 78, September 1983, pages 605-610.**
- **Eskew, H. and K. Lawler (1994), "Correct and Incorrect Error Specifications in Statistical Cost Models,"** ***Journal of Cost Analysis*****, Spring 1994, page 105-123.**
- **Heien, D. (1968), "A Note on Log-Linear Regression,"** ***Journal of the American Statistical Association*****, Vol. 63, September 1968, pages 1034-1038.**
- **Hu, S.-P. and A. Sjovold (1987), "Error Corrections for Unbiased Log-Linear Least-Square Estimates," Tecolote Research under contract to USAF/SMC/FMC, October 1987, 52 pages.**

MCR
Critical Thinking. Solutions Delivered.

References, Chart 2 of 2

- Jarvis, W. and R. Rozzo (2003), "Log-Linear Models: Transformation Bias and Goodness-of-Fit," DoD Cost Analysis Symposium (DoDCAS), January 2003, 15 charts.
- Larsen, R. and M. Marx (2001), *An Introduction to Mathematical Statistics and Its Applications*, Prentice-Hall, pages 338-344.
- Miller, D. (1984), "Reducing Transformation Bias in Curve Fitting," *The American Statistician*, Vol. 38, May 1984, pages 124-126.
- Meyer, H. (1941), "A Correction for Systematic Error Occurring in the Application of the Logarithmic Value Equation," *Journal Series of the Pennsylvania Agricultural Experiment Station*, Paper 1058, pages 1-3.
- Neyman, J. and E. Scott (1960), "Correction for Bias Introduced by a Transformation of Variables," *Annals of Mathematical Statistics*, Vol.31, September 1960, pages 643-655.
- United States Air Force Space and Missile Systems Center (1988), *Unmanned Space Vehicle Cost Model, Sixth Edition* (USCM6).
- United States Air Force Space and Missile Systems Center (1994), *Unmanned Space Vehicle Cost Model, Seventh Edition* (USCM7).
- Young, P. (1999), "The Meaning of the Standard Error of the Estimate in Logarithmic Space," *The Journal of Cost Analysis & Management*, Winter 1999, pages 59-65.

CRITICAL THINKING.
SOLUTIONS DELIVERED.

Speaker's Bio

Dr. Stephen A. Book vacated the position of Chief Technical Officer of MCR, LLC in 2010 (after serving in that position for almost a decade) to concentrate on research, training, and subject-matter-expert customer support. In his former capacity, he was responsible for ensuring technical excellence of MCR products, services, and processes by encouraging process improvement, maintaining quality control, and training employees and customers in cost and schedule analysis and associated program-control disciplines. Earlier, at The Aerospace Corporation, he was a principal contributor to several Air Force cost studies of national significance, including the DSP/FEWS/BSTS/AWS/Brilliant Eyes Sensor Integration Study (1992) and the ALS/Spacelifter/EELV Launch Options Study (1993). He has served on national panels as an independent reviewer of NASA programs, for example the 2005 Senior External Review Team on cost-estimating methods for the Exploration Systems Mission Directorate, the 1997-98 Cost Assessment and Validation Task Force on the International Space Station ("Chabrow Committee"), and the 1998-99 National Research Council Committee on Space Shuttle Upgrades. Dr. Book joined MCR in January 2001 after 21 years with Aerospace, where he held the title "Distinguished Engineer" during 1996-2000 and served as Director, Resource and Requirements Analysis Department, during 1989-1995. Dr. Book is co-editor of the ISPA/SCEA technical journal, *The Journal of Cost Analysis and Parametrics*. He received the 2010 SCEA Lifetime Achievement Award, the 2009 NASA Cost Contractor of the Year award, the 2005 ISPA Freiman Award for Lifetime Achievement, and the 1982 Aerospace Corporation President's Award for Analytic Achievement. Dr. Book earned his Ph.D. in mathematics, with concentration in probability and statistics, at the University of Oregon.