

Statistical Foundations of Adaptive Cost-Estimating Relationships

Stephen A. Book
MCR LLC
sbook@mcri.com

Melvin A. Broder
The Aerospace Corporation
melvin.a.broder@aero.org

Daniel I. Feldman
MCR LLC
dfeldman@mcri.com

SCEA-ISPA Joint Annual Conference & Training Workshop
Society of Cost Estimating and Analysis
International Society of Parametric Analysts
St Louis MO
2-4 June 2009

Acknowledgments

The authors are grateful to the members of MCR's quality-review team: Dr. Neal D. Hulkower, Dr. Jerry D. Hofmann, Raymond P. Covert, Nathan J. Menton, Timothy P. Anderson, and Jan B. Sterbutzel. The combined efforts of the team resulted in a significant improvement in the clarity of explanation of specific details of the conclusions drawn, as well as the elimination of various typos and other errors.

CER Error Sources

- **Inability of Any Cost-Estimating Relationship (CER) to Account for All Influences On Cost, No Matter How Many Cost Drivers it Contains – Too Bad, We Usually Can't Do Anything About This**
- **Incorrectness of Algebraic CER Model to Which Cost Numbers in its Data Base are Statistically Fit – Tough Luck, Try Another Algebraic Form**
- **When CER is Applied to Estimate Cost at a Particular Cost-Driver Value x , Location of that Cost Driver Value with Respect to Cost-Driver Values Comprising Historical Cost Data Base – **This Issue is What We Will Try to Resolve in This Briefing**
 - **If x is Located Near Center of Range of Parameter Values, CER Will Provide Fairly Precise Estimate of Cost**
 - **If x Is Located Far From Center of Range, CER-based Estimate Will Be Considerably Less Precise****

Starting Point for the Study

- **Up to Now, Most Cost-Estimating Relationships (CERs) Have Been Based on Full Data Sets Consisting of All Cost and Technical Data Associated with a Particular Class of Products of Interest (e.g., Components or Subsystems of Satellites, Ground Systems, Aircraft, etc.)**
- **Exceptions Have Typically Involved Only ...**
 - **The Analyst’s Ability to Choose Particular Analogous Entities from a Full Data Set and Derive a CER from the Reduced Data Set, e.g. NAFCOM**
 - **Preliminary Removal of “Outliers” from the Full Data Set**
- **Last Year, the First Two Authors Described the Process of Deriving “Adaptive” CERs, Namely CERs that are “Tuned” to Specific Cost-Driver Values and Presented Several Examples of How to Carry Out that Process**

Today's Presentation

- **The Goal of Such Specialization is to be Able to Apply CERs that Have Smaller Estimating Error and Narrower Prediction Bounds**
- **In Today's Presentation, We Will Explain the Mathematics behind the Analysis of Last Year's Adaptive CER Report, which will Point the Way Toward More Advanced Results**
 - **Adaptive CERs that are Applicable over the Entire Cost-Driver Range**
 - **Improvements in Our Understanding of Adaptive CER Standard Errors and Prediction Bounds**

Contents

- **Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics**
- **Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics**
- **Adaptive CERs via Quadratic-Distance Weighting**
- **The “Universal Adaptive CER”**
- **Prediction Bounds for OLS, WLS, and Adaptive CERs**
- **Conclusions**

Contents

- **Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics**
- **Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics**
- **Adaptive CERs via Quadratic-Distance Weighting**
- **The “Universal Adaptive CER”**
- **Prediction Bounds for OLS, WLS, and Adaptive CERs**
- **Conclusions**

Choice of Additive-Error Model

- **Normally, the Multiplicative-Error Regression Model is Preferred for CERs**
 - Typically, Data-Base Cost Values Range over Large Intervals – Often by Two or Three Orders of Magnitude
 - So the Possible Output Values of the CER Will Range Over A Similar Interval
 - Therefore $\pm 30\%$ is More Meaningful as a Standard-Error Metric Than $\pm \$30,000$
- **For What We are Considering in this Study, However, CER Output Will Range Over a Relatively Short (or even zero-length) Interval**
 - The CER Will Be Usable for Only a Small Set of Cost-Driver Values (or maybe even only one cost-driver value)
 - Therefore, a Dollar-Valued Standard-Error Metric is Just as Meaningful as a Percentage Standard-Error Metric (and the mathematics is simpler)
- **We Will Therefore Apply the Additive-Error Regression Model**

Ordinary Least Squares

- **OLS “Best” Fits a Straight Line $y = a+bx$ to a Set of Data Points (x_k, y_k) in Two-Dimensional Space**
 - x_k is the Value of the Cost Driver
 - y_k is the Cost
- **The OLS Criterion is That the Coefficients a and b are Selected so That the Sum of Squares of the Differences $d_k = y_k - (a + bx_k) = y_k - a - bx_k$ Between the Actual Costs and their Estimates is as Small as Possible**
- **The Mathematics Results in Numerical Values of a and b That Minimize the Quantity**

$$f(a, b) = \sum_{k=1}^n d_k^2 = \sum_{k=1}^n (y_k - a - bx_k)^2$$

Coefficients of the OLS CER

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

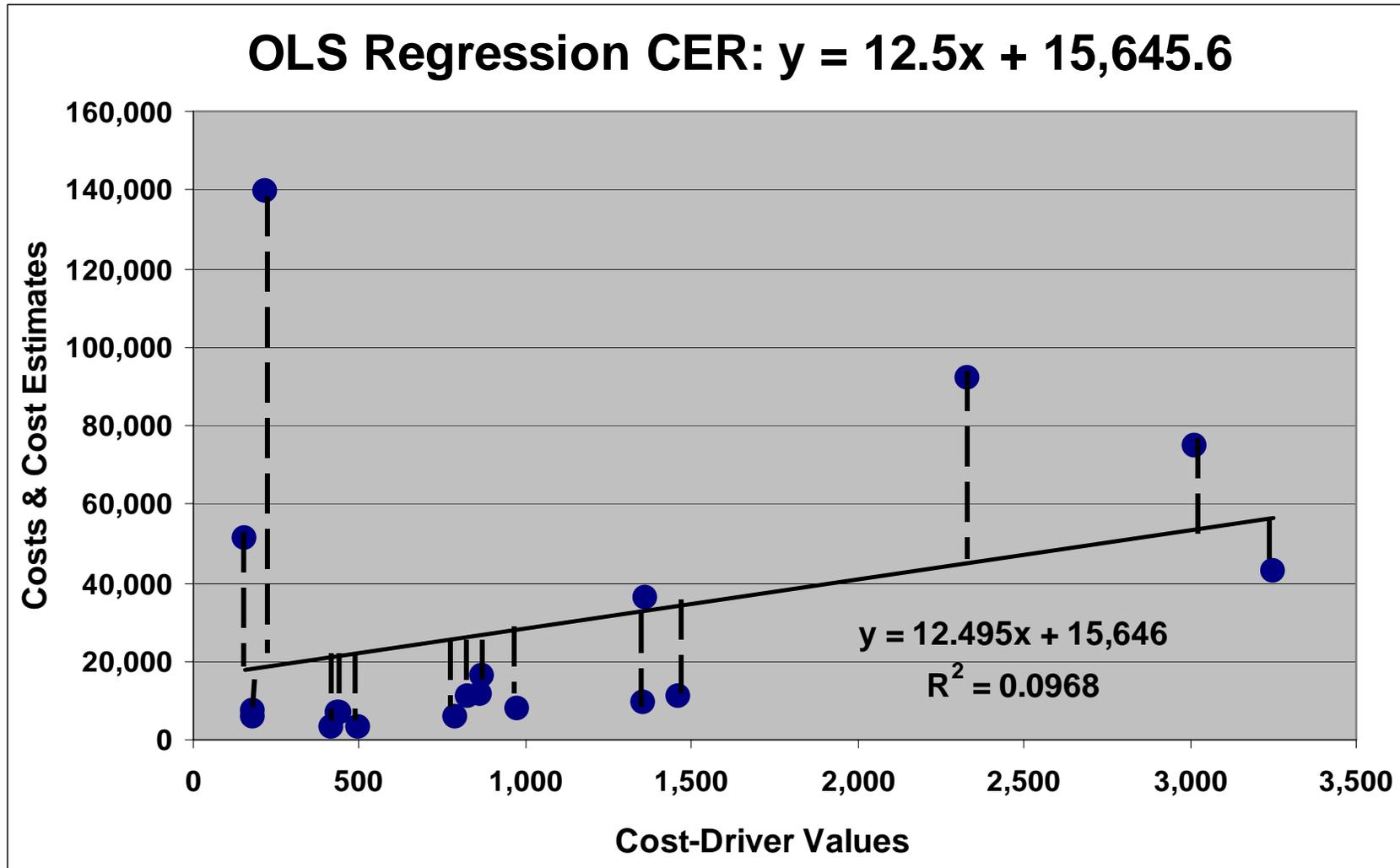
$$a = \frac{\sum_{k=1}^n y_k}{n} - b \frac{\sum_{k=1}^n x_k}{n}$$

A Set of Actual Data

Program	Cost-Driver Value x	Unit Cost y
A	156.12	51,367.22
B	179.40	5,885.00
C	180.30	7,060.00
D	217.50	139,483.12
E	419.14	3,386.00
F	437.09	6,738.00
G	440.93	6,812.00
H	494.45	3,291.34
I	789.90	5,723.14
J	826.10	10,992.00
K	864.30	11,590.00
L	869.30	15,973.00
M	976.50	7,970.67
N	1,355.80	9,524.10
O	1,360.90	35,927.22
P	1,463.21	11,238.73
Q	2,332.10	92,059.97
R	3,017.73	74,649.00
S	3,253.00	42,915.23

Note: This data set is a set of actual cost data; due to proprietary issues, however, the exact descriptions of the data points cannot be revealed.

OLS CER Based on Actual Data



OLS CER Quality Metrics

- OLS Standard-Error of the Estimate

$$SEE = \sqrt{\frac{\sum_{k=1}^n (y_k - a - bx_k)^2}{n-2}} = \sqrt{\frac{\sum_{k=1}^n y_k^2 - a \sum_{k=1}^n y_k - b \sum_{k=1}^n x_k y_k}{n-2}}$$

- OLS Bias

$$B = \frac{1}{n} \sum_{k=1}^n (a + bx_k - y_k) = 0.$$

- OLS R²

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

OLS CER Quality Metrics Based on Set of Actual Data

- $SEE = 34,336.83$
- $B = 0$
- $R^2 = 0.0968 = 9.68\%$

Contents

- Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics
- **Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics**
- Adaptive CERs via Quadratic-Distance Weighting
- The “Universal Adaptive CER”
- Prediction Bounds for OLS, WLS, and Adaptive CERs
- Conclusions

Weighted Least Squares

- In “Weighted” Least Squares (WLS), the Problem is the Same, Except that the Points are not Considered of Equal Value
- Accompanying Each Data Point (x_k, y_k) is a “Weight” w_k , so that the Data Set Consists of “Triples” (x_k, y_k, w_k) , Rather than Pairs (x_k, y_k)
- The WLS Criterion is that the Coefficients a and b are Selected so that the Weighted Sum of Squares of the Differences $(y_k - a - bx_k)$ is as Small as Possible
- The Mathematics Results in Numerical Values of a and b that Minimize the Quantity

$$f(a, b) = \sum_{k=1}^n w_k d_k^2 = \sum_{k=1}^n w_k (y_k - a - bx_k)^2$$

- **Weights w_k are Chosen as Follows:**
 - Large When the Data Point is to Contribute Heavily to the CER
 - Small When the Data Point is to Contribute Only in a Minor Way, if at All, to the CER

The Weighted Least-Squares Solution

- Applying some Calculus, we Can Derive Explicit Formulas for the Numerical Values of a and b that Minimize the Quantity

$$f(a,b) = \sum_{k=1}^n w_k d_k^2 = \sum_{k=1}^n w_k (y_k - a - bx_k)^2$$

- The Resulting Expressions for a and b are as Follows:

$$b = \frac{\left(\sum_{k=1}^n w_k \right) \left(\sum_{k=1}^n w_k x_k y_k \right) - \left(\sum_{k=1}^n w_k x_k \right) \left(\sum_{k=1}^n w_k y_k \right)}{\left(\sum_{k=1}^n w_k \right) \left(\sum_{k=1}^n w_k x_k^2 \right) - \left(\sum_{k=1}^n w_k x_k \right)^2}$$

$$a = \frac{\sum_{k=1}^n w_k y_k}{\sum_{k=1}^n w_k} - b \frac{\sum_{k=1}^n w_k x_k}{\sum_{k=1}^n w_k}$$

Reference: S.A. Book, "Deriving Cost-Estimating Relationships Using Weighted Least-Squares Regression," IAA/ISPA/AIAA Space Systems Cost Methodologies and Applications Symposium, San Diego CA, 10-11 May 1990. (Reference 1)

Weighting Data Points *A Priori*

- **Data Points Supporting CER Development May Not All be of Equal Value**
 - Some May be Known With Greater Precision Than Others
 - Some May Be More Relevant to the Particular Estimating Task Than Others
 - Some May Be Very Far From the Cost-driver Region where Estimating is Most Commonly Done
- **Should All Data Points Contribute Equally to the Computation of the CER?**

The Actual Data Set

Unweighted, i.e. All Weights = 1

Program	Statistical Weight w	Cost-Driver Value x	Unit Cost C	Estimated Cost EST C
A	1	156.12	51,367.22	17,596.30
B	1	179.40	5,885.00	17,887.18
C	1	180.30	7,060.00	17,898.42
D	1	217.50	139,483.12	18,363.23
E	1	419.14	3,386.00	20,882.67
F	1	437.09	6,738.00	21,106.95
G	1	440.93	6,812.00	21,154.93
H	1	494.45	3,291.34	21,823.65
I	1	789.90	5,723.14	25,515.22
J	1	826.10	10,992.00	25,967.53
K	1	864.30	11,590.00	26,444.83
L	1	869.30	15,973.00	26,507.30
M	1	976.50	7,970.67	27,846.74
N	1	1,355.80	9,524.10	32,586.00
O	1	1,360.90	35,927.22	32,649.72
P	1	1,463.21	11,238.73	33,928.06
Q	1	2,332.10	92,059.97	44,784.62
R	1	3,017.73	74,649.00	53,351.39
S	1	3,253.00	42,915.23	56,291.03
Sums =	19.00	19,633.77	542,585.74	542,585.74
a = 15,645.619		Std Error = 36,300.49		
b = 12.495		Bias = 0.00		
		R ² = 9.68%		

Contents

- Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics
- Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics
- **Adaptive CERs via Quadratic-Distance Weighting**
- The “Universal Adaptive CER”
- Prediction Bounds for OLS, WLS, and Adaptive CERs
- Conclusions

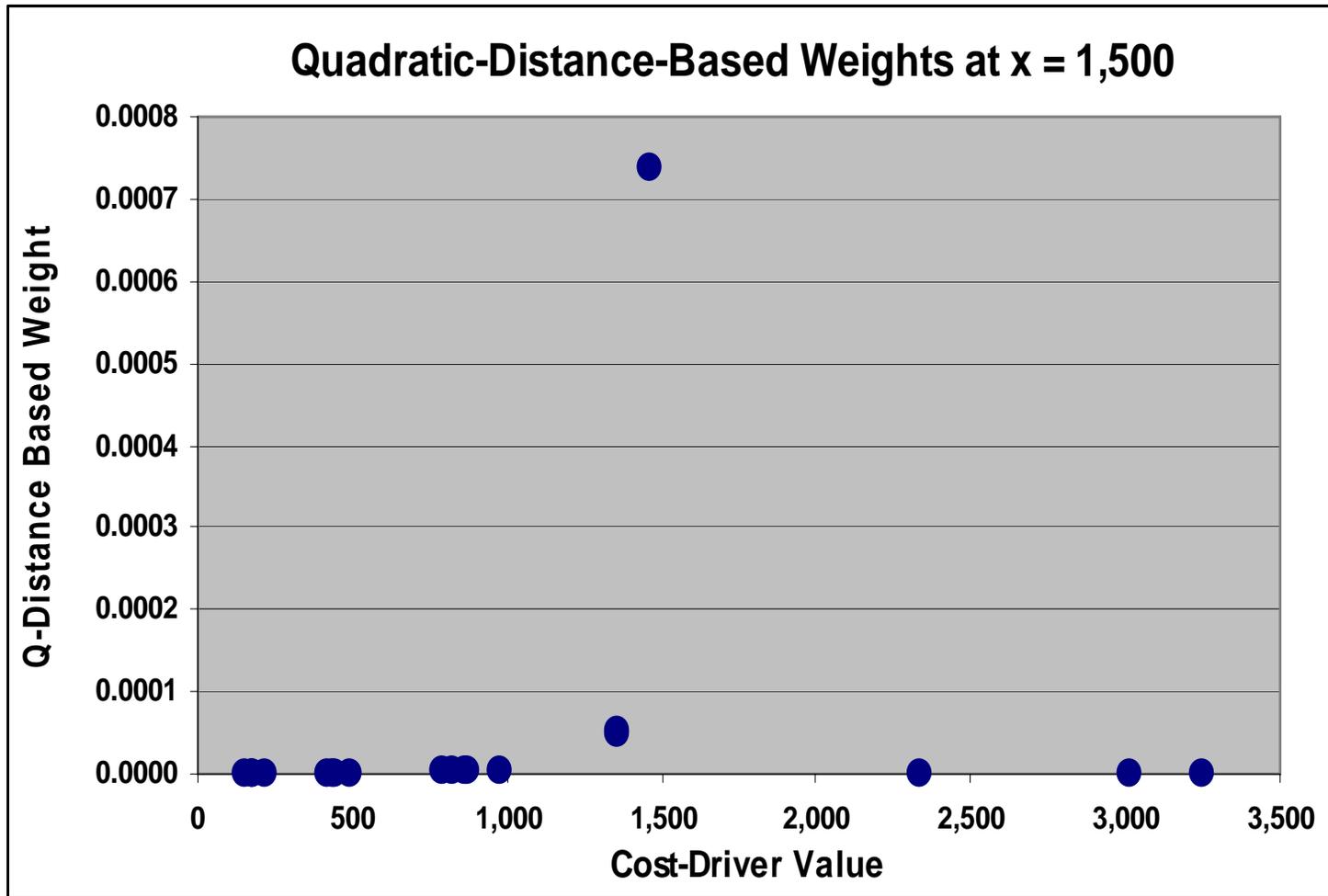
Quadratic-Distance Weighting

- **The “Q-Distance” Method Calls for Weighting Points by the Reciprocal of their Squared Distance Along the x-axis from a Cost-Driver Value of Interest**
- **Given that ...**
 - x_0 is the Value of the Cost Driver for the Product of Interest Being Investigated
 - x_k is the x Value of the k^{th} Data Point
 - d_k is the Distance from the x_0 Value to the Data Point x_k
- **Then the Weight of the Data Point x_k is the Reciprocal of the Quadratic (Squared) Distance, Namely $w_k = d_k^{-2} = (x_k - x_0)^{-2}$**

Why d_k^{-2} ?

- **There is an Infinite Number of Ways to Define the Weighting in Terms of a Cost Driver's Distance from x_0**
- **We Chose the Squared (Quadratic) Distance, because OLS Calculations Use the Squares of Residuals for Best Fit – this Forces the CER to Pass Through the Point (\bar{x}, \bar{y}) , where \bar{x} is the Mean of the Cost-Driver Values and \bar{y} is the Mean of the Cost Values**
- **However, Other Weighting Schemes Can be Used**

Quadratic-Distance-Based Weights at $x = 1,500$

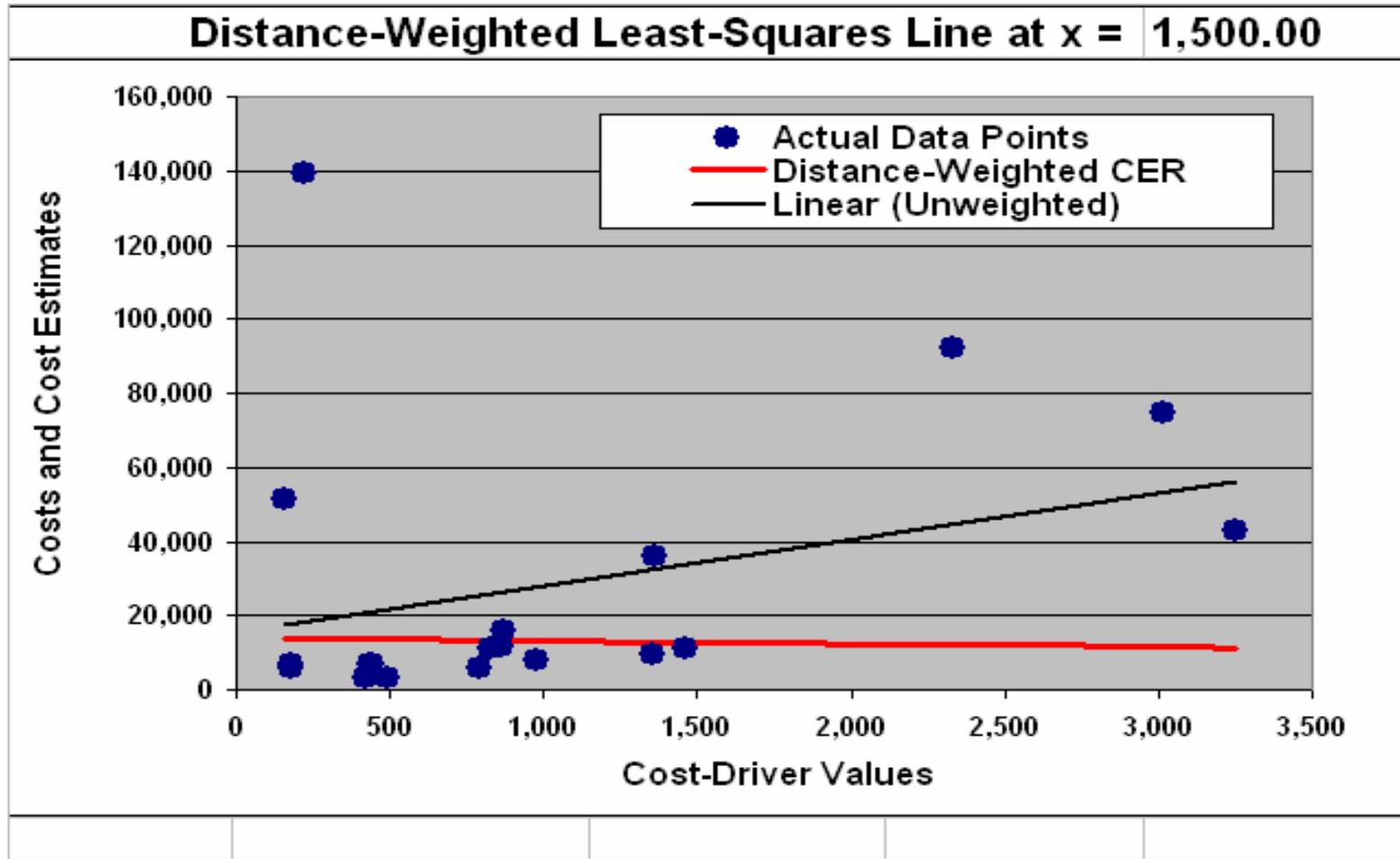


Weighting Points by their Q-Distance from $x = 1,500$

- Quadratic-Distance Weighting Anchored at $x = 1,500$

Program	Statistical Distance d	Cost-Driver Value x	Unit Cost C	Estimated Cost EST C	Weights $w^* = 1/d^2$
A	1343.88	156.12	51,367.22	13,969.64	0.0000005537
B	1320.6	179.40	5,885.00	13,949.82	0.0000005734
C	1319.7	180.30	7,060.00	13,949.06	0.0000005742
D	1282.5	217.50	139,483.12	13,917.39	0.0000006080
E	1080.86	419.14	3,386.00	13,745.72	0.0000008560
F	1062.91	437.09	6,738.00	13,730.44	0.0000008851
G	1059.07	440.93	6,812.00	13,727.17	0.0000008916
H	1005.55	494.45	3,291.34	13,681.61	0.0000009890
I	710.1	789.90	5,723.14	13,430.08	0.0000019832
J	673.9	826.10	10,992.00	13,399.26	0.0000022020
K	635.7	864.30	11,590.00	13,366.74	0.0000024745
L	630.7	869.30	15,973.00	13,362.48	0.0000025139
M	523.5	976.50	7,970.67	13,271.22	0.0000036489
N	144.2	1,355.80	9,524.10	12,948.30	0.0000480916
O	139.1	1,360.90	35,927.22	12,943.96	0.0000516828
P	36.79	1,463.21	11,238.73	12,856.86	0.0007388230
Q	832.1	2,332.10	92,059.97	12,117.13	0.0000014443
R	1517.73	3,017.73	74,649.00	11,533.42	0.0000004341
S	1753	3,253.00	42,915.23	11,333.12	0.0000003254
Sums =	17,071.89	19,633.77	542,585.74	251,233.42	0.0008595547
					Σw_k
a =	14,102.56				
b =	-0.85				

Resulting Q-Distance CER for Real Data Set at $x = 1,500$



“Normalizing” the Weights

- $\{w_1^*, w_2^*, \dots, w_n^*\}$ is the Initial Set of Weights Assigned to the Data Points
- We Want to Find a Set of Weights $\{w_1, w_2, \dots, w_n\}$ that is Equivalent to the Initial Set in the Sense that the Relative Weight of the Data Points to Each Other is Preserved
- In Addition, We Want the “Normalized” Weights to Sum to n , i.e. $\sum_{k=1}^n w_k = n$
- The Definition of Normalized Weights that Works is

$$w_j = n w_j^* \left(\sum_{k=1}^n w_k^* \right)^{-1}$$

- Note, for Each Pair i and j , that $w_i/w_j = w_i^*/w_j^*$, so the Relative Weighting of the Data Points is Preserved

Using the Normalized Weights

- From this Point on, We will be Using the Normalized Weights in Lieu of the Initial Weights
- The Resulting Expressions for a and b are as Follows:

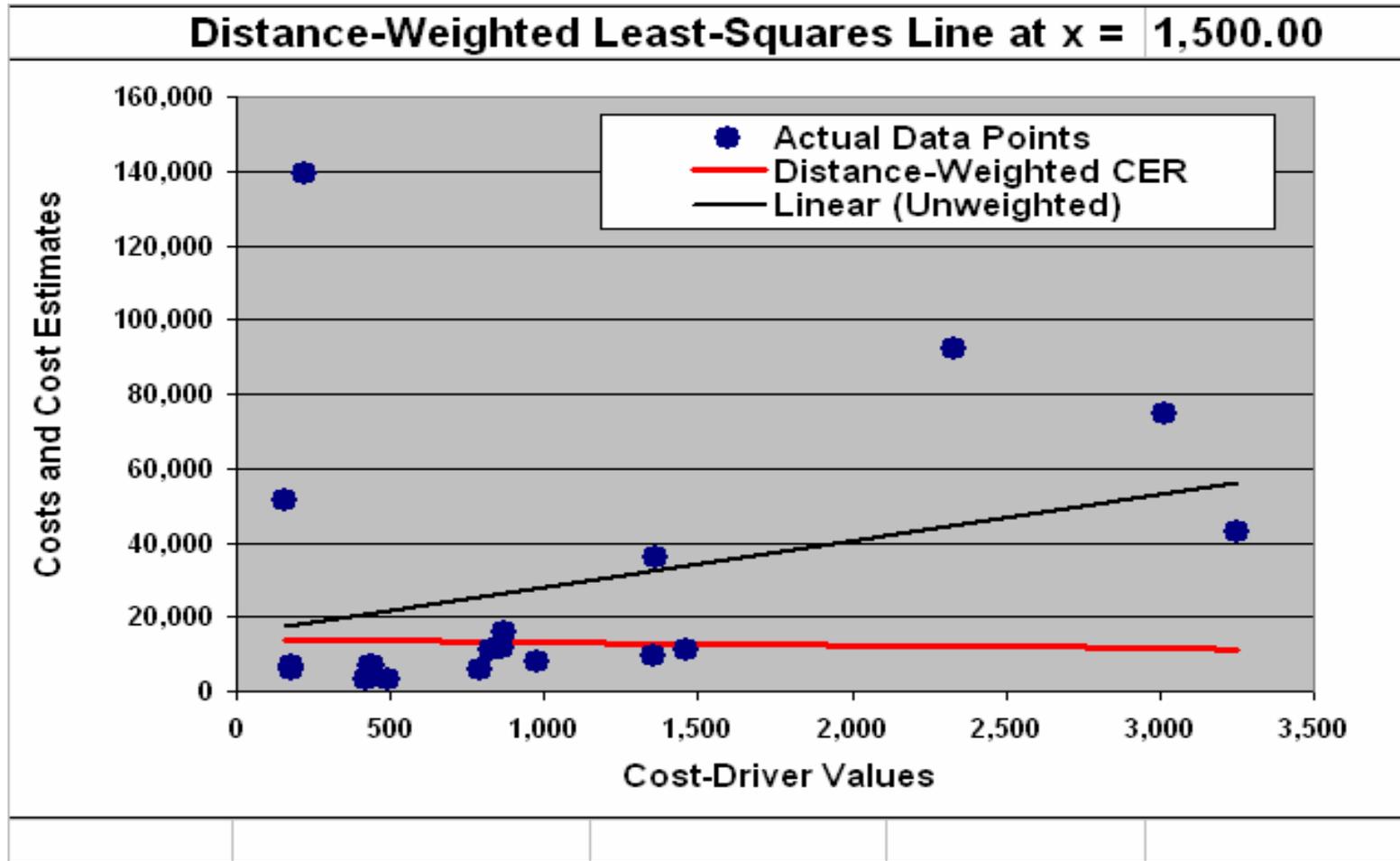
$$b = \frac{n \left(\sum_{k=1}^n w_k x_k y_k \right) - \left(\sum_{k=1}^n w_k x_k \right) \left(\sum_{k=1}^n w_k y_k \right)}{n \left(\sum_{k=1}^n w_k x_k^2 \right) - \left(\sum_{k=1}^n w_k x_k \right)^2}$$
$$a = \frac{\sum_{k=1}^n w_k y_k}{n} - b \frac{\sum_{k=1}^n w_k x_k}{n}$$

Normalized Weighting of Points by their Q-Distance from $x = 1,500$

- Quadratic-Distance Weighting Anchored at $x = 1,500$

Program	Statistical Distance d	Cost-Driver Value x	Unit Cost C	Estimated Cost EST C	Weights $w^* = 1/d^2$	Normalized Weights w
A	1343.88	156.12	51,367.22	13,969.64	0.0000005537	0.0122393710
B	1320.6	179.40	5,885.00	13,949.82	0.0000005734	0.0126746943
C	1319.7	180.30	7,060.00	13,949.06	0.0000005742	0.0126919878
D	1282.5	217.50	139,483.12	13,917.39	0.0000006080	0.0134389498
E	1080.86	419.14	3,386.00	13,745.72	0.0000008560	0.0189208738
F	1062.91	437.09	6,738.00	13,730.44	0.0000008851	0.0195653262
G	1059.07	440.93	6,812.00	13,727.17	0.0000008916	0.0197074643
H	1005.55	494.45	3,291.34	13,681.61	0.0000009890	0.0218611367
I	710.1	789.90	5,723.14	13,430.08	0.0000019832	0.0438370231
J	673.9	826.10	10,992.00	13,399.26	0.0000022020	0.0486731179
K	635.7	864.30	11,590.00	13,366.74	0.0000024745	0.0546985302
L	630.7	869.30	15,973.00	13,362.48	0.0000025139	0.0555692350
M	523.5	976.50	7,970.67	13,271.22	0.0000036489	0.0806578623
N	144.2	1,355.80	9,524.10	12,948.30	0.0000480916	1.0630398840
O	139.1	1,360.90	35,927.22	12,943.96	0.0000516828	1.1424200586
P	36.79	1,463.21	11,238.73	12,856.86	0.0007388230	16.3312905022
Q	832.1	2,332.10	92,059.97	12,117.13	0.0000014443	0.0319248625
R	1517.73	3,017.73	74,649.00	11,533.42	0.0000004341	0.0095960177
S	1753	3,253.00	42,915.23	11,333.12	0.0000003254	0.0071931025
Sums =	17,071.89	19,633.77	542,585.74	251,233.42	0.0008595547	19.0000000000
					Σw_k	
a =	14,102.56					
b =	-0.85					

Resulting Q-Distance CER Anchored at $x = 1,500$



WLS CER Quality Metrics

- **WLS Standard-Error of the Estimate**

$$SEE_w = \sqrt{\frac{\sum_{k=1}^n w_k (y_k - a - bx_k)^2}{\sum_{k=1}^n w_k - 2}} = \sqrt{\frac{\sum_{k=1}^n w_k y_k^2 - a \sum_{k=1}^n w_k y_k - b \sum_{k=1}^n w_k x_k y_k}{n - 2}}$$

- **WLS Bias (not necessarily 0)**

$$B_w = \frac{\left(\sum_{k=1}^n w_k (y_k - bx_k) \right)}{n} - \left(\sum_{k=1}^n (x_k - y_k) \right) = a - \frac{\left(\sum_{k=1}^n y_k \right)}{n} - b \left(\frac{\sum_{k=1}^n x_k}{n} \right)$$

- **WLS R²**

$$R_w^2 = \frac{\left\{ n \left(\sum_{k=1}^n w_k x_k y_k \right) - \left(\sum_{k=1}^n w_k x_k \right) \left(\sum_{k=1}^n w_k y_k \right) \right\}^2}{\left\{ n \left(\sum_{k=1}^n w_k x_k^2 \right) - \left(\sum_{k=1}^n w_k x_k \right)^2 \right\} \left\{ n \left(\sum_{k=1}^n w_k y_k^2 \right) - \left(\sum_{k=1}^n w_k y_k \right)^2 \right\}}$$

WLS CER Quality Metrics of Adaptive CER Based at 1,500

- $SEE = 8,226.72$
- $B = -15,334.33$
- $R^2 = 0.000218 = 0.0218\%$

Contents

- Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics
- Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics
- Adaptive CERs via Quadratic-Distance Weighting
- **The “Universal Adaptive CER”**
- Prediction Bounds for OLS, WLS, and Adaptive CERs
- Conclusions

The “Universal” Adaptive CER

- The “Universal” Adaptive CER is Formed by Combining* the Various Individual Adaptive CERs of the Sort Derived Above Over the Range of Cost Drivers into One CER that Applies Over the Entire Range
- This Universal Adaptive CER is, as P. Foussier (Reference 3, Chart 5) Presciently Noted, “Highly Nonlinear”
- For the Data Set We Have Been Working with, We Can Consider the Cost-driver Range to Go from 50 to 3,500
- We Calculate a Quadratic-Distance Weighted CER and an Estimated Cost at Each Increment of 50 for Each of Those Cost-Driver Values
- Then We String All these Estimates Together and Interpolate between Successive Ones to Form the Universal Adaptive CER

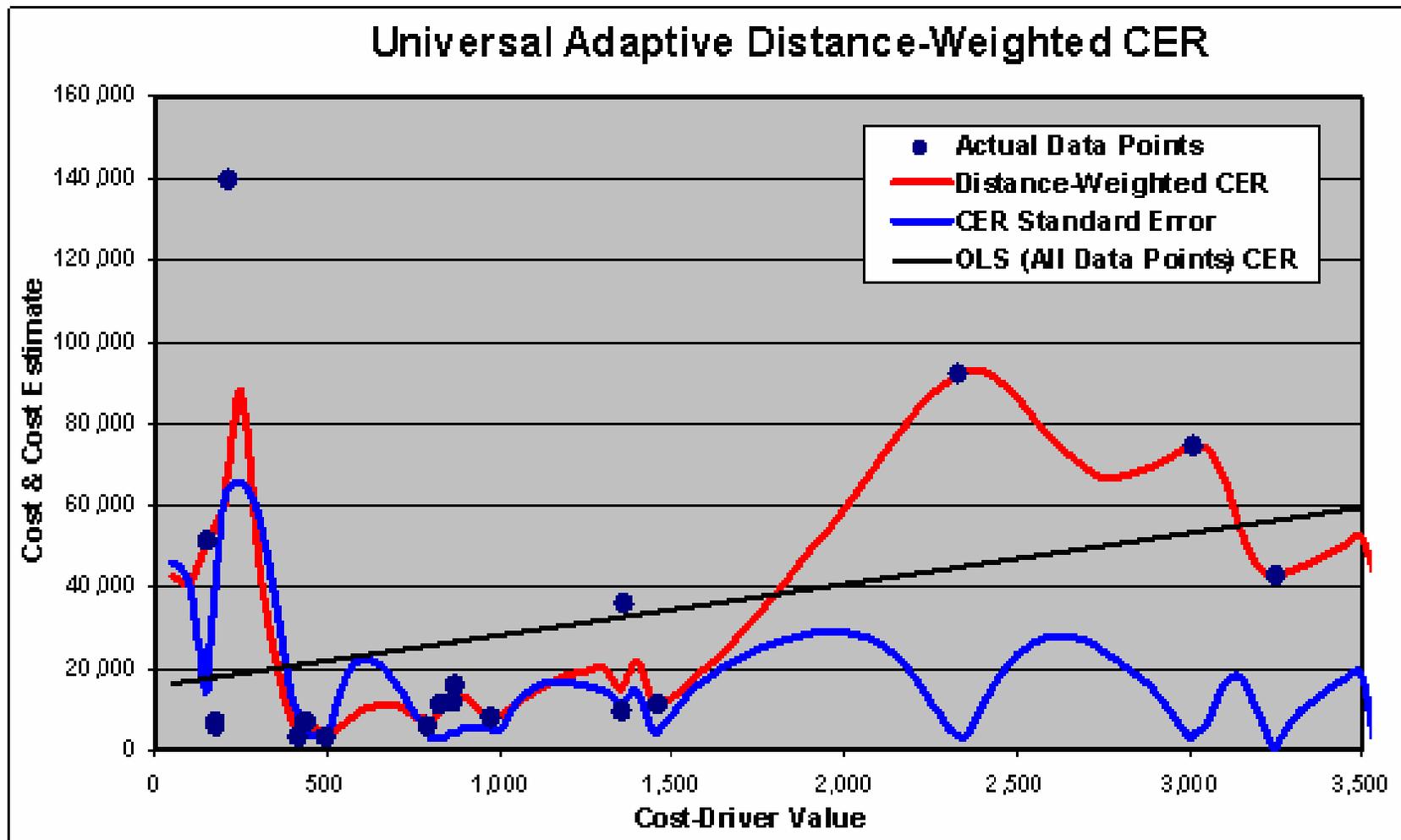
* The idea of combining estimates at various points of the cost-driver range into one all-inclusive CER was suggested to us by Paul Wetzel of OpsConsulting LLC.

Universal Adaptive CER-Based Estimates at 50-Unit Increments

Driver	EST Cost	Std Error
50.00	42,739.31	46,098.71
100.00	40,817.29	41,490.92
150.00	49,546.82	15,013.91
156.12	50,880.53	20,862.57
179.40	55,953.88	43,110.41
180.30	56,150.02	43,970.50
200.00	60,443.18	62,797.07
217.50	69,749.17	63,712.78
250.00	87,031.73	65,413.39
300.00	46,425.71	57,676.55
350.00	22,733.56	36,873.63
400.00	7,006.95	11,986.04
419.14	6,760.42	9,109.80
437.09	6,529.22	6,412.39
440.93	6,479.76	5,835.34
450.00	6,362.94	4,472.36
494.45	3,589.46	3,084.58
500.00	3,243.16	2,911.31
550.00	6,829.12	17,776.83
600.00	9,959.40	22,010.11
650.00	11,310.17	21,033.96
700.00	10,929.01	16,492.92
750.00	8,652.67	9,456.12
789.90	7,175.24	4,565.75
800.00	6,801.25	3,327.84
826.10	9,756.59	3,386.63
850.00	12,462.82	3,440.47
864.30	12,666.50	4,059.71
869.30	12,737.72	4,276.23
900.00	13,174.99	5,605.64
950.00	9,208.15	5,651.88
976.50	8,832.68	5,342.38
1,000.00	8,499.71	5,067.91
1,050.00	11,462.16	11,841.54
1,100.00	14,296.49	15,323.02
1,150.00	16,537.15	16,912.27
1,200.00	18,230.99	17,020.52
1,250.00	19,495.31	16,029.95
1,300.00	20,310.23	14,631.94
1,350.00	14,974.31	11,522.07
1,355.80	15,774.27	11,821.74
1,360.90	16,477.67	12,085.24
1,400.00	21,870.45	14,105.41
1,450.00	11,840.86	4,214.92
1,463.21	12,101.01	5,274.84

Driver	EST Cost	Std Error
1,500.00	12,825.54	8,226.72
1,550.00	16,621.72	13,974.93
1,600.00	20,492.26	17,569.25
1,650.00	24,526.56	20,350.34
1,700.00	28,831.03	22,668.31
1,750.00	33,415.50	24,632.61
1,800.00	38,247.16	26,275.33
1,850.00	43,285.50	27,589.48
1,900.00	48,497.85	28,534.71
1,950.00	53,862.57	29,032.00
2,000.00	59,364.10	28,954.26
2,050.00	64,981.01	28,118.23
2,100.00	70,666.52	26,286.86
2,150.00	76,319.27	23,197.58
2,200.00	81,744.09	18,634.07
2,250.00	86,609.89	12,543.91
2,300.00	90,430.47	5,163.31
2,332.10	91,836.14	3,730.10
2,350.00	92,619.98	2,930.89
2,400.00	92,676.25	10,907.76
2,450.00	90,463.37	17,895.26
2,500.00	86,410.39	23,227.16
2,550.00	81,412.53	26,603.62
2,600.00	76,466.46	28,091.64
2,650.00	72,322.92	27,995.50
2,700.00	69,366.76	26,697.11
2,750.00	66,431.86	24,540.98
2,800.00	67,242.40	21,772.29
2,850.00	67,904.22	18,495.58
2,900.00	69,545.45	14,613.82
2,950.00	71,913.26	9,720.21
3,000.00	74,219.40	3,000.69
3,017.73	74,164.83	4,164.89
3,050.00	74,065.53	6,283.82
3,100.00	67,141.02	15,848.64
3,150.00	54,415.99	17,689.83
3,200.00	45,424.15	9,943.35
3,250.00	42,927.10	501.90
3,253.00	42,978.65	868.74
3,300.00	43,786.36	6,615.99
3,350.00	45,762.39	11,482.72
3,400.00	47,971.96	14,864.14
3,450.00	50,126.95	17,319.87
3,500.00	52,149.51	19,185.52

Universal Adaptive CER Graphed with Standard Error and Actual Data



Contents

- **Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics**
- **Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics**
- **Adaptive CERs via Quadratic-Distance Weighting**
- **The “Universal Adaptive CER”**
- **Prediction Bounds for OLS, WLS, and Adaptive CERs**
- **Conclusions**

Prediction Bounds on CER-Based Estimates

- **Estimating the Cost of Developing or Producing a New Subsystem or Component is Essentially Trying to Predict the Future, which Means that Any Such Estimate Contains Uncertainty**
- **Uncertainty in CER-Based Estimates Can be Expressed in Terms of Prediction Bounds, the Algebra of which is Completely Understood Only in the Case of CERs Derived by Classical OLS Linear Regression**
- **As a Result, an Explicit Formula Exists for “Prediction Intervals” that Bound Cost Estimates Based on CERs that Have Been Derived by Applying OLS to Historical Cost Data**

The Formula for Prediction Bounds in the Case of OLS-Derived CERs

- The Formula for the $(1-\alpha)$ th Percent Upper and Lower Prediction Bounds on the True Cost y , Based on the Estimate $ESTy$ from the CER is the Following:

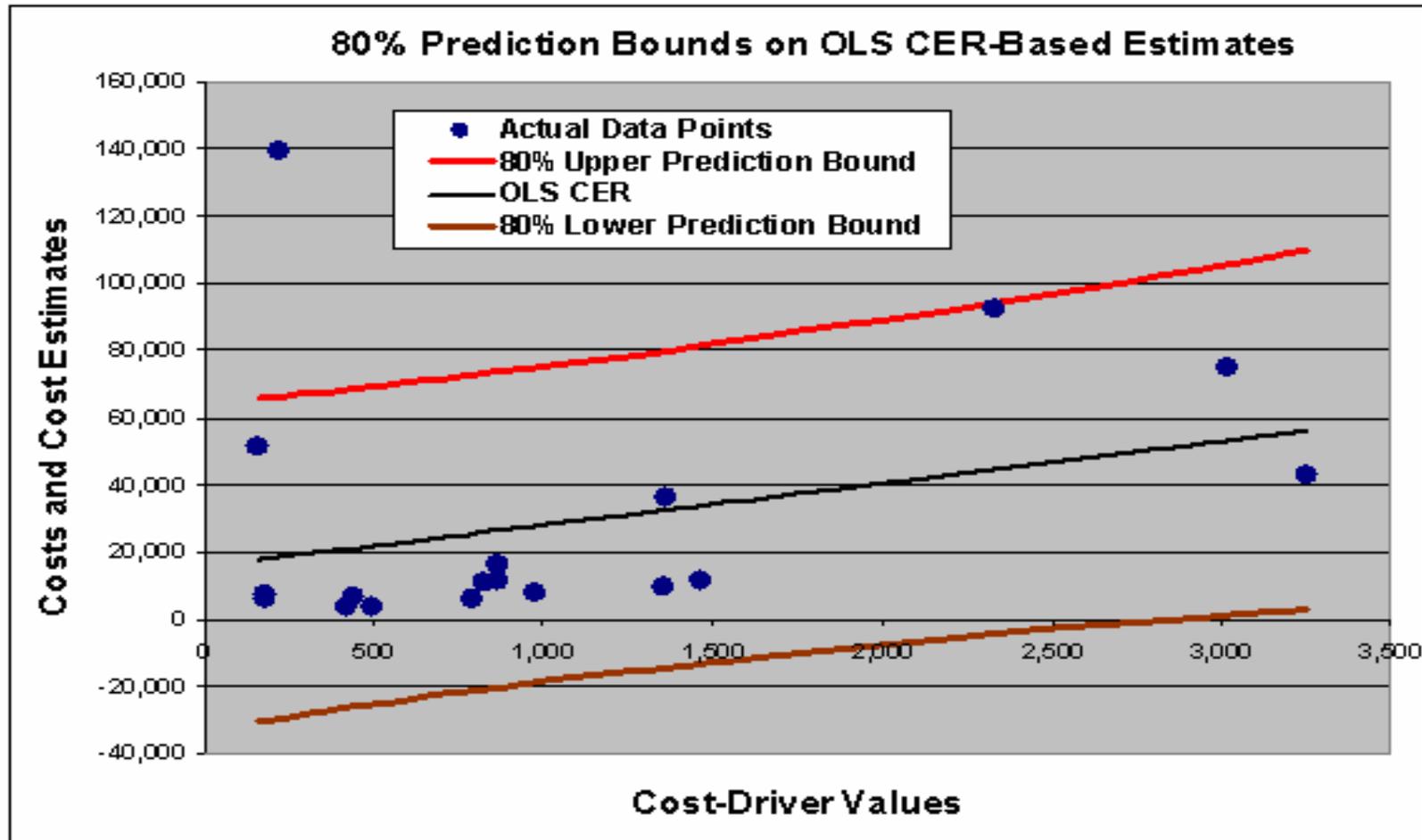
$$ESTy \pm t_{\alpha/2, n-2} * SEE \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

- Here $t_{\alpha/2, n-2}$ is the $(1-\alpha)$ th Percentage Point of the t Distribution, \bar{x} is the Mean of the Cost-Driver Values in the Data Base, x is the Cost-Driver Value at which the Estimate is Being Made, and SEE is the Standard Error of the CER

80% Upper and Lower Prediction Bounds on the OLS CER

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	OLS EST y	80% Lower Bound
A	156.12	51,367.22	65,673.53	17,596.30	-30,480.93
B	179.40	5,885.00	65,907.23	17,887.18	-30,132.88
C	180.30	7,060.00	65,916.29	17,898.42	-30,119.45
D	217.50	139,483.12	66,292.88	18,363.23	-29,566.43
E	419.14	3,386.00	68,400.42	20,882.67	-26,635.08
F	437.09	6,738.00	68,593.51	21,106.95	-26,379.62
G	440.93	6,812.00	68,634.94	21,154.93	-26,325.09
H	494.45	3,291.34	69,216.65	21,823.65	-25,569.35
I	789.90	5,723.14	72,574.56	25,515.22	-21,544.12
J	826.10	10,992.00	73,003.23	25,967.53	-21,068.17
K	864.30	11,590.00	73,459.69	26,444.83	-20,570.03
L	869.30	15,973.00	73,519.75	26,507.30	-20,505.14
M	976.50	7,970.67	74,824.83	27,846.74	-19,131.35
N	1,355.80	9,524.10	79,710.04	32,586.00	-14,538.05
O	1,360.90	35,927.22	79,778.56	32,649.72	-14,479.12
P	1,463.21	11,238.73	81,168.85	33,928.06	-13,312.74
Q	2,332.10	92,059.97	94,145.23	44,784.62	-4,576.00
R	3,017.73	74,649.00	105,728.61	53,351.39	974.17
S	3,253.00	42,915.23	109,940.12	56,291.03	2,641.94

80% Prediction Bounds Graphed with OLS CER and Actual Data



The Formula for Prediction Bounds in the Case of WLS-Derived CERs

- When the Weights are Normalized, the Expressions for the $(1-\alpha)$ th Percent Upper and Lower Prediction Bounds on the True Cost y at the Cost-Driver Value x_p , Based On Estimates $ESTy$ from WLS-Based Adaptive CERs are the Following:

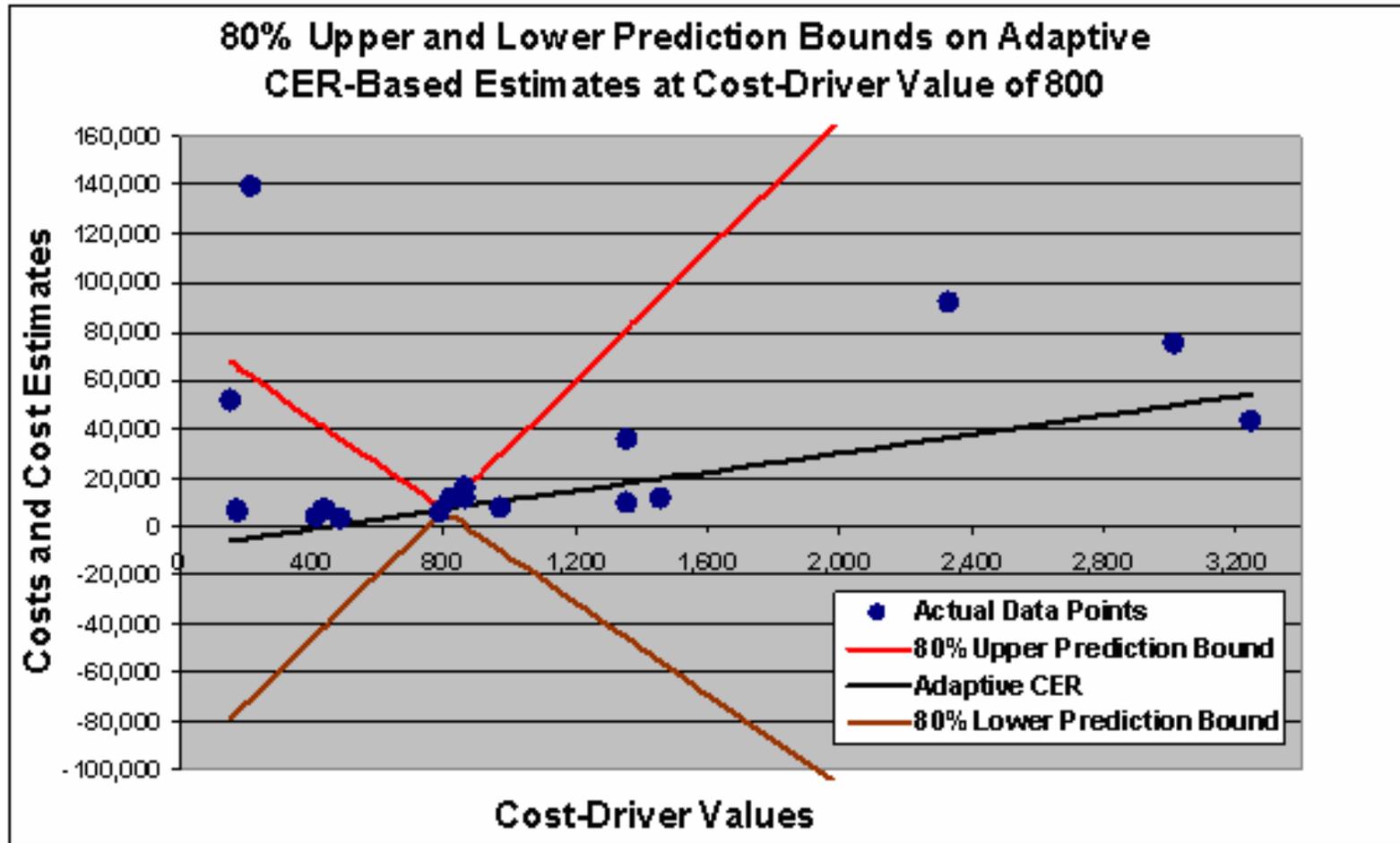
$$ESTy \pm t_{\alpha/2, n-2} * SEE_w \sqrt{\frac{1}{w_p} + \frac{1}{n} + \frac{n(x_p - \bar{x})^2}{n \left(\sum_{k=1}^n w_k x_k^2 \right) - \left(\sum_{k=1}^k w_k x_k \right)^2}}$$

- One Way to Obtain a Usable Value, if Needed, for w_p When x_p is Not in the Data Base from which the Adaptive CERs are Derived is to Interpolate between the Weights of the Nearest Data-base Points

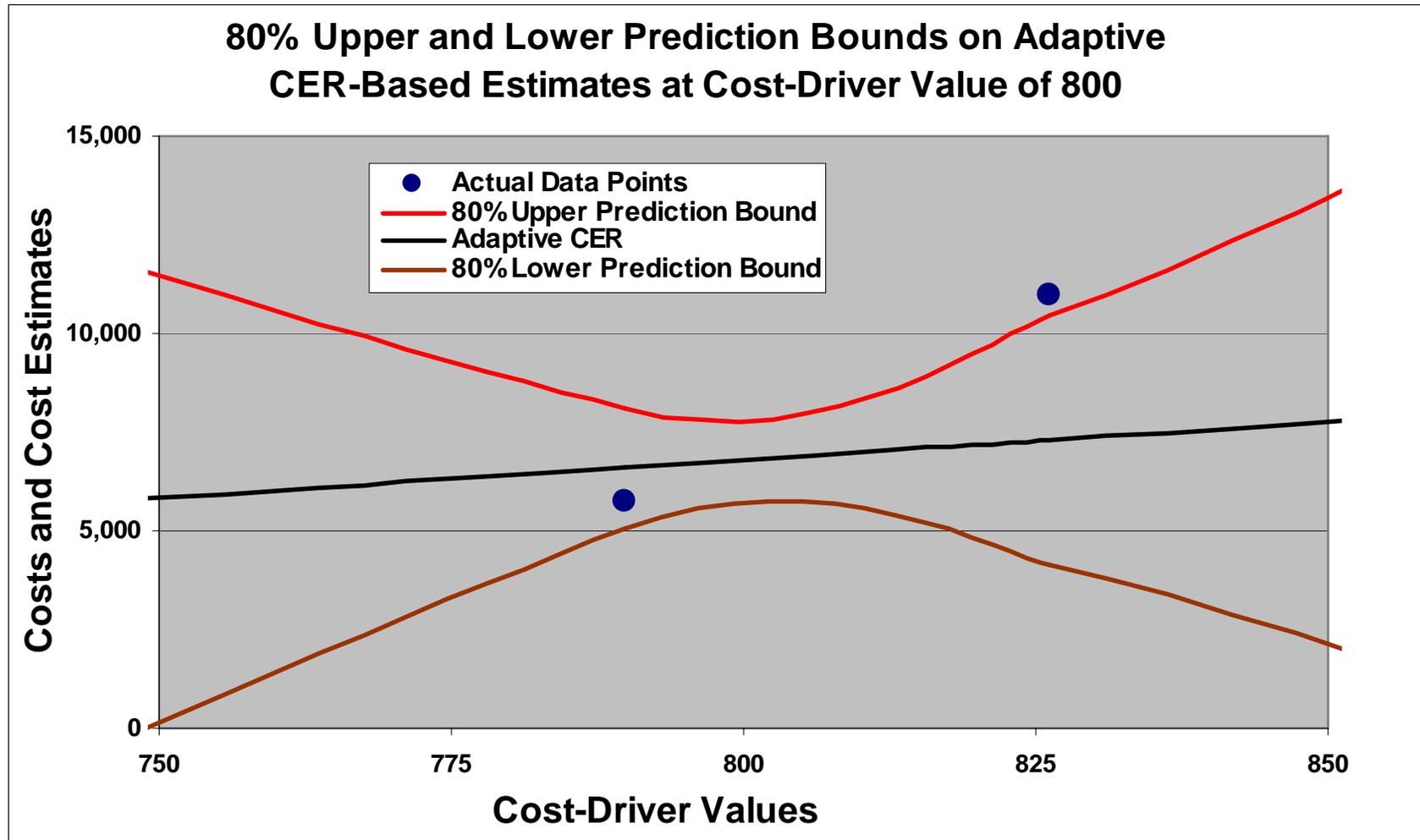
80% Prediction Bounds on the Adaptive CER Based at $x = 800$

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	WLS EST y	80% Lower Bound
A	156.12	51,367.22	67,335.731428	-5,734.14	-78,804.008697
B	179.40	5,885.00	65,146.948025	-5,280.91	-75,708.771200
C	180.30	7,060.00	65,062.330513	-5,263.39	-75,589.110360
D	217.50	139,483.12	61,564.835765	-4,539.16	-70,643.158038
E	419.14	3,386.00	42,608.518817	-613.53	-43,835.578046
F	437.09	6,738.00	40,921.251654	-264.07	-41,449.391167
G	440.93	6,812.00	40,560.306422	-189.31	-40,938.927733
H	494.45	3,291.34	35,529.986321	852.64	-33,824.697703
I	789.90	5,723.14	8,126.533982	6,604.62	5,082.700610
J	826.10	10,992.00	10,459.318778	7,309.38	4,159.436356
K	864.30	11,590.00	15,439.587849	8,053.07	666.561891
L	869.30	15,973.00	16,099.371097	8,150.42	201.463800
M	976.50	7,970.67	30,313.734118	10,237.44	-9,838.849438
N	1,355.80	9,524.10	80,730.945765	17,621.85	-45,487.245014
O	1,360.90	35,927.22	81,409.009710	17,721.14	-45,966.730098
P	1,463.21	11,238.73	95,011.748000	19,712.96	-55,585.820690
Q	2,332.10	92,059.97	210,542.762967	36,628.96	-137,284.838305
R	3,017.73	74,649.00	301,708.981386	49,977.16	-201,754.659776
S	3,253.00	42,915.23	332,992.265384	54,557.52	-223,877.228359
Sums	19,633.77	542,585.74		215,542.66	

80% Prediction Bounds on the Adaptive CER Based at $x = 800$



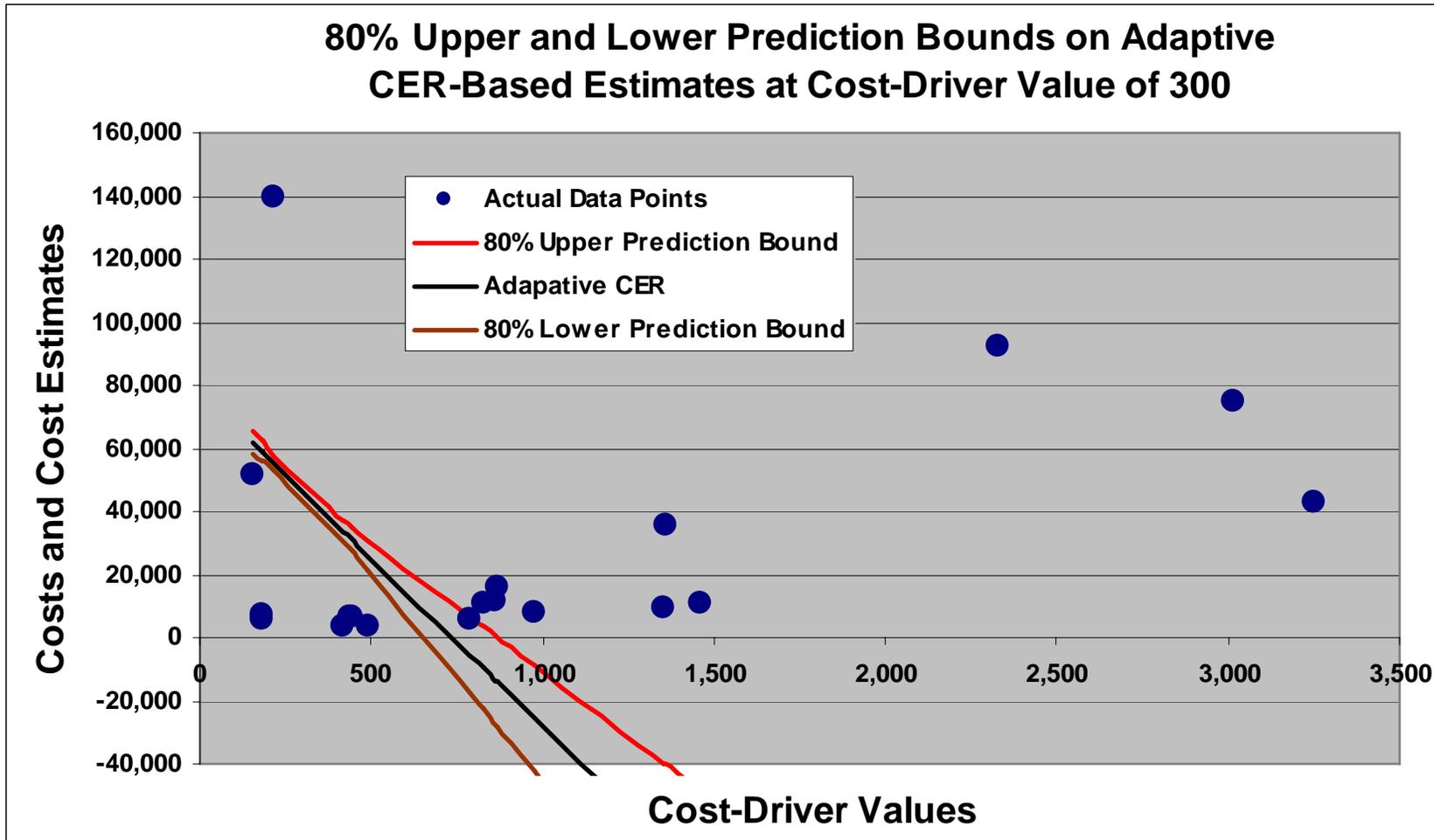
Prediction Bounds Near Cost-Driver Value of 800



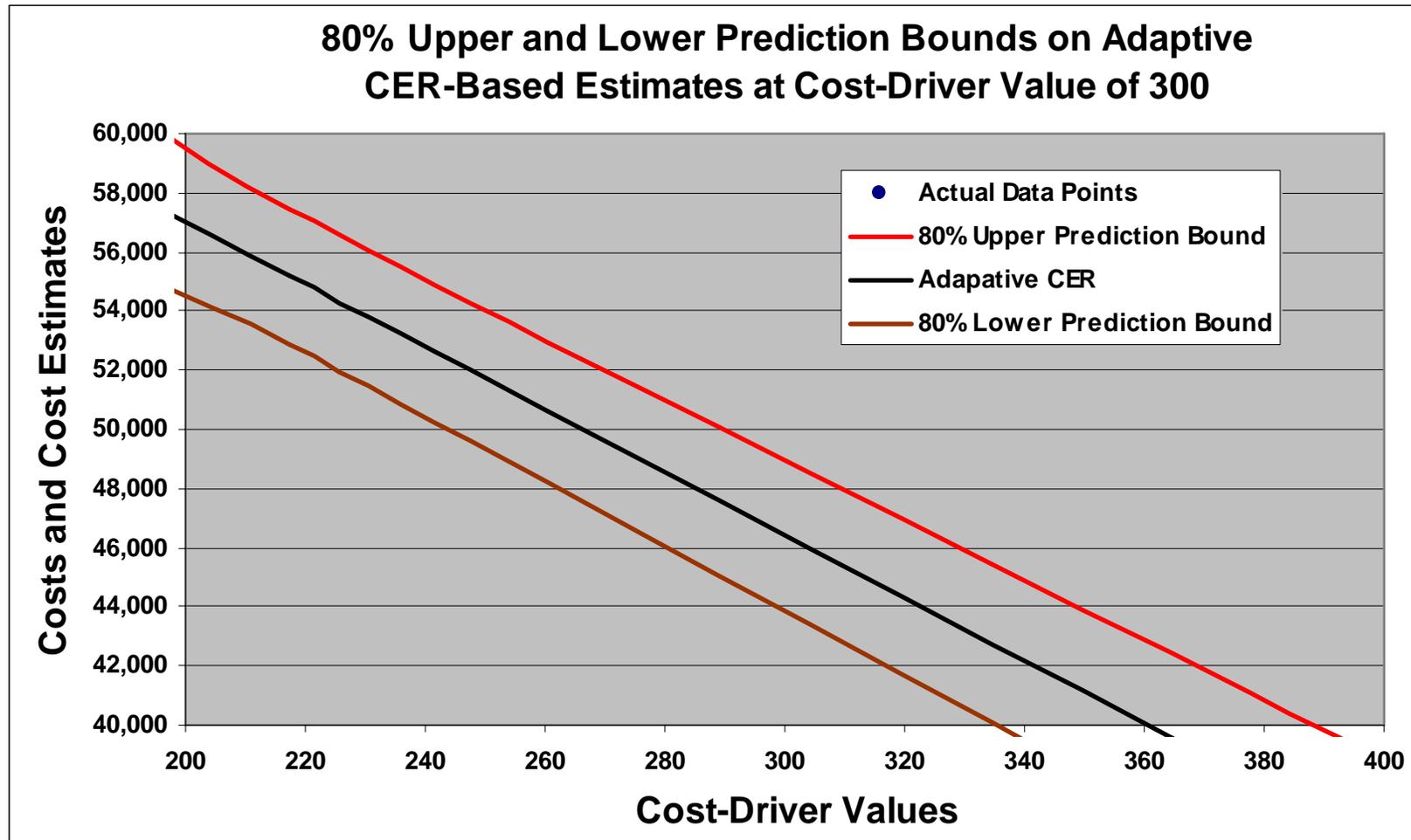
80% Prediction Bounds on the Adaptive CER Based at $x = 300$

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	WLS EST y	80% Lower Bound
A	156.12	51,367.22	65,389.279544	61,698.97	58,008.663971
B	179.40	5,885.00	62,372.227016	59,227.74	56,083.244080
C	180.30	7,060.00	62,255.776784	59,132.20	56,008.619347
D	217.50	139,483.12	57,462.441876	55,183.32	52,904.189048
E	419.14	3,386.00	36,867.788626	33,778.67	30,689.557986
F	437.09	6,738.00	35,381.736102	31,873.23	28,364.726492
G	440.93	6,812.00	35,064.501531	31,465.60	27,866.707881
H	494.45	3,291.34	30,658.711130	25,784.31	20,909.907048
I	789.90	5,723.14	6,491.040727	-5,578.52	-17,648.087346
J	826.10	10,992.00	3,534.857637	-9,421.25	-22,377.363947
K	864.30	11,590.00	415.759782	-13,476.29	-27,368.336816
L	869.30	15,973.00	7.527753	-14,007.05	-28,021.632368
M	976.50	7,970.67	-8,743.802865	-25,386.63	-42,029.453100
N	1,355.80	9,524.10	-39,698.603983	-65,650.37	-91,602.134324
O	1,360.90	35,927.22	-40,114.762116	-66,191.75	-92,268.734323
P	1,463.21	11,238.73	-48,463.042557	-77,052.24	-105,641.431258
Q	2,332.10	92,059.97	-119,355.526647	-169,287.31	-219,219.087245
R	3,017.73	74,649.00	-175,292.373781	-242,068.82	-308,845.271266
S	3,253.00	42,915.23	-194,486.501042	-267,043.38	-339,600.262830
Sums	19,633.77	542,585.74		-597,019.57	

80% Prediction Bounds on the Adaptive CER Based at $x = 300$



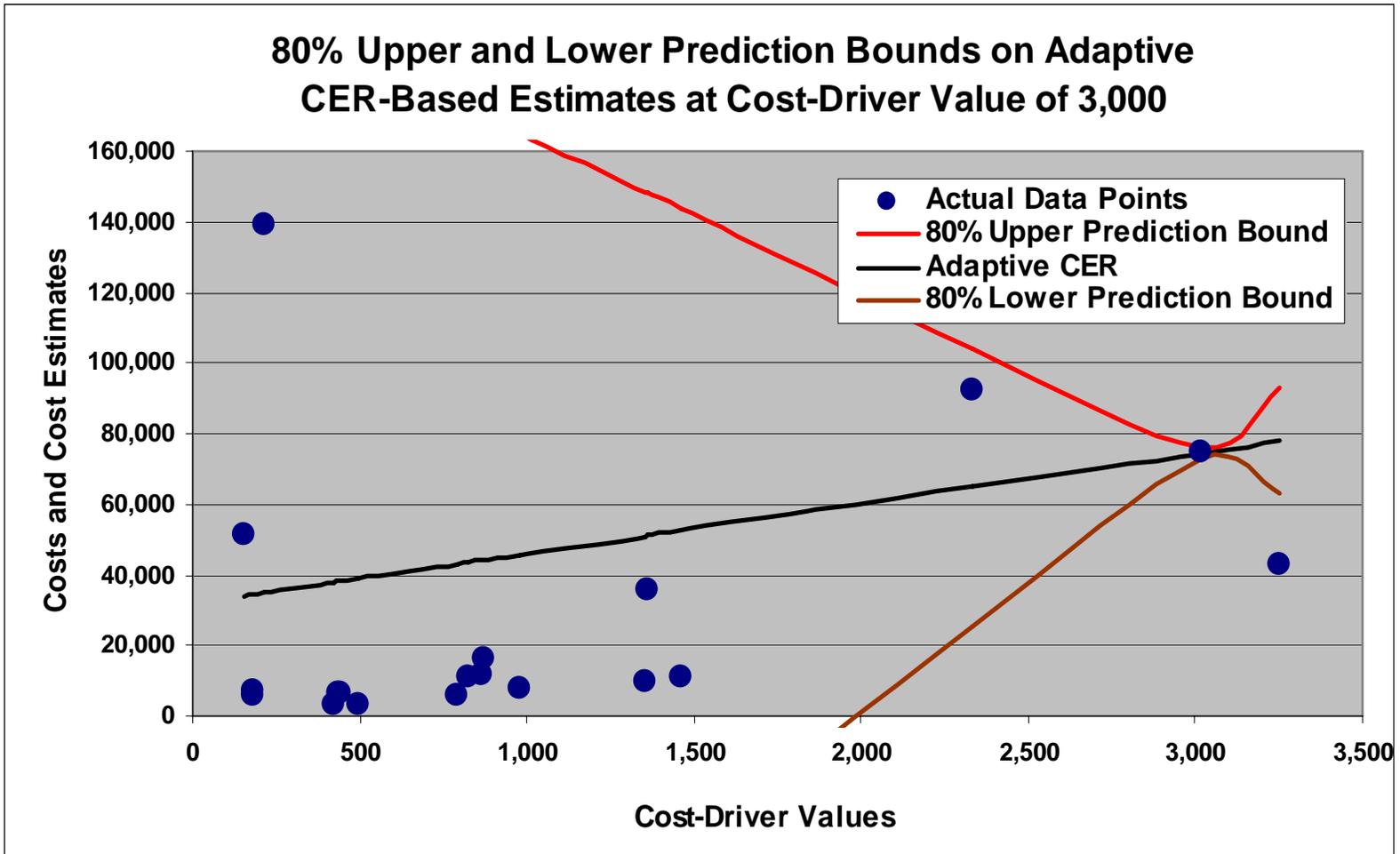
Prediction Bounds Near Cost-Driver Value of 300



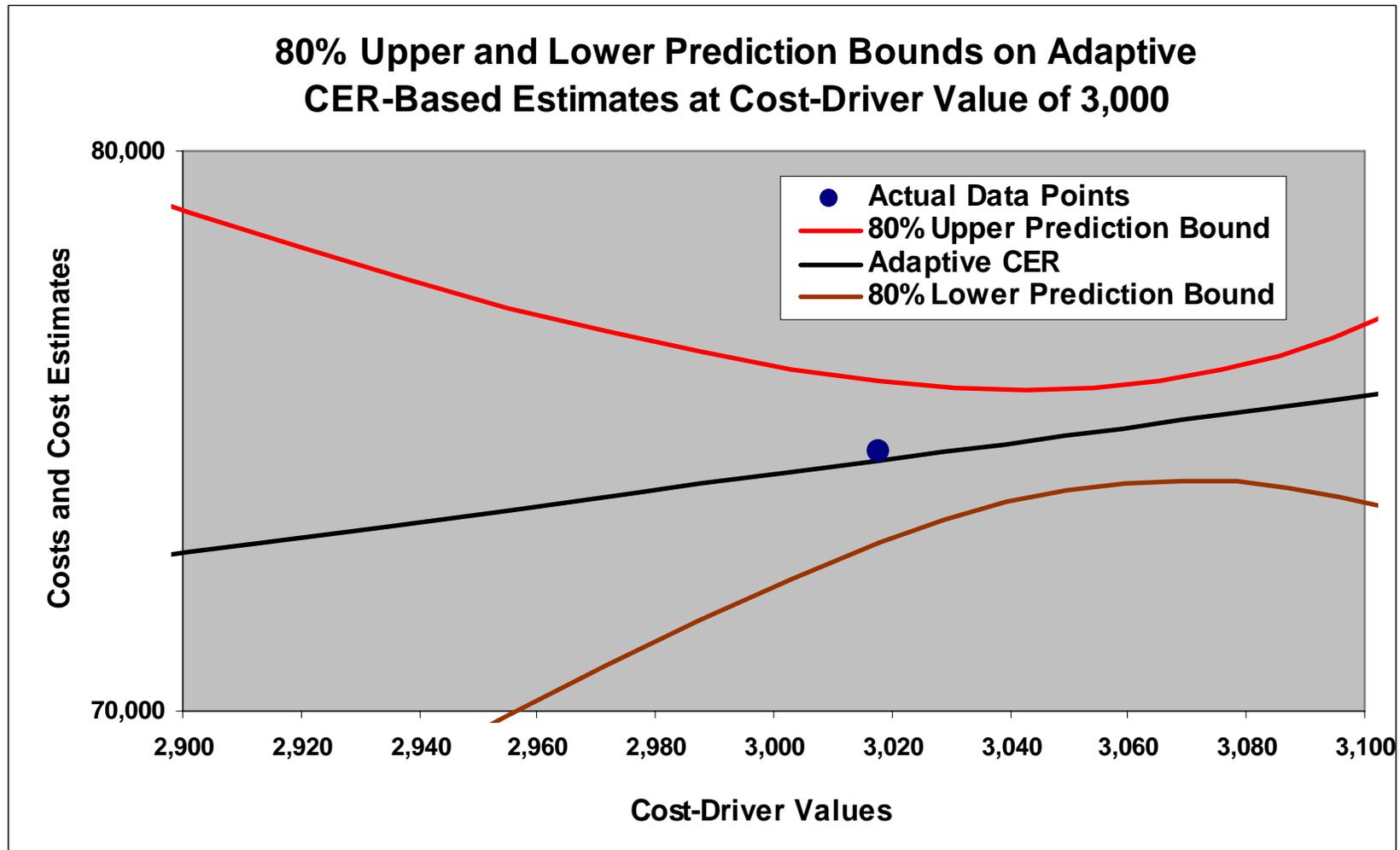
80% Prediction Bounds on the Adaptive CER Based at $x = 3,000$

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	WLS EST y	80% Lower Bound
A	156.12	51,367.22	202,434.005312	34,104.71	-134,224.591913
B	179.40	5,885.00	201,384.901034	34,433.09	-132,518.729992
C	180.30	7,060.00	201,344.342887	34,445.78	-132,452.781730
D	217.50	139,483.12	199,667.940092	34,970.51	-129,726.920845
E	419.14	3,386.00	190,581.137616	37,814.77	-114,951.604146
F	437.09	6,738.00	189,772.232090	38,067.96	-113,636.306880
G	440.93	6,812.00	189,599.184936	38,122.13	-113,354.928569
H	494.45	3,291.34	187,187.341936	38,877.06	-109,433.220060
I	789.90	5,723.14	173,873.151720	43,044.57	-87,784.019292
J	826.10	10,992.00	172,241.840172	43,555.19	-85,131.460894
K	864.30	11,590.00	170,520.403443	44,094.02	-82,332.354836
L	869.30	15,973.00	170,295.084698	44,164.55	-81,965.979897
M	976.50	7,970.67	165,464.262738	45,676.67	-74,110.913120
N	1,355.80	9,524.10	148,371.862469	51,026.94	-46,317.989913
O	1,360.90	35,927.22	148,142.044389	51,098.87	-45,944.294515
P	1,463.21	11,238.73	143,531.737673	52,542.02	-38,447.695941
Q	2,332.10	92,059.97	104,382.272484	64,798.25	25,214.232669
R	3,017.73	74,649.00	75,911.693364	74,469.49	73,027.283557
S	3,253.00	42,915.23	92,744.870060	77,788.12	62,831.365052
Sums	19,633.77	542,585.74		883,094.70	

80% Prediction Bounds on the Adaptive CER Based at $x = 3,000$



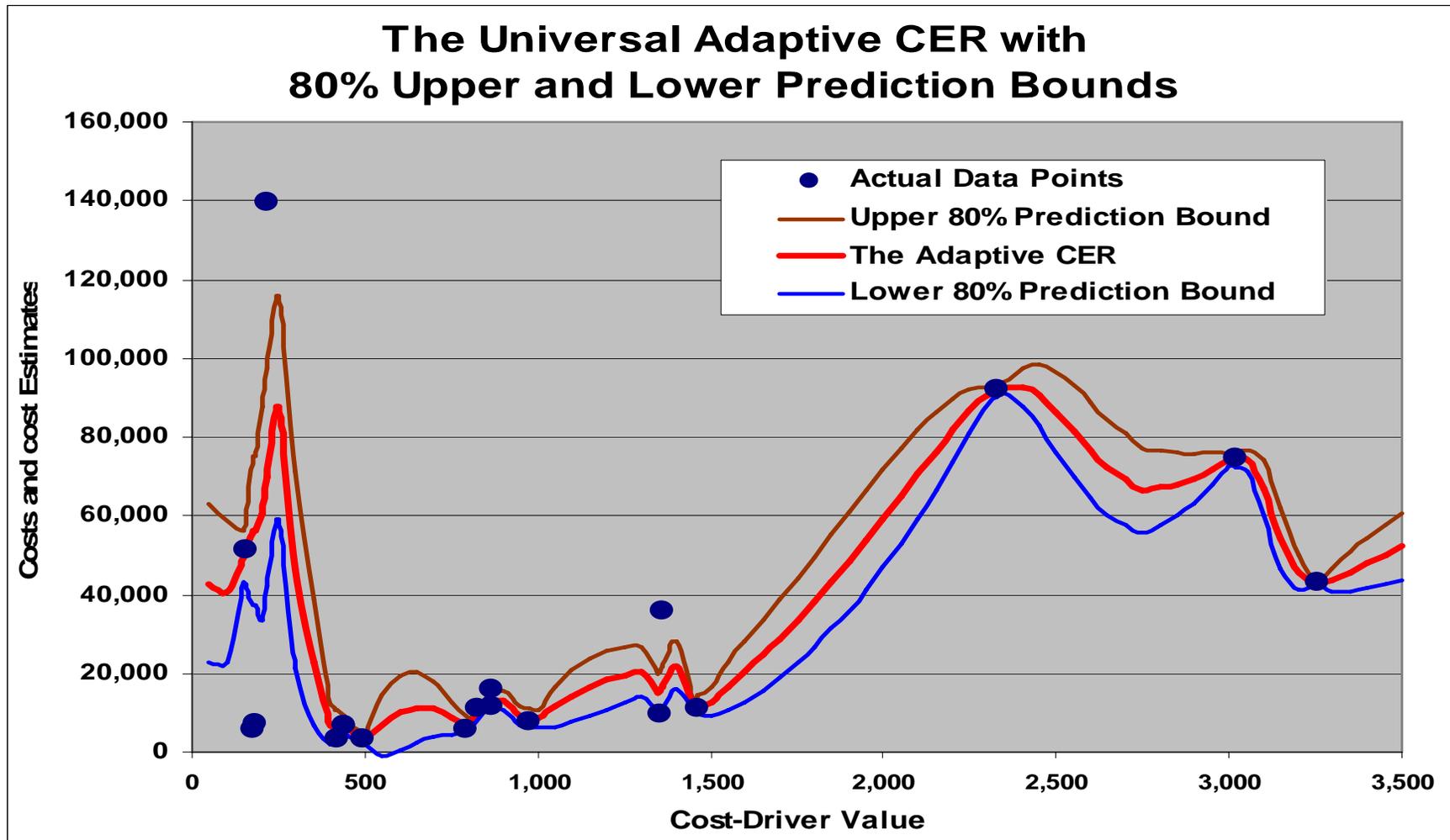
Prediction Bounds Near Cost-Driver Value of 3,000



80% Prediction Bounds on the Universal Adaptive CER

Driver	Cost	80% Upper Bound	EST Cost	80% Lower Bound	Driver	Cost	80% Upper Bound	EST Cost	80% Lower Bound
50.00		62,922.60536	42,739.31	22,556.01954	1,500.00		16,394.47396	12,825.54	9,256.59722
100.00		58,907.24807	40,817.29	22,727.33210	1,550.00		22,698.80390	16,621.72	10,544.64424
150.00		56,054.74733	49,546.82	43,038.89123	1,600.00		28,144.34523	20,492.26	12,840.17489
156.12	51,367.22	59,905.78998	50,880.53	41,855.27867	1,650.00		33,397.70393	24,526.56	15,655.41028
179.40	5,885.00	74,603.67051	55,953.88	37,304.09301	1,700.00		38,715.42463	28,831.03	18,946.63797
180.30	7,060.00	75,171.88754	56,150.02	37,128.14511	1,750.00		44,154.10037	33,415.50	22,676.89870
200.00		87,612.53844	60,443.18	33,273.82964	1,800.00		49,694.94147	38,247.16	26,799.37082
217.50	139,483.12	97,311.65894	69,749.17	42,186.69003	1,850.00		55,295.14895	43,285.50	31,275.84370
250.00		115,347.90219	87,031.73	58,715.55405	1,900.00		60,905.74386	48,497.85	36,089.94751
300.00		71,377.71021	46,425.71	21,473.71561	1,950.00		66,472.33508	53,862.57	41,252.81236
350.00		38,704.87919	22,733.56	6,762.24433	2,000.00		71,925.95418	59,364.10	46,802.23932
400.00		12,204.28249	7,006.95	1,809.62688	2,050.00		77,167.60488	64,981.01	52,794.40890
419.14	3,386.00	10,701.37240	6,760.42	2,819.47622	2,100.00		82,049.60587	70,666.52	59,283.42427
437.09	6,738.00	9,303.25537	6,529.22	3,755.18780	2,150.00		86,358.35570	76,319.27	66,280.18315
440.93	6,812.00	9,004.15958	6,479.76	3,955.36231	2,200.00		89,805.61668	81,744.09	73,682.56956
450.00		8,300.59270	6,362.94	4,425.27919	2,250.00		92,036.69169	86,609.89	81,183.08854
494.45	3,291.34	4,923.86478	3,589.46	2,255.05196	2,300.00		92,664.98297	90,430.47	88,195.96100
500.00		4,503.97498	3,243.16	1,982.35231	2,332.10	92,059.97	93,449.79687	91,836.14	90,222.47807
550.00		14,529.66385	6,829.12	-871.42873	2,350.00		93,889.12293	92,619.98	91,350.84125
600.00		19,484.26578	9,959.40	434.52824	2,400.00		97,402.84769	92,676.25	87,949.65291
650.00		20,409.64947	11,310.17	2,210.70010	2,450.00		98,222.52317	90,463.37	82,704.22441
700.00		18,067.87906	10,929.01	3,790.13759	2,500.00		96,484.73846	86,410.39	76,336.03984
750.00		12,749.77204	8,652.67	4,555.56975	2,550.00		92,951.10708	81,412.53	69,873.94518
789.90	5,723.14	9,150.40455	7,175.24	5,200.06839	2,600.00		88,646.33546	76,466.46	64,286.59020
800.00		8,241.00254	6,801.25	5,361.49607	2,650.00		84,454.68294	72,322.92	60,191.16611
826.10	10,992.00	11,221.66628	9,756.59	8,291.51518	2,700.00		80,929.09901	69,366.76	57,804.41474
850.00		13,951.60979	12,462.82	10,974.03604	2,750.00		77,054.82704	66,431.86	55,808.89003
864.30	11,590.00	14,422.75320	12,666.50	10,910.25030	2,800.00		76,663.27434	67,242.40	57,821.52197
869.30	15,973.00	14,587.63569	12,737.72	10,887.80057	2,850.00		75,905.67799	67,904.22	59,902.76737
900.00		15,604.93947	13,174.99	10,745.03389	2,900.00		75,867.66447	69,545.45	63,223.22554
950.00		11,653.20568	9,208.15	6,763.08930	2,950.00		76,119.31586	71,913.26	67,707.21079
976.50	7,970.67	11,143.81693	8,832.68	6,521.53760	3,000.00		75,518.29497	74,219.40	72,920.49668
1,000.00		10,696.45067	8,499.71	6,302.97553	3,017.73	74,649.00	75,966.58830	74,164.83	72,363.08019
1,050.00		16,599.85083	11,462.16	6,324.47866	3,050.00		76,786.35756	74,065.53	71,344.69813
1,100.00		20,939.42063	14,296.49	7,653.56932	3,100.00		74,002.10190	67,141.02	60,279.92945
1,150.00		23,860.71099	16,537.15	9,213.58728	3,150.00		62,069.86209	54,415.99	46,762.11593
1,200.00		25,595.54200	18,230.99	10,866.44026	3,200.00		49,725.94543	45,424.15	41,122.36282
1,250.00		26,430.13239	19,495.31	12,560.49388	3,250.00		43,144.36743	42,927.10	42,709.82647
1,300.00		26,643.54266	20,310.23	13,976.92318	3,253.00	42,915.23	43,354.47526	42,978.65	42,602.82964
1,350.00		19,965.36192	14,974.31	9,983.26614	3,300.00		46,653.41208	43,786.36	40,919.29842
1,355.80	9,524.10	20,888.41315	15,774.27	10,660.11771	3,350.00		50,744.79550	45,762.39	40,779.98882
1,360.90	35,927.22	21,705.81036	16,477.67	11,249.53159	3,400.00		54,430.68793	47,971.96	41,513.24151
1,400.00		27,979.59574	21,870.45	15,761.29785	3,450.00		57,664.17277	50,126.95	42,589.72220
1,450.00		13,664.60151	11,840.86	10,017.11120	3,500.00		60,512.13570	52,149.51	43,786.89143
1,463.21	11,238.73	14,382.93075	12,101.01	9,819.08646					

80% Prediction Bounds on the Universal Adaptive CER



Contents

- **Ordinary Least-Squares (OLS) Regression CERs and their Quality Metrics**
- **Weighted Least-Squares (WLS) Regression CERs and their Quality Metrics**
- **Adaptive CERs via Quadratic-Distance Weighting**
- **The “Universal Adaptive CER”**
- **Prediction Bounds for OLS, WLS, and Adaptive CERs**
- **Conclusions**

Concluding Remarks

- **CERs are the Mainstay of Parametric Cost Estimating – Their Major Drawback is the Uncertainty of Applicability in Any Particular Estimating Situation**
- **Deriving Adaptive CERs Requires More Work Than Deriving Full Data Set CERs, but it Offers the Possibility of More Credible and Precise Estimates**
 - Adaptive CERs Can be “Tuned” to Specific Cost-Driver Ranges, or
 - They Can be Merged to Form a Universal Adaptive CER that is Valid over the Entire Cost-Driver Range
- **In This Report, We Have Provided the Mathematical Foundations of Calculating and Applying Adaptive CERs**

References

- 1. Book, S.A., “Deriving Cost-Estimating Relationships Using Weighted Least-Squares Regression,” IAA/ISPA/AIAA Space System Cost Methodologies and Applications Symposium, San Diego CA, 10-11 May 1990, 19 pages.**
- 2. Book, S.A.and Broder, M.A., “Adaptive Cost-Estimating Relationships,” Space Systems Cost Analysis Group (SSCAG) Meeting in Noordwijk, Holland, 15-16 May 2008, and ISPA/SCEA Joint Annual Conference & Training Workshop, Industry Hills CA, 24-27 June 2008, 80 Charts.**
- 3. Foussier, P., “Space System Study,” Space Systems Cost Analysis Group (SSCAG), Montreal QB, 16 May 2007, 41 Charts.**