Prediction Bounds for General-Error-Regression CERs

Stephen A. Book
Chief Technical Officer
MCR, LLC
390 No. Sepulveda Blvd.
El Segundo, CA 90254
(310) 640-0005 x244
sbook@mcri.com

39th Annual DoD Cost Analysis Symposium
Williamsburg VA
14-17 February 2006

Abstract

Estimating the cost of a system under development is essentially trying to predict the future, which means that any such estimate contains uncertainty. A portion of this uncertainty is described by the "standard error of the estimate" of a cost-estimating relationship (CER), which is basically the standard deviation of errors made (the "residuals") in using that CER to estimate the (known) costs of the systems comprising the historical data base. The standard error of the estimate depends primarily on the extent to which those (known) costs fit the CER that purports to model them. However, additional uncertainty arises from the location of the particular cost-driver value (*x*) within or without the range of cost-driver values for programs comprising the historical cost data base. For example, if *x* were located near the center of the range of its historical values, the CER would provide a more precise measure of the element's cost than if *x* were located far from the center of the range. The total uncertainty in the estimate can then be expressed in terms of prediction bounds that involve both sources of uncertainty.

The first kind of uncertainty, represented by only one number characteristic of the CER, is fairly easy to measure for any CER shape or error model. The second kind, which involves both the CER itself and the value of the cost-driving parameter, however, is more complicated, and the way to calculate it is completely understood only in the case of classical linear regression, i.e. "ordinary least squares" (OLS). As a result, explicit formulas exist for "prediction intervals" that bound cost estimates based on CERs that have been derived by applying OLS to historical cost data. For CERs derived by other statistical methods, there appears to be no general method of solution described in the theoretical statistical literature. This presentation demonstrates the application of bootstrap random sampling, a 28-year-old statistical process, to the problem of estimating prediction bounds for multiplicative-error CERs and other CERs derived by non-OLS methods. After the bootstrap method is shown to yield prediction bounds that approximate the known OLS bounds fairly well, it is applied analogously to non-OLS-derived CERs. Although statistical sampling can yield only approximations to the "true" prediction bounds, the bootstrap technique appears to be a practical and theoretically credible method of approaching this currently unsolved estimating problem.

Acknowledgments

This Research was Supported and Funded by the

NRO Cost Group
Chantilly VA

and the

IC CAIG
Langley VA

The presenter would like to thank Erik Burgess of Burgess Consulting, Inc., and Tim Anderson of The Aerospace Corporation for their advice and assistance, and NCG contractor team members Steve Rozzi of Key Technologies, Don MacKenzie and Mitch Robinson of Wyle Labs, and Nathan Menton of MCR for their valuable comments and suggestions during the course of this work.

Contents

- Prediction Bounds for Cost-Estimating Relationships (CERs) Derived by "Ordinary Least Squares" (OLS) Regression
- The "Bootstrap" Method of Statistical Sampling
- Minimum-Percentage-Error/Zero-Percentage-Bias (MPE-ZPB) CERs
- Using Bootstrap Sampling to Establish Prediction Bounds in the MPE-ZPB Situation
  - Factor CERs
  - Linear CERs
  - Power CERs
  - Triad CERs
- Summary

Error Sources of CER-Based Cost Estimates

1. **Inability of Any CER to Account for All Influences on Cost, No Matter How Many Inputs it Allows**
2. **Incorrectness of Algebraic CER Model to which Cost Numbers in Data Base are Statistically Fit**
   - Explicit CERs are Derived from Historical Cost Data by Minimizing a Quality Metric, Often the Standard Error of the Estimate (SEE), <u>that Depends on the Algebraic Model</u>
   - SEE is an Estimator of "True" Standard Deviation σ of "Errors" in the Knowledge Base of Historical Cost Data Points, <u>Assuming the Algebraic Model is Correct</u>
3. **Location of Cost Driver Value x among Parameter Values Comprising Historical Cost Data Base**
   - If x is Located Near Center of Range of Parameter Values, CER will Provide More Precise Estimate of the System's Cost
   - If x is Located Far From Center of Range, CER-based Estimate will be Considerably Less Precise

State of the Art in CER-Based-Estimating

- **Ordinary Least Squares (OLS)**
  - Model Cost as an Additive-Error Linear Function of One or More Cost Drivers
  - Estimating Problem is Completely Solved
  - Explicit Algebraic Formulas Exist for the Upper and Lower Bounds on the Cost ("Prediction Intervals") for Any Value of Cost-Driving Parameter at Any Level of Confidence
  - Width of Prediction Interval Depends on Both the CER's Standard Error of the Estimate (SEE) and the Location of the Cost-Driver Value x
- **Special Nonlinear CER Forms**
  - Model Cost as One of a Particular Class of Nonlinear Functions
  - Such Nonlinear Forms Can be Made OLS-Solvable by an Algebraic Transformation (usually Logarithmic of a Multiplicative-Error Model)
  - Non-Optimal Prediction Intervals Can be Calculated in a Roundabout Way by Applying an Inverse Transform
- **General Nonlinear CER Forms**
  - Model Cost Using Any Additive- or Multiplicative-Error, Linear or Nonlinear Functional Form
  - Standard Error of the Estimate Can be Calculated, as Well as Some Information About Variances of the Coefficients
  - But Prediction-Interval Problem Appears Not to Have Been Solved

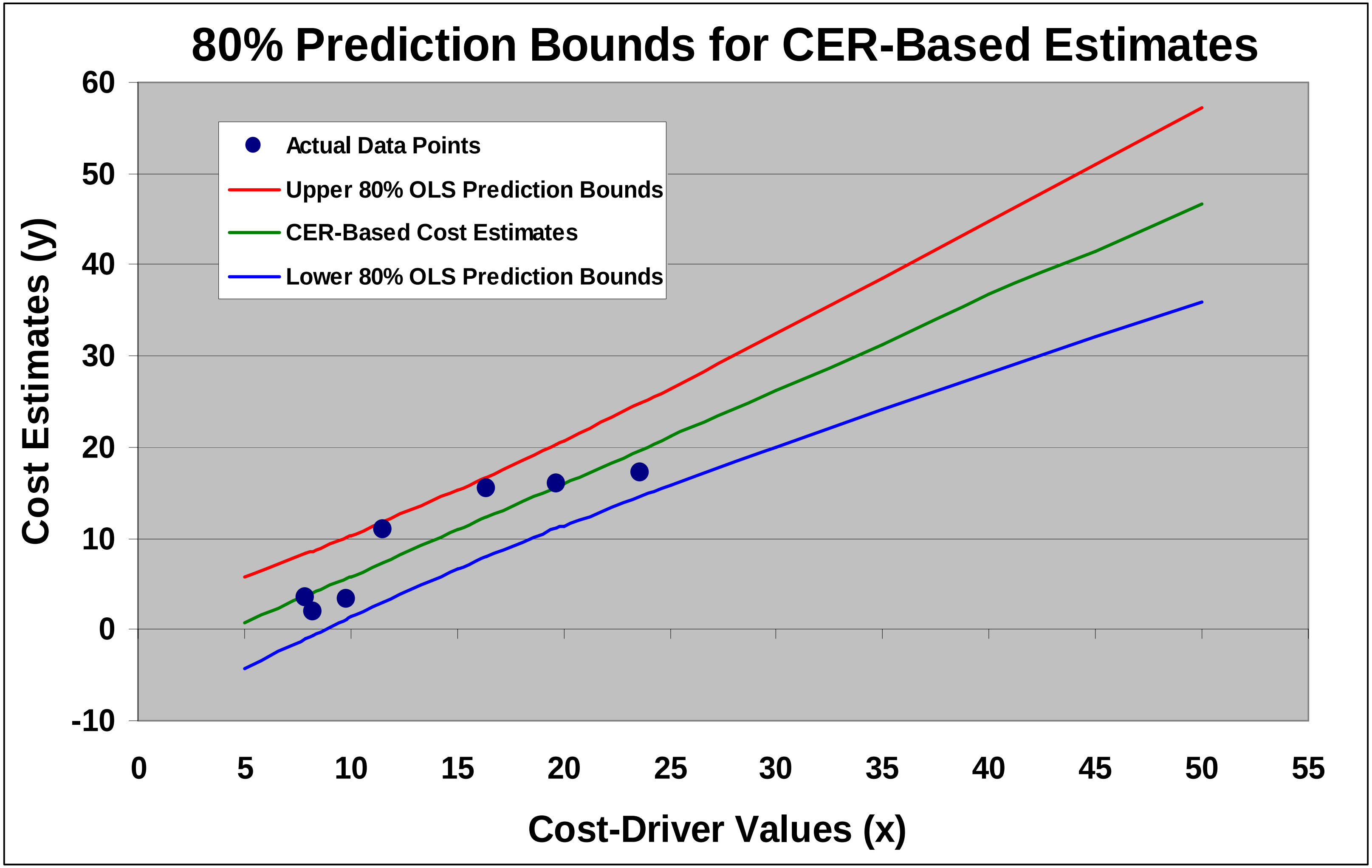
Bounding the Cost of an Element of the Program You're Estimating

- **For CERs Derived by OLS Regression, the Prediction Interval for the Actual Cost Y is Based on the Estimated Cost $\hat{Y}$ and the SEE in the Following Way:**

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

- **The Degree of Confidence Associated With this Interval is (1-α)100%, Enforced by Choice of the "Percentage Point of the *t* Distribution," Namely $t_{\alpha/2,n-2}$**

Prediction Bounds Calculated by the OLS Statistical Formula



The Bootstrap Sampling Approach when a Formula is Unavailable

- While We Wait for the "Exact" Theoretical Solution to be Found for non-OLS CERs (which, if history is a guide, could take decades), It Would be Useful to Have Available a Practical "Ad Hoc" Method that We Can Apply to Generate Prediction Intervals in Any Particular Case
- "Bootstrap" Statistical Sampling Appears to be an Appropriate Technique to Consider
  - The Bootstrap Method of Error Estimation was Introduced by B. Efron in 1977 and Has a 28-Year History Behind It
  - It is a "Distribution-Free" Method, so It Does <u>not</u> Require the Usual (and questionable) Distributional Assumptions, e.g., Normal or Lognormal Error Distributions or even Homoscedasticity
  - It Works with Additive- or Multiplicative-Error Models and all Algebraic Functional Forms

The Bootstrap Philosophy

- **The Bootstrap Philosophy Parallels the Philosophy of OLS and General-Error Regression**
- **It Assumes that …**
  - **The Relationship between y (Cost) and x (Cost-Driving Parameter) is Exactly the Algebraic Relationship that is Being Modeled**
  - **The x Values are Known Precisely, but "Actual" y Values are Known Accurately Only within Some Statistical Error Distribution**
    - **The Error Distribution Depends on How Well the Algebraic Relationship with x Accounts for the Various Influences on y**
    - **The Set of "Residuals" (namely, the differences between "estimated" and "actual" costs) Represents the Distribution of Error in the Actual y Values**

A Major Consequence of the Bootstrap Philosophy

- The Combined Assumptions of the Bootstrap Philosophy Imply that the Residual that Happens to be Matched with any Particular x Value is Merely a Matter of Chance
- Residuals are Assumed to be Randomly (i.e., equally likely) Selected from an (unknown) Error Distribution
  - In OLS this Error Distribution is Assumed to be the Normal
  - Bootstrapping Does Not Require the Normal Distribution – its Error Distribution is Defined Solely by the Residuals
- So if We Had Collected Our Data in a Different Way or at a Different Time, We Might have Obtained Any One of the Residuals for Any of the x Values
- Therefore, if We Randomly Choose a Residual for Each x Value, We Will be Able Construct a Set of y Values that Could Very Well Have Been the "Actual" y Values

What is Bootstrap Sampling?

- **Bootstrap Sampling is a "Resampling" Method**
- **Random Samples are Taken, not from a Probability Distribution, but from the Set of Residuals**
  - **Residuals are Calculated from the <u>Actual</u> Data Base with which We are Working (<u>not</u> from an assumed probability distribution such as the normal or lognormal)**
  - **Bootstrap Theory Assumes that Each of the n Residuals (n = number of data pairs) has Probability $1/n$ of Being <u>the</u> Residual Associated with any Given x Value**
  - **This Assumption Requires Sampling "with Replacement," because a Residual's Association with One x Value Does not Preclude its Association with Another in the Same Data Set**
  - **Only the Probability $1/n$ is Hard-wired into the Process**
- **Many Sets of n Sample Residuals are Generated**
  - **Then a New Data Set ("the Bootstrap Sample") that "Could Have Been" the Actual Data Set is Calculated from Each Set of n Sample Residuals**
  - **Next, a "CER" that "Could Have Been" the Real CER is Calculated from Each Bootstrap Sample**

Prepare Data Base for Bootstrap Sampling – Calculate the Residuals

x Values (Cost Driver)	y Values (Actual Costs)	Predicted y Values (Cost Estimates)	Residuals = Actuals-Estimates
7.9	3.595	3.699	0.104
8.2	1.900	4.005	2.105
9.8	3.300	5.635	2.335
11.5	10.900	7.367	-3.533
16.4	15.434	12.358	-3.076
19.7	16.074	15.720	-0.354
23.6	17.274	19.693	2.419

Note: CER derived from x and actual y values is $y = a + bx$, where a = -4.348 and b = 1.0187.

Draw Random Samples of Residuals

x Values (Cost Driver)	Residual Samples:	#1	#2	#3	#4	#5	#6	...
7.9	1st Residual	0.104	2.105	2.335	-3.533	-3.076	2.419	...
8.2	2nd Residual	-0.354	2.419	0.104	-0.354	2.335	2.419	...
9.8	3rd Residual	0.104	-3.533	-3.533	2.105	2.335	-0.354	...
11.5	4th Residual	2.419	-3.533	2.419	2.419	2.105	2.419	...
16.4	5th Residual	2.105	2.105	-3.533	2.105	-0.354	-0.354	...
19.7	6th Residual	2.105	-3.533	2.105	2.105	2.419	-3.076	...
23.6	7th Residual	-3.533	2.419	-0.354	0.104	2.419	0.104	...

Note: Sampling is done "with replacement", so some residuals will appear more thar once in the same sample.

Compute Bootstrap Samples of "Possible" Actual y Values

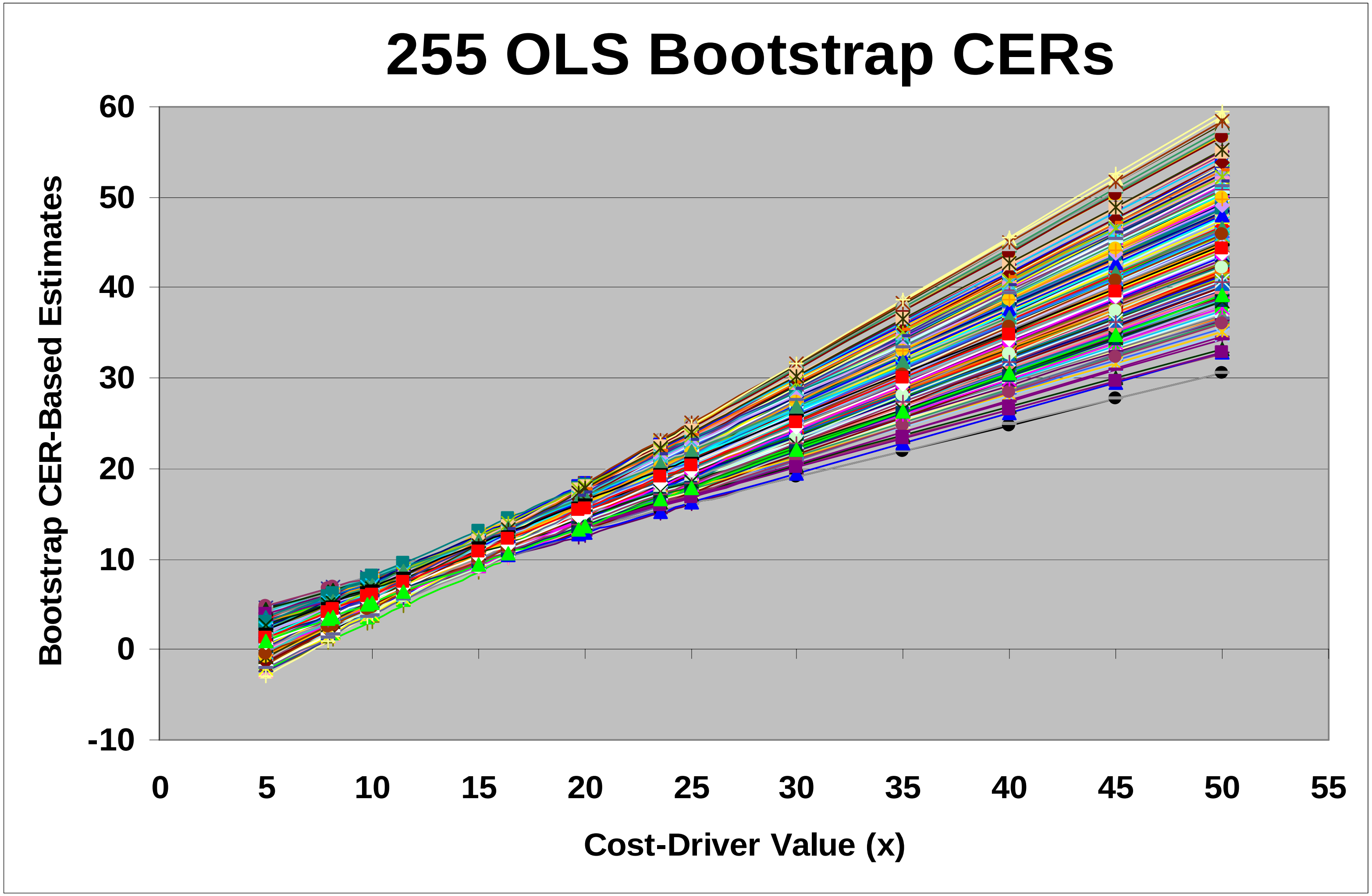
x Values (Cost Driver)	Bootstrap Actual = Estimate + Residual	#1	#2	#3	#4	#5	#6	...
7.9	3.699+1st Residual	3.804	5.804	6.034	0.166	0.624	6.118	...
8.2	4.005+2nd Residual	3.651	6.424	4.109	3.651	6.340	6.424	...
9.8	5.635+3rd Residual	5.739	2.102	2.102	7.740	7.970	5.281	...
11.5	7.367+4th Residual	9.785	3.833	9.785	9.785	9.472	9.785	...
16.4	12.358+5th Residual	14.463	14.463	8.825	14.463	12.004	12.004	...
19.7	15.720+6th Residual	17.825	12.187	17.825	17.825	18.139	12.644	...
23.6	19.693+7th Residual	16.159	22.112	19.339	19.797	22.112	19.797	...

Note: Each bootstrap sample is treated as if it were a set of "actual" data. The only use made of the real actual data set is to calculate the estimates and residuals.

Use Each Bootstrap Sample to Calculate an OLS "CER"

x Values (Cost Driver)	Bootstrap Actuals = Estimate + Residual #1	#2	#3	#4	#5	#6	...
7.9	3.804	5.804	6.034	0.166	0.624	6.118	...
8.2	3.651	6.424	4.109	3.651	6.340	6.424	...
9.8	5.739	2.102	2.102	7.740	7.970	5.281	...
11.5	9.785	3.833	9.785	9.785	9.472	9.785	...
16.4	14.463	14.463	8.825	14.463	12.004	12.004	...
19.7	17.825	12.187	17.825	17.825	18.139	12.644	...
23.6	16.159	22.112	19.339	19.797	22.112	19.797	...
a (Intercept)	-2.598	-4.862	-4.044	-5.330	-4.866	-0.739	...
b (Slope)	0.923	1.040	0.992	1.140	1.140	0.795	...
r	0.943	0.899	0.919	0.960	0.962	0.954	...
r Squared	88.86%	80.86%	84.49%	92.15%	92.50%	90.92%	...
Std Error of Estimate	2.193	3.393	2.851	2.233	2.179	1.686	...

Graphs of 255 Bootstrap "CERs"



Use Each Bootstrap CER to Estimate Cost at Various x Values

Cost Driver (x) Values	Bootstrap Cost Estimates (y Values) 1	2	3	4	5	6	...
5	2.017	0.337	0.916	0.372	0.836	3.238	...
7.9	4.693	3.352	3.793	3.679	4.142	5.544	...
8.2	4.970	3.664	4.091	4.022	4.484	5.783	...
9.8	6.446	5.327	5.678	5.846	6.309	7.055	...
10	6.631	5.535	5.876	6.074	6.537	7.214	...
11.5	8.015	7.095	7.365	7.785	8.247	8.407	...
15	11.245	10.734	10.837	11.777	12.238	11.191	...
16.4	12.537	12.190	12.226	13.373	13.835	12.304	...
19.7	15.583	15.621	15.499	17.137	17.597	14.929	...
20	15.860	15.933	15.797	17.479	17.940	15.168	...
23.6	19.182	19.676	19.368	21.585	22.044	18.031	...
25	20.474	21.131	20.757	23.182	23.641	19.144	...
30	25.089	26.330	25.717	28.884	29.342	23.121	...
35	29.703	31.529	30.677	34.586	35.043	27.098	...
40	34.317	36.727	35.638	40.289	40.745	31.074	...
45	38.932	41.926	40.598	45.991	46.446	35.051	...
50	43.546	47.125	45.558	51.694	52.147	39.028	...

Rank All 255 Estimates Associated with Each Cost-Driver Value x

Estimate Ranks	x = 5	x = 15	x = 50
1	-3.021	8.583	30.604
2	-2.434	8.586	30.629
3	-2.387	8.805	32.750
4	-2.381	8.918	32.791
5	-2.378	9.057	32.913
6	-2.311	9.096	32.922
7	-2.112	9.108	33.173
8	-2.100	9.273	34.280
9	-1.889	9.301	34.672
10	-1.889	9.352	34.862
11	-1.848	9.390	35.151
12	-1.717	9.426	35.517
13	-1.681	9.534	36.022
14	-1.680	9.581	36.132
15	-1.608	9.582	36.417
16	-1.562	9.589	36.497
17	-1.526	9.652	36.763
18	-1.489	9.655	36.819
19	-1.376	9.677	36.843
20	-1.338	9.715	36.857
21	-1.334	9.728	37.228
22	-1.330	9.731	37.388
23	-1.293	9.750	37.731
24	-1.281	9.762	37.754
25	-1.272	9.780	37.848
26	-1.235	9.787	38.052
27	-1.206	9.796	38.122
28	-1.192	9.813	38.444
29	-1.144	9.822	38.449
30	-1.101	9.851	38.471

Lower 10th Percentile

Estimate Ranks	x = 5	x = 15	x = 50
226	2.861	11.970	52.659
227	2.942	11.971	52.908
228	2.951	11.975	52.953
229	2.989	11.977	53.272
230	2.989	12.009	53.285
231	3.008	12.014	53.289
232	3.012	12.069	53.303
233	3.042	12.073	53.346
234	3.045	12.080	53.437
235	3.205	12.084	53.455
236	3.238	12.098	53.493
237	3.254	12.103	53.535
238	3.289	12.110	53.752
239	3.350	12.114	54.388
240	3.415	12.119	54.509
241	3.439	12.136	54.650
242	3.577	12.146	54.852
243	3.598	12.171	55.028
244	3.890	12.202	55.142
245	3.909	12.209	55.163
246	4.006	12.238	56.649
247	4.161	12.270	56.747
248	4.173	12.305	56.915
249	4.174	12.336	57.034
250	4.274	12.480	57.600
251	4.461	12.534	57.894
252	4.520	12.539	58.082
253	4.556	12.613	58.398
254	4.620	12.659	58.901
255	4.896	13.155	59.453

Upper 10th Percentile

Interpolate, if Necessary, to Find 10th and 90th Percentile Estimates

Estimate Ranks	x = 5	x = 15	x = 50
Rank 25	-1.272	9.780	37.848
Lower 80% Bound	-1.253	9.783	37.950
Rank 26	-1.235	9.787	38.052
Rank 229	2.989	11.977	53.272
Upper 80% Bound	2.989	11.993	53.279
Rank 230	2.989	12.009	53.285

Note: The 10th percentile of the estimates at any value of x serves as the lower 80% bootstrap bound, and the 90th percentile serves as the upper 80% bootstrap bound. 80% of the bootstrap estimates lie between those two numbers.

Can We Derive Prediction Bounds from Bootstrap Bounds?

- It Turns Out that the "Bootstrap Bounds," namely the 10^{th} and 90^{th} Percentile Bootstrap Estimates, for any Given x Value are Closer Together than the Known 80% Lower and Upper OLS Prediction Bounds
- After some Theoretical Investigations and Numerical Experimentation, We Found that Adjusting the OLS Bootstrap Bounds Outward by an Additive Amount Equal to the SEE of the "Real" CER Brought the so-called "Bootstrap-Based" Bounds Closer to the OLS Prediction Bounds

Lower 80% Bootstrap-Based Bounds vs. OLS Prediction Bounds

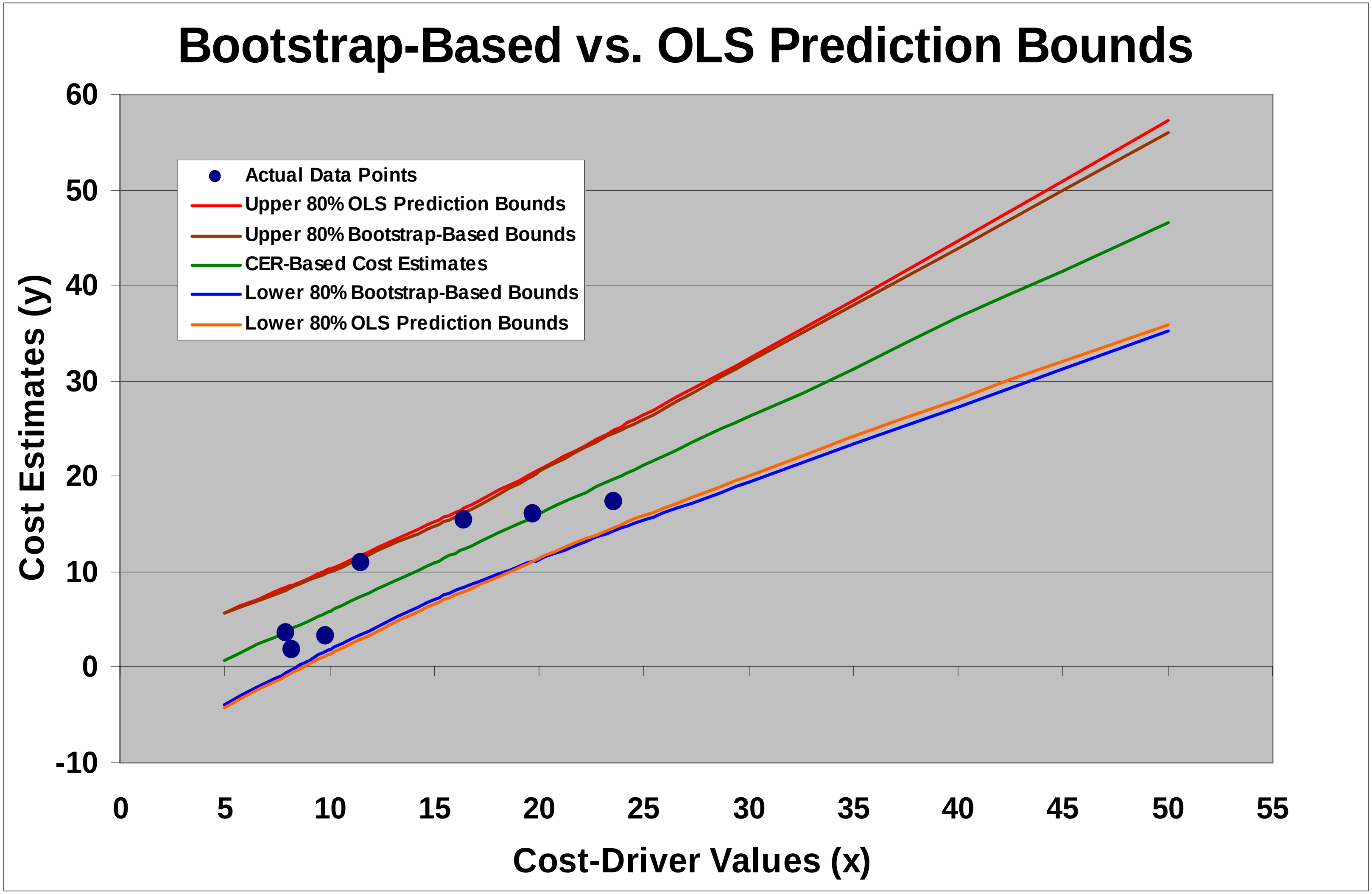
Cost Driver (x) Values	Bootstrap 80% Lower Bounds	Bootstrap-Based 80% Lower Bounds	OLS Prediction 80% Lower Bounds	Differences: B-BB vs. OLS Absolute	Percentage
5	-1.253	-4.003	-4.214	0.2105	4.9958%
7.9	2.176	-0.574	-0.931	0.3572	38.3653%
8.2	2.546	-0.204	-0.598	0.3938	65.8887%
9.8	4.358	1.608	1.158	0.4501	38.8756%
10	4.569	1.819	1.375	0.4443	32.3172%
11.5	6.222	3.472	2.980	0.4922	16.5158%
15	9.783	7.033	6.582	0.4513	6.8568%
16.4	11.110	8.360	7.965	0.3946	4.9537%
19.7	13.795	11.045	11.103	0.0576	0.5188%
20	14.037	11.287	11.380	0.0932	0.8189%
23.6	17.052	14.302	14.617	0.3158	2.1603%
25	18.238	15.488	15.837	0.3489	2.2030%
30	22.188	19.438	20.058	0.6196	3.0891%
35	26.097	23.347	24.128	0.7810	3.2368%
40	30.039	27.289	28.104	0.8142	2.8973%
45	33.945	31.195	32.018	0.8223	2.5683%
50	37.950	35.200	35.890	0.6905	1.9239%
Mean(x) =	13.871				
StdDev(x) =	5.670				
Std Error =	2.750				

Note: "B-BB" = Bootstrap-Based Bounds"

Upper 80% Bootstrap-Based Bounds vs. OLS Prediction Bounds

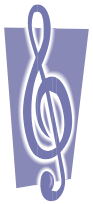
Cost Driver (x) Values	Bootstrap 80% Upper Bounds	Bootstrap-Based 80% Upper Bounds*	OLS Prediction 80% Upper Bounds	Differences: B-BB vs. OLS Absolute	Percentage
5	2.989	5.739	5.704	0.0349	0.6121%
7.9	5.382	8.132	8.330	0.1978	2.3749%
8.2	5.658	8.408	8.608	0.2000	2.3235%
9.8	7.072	9.823	10.112	0.2894	2.8621%
10	7.223	9.973	10.303	0.3296	3.1995%
11.5	8.586	11.336	11.753	0.4175	3.5525%
15	11.993	14.743	15.282	0.5392	3.5282%
16.4	13.516	16.267	16.751	0.4847	2.8933%
19.7	17.358	20.108	20.337	0.2294	1.1278%
20	17.696	20.446	20.671	0.2249	1.0878%
23.6	21.758	24.508	24.768	0.2600	1.0496%
25	23.268	26.018	26.401	0.3828	1.4499%
30	29.247	31.997	32.367	0.3694	1.1412%
35	35.148	37.898	38.483	0.5851	1.5204%
40	41.147	43.897	44.695	0.7978	1.7849%
45	47.148	49.898	50.968	1.0698	2.0989%
50	53.279	56.029	57.282	1.2529	2.1873%
Mean(x) =	13.871				
StdDev(x) =	5.670				
Std Error =	2.750				

Bootstrap-Based Bounds as *Ersatz* Prediction Bounds



Apply Analogous Adjustment in Other Contexts

- **Prediction-Bound Formulas Do Not Exist in Any Other Cost-Modeling Context**
- **The "$B\text{-}BB = BSB^{*} \pm StdError$" Adjustment Seems Pretty Good in the OLS Case**
- **So Let's Agree to Use that Adjustment for All Additive-Error-CER Scenarios**
  - **It Won't Yield "Exact" Prediction Bounds (we don't know what those are yet!)**
  - **But It'll Have to Do Until the Real Thing Comes Along**



- **Then Extend the Analogy to Multiplicative-Error CERs, i.e.,**
  - **First Apply Percentage Adjustment to Each Estimate**
  - **Then $B\text{-}BB(y) = BSB(y) \pm (\%StdError) \times ESTy$**
  - **This Makes Sense, because the Term "Percentage Standard Error" Refers to a Percentage of the Estimate**

*** "BSB" = "Bootstrap Bounds"**

Common CER Forms

- y = Cost
  x = Technical Parameter (Cost Driver)
- Factor CER: $y = ax$
  Linear CER: $y = a + bx$
  "Nonlinear" CERs: $y = ax^b$
  $y = ab^x$
  $y = a + bx^c$
- a, b, c are Constant Coefficients Derived from Historical Data
- Although We Will Discuss the Case of Only One Cost Driver per CER, the Concepts are the Same for Multiple Cost Drivers, but the Mathematics is More Complicated

General-Error Regression

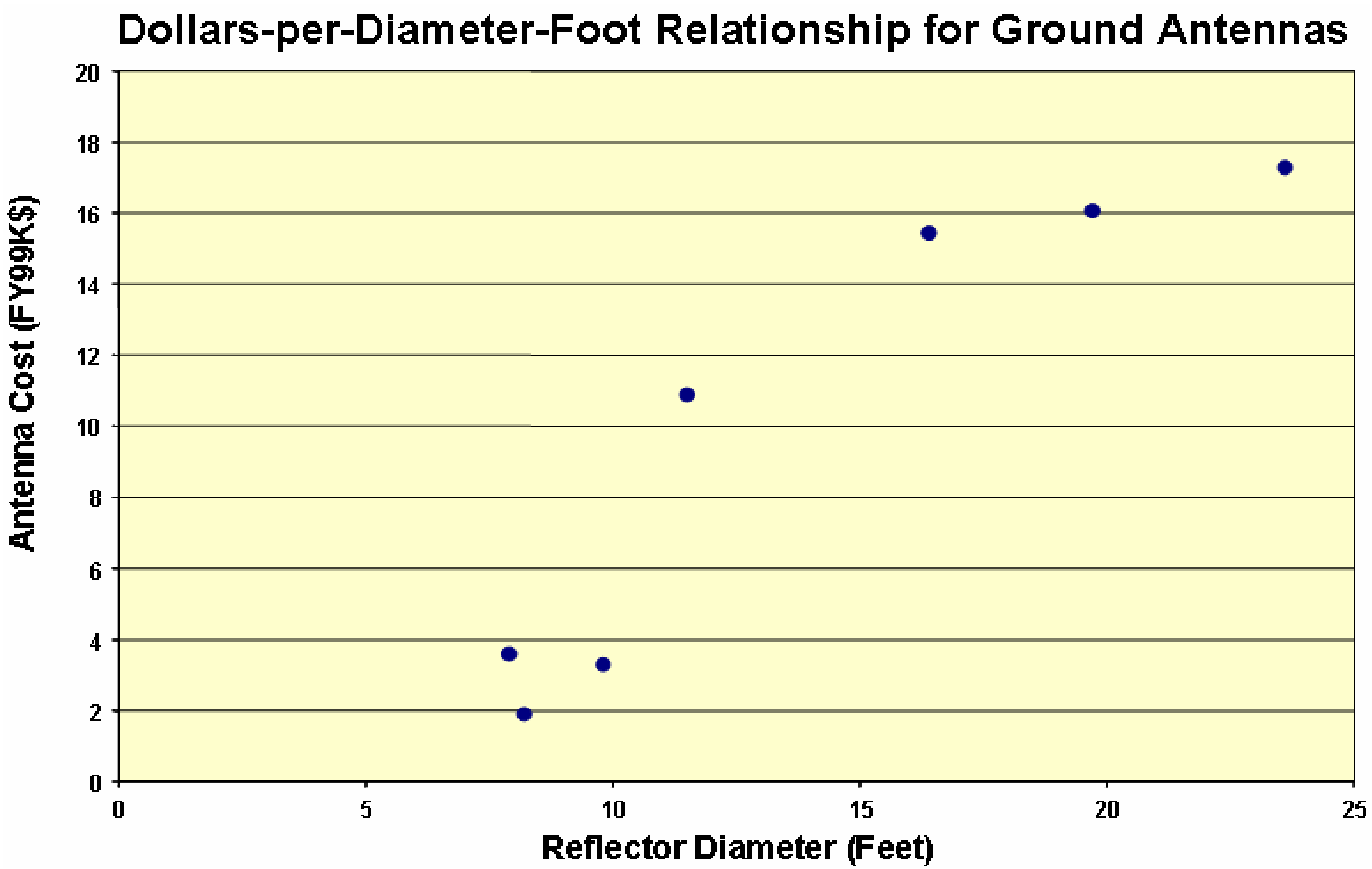
- OLS Offers the Opportunity to Derive Only Two Kinds of CERs
  - Linear (and Other Polynomial) Additive-Error CERs
  - Log-Linear Multiplicative-Error CERs
- In 1974, R.W.M. Wedderburn Established "Iteratively Reweighted Least Squares" (IRLS) as a Technique for Deriving …
  - Unbiased CERs
  - Multiplicative-Error CERs
  - CERs Having Arbitrary Functional Form $y = f(a,b,c)$
- In 1998, the "Minimum Percentage Error - Zero Percentage Bias" (MPE-ZPB) Technique was Introduced to Yield CERs that also Had Minimum Possible Percentage Error among all Unbiased CERs of the Functional Form being Considered

Minimum-Percentage-Error, Zero-Percentage-Bias CERs

- **Experience and Theory Indicates that IRLS CERs Do Not Necessarily Have Minimum Possible Standard Error among all Zero-Bias CERs**
- **IRLS CERs Maximize "Quasi-Likelihood" – They do not Minimize Percentage Error**
- **MPE-ZPB Technique Pursues the Minimum-Percentage-Error Goal Directly**
  - **Computes Minimum-Percentage-Error CER, Subject to Constraint that Percentage Bias be Exactly Zero**
  - **CERs Derived Using "Constrained Optimization" – Another Capability of *Excel Solver***
- **Minimize** $F(a,b,c) = \sum_{k=1}^{n} \left(\frac{y_k - a - bx_k^c}{a + bx_k^c} \right)^2$**, Subject to**

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

CER Example: Antenna Cost vs. Reflector Diameter



Case 1: Multiplicative-Error Factor CER $y = ax \times \varepsilon$ Using MPE-ZPB

- **Minimum-Percentage Error CERs, Subject to the Constraint that Percentage Bias be Zero**
- **Minimize the Sum of Percentage Squared Errors**

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

Subject to the Constraint that the Percentage Bias

$$B(a) = \sum_{k=1}^{n} \left(\frac{ax_k - y_k}{ax_k} \right) = 0$$

Analysis of the Bias Expression

- **We First Rearrange the Equation $B(a) = 0$:**

$$0 = B(a) = \sum_{k=1}^{n} \left(\frac{ax_k - y_k}{ax_k} \right) = \sum_{k=1}^{n} \left(1 - \frac{y_k}{ax_k} \right)$$

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

- **It then Follows that**

$$\sum_{k=1}^{n} \frac{y_k}{x_k} = na$$

- **This is the Mathematical Form of the MPE-ZPB Factor CER's Constraint**

Deriving the "Factor" a in the CER $y = ax \times \varepsilon$

- The Constraint that Percentage Bias be Zero is Expressed Mathematically as

$$\sum_{k=1}^{n} \frac{y_k}{x_k} = na$$

- This Expression Leads Uniquely to the MPE-ZPB Numerical Value of the Coefficient a, which is

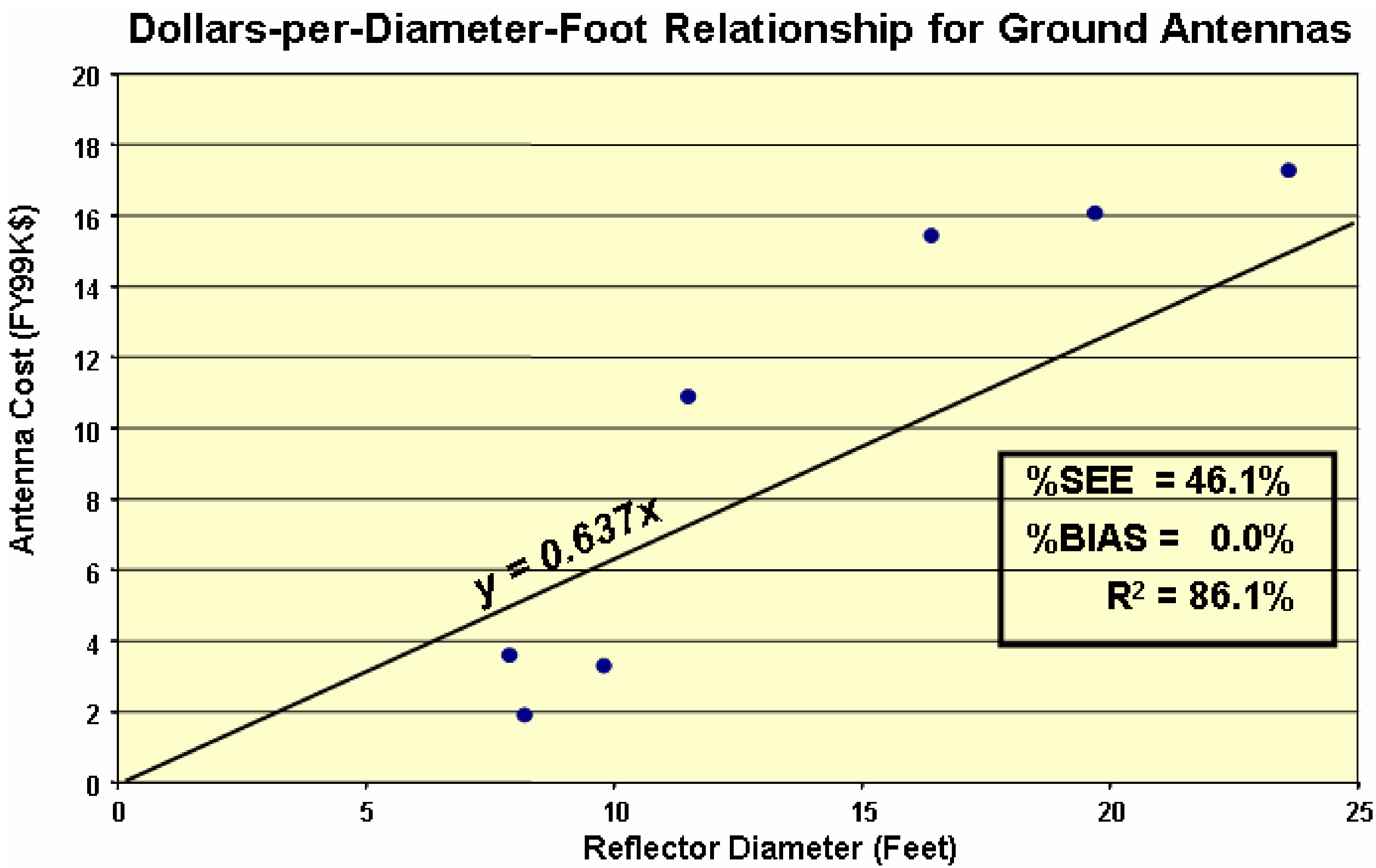
$$a = \frac{1}{n} \sum_{k=1}^{n} \frac{y_k}{x_k}$$

- No Optimization Using *Excel Solver* is Needed in this Simple Case

Dollars-per-Foot Data Base for MPE-ZPB Factor CER

Number of Data Points	Diameter x (feet)	Cost y (FY99$K)	y/x	a	ESTy = ax	%Bias	%Std Error
7	7.9	3.595	0.455	0.637	5.034	28.5826%	8.1696%
	8.2	1.900	0.232	0.637	5.225	63.6360%	40.4954%
	9.8	3.300	0.337	0.637	6.244	47.1530%	22.2341%
	11.5	10.900	0.948	0.637	7.328	-48.7514%	23.7670%
	16.4	15.434	0.941	0.637	10.450	-47.6954%	22.7485%
	19.7	16.074	0.816	0.637	12.553	-28.0531%	7.8698%
	23.6	17.274	0.732	0.637	15.038	-14.8717%	2.2117%
Sums	97.1	68.477	4.460		61.871	0.0000%	127.4960%
EST = Estimated						%Bias =	0.0000%
FY = Fiscal Year						%Std Error =	46.0970%
						R^2 =	86.0636%

MPE-ZPB Factor CER and its Quality Metrics Superimposed on Data Base



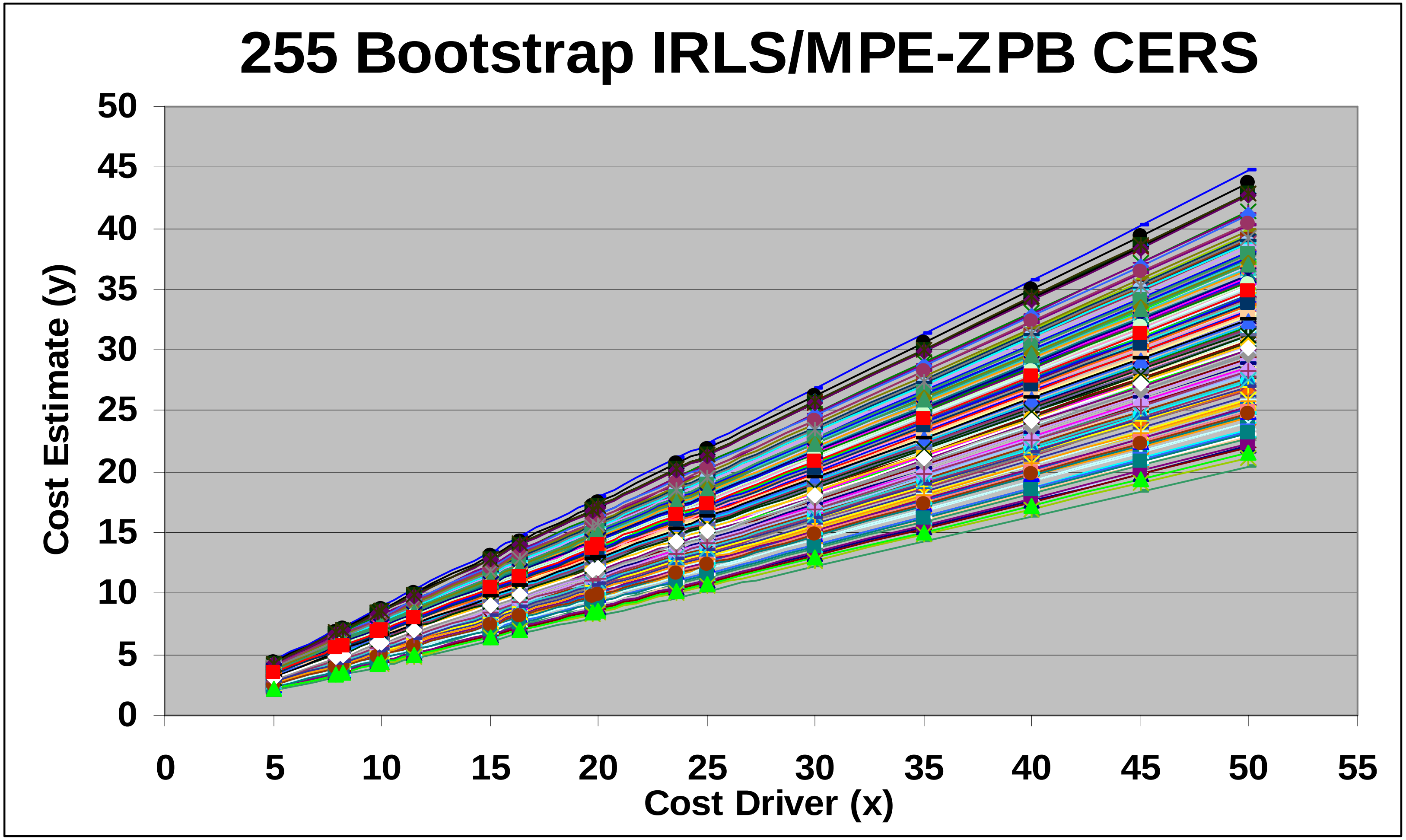
MPE-ZPB Factor CER Percentage Residuals (= Actual÷Estimate)

x Values	Actual y Values	Estimated y Values	Residuals = Actual/Estimated
7.9	3.595	5.034	0.714
8.2	1.900	5.225	0.364
9.8	3.300	6.244	0.528
11.5	10.900	7.328	1.488
16.4	15.434	10.450	1.477
19.7	16.074	12.553	1.281
23.6	17.274	15.038	1.149

Table of Bootstrap y Values, with MPE-ZPB Bootstrap Factor CERs

x Values (Cost Driver)	Estimated y Values	Bootstrap y Value = (Estimated y Value)x(Bootstrap Residual)						
		1	2	3	4	5	6	...
7.9	5.034	1.830	5.782	5.782	2.660	6.446	1.830	...
8.2	5.225	2.761	7.772	1.900	6.691	7.717	7.717	...
9.8	6.244	3.300	7.996	7.173	4.460	9.289	9.289	...
11.5	7.328	8.417	5.233	9.383	10.900	2.665	9.383	...
16.4	10.450	15.544	3.800	7.463	13.381	15.434	5.522	...
19.7	12.553	14.419	8.965	16.074	16.074	18.540	14.419	...
23.6	15.038	22.210	7.947	22.369	10.739	5.468	7.947	...
a (Factor)	0.637	0.608	0.568	0.676	0.663	0.721	0.620	...
% Std Error	46.10%	49.59%	46.86%	36.52%	36.12%	46.82%	49.81%	...
%Bias	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	...
R^2	86.06%	94.92%	5.18%	84.04%	56.53%	12.61%	18.29%	...

Graphs of 255 Bootstrap MPE-ZPB Factor CERs



For Each CER, Estimate Cost at Each of a Range of Cost-Driver Values

Cost Driver (x Values)	Bootstrap Cost Estimates (y Values)						
	#1	#2	#3	#4	#5	#6	...
5	3.041	2.839	3.379	3.316	3.607	3.101	...
7.9	4.805	4.485	5.339	5.239	5.700	4.900	...
8.2	4.988	4.656	5.541	5.438	5.916	5.086	...
9.8	5.961	5.564	6.623	6.500	7.071	6.079	...
10	6.083	5.678	6.758	6.632	7.215	6.203	...
11.5	6.995	6.529	7.771	7.627	8.297	7.133	...
15	9.124	8.516	10.137	9.948	10.822	9.304	...
16.4	9.976	9.311	11.083	10.877	11.833	10.173	...
19.7	11.983	11.185	13.313	13.065	14.213	12.220	...
20	12.166	11.355	13.515	13.264	14.430	12.406	...
23.6	14.356	13.399	15.948	15.652	17.027	14.639	...
25	15.207	14.194	16.894	16.580	18.037	15.507	...
30	18.249	17.033	20.273	19.896	21.645	18.609	...
35	21.290	19.871	23.652	23.213	25.252	21.710	...
40	24.331	22.710	27.031	26.529	28.860	24.811	...
45	27.373	25.549	30.410	29.845	32.467	27.913	...
50	30.414	28.388	33.788	33.161	36.075	31.014	...

Rank All 255 Estimates Associated with a Cost-Driver Value of x

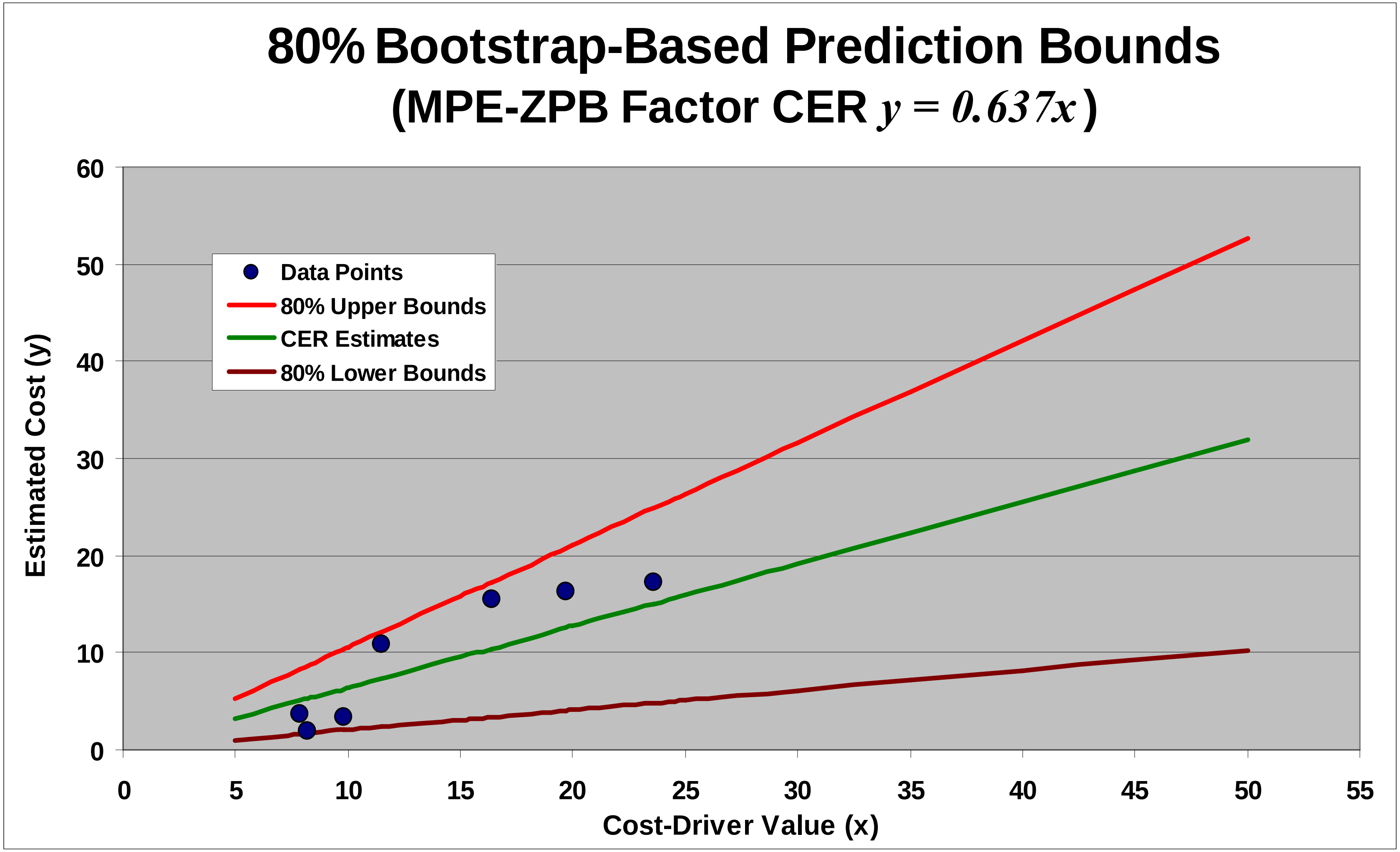
Estimate Ranks	x = 5	x = 15	x = 50
1	2.045	6.135	20.451
2	2.120	6.360	21.200
3	2.153	6.458	21.527
4	2.195	6.585	21.950
5	2.205	6.614	22.045
6	2.209	6.628	22.094
7	2.229	6.687	22.290
8	2.280	6.839	22.795
9	2.289	6.867	22.890
10	2.308	6.923	23.076
11	2.317	6.951	23.171
12	2.322	6.966	23.219
13	2.331	6.994	23.315
14	2.337	7.011	23.369
15	2.337	7.011	23.369
16	2.341	7.023	23.411
17	2.402	7.207	24.023
18	2.407	7.221	24.072
19	2.412	7.236	24.120
20	2.453	7.358	24.528
21	2.460	7.379	24.597
22	2.477	7.430	24.767
23	2.482	7.447	24.822
24	2.486	7.459	24.863
25	2.491	7.473	24.911
26	2.491	7.473	24.911
27	2.496	7.489	24.965
28	2.520	7.559	25.197
29	2.520	7.559	25.197
30	2.535	7.604	25.347

Estimate Ranks	x = 5	x = 15	x = 50
226	3.782	11.346	37.820
227	3.786	11.358	37.860
228	3.787	11.361	37.869
229	3.792	11.375	37.917
230	3.796	11.388	37.961
231	3.801	11.403	38.010
232	3.842	11.526	38.420
233	3.875	11.626	38.754
234	3.880	11.641	38.802
235	3.881	11.642	38.807
236	3.890	11.671	38.904
237	3.900	11.700	39.000
238	3.910	11.729	39.096
239	3.914	11.743	39.144
240	3.941	11.822	39.407
241	3.946	11.837	39.455
242	3.980	11.939	39.798
243	4.030	12.090	40.301
244	4.045	12.134	40.446
245	4.105	12.314	41.046
246	4.124	12.371	41.238
247	4.134	12.402	41.339
248	4.267	12.801	42.669
249	4.283	12.850	42.833
250	4.283	12.850	42.833
251	4.288	12.864	42.879
252	4.288	12.864	42.881
253	4.293	12.878	42.927
254	4.378	13.133	43.775
255	4.472	13.415	44.715

Bootstrap-Based Prediction-Interval 80% Bounds for Factor CER

Cost Driver x Values	Lower 80% BS Bounds	Lower - %SEE*ESTy	MPE-ZPB Factor Estimate	Upper + %SEE*ESTy	Upper 80% BS Bounds
5	2.491	1.023	3.186	5.263	3.794
7.9	3.936	1.616	5.034	8.315	5.994
8.2	4.085	1.677	5.225	8.631	6.222
9.8	4.883	2.004	6.244	10.315	7.436
10	4.982	2.045	6.372	10.525	7.588
11.5	5.730	2.352	7.328	12.104	8.726
15	7.473	3.068	9.558	15.788	11.382
16.4	8.171	3.354	10.450	17.261	12.444
19.7	9.815	4.029	12.553	20.734	14.948
20	9.965	4.090	12.744	21.050	15.176
23.6	11.758	4.826	15.038	24.839	17.907
25	12.456	5.113	15.930	26.313	18.970
30	14.947	6.135	19.116	31.575	22.763
35	17.438	7.158	22.302	36.838	26.557
40	19.929	8.180	25.488	42.100	30.351
45	22.420	9.203	28.673	47.363	34.145
50	24.911	10.225	31.859	52.625	37.939

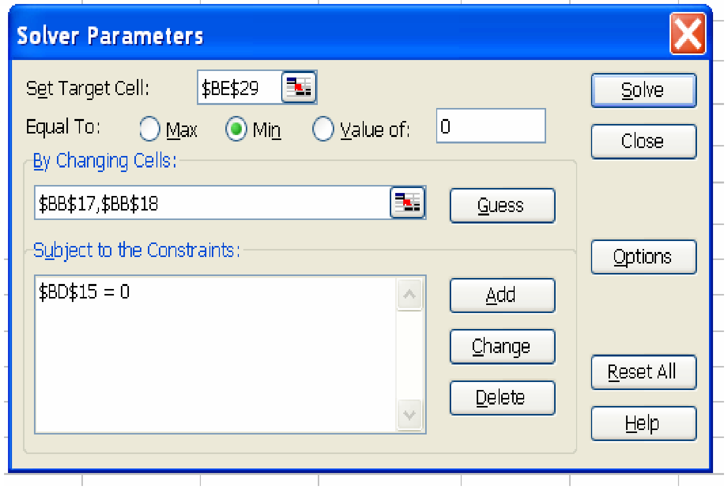
Precision of Estimate Over Range of Possible Cost-Driver Values



Case 2: Multiplicative-Error Linear CER $y = (a+bx) \times \varepsilon$ Using MPE-ZPB

x Values	Actual y Values	Estimated y Values	Percentage Bias	Percentage Squared Error	Actual y^2	Estimated y^2	Actual y * Estimated y
7.9	3.595	2.852	-26.0399%	6.7808%	12.924	8.135	10.254
8.2	1.9	3.206	40.7414%	16.5986%	3.610	10.280	6.092
9.8	3.3	5.094	35.2226%	12.4063%	10.890	25.953	16.811
11.5	10.9	7.100	-53.5112%	28.6345%	118.810	50.417	77.395
16.4	15.434	12.883	-19.8040%	3.9220%	238.208	165.964	198.832
19.7	16.074	16.777	4.1896%	0.1755%	258.373	281.464	269.672
23.6	17.274	21.379	19.2014%	3.6869%	298.391	457.065	369.302
97.1	68.477	69.292	0.0000%	72.2047%	941.207	999.278	948.358

a =	-6.470138		n =	7
b =	1.180052		$n\Sigma x^2-(\Sigma x)^2$ =	1899.3490
%Std Error =	38.0012%		$n\Sigma y^2-(\Sigma y)^2$ =	2193.5568
%Bias =	0.0000%		$n\Sigma xy-(\Sigma x)^2(\Sigma y)^2$ =	1893.5931
R^2 =	86.0636%		$\therefore R$ =	0.927705



MPE-ZPB Linear CER Details for Diameter-vs.-Cost

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

$$a = -6.4701\,; \quad b = 1.1801$$

- **Linear Multiplicative-Error CER**

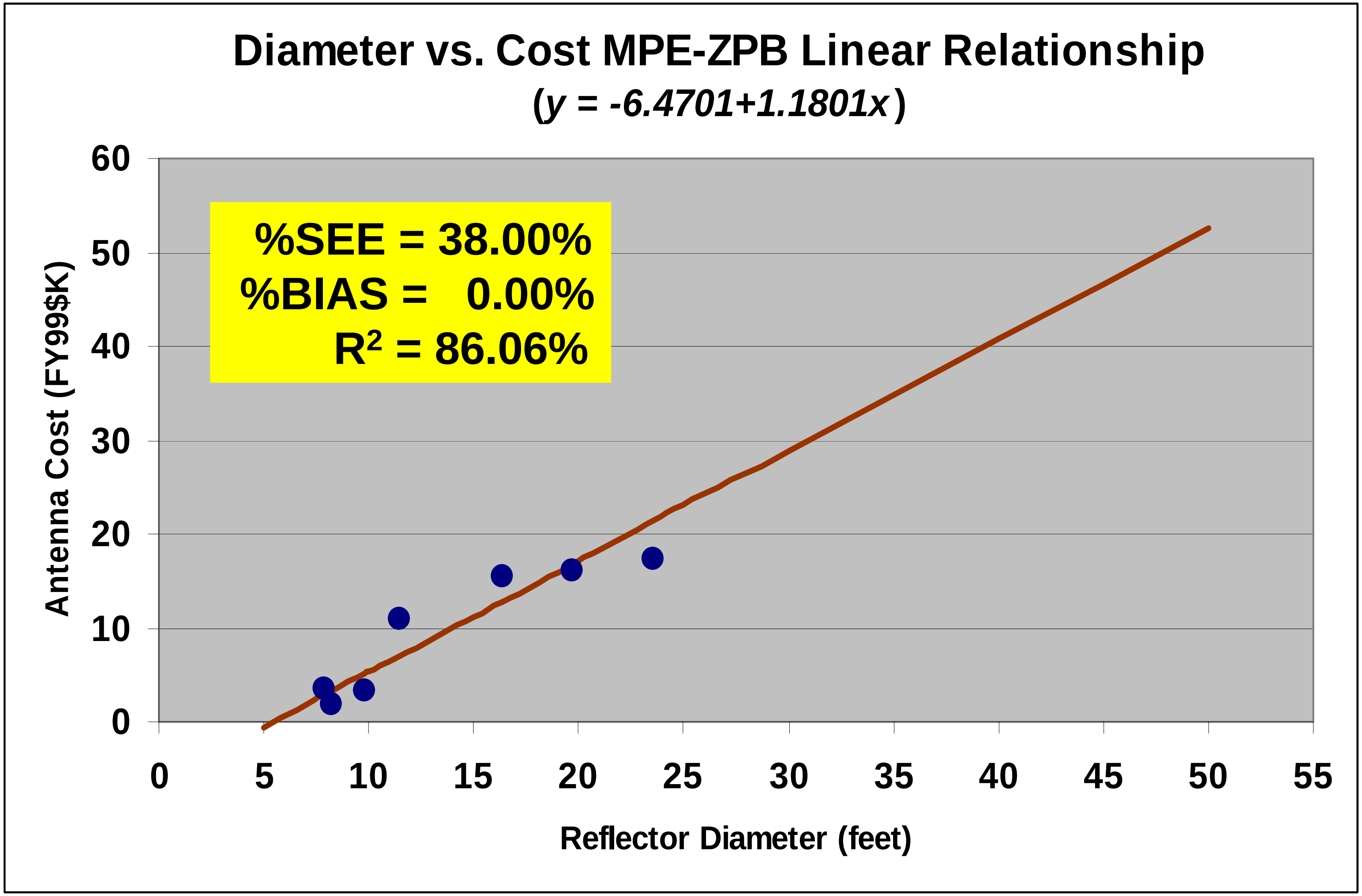
$$y = -6.4701 + 1.1801x$$

- **Percentage Standard Error of the Estimate**

$$Standard\ Error = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\frac{y_i - a - bx_i}{a + bx_i}\right)^2} = \sqrt{\frac{1}{7-2}(0.7220)} = 0.3800$$

(Average 38.00% Across the Data Range)

MPE-ZPB Linear CER and Its Quality Metrics Superimposed on Data Base

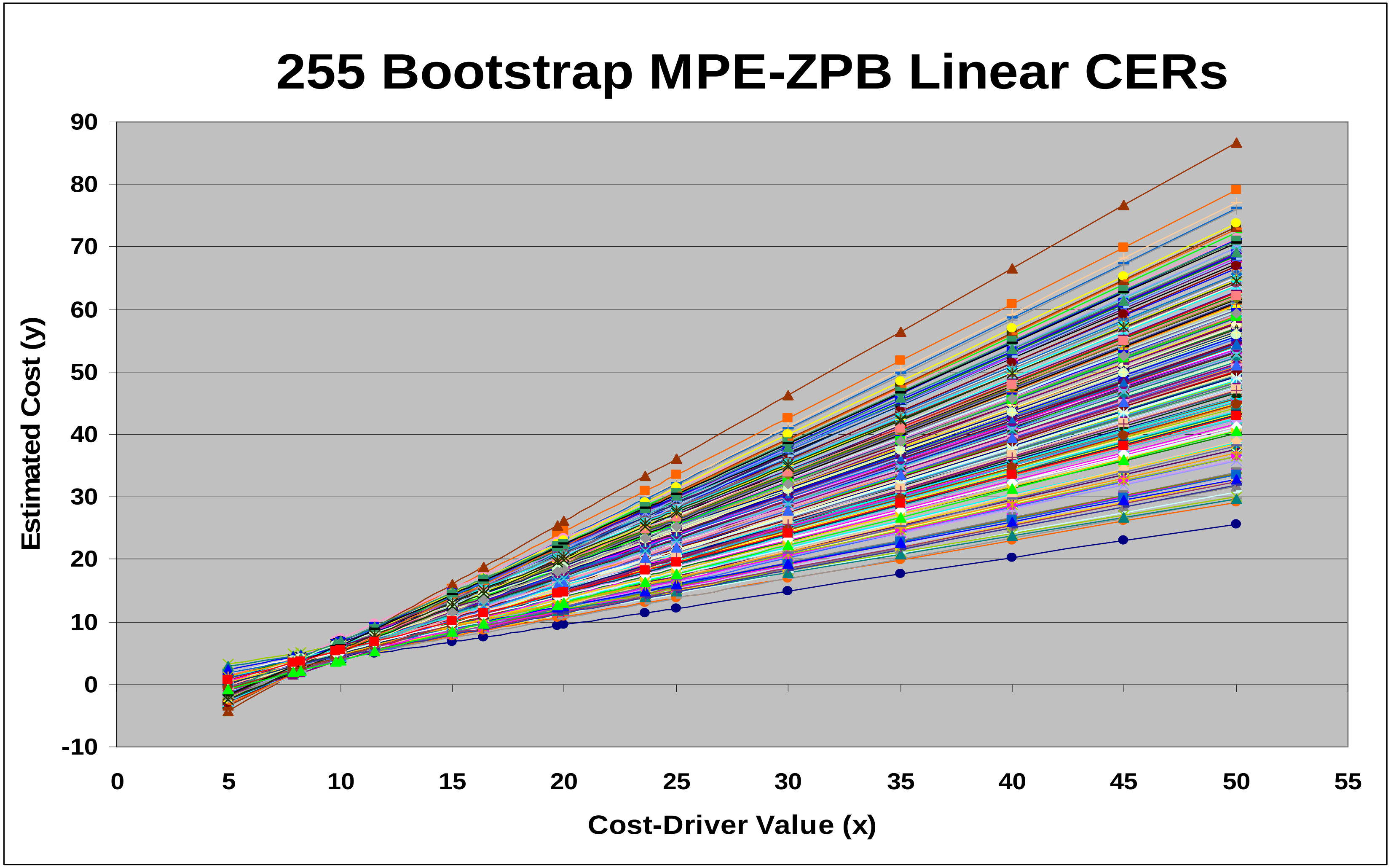


Linear CER Percentage Residuals (= Actual÷Estimate)

x Values (Cost Drivers)	y Values (Actual Costs)	Estimated y Values (Estimated Costs)	Residuals = Actual/Estimated
7.9	3.595	2.852	1.260
8.2	1.900	3.206	0.593
9.8	3.300	5.094	0.648
11.5	10.900	7.100	1.535
16.4	15.434	12.883	1.198
19.7	16.074	16.777	0.958
23.6	17.274	21.379	0.808
a = -6.4701		b = 1.1801	
% SEE = 38.0012%		%Bias = 0.0000%	

Note: Estimated y Values are Calculated Using the MPE-ZPB Linear Regression Equation $y = -6.4701+1.1801x$.

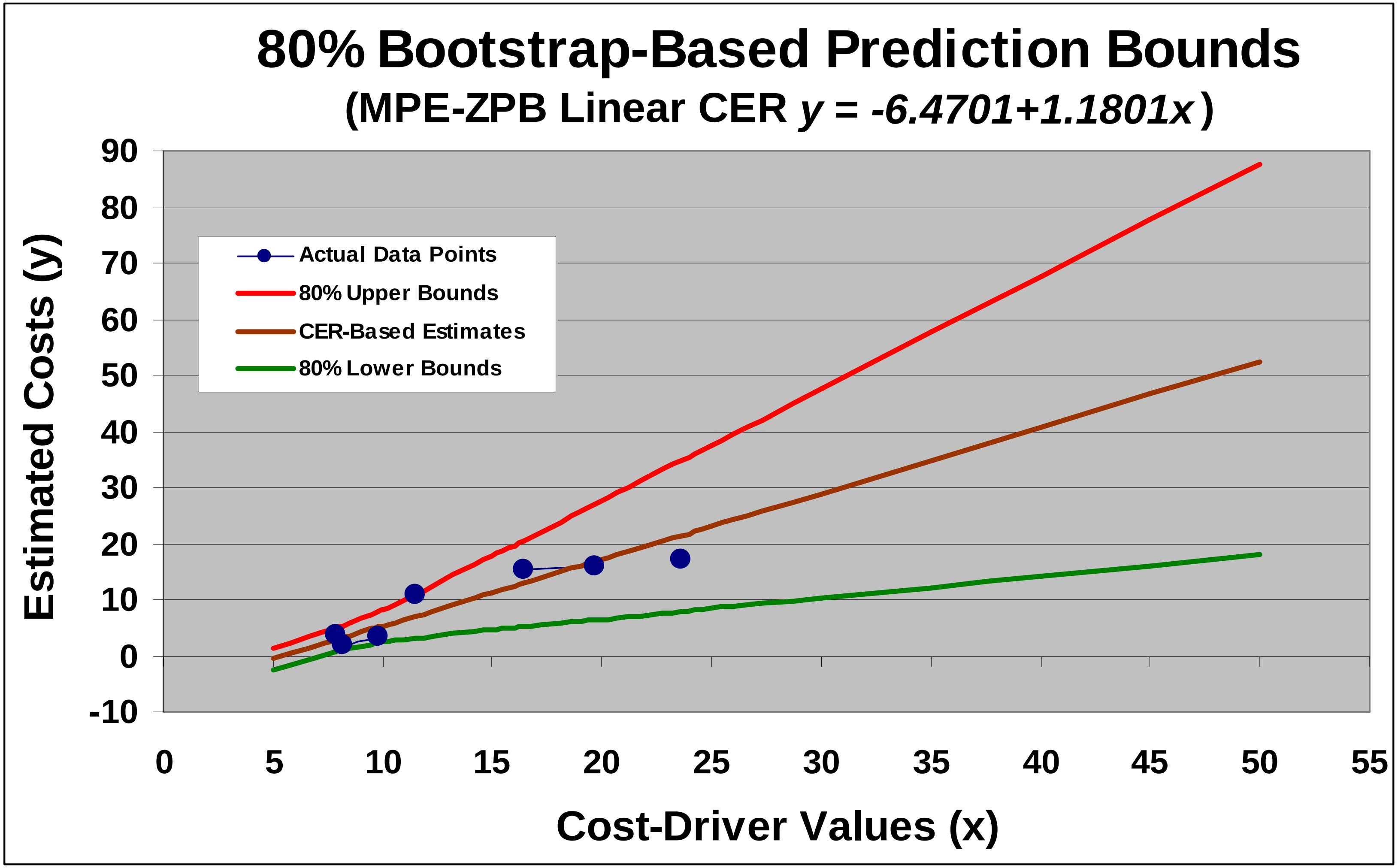
Graphs of 255 Bootstrap MPE-ZPB Linear CERs



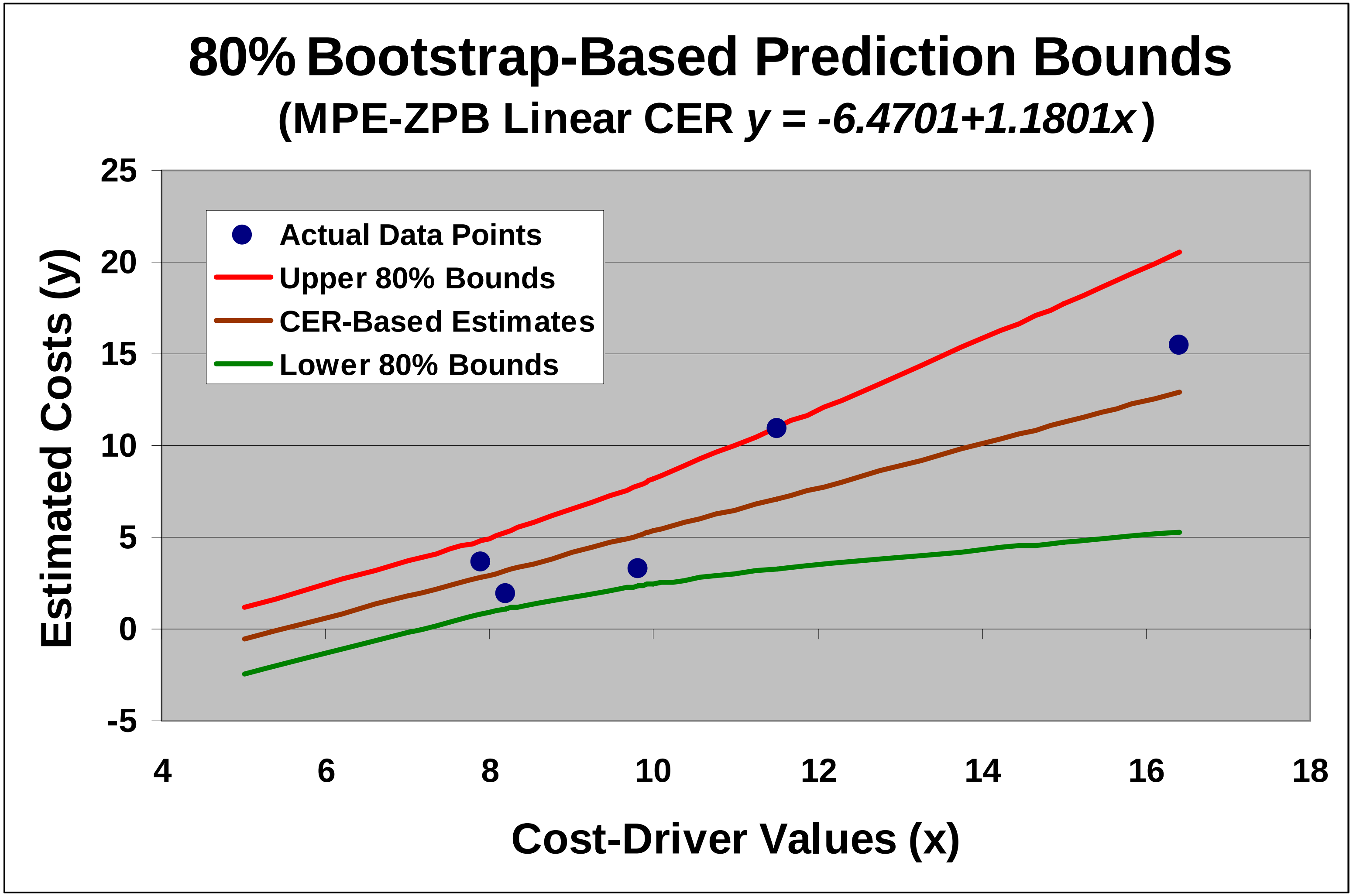
Bootstrap-Derived Prediction-Interval 80% Bounds for MPE-ZPB Linear CER

Cost Driver (x Value)	Lower 80% Bootstrap Bound	Lower 80% Bound − %SEE*ESTy	MPE-ZPB Linear Estimate	Upper 80% Bound + %SEE*ESTy	Upper 80% Bootstrap Bound
5	-2.222	-2.439	-0.570	1.215	0.998
7.9	1.934	0.850	2.852	4.797	3.713
8.2	2.304	1.086	3.206	5.261	4.042
9.8	4.282	2.346	5.094	7.833	5.897
10	4.469	2.443	5.330	8.188	6.163
11.5	5.935	3.237	7.100	11.008	8.309
15	9.007	4.739	11.231	17.767	13.500
16.4	10.158	5.263	12.883	20.541	15.645
19.7	12.782	6.406	16.777	26.913	20.538
20	13.037	6.527	17.131	27.519	21.009
23.6	16.030	7.905	21.379	34.787	26.663
25	17.181	8.429	23.031	37.602	28.850
30	21.398	10.404	28.931	47.588	36.594
35	25.455	12.219	34.832	57.708	44.471
40	29.556	14.078	40.732	67.734	52.255
45	33.722	16.001	46.632	77.635	59.914
50	37.887	17.925	52.532	87.512	67.549
%SEE =	38.001227%				

80% Prediction Bounds Over Range of Possible Cost-Driver Values



80% Prediction Bounds for "Small" Values of the Cost Driver x



Case 3: Multiplicative-Error Power CER $y = ax^b \times \varepsilon$ Using MPE-ZPB

- **Based on MPE-ZPB Computations on the Historical Data Using *Excel Solver* …**

$$a = 0.1037; \quad b = 1.6934$$

- **Multiplicative-Error CER**

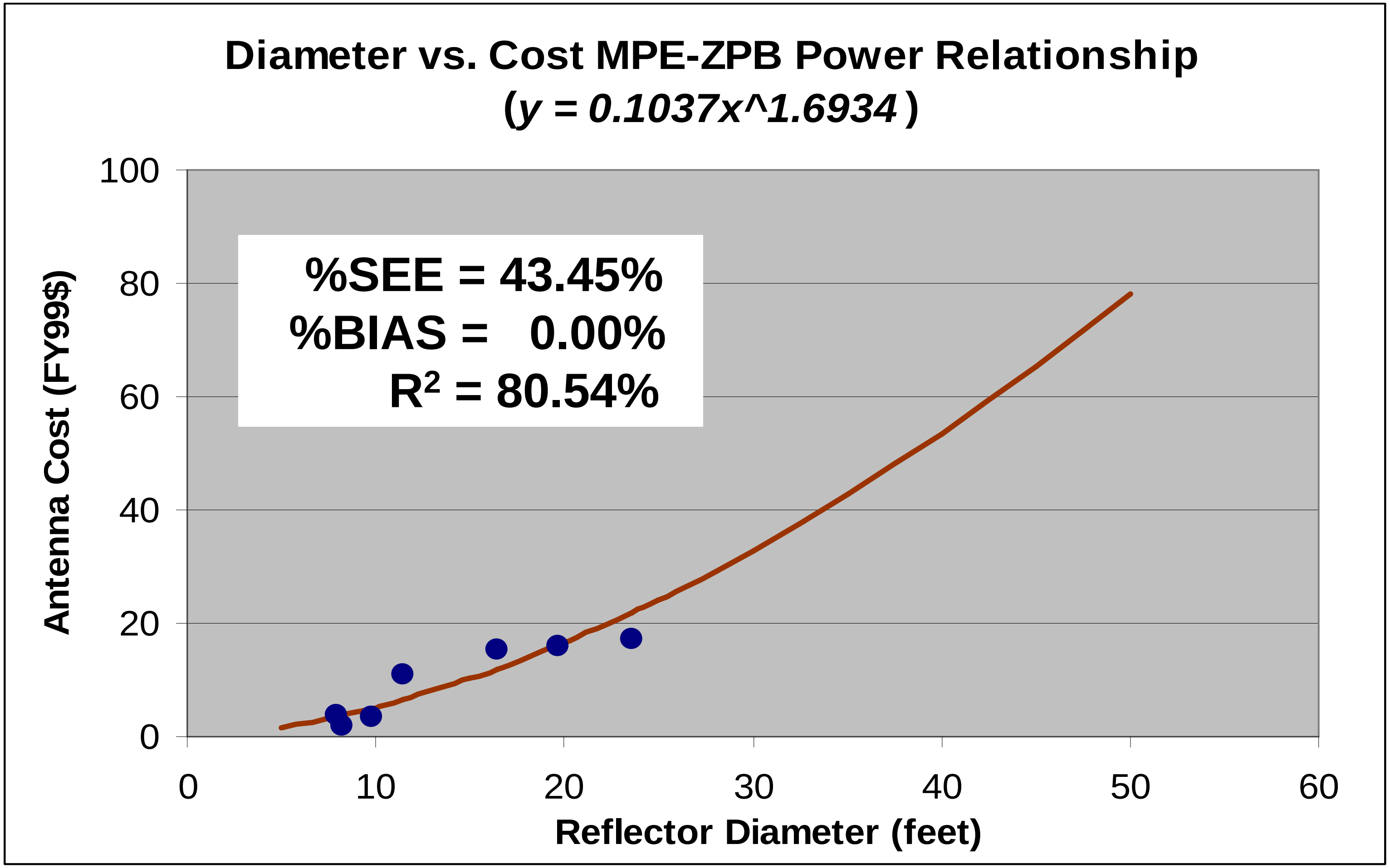
$$y = 0.1037x^{1.6934}$$

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

$$Standard\ Error = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\left(\frac{y_i - a - bx_i}{a + bx_i}\right)^2} = \sqrt{\frac{1}{7-2}(0.9439)} = 0.4345$$

(Average 43.45% Across the Data Range)

MPE-ZPB Power CER and Its Quality Metrics Superimposed on Data Base



Power CER Percentage Residuals (= Actual÷Estimate)

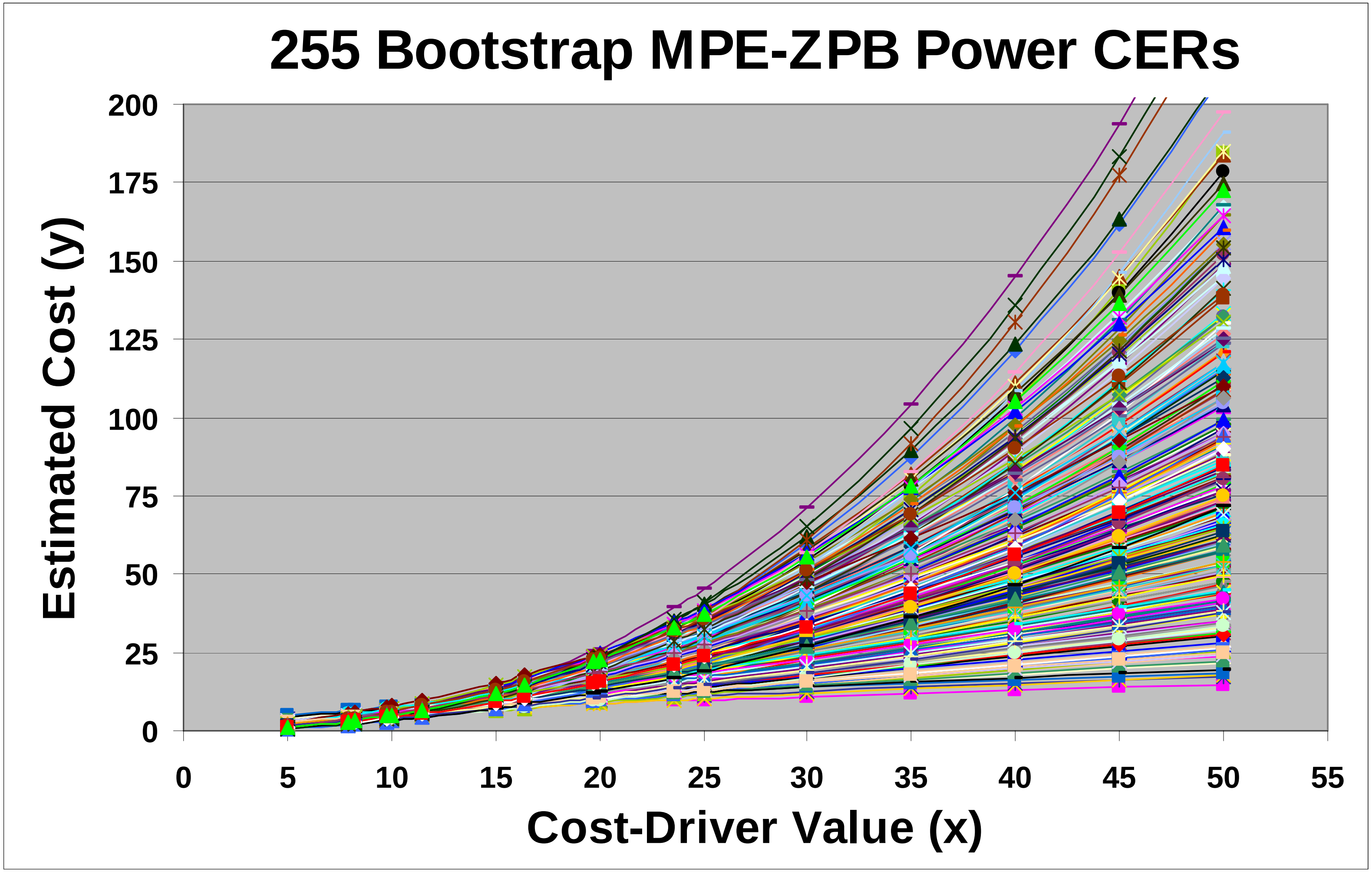
x Values (Cost Drivers)	y Values (Actual Costs)	Estimated y Values (Cost Estimates)	Residuals = Actual/Estimate
7.9	3.595	3.436	1.046
8.2	1.900	3.659	0.519
9.8	3.300	4.949	0.667
11.5	10.900	6.489	1.680
16.4	15.434	11.835	1.304
19.7	16.074	16.144	0.996
23.6	17.274	21.921	0.788
a =	0.103734	b =	1.693439
% SEE =	43.450063%	%Bias =	0.000000%

Note: Estimated y Values are Calculated Using the MPE-ZPB Power Regression Equation $y = 0.1037x^{1.6934}$.

Use Each Bootstrap Sample to Derive a Power CER Using MPE-ZPG

x Values (Cost Driver)	Estimated y Values	Bootstrap "Actuals" = Estimates x Residuals						
		#1	#2	#3	#4	#5	#6	...
7.9	3.436	3.595	3.421	3.595	1.784	5.771	3.595	...
8.2	3.659	3.643	3.643	1.900	2.440	1.900	3.643	...
9.8	4.949	5.178	8.313	4.927	6.454	6.454	3.300	...
11.5	6.489	5.113	10.900	3.369	8.461	5.113	10.900	...
16.4	11.835	12.385	11.784	7.892	6.145	12.385	15.434	...
19.7	16.144	21.053	10.765	16.894	27.121	16.074	21.053	...
23.6	21.921	11.382	11.382	21.826	11.382	11.382	11.382	...
	a (Constant)	0.096	1.061	0.067	0.031	0.422	0.095	...
	b (Exponent)	1.710	0.797	1.788	2.135	1.129	1.756	...
	% Std Error	28.056%	35.287%	32.403%	54.319%	38.551%	40.599%	...
	% Bias	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	...
	R^2	58.393%	61.167%	94.945%	39.494%	68.860%	49.636%	

Graphs of 255 Bootstrap MPE-ZPB Power CERs

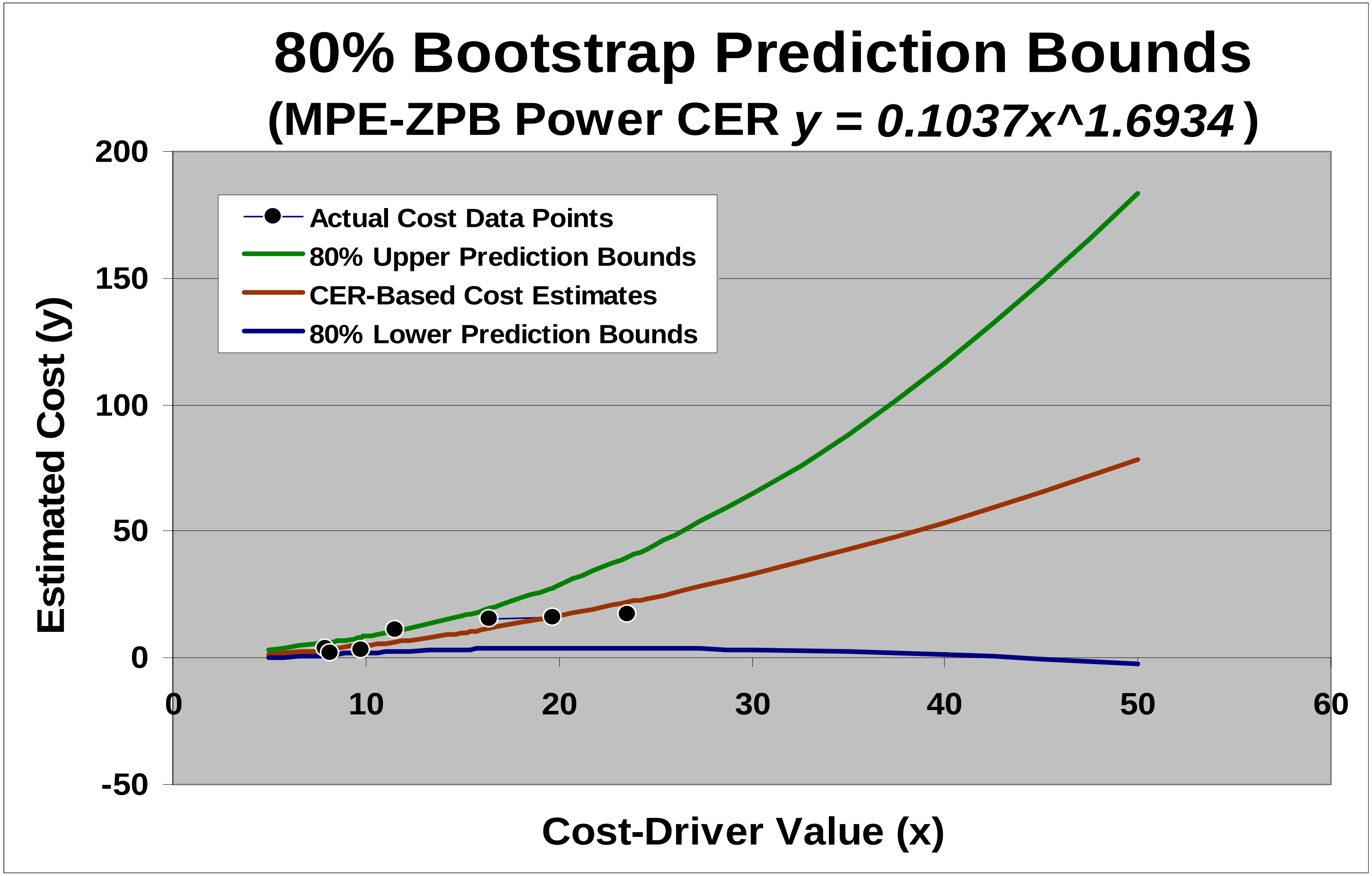


Bootstrap-Derived 80% Prediction Bounds for MPE-ZPB Power CER

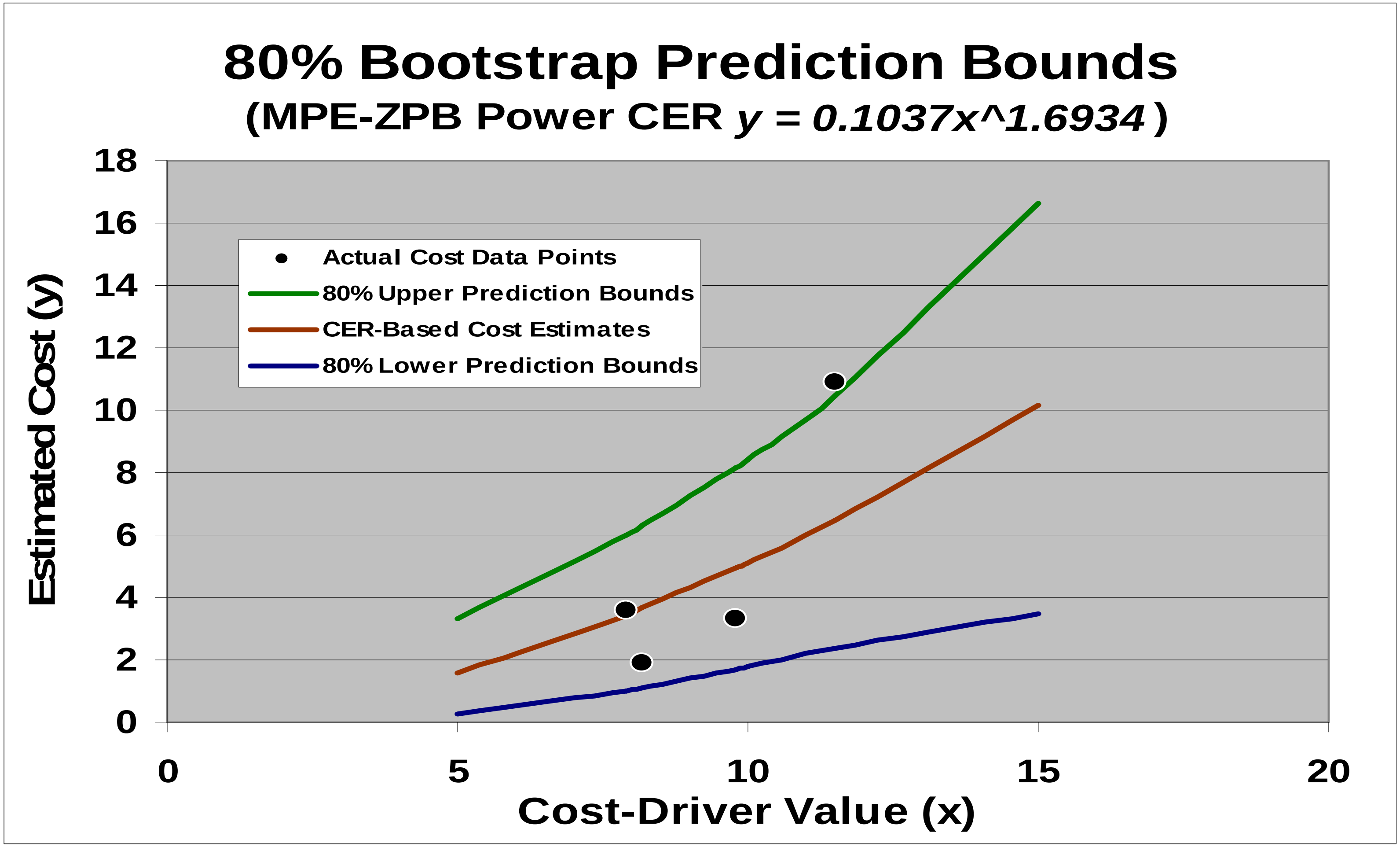
Cost Driver (x Values)	Lower 80% Bootstrap Bound	Lower 80% Bootstrap Prediction Bound	MPE-ZPB Power Estimate of y	Upper 80% Bootstrap Prediction Bound	Upper 80% Bootstrap Bound
5	0.945	0.257	1.583	3.304	2.616
7.9	2.468	0.975	3.436	5.982	4.489
8.2	2.682	1.092	3.659	6.325	4.735
9.8	3.847	1.697	4.949	8.155	6.005
10	4.006	1.781	5.121	8.397	6.172
11.5	5.171	2.352	6.489	10.493	7.674
15	7.895	3.473	10.176	16.616	12.194
16.4	8.766	3.624	11.835	19.577	14.435
19.7	10.994	3.980	16.144	27.752	20.737
20	11.161	3.965	16.563	28.579	21.383
23.6	13.338	3.813	21.921	39.809	30.284
25	14.340	3.839	24.168	44.584	34.083
30	17.521	3.221	32.911	64.723	50.423
35	20.865	2.300	42.727	88.358	69.793
40	24.478	1.202	53.569	116.486	93.210
45	28.057	-0.357	65.394	148.617	120.203
50	31.700	-2.263	78.167	183.698	149.735

Note: Bootstrap Prediction Bound = Bootstrap Bound ± %SSE*Est$y$

Precision of Estimate Over Range of Possible Cost-Driver Values



80% Prediction Bounds for "Small" Values of the Cost Driver x



Case 4: Multiplicative-Error Triad CER $y = (a+bx^c) \times \varepsilon$ Using MPE-ZPB

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

$$a = -236.11; \quad b = 212.42; \quad c = 0.06$$

- **Multiplicative-Error CER**

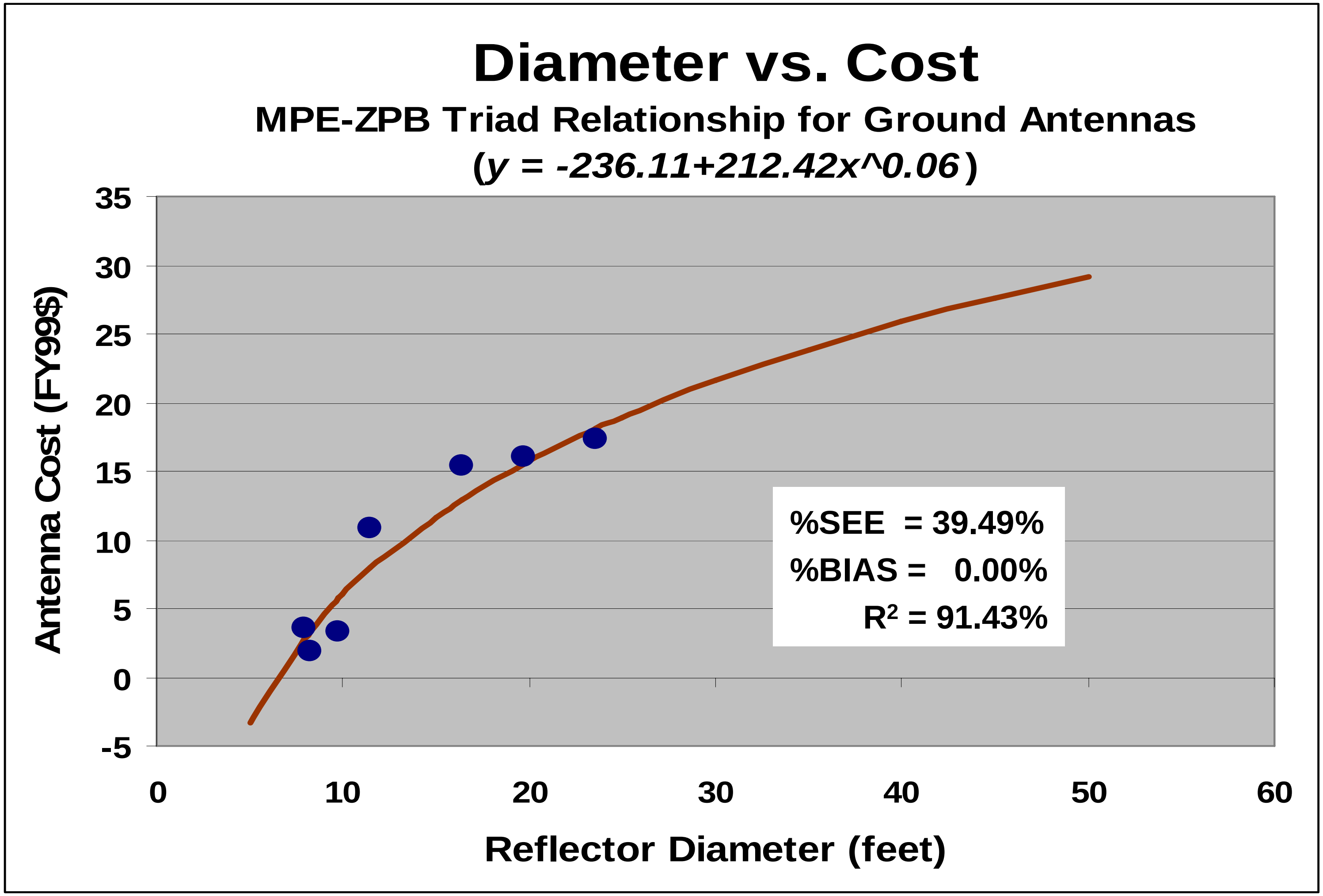
$$y = -236.11 + 212.42x^{0.06}$$

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

$$Standard\ Error = \sqrt{\frac{1}{n-3}\sum_{i=1}^{n}\left(\frac{y_i - a - bx_i^c}{a + bx_i^c}\right)^2} = \sqrt{\frac{1}{7-3}(0.6236)} = 0.3949$$

(Average 39.49% Across Data Range)

MPE-ZPB Triad CER, Quality Metrics Superimposed on Data Base



MPE-ZPB Triad CER Percentage Residuals (= Actual÷Estimate)

x Values	Actual y Values	Estimated y Values	Residuals = Actual/Estimated
7.9	3.595	2.786965	1.2899336721
8.2	1.900	3.293516	0.5768911567
9.8	3.300	5.730955	0.5758203023
11.5	10.900	7.939498	1.3728828226
16.4	15.434	12.912169	1.1953065162
19.7	16.074	15.520302	1.0356757065
23.6	17.274	18.116607	0.9534898236

Bootstrap Sampling of MPE-ZPB Residuals

x Values	Residual Samples #1	#2	#3	#4	#5	#6	...
7.9	1.195	1.373	1.290	0.953	1.195	0.953	...
8.2	0.577	1.290	1.290	1.290	1.036	0.576	...
9.8	1.195	1.373	1.373	0.577	0.953	0.576	...
11.5	0.577	1.036	1.036	0.953	1.373	1.373	...
16.4	1.290	1.036	0.953	0.577	1.036	0.953	...
19.7	1.036	0.577	1.195	1.195	1.290	0.953	...
23.6	0.577	1.036	0.577	1.290	0.576	0.953	...

Use Each Bootstrap Sample to Calculate an MPE-ZPB Triad CER

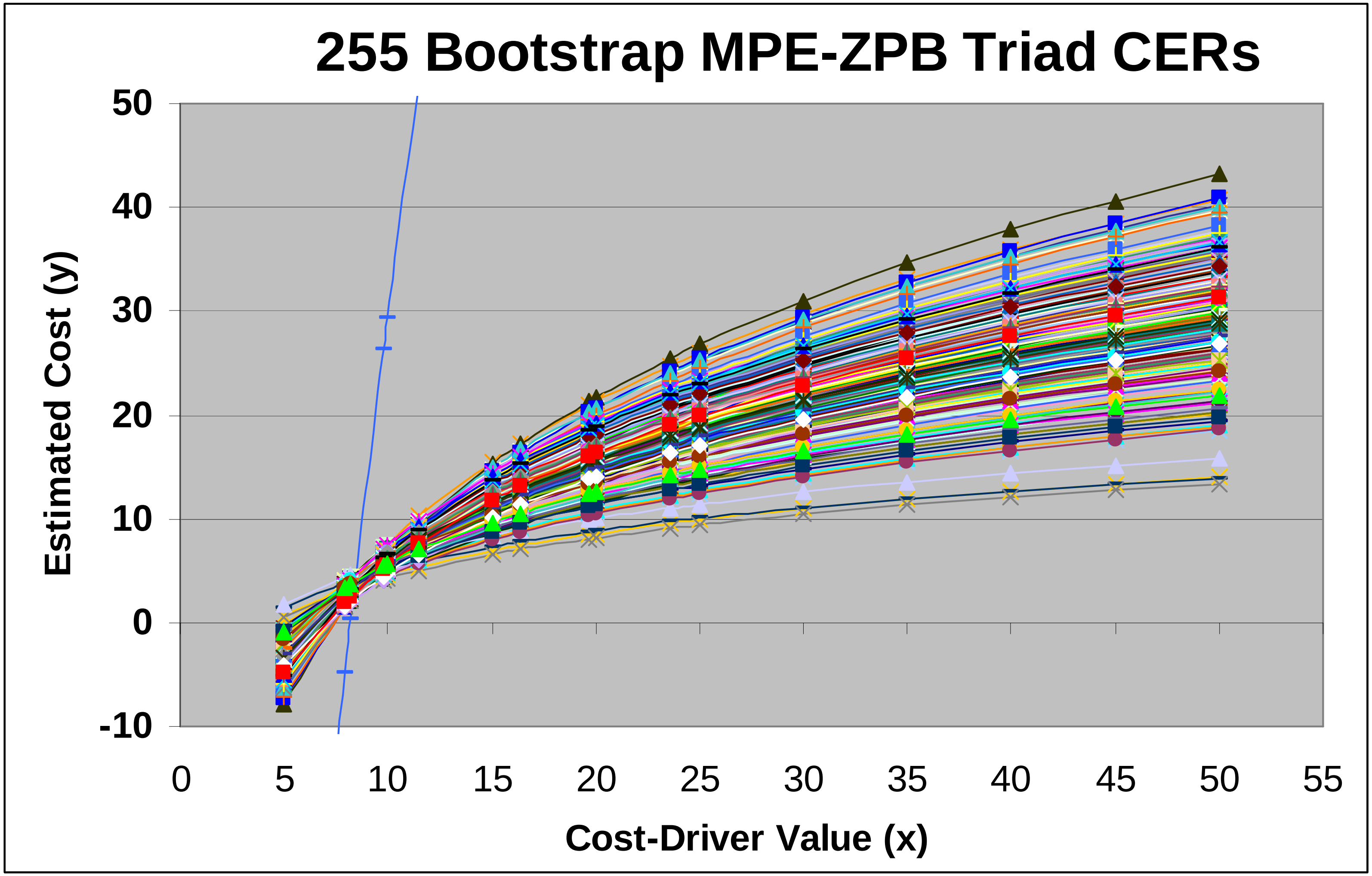
x Values	Bootstrap y Values #1	#2	#3	#4	#5	#6	...
7.9	3.331	3.826	3.595	2.657	3.331	2.657	...
8.2	1.900	4.248	4.248	4.248	3.411	1.896	...
9.8	6.850	7.868	7.868	3.306	5.464	3.300	...
11.5	4.580	8.223	8.223	7.570	10.900	10.900	...
16.4	16.656	13.373	12.312	7.449	13.373	12.312	...
19.7	16.074	8.954	18.552	18.552	20.020	14.798	...
23.6	10.451	18.763	10.451	23.369	10.432	17.274	...
a (Intercept)	-234.930	-233.629	-234.175	-235.904	-236.560	-236.531	...
b (Coefficient)	213.612	215.212	214.377	212.627	211.969	211.998	...
c (Exponent)	0.051	0.048	0.050	0.056	0.059	0.057	...
% Std Error	43.87%	23.68%	26.03%	38.48%	30.23%	35.10%	...
% Bias	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	...
R^2	91.45%	91.46%	91.45%	91.43%	91.42%	91.42%	...

Note: Bootstrap y Value = Estimated y Value × Sample Residual

For Each CER, Estimate Cost at Each of a Range of Cost-Driver Values

Cost Driver (x Values)	Bootstrap Cost Estimates (y Values)						
	1	2	3	4	5	6	...
5	-2.884	-1.207	-1.664	-3.302	-3.379	-4.011	...
7.9	2.640	3.931	3.765	2.710	3.028	2.177	...
8.2	3.096	4.355	4.213	3.207	3.558	2.688	...
9.8	5.288	6.391	6.367	5.597	6.108	5.149	...
10	5.538	6.623	6.612	5.869	6.398	5.429	...
11.5	7.272	8.233	8.316	7.762	8.419	7.379	...
15	10.605	11.324	11.589	11.400	12.307	11.128	...
16.4	11.735	12.371	12.698	12.635	13.626	12.400	...
19.7	14.071	14.537	14.992	15.190	16.360	15.034	...
20	14.265	14.716	15.182	15.402	16.586	15.252	...
23.6	16.396	16.689	17.273	17.733	19.081	17.657	...
25	17.142	17.379	18.005	18.550	19.956	18.499	...
30	19.516	19.576	20.336	21.151	22.742	21.183	...
35	21.542	21.449	22.323	23.371	25.122	23.474	...
40	23.309	23.082	24.057	25.310	27.201	25.474	...
45	24.878	24.531	25.596	27.032	29.048	27.252	...
50	26.290	25.835	26.981	28.582	30.712	28.853	...

Graphs of 255 Bootstrap MPE-ZPB Triad CERs



Rank All 255 Estimates Associated with a Cost-Driver Value of x

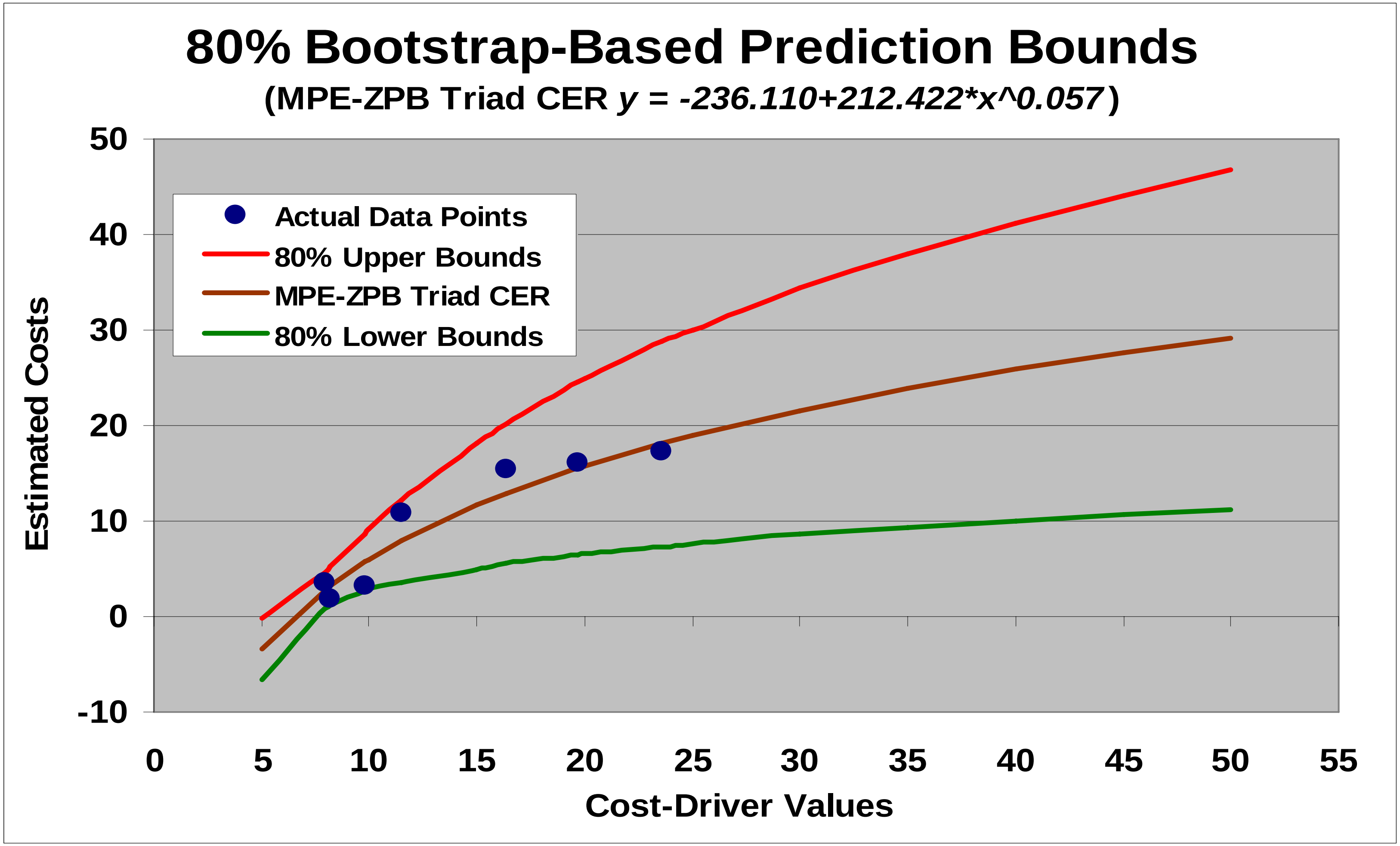
MPE-ZPB Estimates	5	15	50
1	-61.076	6.522	13.361
2	-7.833	6.886	13.863
3	-7.418	7.289	13.979
4	-7.118	7.924	15.776
5	-6.582	8.185	18.484
6	-6.464	8.189	18.603
7	-6.433	8.308	18.854
8	-6.229	8.310	18.979
9	-6.228	8.506	19.364
10	-6.151	8.725	19.781
11	-5.992	8.804	20.118
12	-5.942	8.857	20.181
13	-5.870	8.930	20.221
14	-5.849	9.010	20.244
15	-5.763	9.103	20.594
16	-5.649	9.106	21.085
17	-5.635	9.127	21.173
18	-5.624	9.184	21.308
19	-5.618	9.195	21.315
20	-5.597	9.297	21.343
21	-5.538	9.401	21.366
22	-5.480	9.417	21.837
23	-5.328	9.482	22.159
24	-5.298	9.482	22.303
25	-5.220	9.484	22.376
26	-5.220	9.523	23.008
27	-5.209	9.672	23.252
28	-5.192	9.767	23.298
29	-5.179	9.815	23.538
30	-5.119	9.838	23.878

MPE-ZPB Estimates	5	15	50
226	-1.625	13.477	35.033
227	-1.603	13.505	35.138
228	-1.574	13.511	35.163
229	-1.560	13.590	35.227
230	-1.472	13.632	35.302
231	-1.423	13.633	35.398
232	-1.403	13.700	35.641
233	-1.262	13.702	35.832
234	-1.207	13.704	36.019
235	-1.192	13.704	36.028
236	-1.179	13.769	36.038
237	-1.167	13.787	36.040
238	-1.153	13.838	36.099
239	-1.029	13.861	36.476
240	-1.002	13.880	36.597
241	-0.938	13.889	36.611
242	-0.919	13.946	36.882
243	-0.833	13.960	36.925
244	-0.833	14.039	37.133
245	-0.802	14.069	37.534
246	-0.799	14.088	37.778
247	-0.777	14.366	38.229
248	-0.608	14.482	39.490
249	-0.535	14.531	39.708
250	-0.443	14.740	40.103
251	-0.169	14.802	40.142
252	0.450	14.839	40.805
253	0.595	15.270	40.834
254	1.446	15.491	43.202
255	1.694	98.704	409.270

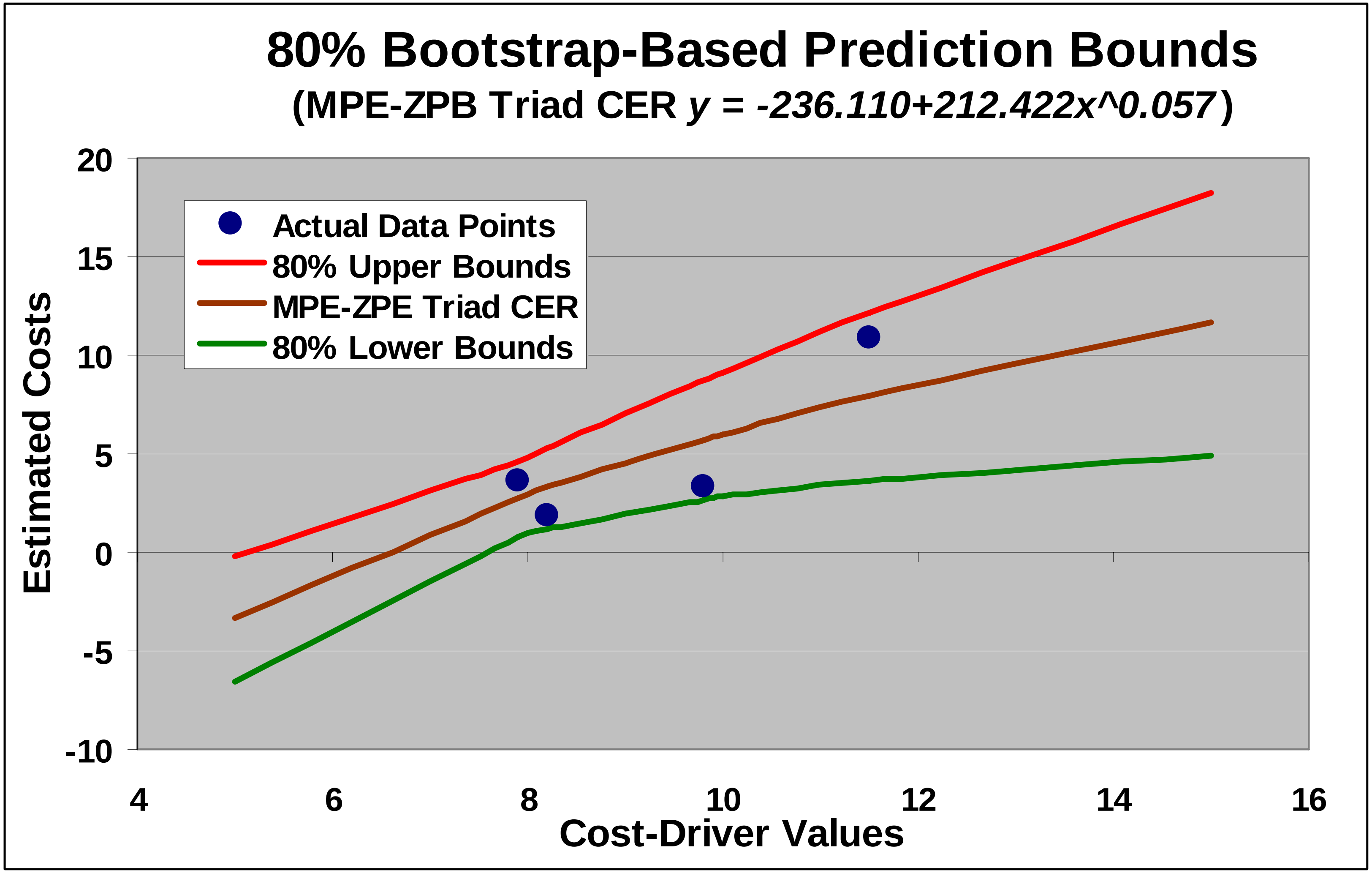
Bootstrap-Derived Prediction-Interval 80% Bounds for MPE-ZPB Triad CER

Cost Driver (x Values)	Lower 80% BS Bounds	Lower - SEE%	MPE-ZPB Linear Estimates	Upper + SEE%	Upper 80% BS Bounds
5	-5.220	-6.540	-3.343	-0.196	-1.516
7.9	1.932	0.832	2.787	4.604	3.503
8.2	2.501	1.201	3.294	5.320	4.020
9.8	4.958	2.695	5.731	8.702	6.439
10	5.213	2.841	6.009	9.131	6.759
11.5	6.763	3.628	7.939	12.151	9.016
15	9.504	4.903	11.653	18.212	13.611
16.4	10.628	5.529	12.912	20.245	15.147
19.7	12.586	6.458	15.520	24.538	18.410
20	12.741	6.527	15.737	24.894	18.680
23.6	14.491	7.338	18.117	28.736	21.583
25	15.117	7.635	18.951	30.070	22.587
30	17.115	8.583	21.607	34.353	25.821
35	18.800	9.373	23.875	38.014	28.587
40	20.250	10.041	25.855	41.240	31.031
45	21.536	10.632	27.614	44.143	33.239
50	22.692	11.163	29.198	46.794	35.265

Precision of Estimate Over Range of Possible Cost-Driver Values



80% Prediction Bounds for "Small" Values of the Cost Driver *x*



Summary

- **Explicit Formulas Exist for Prediction Intervals Corresponding to OLS-Derived Linear CERs**
  - **These Intervals Can be Reproduced Fairly Well by Adjusting Bounds Derived by Bootstrap Sampling**
  - **For General-Error CERs, Algebraic Expressions for Prediction Bounds Do Not Appear to be Available**
- **Although Theoretical Research in this Direction May be Worthwhile and May Lead to Future Algebraic Solutions, There is a Need to Estimate Prediction Bounds Right Now**
- **While We Await the "Exact" Solution, the Bootstrap Sampling Technique Appears to Offer an Opportunity to Compute Prediction Bounds for Any Specific CER**
  - **By Analogy with the OLS Adjusted Bootstrap Bounds**
  - **Provides Solution on a "One-Time-Only" Basis for Each CER, Rather than via an Algebraic Formula of Wide Applicability**

References

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.

Casella, G., "Introduction to the Silver Anniversary of the Bootstrap," *Statistical Science*, Vol. 18, Number 2 (2003), pages 133-134.

Deflandre, David and Kleijnen, Jack P.C., "Statistical Analysis of Random Simulations: Bootstrap Tutorial," Center for Economic Research, Tilburg University, The Netherlands, Discussion Paper No. 2002-58, June 2002, 23 pages, available at http://greywww.kub.nl:2080/greyfiles/center/2002/doc/58.pdf.

Efron, B., "Second Thoughts on the Bootstrap," *Statistical Science*, Vol. 18, Number 2 (2003), pages 135-140.

Goldberg, Matthew S. and Tuow, Anduin, "Statistical Considerations in Estimating Learning Curves and Multiplicative CERs," 32nd Annual DoD Cost Analysis Symposium, Williamsburg VA, 2-5 February 1999, 44 charts.

Jørgensen, Bent, *The Theory of Dispersion Models*, London: Chapman & Hall, 1997, xii+237 pages.

Léger, Christian; Politis, Dimitris N.; and Romano, Joseph P., "Bootstrap Technology and Applications," *Technometrics*, Vol. 34 (November 1992), pages 378-398.

Wedderburn, R.W.M., "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, Vol. 61, Number 3 (1974), pages 439-447.

Acronyms

B-BB	Bootstrap-Based Bounds
BSB	Bootstrap Bounds
CAIG	Cost Analysis Improvement Group
CER	Cost-Estimating Relationship
EST	Estimated
FY	Fiscal Year
IC	Intelligence Community
IRLS	Iteratively Reweighted Least Squares
K	Thousands (usually of dollars)
MPE-ZPB	Minimum Percentage Error – Zero Percentage Bias
NCG	NRO Cost Group
NRO	National Reconnaissance Office
OLS	Ordinary Least Squares
SEE	Standard Error of the Estimate

CLEARANCE REQUEST FOR PUBLIC RELEASE OF DEPARTMENT OF DEFENSE INFORMATION

(See instructions on back.)

(This form is to be used in requesting review and clearance of DoD information proposed for public release in accordance with DoDD 5230.9.)

TO: Director, Freedom of Information & Security Review, Rm. 2C757, Pentagon

1. DOCUMENT DESCRIPTION

a. TYPE Presentation for Posting to Website	b. TITLE Prediction Bounds for General-Error Regression CERs
c. PAGE COUNT 71	d. SUBJECT AREA Cost Analysis

2. AUTHOR/SPEAKER

a. NAME *(Last, First, Middle Initial)* Stephen A. Book	b. RANK Civilian	c. TITLE Chief Technical Officer
d. OFFICE MCR, LLC		e. AGENCY Contractor

3. PRESENTATION/PUBLICATION DATA *(Date, Place, Event)*

Request permission to post the presentation to the DoD Cost Analyst Symposium website at http://www.dodcas.osd.mil. The purpose is to make symposium presentations available to interested parties. The symposium was held on February 14-17, 2006 at the Williamsburg Marriott in Williamsburg, VA.

4. POINT OF CONTACT

a. NAME *(Last, First, Middle Initial)* Angers, Jeffrey P.	b. TELEPHONE NO. *(Include Area Code)* (703) 697-3845

5. PRIOR COORDINATION

a. NAME *(Last, First, Middle Initial)*	b. OFFICE/AGENCY	c. TELEPHONE NO. *(Include Area Code)*

6. REMARKS

7. RECOMMENDATION OF SUBMITTING OFFICE/AGENCY

a. THE ATTACHED MATERIAL HAS DEPARTMENT/OFFICE/AGENCY APPROVAL FOR PUBLIC RELEASE *(qualifications, if any, are indicated in Remarks section)* AND CLEARANCE FOR OPEN PUBLICATION IS RECOMMENDED UNDER PROVISIONS OF DODD 5320.9. I AM AUTHORIZED TO MAKE THIS RECOMMENDATION FOR RELEASE ON BEHALF OF:

ODNI/CAIG

b. CLEARANCE IS REQUESTED BY 20060130 *(YYYYMMDD)*.

c. NAME *(Last, First, Middle Initial)* Georges, John G.	d. TITLE Acting Dir, CAIG
e. OFFICE CAIG	f. AGENCY DNI
g. SIGNATURE John Georges	h. DATE SIGNED *(YYYYMMDD)* 20060120

DD FORM 1910, MAR 1998 (EG) PREVIOUS EDITION MAY BE USED. WHS/DIOR, Mar 98

Covert William

From: Fiume James
Sent: Friday, February 03, 2006 5:26 PM
To: Covert William
Subject: FW: Security Review Case S06-0081 - Prediction Bounds for General-Error-Regression CERs --- UNCLASSIFIED//FOUO
Classification: UNCLASSIFIED//FOUO

classification: UNCLASSIFIED//FOUO

From: Carey Regina
Sent: Friday, February 03, 2006 3:13 PM
To: Carey Regina; Martini Mark (WF); FOIA (IART)
Cc: Fiume James; Lugo Gretchen; DDNS-security
Subject: RE: Security Review Case S06-0081 - Prediction Bounds for General-Error-Regression CERs --- UNCLASSIFIED//FOUO

classification: UNCLASSIFIED//FOUO

Mark,

The Information Access and Release Team (IART) has completed the pre-publication/security review you requested. The "Prediction Bounds for General-Error-Regression" presentation has been approved for release. Please keep this e-mail as a record of the approval process.

Please contact me at 853-6122 if you have any questions and reference case # S06-0081.

Regina Carey
Information Access & Release Team
853-6122 (R)
227-9122 (B)

From: Carey Regina
Sent: Friday, January 20, 2006 1:23 PM
To: Martini Mark (WF); FOIA (IART)
Cc: Fiume James; Lugo Gretchen; DDNS-security
Subject: Security Review Case S06-0081 - Prediction Bounds for General-Error-Regression CERs --- UNCLASSIFIED

classification: UNCLASSIFIED

Mark,