Presented at the 2016 ICEAA Professional Development & Training Workshop

Naval Center for Cost Analysis (NCCA)

Cost Estimating Relationship (CER) Development Handbook

Study Team:

Bruce Parker, Patrick Staley, John Fitch (NCCA)

Adam James, Anna Irvine, Jeff Cherwonik, Brian Flynn (Technomics)

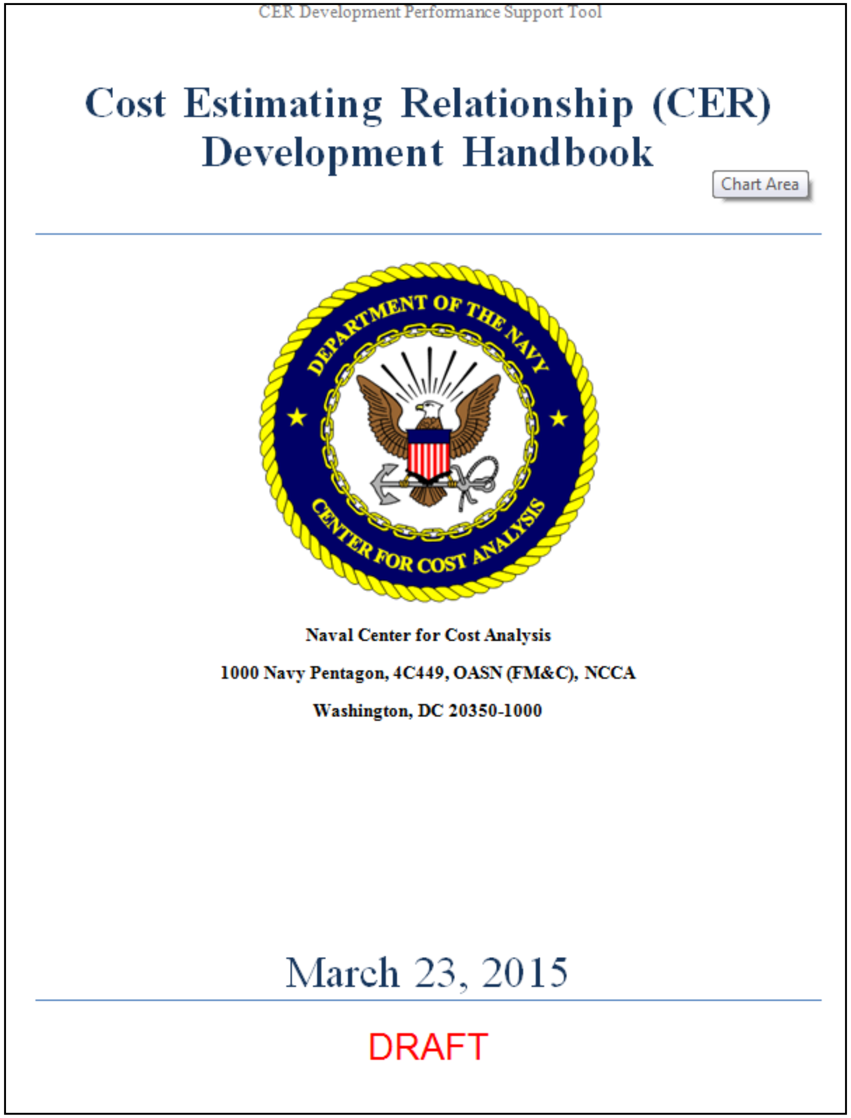
Alf Smith, Jeff McDowell (Tecolote)

Motivation/ Objective

- A handbook to provide cost estimators with current best practices analyzing data and developing CERs
- Incorporates traditional and modern methodologies in a logical flow
- Provides well-developed examples to follow
- Part of NCCA Tool Roadmap

Outline of Presentation

- Basic Steps
  - Collect data
  - Normalize
  - Analyze
  - Find cost drivers
  - Generate CER
  - Validate CER
  - Characterize uncertainty
  - Document
- Sample of Worked Examples
- Concluding Remarks


CER Development Performance Support Tool

Cost Estimating Relationship (CER) Development Handbook

Chart Area

Naval Center for Cost Analysis

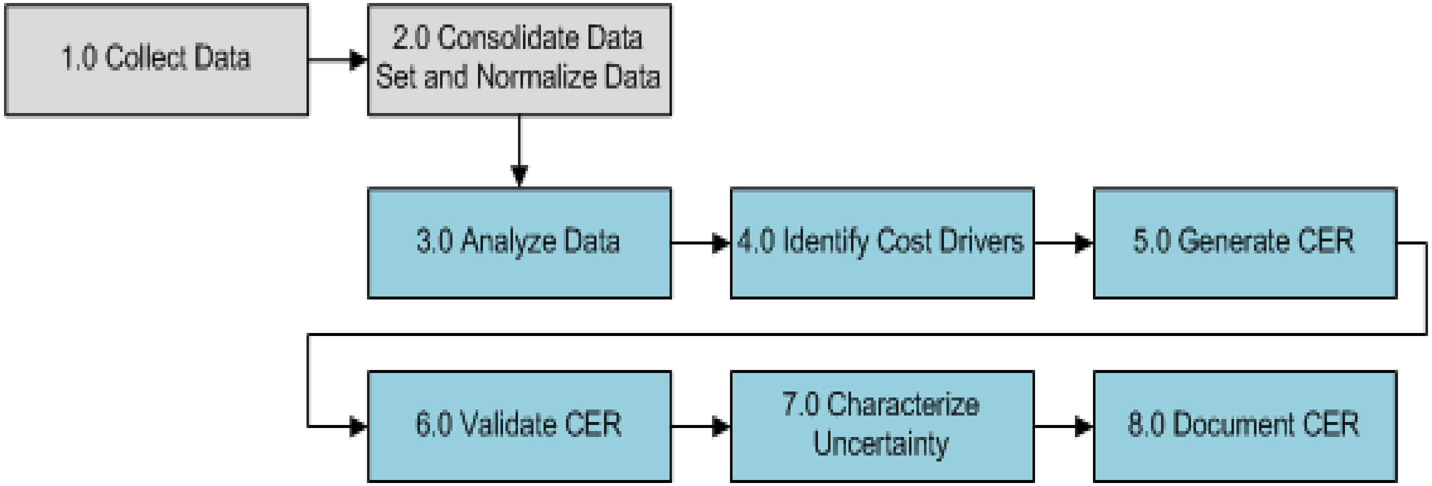
1000 Navy Pentagon, 4C449, OASN (FM&C), NCCA

Washington, DC 20350-1000

March 23, 2015

DRAFT

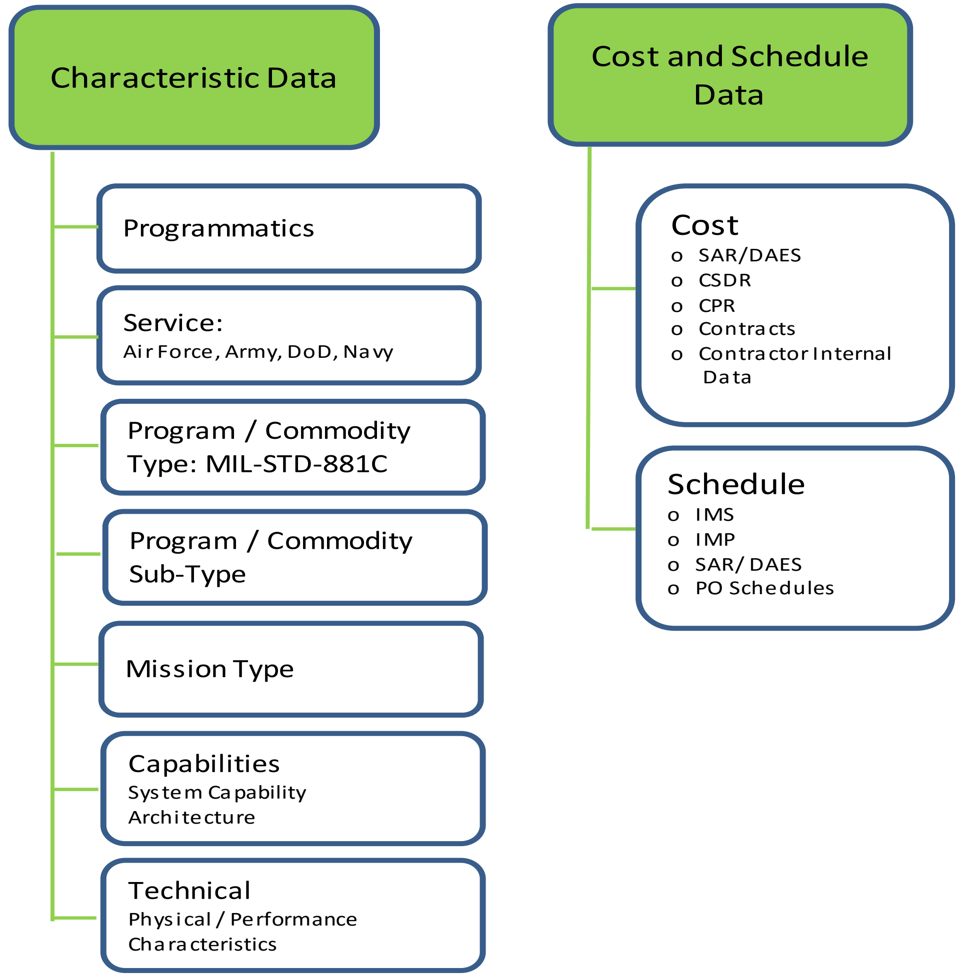
Basic Steps in CER Development Process



- The blue boxes correspond to the six core steps of the process, which are the focus of the HB.
- The preceding steps, denoted by gray boxes, are important to the CER development process, and a discussion of how to approach each of these steps is included.
- The analysis, in practice, is often iterative in nature, therefore the utility of the guide allows the user to move from one section to another to find the information most relevant to a particular question. The flow charts guide the flow of the document.

Step 1: Collect Data

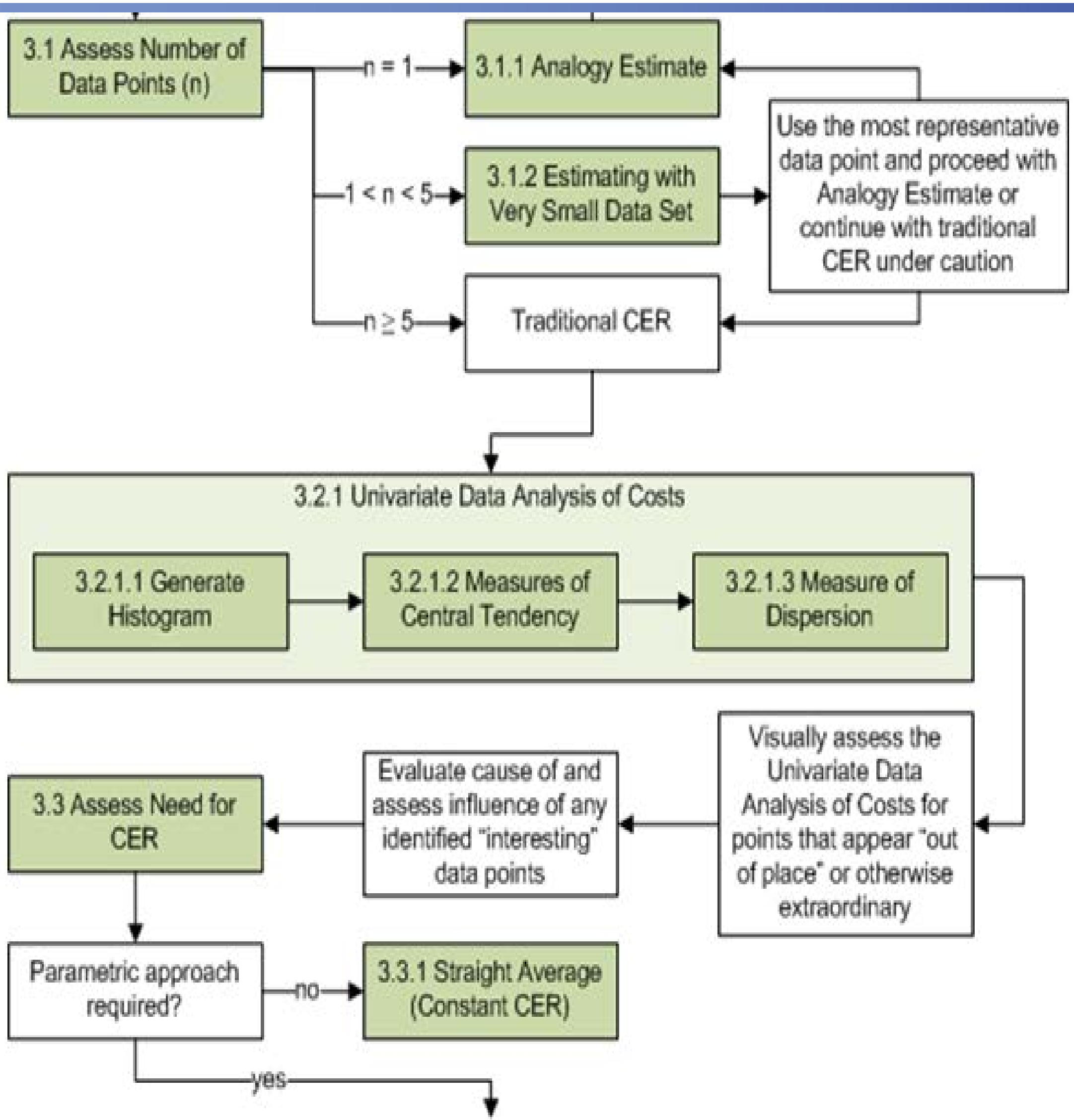
- Data Collection is the foundation of a reliable cost estimate
- Resources outlining best practices for cost estimating, including data collection and analysis:
  - GAO Cost Analysis Handbook (http://www.gao.gov/products/GAO-09-3SP)
  - Department of the Navy Cost Estimating Guide. (https://www.ncca.navy.mil/references/DON_Cost_Estimating_Guide.pdf)
- Familiarity with the programmatic, technical, manufacturing and supportability aspects of the program of interest is critical to understand the cost drivers



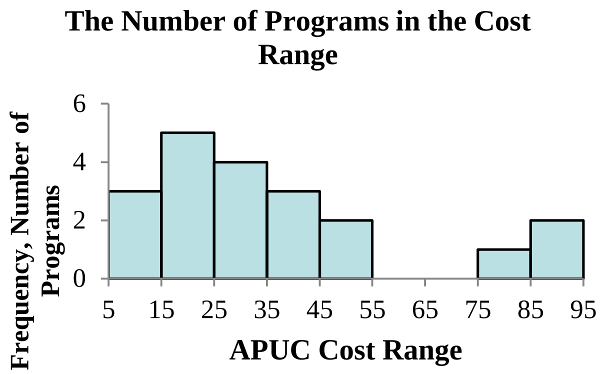
Step 2: Consolidate and Normalize Data

- Ensure that analogous systems or system components are consistent
  - May need to go to component level to make them more comparable
- Normalizing data for cost, quantity, and sizing; content; and physical parameters. Examples include (CEBoK Module 4):
  - Escalating historical costs to a common base year including overhead, Cost of Money (COM), General and Administrative (G&A) expenses, and fee
  - Normalize data to a single unit cost where applicable, dividing by the quantity and standardizing along the cost improvement curve
  - Addressing differences in manufacturing processes across states of development
  - Differences in technology across systems
  - Significant digits of numerical data (greater than overall estimate precision)

Step 3: Analyze Data



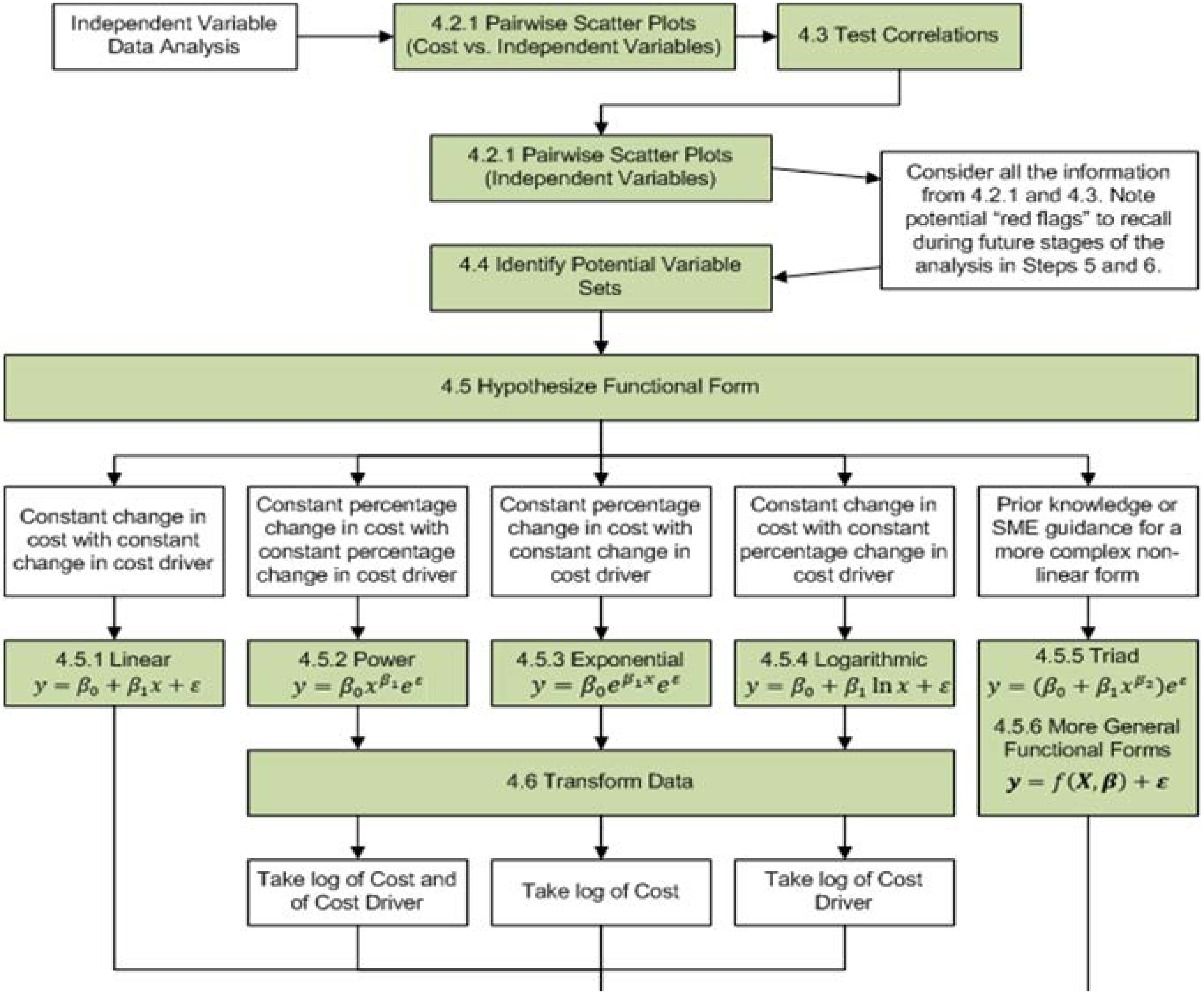
- Assess # of Data Points
  - Degrees of Freedom
- Univariate Data Analysis



- Mean, Median, Mode, Variance, StDev, CV, Range
- Assess need for CER
  - Constant CER
  - Parametric

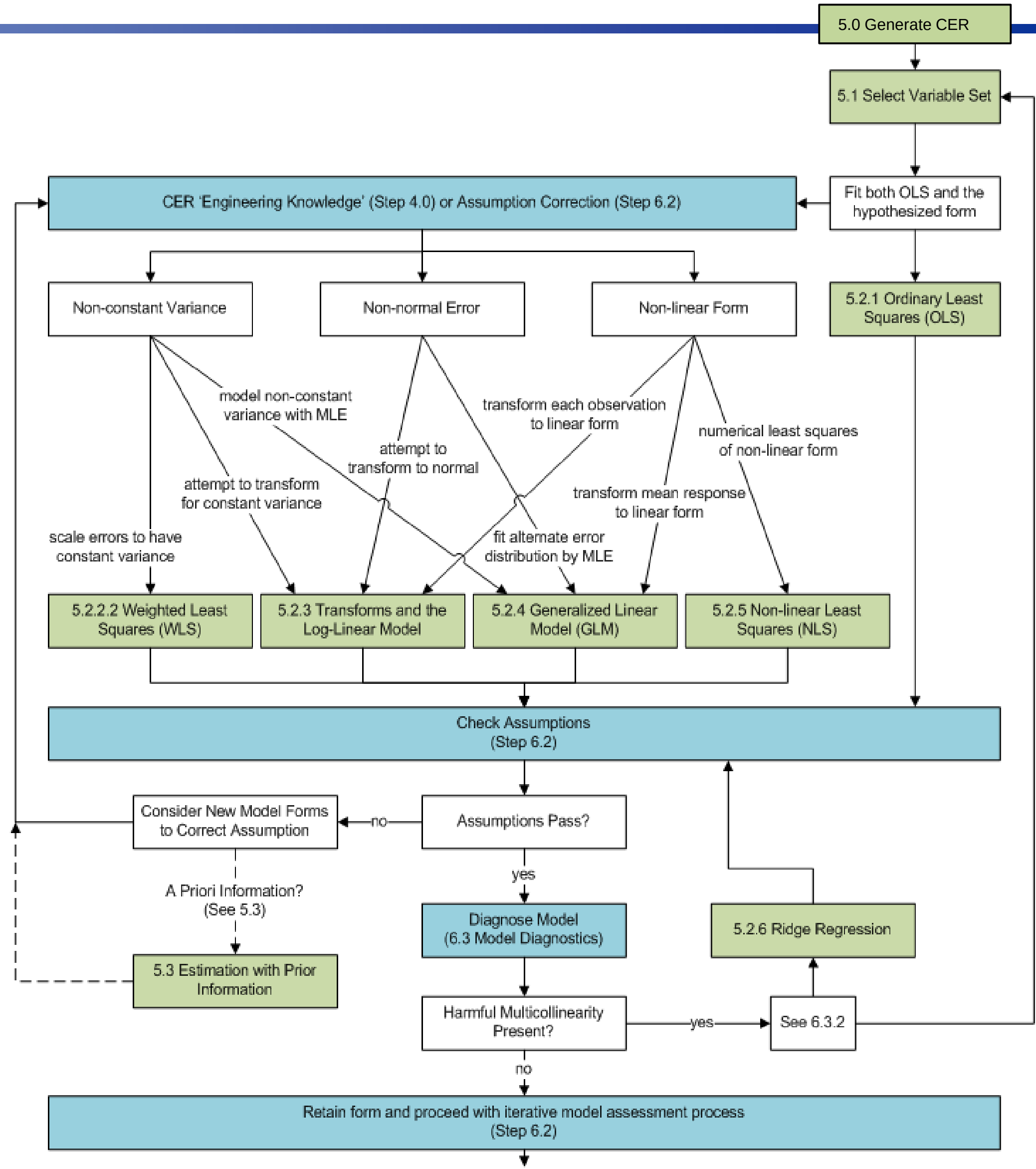
Step 4: Identify Cost Drivers

- Independent Variable Data Analysis
- Test Correlations
- ID Variables
- Hypothesize Functional Form
- Transform data, if needed



Step 5: Generate CER

- Select Variables to use
- Always fit the OLS model, even if hypothesizing another form
- Regression Methods
  - OLS, WLS, Transforms, GLM (MLE), NLS
  - Different tools to solve different model forms
- Check Assumptions to validate use of the model form
- A priori information as a tool
- Harmful multicollinearity
  - Ridge Regression
- Steps 5 and 6 can be (and often are) iterative



Step 5: Generate CER

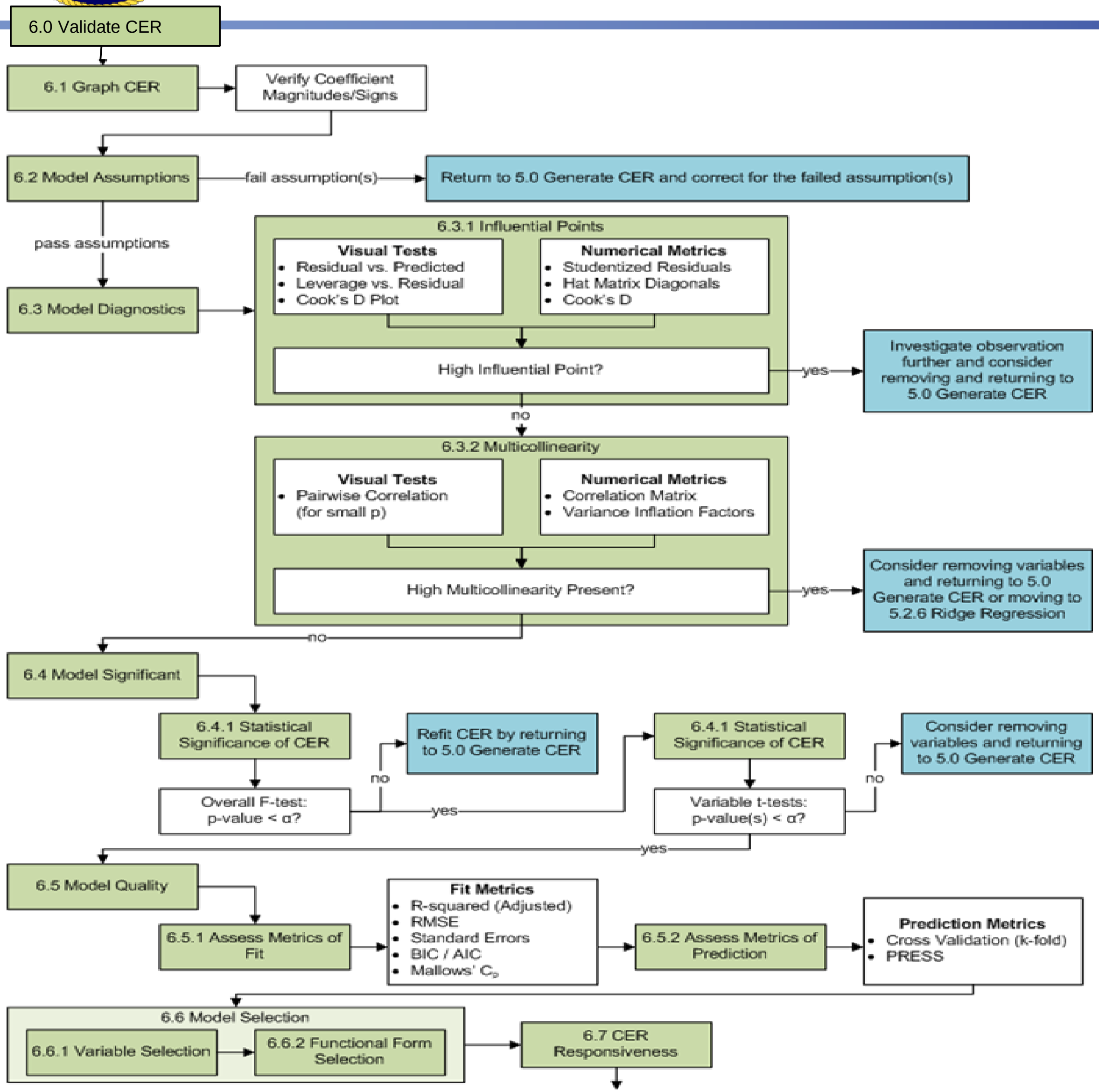
- Ordinary Least Squares (OLS) regression serves as the foundational model for which other models can be thought of as "remedies" to OLS deficiencies and/or assumption violations
  - "Core" OLS assumptions
    1. Independence of Errors
    2. Homoscedasticity
    3. Normality of Errors
    4. Linearity
  - Other issues such as multicollinearity, physical interpretations, computational efficiencies, etc. may drive model choice
- Models are specified by both a functional form and error structure

$$\boldsymbol{y} = f(\boldsymbol{X}; \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

some function of your predictors

some error term or function (can be "multiplicative" in nature)

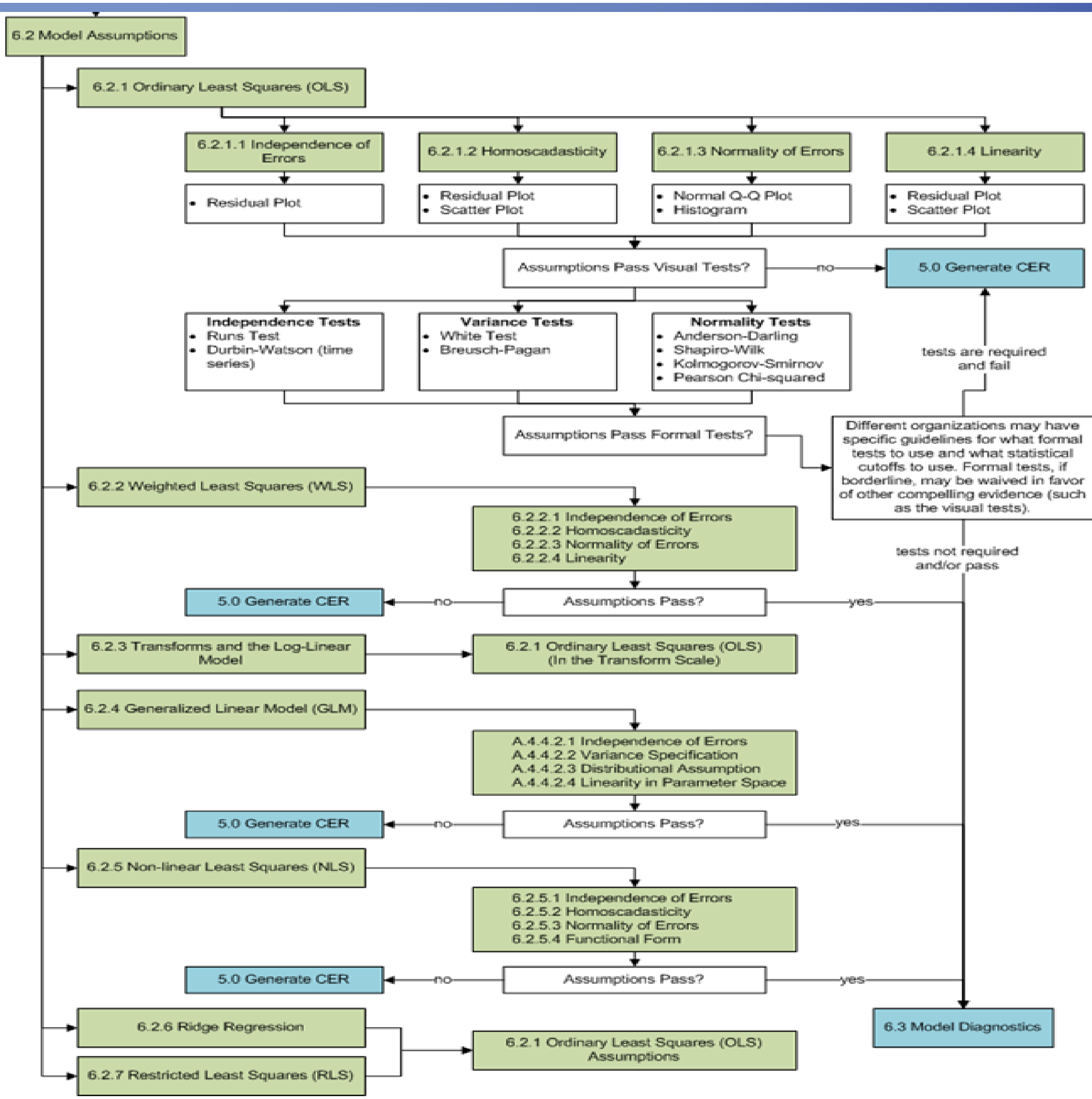
Step 6: Validate CER



Validating a CER is far more than just looking at an R-squared, F-statistic, or a single graph

- Understand the data relationships and coefficient estimates
- Determine consistency with engineering and physical principles
- Assess and validate the statistical model assumptions
- Identify and review high influence points such as leverage points and potential outliers
- Assess the impacts of multicollinearity
- Determine the significance of the model and independent variables
- Quantify metrics of best fit and prediction strength
- Compare and contrast multiple competing CERs to identify the "best" model

Step 6: Validate CER



Steps through each assumption and provides discussions of options to diagnose and remedy problems

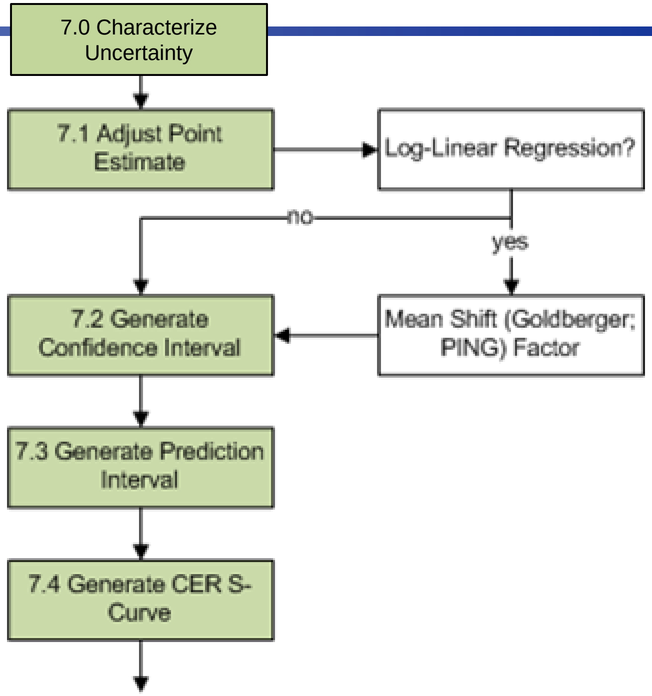
1. Independence of errors
   – Each error is distributed independently
2. Homoscedasticity
   – Each error is distributed identically (with the same variance)
3. Normality of Errors
   – Each error is distributed according to the normal distribution
4. Linearity
   – $\boldsymbol{y}$, is a linear function of the predictors

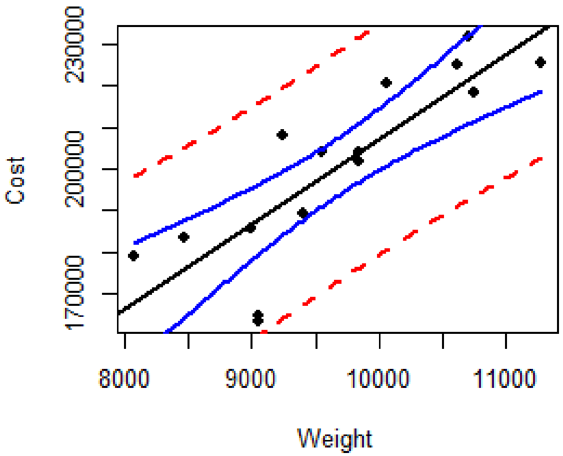
Step 6: Validate CER

- Model Assumptions
  - Assess and validate (both graphically and formally) the underlying assumptions
  - Violations of assumptions, depending on the severity, can completely invalidate a model
- Model Diagnostics
  - Diagnose and assess influential points (i.e., potential outliers and leverage points)
  - Assess multicollinearity which can potentially devastate a model
- Model Significance
  - Assess statistical significance of both predictors and the model
  - F-tests, t-tests, p-values, etc…
- Model Quality
  - Assess metrics of fit and prediction for the newly constructed CER
  - Often assess metrics for fit, when prediction is really of interest
- Model Selection
  - Choose a set of variables within one model form (such as linear or log-linear)
  - Select between different model functional forms

Step 7: Characterize Uncertainty

- Understanding the risk and uncertainty associated with a CER is crucial to accurate implementation
- Adjust Point Estimate: NCCA policy to estimate using the sum of the means
- The HB references the JA CSRUH





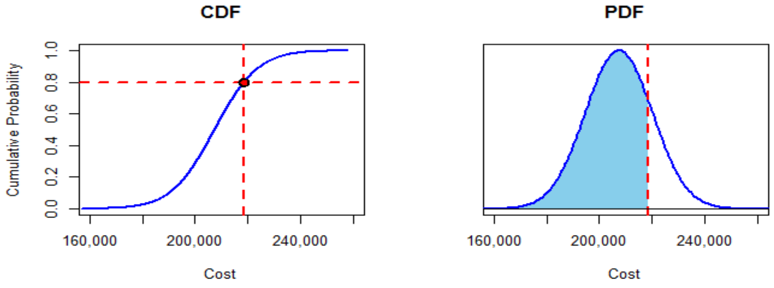
IV. Prediction Intervals

Estimate Inputs

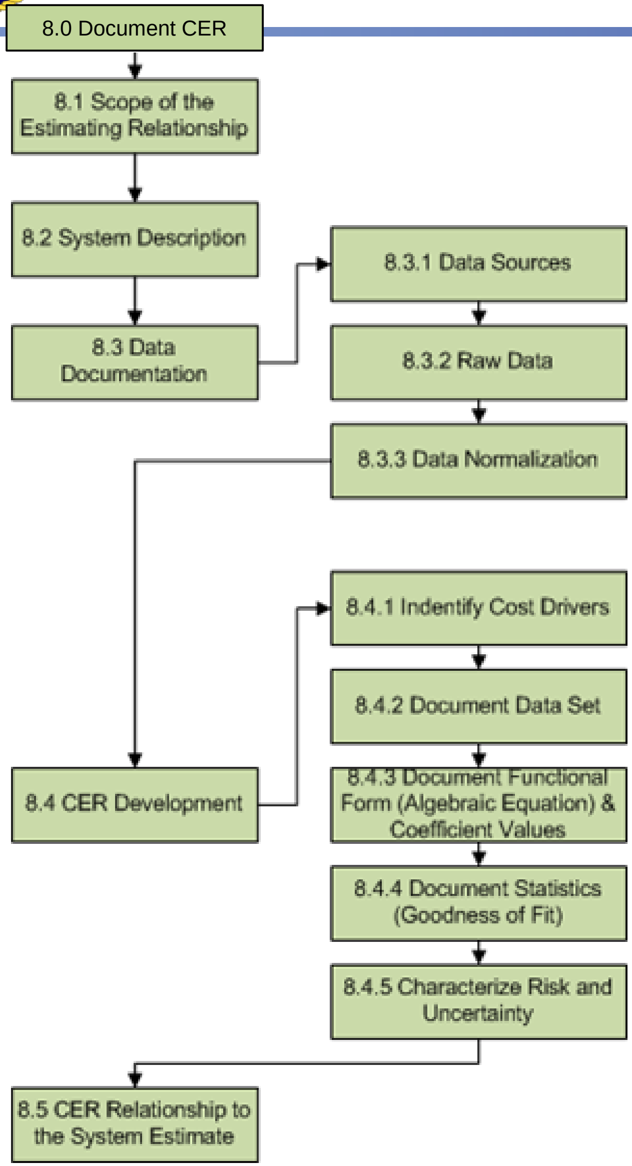
Input	x_null
HP	500.0000
Weight	10000.0000
Confidence Level (%)	95.00%

Prediction Results

Result	x_null
Lower Bound	179539.6779
Estimate	207370.4389
Upper Bound	235201.1998
Delta(%)	
Lower Bound	13.4208
Upper Bound	13.4208
RI$K(%) Multiplier	
Lower Bound	86.5792
Upper Bound	113.4208



Step 8: Document CER



- Comprehensive, accurate documentation is critical to enable traceability and auditability of a CER, and more broadly a cost estimate

CER HANDBOOK EXAMPLES

- First principles calculations for most examples are in an Excel workbook
- CO$TAT is used to demonstrate typical statistical package behavior

Notional Electronics Data

- Most example calculations make use of this data
- Every chart and tabular result can be found in an Excel workbook shipped with the handbook

Table 3

Observation	Cost (FY16$M)	Power (kW)	Cost per Unit Power ($M/kW)	Aperture (cm^2)	Power per Unit Aperture (kW/cm^2)	FFP (1) or T&M (0)
Project 1	$390	10.00	39.0000	8.70	1.149	1
Project 2	$200	5.00	40.0000	8.00	0.625	0
Project 3	$240	5.20	46.1538	8.20	0.634	1
Project 4	$300	7.00	42.8571			0
Project 5	$460	12.00	38.3333	9.00	1.333	1
Project 6	$560	17.80	31.4607	9.50	1.874	0
Project 7	$700	21.00	33.3333	9.20	2.283	0
Project 8	$800	25.00	32.0000	9.70	2.577	1
Project 9	$500	18.00	27.7778			0

Univariate Descriptive Statistics

Table 4

Observation	Cost (FY16$M)
Project 1	$390
Project 2	$200
Project 3	$240
Project 4	$300
Project 5	$460
Project 6	$560
Project 7	$700
Project 8	$800
Project 9	$500

Table 5

Measure of Central Tendancy	Cost (FY16$M)
Mean	$461
Median	$460
Mode	N/A

Table 6

Pecentile	Cost (FY16$M)
Minimum (0%)	$200
First Quartile (25%)	$300
Median (50%)	$460
Third Quartile (75%)	$560
Maximum (100%)	$800

Figure 7

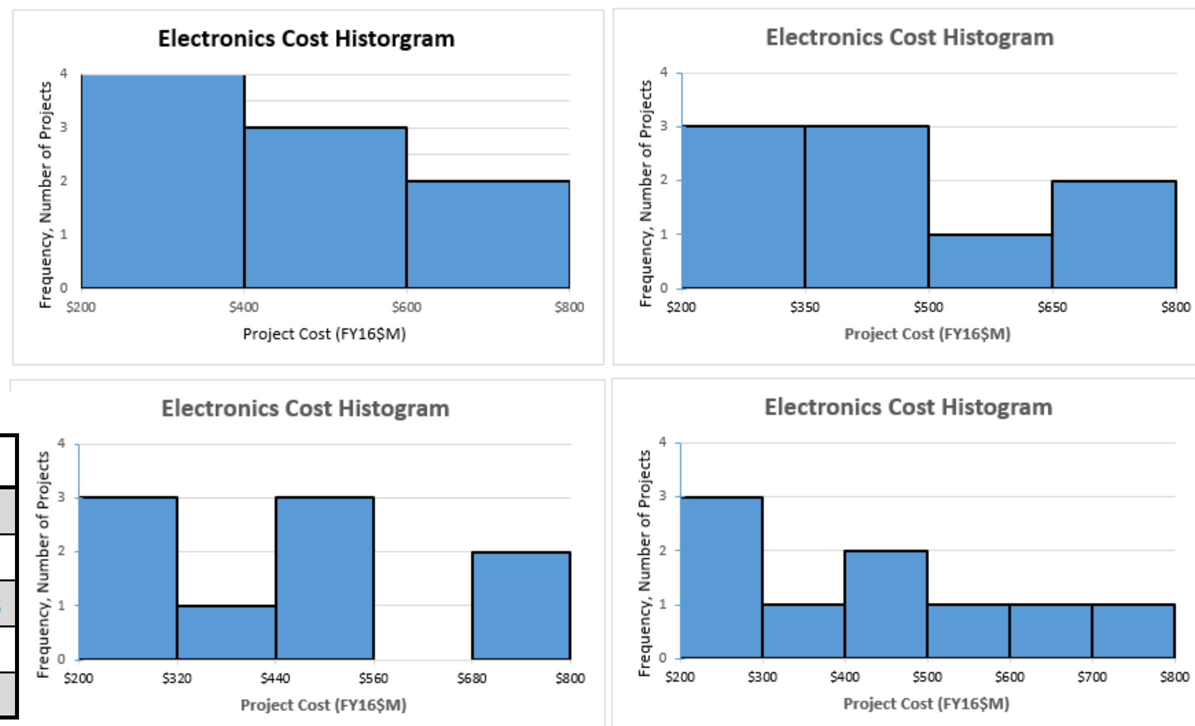


Table 7

Measures of Dispersion	Value	Units
Variance	41,411.11	$M^2
Standard Deviation	203.50	$M
Coefficient of Variation (CV)	0.441	Unitless
Range	600.00	$M
Interquartile Range (IQR)	260.00	$M

Univariate Inferential Statistics

- Excel workbook contains detailed calculations (sample to right) for every figure and table
- This figure illustrates how to calculate the confidence and prediction interval if your point estimate is the mean of a set of data

Confidence Interval of the Mean

$$\left(\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Prediction Interval of an Estimate

$$\left(\bar{X} - t_{\alpha/2,n-1}S\sqrt{1+\frac{1}{n}},\ \bar{X} + t_{\alpha/2,n-1}S\sqrt{1+\frac{1}{n}}\right)$$

		Confidence Interval			Prediction Interval		
Prob	t INV	CI	Bound	Height	PI	Bound	Height
99.9%	4.5008	305.30	$766.41	0.0013	965.44	1,426.55	0.0007
99.0%	2.8965	196.47	$657.58	0.0153	621.30	1,082.42	0.0077
97.5%	2.3060	156.42	$617.53	0.0390	494.65	$955.76	0.0195
95.0%	1.8595	126.14	$587.25	0.0768	398.88	$859.99	0.0384

Figure 8

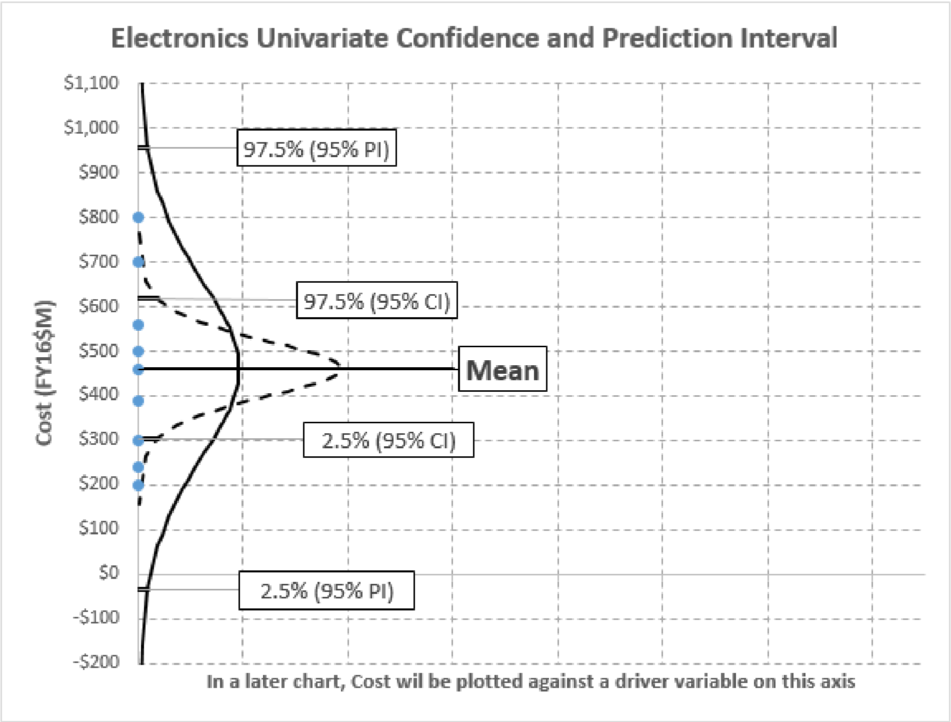


Table 8

Mean	$461.11
Std Dev	$203.50
Confidence Interval Standard Error of the Mean	$67.83
Prediction Interval Standard Error of the Mean	$214.50
t distribution at 97.5%, 8 degrees of freedom	2.3060
97.5% bound for the 95% Confidence Interval	$617.53
97.5% bound for the 95% Prediction Interval	$955.76

OLS Regression

Table 12

Observation	Cost (FY16$M)	Power (kW)
Project 1	$390	10.00
Project 2	$200	5.00
Project 3	$240	5.20
Project 4	$300	7.00
Project 5	$460	12.00
Project 6	$560	17.80
Project 7	$700	21.00
Project 8	$800	25.00
Project 9	$500	18.00

Figure 24 (26?)

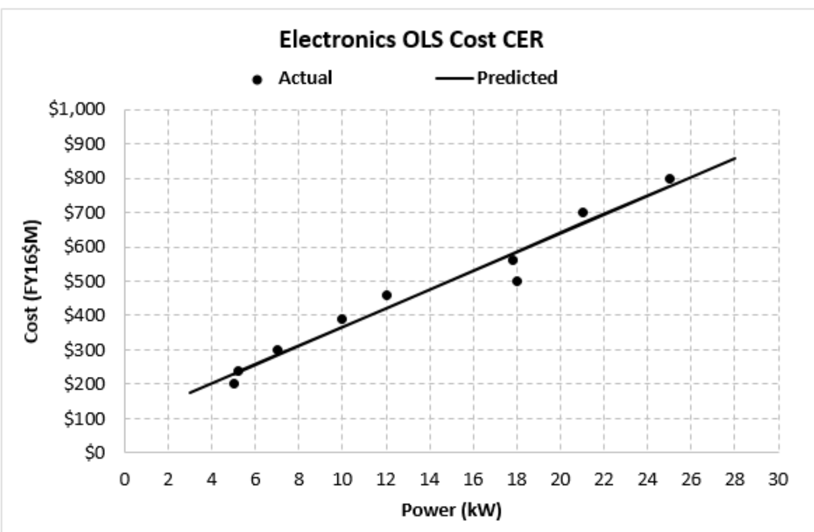
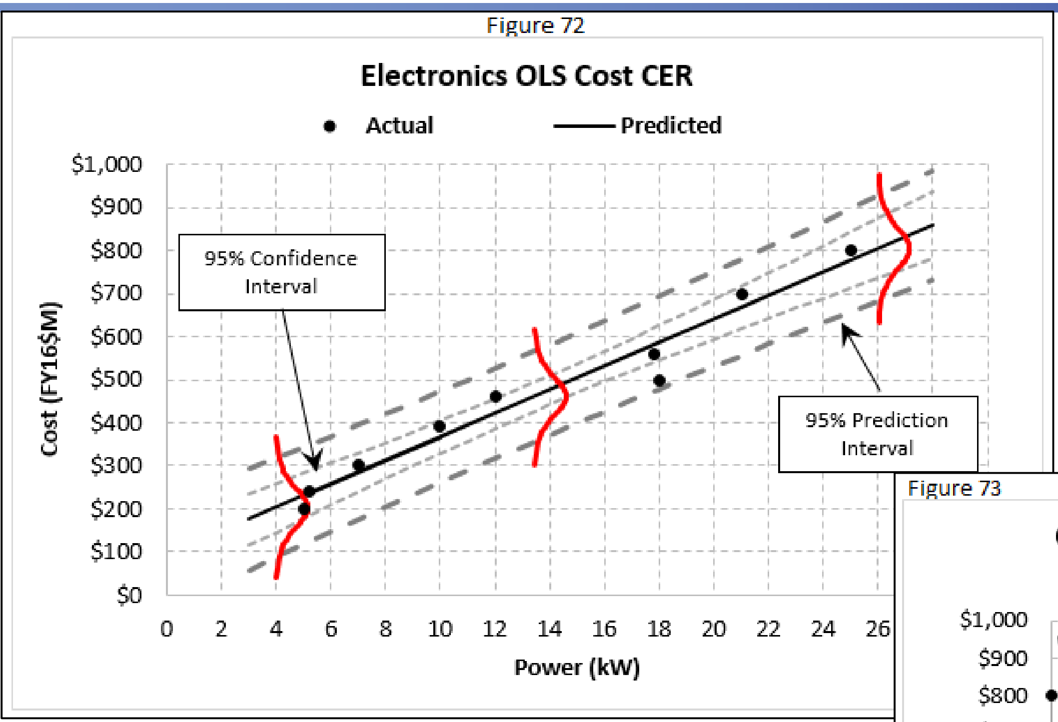


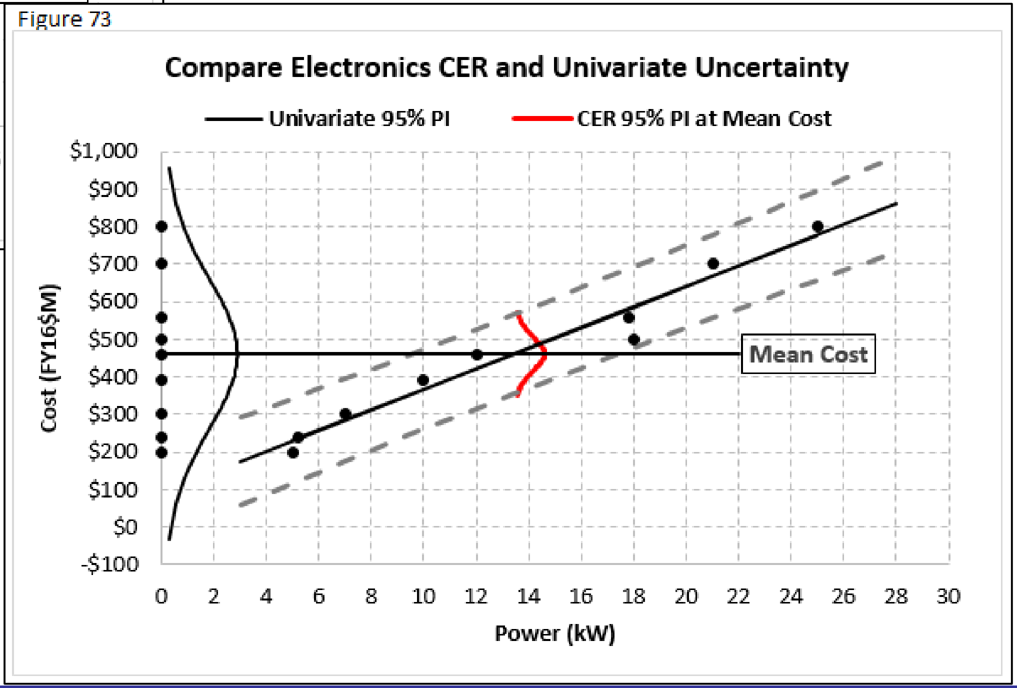
Figure 69

Parameter	Formula	Result
Slope	$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$	27.3853
Intercept	$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$	92.9309
Standard Error of the Estimate	$S_e = \sqrt{\frac{\sum(y - \hat{y})}{n - 2}}$	$42.2261
Standard Error of the Intercept	$S_y = S_e\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\sum(x_i - \bar{x})^2}}$	$30.9235
Standard Error of the Slope	$S_b = \frac{S_e}{S_x\sqrt{n - 1}}$	$2.0480
Confidence Interval (CI) Standard Error	$S_{yc} = S_e\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$	*Power = 13.44* $14.08 Power = 26 $29.31
Prediction Interval (PI) Standard Error	$S_{yp} = S_e\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$	*Power = 13.44* $44.51 Power = 26 $53.57

Compare Univariate to Linear Regression



Excel workbook has detailed calculations to produce these charts



- By using same data throughout, can compare results
- Figure 73 demonstrates increased precision of the OLS CER over using the mean of the cost as an estimate

Weighted Least Squares

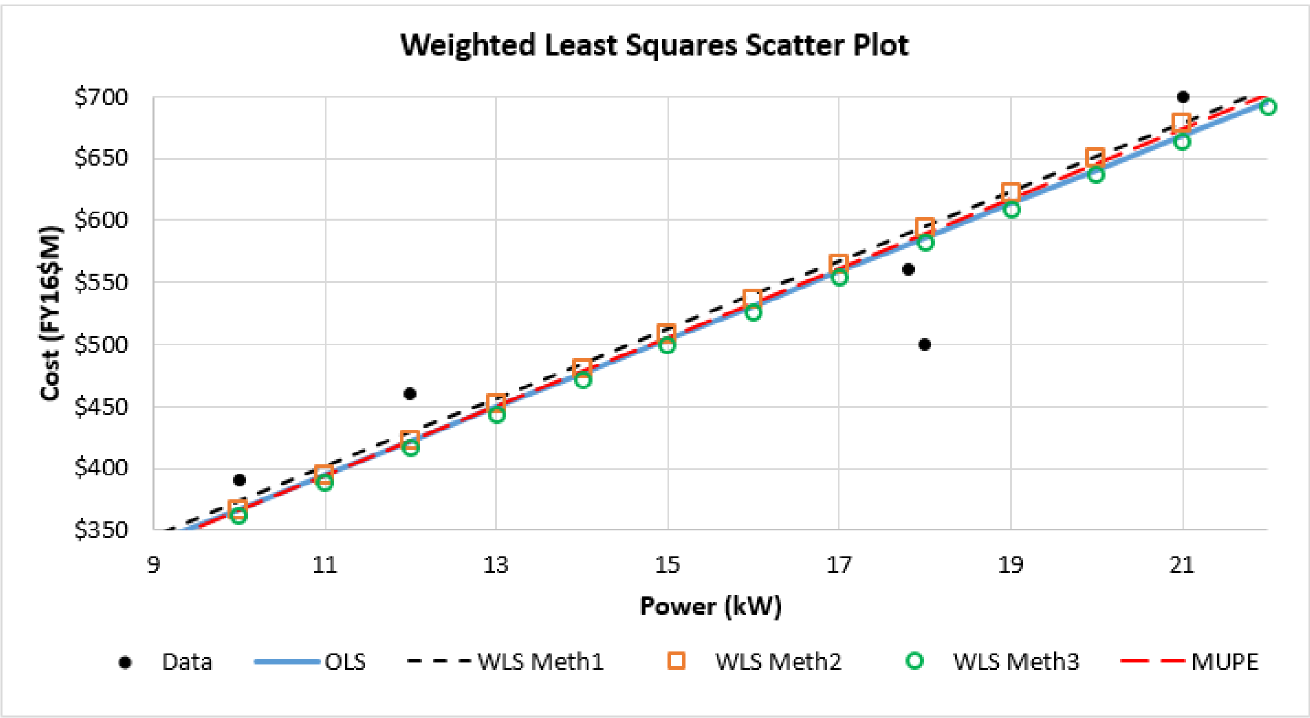
Table 14

Observation	Cost (FY16$M)	Power (kW)	Weight Methods 1, 2, 3: 1/OLS Err^2	1/Pwr^2	1/Cost^2
Project 1	$390	10.00	0.0018553	0.0100000	0.0000066
Project 2	$200	5.00	0.0011218	0.0400000	0.0000250
Project 3	$240	5.20	0.0459405	0.0369822	0.0000174
Project 4	$300	7.00	0.0042319	0.0204082	0.0000111
Project 5	$460	12.00	0.0006766	0.0069444	0.0000047
Project 6	$560	17.80	0.0024054	0.0031562	0.0000032
Project 7	$700	21.00	0.0009779	0.0022676	0.0000020
Project 8	$800				
Project 9	$500				

Table 15

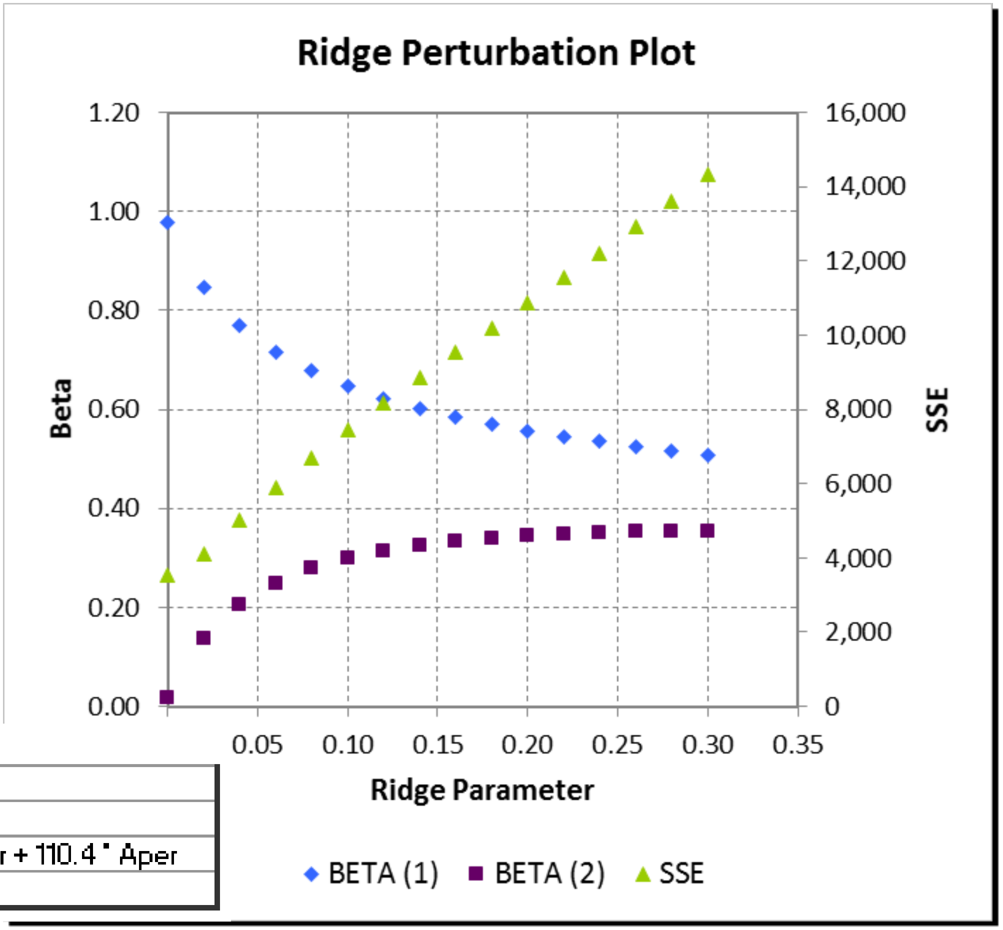
Name	Equation	Intercept Prob not 0	Slope Prob not 0	In Unit Space: R^2 Adj	SE	MAD
OLS	92.93 + 27.39 * Power	98.02%	100.00%	95.69%	42.23	7.17%
Method 1	95.74 + 27.78 * Power	100.00%	100.00%	95.47%	43.32	6.71%
Method 2	81.8 + 28.43 * Power	99.62%	100.00%	95.51%	43.12	7.17%
Method 3	85.13 + 27.58 * Power	99.47%	100.00%	95.60%	42.66	7.56%
Method 4 (MUPE)	86.65 + 27.95 * Power	99.60%	100.00%	95.61%	42.48	6.89%

Figure 29



Ridge Regression

- Handbook goes through the process to identify that Power and Aperture are correlated (multicollinearity)
- Ridge regression explained to address the issue
- Ridge plot shows that as ridge parameter increases, SSE also increases



I. Model Form and Equation Table

Model Form:	Unweighted Linear model
Number of Observations Used:	7
Equation in Unit Space:	Cost = (-751) + 17.98 * Power + 110.4 * Aper
Ridge Parameter Specified:	0.1200

Other Examples

- The handbook is your one stop shop for cost analysis regression and statistics
- Worked examples include:
  - Dummy Variables
  - Generalized Least Squares (GLS) (IRLS, MUPE)
  - Generalized regression (ZMPE)
  - Transformable Linear and the Log-Linear Model
  - Generalized Linear Model
  - Estimation with Prior Information
    - Exact Prior Information on Parameter Relationships
    - Pseudo-Exact Prior Information on Parameter Values

Concluding Remarks

- Initial draft of handbook completed Mar 2015
- Leads the analyst through a logical and systematic process to perform cost statistical analysis and regression
  - Core concepts, mathematics and worked examples are described in detail
  - The more esoteric technical content is found in the appendices
- Draft now being refined and a single set of examples being established throughout
- Approved NCCA CER Handbook will be published at www.ncca.navy.mil on or before 26 August 2016
- Future goals: broader coordination of handbook, training course development, and tool enhancements

CER HANDBOOK BACKUP

Multivariate Using Matrix Math

- Excel contains the matrix math to solve for Cost = 37.31 + 28.21 * Power + 6.105 * Aper
- ANOVA and t statistics also provided

=MMULT(H24:J26,L19:L21)

Design Matrix		
X	Power	Aperture
1	10.00	8.70
1	5.00	8.00
1	5.20	8.20
1	12.00	9.00
1	17.80	9.50
1	21.00	9.20
1	25.00	9.70

Cost
Y
$390.0
$200.0
$240.0
$460.0
$560.0
$700.0
$800.0

X'X		
7.000	96.000	62.300
96.000	1678.880	882.440
62.300	882.440	556.910

X'Y
3,350.000
56,336.000
30,621.000

INV(X'X)		
228.306	2.210	-29.041
2.210	0.025	-0.287
-29.041	-0.287	3.705

b
37.313
28.213
6.105