PT 02 – Cost Estimating Relationships

Adapted from:
CEBOK Module 3 &
ISPA Parametric Estimating Handbook

"Parametric estimating is a technique that develops cost estimates based upon the examination and validation of the relationships which exist between a project's technical, programmatic, and cost characteristics as well as the resources consumed during its development, manufacture, maintenance, and/or modification."
-Parametric Estimating Handbook

Parametric Estimating Handbook (4th ed.), ISPA, 2007.

ISPA ICEAA CEBoK

Unit Index

Unit I – Cost Estimating

1. Cost Estimating Basics
2. Costing Techniques
3. *Parametric Estimating*

Unit II – Cost Analysis Techniques

Unit III – Analytical Methods

Unit IV – Specialized Costing

Unit V – Management Applications

Parametric Estimating Overview

- Key Ideas
  - Cost Drivers (and "Cost Passengers")
  - Inputs and Outputs
  - Parametric Models
- Practical Applications
  - CER Development
  - Cost Response Curves (CRCs)
  - Sensitivity Analysis
  - Schedule or Weight Estimating Relationships
- Analytical Constructs
  - Linear equations
  - Other functional forms
    - Power, exponential, log, polynomial
  - Curve fitting
- Related Topics
  - COTS Cost Models 3
  - Manufacturing Cost Estimating 11
  - Software Cost Estimating 12
  - Trade Studies 16

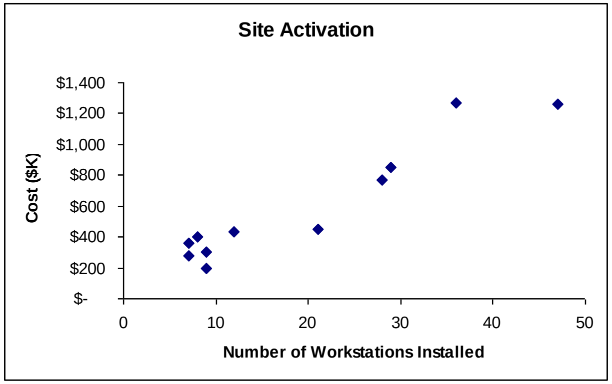
NEW!

Parametric Estimating Within The Cost Estimating Framework

v1.0

Past

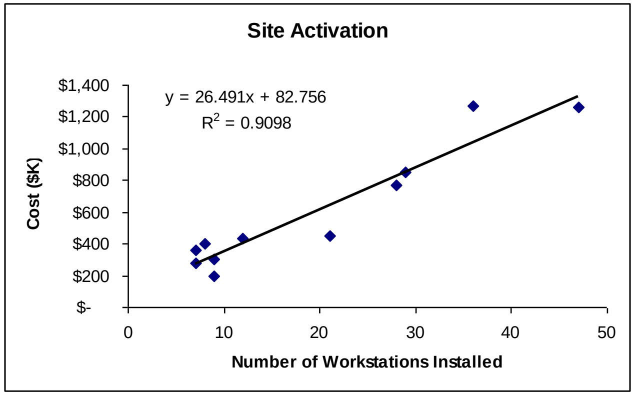
Understanding your historical data



Historical costs for similar systems

Present

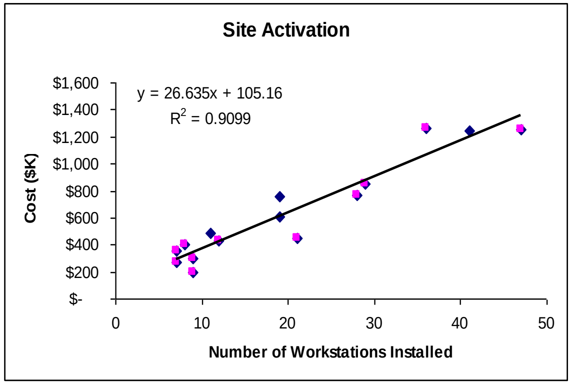
Developing estimating tools



Cost Estimating Relationships

Future

Estimating the new system



Applying, Validating, and Updating CERs

4

NEW!

Parametric Estimating Outline

- Core Knowledge
  - Basics of Parametric Estimating
  - Parametric Estimating Process
    - Collecting data
    - Identifying cost drivers
    - Developing CERs
    - Building a parametric model
  - Parametric Estimating Examples
- Summary
- Resources
- Related and Advanced Topics

Parametric Estimating Technique

3

- Parametric Estimating – using cost estimating relationships (CERs) based on historical data to estimate a project's cost
  - Uses cost drivers to estimate cost and production schedules
    - Technical parameters, like size and weight
    - Performance parameters, like speed and accuracy
  - Advantages
    - Model is easily traceable and objective
    - Can be easily adjusted for changes to the system or sub-system by modifying input parameters
      - Important for sensitivity analysis and design trades
    - Statistical results relating to the model
      - Includes t statistics, F statistics, R^2, coefficient of variation (CV)
      - Objective measures of validity
      - Used in risk analysis

6

8

9

Tip: Many cost estimators confuse parametric estimating with the use of COTS models. Most COTS cost models are parametric, but many parametric models are not COTS.

Cost Estimating Techniques

2

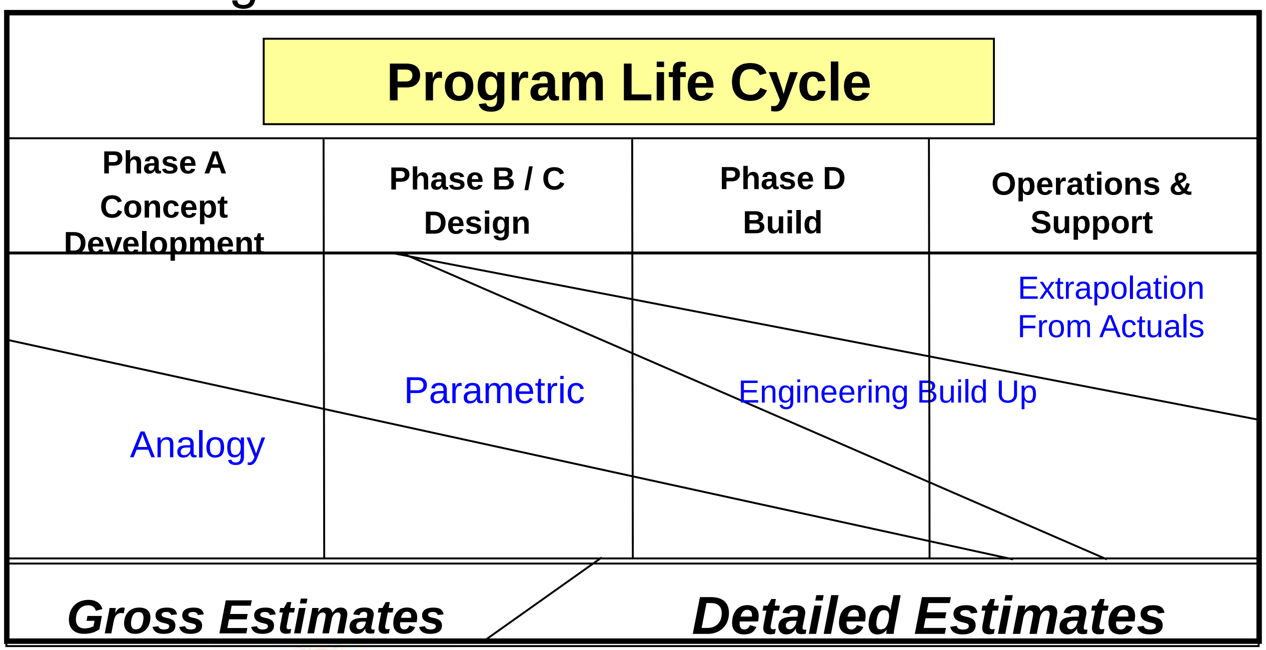
10

Technique	Approach	Considerations
Analogy	Comparative analysis of similar systems	Can be used early in programs before detailed requirements are known, but there is no objective test of validity
Parametric Estimating	A mathematical relationship between a parameter and a cost	Historical data is difficult to obtain, but CERs can be easily adjusted for requirements changes and provide statistical results
Engineering Build-Up	Estimating is done at lower levels and results rolled up to produce higher-level estimates	It is easy to see exactly what the estimate includes, but it is expensive and requires detailed data to be collected, maintained and analyzed

Parametric Estimating and the Life Cycle

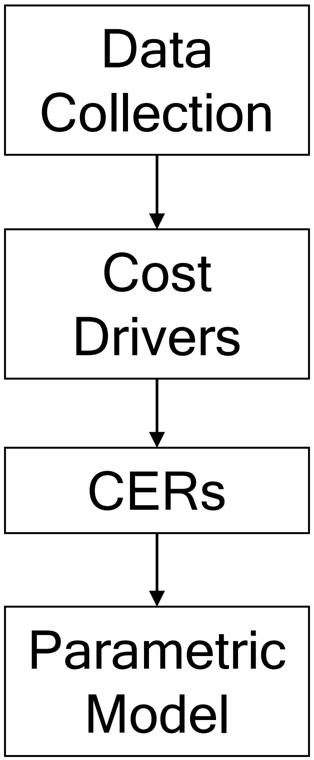
1

- Different techniques are used for the different stages in the acquisition/product development cycle
  - Early in the program, cost estimates are based on analogies and parametrics
  - As the program matures, the estimate will incorporate greater amounts of actuals



Adapted from Integrated Defense Acquisition, Technology, and Logistics Life Cycle Management Framework chart, http://akss.dau.mil/ifc

Parametric Estimating Process



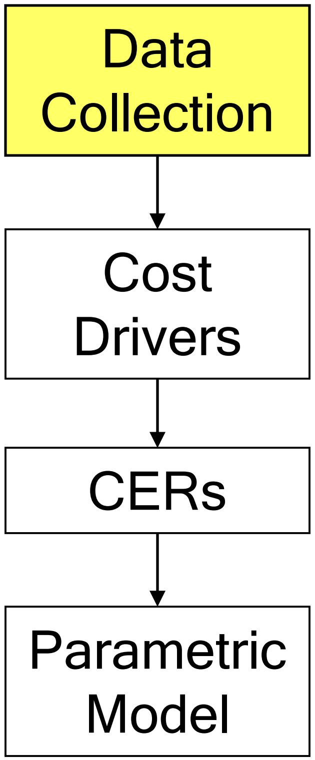
- Parametric estimating is a process

1. Collecting data (cost, schedule, and technical) from historical programs
2. Identifying cost drivers
3. Developing Cost Estimating Relationships (CERs)
4. Building a parametric model

11

Tip: Data collection and cost driver identification are an iterative process.

Process – Data Collection

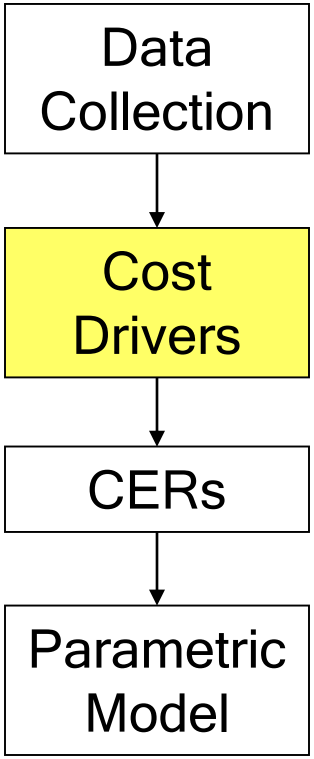


4

1. Collecting data (cost, programmatic, schedule, and technical) from similar historical programs
   - Corresponding data from current program can be obtained from program cost reports, schedules, and requirements documents
   - Some examples of the types of specific types of data are lines of code, site deployment plans, sub-system procurement, test and evaluation (T&E) schedules, weight, speed, crew size, steaming hours underway

Tip: It is important to document sources, reasoning, assumptions, and raw data so that the development of the parametrics can be traced.

Process – Cost Drivers



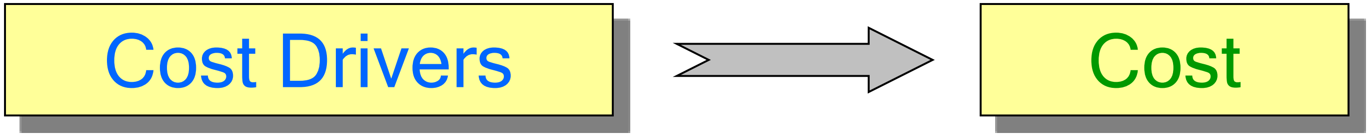
2. Identifying cost drivers

- Cost drivers will be the independent variables in your CERs
- How to identify:
  - Talk to subject matter experts
  - Understand technical and operational parameters
  - Scatter plots

2

6

Warning: Parameters that drive design are *not* always the parameters that drive cost.



Typical Cost Drivers

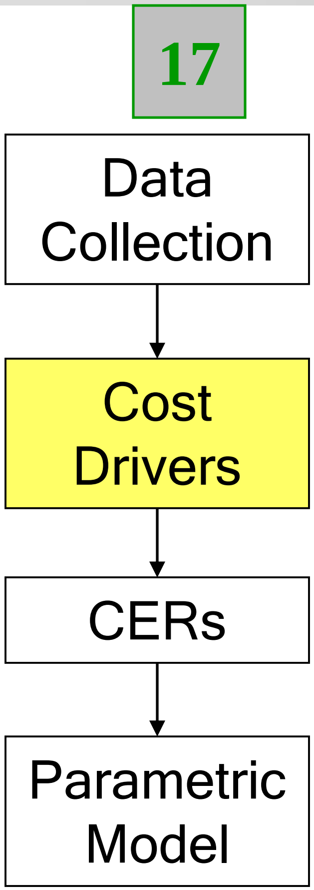
- Mass or Weight
- Software Lines of Code or Function Points
- Processing Speed
- Memory Capacity
- Signal Frequency
- Data-Processing and Downlink Throughput
- Power Requirements
- Antenna Diameter
- Special Mission Equipment

Developing a CER – Logic Testing

- Logic Testing
  - CER must be logical
  - Which comes first:
    - Hypothesis?
    - Database?
  - Subject Matter Experts (SME) are invaluable
  - Judgment is often statistical or results based

- The CER Model
  - Pronounced model: Based on the hypothesis of an SME; e.g. $1M for every foot of documentation.
  - Regressed Model: Based upon statistical methods to fit an equation to data.

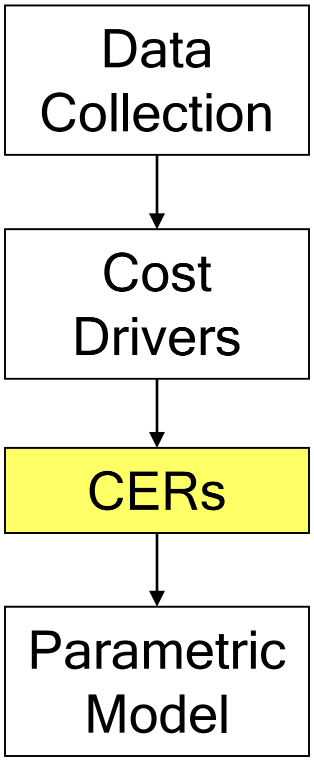
Cost Drivers vs. Cost Passengers



16

- Cost drivers are characteristics that drive cost
  - Parameters with greatest leverage on cost – WHY and HOW
  - Identifying usually requires understanding, expertise
  - Most potential for cost savings, may be hard to quantify
  - May include armament, manning, automation, speed, accuracy, reliability, maintenance philosophy, etc.
- Cost passengers are the “big ticket items”
  - Cost elements in WBS with highest values – WHAT and WHERE
  - Can be found with comb charts or Pareto charts
  - Not always the elements with biggest potential for cost savings

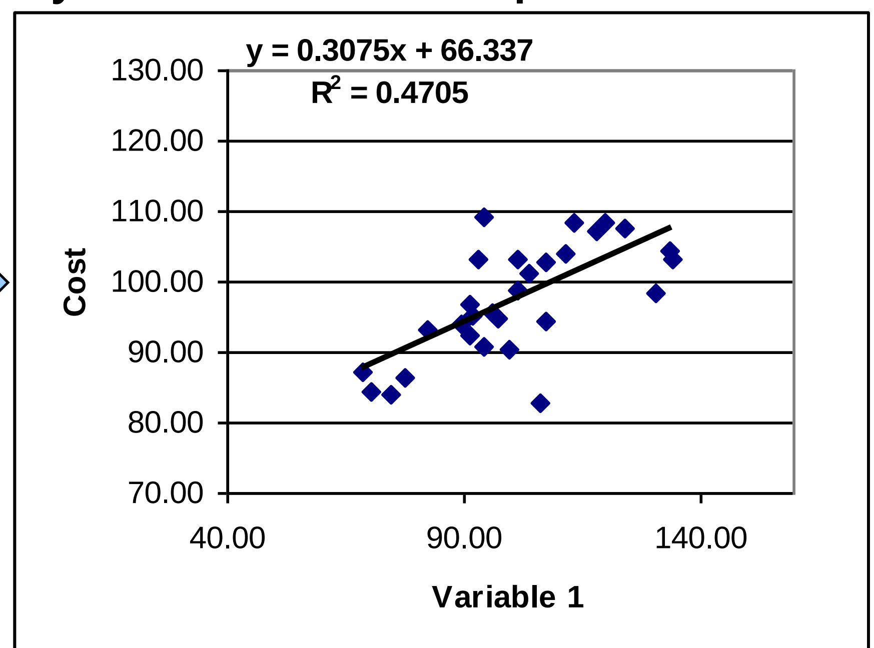
Process - CERs



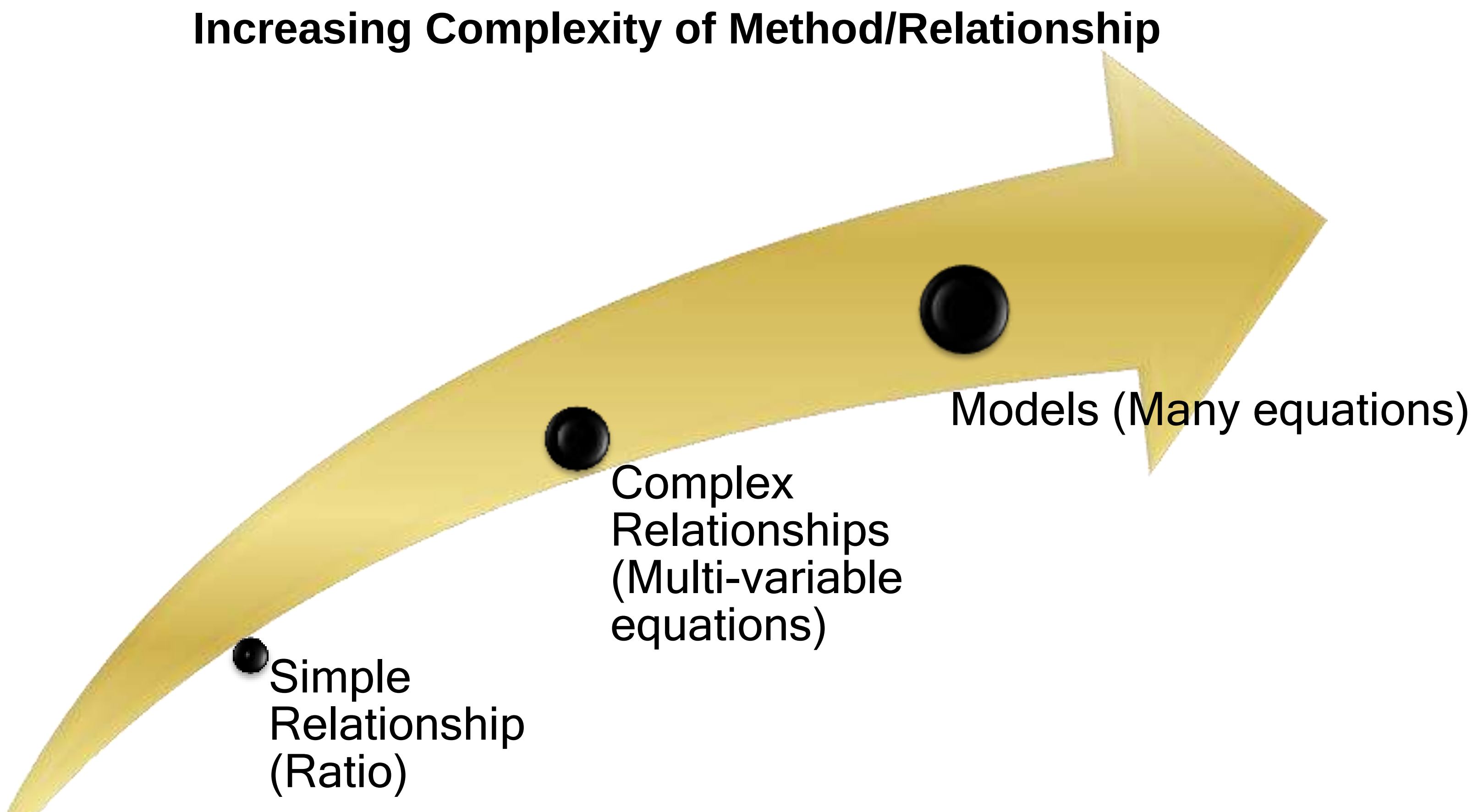
3. Developing Cost Estimating Relationships (CERs)

- After the available historical and industry data are collected for the system, the cost analyst then analyzes "relationships" within the data

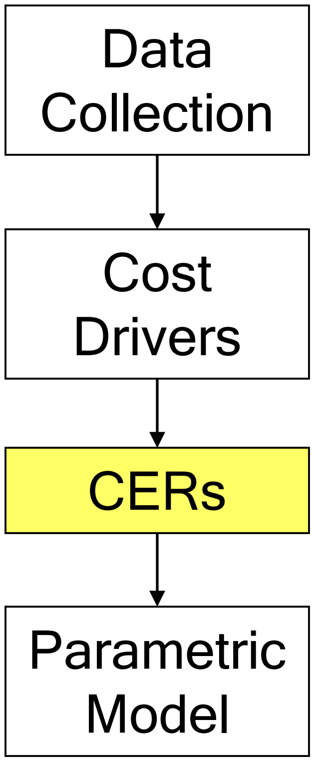
6



Continuum of CER Complexity



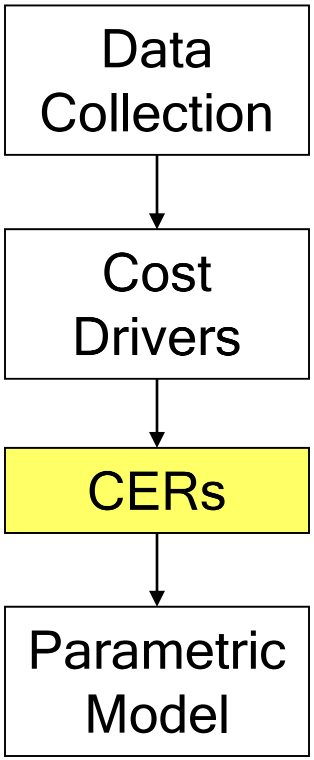
CERs - Rates, Factors, and Ratios



- Rates: Cost on Parameter 4
  - Example: The average cost for 1 GB RAID storage is $105
- Factors: Cost on Cost
  - Example: System Engineering/ Program Management (SE/PM) = 20% of the program's prime mission equipment (hardware and software) 5
- Ratios: Parameter on Parameter 6
  - Example: 1,200 lines of code to integrate a COTS software package (industry average)

Warning: These three terms are sometimes carelessly used interchangeably.

CERs - Regression



- The preferred method for deriving CERs is via regression analysis:
  1. Select Variables
  2. Test Relationships
  3. Perform Regression (Ordinary Least Squares (OLS))
     - Examples of equations:
       - Linear: $y=a+bx$
       - Power: $y=ax^b$
       - Logarithmic: $y=a+b \ln x$
       - Exponential: $y=ae^{bx}$
       - Other
         - Polynomial: $y=a+b_1x+b_2x^2+b_3x^3+\ldots+b_nx^n$

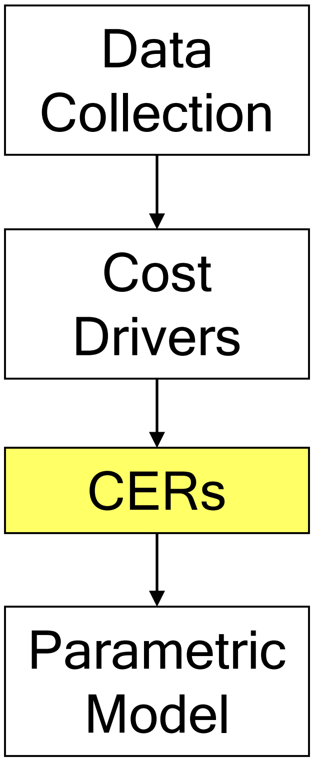
6

8

7

16

CERs - Regression



- The preferred method for deriving CERs is via regression analysis:

4. Select CERs
5. Validate CERs
   - Graphical & Ordinary Lease Squares (OLS)
   - Significance Tests
6. Document CERs

Warning: The derived CER can readily be applied to values both within and outside the range of the data used to derive the relationship. Use the Prediction Interval methodology to accurately capture uncertainty.

8

8 **9**

Traditional Linear Regression

- Linear CER Additive-Error Model

$$y = a + bx + \varepsilon$$

(True Cost = Estimated Cost + Error of Estimation)

- Ordinary Least-Squares (OLS) Regression Minimizes Sum of Squared Errors
  - **Actual cost for data point i is y_i**
  - **Estimated cost for data point i is $a + bx_i$**
  - **Error of estimation for data point i is $\varepsilon_i = y_i - (a + bx_i)$**
  - **Choose values for a and b that minimize $\sum (y_i - a - bx_i)^2 = \sum \varepsilon_i^2$**
  - **Resulting estimates are unbiased**
- OLS Solution

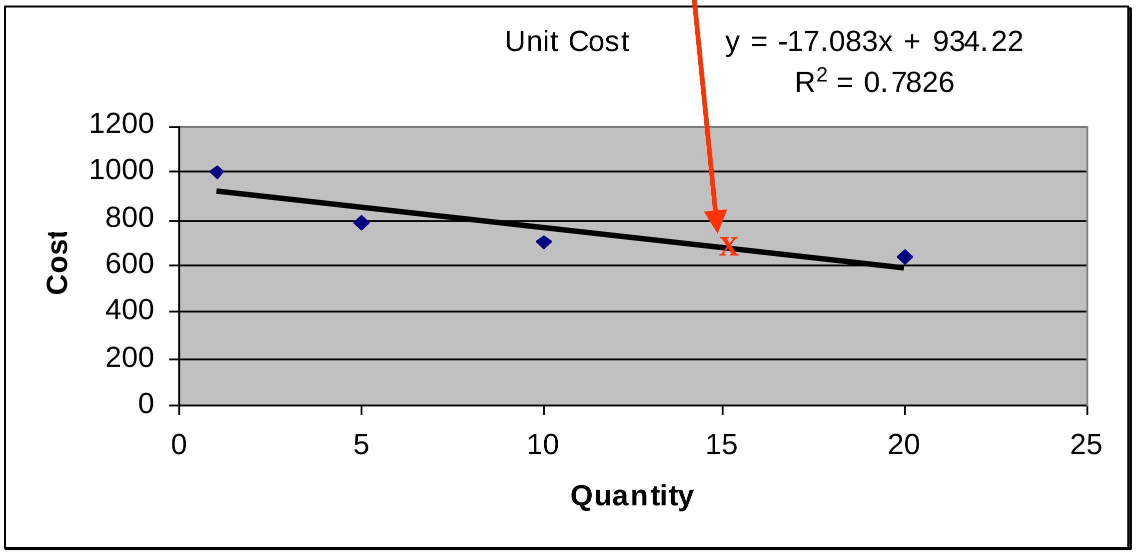
$$b = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

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

Curve Fitting - Least-Squares Bet Fit (LSBF)

- The root of regression analysis
- Fits independent variables to a dependent variable or variables
- Resulting mathematical form may be linear or curvilinear
- Fit minimizes the square of errors between the specific equation and the data (least errors squared yields best fit)

$$-17.083 \times 15 + 934.22 = 678$$



Curve Fitting - Least-Squares Best Fit (LSBF)

ASSUMPTIONS

- Mean value of distribution lies on the LSBF line;
  e.g. (9, 780) in example
- Independent variables are free of measurement error
- Error term (ε) is random and normally distributed about the LSBF line

$$Y = (A + b * X) + \varepsilon$$

- Extrapolation outside the range of historical data can lead to less accurate estimates

Curve Fitting – Least-Squares Bet Fit (LSBF)

Computational Example

	X	Y	X * Y	X ^ 2	Y ^ 2
	1	1000	1000	1	1000000
	5	783	3915	25	613068
	10	705	7047	100	496585
	20	634	12684	400	402234
Sum	***36***	***3122***	***24646***	***526***	***2511888***

Y = *a* + *bX*

$$b = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sum X^2 - n\overline{X}^2}$$

$$\frac{24645 - 4 * (36 / 4) * (3122 / 4)}{526 - 4 * (36 / 4)\text{^}2}$$

b = -17.083

$a = \overline{Y} - b\overline{X}$ = (3122 / 4) - (-17.083) * (36 / 4)

a = 934.22

CER Significance Testing

- How good is the CER?
- How likely is it to estimate well?
- What is the confidence level?
- Is the database sufficiently large?
- Are there few independent variables?
- Is the range of applicability wide?
- Is the prediction error small?
- Quantitative & qualitative measures apply

CER Significance Testing – Validation Measures and Prediction Measures

- Validation Measures
  - R2: Y variation explained by X (higher is better)
  - SE: Avg. estimating error
  - CV: SE ÷ Avg. Y (smaller is better)
  - F-stat: Equation validity test (close to 0)
  - t-stat: X variable validity test (close to 0)
  - Data: Narrative description of amount and quality
  - Logical Relationships: Narrative description of rationale
- Prediction Measures
  - N: Number of observations (many)
  - d.f.: N - # Independent variables (close to N)
  - Outliers: Poorly predicted Y values (none is ideal)
  - Data Range: Large enough to cover expected future observations (complete coverage is not essential)

Appropriate Use of CERs - General

- CERs targeted to specific forecasts invoke more confidence
- Generic CERs have wider applicability, but care must be exercised to:
  - Adapt to different assumptions through incorporation of new data and/or variables
  - Document the adaptation, no matter how slight
- Application domain is client driven
- Confidence in generalized CERs is gained by thorough testing and the wide applicability of the development database and method of construction

Appropriate Use of CERs – Strengths and Weaknesses

- Strengths
  - Excellent predictor if used appropriately
  - Time Saving Method
    - Proposal cycle time
    - Evaluation & negotiation cycle time
  - Easy to apply
  - Functions with minimal information
  - Avoids over-reliance on the strength of a single analogy
- Weaknesses
  - Sometimes too simplistic
  - Strong database dependencies
    - Developer should validate CER and database
    - Applier must validate use; i.e., database applicability, database age, data normalization
    - CER source documentation must be complete

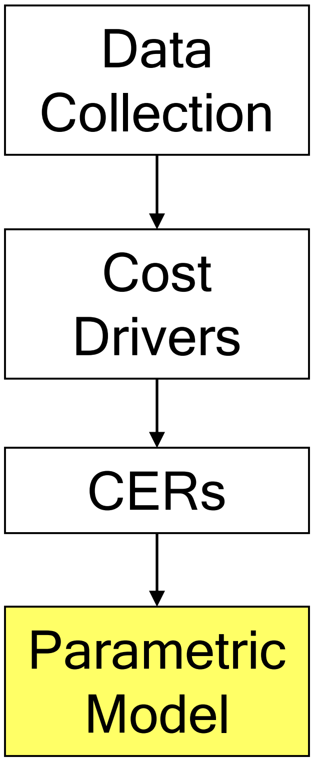
Documenting a CER

CER documentation should include:

- A clear explanation of the types of efforts or products to be estimated by the CER including:
  - Identification, explanation, and rationale for the CER database and CER functional relationships
  - Calculation and description of effort (hours, dollars, etc.) in the pool and base
- Information on when and how to use the CER
- Complete actual cost information for all accounting data used. This provides an audit trail that is necessary to identify the data used
- Noncost information (technical data)

Process - Parametric Model

19



4. Building a parametric model

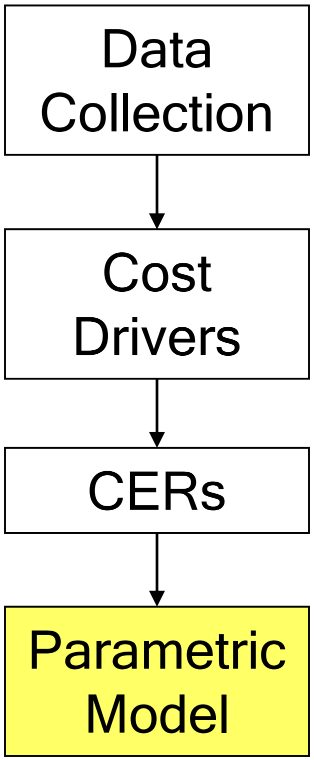
- After CERs are developed, they are gathered into an integrated, automated structure that makes up the cost analyst's model for estimating the system

9

Warning: If you don't properly account for correlation between cost elements, your uncertainty will be understated!

- Updating the model with actuals
  - After the model is built, the analyst can use actuals to update and re-run the CERs in the model with program-specific data
    - 1,200 lines of code per package for COTS software integration based on industry data (previous example)
    - 1,500 lines per package based on actuals (update)

Complex Parametric Models



- Consist of more than a single CER
- May contain a number of different estimating techniques
  - CERs, discrete estimates, etc.
- May incorporate time phasing and inflation
- May provide a range of cost estimates to account for risk and uncertainty
- May incorporate anticipated uses and goals of the analysis into the model logic
  - DTC or CAIV trades, e.g.

NEW!

CERs – Calibration

- Calibration: Resetting the y-intercept so that the CER passes through a desired point
- Reasons to calibrate a CER:
  - To correct an Analogy
  - To make more applicable to a specific subset of data
  - To adjust input factors in a commercial cost model
  - To support CAIV trades more accurately

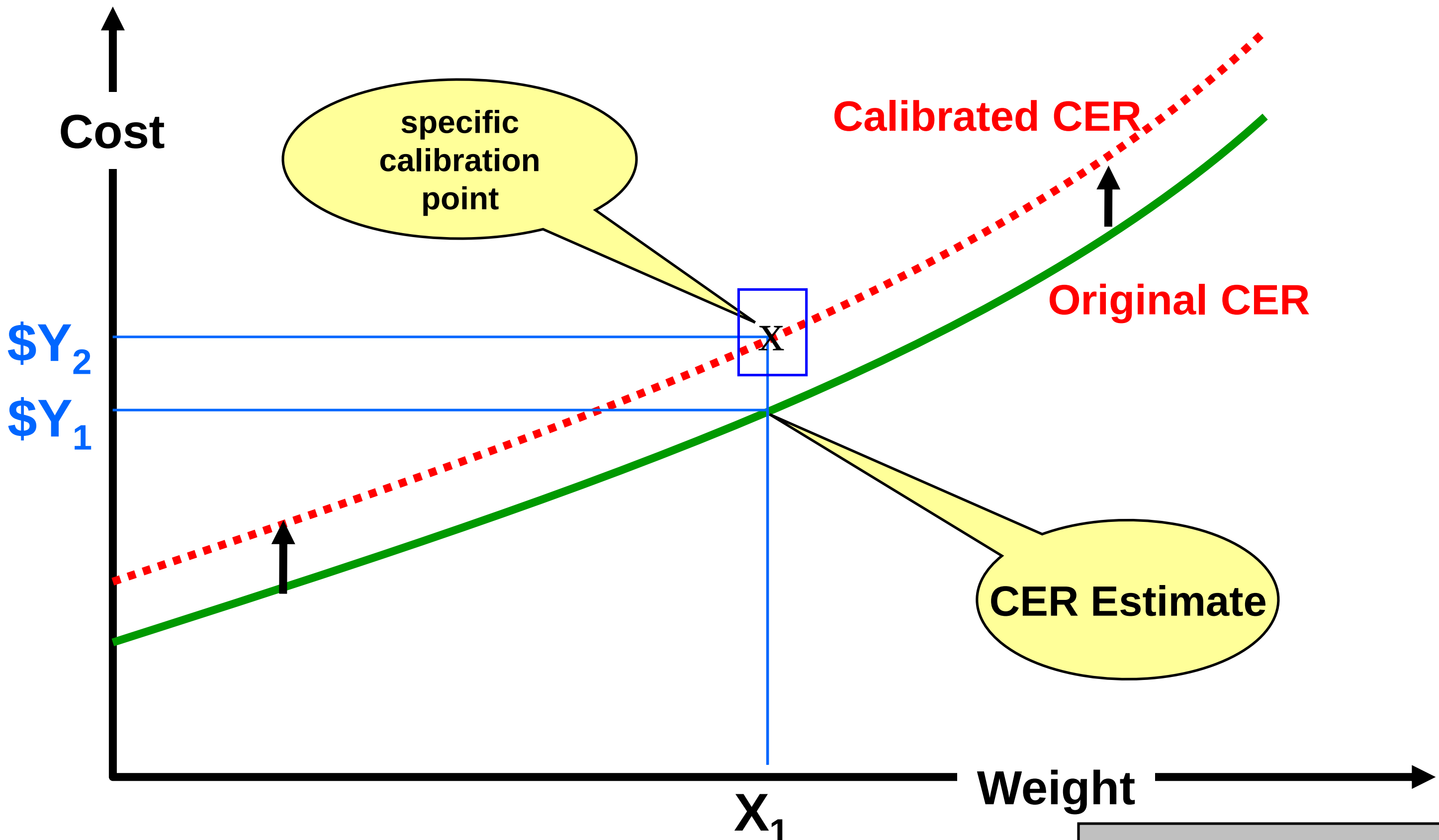
16

Warning: Never calibrate to a single point in the data set on which the CER is based!

Calibration should always be done with care and with a clear purpose in mind

Calibrating Parametric CERs

18



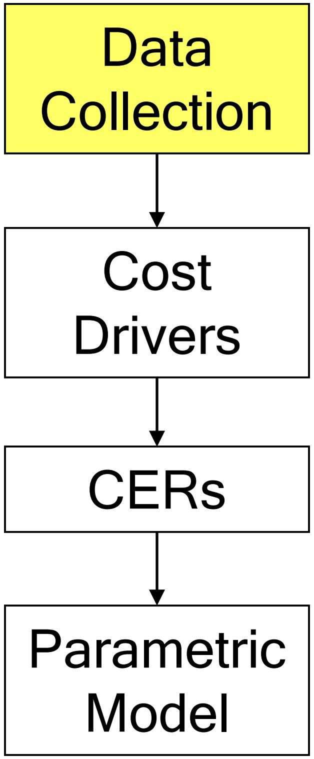
Tip: A calibrated CER (shifted y-intercept) and an adjusted analogy (borrowed slope) are mathematically equivalent!

"Analogies: Techniques for Adjusting Them," R. L. Coleman, J. R. Summerville, S. S. Gupta, SCEA 2004 & MORSS 2004

NEW!

Parametric Process Examples

Example – Data Collection



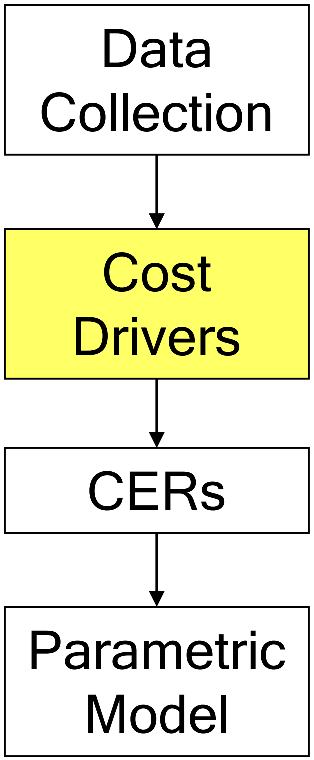
- Example: Site Activation as a function of the number of workstations
- Hypothesis: There exists a cost relationship for Site Activation to be calculated as a function of the quantity of workstations per site.

Independent Variable: X (Cost Driver). Dependent Variable: Y.

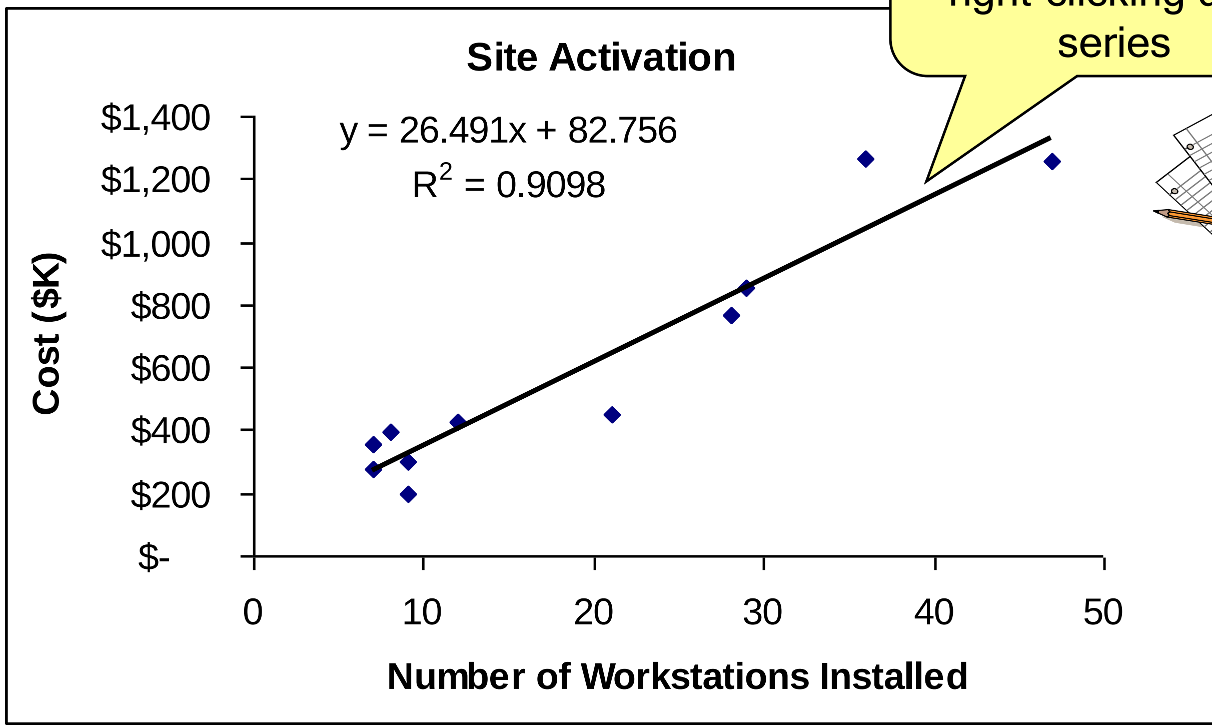
	X (Cost Driver) Number of Workstations	Y Cost ($K)
Site 1	8	$ 398
Site 2	28	$ 769
Site 3	12	$ 430
Site 4	9	$ 199
Site 5	21	$ 447
Site 6	29	$ 851
Site 7	47	$ 1,258
Site 8	7	$ 359
Site 9	36	$ 1,267
Site 10	9	$ 300
Site 11	7	$ 275

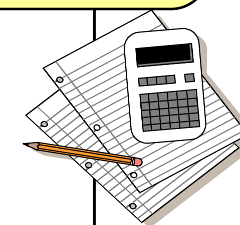
Example - Cost Drivers

- Site Activation Example:



Add Trendline and equation in Excel by right-clicking data series





Tip: Always be sure to scatter plot your data first. Scatter plots can help identify cost drivers.

6

Example - CERs

12

13

14

- Site Activation Example:

Data Collection → Cost Drivers → **CERs** → Parametric Model

$$y = 26.491x + 82.756$$

$$R^2 = 0.9098$$

Equation Parameters

Ratio of explained variation to total variation in the data set

CV = std error/mean
= 121.66/595.73 = 20.4%

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.953833897
R Square	0.909799102
Adjusted R Square	0.89977678
Standard Error	121.6606251
Observations	11

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1343622.413	1343622.413	90.77727729	5.34455E-06
Residual	9	133211.7692	14801.30769		
Total	10	1476834.182			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	82.75641026	65.14834309	1.270276516	0.235841095	-64.61949303	230.1323135	-64.61949303	230.1323135
X Variable 1	26.49145299	2.780463527	9.527711022	5.34455E-06	20.20160271	32.78130327	20.20160271	32.78130327

8

10

Inspect the t and F statistics for significance
If p-value < significance level (α=0.05), then coefficient is significant

9

Example – Updated CER

- Updating and re-running the CER Example for Site Activation

Data Collection → Cost Drivers → CERs → **Parametric Model**

Adding Sites 12 - 15:

	X (Cost Driver) Number of Workstations	Y Cost ($K)
Site 1	8	$ 398
Site 2	28	$ 769
Site 3	12	$ 430
Site 4	9	$ 199
Site 5	21	$ 447
Site 6	29	$ 851
Site 7	47	$ 1,258
Site 8	7	$ 359
Site 9	36	$ 1,267
Site 10	9	$ 300
Site 11	7	$ 275
Site 12	11	$ 488
Site 13	19	$ 755
Site 14	41	$ 1,247
Site 15	19	$ 605

ICEAA CEBoK

Example – Updated CER Graph

– Site Activation Example:

Data Collection

↓

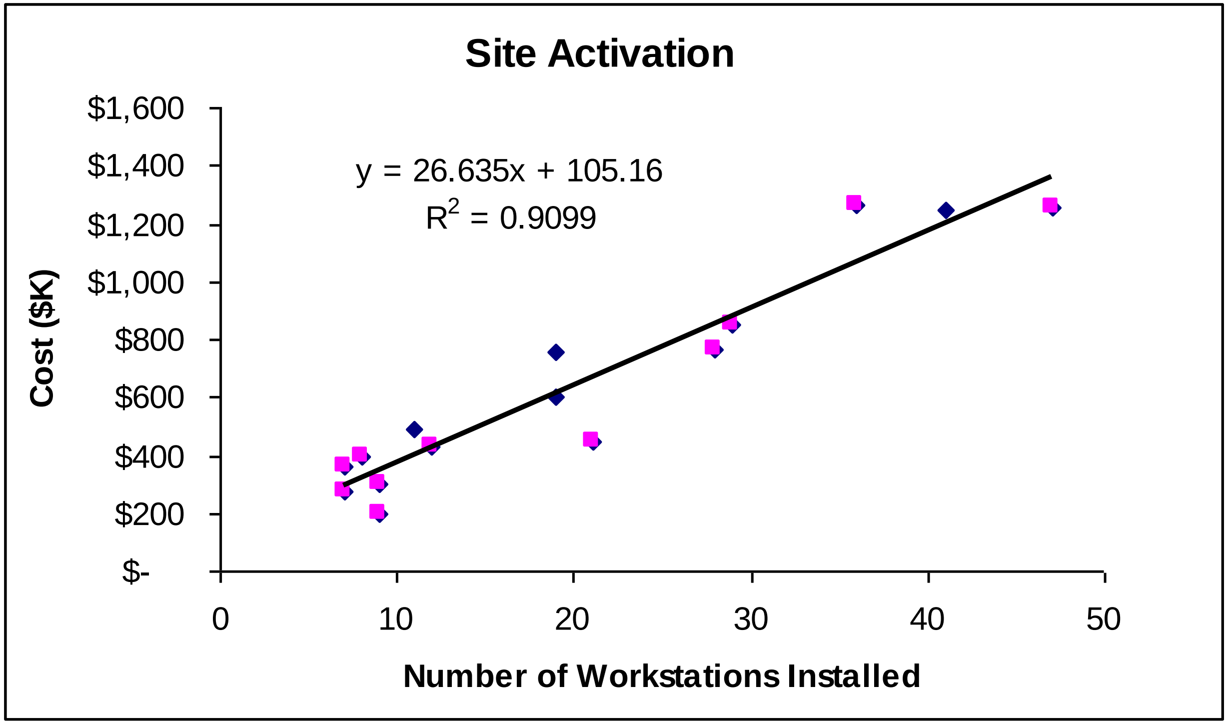
Cost Drivers

↓

CERs

↓

Parametric Model



$y = 26.491x + 82.756$

$R^2 = 0.9098$

→

$y = 26.635x + 105.16$

$R^2 = 0.9099$

Example - Updated CER Regression

- Site Activation Example:

Data Collection → Cost Drivers → CERs → Parametric Model

$$y = 26.635x + 105.16$$

$$R^2 = 0.9099$$

Equation Parameters

Ratio of explained variation to total variation in the data set

CV = std error/mean
= 114.86/643.20 = 17.9%

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.95
R Square	0.91
Adjusted R Square	0.90
Standard Error	114.86
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1732752.60	1732752.60	131.35	0.000000036
Residual	13	171491.80	13191.68		
Total	14	1904244.40			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	105.164101	55.527605	1.893907	0.080704	-14.795974	225.124176	-14.795974	225.124176
X Variable 1	26.635441	2.324029	11.460888	0.000000	21.614681	31.656200	21.614681	31.656200

8 10

Inspect the t and F statistics for significance
If p-value < significance level (α=0.05), then coefficient is significant

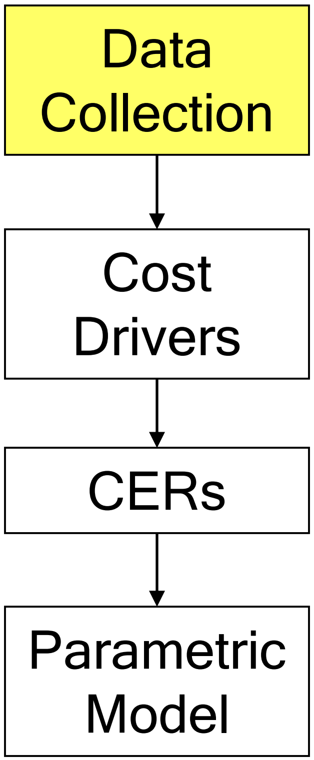
Multivariate Example

- With the site activation example, a single independent variable provided a tight fit to estimate cost
- In some cases, multiple cost drivers (independent variables) are used
  - Dictated by expert opinion, logic, and data analysis
  - Requires multiple regression
  - Brief example follows

15

8

Multivariate Example – Data Collection

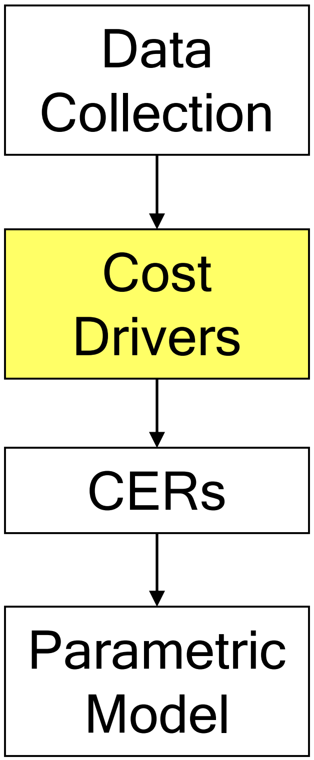


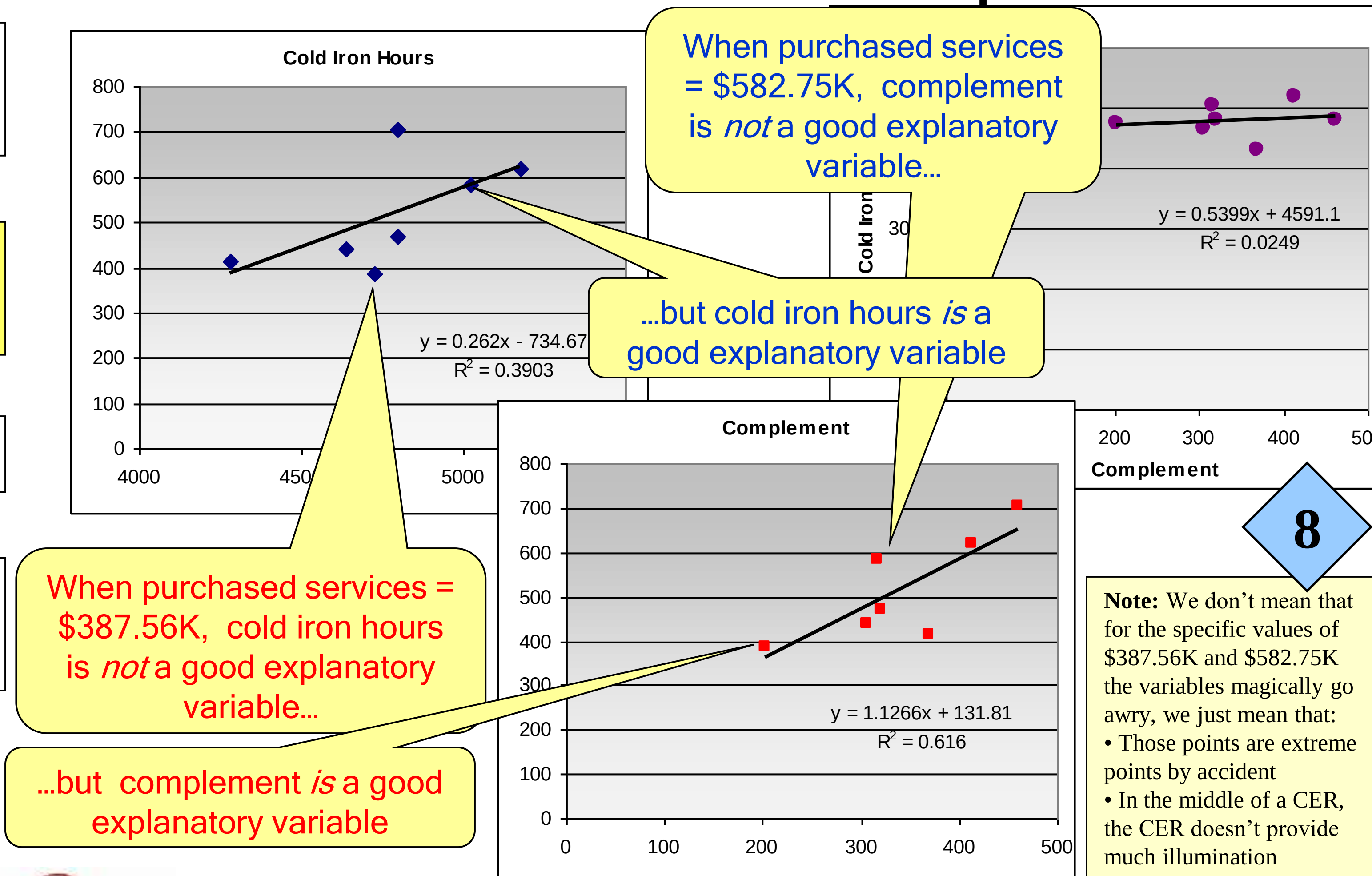
- Data collected with two independent variables:

Multivariate Example Data			
	Purchased Services ($K)	Cold Iron Hours	Comple-ment
Frigate A	387.56	4727	203
Destroyer B	439.66	4635	306
Destroyer C	582.75	5021	317
Destroyer A	469.24	4796	321
Cruiser C	413.00	4280	369
Cruiser A	617.91	5174	413
Cruiser B	705.16	4795	461

Multivariate Example - Cost Drivers

- Purchased Services Example:





Note: We don't mean that for the specific values of $387.56K and $582.75K the variables magically go awry, we just mean that:
- Those points are extreme points by accident
- In the middle of a CER, the CER doesn't provide much illumination

Multivariate Example - CERs

20

Data Collection → Cost Drivers → CERs → Parametric Model

- Purchased Services Example:

$$y = 0.215x_1 + 1.01x_2 - 854.3$$

$$R^2 = 0.8732$$

Equation Parameters

Tip: Intercept does *not* represent fixed cost!

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.934462283
R Square	0.873219758
Adjusted R Square	0.809829638
Standard Error	52.13001587
Observations	7

Ratio of explained variation to total variation in the data set

CV = std error/mean
= 52.13/574.58 = 9.1%

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	74869.97435	37434.98717	13.77532882	0.01607323
Residual	4	10870.15422	2717.538555		
Total	6	85740.12857			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-854.2676929	358.213134	-2.384802822	0.075593456	-1848.828855	140.2934695	-1848.828855	140.2934695
Cold Iron Hours	0.214824888	0.075578629	2.842402553	0.046753766	0.004984539	0.424665237	0.004984539	0.424665237
Complement	1.009958432	0.258711303	3.903804826	0.017485119	0.291659214	1.728257649	0.291659214	1.728257649

8 10

Inspect the t and F statistics for significance:
If p-value < significance level (α=0.05),
then coefficient is significant

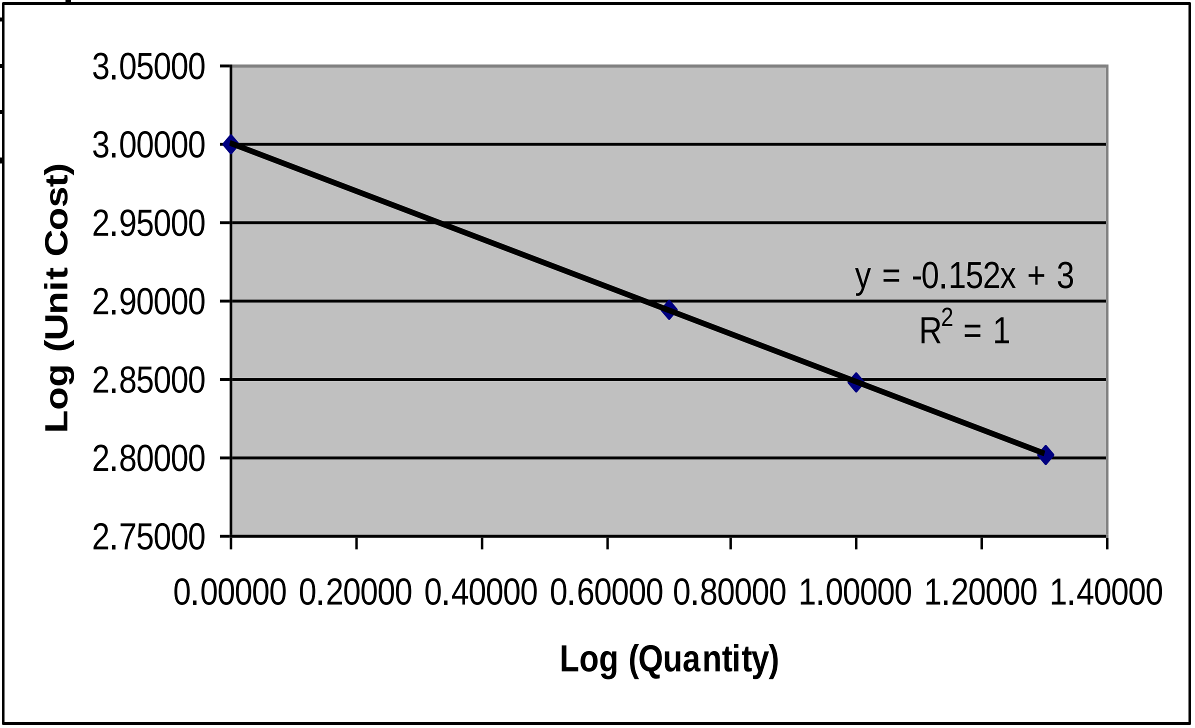
ICEAA CEBoK

Multiple and Curvilinear Regression

- Multiple Regression
  - Two or more independent variables
  - Step-wise regression often used identify appropriate variables.
  - Statistical “F-test” can be employed in decision to add variables.
  - Multicollinearity Problem
    - Independent variables may not really be so
    - Regression analysis doesn’t identify this
    - Must rely on pair-wise correlation to solve
- Curvilinear Regression
  - Transformations used for exponential, logarithmic, and general power series functions
  - e.g. Y = A * Xb can be transformed to
    - Y´ = A´ + b x X´ ;
    - where
    - Y´ = Log (Y); A´ = Log (A); & X´ = Log (X)

Curvilinear Regression

Log (Unit Cost)	Log (Quantity)
3.00000	0.00000
2.89375	0.69897
2.84800	1.00000
2.80224	1.30103



Regressed Equation: $\text{Log}(Y) = 3 - .152 * \text{Log}(X)$

or $Y = 1000 * X^{-.152}$

Other CER Examples

Cost Improvement Curve

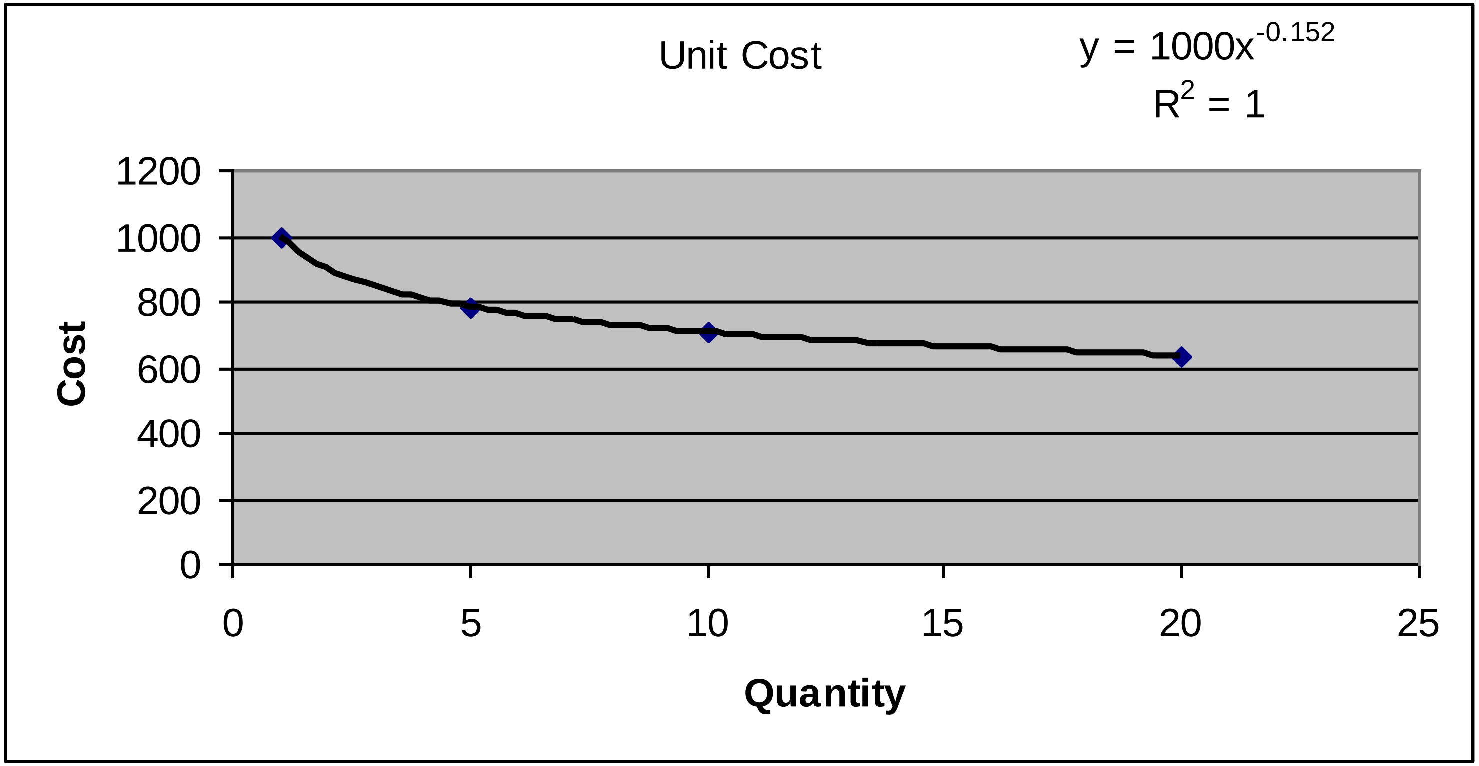
- Also known as the Learning Curve
- Mathematical form: Y = A * X b
  - Y is either unit cost or average unit cost
  - A is first unit cost (T1)
  - X is unit number
  - b is improvement or learning factor
- Learning Curve Interpretation
  - Constant cost improvement with each doubling of quantity (e.g. 90% LC = 10% improvement with every doubling of quantity)

    $b = \ln(LC) \div \ln(2)$

Cost Improvement Curve

- Unit Learning Curve: Equation describes unit cost behavior
  - Unit Cost = T1 * Unit # b
- Cumulative Average Learning Curve: Equation describes average cost behavior
  - AUC = T1 * Cum Qty b
- For a fixed data set, the difference is the T1

Cost Improvement Curve



$-.152 = \ln(LC) \div \ln(2); \ln(LC) = -.152 * .692 = -.105$

$\therefore$ LC = $e^{-.105}$ = .90 or a 90% Learning Curve

ICEAA CEBoK

Electronics and Software

- Electronics
  - Variables & Constants
    - X: Number of ICs
    - Y: Electronic Item Unit Production Cost
    - A: Set-up cost ($57)
    - b: Recurring Unit Cost per IC ($1.10)
  - CER: $Y = A + b \times X$
  - Example: for a 30 IC unit,
    - $Y = \$57 + \$1.10 \times 30 = \$90$

Construction/Reconstruction

- Variables & Constants
  - X1: Area (square feet)
  - a: area coefficient ($75 per square foot)
  - X2: # Rooms
  - b: rooms coefficient ($5,000 per room)
  - C: Home type (1 for tract, 1.33 for semi-custom, 1.5 for custom)
  - Y: Construction Cost
- CER: Y = C x (a x X1 + b x X2)
- Example: A 12 room, 2500 square foot, semi-custom built home,
- 1.5 x ($75 x 2500 + $5000 x 12) = $371,250

Software

- Variables & Constants
  - X: Thousands of Delivered Source Instructions
  - A: Type coefficient; 3.2, 3.0, or 2.8
  - b: Type Exponent; 1.05, 1.12, or 1.2
  - Y: Labor months of development effort
- CER: Y = A x Xb
- Example: For a 40,000 DSI, Type 3 program
- 2.8 x 401.2 = 234.2 labor months

CER Validation

Evaluating CERs - Criteria

- Logical data relationships
- Uses adequate data
- Evidence of strong relationships
- Consistent and valid application
- Cost drivers (variables) make logical sense
- Driving data should be accessible for both CER development and implementation
- Driver list should be comprehensive
- Excluded outliers and rationale for exclusion should be identified
- Meets customer needs

Evaluating CERs – Criteria

- Validation
  - Establishes the CER as a credible estimating method
  - Validation = Documented Consistent Accuracy
    - Requires periodic testing
    - CER updated when testing indicates need
  - Validation falls to CER developer
  - Certification falls to CER applier

Evaluating CERs – Credible Data

- **Credible Data**
  - Assembling a quality database is the single most important and time consuming step
  - Database must be encompassing but not overwhelming
  - Historical data is preferred
  - Data normalization must be consistent
  - Formats for data types should be consistent (Cost, Technical, Programmatic)
  - Database must be fed and cared for
- **Strength of Data Relationships**
  - Can be judged with Statistical tests
  - R2 and Standard Error are most popular strength tests
    - Complementary
    - R2 measures the degree to which the independent variables describe variation in the historical data
    - Standard error measures average error |predicted - actual| in the historical data

Parametric Estimating Summary

- Parametric Estimating uses cost estimating relationships (CERs) based on historical data to predict cost
- Until actual cost data are available, the use of parametric costing techniques is the preferred approach
- Parametric estimating process
  - Collecting Data
  - Identifying Cost Drivers
  - Developing CERs
  - Building a parametric model
- OTS Cost Models are convenient to use but must be applied with caution

4

6

8

Resources

- DoD 5000.4-M, Department of Defense Manual Cost Analysis Guidance and Procedures, December 1992
  - http://www.dtic.mil/whs/directives/corres/pdf/500004m.pdf
- International Society of Parametric Analysts (ISPA), Parametric Estimating Handbook, 4th Edition, April 2008
  - http://www.iceaaonline.com/ready/wp-content/uploads/2014/03/2008-Parametric-Estimating-Handbook-4th-Edition.pdf
- "Cost Response Curves - Their generation, their use in IPTs, Analyses of Alternatives, and Budgets," K.J. Allison, K.E. Crum, R.L. Coleman, R.G. Klion, DoDCAS, 1996
- "Analogies: Techniques for Adjusting Them," R. L. Coleman, J. R. Summerville, S. S. Gupta, SCEA 2004 & MORSS 2004