NORTHROP GRUMMAN

DEFINING THE FUTURE

CUMAV or Unit?

Is Cum Average vs. Unit Theory a Fair Fight?

SCEA National Conference
June 2008

Bethia L. Cullis, Richard L. Coleman
Peter J. Braxton, Jacquelyn T. McQueston

Northrop Grumman Corporation

Outline

- Introduction
- Background
- Investigation
  - Perfect Data
  - Imperfect Data
- Results
- Conclusions

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Introduction

- Two learning theories are commonly used in practice
  - Cumulative Average Theory
  - Unit Theory
- Parameters for each typically investigated
  - Slope
  - T1
  - Position on the Curve
- Less attention is paid to selection of theory
  - Statistics of the regression are used to judge appropriate theory
  - The statistics are inherently biased towards Cum Ave

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Background- CUMAV and Unit Theories

NORTHROP GRUMMAN

- Cumulative Average Theory (CUMAV)
  - Introduced in 1936 by T. P. Wright ("Factors Affecting the Cost of Airplanes")
  - Cumulative average cost of production decreases at a constant rate when production volume doubles
- Unit Theory
  - Also called Crawford Curves
  - Unit cost of production decreases at a constant rate when production volume doubles

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Presented at the 2008 SCEA-ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

Investigation- Comparing CUMAV and Unit

NORTHROP GRUMMAN

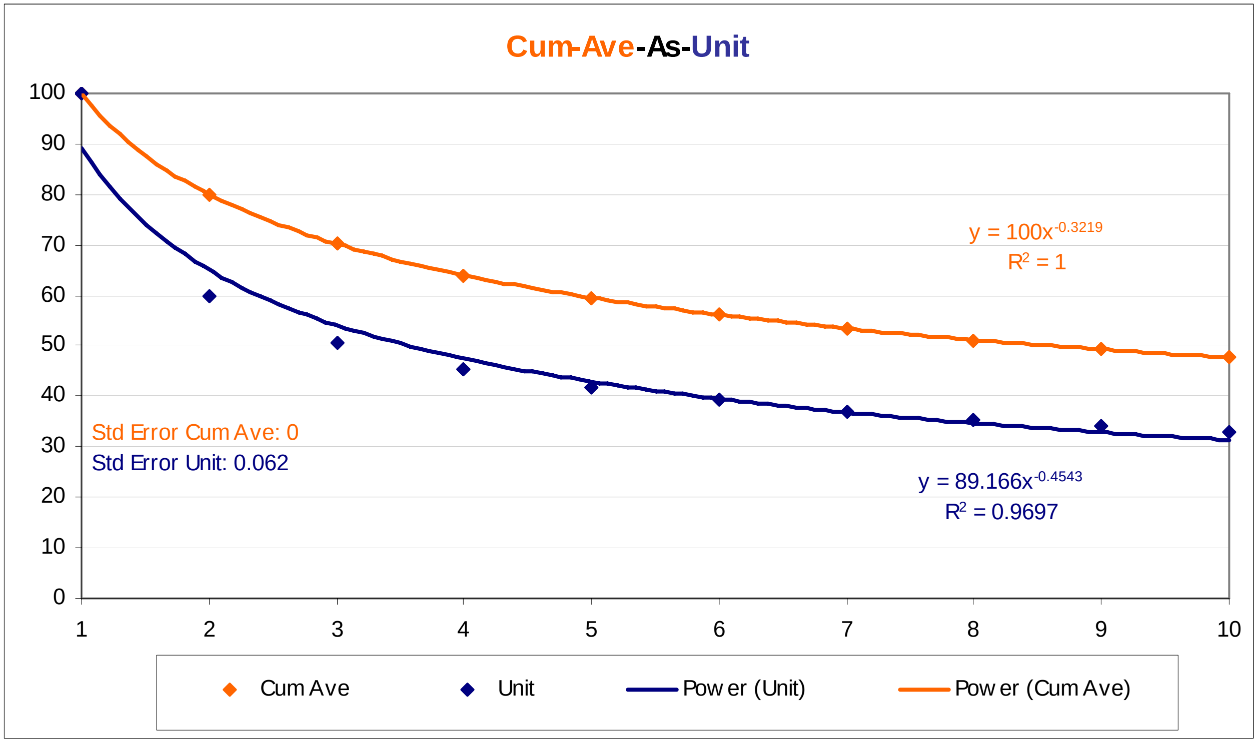
- Data for two perfect curves was generated
  - Parameters:
    - Theoretical First Unit: 100
    - Learning Curve Slope: 80%
  - This allows the advantage of knowing the underlying theory before examining the statistics.
- Data were then regressed twice
  - First, assuming the known underlying theory (i.e., CUMAV-as-CUMAV and Unit-as-Unit)
  - Then assuming the other theory (i.e. CUMAV-as-Unit and Unit-as-CUMAV)
- Statistics of the regression (standard error) were compared to determine which theory best fits the data set
- Perfect data were examined first, followed by data with injected error

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Presented at the 2008 SCEA-ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

CUMAV Data without Error

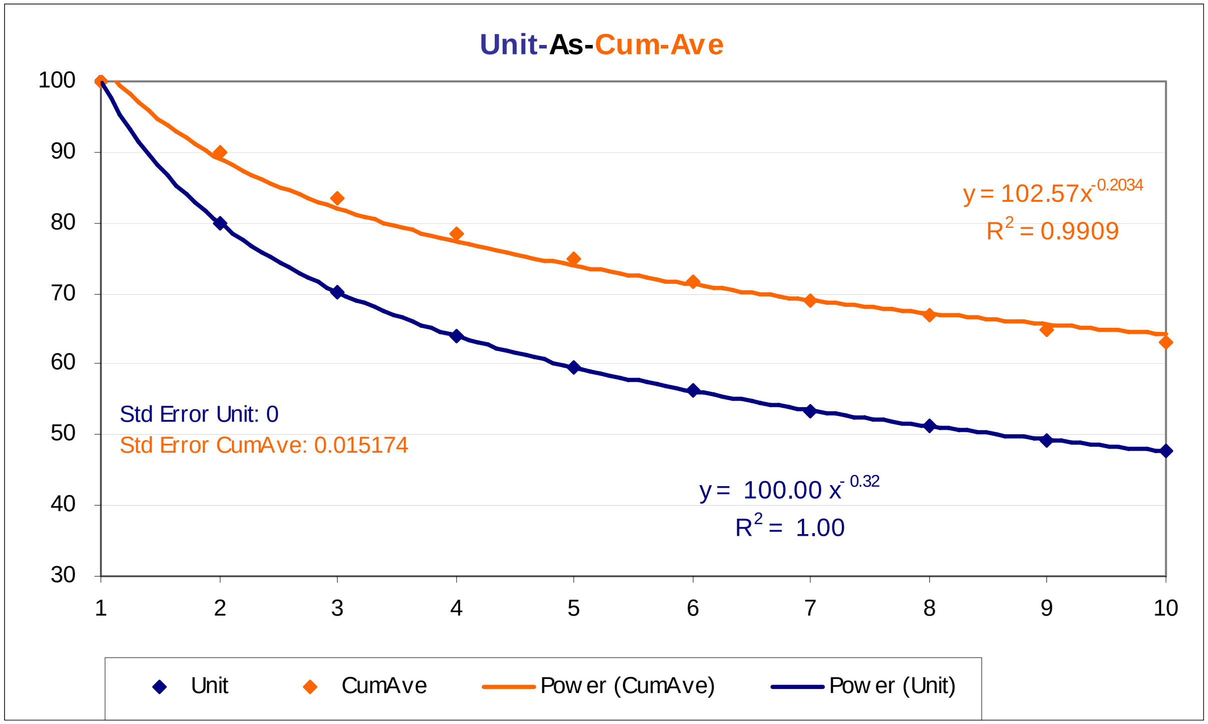
- Plotted error free CUMAV data
  - Fitted using the CUMAV Model
  - Fitted using the Unit Model
- Results
  - As expected, CUMAV Model fits perfectly
  - Unit Model fits with
    - R2: 0.97
    - Standard Error: 0.052



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Unit Data Without Error

- Plotted error free CUMAV data
  - Fitted using the Unit Model
  - Fitted using the CUMAV Model
- Results
  - As expected, Unit Model fits perfectly
  - CUMAV Model fits with
    - R2: 0.99
    - Standard Error: 0.015174



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Model with Error

- With perfect data, statistics reliably direct the analyst to the correct model
- In practice, data is never perfect
  - To generate imperfect data that has the characteristics of real data, random error was injected into each model;
  - The initial size of the random error was based on observed error
    - In a learning study3 produced previously, a coefficient of variation (CV) at the 8th Unit (of a 17 Unit series) of 0.0187 was observed
- Using this information, the curves with error were generated
  - Unit
    - A perfect curve was generated using the same parameters as before:
      - T1 = 100
      - LC = 80%
    - 17 units of data were projected
    - An random normal error term was generated for each unit with a
      - Mean of 0
      - Standard deviation computed using the CV of 0.0187 multiplied by the cost of the eighth "perfect" unit
    - The error term was added to each "perfect" unit cost
    - 500 imperfect data sets were found (with a new random error term set for each run)

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Presented at the 2008 SCEA-ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

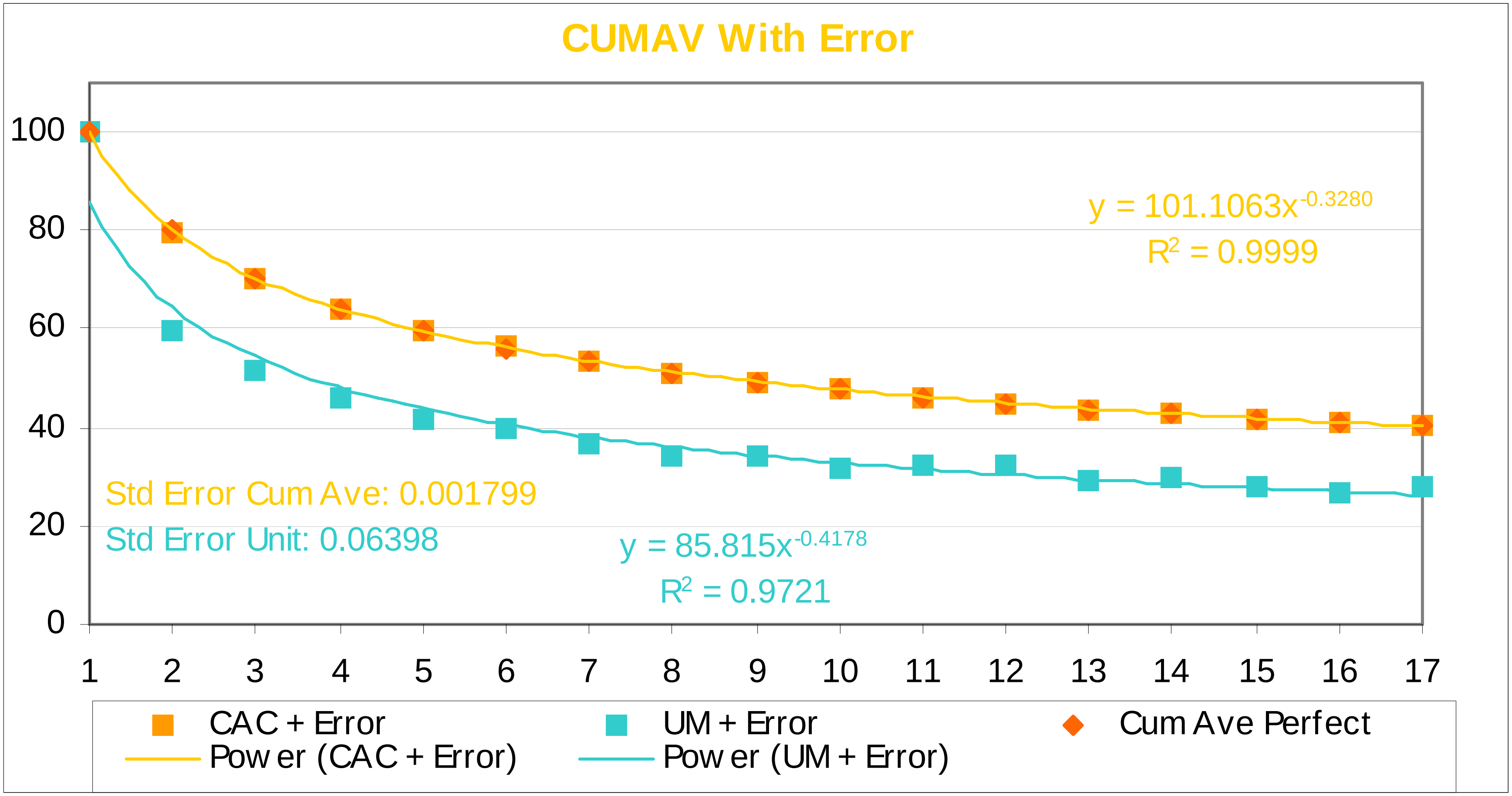
Model with Error Continued...

- CUMAV
  - A perfect curve was generated using the same parameters as before:
    - T1 = 100
    - LC = 80%
  - 17 cumulative average data points were projected.
  - The unit metric for each point was found- for example:
    - If X=1, Y=100 (cumulative average cost)
    - If X=2, Y=80 (cumulative average cost)
    - (100 + 80)/2 = 60 (unit metric of unit 2)
  - The error term was generated as it was in the Unit Curve with error.
  - The error term was added to each unit metric derived from the “perfect” CUMAV data
  - 500 imperfect data sets were found (with a new random error term set for each run)

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

CUMAV with Error: Sample Output

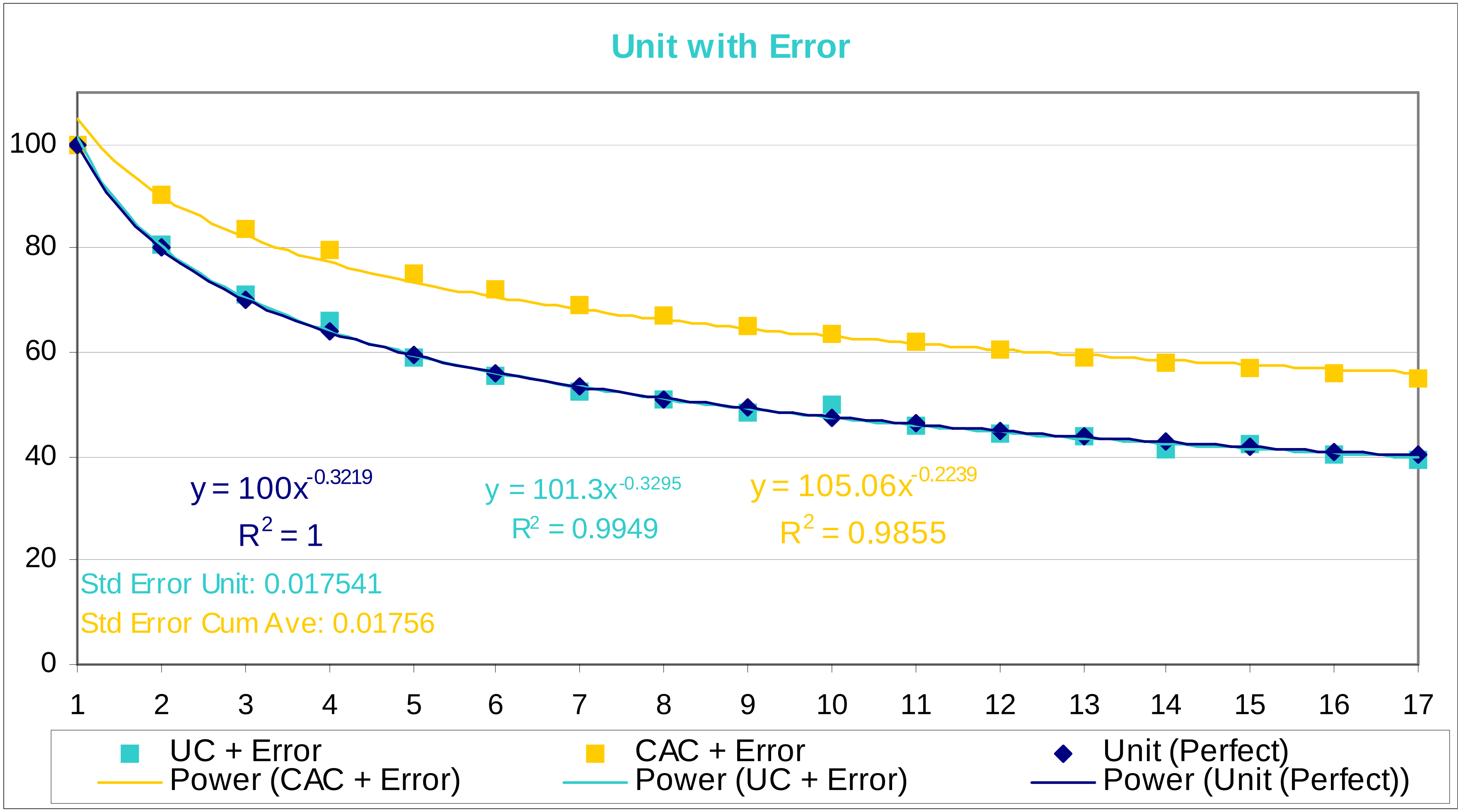
- One of the 500 runs is shown below
- The error in the CUMAV curve is minimal



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Unit with Error: Sample Output

- One of the 500 runs is shown below
- The error in the Unit curve is minimal



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Presented at the 2008 SCEA-ISPA Joint Annual Conference and Training Workshop - www.iceaaonline.com

CUMAV and Unit with Error

- 500 runs of each trial yields the following on average

CV: .0187	Average SE	# of "false positives"	% of "false positives"
Unit-as-Unit	0.018963337		
Unit-as-Cum-Ave	0.017925980	295	59%
Cum-Ave-as-Cum-Ave	0.002934640		
Cum-Ave-as-Unit	0.061955773	0	0%

- From the above, with this error term, an analyst should never mistake CUMAV data for Unit data
- 59% of the time the standard error indicates that CUMAV is a better fit for Unit data than Unit Theory is
  - On average, the standard error for CUMAV Theory as applied to Unit data is lower than the standard error of Unit Theory applied to the same data

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

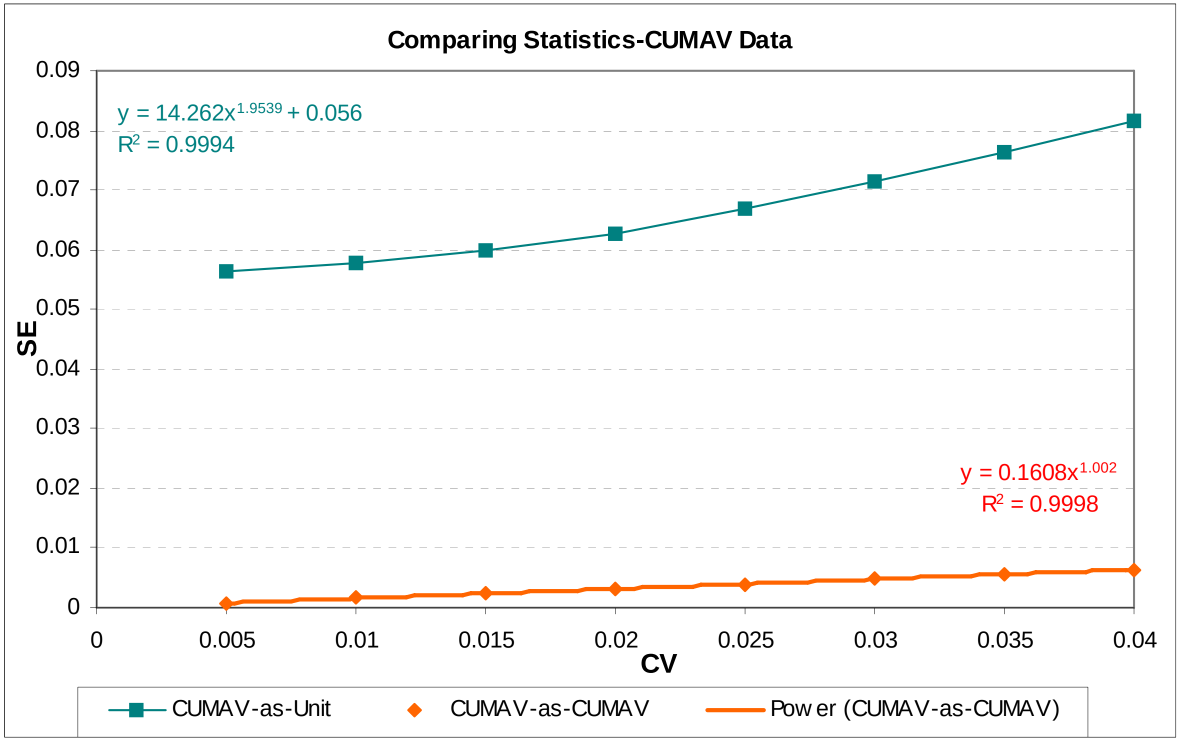
CUMAV and Unit with Error continued...

- Error term injected to generate imperfect data was relatively low compared to residual error in "real" data
- In order to determine how low the CV would have to be in Unit data in order to test as Unit data (and not CUMAV), the above process was repeated with an error term ranging from
  - a standard deviation of 0.005 X 8th unit of cost up to
  - a standard deviation of 0.04 X 8th unit of cost

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Average Standard Error

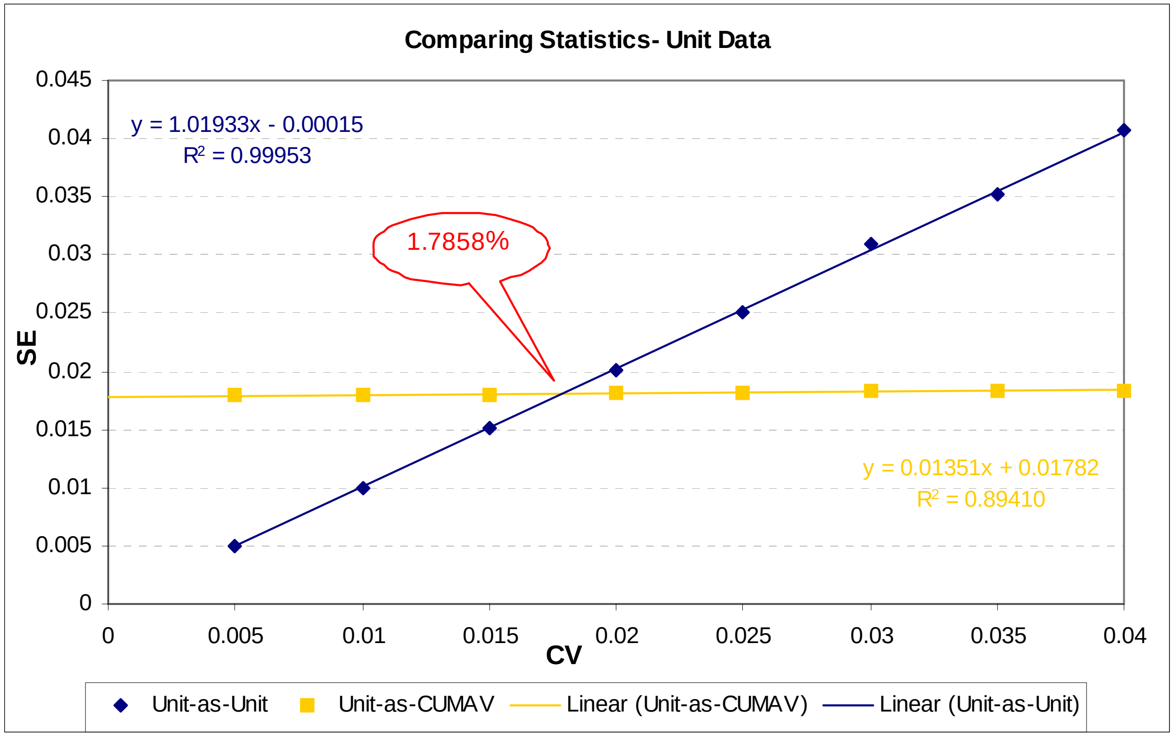
- Plotted average Standard Error of CUMAV data
  - Fitted using the CUMAV Model
  - Fitted using the Unit Model
- Results
  - CUMAV data *never tests* as though it followed Unit Theory



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Average Standard Error Continued...

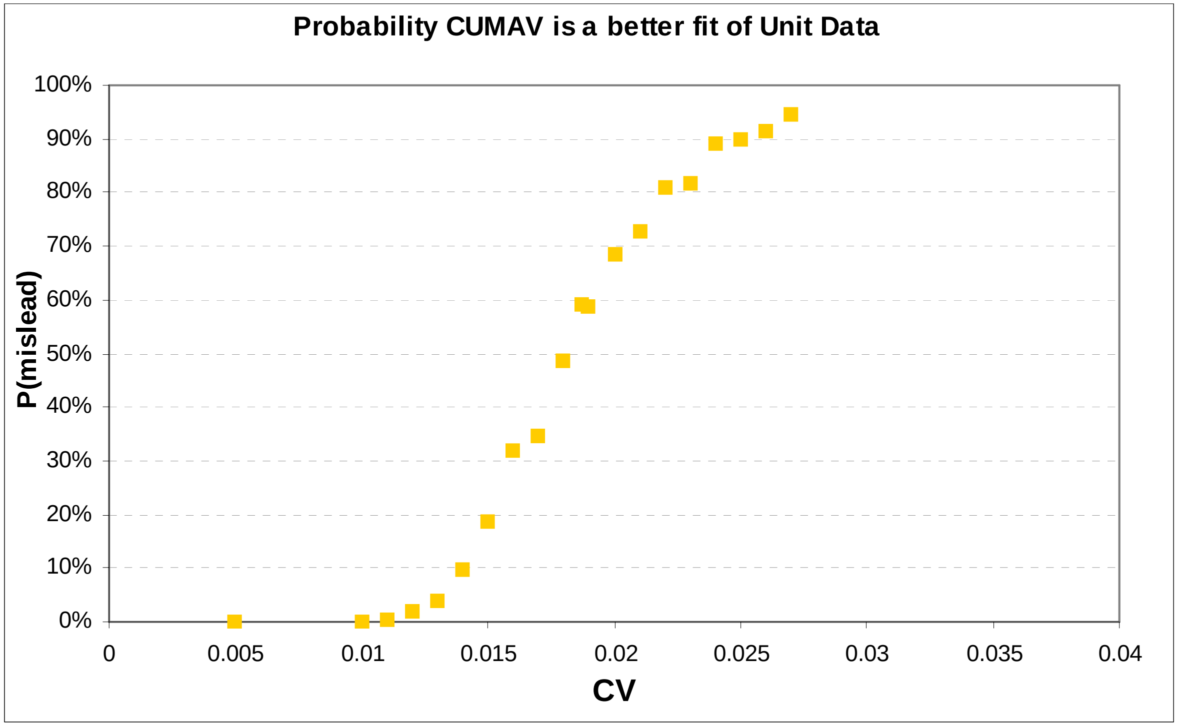
- Plotted average Standard Error of Unit data when
  - Fitted using the Unit Model
  - Fitted using the CUMAV Model
- Results
  - When the underlying data have a CV of 1.786% or greater, the CUMAV Theory returns a lower standard error than Unit Theory



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Probability of Being Mislead

- Plotted the % of "false positives" when fitting CUMAV to Unit data for each set of trials against the CV
  - The probability that the analyst will be mislead by statistics increases as the CV increases
- When the CV of the underlying data is ~0.018, the probability that an analyst will be mislead by the statistics is 50%
- If the CV is ~0.03, that probability increases to 98%



R. L. Coleman, richard.coleman@ngc.com, 703-615-4482

Conclusions

- Analysis in this paper is clearly summarized in the preceding graphs
- Using statistics, an analyst should never accidentally select Unit Theory when the underlying data follow a Cumulative Average curve.
- When the underlying data have a CV of 1.786% or greater, however, applying CUMAV Theory to Unit data returns, on average, a lower standard error than does Unit Theory
- This is problematic, particularly since there is no better way to determine which theory the data truly follow

R. L. Coleman, richard.coleman@ngc.com, 703-615-4482