Learning Curve Analysis of Small Data Sets Spacecraft Bus Cost Improvement Analysis

Brian Welsh
James York

Sponsors:
Greg Hogan, AFCAA
Bruce Thompson, SMC

ICEAA 2013 Annual Professional Development & Training Workshop
New Orleans, Louisiana 18-21 June 2013

Technomics
The Science of Informed Decision Making

Presentation Outline

- Task Summary
- USCM Database
  - Data used for analysis
  - Data Summary
- Analysis Methodology
  - Scaling the data
  - Dependent Variables Analyzed
  - Independent Variables; Regression equation
- Results
- Summary

Task Summary

- Develop cost improvement curves at the satellite level using the Unmanned Spacecraft Cost Model (USCM) data set.
- Where possible, decompose cost improvement into both a learning and rate component.
- Identify production breaks and how they may impact cost improvement curves.

Summary of USCM DB Programs with Multiple Data Points

- Filtered USCM data for ~63 Bus and ~15 Payload Sensor Data Packages (DP) to:
  - 1 Program with 3 data points: Payload Program 7
  - 7 Programs with 2 data points: Program 1, Program 2, Program 3, Program 4, Program 5, Program 6, Payload Program 2
    - Program 3 and Program 2 had production breaks
    - Program 5 and Program 4 had production breaks within a data package
  - Payload Program 2 and Payload Program 7 Sensor are payload sensors

Bus Programs	# data packages/ data points	Data points = # units	Total Units
Program 1	2	11	11
Program 2	2	9	9
Program 3	2	8	8
Program 4	2	10	10
Program 5	2	7	7
Program 6	2	10	10

Payload Programs	# data packages/ data points	Data points = # units	Total Units
Payload Program 2	2	9	9
Payload Program 7	3	3	4

USCM Database Summary

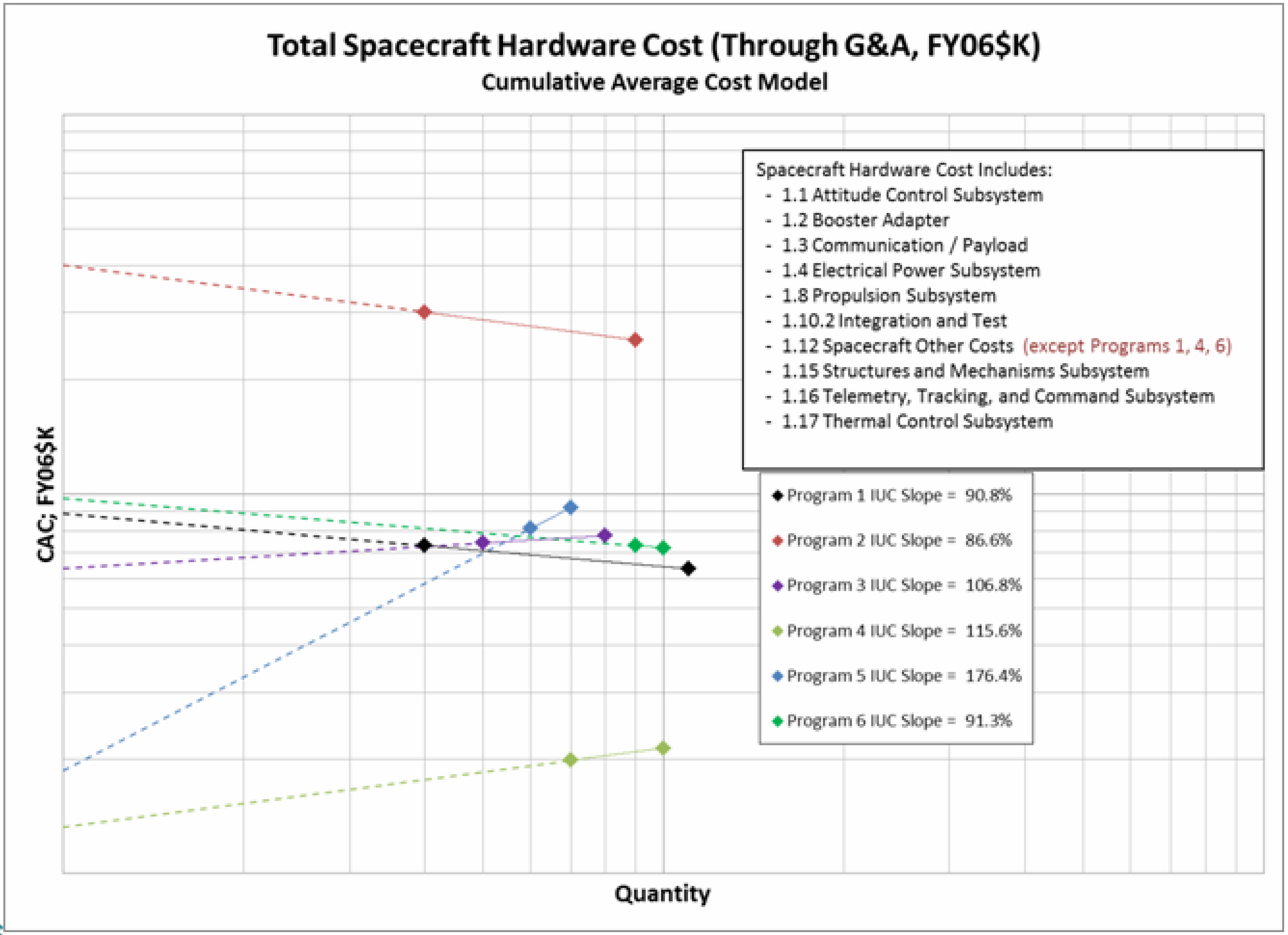
- USCM Cost Data Package Composition
  - USCM compiles cost for quantity groups based on various factors
    - Production block quantity
    - Grouping of contract/option quantities,
  - The cost data package quantities reported in USCM are not necessarily production lot quantities
  - Thus, the data package quantities cannot be used as annual production quantities/lot sizes or as a proxy for production rate

USCM Database Summary

- Cost breakdown
  - Analyzed cost (i.e., DP average (Individual) Unit Cost (IUC) and Cumulative Average Cost (CAC)) for total bus recurring hardware.
  - Not enough degrees of freedom to analyze sensors
  - Used cost data at WBS level 2 to make cost and weight adjustments
- Weight breakdown
  - Used weight as an independent variable
  - Used weight data at WBS level 2 to make cost and weight adjustments
- Programmatic data
  - Used extensively to compile data sets appropriately as to:
    - Bus recurring cost and weight
    - Bus configuration by data package
    - Production rate
    - Unit sequence of data package quantities
    - Delivery dates and launch dates to determine production breaks

Graphical View of USCM Bus Data Set

- Graphic depicts cumulative average learning curve plots for the six bus programs with multiple data points
  - High variability: Learning slopes from 87% to 176%
  - Individual unit learning also analyzed; naturally with higher variability



Methodology – Scaling the Data Set for Regression Analysis

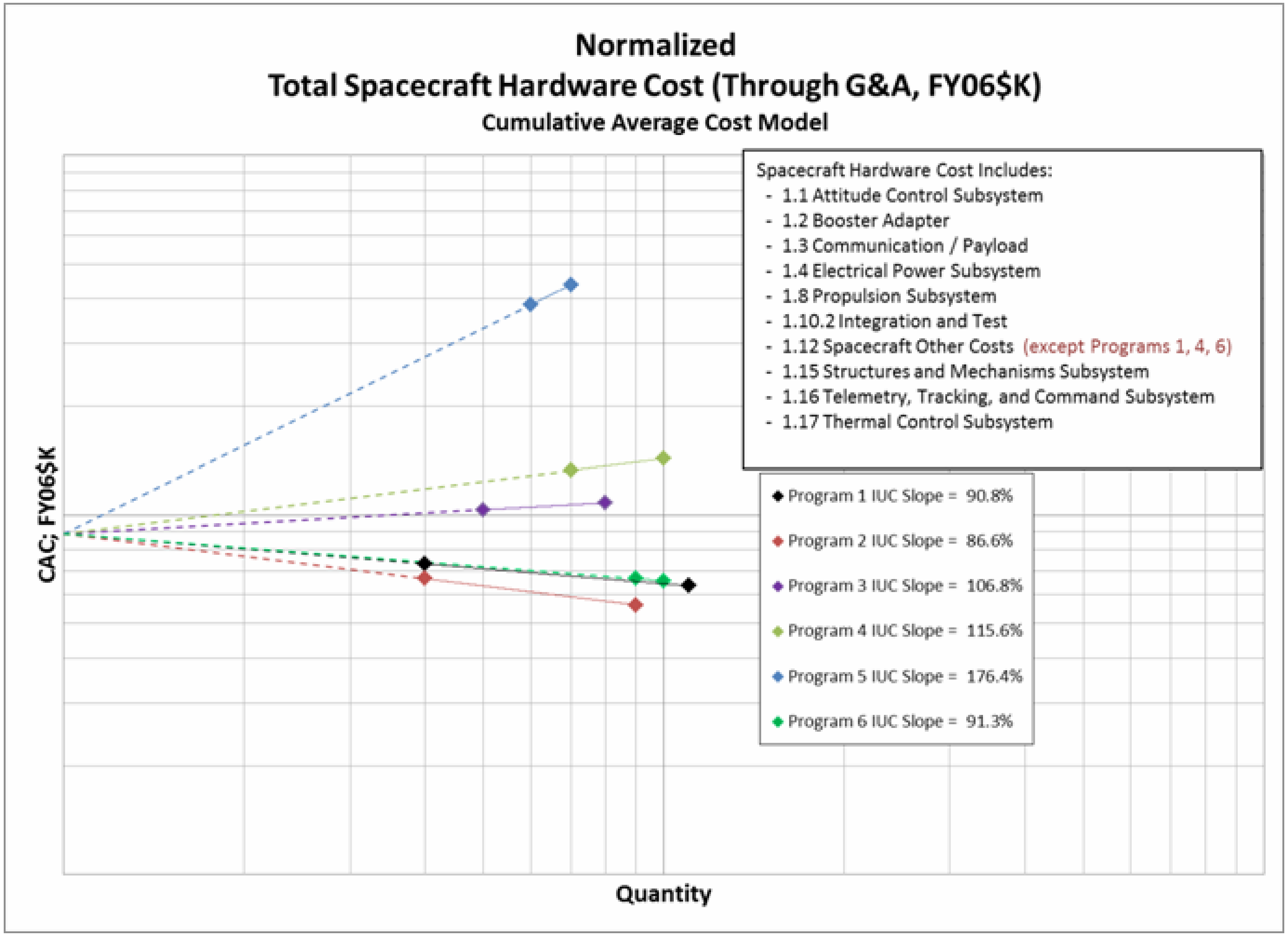
1. Used weight to scale program unit cost
   - Allows regression of all data points simultaneously
   - Provides an estimator of unit cost or CAC
   - Provides estimators for learning slope, rate, etc.
2. Normalized program unit cost to a baseline T1
   - Allows regression of all data points simultaneously
   - Provides estimators for learning slope, rate, etc.
3. Use program dummy variables to scale program unit cost
   - Allows regression of all data points simultaneously
   - Designates program data points as being correlated, thereby potentially reducing standard error
   - Provides estimators for learning slope, rate, etc.

Methodology – Dependent Variables Analyzed

- Cost impacts were analyzed using two cost model frameworks
  - Cumulative Average Cost (CAC) (Cum Cost divided by Cum Quantity)
  - Individual Unit Cost (IUC) (Average Lot Cost assigned to Lot Midpoint)
- CAC, IUC defined by USCM Cost Data Package Composition
  - USCM compiles spacecraft program cost for quantity groups based on contract/option quantities
  - The cost data package quantities reported in USCM are not necessarily production lot quantities
- CAC, IUC defined as recurring cost of Bus hardware WBS elements
  - Hardware WBS elements (e.g., structure, propulsion, etc.) plus I&T
  - SE, PM and support elements excluded
  - CAC, IUC defined as USCM FY06$ through G&A

Graphical View of USCM Data Set - Normalized

- Graphic depicts cumulative average learning curve plots for the six bus programs, normalized to a baseline T1
  - Provides a better perspective on variability of data



Methodology – Independent Variables Tested

- Midpoint or Cum Quantity
  - Data package midpoint (unit theory) or cum quantity (CAC theory)
  - Provides coefficient for learning slope
- Production Rate – annual rate from delivery schedule
- Production Break – a dummy variable to capture impact of break
  - Based on timeline analysis of deliveries and programmatic data
- Weight by Data Package – used to scale programs by cost, and capture impact of weight changes
- Weight ratio by Data Package (DP Wt/1st DP Wt) – used with normalized data set to capture impact of weight changes
- Last lot – a dummy variable to capture impact of last lot
- Lot size – production lot size to model annual production rate or economic order quantity impacts

Methodology – Regression Equation

- Cost (FY06$) = T_1 (Q^b) ~~(PR^c)~~ ~~(LS^s)~~ (PB^p) (LL^a) ~~(Wt^w)~~ (D_1^{d1}) . . . (D_n^{dn})
  - Cost = DP unit cost or DP CAC
  - T_1 = constant
  - Q = DP cumulative quantity or DP midpoint
  - PR = annual production rate
  - LS = DP Quantity
  - PB = a dummy variable; e if production break precedes DP, 1 if not
  - LL = a dummy variable; e if Last Lot, 1 if not Last Lot
  - Wt = DP configuration weight for non-normalized data sets; weight ratio of DP weight divided by 1st DP weight
  - D_1 . . . D_n = dummy variables; one per program except for a baseline program – dummy variable used to scale data, coefficients are not used

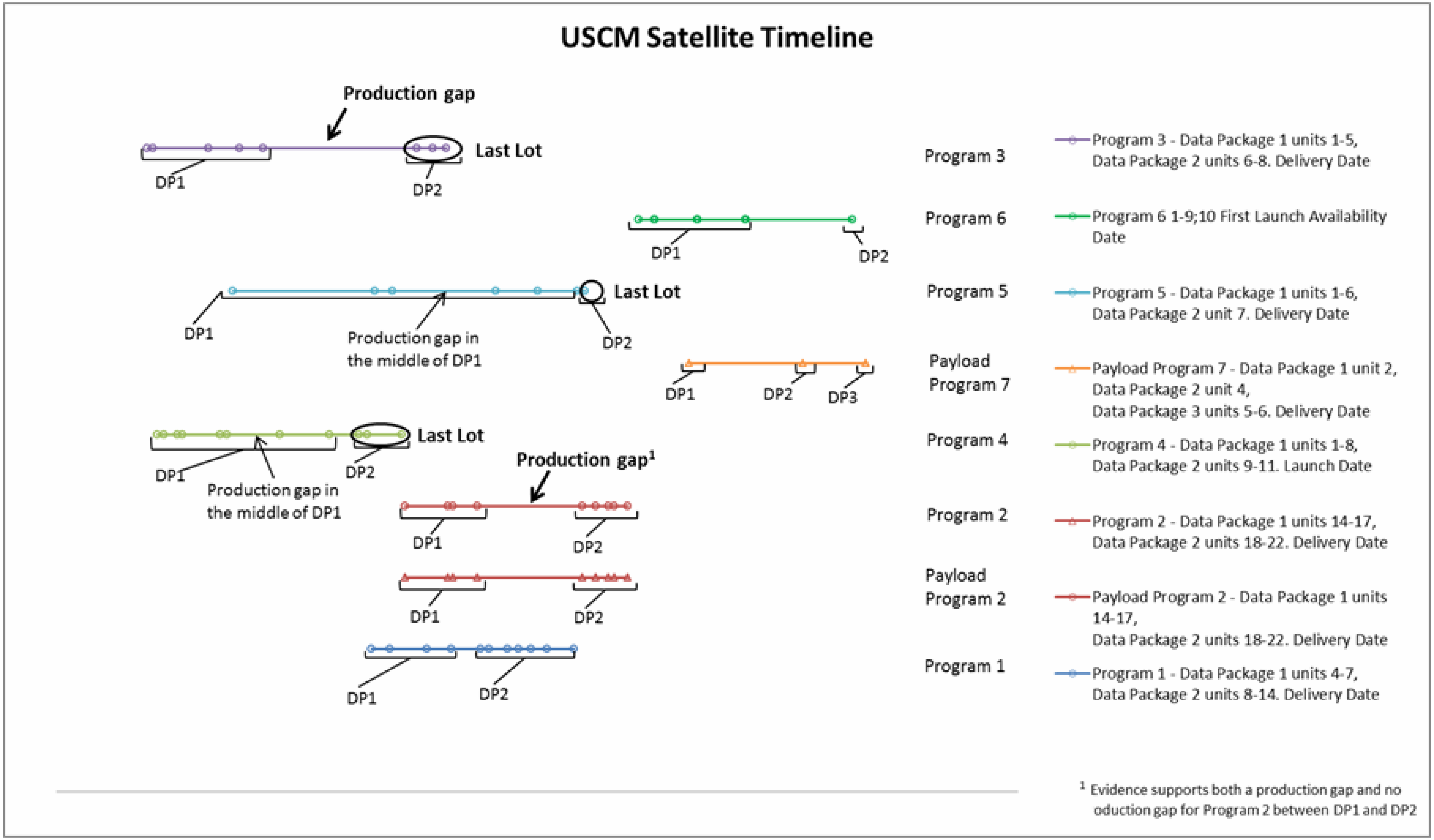
Green – production rate and economic ordering quantity
Purple – binary variable representing last lot and production break
Blue – scaling factor

- Not all variables tested simultaneously
  - Program dummy variables excluded when weight is used to scale data
  - Weight excluded when program dummy variables used
  - Lot size was eventually excluded from use due to data package features

*strikethrough of variables indicates they were not used in the final equation

USCM Satellite Data Packages – Timeline Summary

- Timeline used to identify production rate, production breaks and last lots



Variable Observations

Program	DP Quantity	Lot Quantities	Annual Production Rate	Production Break	Last Lot	Config Change
Program 3 DP1	5	1, 1, 1, 2	1			
Program 3 DP2	3	3	3	Yes	Yes	
Program 6 DP1	9	1, 2, 3, 3	2			Yes[1]
Program 6 DP2	1	1	1			
Program 5 DP1	6	Unknown	1			
Program 5 DP2	1	Unknown	2		Yes	
Program 4 DP1	7	Unknown	2			
Program 4 DP2	3	Unknown	1.5		Yes	Yes
Program 2 DP1	4	2, 2	1			
Program 2 DP2	5	1, 2, 1, 1	2	Yes or No		
Program 1 DP1	4	2, 2	1			
Program 1 DP2	7	2, 2, 2, 1	2			Yes

- Annual production rate never exceeded approximately one or two per year (Program 3 DP2 is one exception), and was not statistically significant
- Production Break – Program 5 and Program 4 had breaks within DP that may have influenced unit cost; Program 2 was analyzed under both conditions
- Lot size was eventually excluded from use due to data package features

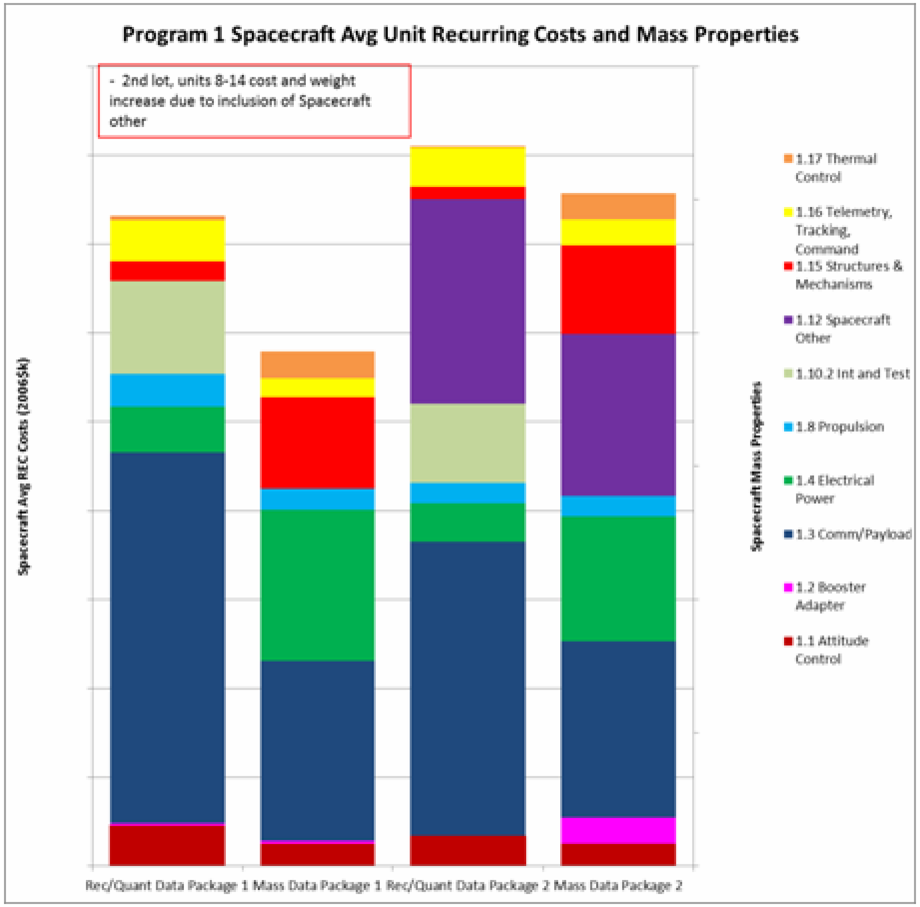
[1] Program 6 DP1 included spacecraft other costs which were all launch vehicle related costs

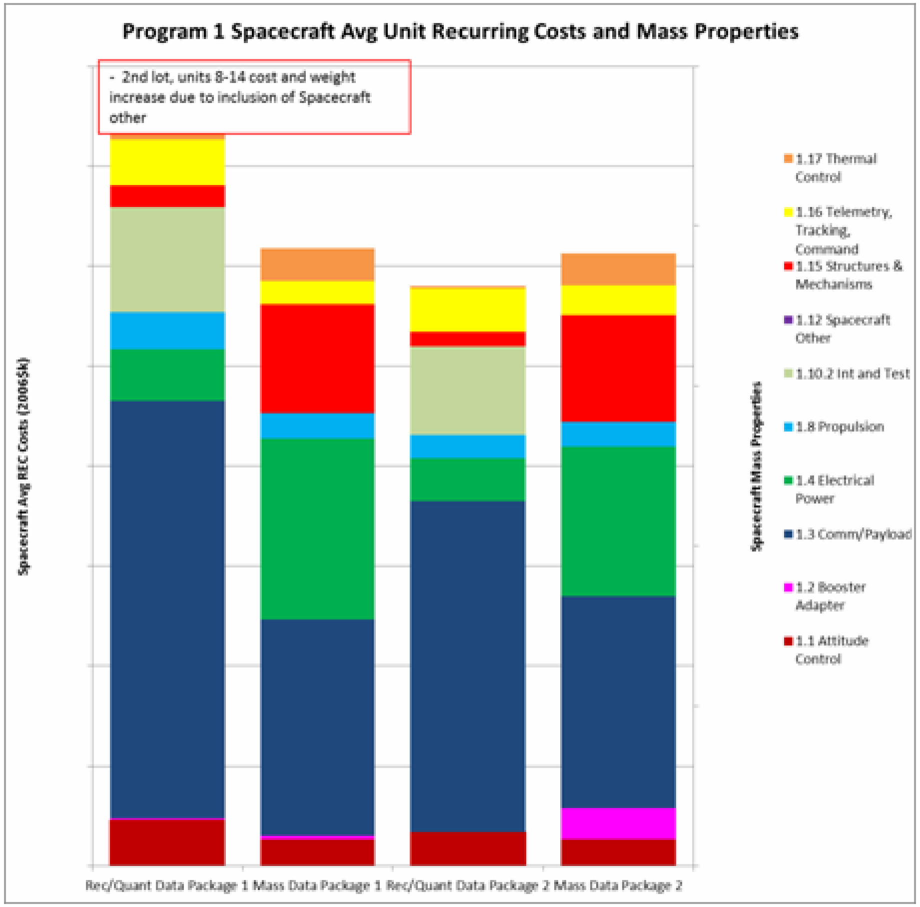
Adjustments to USCM Cost Values

- Program 1 and Program 4 cost (and weight) values were adjusted to exclude "Spacecraft Other" cost in order to normalize to a constant configuration
  - Spacecraft Other cost and weight included in USCM DP2 but not DP1
- Program 6 Spacecraft Other cost also excluded from DP1 since it was launch vehicle related
- Results in coefficients for a constant configuration
- Reduces variability due to changing configuration, resulting in reduced regression error

Adjustments to USCM Cost Values

- Program 1 "Spacecraft Other" cost comprised an apogee boost system included in Program 1 DP2 but not included in DP1 (graphic on left)
- Spacecraft Other cost and weight removed from DP2 data to normalize configuration weight and cost to DP1 configuration (graphic on right)





Results – Top Level Summary

- Results for data sets scaled by weight, or normalized to a baseline T1, were not satisfactory.
  - adj R^2 <.7; SEE > 50%
- Results using dummy variables to scale data produced results that can be used by AFCAA and SMC to estimate future spacecraft bus systems.
- Caveats
  - Learning slope based on two data points for each of 6 programs
  - Due to limited data set and multicolinearity, distinguishing impacts of production breaks and last lots is difficult
  - Program 5 and Program 2 drive results for production break and last lot; excursion excluding Program 5 data was performed

Results (With Break in Program 2 production)

- Regression equation for selected results
  - Cost (FY06$) = $T_l (Q^b) (PB^p) (LL^a) (D_l^{dl}) \ldots (D_n^{dn})$

Only results for Quantity, Production Break and Last Lot are used

Model	Adj. R^2	SEE	Ridge k	Learning		Production Break		Last Lot	
				Slope	p-value*	Cost Factor	p-value	Cost Factor	p-value
CAC	.993	.068	.04	87%	.001	+5.6%	.014	+10.5%	.001
IUC	.839	.312	.20	90%	.088	+7.2%	.493	+31.6%	.039
CAC w/o Program 5	.986	.101	.05	87%	.003	+9.2%	.019	+5.2%	.046
IUC w/o Program 5	.946	.192	.10	88%	.006	+8.4%	.07	+20.7%	.014

Production Break: Program 2, Program 3. Last Lot: Program 3, Program 4, Program 5 (when included)

Since OLS regressions exhibited multicolinearity, coefficients and statistics are based on Ridge Regressions, used to specify stable coefficients for independent variables.

*p-value = 1 - probability not zero

Technomics

Results (No Break in Program 2 production)

- Regression equation for selected results
  - $\text{Cost (FY06\$)} = T_1\,(Q^b)\,(PB^p)\,(LL^a)\,(D_1^{d1})\ldots(D_n^{dn})$

Only results for Quantity, Production Break and Last Lot are used

Model	Adj. R^2	SEE	Ridge k	Learning		Production Break		Last Lot	
				Slope	p-value	Cost Factor	p-value	Cost Factor	p-value
CAC	.994	.059	.04	88%	.003	+2.4%	.456	+10.7%	.009
IUC	.851	.301	.20	91%	.112	-3.2% [1]	.807	+33.6%	.034
CAC w/o Program 5	.990	.087	.05	89%	.016	+7.9%	.153	+4.3%	.232
IUC w/o Program 5	.952	.180	.10	89%	.041	+9.95%	.347	+18.3%	.103

Production Break: Program 3. Last Lot: Program 3, Program 4, Program 5 (when included)

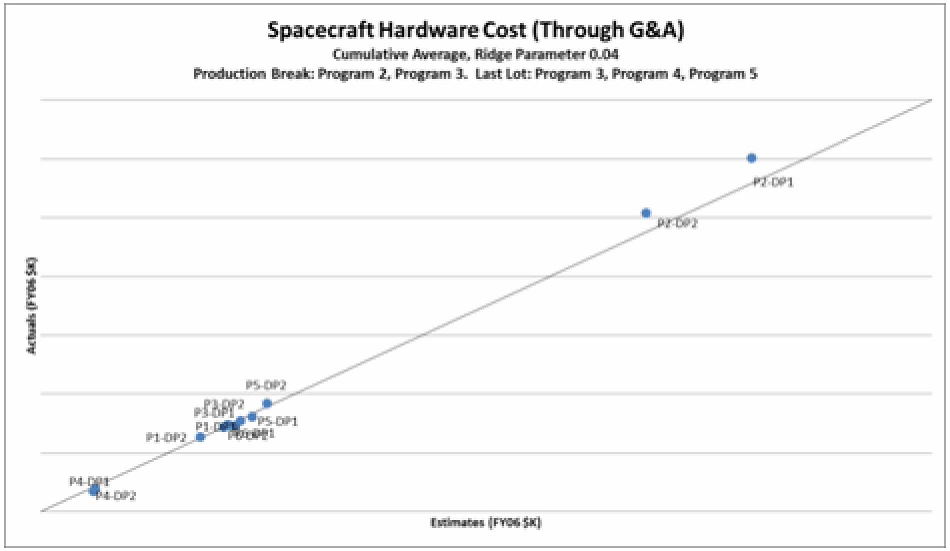
- Since OLS regressions exhibited multicolinearity, coefficients and statistics are based on Ridge Regressions, used to specify stable coefficients for independent variables.

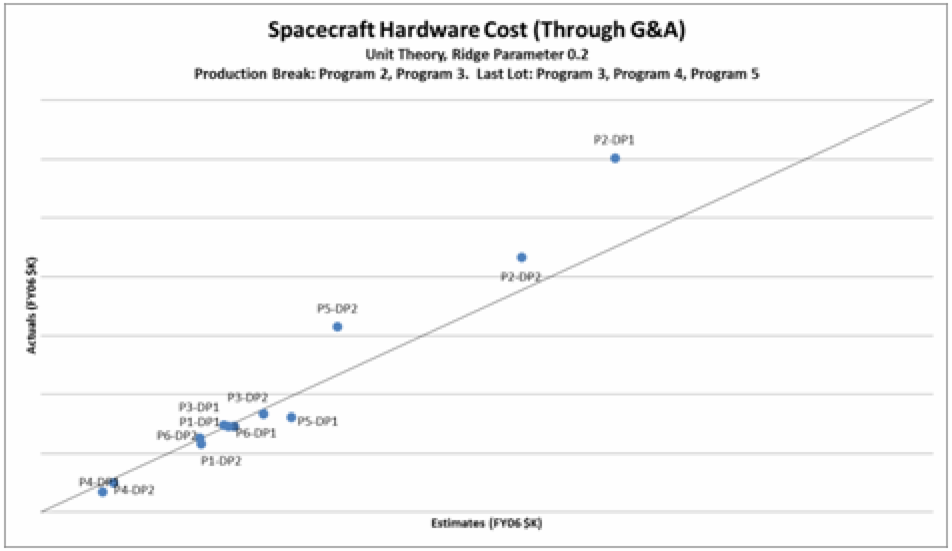
[1]The wrong sign on this coefficient indicates results for this scenario should not be used.

Error Plots

Includes Program 5, Ridge Regression, Program 2 w/ Break

- IUC error plots show how Program 5 and Program 2 drives error
  - Second Program 5 point is an outlier due to last lot impact
  - First Program 2 point is an outlier – Program 2 has production gap, yet has steepest slope
- Error plots demonstrate how CAC model smoothes variability



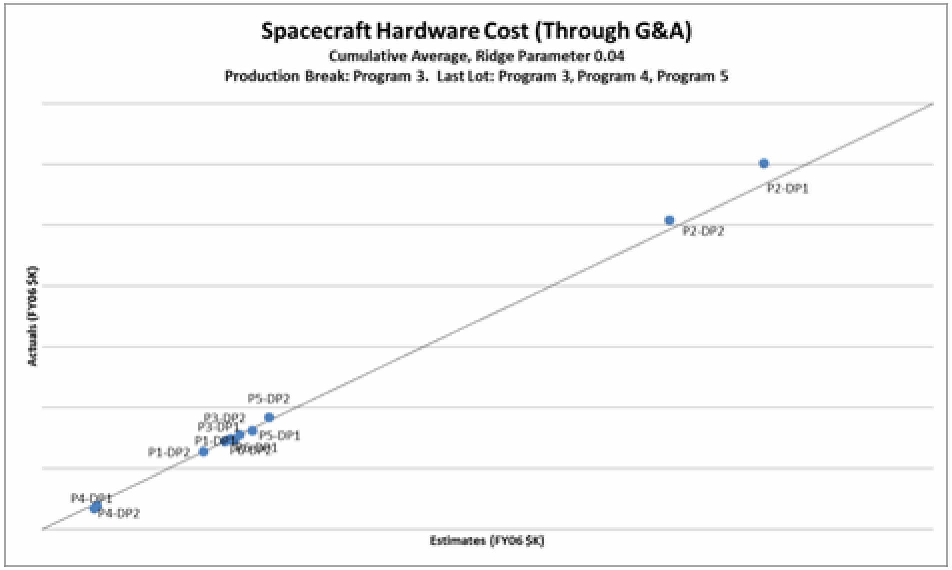


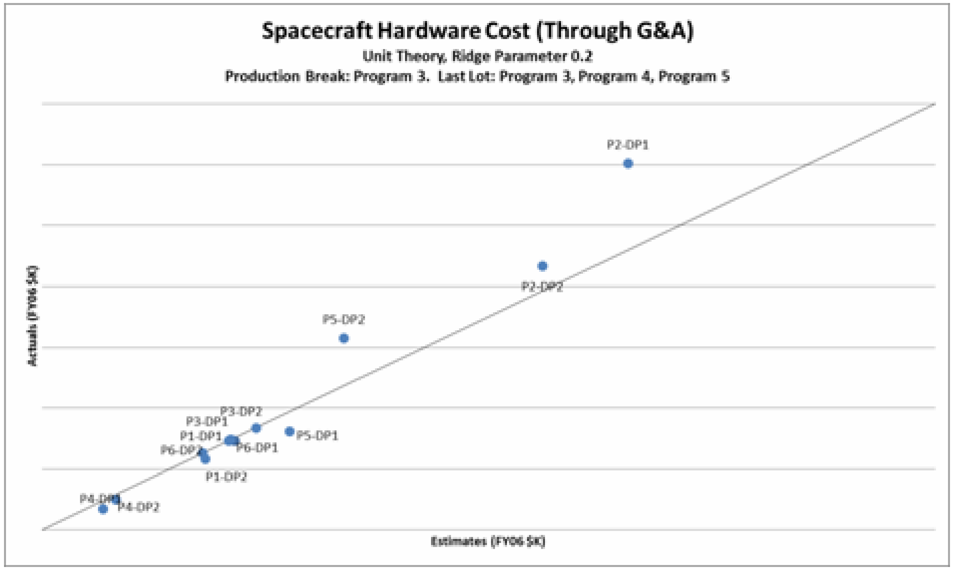
Production Break: Program 2, Program 3. Last Lot: Program 3, Program 4, Program 5

Error Plots

Includes Program 5, Ridge Regression, Program 2 w/o Break

- IUC error plots show how Program 5 and Program 2 drives error
  - Second Program 5 point is an outlier due to last lot impact
  - First Program 2 point is an outlier – Program 2 has production gap, yet has steepest slope
- Error plots demonstrate how CAC model smoothes variability





Production Break: Program 3. Last Lot: Program 3, Program 4, Program 5

How Results are Used

- Regression coefficients specify a cost improvement slope and cost estimating factors for production break and last lot impact.
- Results are applied to spacecraft bus recurring hardware WBS cost, including I&T, but excluding Systems Engineering and Program Management.
- Results apply to constant configurations. Impacts due to major (> ~5% weight) configuration changes should be estimated independently.
- Cost improvement slopes can be applied to point estimates for a specified quantity, (e.g., T_1, CAC_{10}) to estimate the cost of individual production lots. Point estimates must be developed independently.
- Production break and last lot cost factors are multiplicative factors applied to appropriate production lots (i.e., last lots or lots subsequent to a production break); either applied to lot average cost or total lot cost.
- Regression coefficients are interdependent - production break and last lot factors should not be used independent of respective improvement slopes.
- Production break and last lot cost factors are applied only to appropriate lot cost; thus, improvement slopes can be used exclusively, although most programs end with last lot effects.

Summary

- USCM data is fairly limited for learning analysis
  - Few programs with multiple data points
  - Almost all of these have only two data points
- Nature of the data (i.e., grouping of individual production lot cost into data packages) prevented results for production rate and lot size (although estimated production rate was not a significant variable for the data set analyzed).
- Results for learning slope were consistently 87 – 91%, for the data set analyzed
  - CAC: 87-89%; IUC: 88 – 91%
  - These are supported by the slopes for the three programs without production break or last lot impacts
- Results for production break and last lot effects are less certain and should be validated with additional data if possible

Learning Curve Analysis of Small Data Sets
Spacecraft Bus Cost Improvement Analysis

Questions?

Brian Welsh +1 571 366 1454 bwelsh@technomics.net
James York +1 571 366 1425 jyork@technomics.net

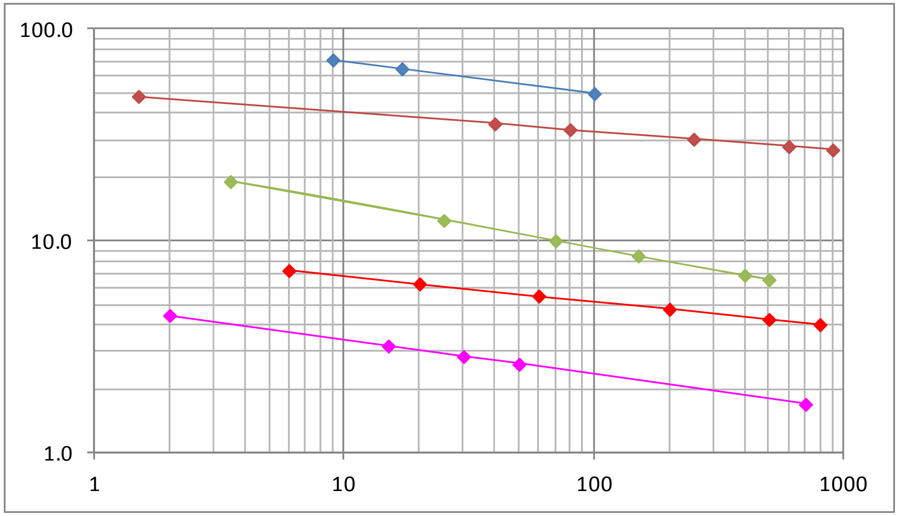
Technomics
The Science of Informed Decision Making

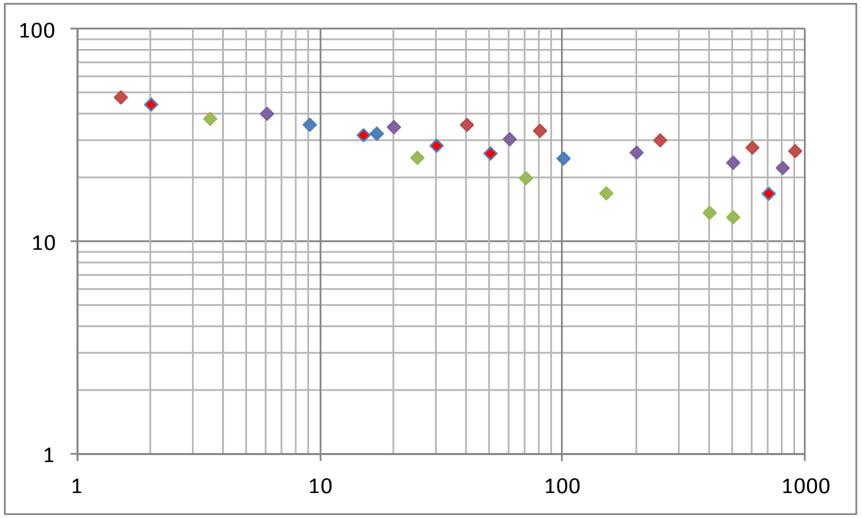
Back-up

Potential Follow-on Analysis

- Augment current analysis of USCM data with CCDR
  - CCDR repository can potentially produce two or more data points for: EELV, NAVSTAR GPS, GPS IIF, GPS IIR, WGS, DSCS II, DSCS III, FLTSATCOM, DSP, DMSP
- Analysis of cost improvement for spacecraft bus Systems Engineering and Program Mgmt.
  - SE/PM can be analyzed independently and in conjunction with hardware WBS elements
- Analysis of cost improvement at second level of WBS for spacecraft bus elements.
- Analysis of cost improvement for ground equipment.

Effect of Normalizing Program Learning Curves to a Baseline T1





Ridge Plot for IUC Results, w/TDRSS

Ridge plot shows effect of multicolinearity on coefficients

- PB goes from negative to positive
- LL decreases significantly