Data-Driven Guidelines for Correlation of Cost and Schedule Growth

ICEAA Training Workshop, San Diego

Summer 2015

Sidi Huang
Dr. Mark Pedigo
Cris Shaw
Ken Odom

Booz | Allen | Hamilton

Correlation Effects in Monte Carlo Simulations

- One of the most difficult things to account for in a Monte Carlo simulation is correlation between the independent variables
  - If correlation is not accounted for, the Coefficient of Variation of the top level distribution will be artificially shrunk

- When independent variables are correlated, they will tend to grow and shrink in tandem. Ignoring correlation will result in a poor analysis; generally reporting overly optimistic results.

- Correlation is just an observed relationship, there does not have to be an explanation for why it happens, although often we want to know if there is one

Default Guidance for Correlation Values

- Although there have been empirically driven studies for cost correlation, there have not been considerable empirical studies on task duration correlation

- "Using reasonable nonzero values, such as 0.2 or 0.3, generally leads to a more realistic representation of total cost uncertainty"

  "Estimating System Cost" (Stephen A. Book, Crosslink Winter 2000/2001)

- Perform data-driven research to find a default correlation guidance value for schedule uncertainty

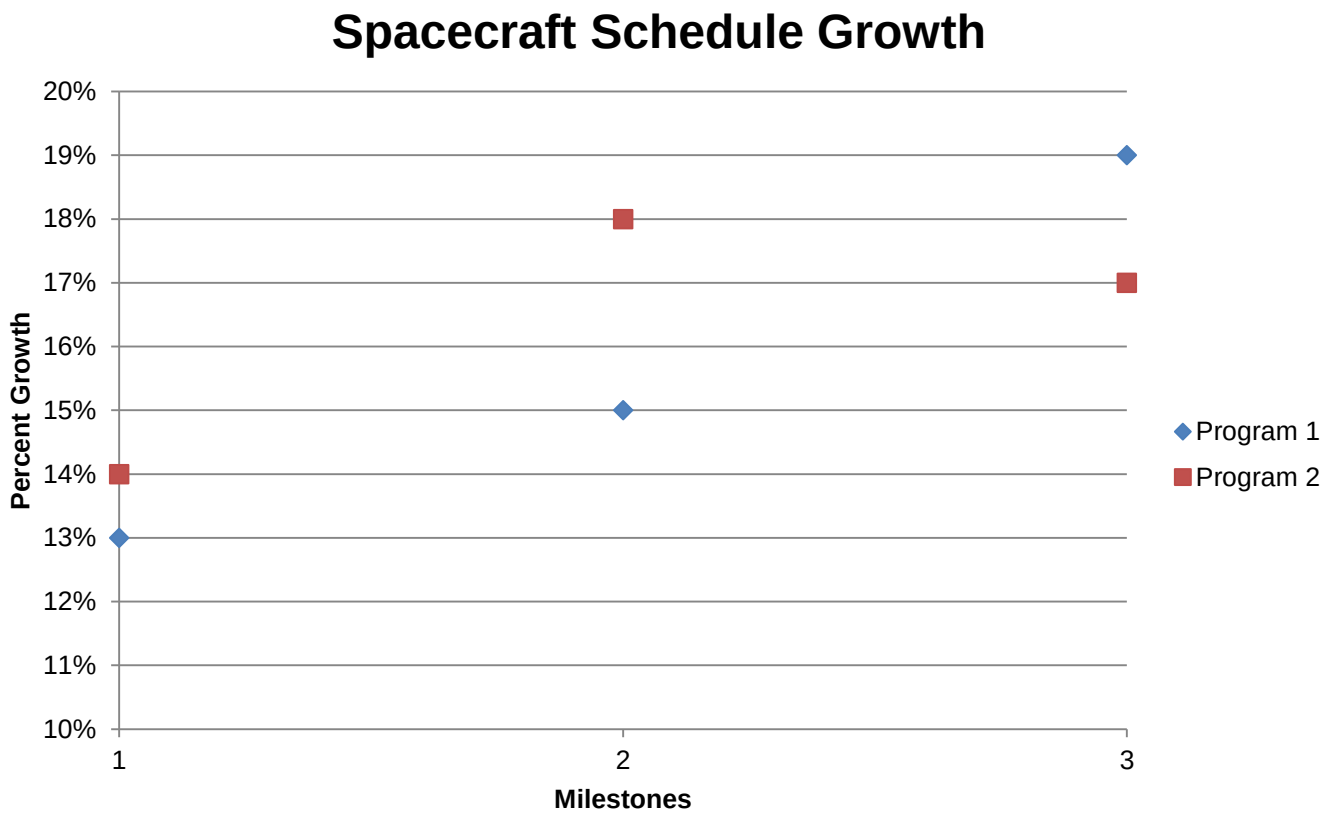
- Data files from historical NASA missions
  - NASA Cost Analysis Data Requirements (CADRe)
  - Mission Milestone Reviews
  - Mission Quarterly or Monthly Status Reviews

Correlation of Schedule Activity Durations
Approach #1: Methodology

- Find consistent milestones for two Missions (e.g., PDR, CDR, and Launch); missions were held constant
- Focus on finding a correlation by Activity (e.g., spacecraft development) between two missions
- Take the many correlation values for each Activity for many missions, pool the correlation results for that Activity (e.g. spacecraft development) and compare to correlation results for another Activity (e.g. instrument development) to find “magic correlation value”
- Spacecraft development growth from PDR to CDR

Correlation of Schedule Activity Durations
Approach #1: Outcome

- Three data points were too few to capture correlation



Correlation of Schedule Activity Durations
Approach #1: Observations

- Not every Mission had data for the same milestone reviews (over 15 different milestone reviews identified)
- With few data points, the correlation was very volatile; required the two missions to react in similar ways for each milestone
- Lesson learned: Find a way to increase the number of data points captured

Correlation of Schedule Activity Durations
Approach #2: Methodology

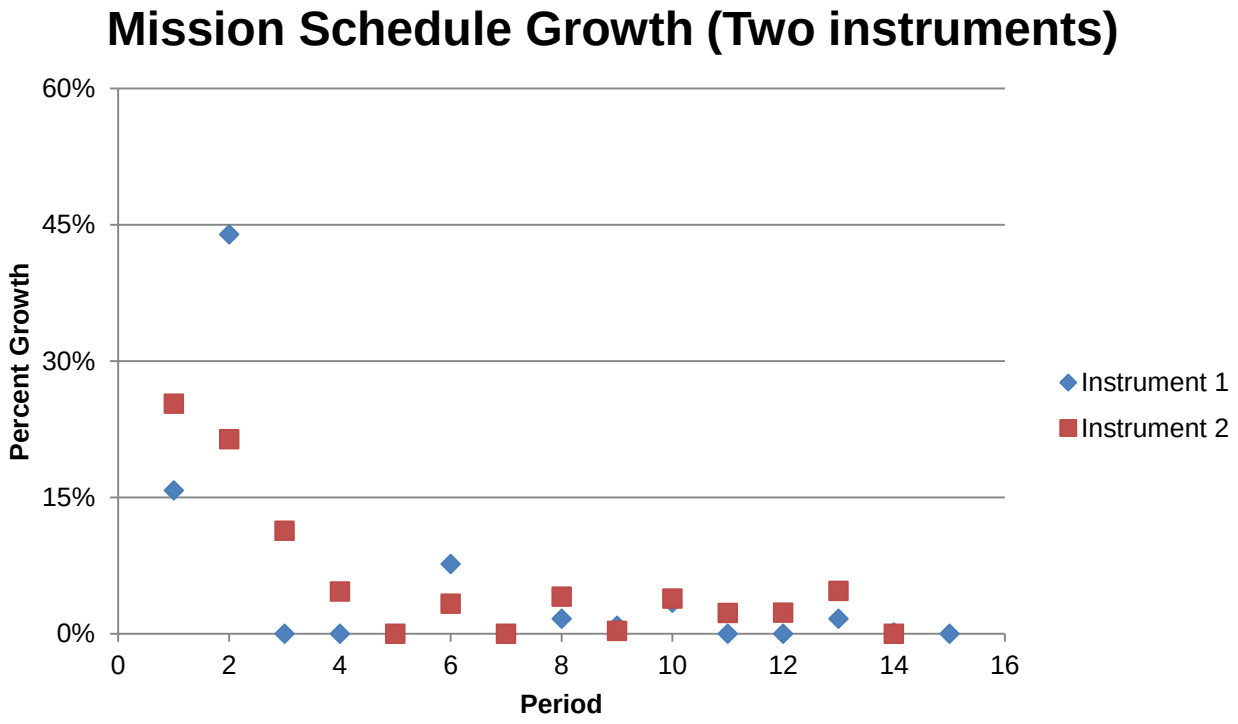
- To solve the problem of too few data points, find more data for one mission (either from multiple milestones or from multiple monthly status reviews) that captured two activities; activities were held constant

- Find a correlation between two activities (e.g., spacecraft and instrument development) between one mission

- After identifying correlation values for many missions, pool correlation results for that activity (e.g. spacecraft development) and compare to correlation results for another activity (e.g. instrument development) to find “magic correlation value”

Correlation of Schedule Activity Durations
Approach #2: Outcome

- Plausible but extraordinarily time intensive
- Some missions might have many years of no growth, then one period of growth (e.g. due to schedule replan)



Correlation of Schedule Activity Durations
Approach #2: Observations

- Extraordinarily time intensive
- Required every monthly or quarterly status review to show the same level of detail
- When an Activity was completed, no more changes were made and correlation data ended
- Lessons learned: Focus on “early” and “late” data

Correlation of Schedule Activity Durations
Approach #3: Methodology

- Find correlation of uncertainty distributions in schedule duration
  - Implying schedule estimates have uncertainty distributions
  - Two estimates for two activities have correlated uncertainty distributions
- Collect an “early” estimate and a “late” estimate
  - The “early” estimate will be an estimate that has a lot of uncertainty
  - The “late” estimate will not have very much uncertainty and will primarily be actuals
  - The difference in estimation is the uncertainty that occurred (i.e. the uncertainty distribution closed)
- Compare the early-estimation to late-estimation differences
  - This comparison will capture the early estimate uncertainty
  - If the estimation differences are correlated then the uncertainty distributions are correlated

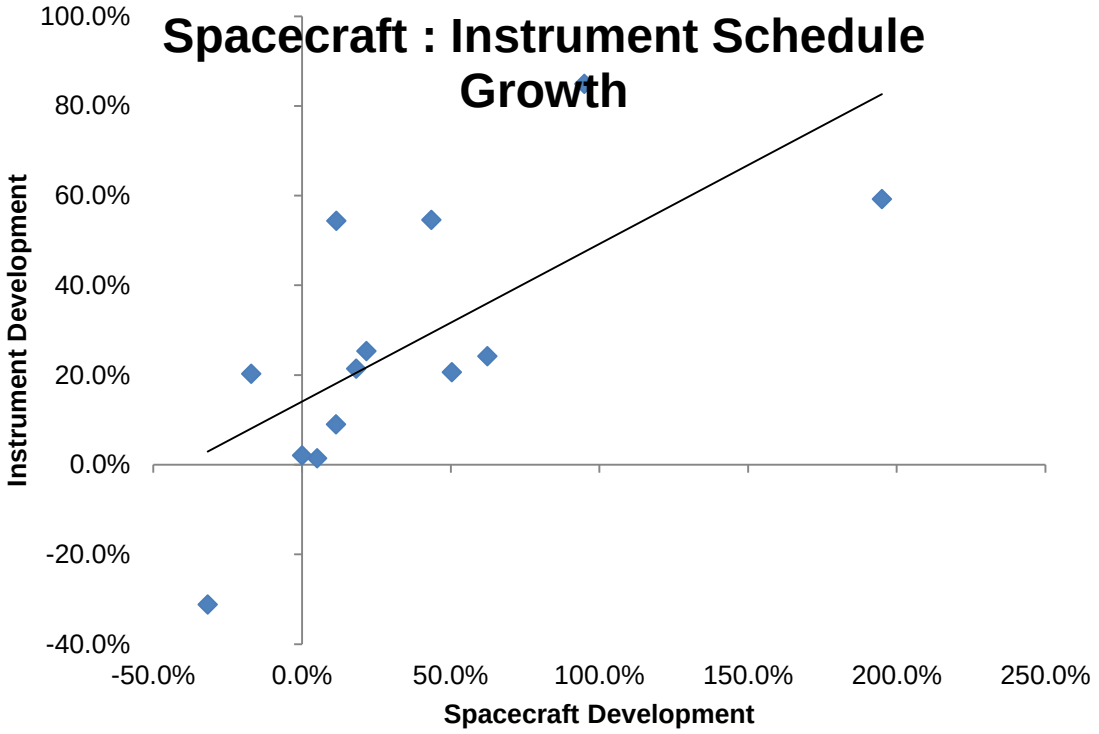
Correlation of Schedule Activity Durations
Approach #3: Outcome

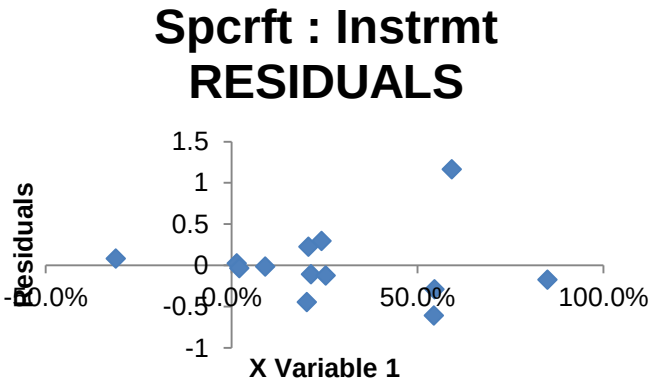
- When missions are loaded with multiple instruments, there is an opportunity for greater instrument:instrument comparisons
- A random sample of up to 3 data points per mission (instrument:instrument comparison) was taken

Mission	Instruments Comparison	Instrument X growth	Instrument Y growth
Mission 1	1:2	11.4%	16.4%
	1:3	11.4%	11.0%
	1:4	11.4%	3.4%
	1:5	11.4%	21.2%
	2:3	16.4%	11.0%
	2:4	16.4%	3.4%
	2:5	16.4%	21.2%
	3:4	11.0%	3.4%
	3:5	11.0%	21.2%
	4:5	3.4%	21.2%
Mission 2	1:2	54.4%	16.5%
	1:3	54.4%	-0.7%
	2:3	16.5%	-0.7%
Mission 3	1:2	24.2%	21.3%

Correlation of Schedule Activity Durations
Approach #3: Empirical Results

- Spacecraft : Instrument schedule growth correlation
  - Pearson's r = **0.679** (13 missions; n = 13)

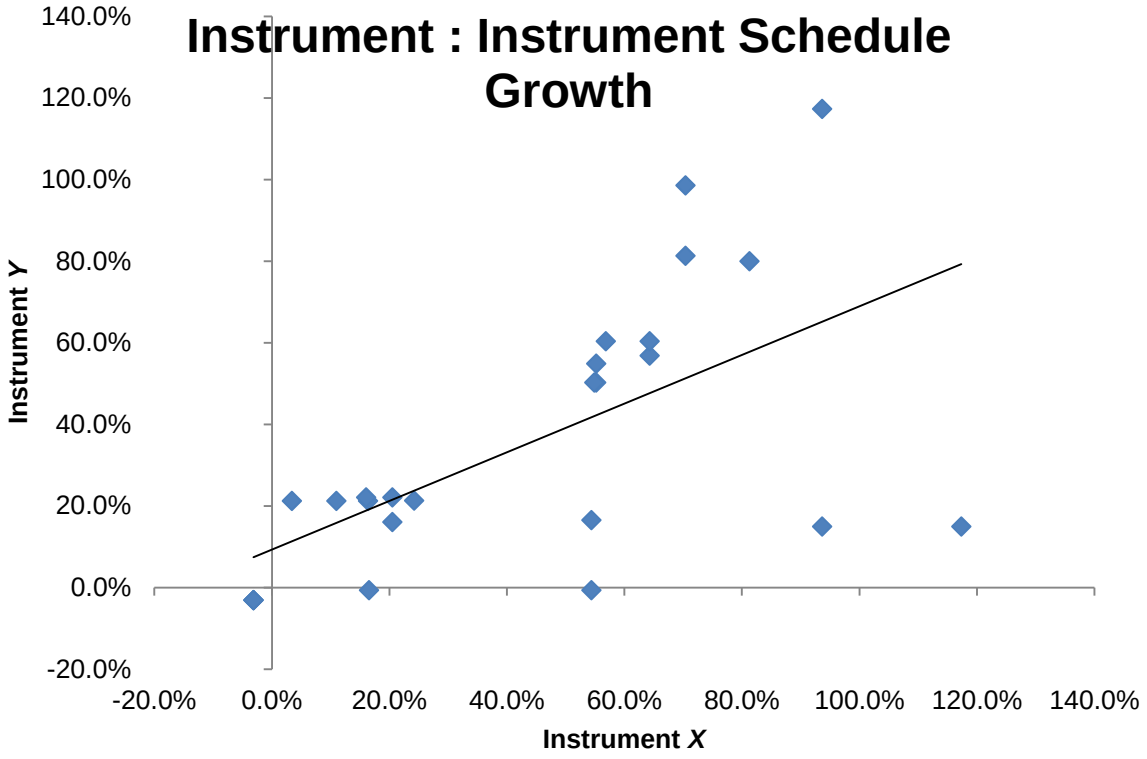


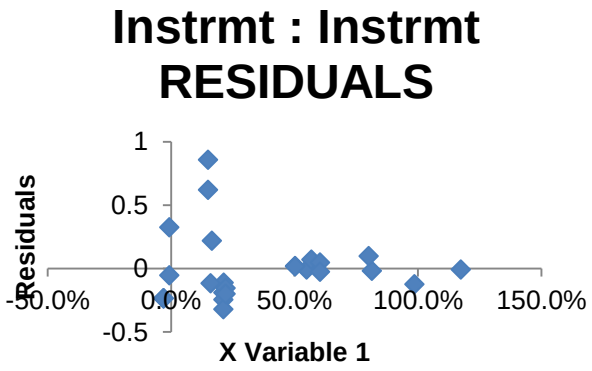


Booz | Allen | Hamilton

Correlation of Schedule Activity Durations
Approach #3: Empirical Results

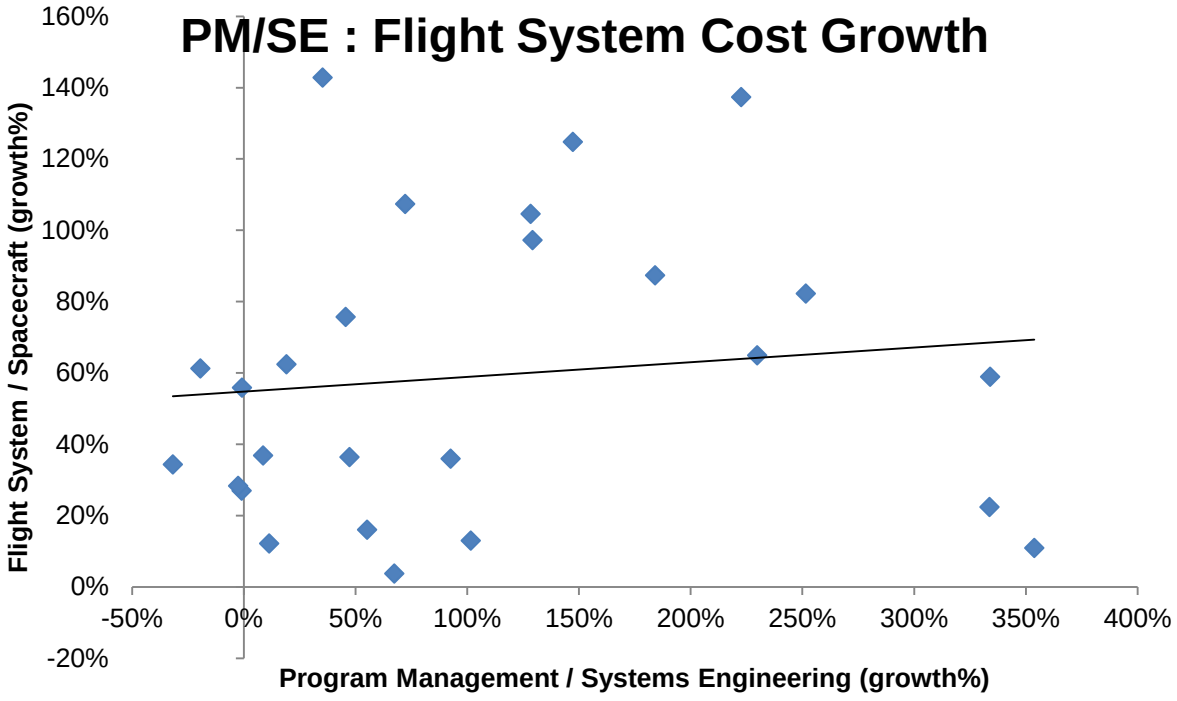
- Instrument : Instrument schedule growth correlation
  - Pearson's r = **0.605** (34 instruments across 9 missions; n = 25)

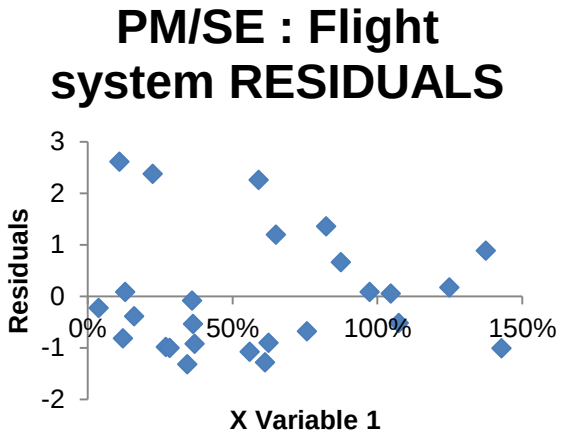




Correlation of Costs
Empirical Results

- PM/SE : Flight System cost growth correlation
  - Pearson's r = **0.117** (26 missions; n = 26)

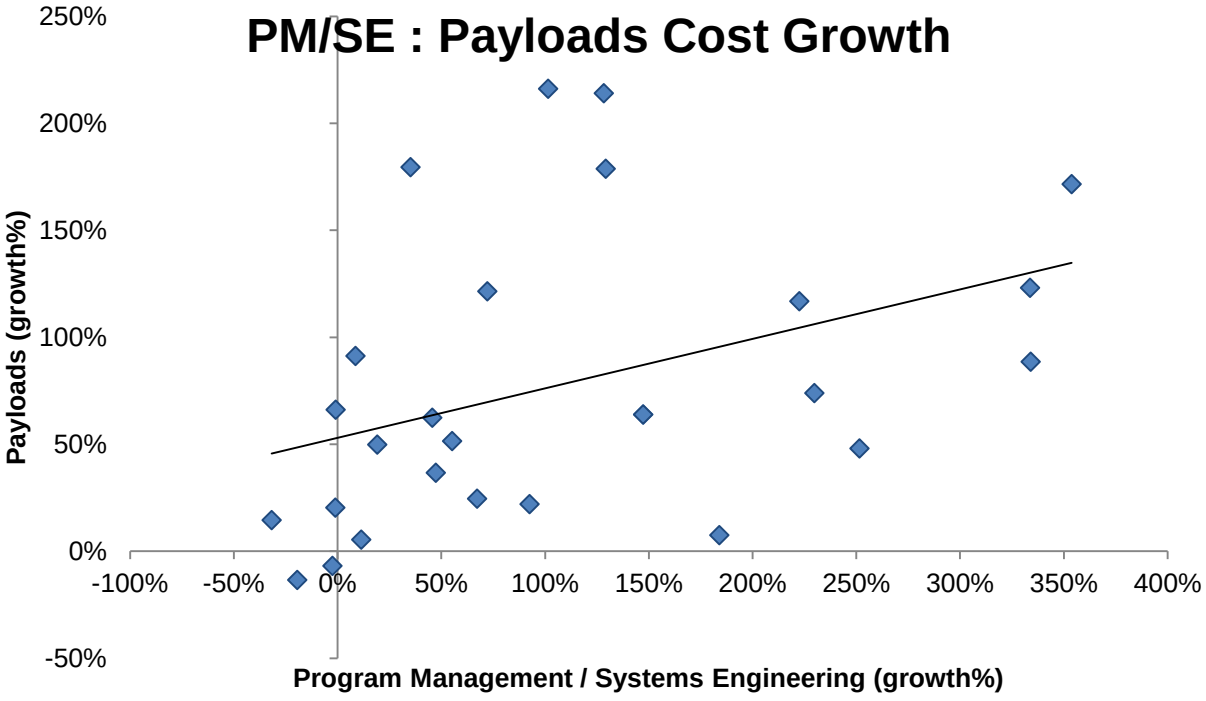


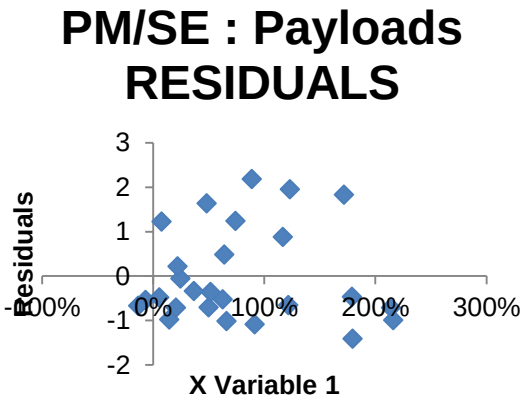


Booz | Allen | Hamilton

Correlation of Costs
Empirical Results

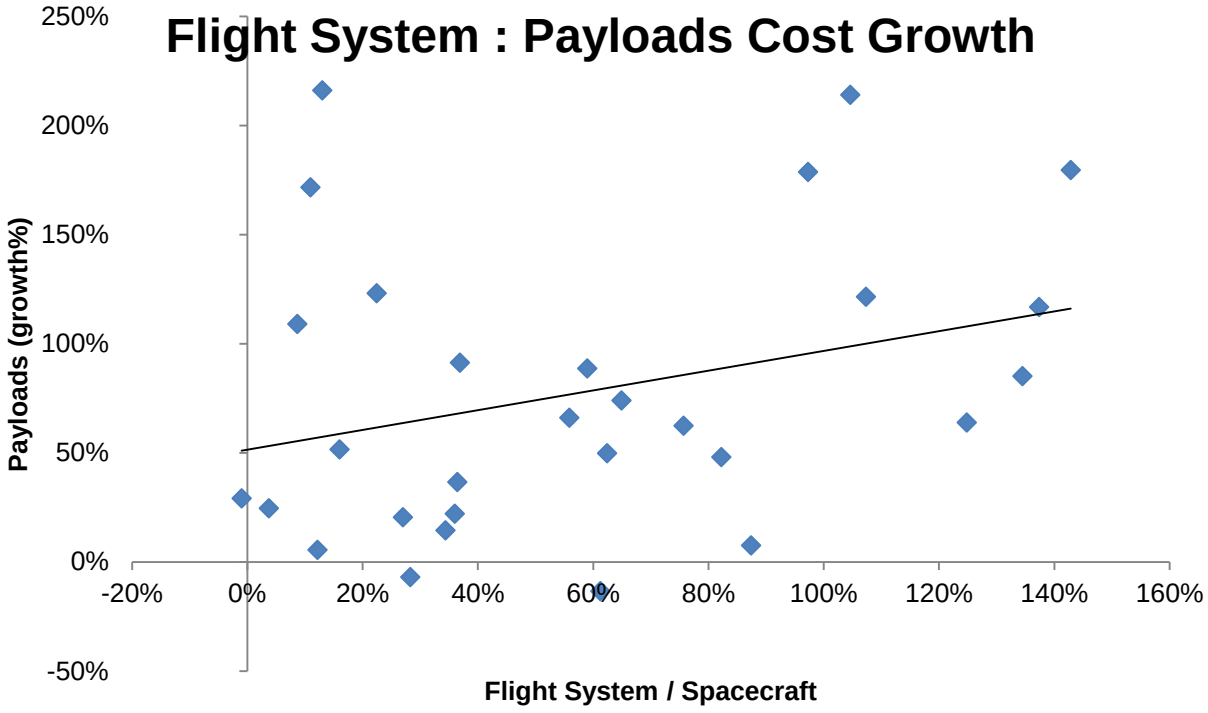
- PM/SE : Payloads cost growth correlation
  - Pearson's r = **0.394** (26 missions; n = 26)

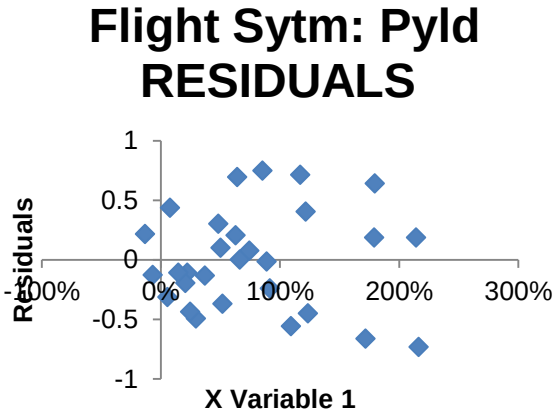




Correlation of Costs
Empirical Results

- Flight System : Payloads cost growth correlation
  - Pearson's r = **0.303** (29 missions; n = 29)





Conclusions and Future Research

- Ignoring correlation in running Monte Carlo simulations generally report overly optimistic results
- For schedule uncertainty, a default correlation value closer to 0.6 is shown to be effective
- Further categorization of missions by certain metrics could find unique correlation values for different types of missions
  - Example metrics: mission duration, cost, launch year, mass, power
  - Advanced metrics: schedule topology, missions where costs are skewed toward flight system versus costs skewed toward payload
- Categorization of cost elements to Time-Independent or Time-Dependent costs, and reconciling differences between WBS elements and CES elements