Early Effort and Schedule Models for Agile Projects in the US Department of Defense (DoD)

Wilson Rosa
Bradford K Clark
Ray Madachy
Barry Boehm

2020 SCAF/ICEAA Virtual Workshop
December 8, 2020

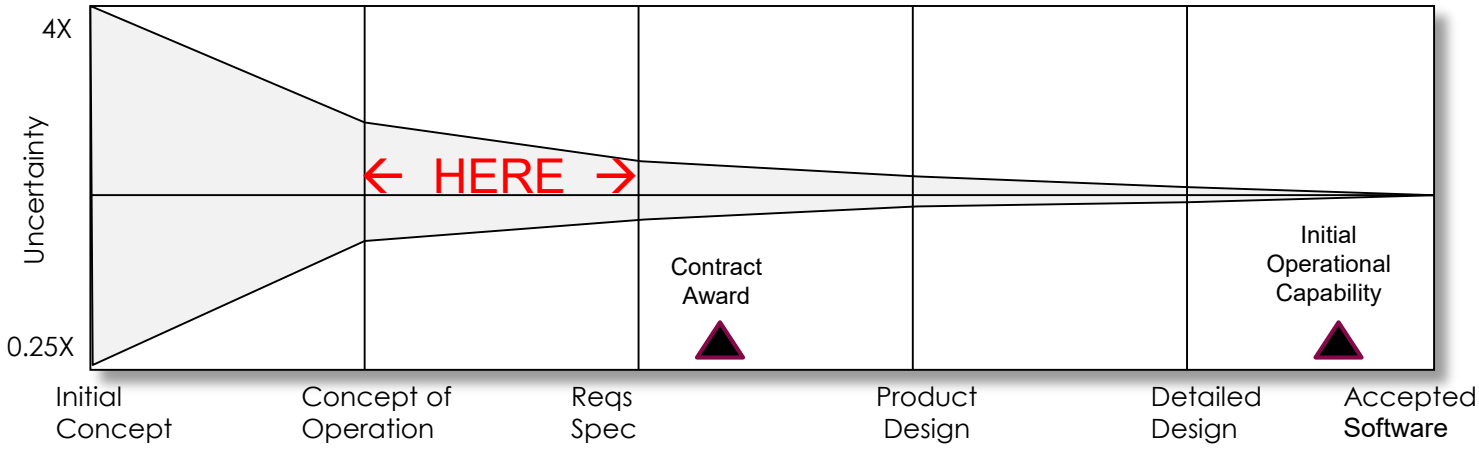
What is the Problem?

- Cost estimates for agile software projects are very critical at early stages to evaluate contract proposals and to establish initial program budgets

- However, mainstream sizing measures are not practical for estimating agile projects at early life cycle as these are generated after contract award

What is the Solution?

- A software sizing approach for model inputs, generally available at early lifecycle
- A set of software effort and schedule estimation models for agile projects that can be used early in the project's lifecycle



Outline

- Research Method
- Dataset Demographics
- Agile vs Traditional Processes
- Effort Estimation Models
- Schedule Estimation Models
- Conclusion

Research Method

DoD Data Collection Form:

Software Resource Data Report (SRDR)

- SRDR is a standardized mechanism to collect objective and measurable data on programs
- SRDRs are required for software development contracts over $20M USD
- Includes fields for agile processes and metrics reporting

Submitted after contract start

SRDR Initial Developer Report

Initial Functional Requirements
Initial External Interfaces
Initial Peak Staff

Submitted after contract end

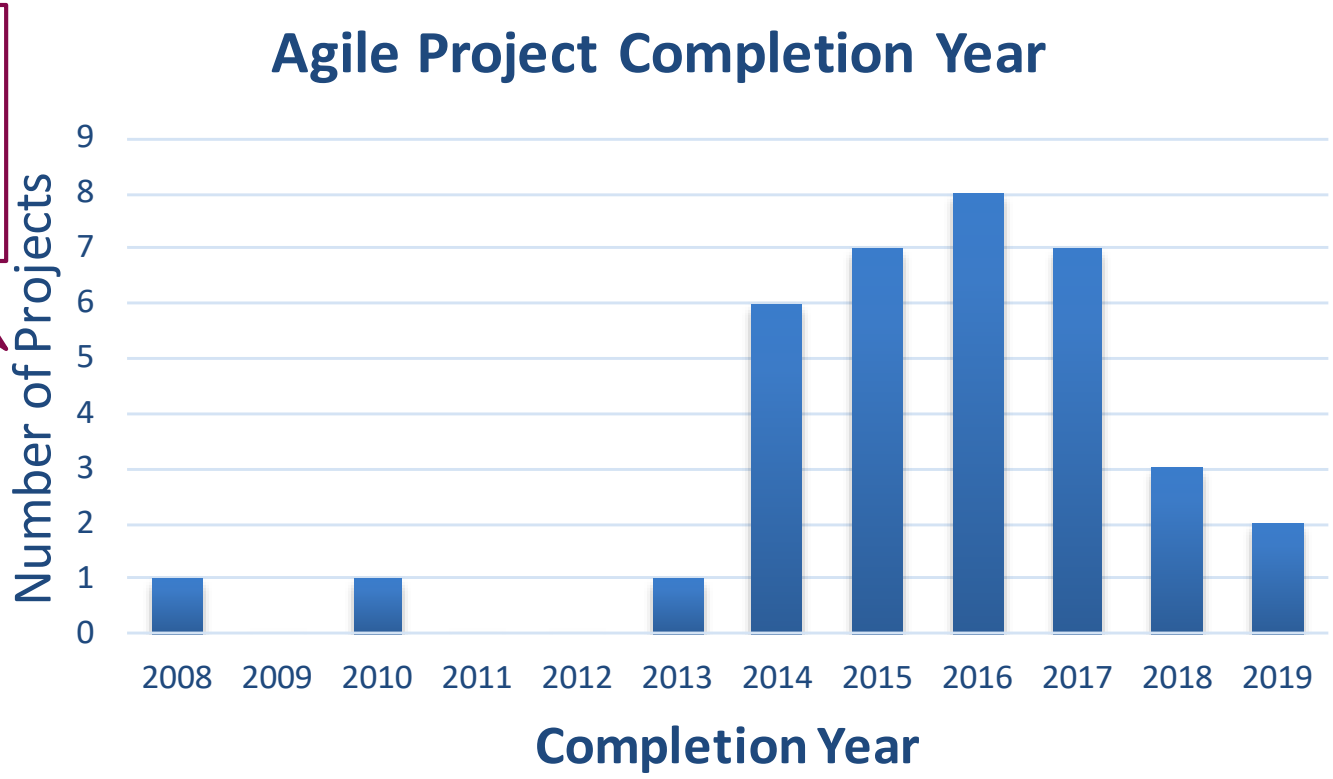
SRDR Final Developer Report

Actual Development Effort
Actual Development Schedule
Actual Development Process

Questionnaire: https://cade.osd.mil/Content/cade/files/csdr/guidance/DI-MGMT-82035A_SRDR%20Report.pdf
Form: https://cade.osd.mil/content/cade/files/csdr/dids/current/dd3026-1_2019.XLSX

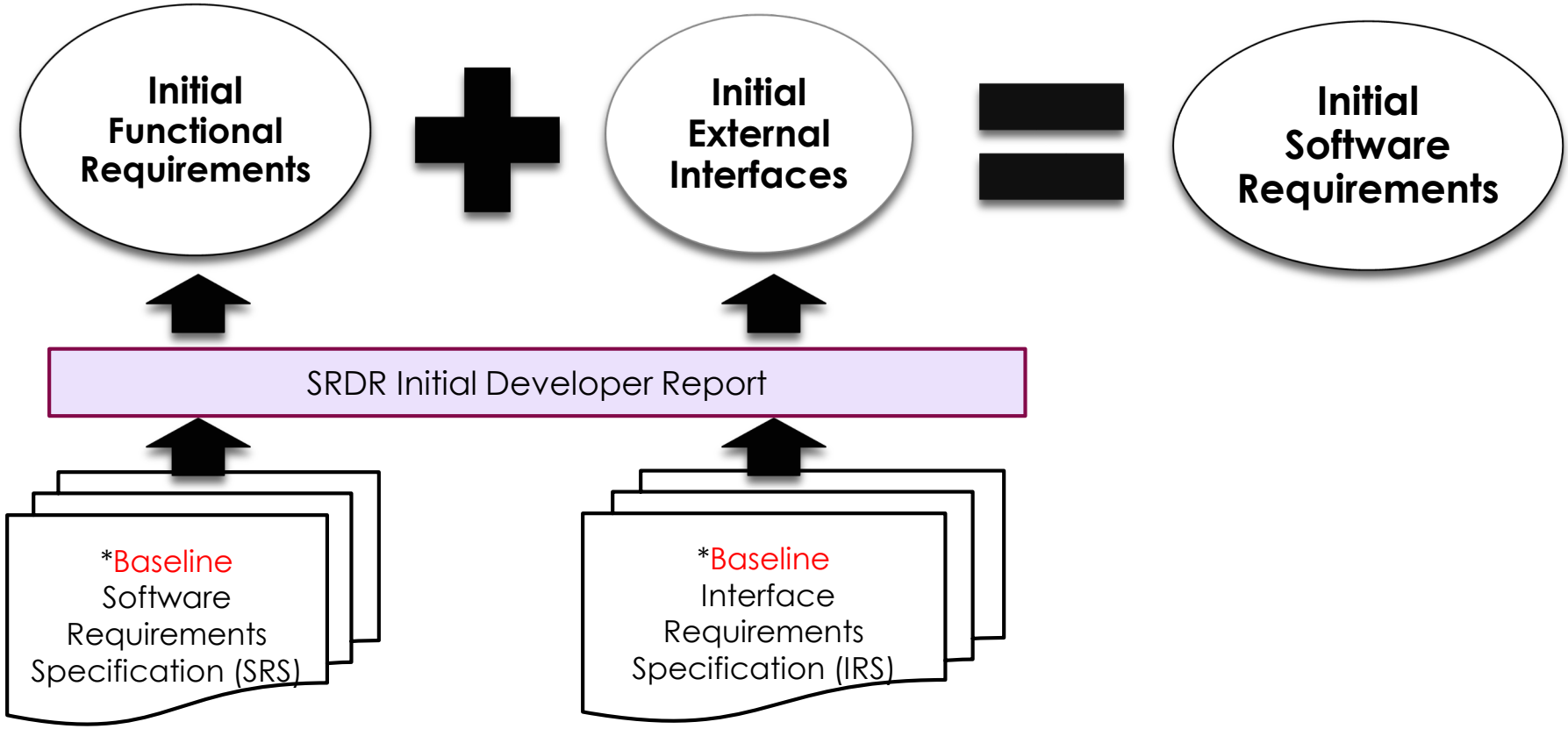
Sample and Population

36 Agile Projects



Earlier projects (2008, 2010, 2013) used a tailored agile process to fit developer practices

Software Sizing Approach



*Baseline SRS and IRS are typically developed by the government before contract award
Developers will report the initial functional and external interface requirements in the SRDR Initial Developer Report

Data Grouping by Super Domain

- Dataset grouped into 4 complexity zones called **Super Domain**

Super Domain		Application Domain
Mission Support	SUPP	Software Tools
		Training
Automated Information System	AIS	Enterprise Services
		Custom AIS
		Mission Planning
		Enterprise Information System
Engineering	ENG	Scientific & Simulation
		Test Measurement & Diagnostic Equipment
		System Software
		Process Control
Real Time Embedded	RTE	Command & Control, Communications
		Real Time Embedded
		Vehicle Control, Vehicle Payload
		Signal Processing, Microcode & Firmware

Variables in the Study

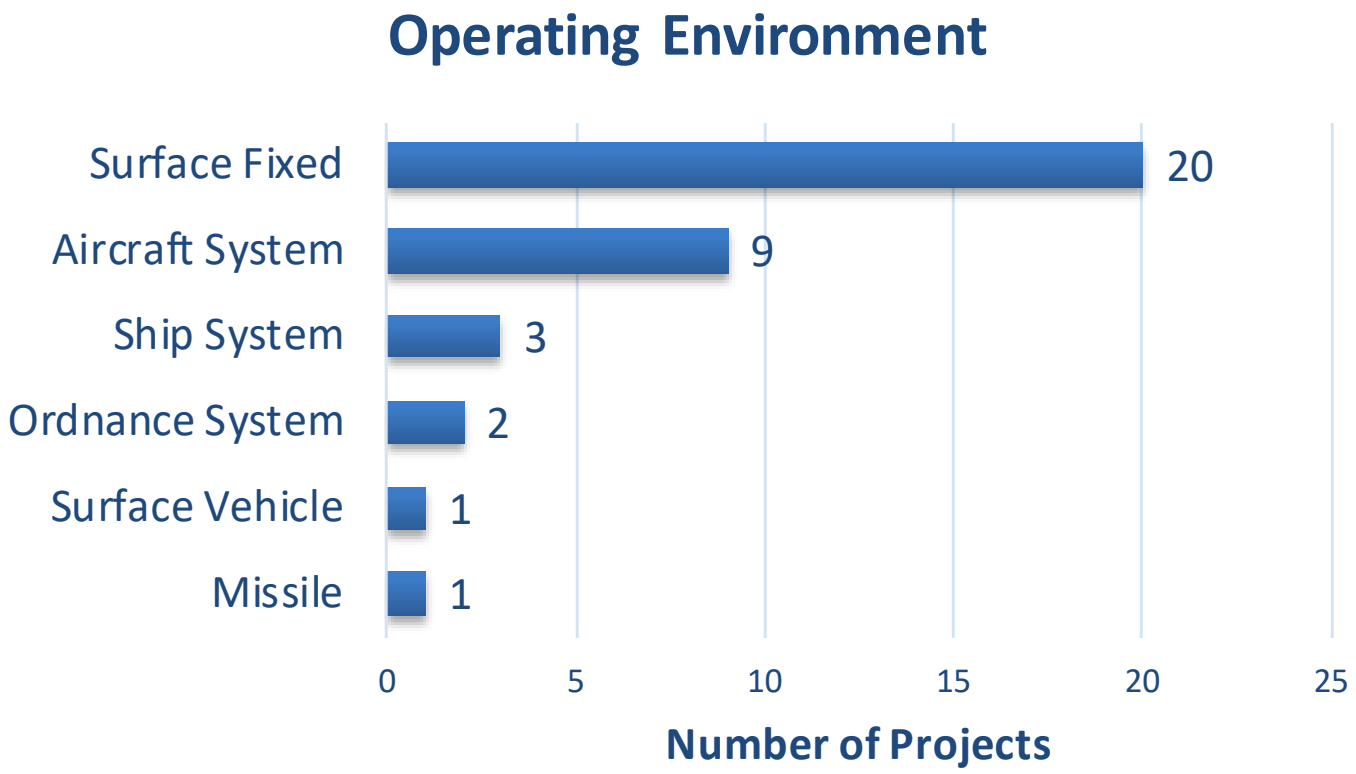
Variable	ID	Type	Definition
Final Effort	E	Dependent	Actual development effort (in Hours) at contract end
Final Schedule	TDEV	Dependent	Actual development time (in Months) at contract end
Initial Software Requirements	REQ	Independent	Sum of Initial Functional Requirements and Initial External Interface Requirements reported at contract award. Counts "system shall" statements from baseline SRS and IRS.
Initial Peak Staff	Staff	Independent	Estimated peak staff (in full-time equivalent) at contract start
Super Domain	SD	Categorical (Dummy)	Treatment of the 4 (r) super domains required the addition of 3 (r-1) dummy variables denoted as: D1 = 1 if AIS, 0 if SUPP or otherwise D2 = 1 if ENG, 0 if otherwise D3 = 1 if RTE, 0 if otherwise

Model Selection Criteria

Measure	ID	Description
Coefficient of Determination	R^2	Coefficient of determination is the percentage of variation in the response explained by the model
Adjusted R^2	R^2 (adj)	Percentage of the variation in the response explained by the model, adjusted for the # of predictors relative to the # of observations.
Predicted R^2	R^2 (pred)	Involves removing each observation from the dataset, estimating the regression equation, determining how well the model predicts the removed observation, and repeats for all data points in the dataset.
Variance Inflation Factor	VIF	Indicates whether multi-collinearity (correlation among predictors) is present in a multi-regression analysis.
P-value	a	Statistical significance established through coefficient alpha (a = 0.05).
Mean Magnitude of Relative Error	MMRE	Mean Magnitude of Relative Error is an indicator of model's accuracy: Low MMRE= high accuracy

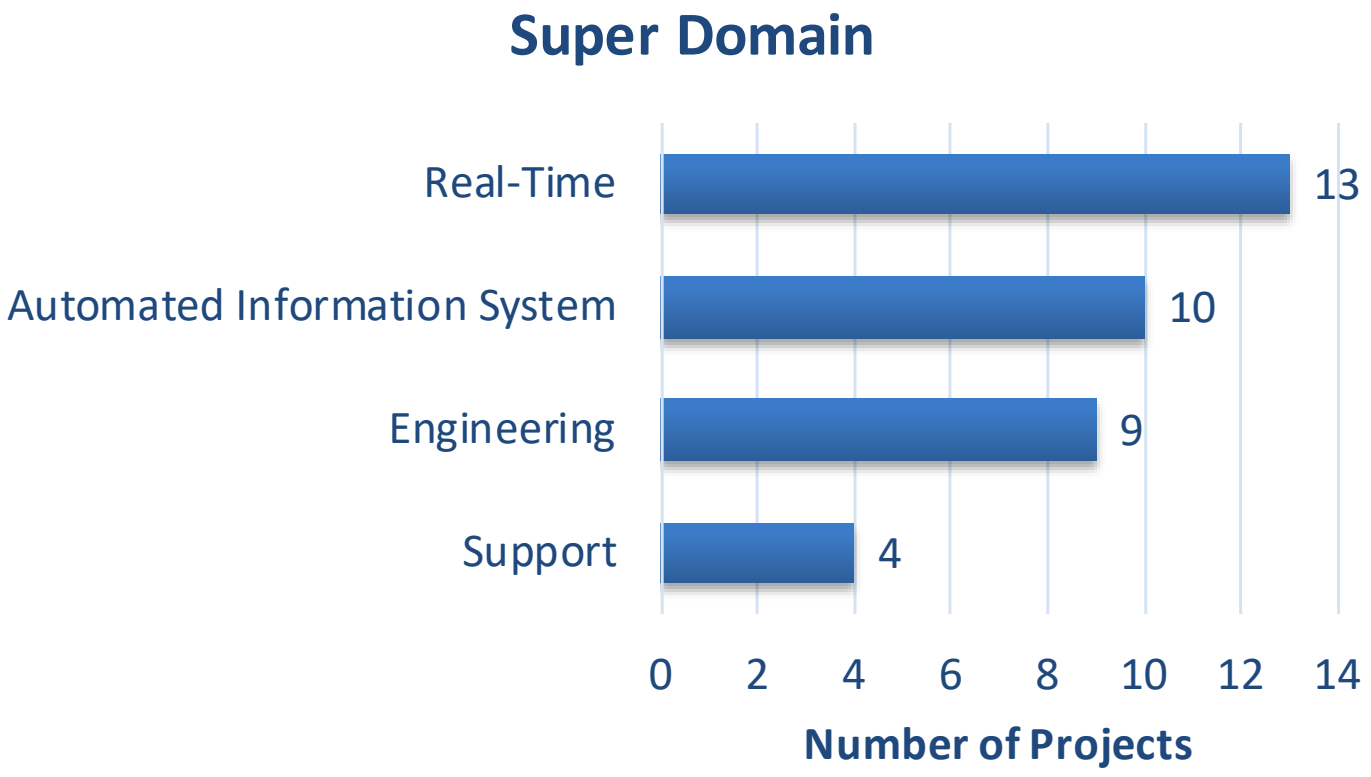
Dataset Demographics

Project Characteristics (1 of 2)

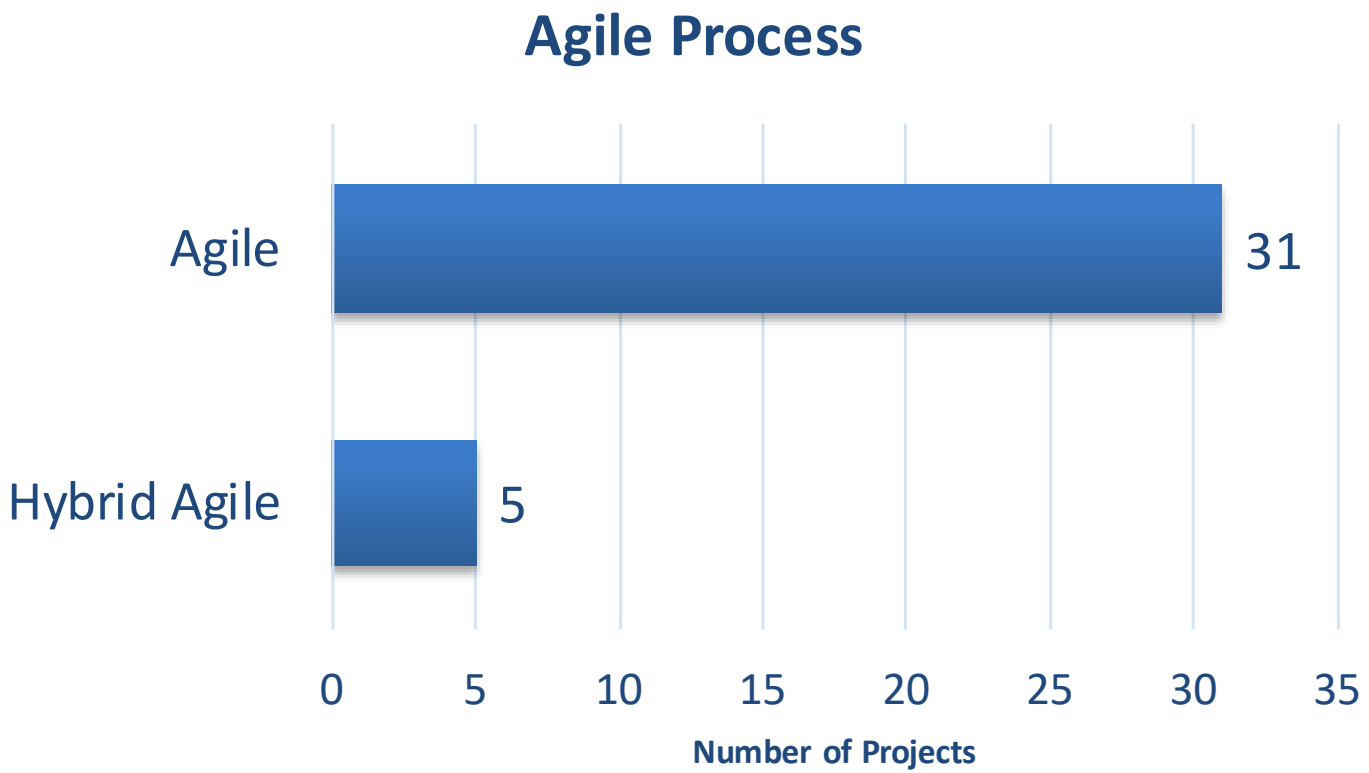


Most projects are hosted at a Surface Fixed or part of an Aircraft System

Project Characteristics (2 of 2)



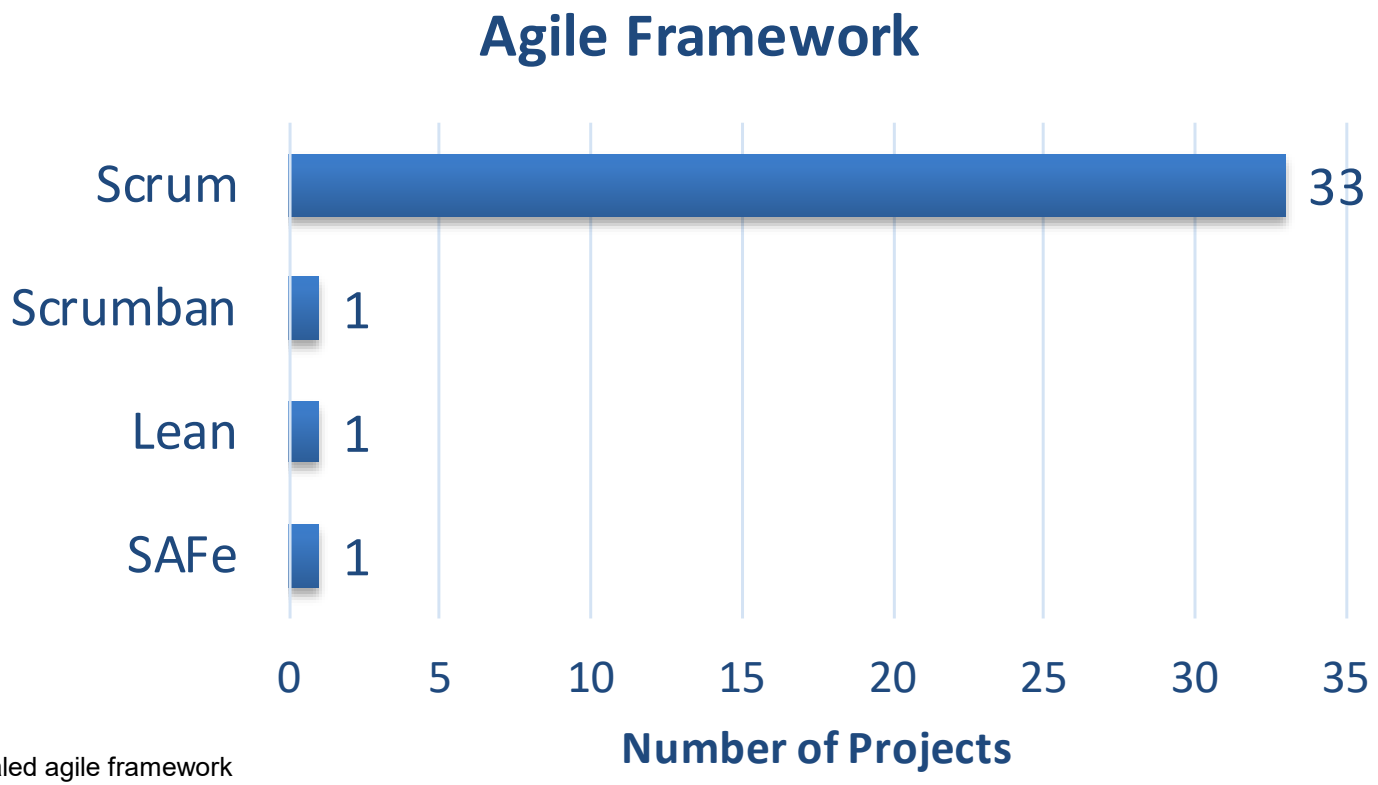
Dataset by Agile Process



Hybrid Agile combines principles of waterfall (for requirements analysis) and agile (for **DCTI**)

DCTI = Design, Code, Test, and Integration

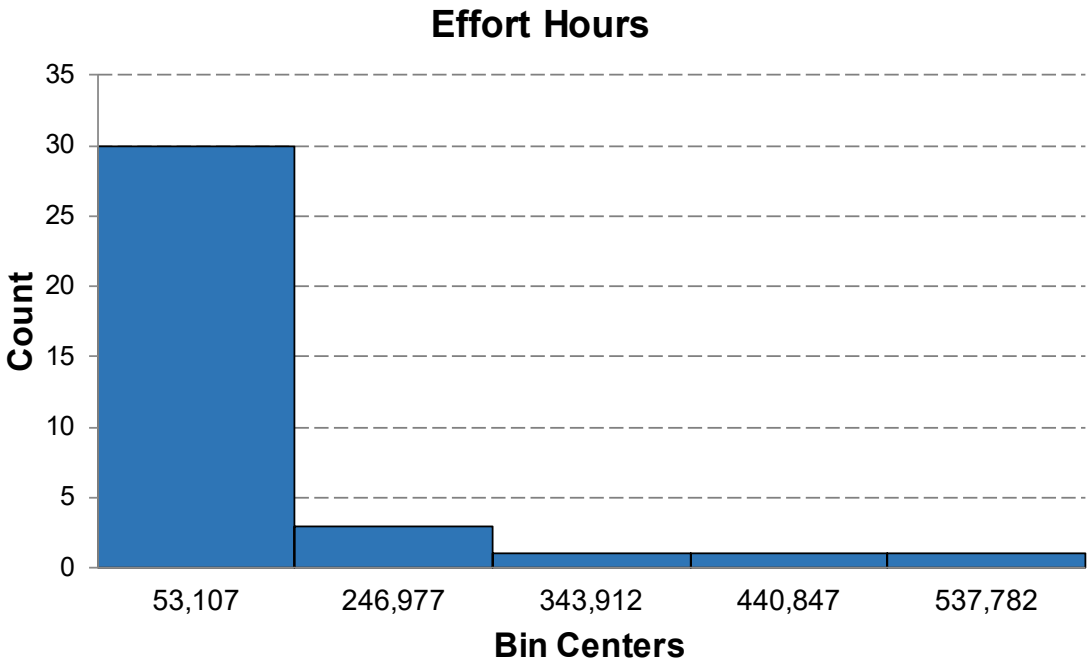
Dataset by Agile Framework



SAFe = scaled agile framework

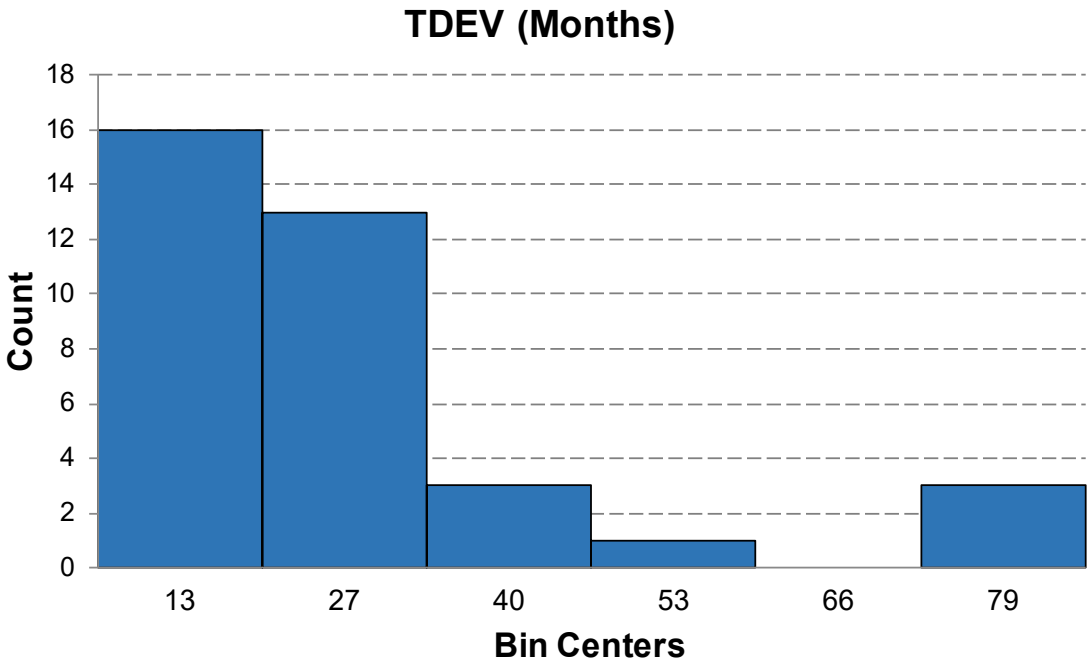
Scrum is the most common framework in DoD

Histogram: Actual Effort Distribution



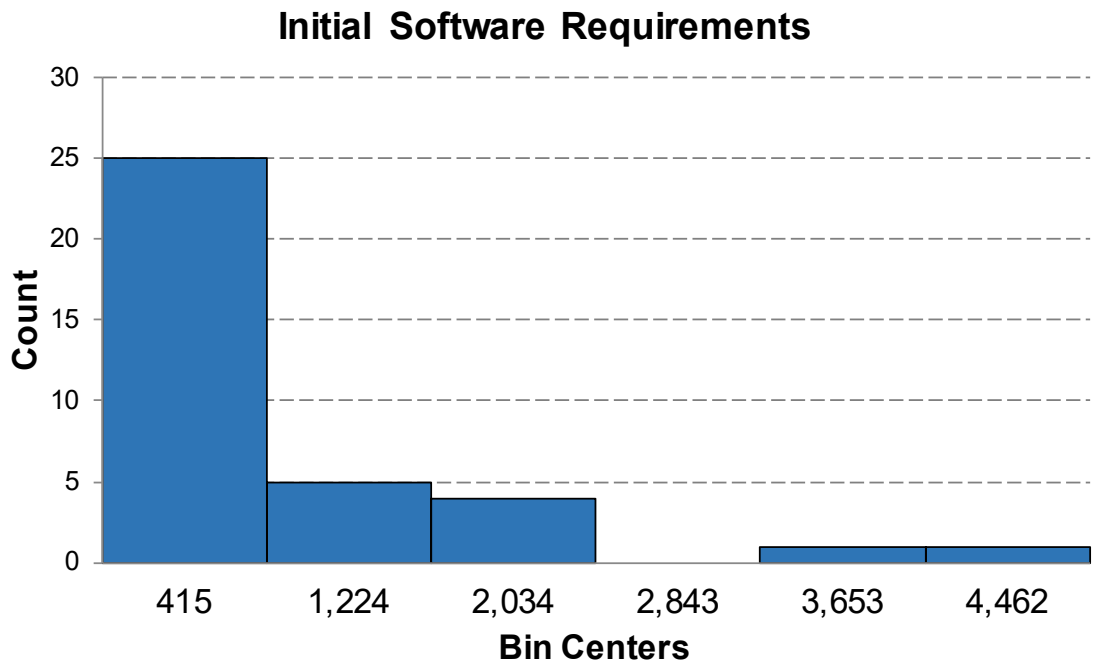
Average development effort for the agile project sample is 99,959 hours

Histogram: Actual Schedule Distribution



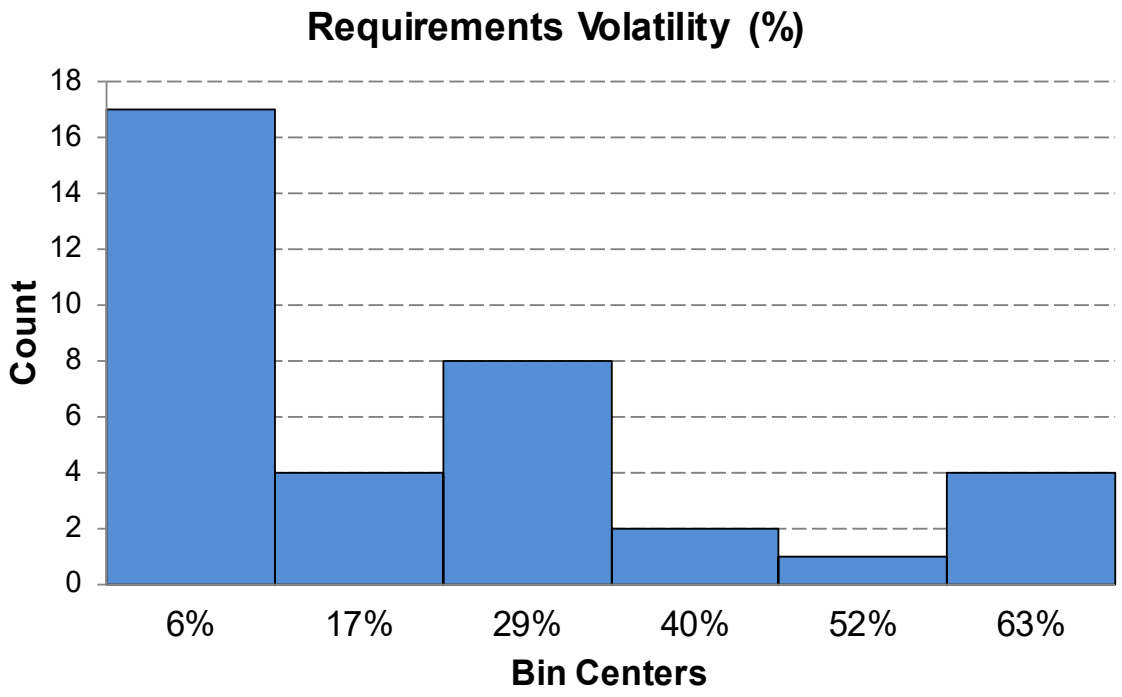
Average development time for the agile project sample is 26 months

Histogram: Initial Software Requirements



Average number of initial software requirements for the sample is 798

Histogram: Requirements Volatility (RVOL)



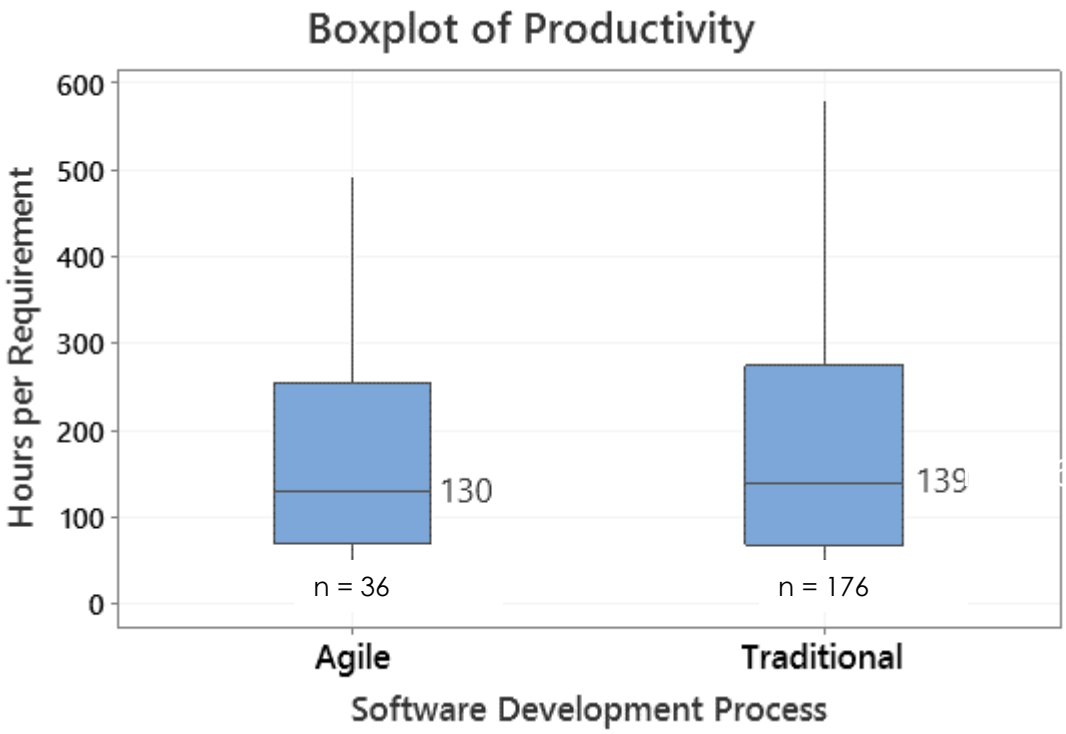
Average RVOL for the agile project dataset is 19%.

RVOL = requirements changes (added, modified, and deleted) divided by total initial software requirements

Agile vs Traditional Processes

Research Question:
How do Agile and traditional development processes compare for productivity, velocity, and cost overruns in the US DoD?

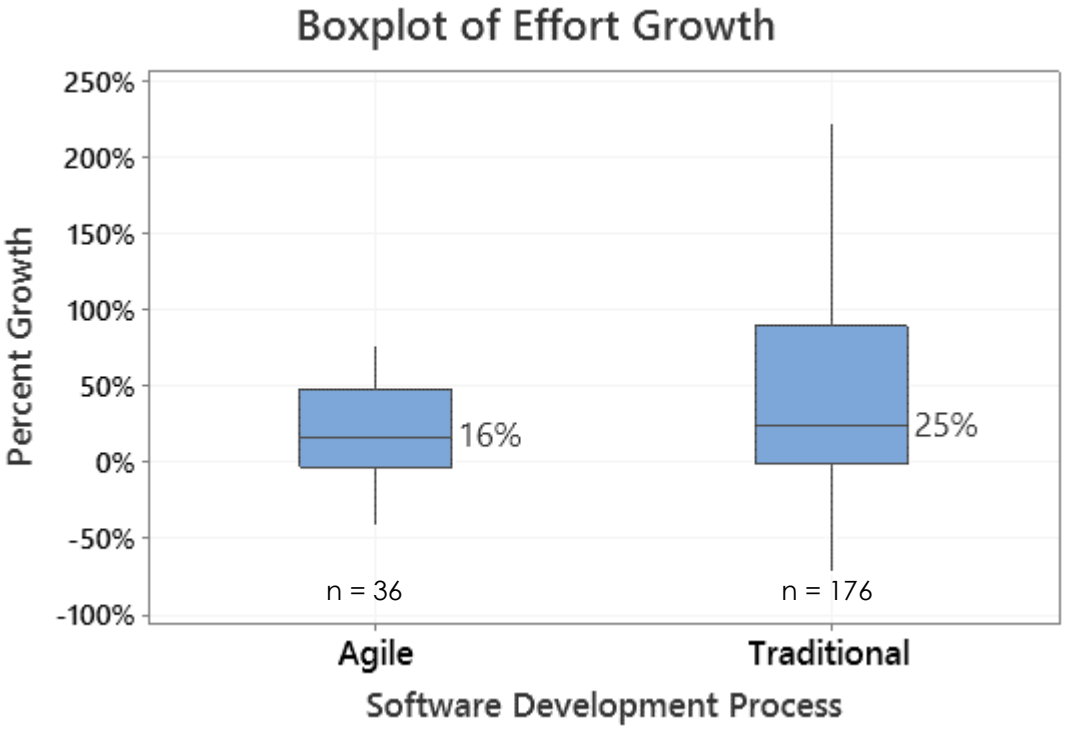
Comparison: Productivity (Median)



agile projects appear to take slightly less effort to develop a software functionality

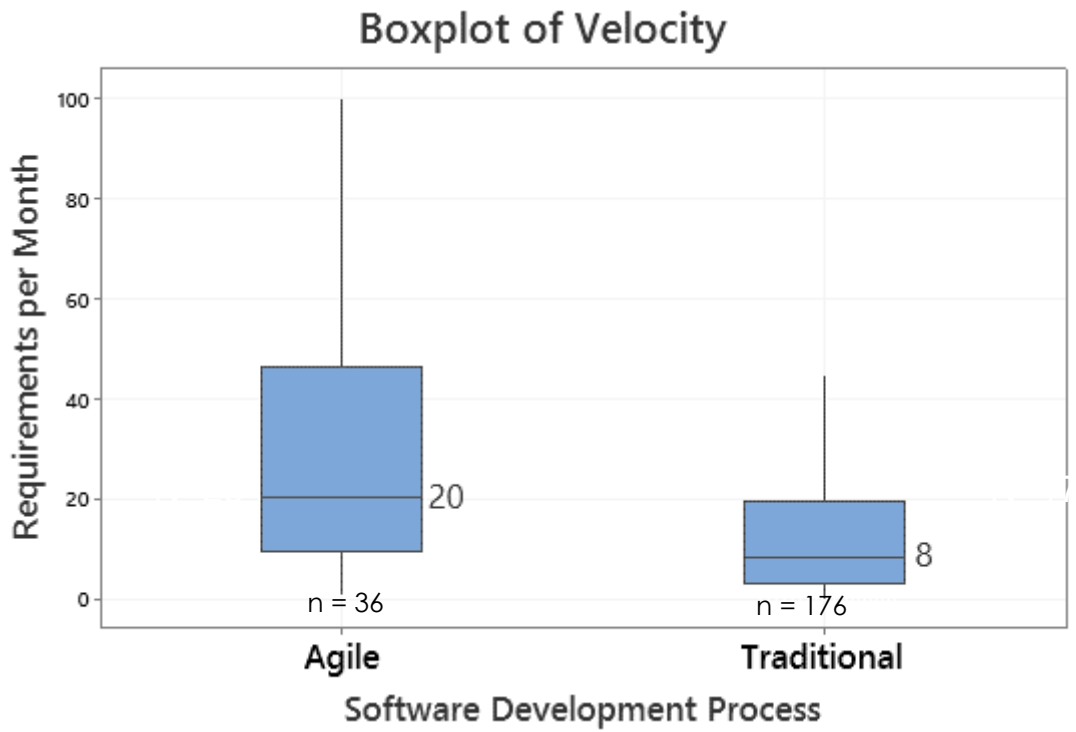
*Final Hours per Initial Software Requirement

Comparison: Effort Growth (Median)



- Agile software projects seem to experience less overruns than traditional

Comparison: Team Velocity (Median)



Agile Projects show a higher rate of progress than traditional

*software requirements per actual month

Effort Estimation Models: Agile Projects

Research Questions:

Do initial ,as opposed to final, software requirements* relate to final effort?

Do initial software requirements along with super domain relate to final effort?

Do initial software requirements along with initial peak staff and super domain relate to final effort?

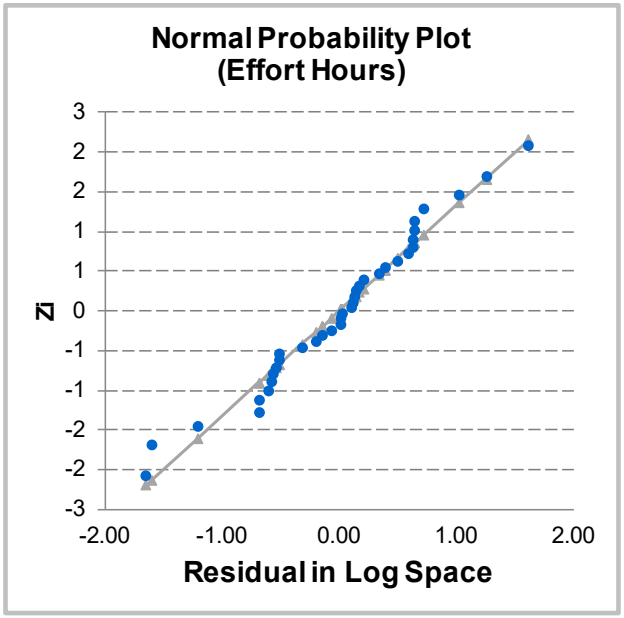
Effort Model 1: One Variable

Model	Equation Form	N	R^2	R^2 (adj)	R^2 (pred)	MMRE
1	$E = 1006 \times REQ^{0.65}$	36	64%	63%	60%	68%

E = Final Effort (in Hours) at contract completion

REQ = Initial Functional Requirements + Initial External Interfaces

Term	T-Statistic	P-value	VIF
Intercept	13.7	0.0000	***
REQ	7.8	0.0000	***



- REQ is strongly corrected to development effort
- However, R^2 (adj) value suggest adding variables to improve model reliability and accuracy

Effort Model 2: Two Variable

Model	Equation Form	N	R^2	R^2 (adj)	R^2 (pred)	MMRE
2	$E = 200 \times REQ^{0.718} \times (3.0^{D1}) \times (3.6^{D2}) \times (5.1^{D3})$	36	80%	77%	73%	47%

E = Final Effort (in Hours) at contract completion

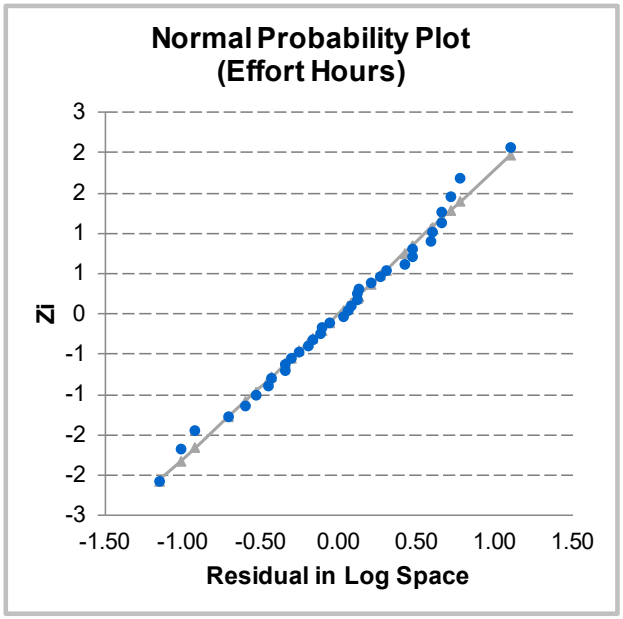
REQ = Initial Functional Requirements + Initial External Interfaces

D1 = 1 if Automated Information System, 0 otherwise

D2 = 1 if Engineering, 0 otherwise

D3 = 1 if Real-Time Embedded, 0 otherwise

Term	T-Statistic	P-value	VIF
Intercept	9.7	0.0000	
REQ	10.2	0.0000	1.2
D1	3.2	0.0028	2.5
D2	3.5	0.0013	2.7
D3	4.9	0.0000	2.8



- Effort Model shows better fit and higher accuracy when super domain is added
- Appropriate for early estimate prior to the release for proposal

Effort Model 3: Three Variable

Model	Equation Form	N	R^2	R^2 (adj)	R^2 (pred)	MMRE
3	$E = 173 \times REQ^{0.539} \times Staff^{0.463} (2.3^{D1}) \times (3.7^{D2}) \times (3.9^{D3})$	36	89%	87%	84%	34%

E = Final Effort (in Hours) at contract completion

REQ = Initial Functional Requirements + Initial External Interfaces

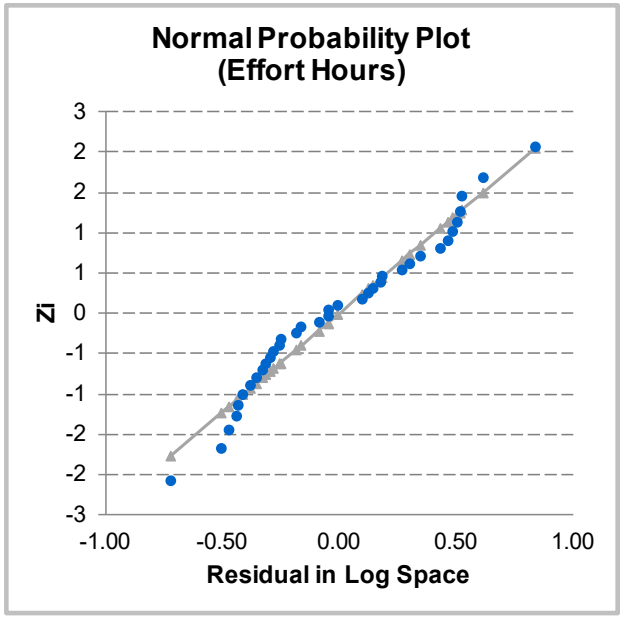
Staff = Initial (or Estimated) Peak Staff at contract start

D1 = 1 if Automated Information System, 0 otherwise

D2 = 1 if Engineering, 0 otherwise

D3 = 1 if Real-Time Embedded, 0 otherwise

Term	T-Statistic	P-value	VIF
Intercept	12.7	0.0000	
REQ	8.6	0.0000	1.7
Staff	5.2	0.0000	1.8
D1	3.3	0.0025	2.6
D2	4.9	0.0000	2.7
D3	5.5	0.0000	2.9



- Effort Model shows best fit and highest accuracy when all three variables are added
- Appropriate for assessing contract cost proposals

Schedule Estimation Models: Agile Projects

Research Questions:

Do initial ,as opposed to final, software requirements* relate to final duration?

Do initial software requirements along with super domain relate to final duration?

Do initial software requirements along with initial peak staff and super domain relate to final duration?

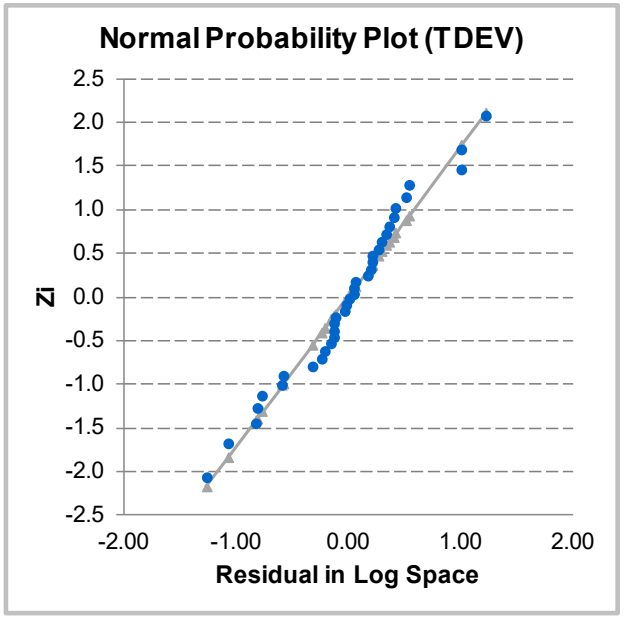
Schedule Model 4: One Variable

Model	Equation Form	N	R^2	R^2 (adj)	R^2 (pred)	MMRE
4	$TDEV = 6.8 \times REQ^{0.202}$	36	22	20	14	46

TDEV = Final Schedule (in Months) at contract completion

REQ = Initial Functional Requirements + Initial External Interfaces

Term	T-Statistic	P-value	VIF
Intercept	4.89	0.0000	***
REQ	3.16	0.0033	***



- REQ is strongly corrected to development duration (months)
- However, low R^2 (adj) value suggest segmenting by super domain to improve model reliability

Schedule Model 5: Two Variable

Model	Equation Form	N	R^2	R^2 (adj)	R^2 (pred)	MMRE
5	$TDEV = 1.6 \times REQ^{0.272} \times (2.1^{D1}) \times (2.9^{D2}) \times (4.0^{D3})$	36	69	65	59	30

TDEV = Final Schedule (in Months) at contract completion

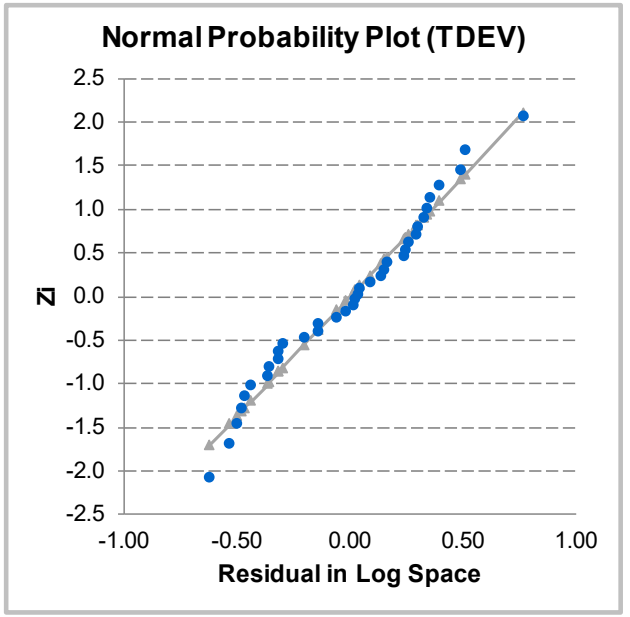
REQ = Initial Functional Requirements + Initial External Interfaces

D1 = 1 if Automated Information System, 0 otherwise

D2 = 1 if Engineering, 0 otherwise

D3 = 1 if Real-Time Embedded, 0 otherwise

Term	T-Statistic	P-value	VIF
Intercept	1.40	0.1724	
REQ	5.97	0.0000	1.2
D1	3.49	0.0015	2.5
D2	4.69	0.0001	2.7
D3	6.56	0.0000	2.8



- Schedule Model shows better fit and higher accuracy when Super Domain (D1, D2, D3) is added
- Appropriate for independent government estimates

Schedule Model 6: Three Variable

Model	Equation Form	N	R^2	R^2 (adj)	R^2 (pred)	MMRE
6	$TDEV = 1.7 \times REQ^{0.34} \times Staff^{-0.19} (2.3^{D1}) \times (3.0^{D2}) \times (4.5^{D3})$	36	75%	70%	63%	27%

TDEV = Final Schedule (in Months) at contract completion

REQ = Initial Functional Requirements + Initial External Interfaces

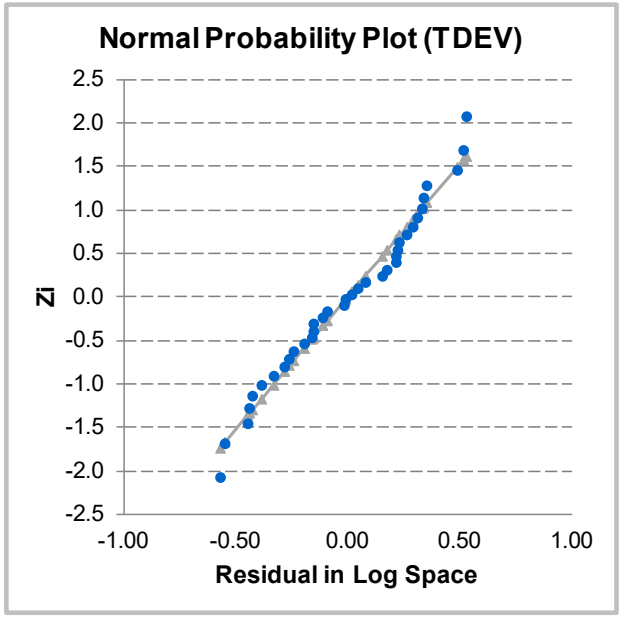
Staff = Initial (or Estimated) Peak Staff at contract start

D1 = 1 if Automated Information System, 0 otherwise

D2 = 1 if Engineering, 0 otherwise

D3 = 1 if Real-Time Embedded, 0 otherwise

Term	T-Statistic	P-value	VIF
Intercept	1.7	0.0986	
REQ	6.9	0.0000	1.7
Staff	-2.6	0.0135	1.8
D1	4.2	0.0002	2.6
D2	5.1	0.0000	2.7
D3	7.5	0.0000	2.9



- Schedule Model shows best fit and highest accuracy when all three variables are added
- Appropriate for assessing realism of cost proposals

Conclusion

Summary of findings

- ✓ Results confirmed the notion that Initial, as opposed to final, functional and external interface requirements, when treated as primary size input along with super domain categorical variable, proved to be effective in predicting development effort and schedule for agile projects at early lifecycle in DoD
- ✓ Approach is pragmatic as mainstream agile sizing metrics are not available for early estimation in the DoD.