Naval Center for Cost Analysis (NCCA)

Modeling with Gumby: Pros and Cons of the Weibull Curve

Jake Mender and Ann Hawpe

07 JUN 2017

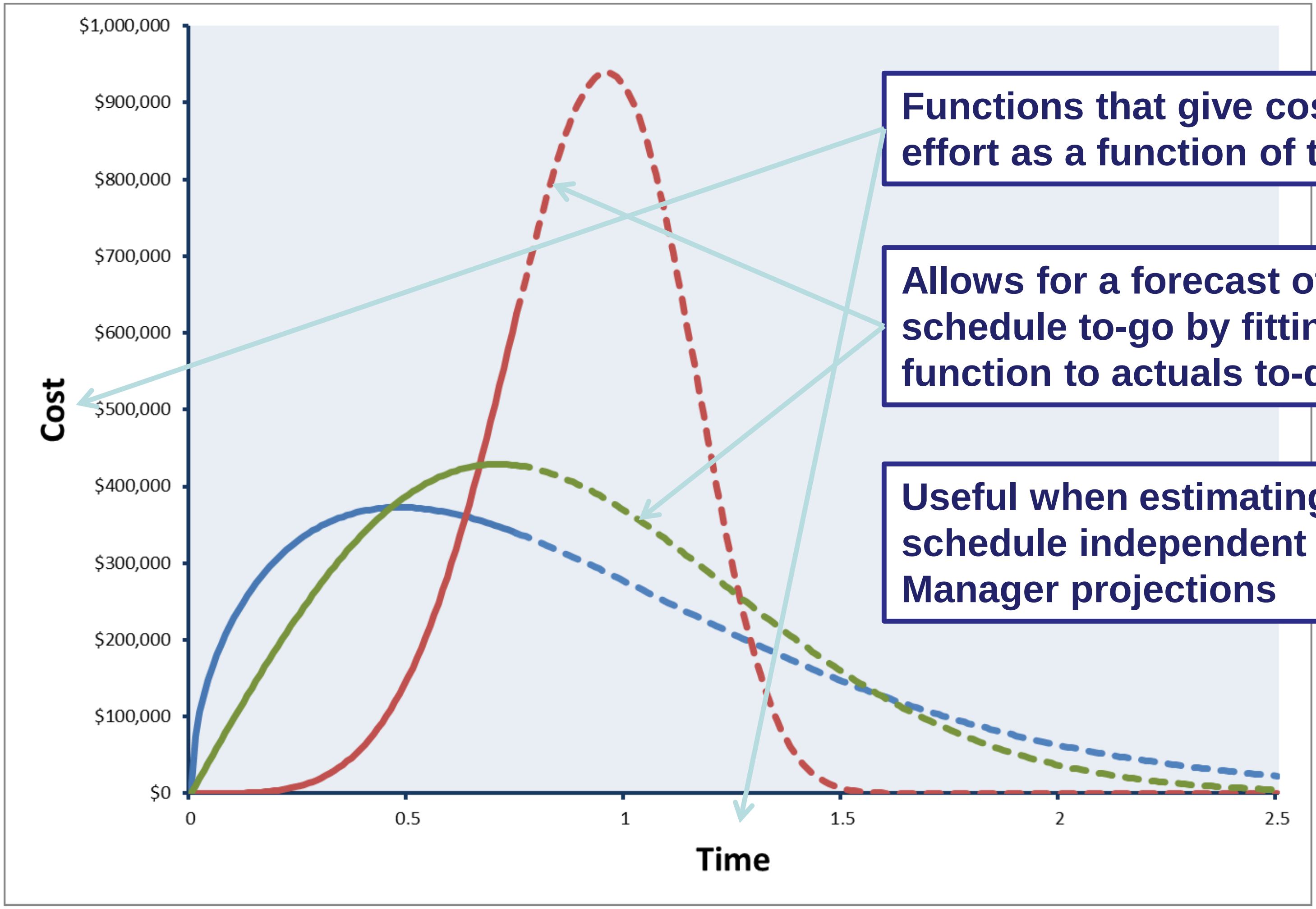
Overview

- Motivation
- Parametric Phasing Curve Overview
- Overfitting
- Rayleigh Function Details
- Rayleigh vs. Weibull
- Problem with Rayleigh
- Solutions
- Take Aways
- Future Study

Motivation

- NCCA was asked to evaluate the use of the Weibull curve, instead of the commonly used Rayleigh curve, as a tool for **early R&D project** estimating
- Our review highlighted some potential pitfalls when using the Weibull for **predictive purposes** with small data sets
- Results are relevant to other modeling problems in the cost estimating domain

Parametric Phasing Curves



Functions that give cost, hours, or effort as a function of time elapsed

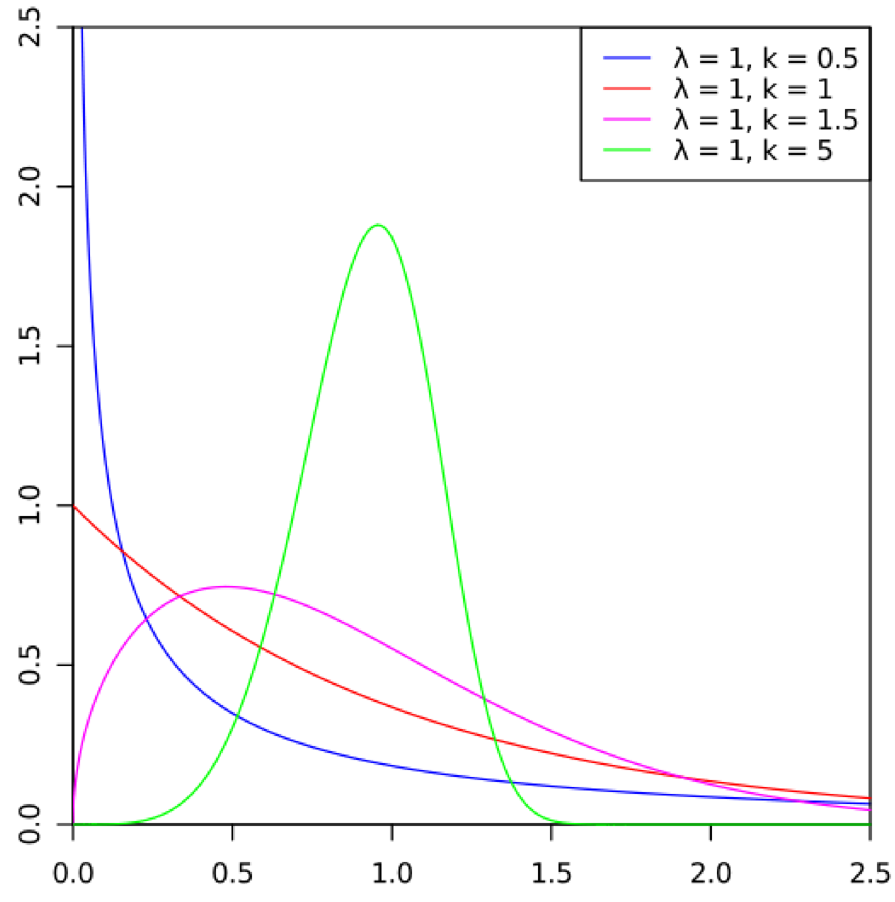
Allows for a forecast of effort and schedule to-go by fitting the chosen function to actuals to-date (by period)

Useful when estimating cost or schedule independent of Program Manager projections

Weibull Distribution

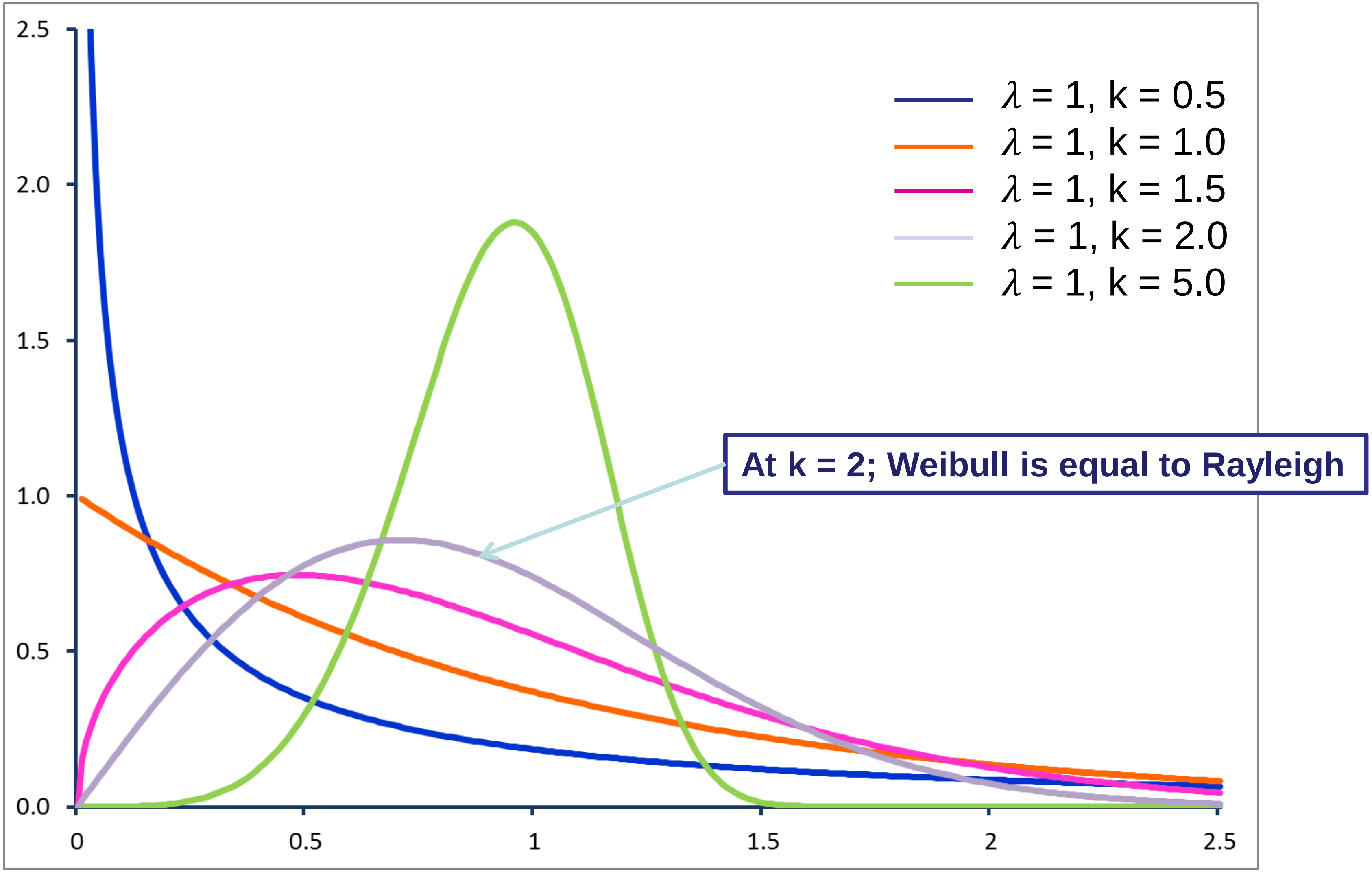
- $f(x) = \frac{\kappa}{\lambda}\left(\frac{x-\mu}{\lambda}\right)^{(\kappa-1)} e^{\left(-\left(\frac{x-\mu}{\lambda}\right)^{\kappa}\right)}$ $x \geq \mu$; $\lambda, \kappa > 0$
  - (κ): shape
  - (μ): location
  - (λ): scale

- Three parameters result in a function that is extremely flexible (like Gumby)
  - Pro: Can fit wide range of projects
  - Con: 30-45 data points required to avoid overfitting



Most popular solution is to use the Rayleigh function, a degenerate of the Weibull

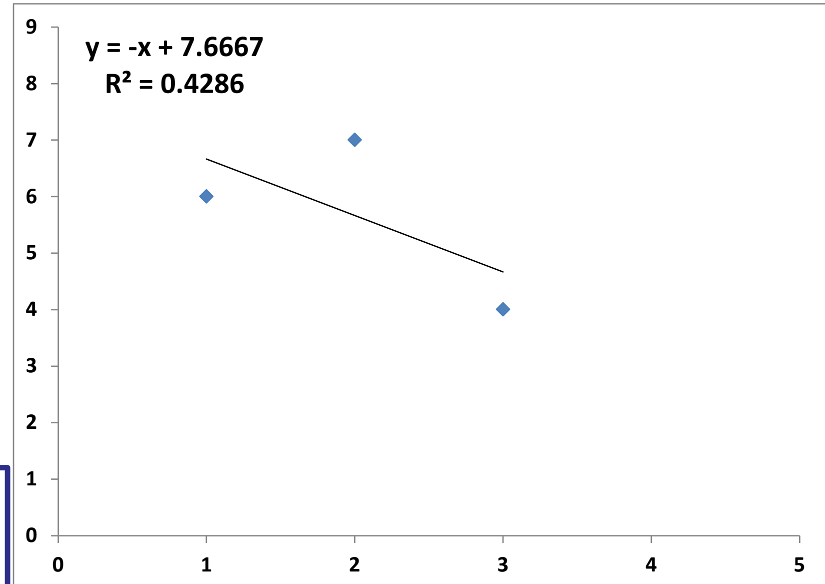
Flexibility of the Weibull Curve

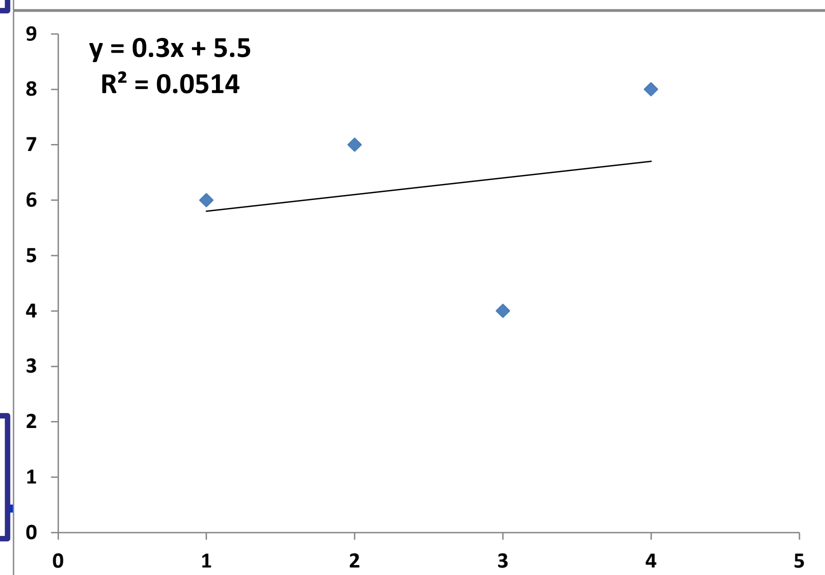


Overfitting

- When a model is too complex for the data set
- An overfitted model will poorly predict performance outside training data set

Is a linear model (2 parameters) accurate with 3 or 4 data points?

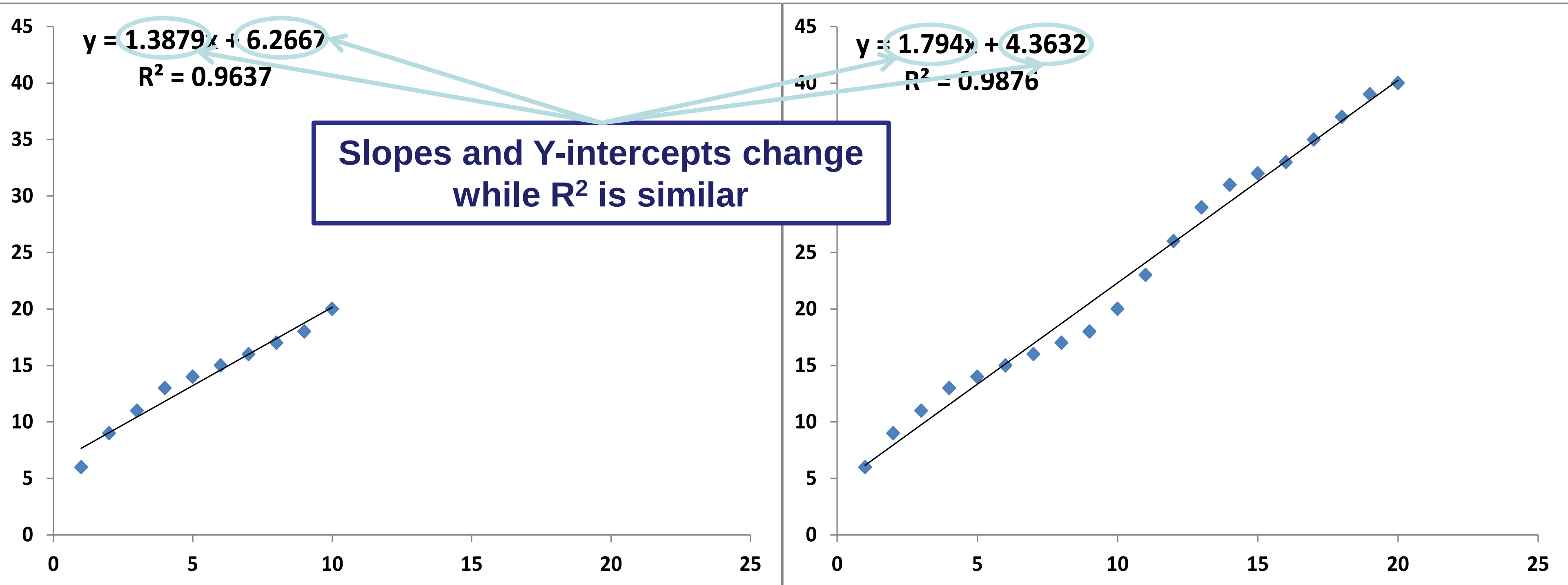




- Rule of thumb: 10-15 data points per parameter
- Defense R&D projects rarely have 30-45 data points available

Weibull model (with 3 parameters) requires a lot of data (30-45)

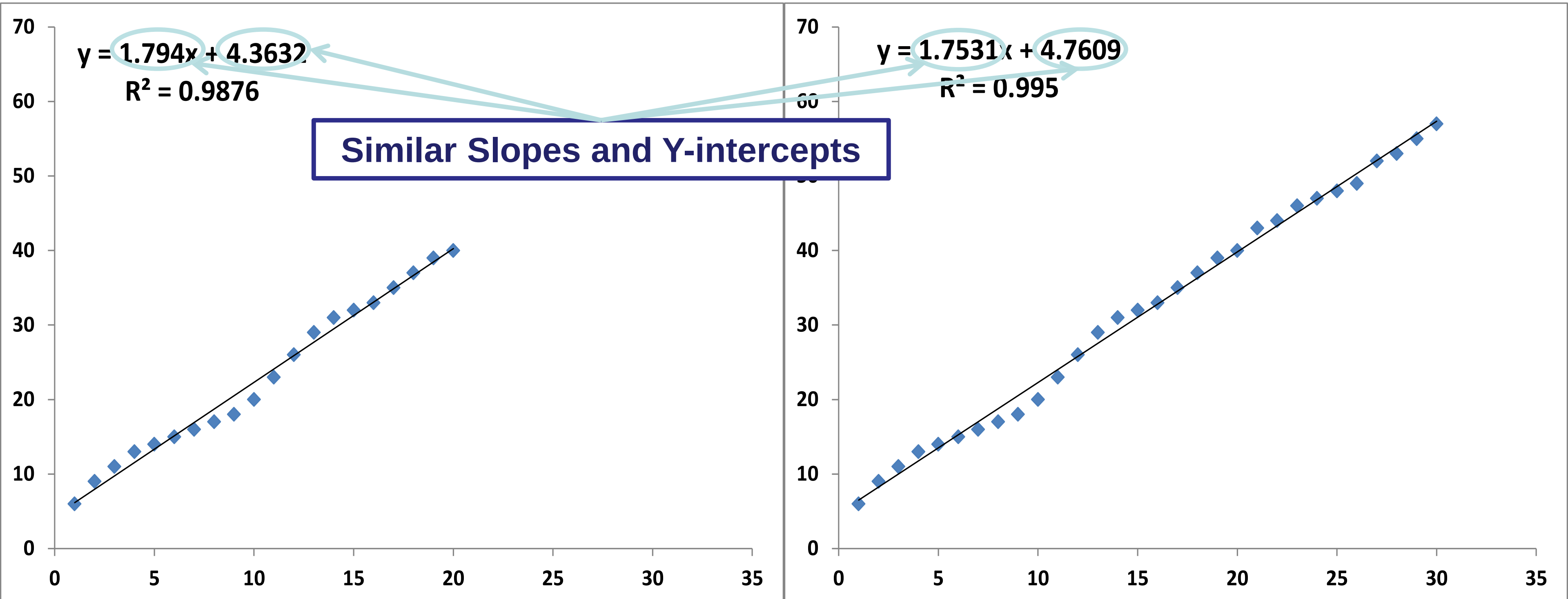
Overfitting: Part 2



10 points → Estimates change greatly → 20 points

Applying rule: linear model requires 20-30 points before parameter estimates will stabilize around the population mean

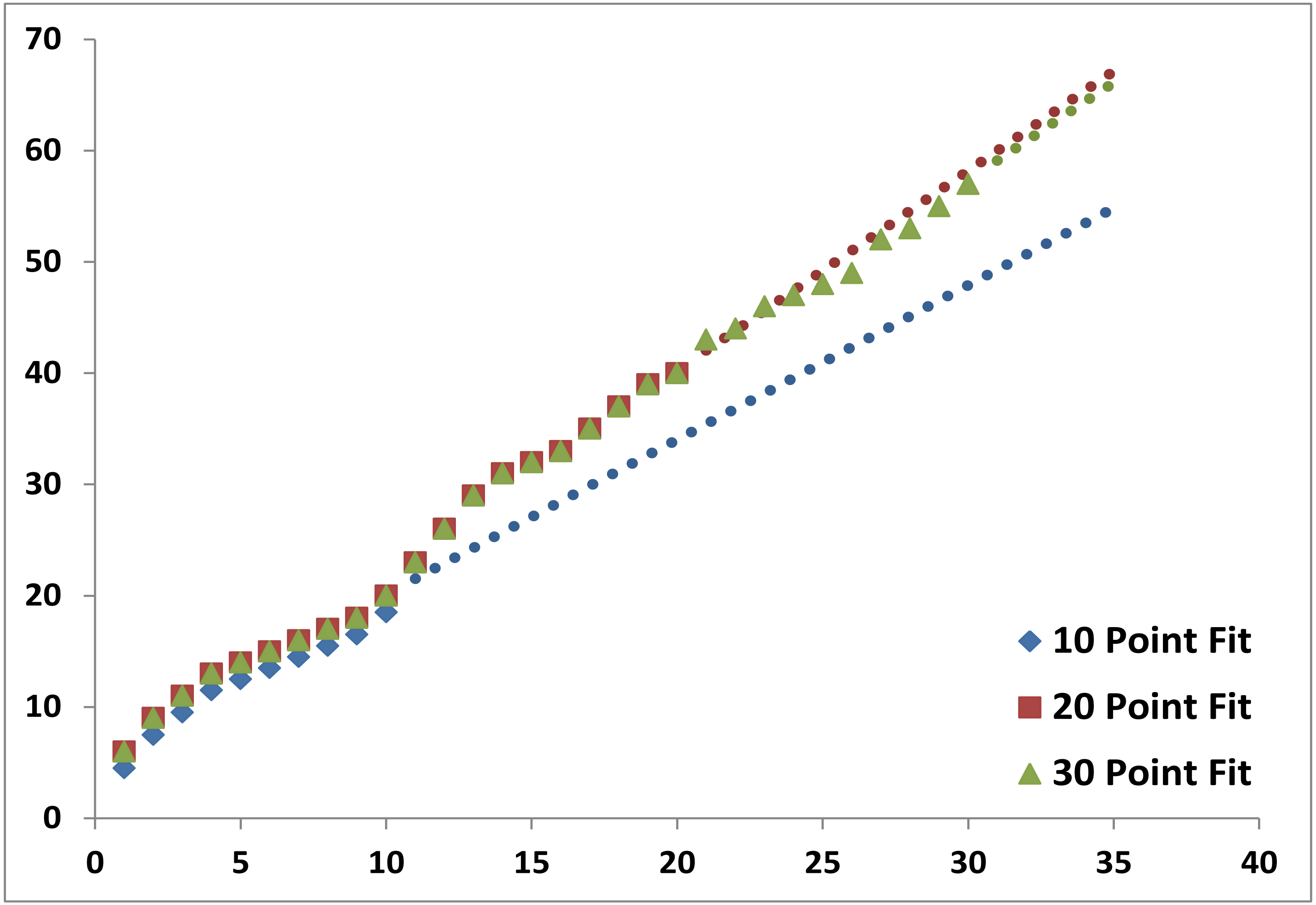
Overfitting: Part 3



20 points → Estimates stable → 30 points

Weibull model (with 4 parameters) requires a lot of data (40-60)

Overfitting: Part 4



Predictions stabilize once sufficient data is available

Overfitting: Solutions

Three ways to deal with this:

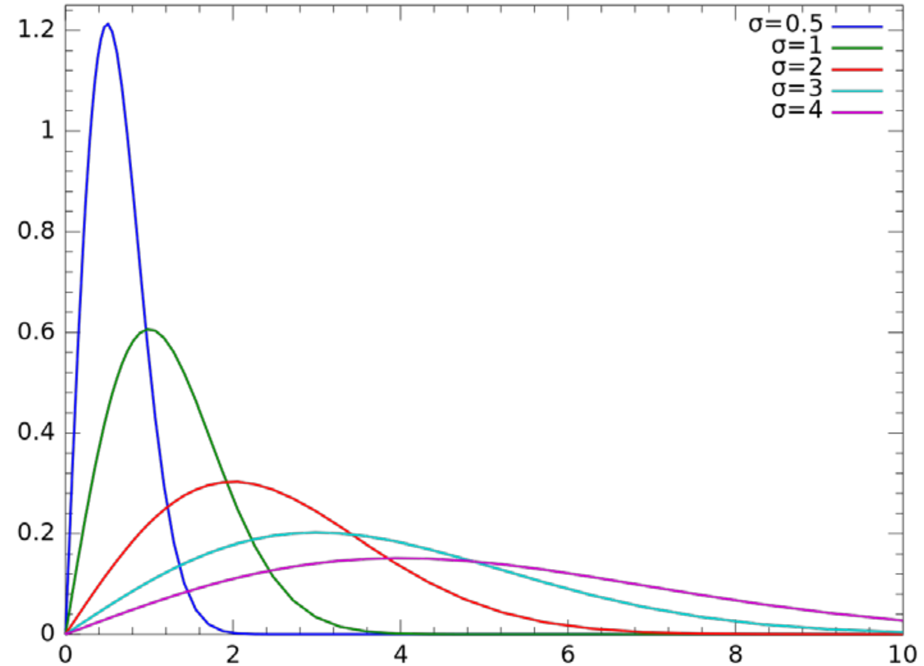
1. Use Weibull degenerates (like the Rayleigh) with fewer parameters
   - Reduce the amount of required data

2. Estimate parameters from other features of the project (AKA, get more data)
   - “Create” more data by considering other information

3. Use a different method
   - Find another method that isn’t as data hungry

The Rayleigh Curve

- Developed from Manpower Utilization model developed by P.V. Norden in the 1960s
- Pattern is approximated by Rayleigh Function
- If true, allows for total effort and duration to be estimated from the trajectory of early effort
  - For our purposes this is conveyed by ACWP as reported in EVM CPRs



Rayleigh simplifies the Weibull by <u>assuming</u> the distribution mode is always at the 39th percentile

Rayleigh is a degenerate (or restricted version) of the Weibull

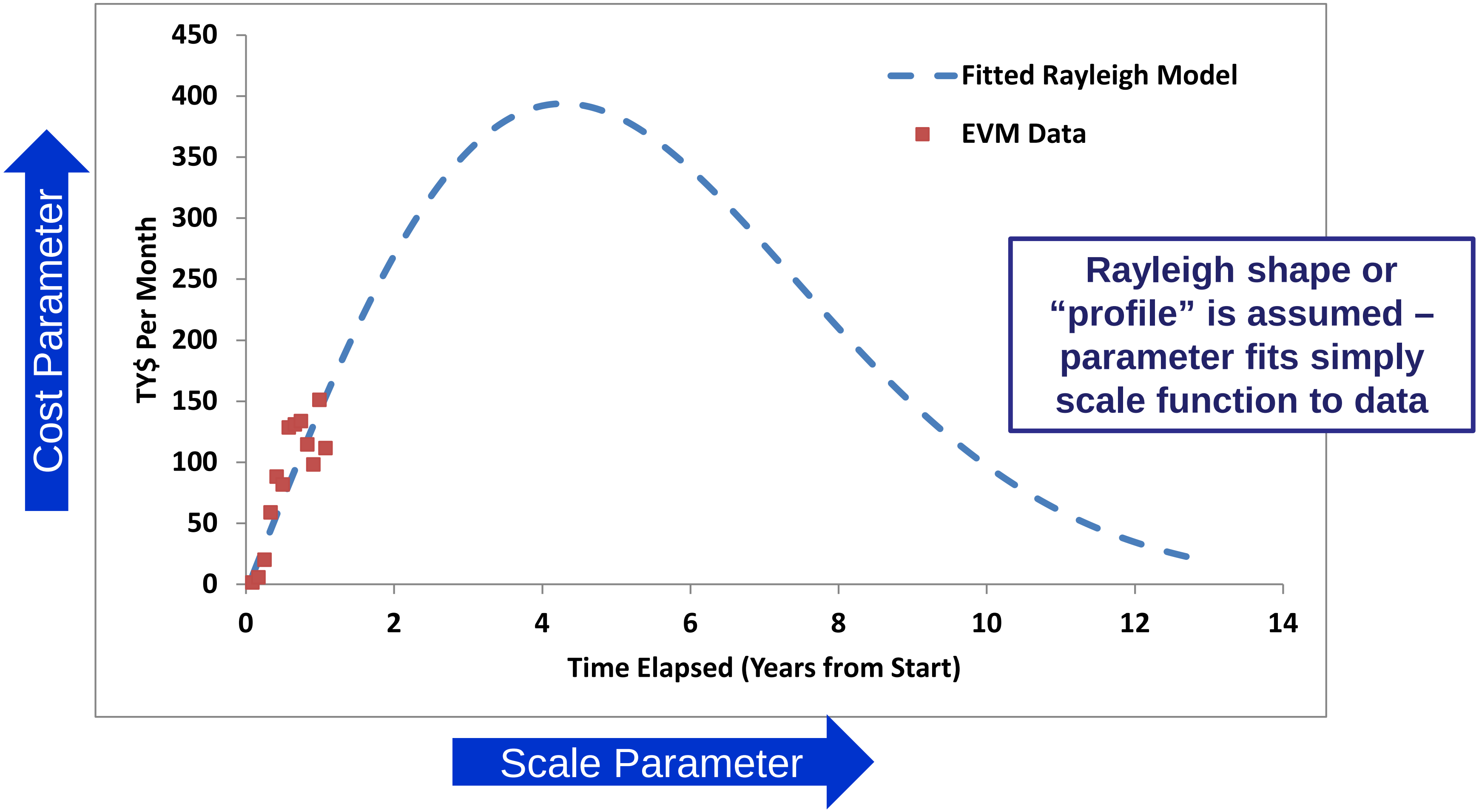
Rayleigh Function Basics

- Restricted version of the Weibull
  - 2 parameters set to fixed value (restricted)
  - 1 parameter added, a cost scalar (d)

1. Effort completed at time (t) is given by the Rayleigh function
2. Rayleigh function defined by CDF:
$$Cumulative\ Effort\ (t) = C(t) = k * \left(1 - e^{-\alpha * t^2}\right)$$
3. Taking derivative of CDF gives PDF:
$$Change\ in\ Effort\ (t) = c(t) = -\alpha * 2 * k * t * e^{-\alpha * t^2}$$
4. Parameter definitions:

t = time elapsed since contract start
α = Rayleigh shape parameter (related to duration)
k = Rayleigh scale parameter (related to Final Cost)

Rayleigh Model



By assuming a fixed shape parameter, we are scaling the cost and duration of the fixed Rayleigh profile to find the combination of the two that best fits our data

Rayleigh vs. Weibull

Demo to be done in Excel

The Problem

- Restricted models, like the Rayleigh, require a major assumption that projects have the Rayleigh profile
- Evidence is mixed on this assumption
- Rayleigh has predictive power because of the assumed profile
- Weibull lacks predictive power because it can take on a variety of forms
  - An effective heuristic to determine which Weibull fits the problem at hand is needed

Possible Solution

- Rayleigh method assumes a profile
- Instead of assuming a profile, why not use two-stage model to predict profile, and then predict cost and duration?
- Won't work off existing CPR data
- Could work if additional independent information about the project allowed for classification

NCCA hasn't done this yet, but other people have tried…

Air Force Institute of Technology (AFIT) Weibull Model

Attempted to address the overfitting problem via a two-stage model

Step 1: Develop regression using program characteristics to predict best Shape and Scale parameters

Shape Model Summary of Fit

RSquare	0.310116
RSquare Adj	**0.274185**
Root Mean Square Error	0.763702
Mean of Response	2.724529
Observations (or Sum Wgts)	102

Shape Model Analysis of Variance (ANOVA)

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	25.169127	5.03383	8.6308
Error	96	55.991124	0.58324	**Prob > F**
C. Total	101	81.160251		**<.0001**

Shape Model Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.299561	0.32514	4.00	0.0001
ln(1/Duration)	-0.973254	0.16037	-6.07	<.0001
Army	-0.423434	0.20643	-2.05	0.0430
Navy	-0.485661	0.18882	-2.57	0.0116
Electronic	-0.545079	0.18152	-3.00	0.0034
Space	-1.100189	0.56290	-1.95	**0.0536**

Scale Model Summary of Fit

RSquare	0.921671
RSquare Adj	**0.920888**
Root Mean Square Error	0.824422
Mean of Response	5.854373
Observations (or Sum Wgts)	102

Scale Model Analysis of Variance (ANOVA)

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	799.75149	799.751	1176.672
Error	100	67.96724	0.680	**Prob > F**
C. Total	101	867.71873		**<.0001**

Scale Model Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0683049	0.187391	0.36	0.7163
Duration	0.7256199	0.021153	34.30	**<.0001**

Step 2: Fit a Weibull model using estimated parameters and optimizing for budget parameter

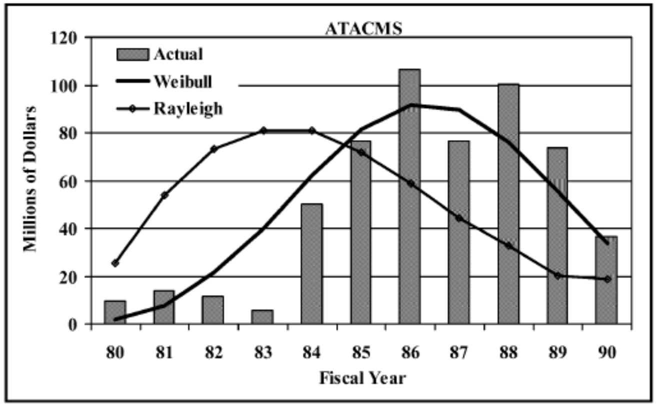
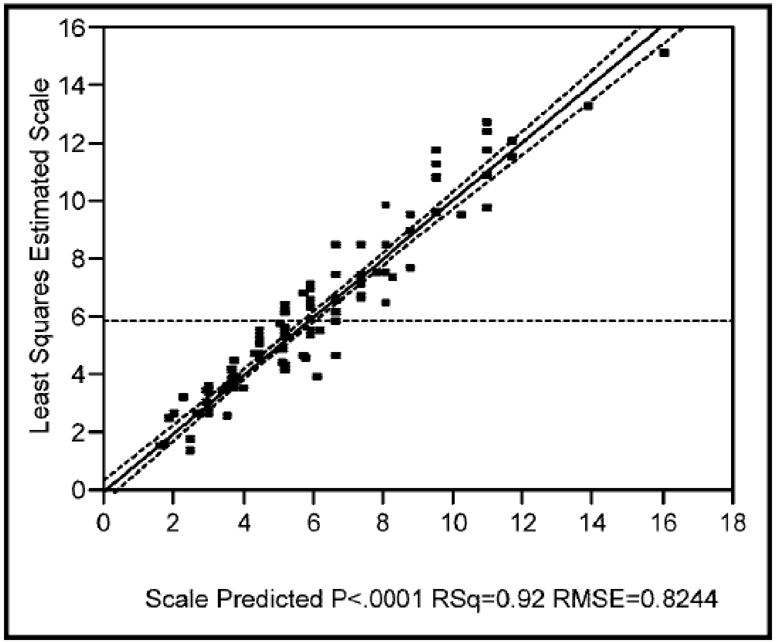


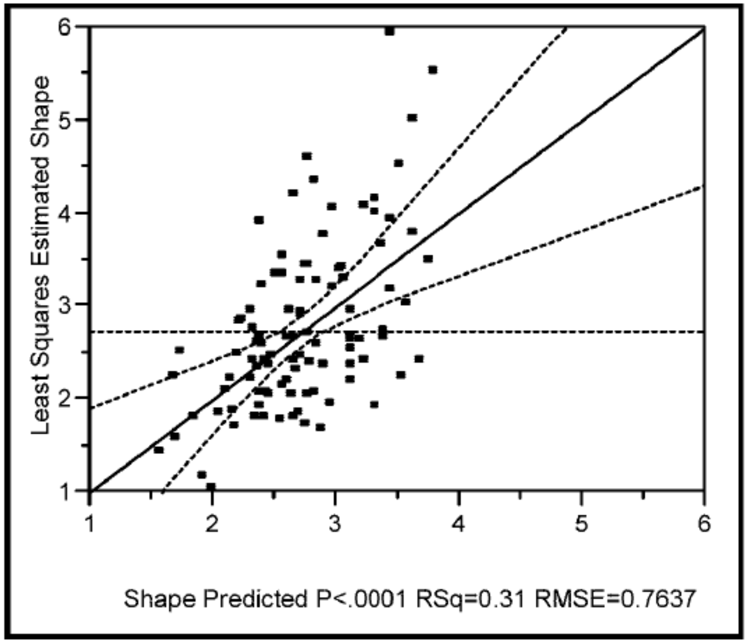
Figure 22. Rayleigh and Weibull Model vs. Actual ATACMS Budget Profile

See: "Forecasting Research and Development Program Budgets Using the Weibull Model" Captain Thomas W. Brown USAF, March 2002, Air Force Institute of Technology

AFIT Model: Results

Predict Scale and Shape as a function of service, system type, total cost, and duration





Results: success predicting scale (and to a lesser extent Shape) – Forecasted Weibull models correlated well to project expenditures, but the associated duration and cost estimates were not evaluated for accuracy

NCCA Evaluation Criteria

- In DoD use, phasing curves are used to make specific predictions regarding three things:
  1. Profile: the period by period expenditures
  2. Duration: The total amount of time required
  3. Total Cost: The total amount required to complete the project
- Any proposed Phasing Model should be evaluated in all 3 dimensions before use

AFIT model appears promising, but was not fully evaluated for this use case

Summary

- Parametric phasing models can be an effective tool for independent evaluation of R&D projects
- Estimators must be careful when using complex models early in a project
  - Complex models require larger amounts of data
- Rayleigh model is well supported by a large body of research, but has known limitations
- The more complex Weibull shows promise, but requires further research regarding the overfitting problem

Analysts must think critically regarding their use case and ensure the model selected is appropriate for the problem at hand

Future Study

- Evaluate AFIT model in the context of the early R&D forecasting use case
- Explore alternate predictors for shape, scale, and location parameters (beyond what was considered in the AFIT study)

Questions?