Naval Center for Cost Analysis (NCCA)

Validation and Improvement of the Rayleigh Curve Method

10-13 JUN 2014

POC: Mr. Jake Mender
Ship & Ship Weapons Branch
Naval Center for Cost Analysis
Michael.Mender@navy.mil

Overview

1. Background and Motivation
2. Parameter Estimation
3. Evaluating Rayleigh Method
4. Conclusions

Background

Rayleigh Hypothesis

- Theory that effort on a project follows a standard pattern
  - Pattern is approximated by Rayleigh Function
  - Developed from Manpower Utilization model developed by P.V. Norden[1] in the 1960s
- If true, allows for total effort and duration to be estimated from the trend of early data
  - For our purposes this is conveyed by ACWP as reported in EVM CPRs

1. Norden, P.V., "Useful Tools for Project Management," *Operations Research in Research and Development*, B.V. Dean, Editor, John Wiley and Sons, 1963

Rayleigh Function Basics

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)

Motivation

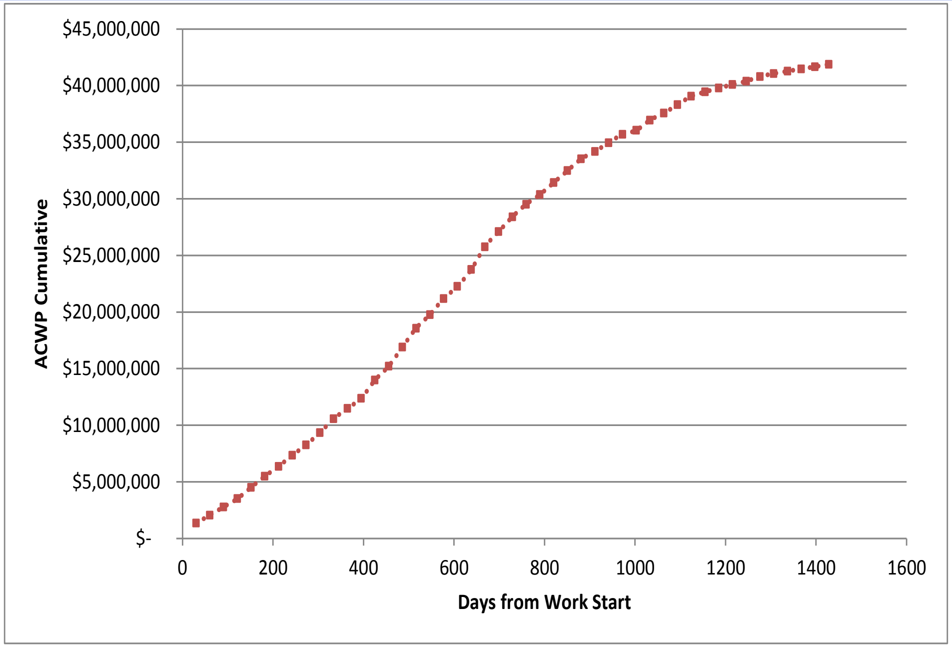
1. Rayleigh is popular, but vintage support
   – NCCA does not have access to original data set

2. Desire to validate and verify theory
   – Does EVM follow Rayleigh “path” ?
   – Does theory still hold for current contracts?
   – How accurate is it?
   – Are there any pitfalls analysts should be aware of?

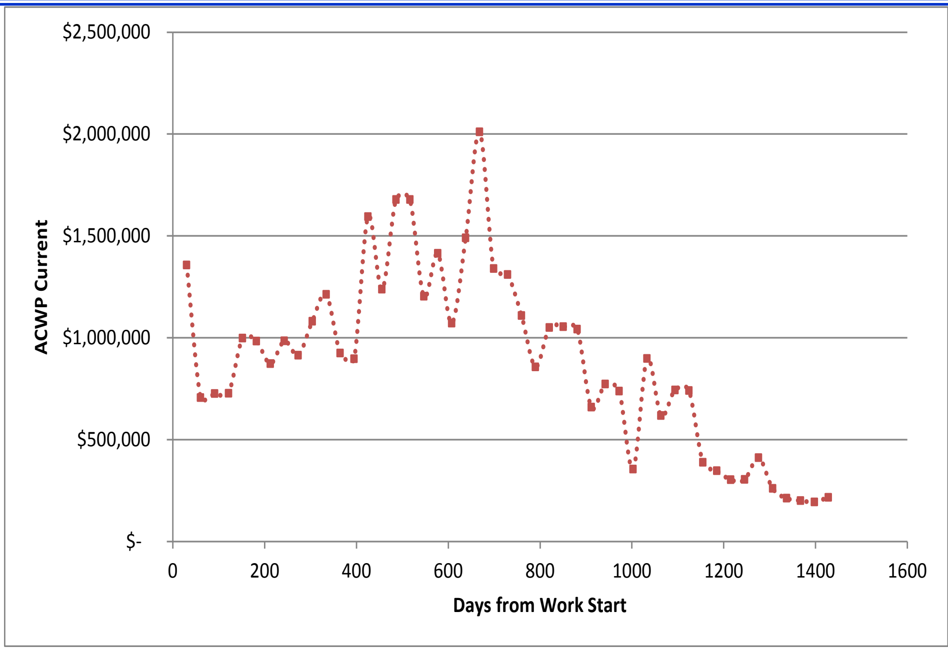
Example EVM Data

Report Date	Days From Start	ACWP Cumulative	ACWP Current	Estimated Completion Date (ECD)	Estimate at Completion (EAC)
6/25/2020	30	$ 3,737,226	$ 3,737,226	12/27/2023	44,862,882
7/26/2020	61	$ 5,682,668	$ 1,945,442	12/27/2023	44,862,882
8/26/2020	92	$ 7,683,822	$ 2,001,154	12/27/2023	47,291,764
9/25/2020	122	$ 9,687,672	$ 2,003,850	12/27/2023	62,556,969
10/26/2020	153	$ 12,435,553	$ 2,747,881	12/27/2023	71,045,926
11/25/2020	183	$ 15,144,794	$ 2,709,241	12/27/2023	71,045,926
12/26/2020	214	$ 17,548,516	$ 2,403,722	12/27/2023	71,142,074
1/26/2021	245	$ 20,261,352	$ 2,712,836	12/27/2023	72,469,288
2/23/2021	273	$ 22,780,991	$ 2,519,639	12/27/2023	73,054,269
3/20/2021	298	$ 25,757,113	$ 2,976,122	12/27/2023	79,993,162
4/20/2021	329	$ 29,099,859	$ 3,342,746	12/27/2023	109,207,141
5/20/2021	359	$ 31,647,355	$ 2,547,496	12/27/2023	109,207,141
6/19/2021	389	$ 34,117,572	$ 2,470,217	12/27/2023	111,012,404
7/24/2021	424	$ 38,510,766	$ 4,393,195	12/27/2023	111,012,404
7/27/2021	427	$ 41,920,008	$ 3,409,241	12/27/2023	113,618,308
8/27/2021	458	$ 46,542,342	$ 4,622,334	12/27/2023	113,618,308
9/26/2021	488	$ 51,164,676	$ 4,622,334	12/27/2023	114,752,325

Plot of ACWP vs. Time



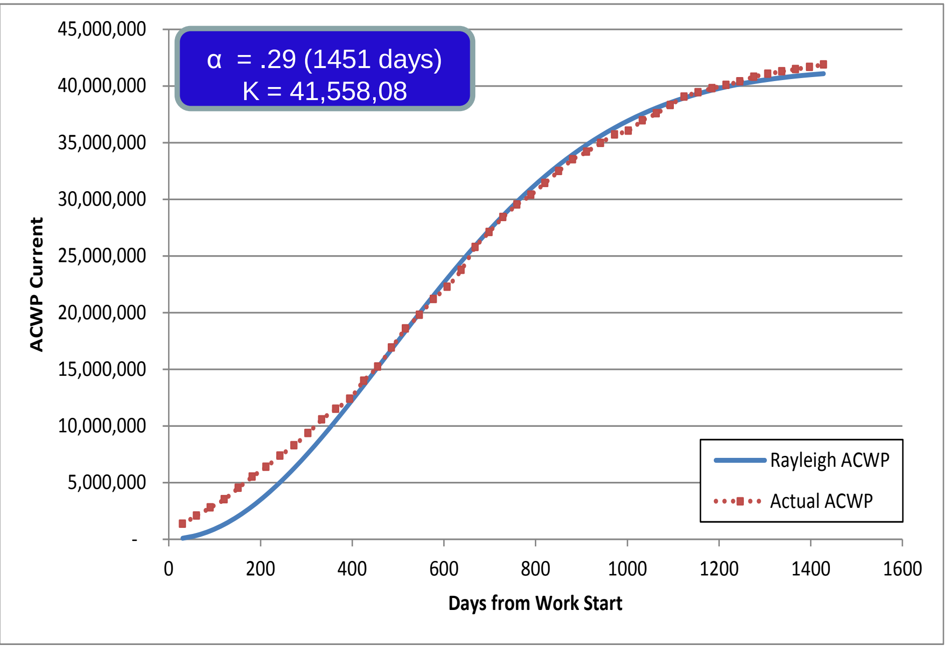
Cumulative ACWP

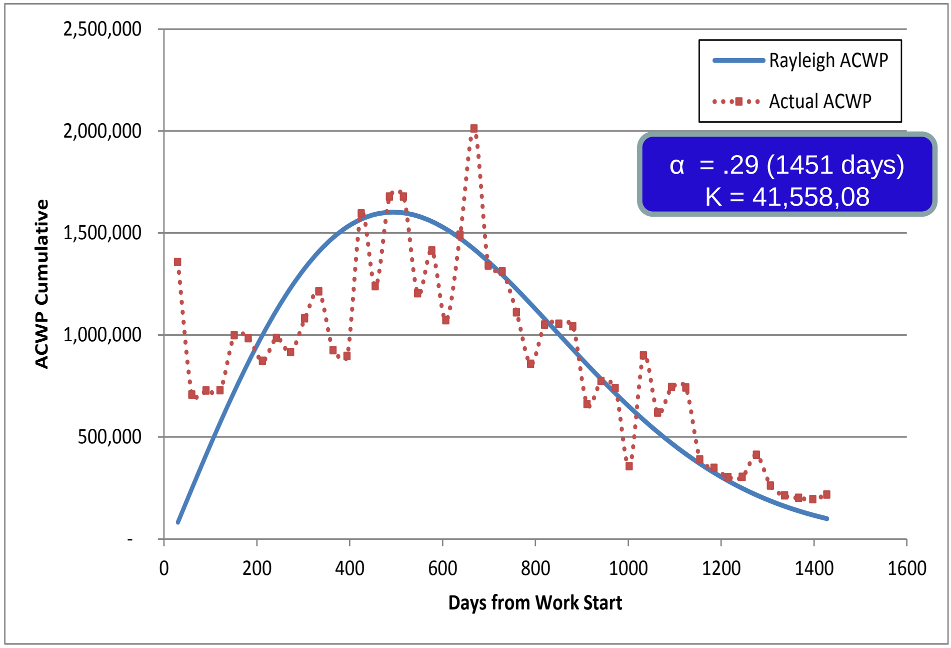


Current ACWP

Plotting EVM data with respect to time reveals trends

Data vs. Function





Cumulative Distribution Function

$$C(t) = K * \left(1 - e^{-\alpha * t^2}\right)$$

Probability Distribution Function

$$c(t) = -\alpha * 2 * K * t * e^{-\alpha * t^2}$$

Estimating Rayleigh Parameters allows estimation of final cost and duration – assuming effort is Rayleigh distributed

Parameter Estimation

Multiple Options

1. Optimization
2. Linear Transform and Regression
3. Method of Moments
4. Maximum Likelihood
5. Bayesian Methods

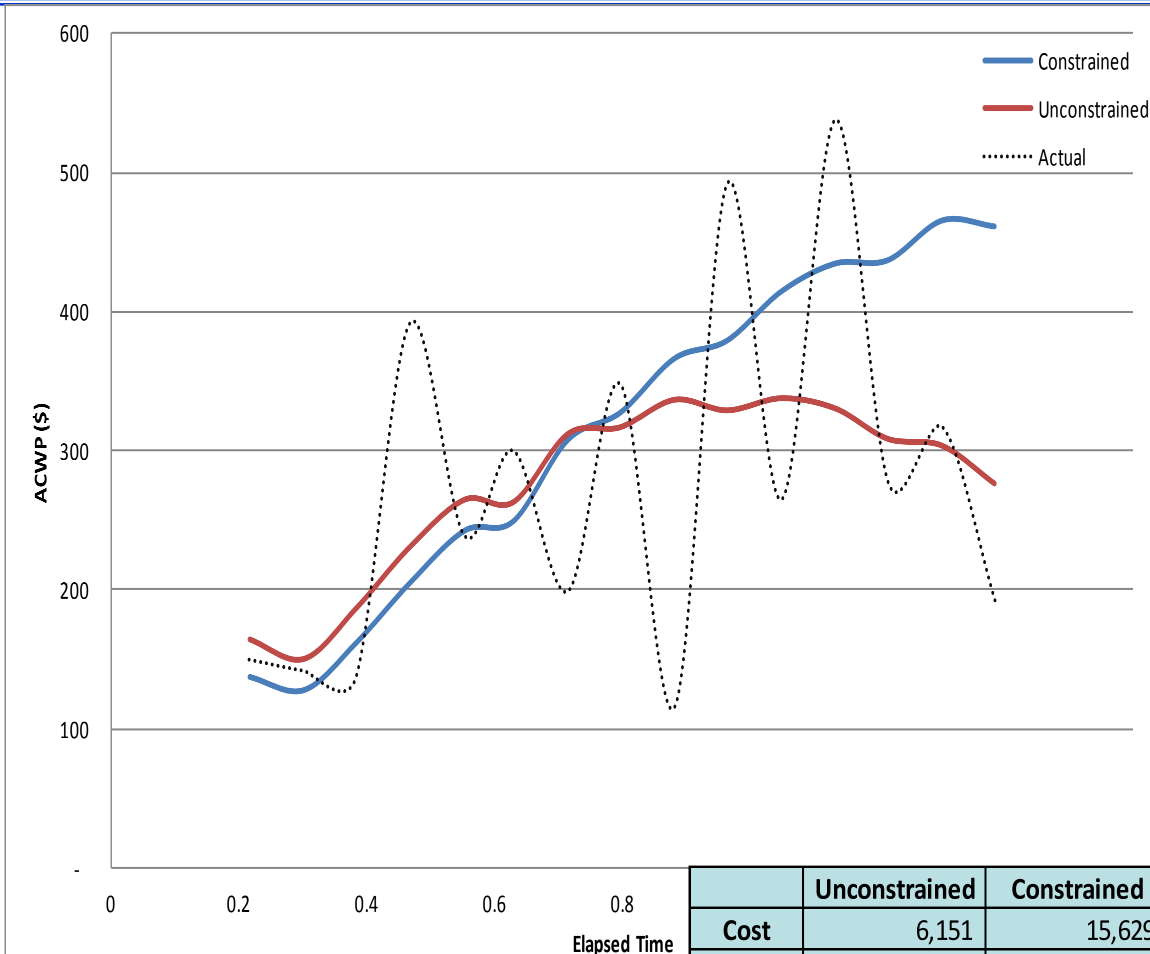
Today’s focus is on Linear Transform and Regression, prior work analyzed Optimization

Non-Linear Optimization

1. Appealing because it's simple to set up
2. Computationally complex, but easy to do with modern technology
3. Main issue is that standard measures of goodness-of-fit cannot be used
4. Has form:
   - $Minimize \sum_{i=1}^{n}(ACWP_i - Prediction_i)^2$
   - Subject to:
     - $Min\ EAC \leq \text{Final Cost (K)} \leq 2 * Max\ EAC$
     - $Min\ ECD \leq \text{Duration } (\alpha) \leq 2 * Max\ ECD$

Optimization Constraints

1. Appealing becau
2. Computationally
   modern technolo
3. Main issue is tha
   of-fit cannot be
4. Has form:
   - $Minimize\ \sum_{i=1}^{n}($
   - Subject to:
     - $Min\ EAC \leq$ Fin
     - $Min\ ECD \leq$ Dur



	Unconstrained	Constrained
Cost	6,151	15,629
Duration	2.82	4.96

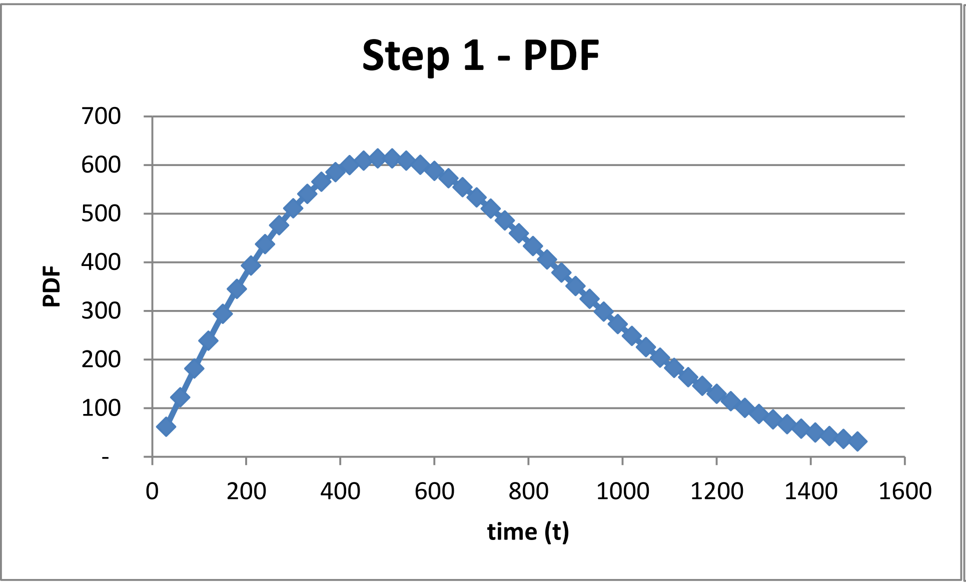
Constraint Selection is really important!

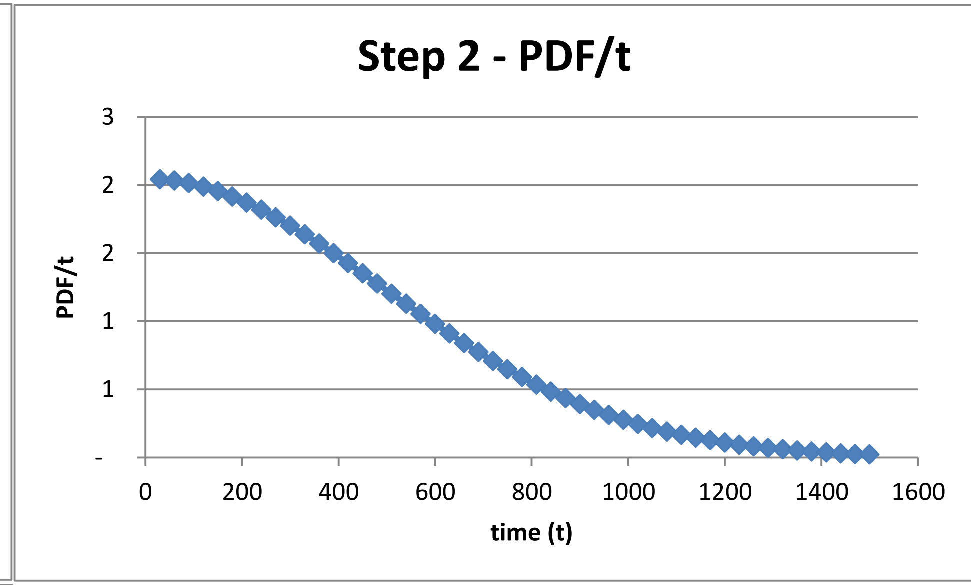
Linear Transform & Regression

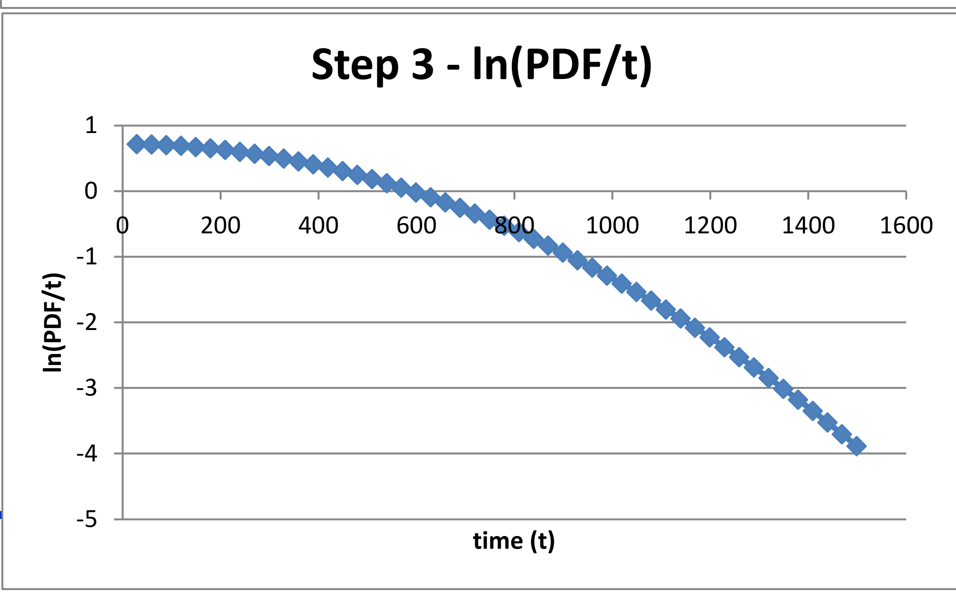
1. Given issues with optimization, desirable to find method that has analytical closed form solution
   - Not easy, as Rayleigh function is not obviously transformed
2. Abernathy (1984)[1] developed a linear transform
   - Uses Rayleigh PDF as starting point
3. Provides straightforward means to generate parameter estimates via linear regression
   - Requires way to estimate empirical derivative

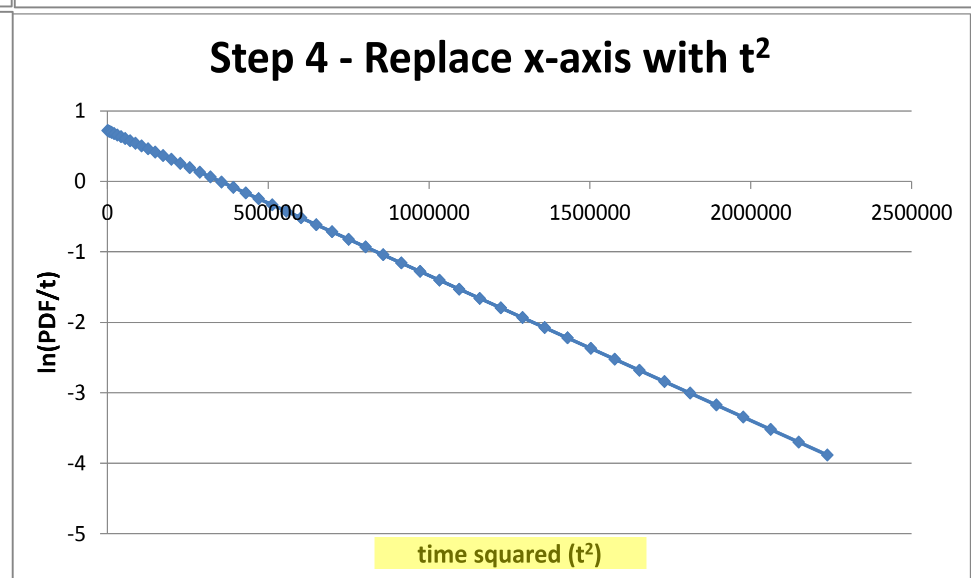
1. Abernathy, T., "An Application of the Rayleigh Distribution to Contract Cost Data," *Master's Thesis*, Naval Postgraduate School, Monterey, California,1984.

Rayleigh Linear Transform









Methods for Derivative Estimation

1. Slope of tangent line as approximation
   1. $f'(x_i) = \frac{\Delta y}{\Delta x} = \frac{\Delta ACWP}{\Delta time}$

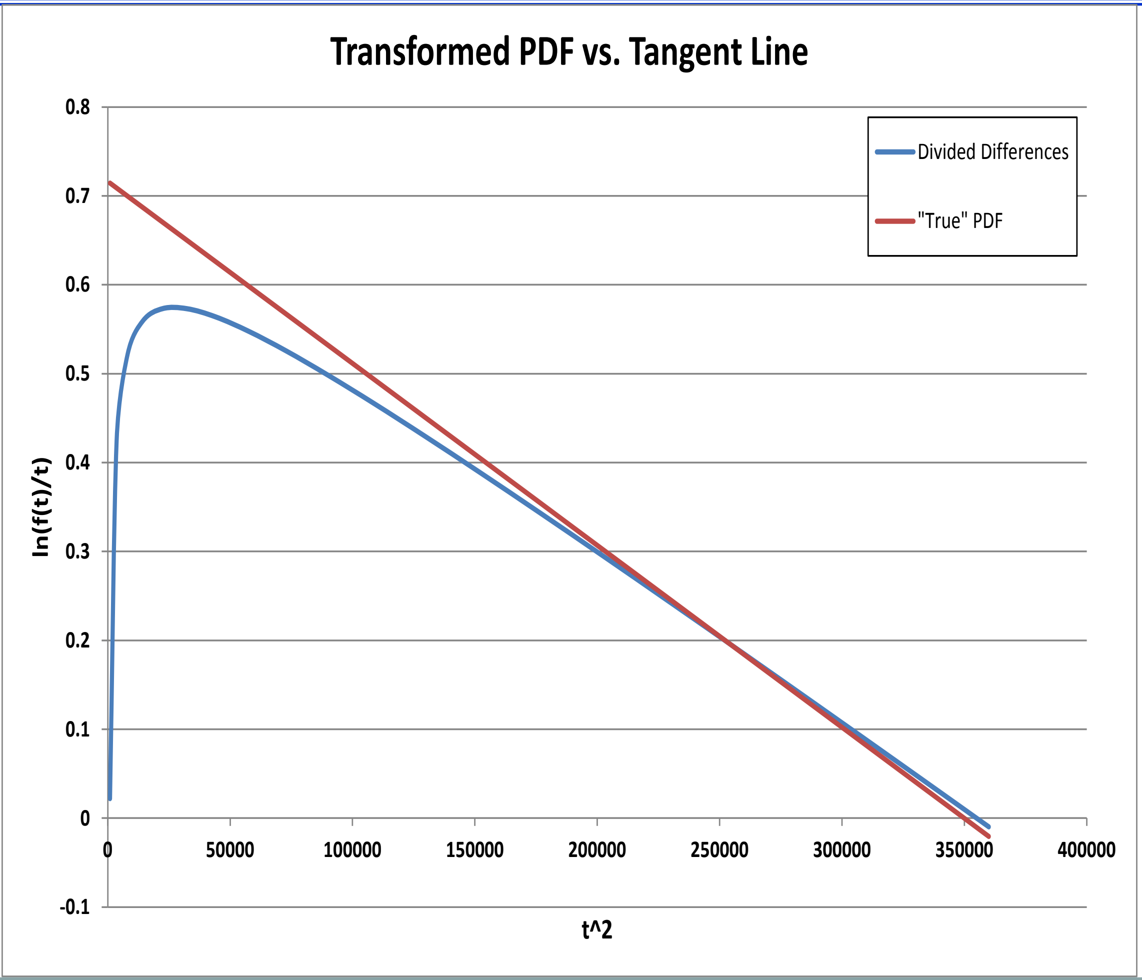
2. Interpolating Polynomials
   1. Will explain later

Tangent Line Accuracy

1. Evaluate accuracy of derivative estimator by comparing to Rayleigh PDF

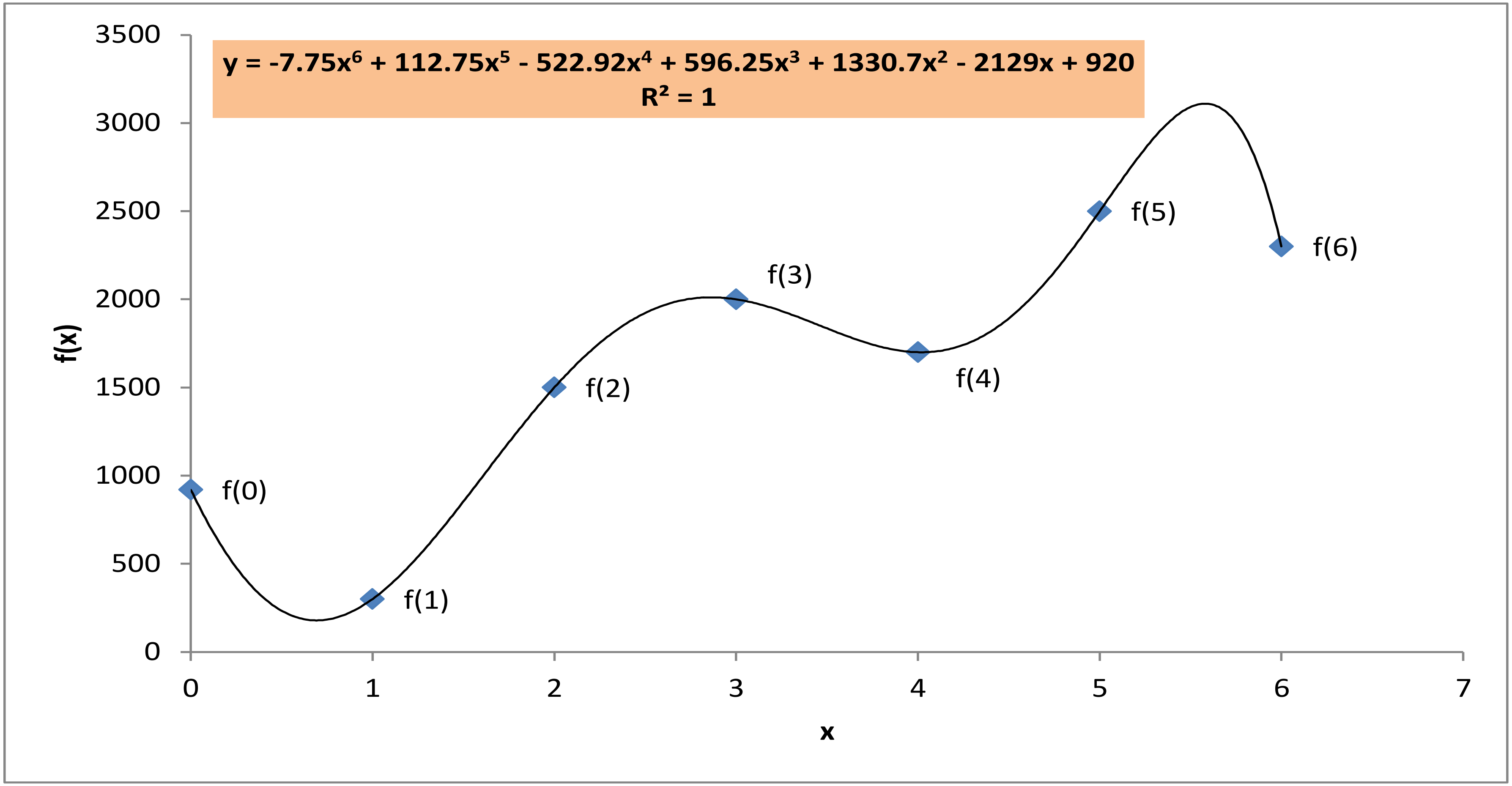
2. Tangent line method struggles with non-constant derivatives

3. Drives need for alternate derivative estimator



Errors mean we need to use a different method – Polynomial Interpolation

Polynomial Interpolation - Visual



1. Can always create a polynomial of order n-1 (6) that passes through all data
2. This polynomial is easily differentiated

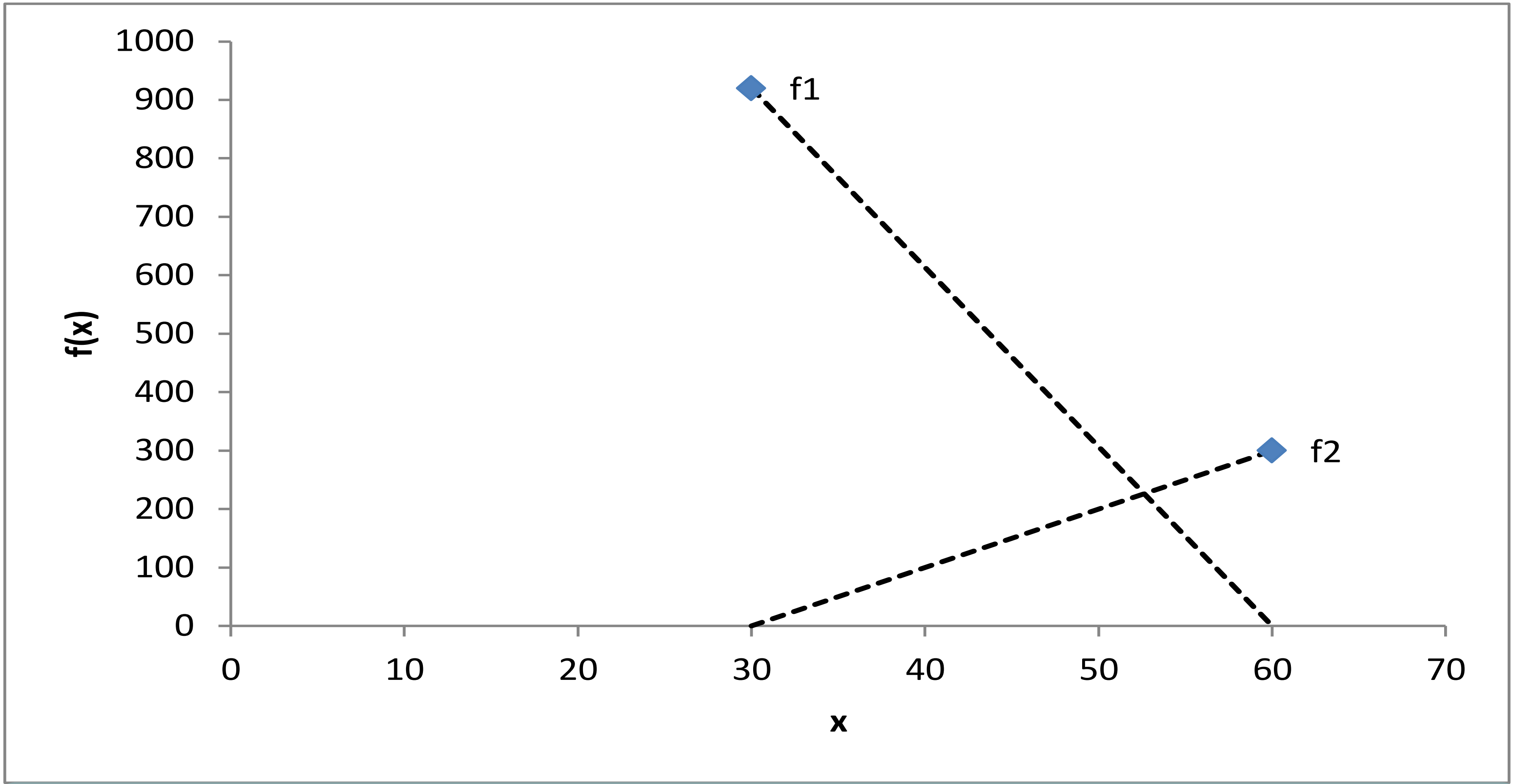
Lagrange Polynomial Construction

1. Polynomial has number of terms equal to number of points being fit

2. Function is comprised of two basic components:
   1. $The\ Lagrange\ Polynomial\ V_i(x_z)$
      - $\rightarrow V(x_i) = 1\ at\ i = \mathrm{z}, 0\ \mathrm{at\ i} \neq z$
   2. $The\ Coefficient\ f(x_i) = data\ point\ (i)$

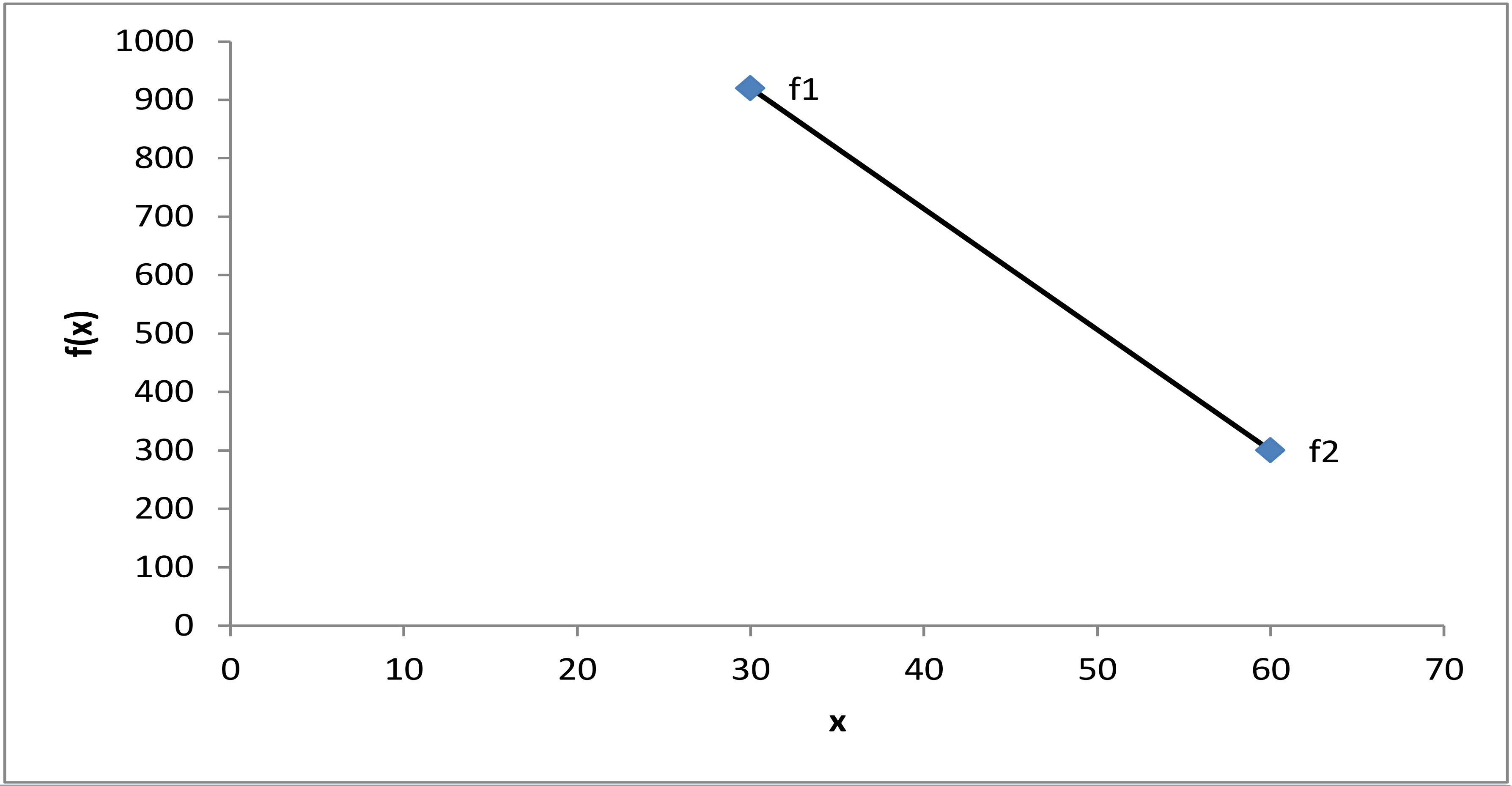
3. Polynomial has form:
   1. $g(x) = \sum_{i=0}^{n} V_i(x) * f(x_i)$

2-Point Example (order 1)



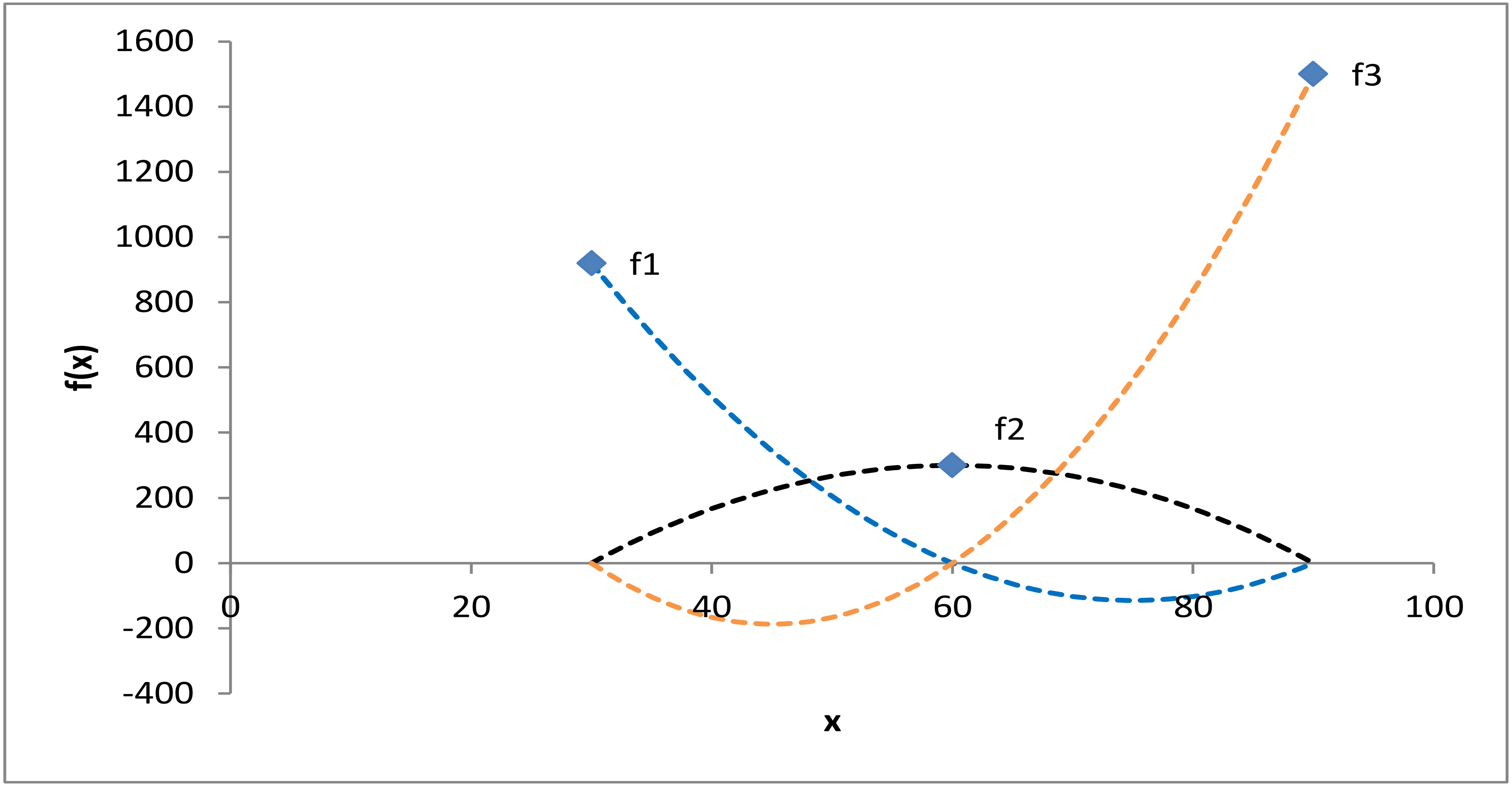
1) Each dashed line is the Lagrange Polynomial for that f(i)
2) Polynomial is order 1 with n=2, results in straight line

2-Point Example (order 1)



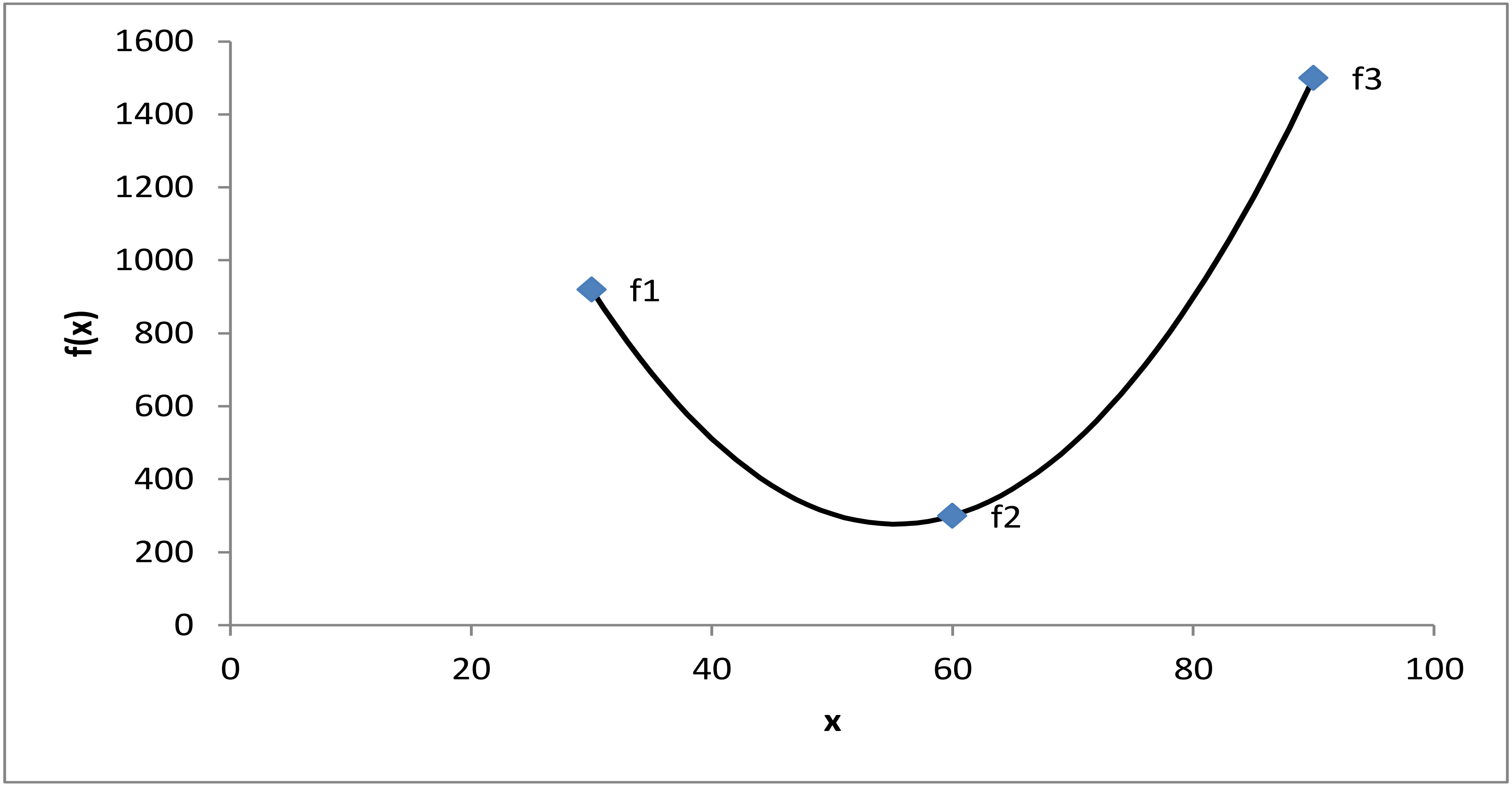
1) Polynomials for each f(i) are summed, giving interpolating function of order 1
2) Replicates tangent line method discussed previously

3-point example (order 2)



1) Each dashed line is the Lagrange Polynomial for that f(i)
2) Polynomial is order 2 with n=2, results in parabola

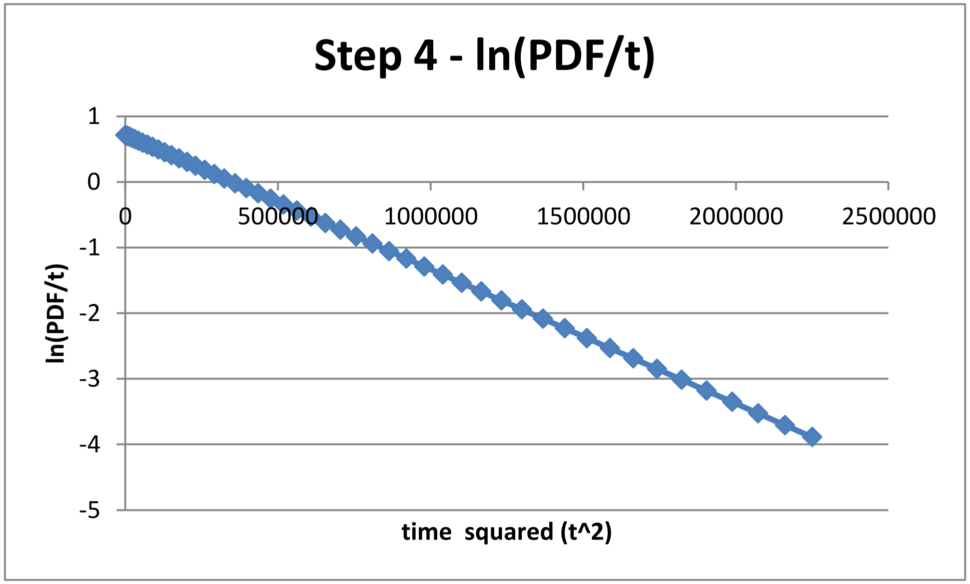
3-point example (order 2)



1) Polynomials for each f(i) are summed, giving interpolating function of order 2
2) Results in different derivative estimate than tangent function

Polynomial Parameter Estimation

1. Use polynomials to estimate derivative
2. Transform derivative estimate
3. Use linear regression to estimate line parameters
4. Transform line parameters back to Rayleigh form



Evaluating Rayleigh Method

Basic Method

1. For each contract in data set, start with first 3 data points
   - Generate Rayleigh fit using only first 3 points
   - Evaluate performance of that Rayleigh curve
2. Repeat step 1 using first 4 data points, continue until the last observation for the contract is reached
3. Repeat steps 1 and 2 using different fitting methods for comparison purposes

Data

- 20 Programs from EVM data base
- Criteria:
  - Must be completed
  - At least 90% complete
  - Less than 25% complete at first report
  - RDT&E Funded, SDD contracts only
- 320 curves to evaluate (per method)

	Minimum	Maximum
Start Year	1994	2004
Duration (Years)	2.00	7.97
Value($K)	6,227	2,139,304

Important Notes

1. Extremely poor fits are excluded
   – Would add support to my conclusion

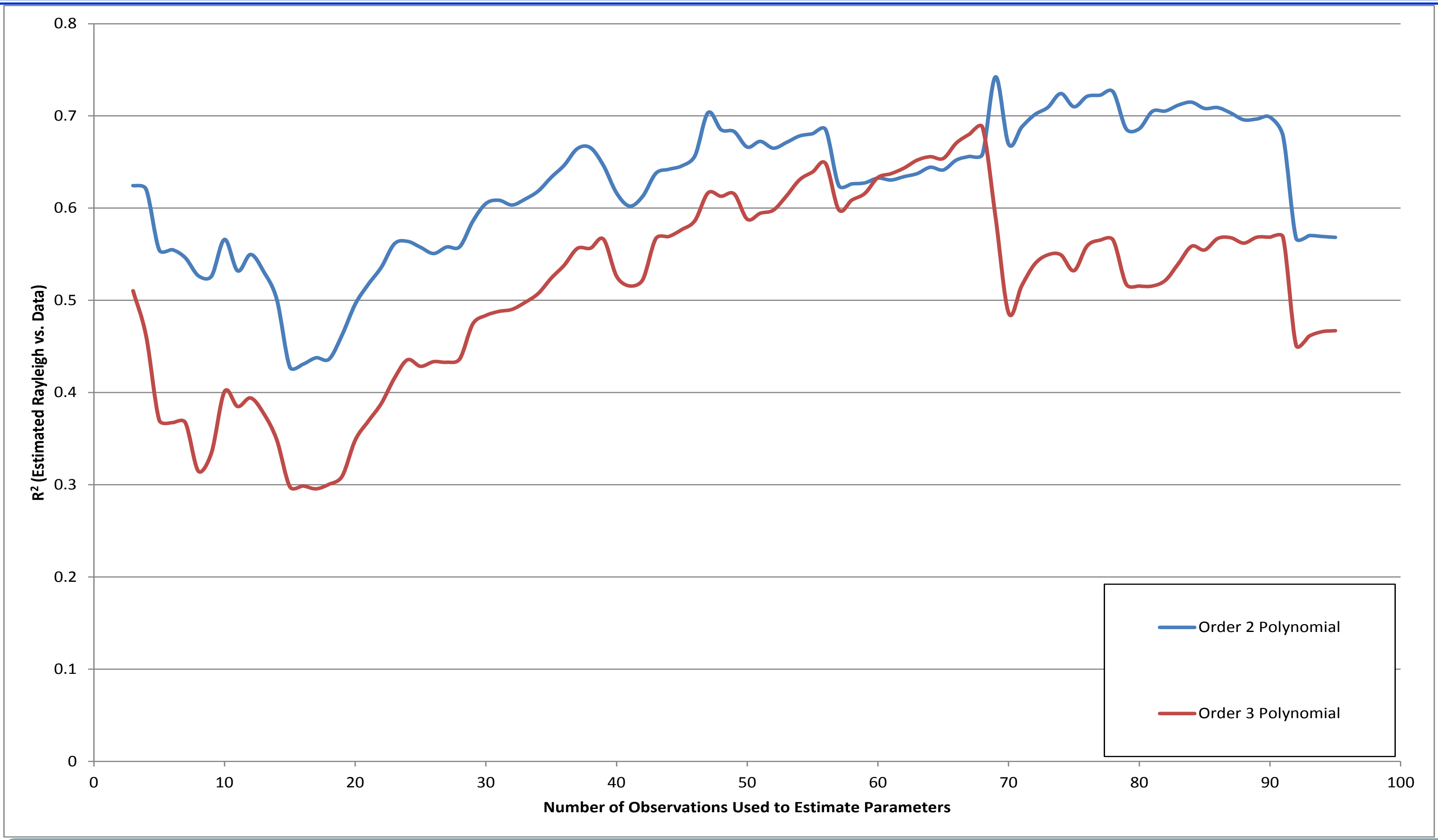
2. Some parameter fits were not calculable
   – i.e. LN(-1)
   – Excluded from analysis

3. Focus of analysis was on first 18 points
   – Evaluating use as an “early warning system”

R^2 of Polynomial Estimates

(average across all contracts reporting at observation (i))

Larger is better

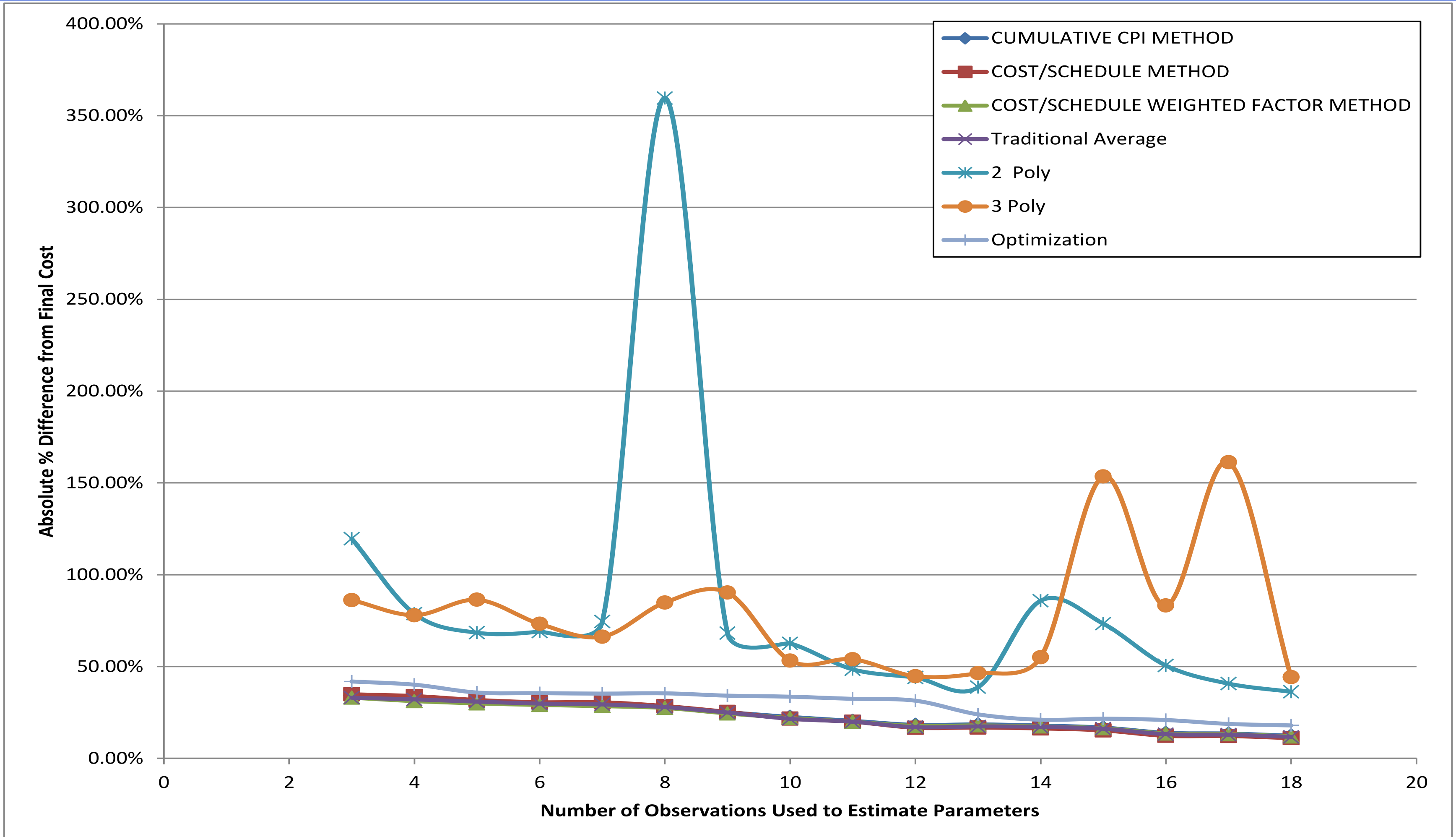


Would expect R^2 very close to 1.0 if data follows Rayleigh Curve

Error of Cost Parameter Estimates

(average across 20 contracts)

Smaller is better

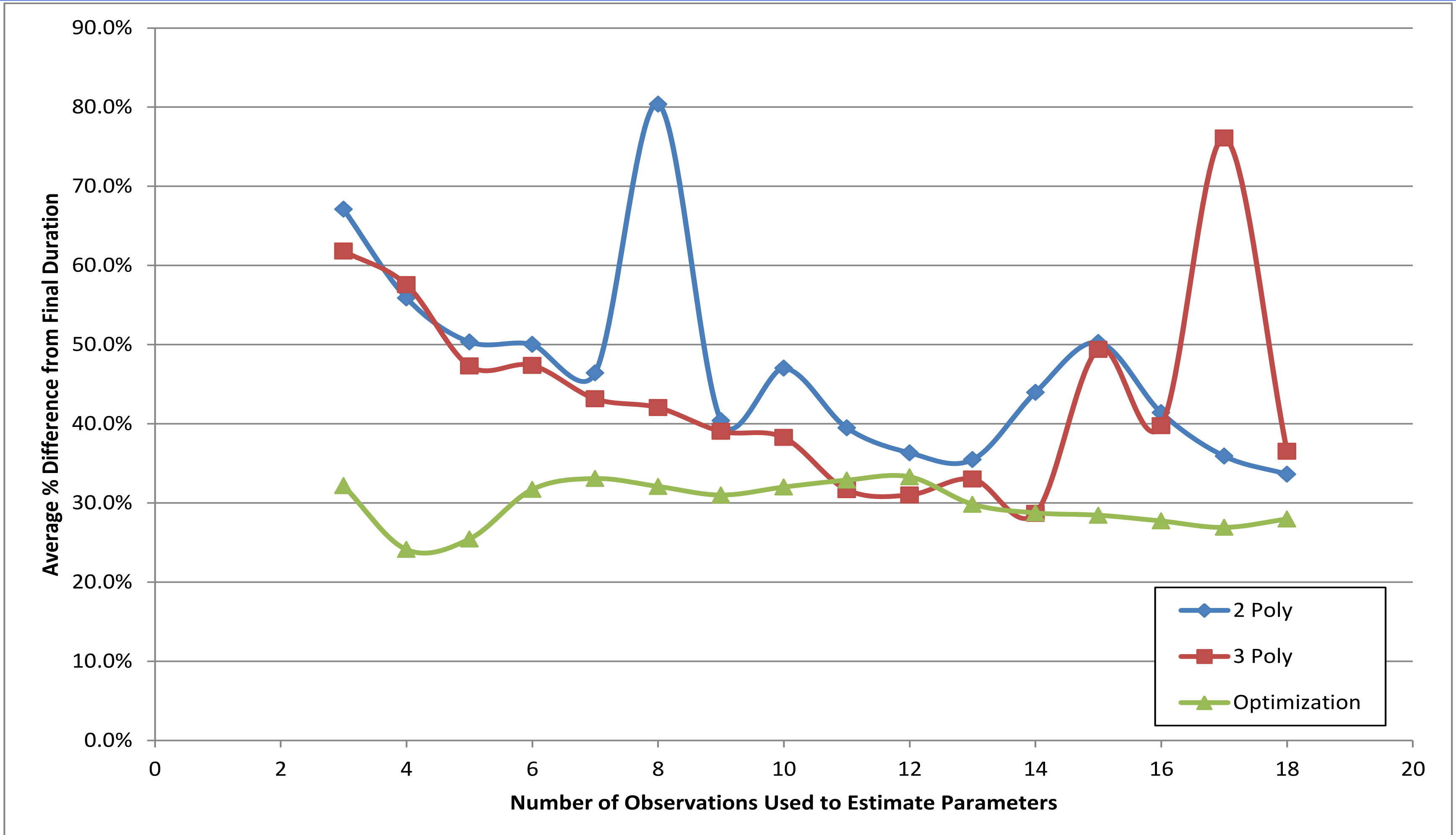


Optimization and Derivative methods perform worse than traditional methods

Error of Time Parameter Estimates

(average across 20 contracts)

Smaller is better

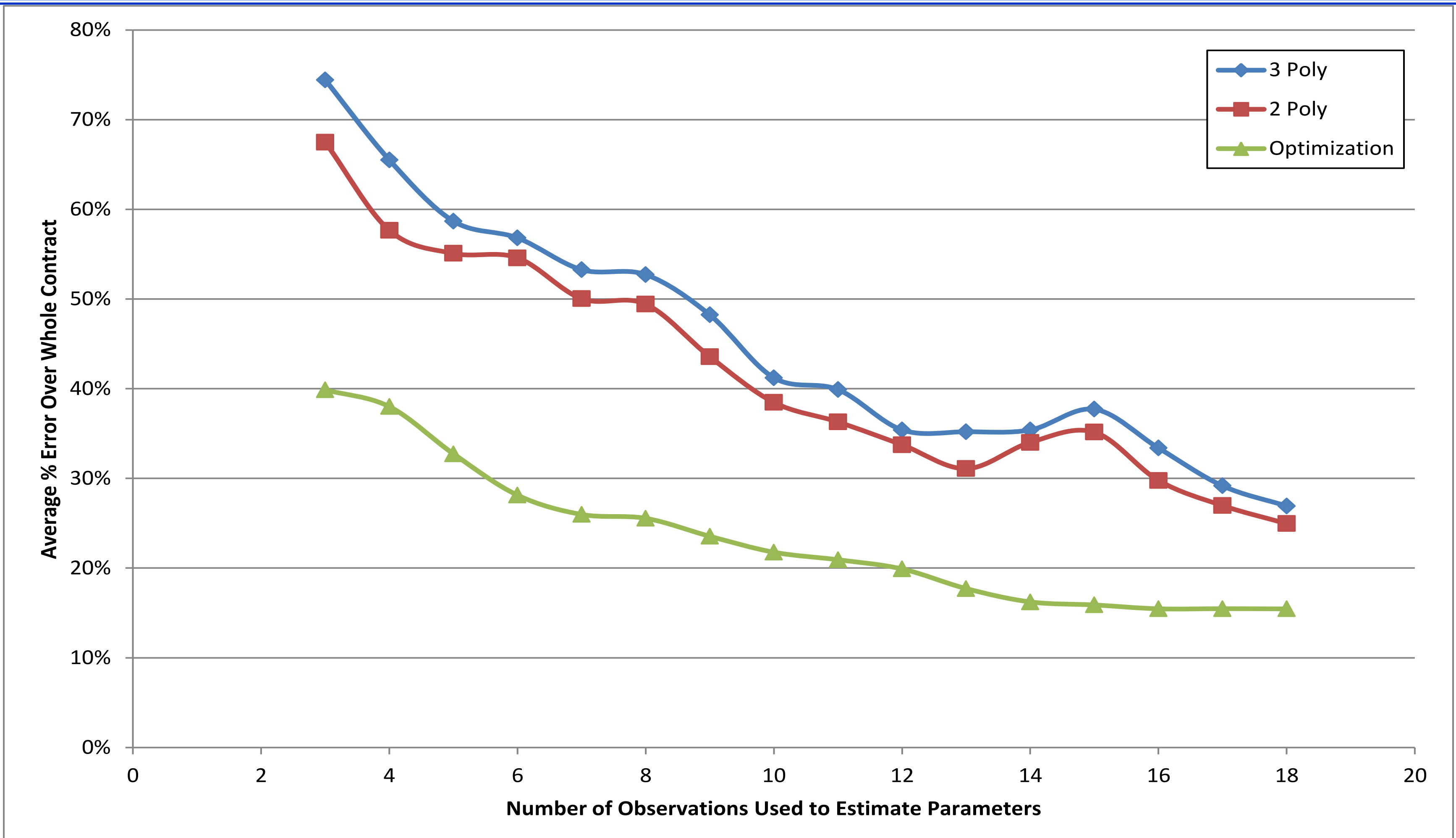


Optimization and Derivative methods don't perform well

Percent Error Over All Points

(Average across 20 contracts)

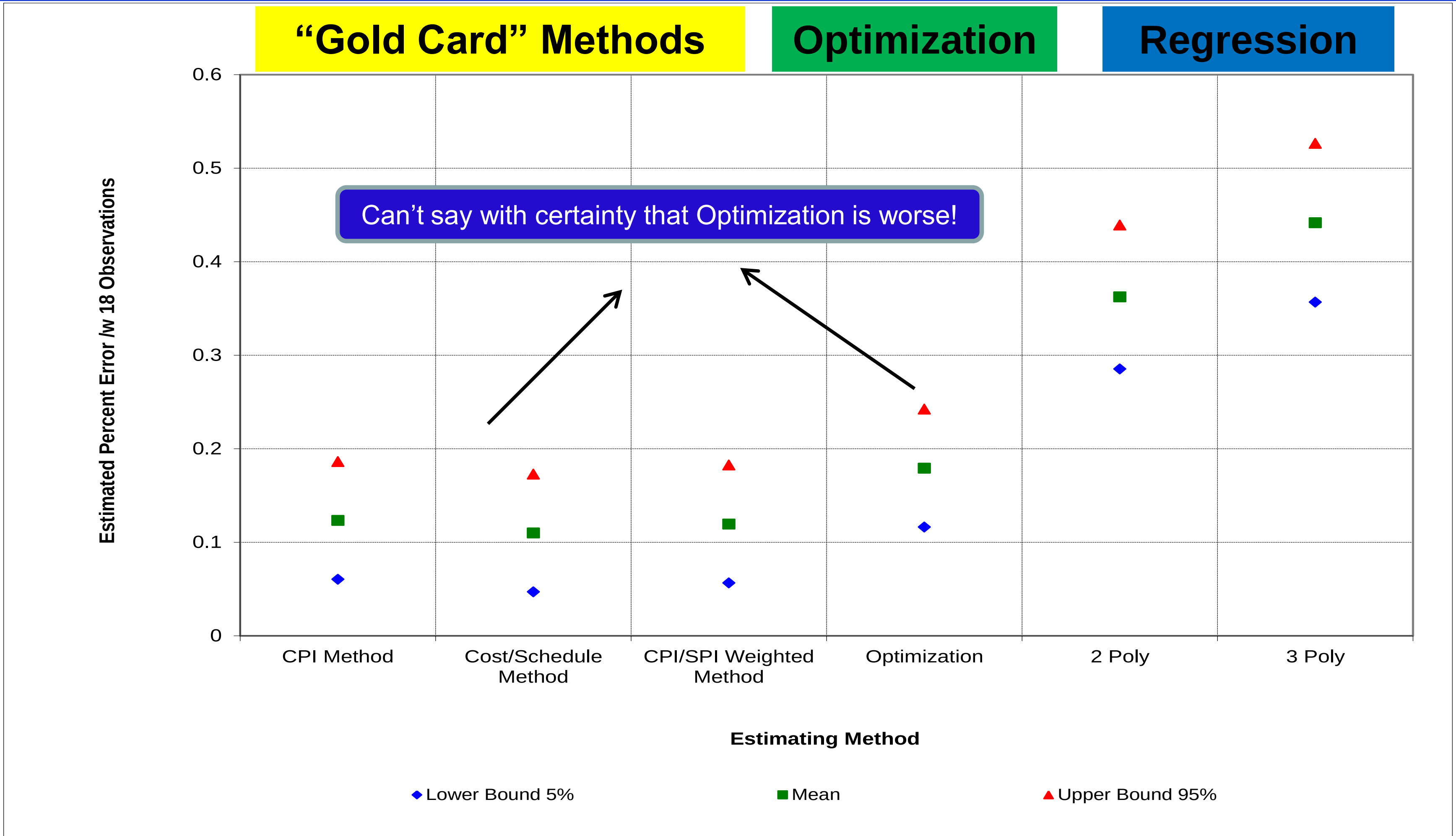
Smaller is better



Similar to R^2 , expect minimal error over all points if EVM follows Rayleigh curve

ANOVA (18 Observations)

Smaller is better



Rayleigh methods appear to perform worse than traditional methods

Conclusions

Conclusions

- For this data set, Rayleigh estimates do not improve on Gold Card methods
- The Rayleigh model does not fit the contracts in this data set
- Analysts need to be careful if using Rayleigh
  - Results need to be supported with actual data
  - Rayleigh is just a mathematical function, not magic
- Consistent with results found by Abernathy
  - Rayleigh parameters can be estimated, but no success as predictor

Future Work

- Expand analysis to include more contracts
- Continue to Refine parameter estimation tool
- Use analysis to generate standard risk distributions for R&D estimates
- Develop guidance for optimization constraints
- Evaluate other mathematical functions to replace Rayleigh

Acknowledgements

1. Mr. Mike Popp
2. Mr. Marc Greenberg
3. Mr. Ben Breaux
4. Ms. Heather Brown
5. Dr. Casey Trail

Questions?