

# ***Practical Software and Systems Measurement***

***Objective Information for Decision Makers***



***The Application of Nonlinear  
Regression Methods and Machine  
Learning to Army Software  
Maintenance Cost Estimation***

September 18, 2019

Dr. Christian Smart and Kimberly  
Roye, Galorath Federal  
Paul Janusz, US Army CCDC-AC

## ***The Application of Nonlinear Regression Methods and Machine Learning to Army Software Maintenance Cost Estimation***

- The scope of software projects varies widely. Due to economies - and more often diseconomies - of scope, nonlinear cost estimating relationships (CERs) are more appropriate than linear ones. The use of log-transformed ordinary least squares is a traditional method, but one of its main drawbacks is that it is biased low
- Nonlinear regression methods are useful in developing CERs while minimizing bias
- The following methods will be compared:
  - *Zero-bias Minimum Percent Error (ZMPE)*
  - *Maximum likelihood Regression of Log Normal error (MRLN)*
  - *Zero-bias Minimum Absolute Percent Error (ZMAPE)*

# ***Objectives of the Workshop***

- **Compare modern nonlinear regression techniques used to develop cost estimating relationships that are unbiased**
- **Gain an introduction to a variety of supervised learning methods that can also be used for predictive analysis and practice applying some of these techniques in estimating**

# ***Workshop Format***

- ***Presentation of Machine Learning (ML) Techniques***
- ***Walkthrough of R-Code for a Subset of ML Techniques and Comparison of Results of an Applicable Example***
- ***Comparison of Nonlinear Regression Techniques***
- ***Demo of MRLN and ZMPE Techniques***

# ***Workshop Background***

- ***Regression, including log-transformed ordinary least squares, to develop cost estimating relationships and trend lines has a long history in cost estimating, including software cost estimating***
  - ***Recent PSM workshop presentations dealing with this topic include:***
    - ***“Using Army SW Sustainment Cost Estimating Results,” Cheryl Jones and James Doswell, 2018***
    - ***“Leveraging the Power of Historical Data Through the Use of Trend Lines,” Taylor Putnam-Majarian, John Staiger***
- ***Where we’re heading***
  - ***Log-transformed ordinary least squares has issues, particularly bias; using modern methods can remove this and improve the fit (lower standard errors)***
  - ***The use of machine learning offers the potential to offer alternative methods for estimating costs that have their own unique advantages***
- ***Issues, questions, and topics***
  - ***Adoption of techniques, training, and tools***

# ***Intended Output***

- ***Recommendations from users on adjustments to ML code to optimize performance***
- ***Provide users with code and Excel spreadsheets for use in conducting future analyses using techniques presented***

# ***Workshop Title***

- ***The Application of Nonlinear Regression Methods and Machine Learning to Army Software Maintenance Cost Estimation***

# ***Workshop Participants***

- ***Joe Dean***
- ***Marilyn Vickers, USAF***
- ***Victoria Brown, USAF***
- ***Katherine Bradshaw, AFCAA***
- ***James Doswell, US Army***
- ***Paul Janusz, US Army***
- ***William Nichols, SEI***
- ***Robert Stoddard, SEI***

## *Summary*

- Presented several Machine Learning (ML) methods and discussed how the methods were applied to Army software sustainment data. Walked through R code on how results were calculated
- To follow-on from the presentation on Nonlinear regression methods, performed a demo in Excel Solver and R on the three nonlinear methods (ZMPE, MRLN, ZMAPE)

## ***Conclusions, Recommendations, and Results***

- ML and Nonlinear regression techniques can be very useful in validating traditional regression results and even allowing prediction when no significant regression models are available
- We recommend users try implementing some of these techniques when analyzing datasets
  - Nonlinear methods can be applied to strengthen cost estimating relationships and removing bias present in log-transformed Ordinary Least Squares models
  - ML techniques can be applied to classify and predict results

## ***Next Steps/Action Items***

- Revise input dataset to scale factors and determine how the results are impacted
- Determine if additional parameters should be added to code to fine tune the performance
- Determine how results can be combined/averaged/weighted to provide strong estimates and crosschecks
- Determine how to help decision makers have more confidence in the results developed by these new methods