The NRO CAAG
CER Analysis Tool

June 9, 2015

Donald MacKenzie

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Topics

- CERAT Overview
- Background
- IDP Analysis Process
- CER Development Aids
- Summary

CERAT Overview

- Developed by the NRO Cost and Acquisition Assessment Group (CAAG)
- Primary purpose: Identify and assess influential data points (IDPs) in CER data sets
- IDP Impact: Percent change in a CER estimate for a target data point due to removal of any data point
  - CER analyst selects the “target” data point from the CER data set
- ZMPE, MUPE, LOLS and AAPE best-fit methods used
- Baseline CER fits with each method are performed first
- Next, the IDP influence analysis is performed
- Focus is on the top three most influential data points (for each best-fit method)

CERAT Overview, Con't

- Also, the largest impact on any data point estimate is determined for each data point removal
- CER stability is assessed by movements in the CER constants with each data point removal
- Several other aids for CER development are included in CERAT output displays, by best fit method
  - Advanced X-Y graphics
  - Residuals plotted vs. continuous IVs (linear & log)
  - Residual histograms (linear and log residuals)
  - Correlation matrices and variable Swing Factors
  - Modified Cook's Distance
  - Skew and specialized R^2 graphs

Topics

- CERAT Overview
- Background
- IDP Analysis Process
- CER Development Aids
- Summary

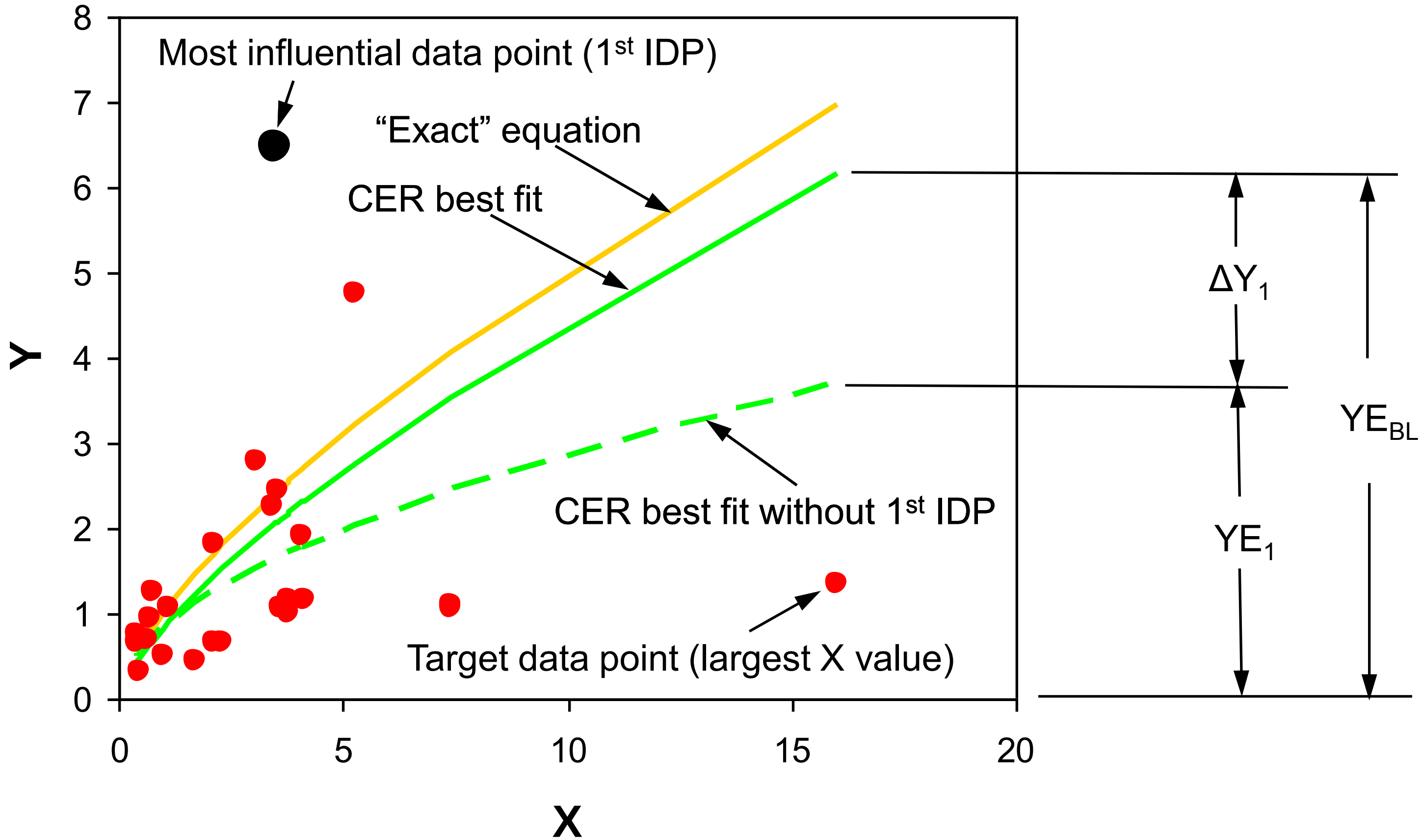
CAAG Influential Data Point Study

- Performed in 2011, giving rise to CERAT development
- Described in 2012 Joint ISPA/SCEA Conference in Brussels
- Monte Carlo simulation of CER data sets
  - CER Form: $Y = AX^B$
  - X and Y lognormally distributed
- Perform LOLS, MUPE, ZMPE and AAPE best fits for each sampled data set
- Calculate 1st, 2nd & 3rd IDP impacts on the target data point,
  - At max value of X in the data set (largest Y estimate)
- Analysis cases:
  - 200 data sets per analysis case
  - 10, 15, 25 & 50 data points per data set
  - 35% , 65% & 100% SPE
  - Exponent B: 0.5, 0.7 & 1.0

IDP Impact Measurement

1st IDP Impact = $DY_1 = (YE_1 - YE_{BL}) / YE_{BL}$

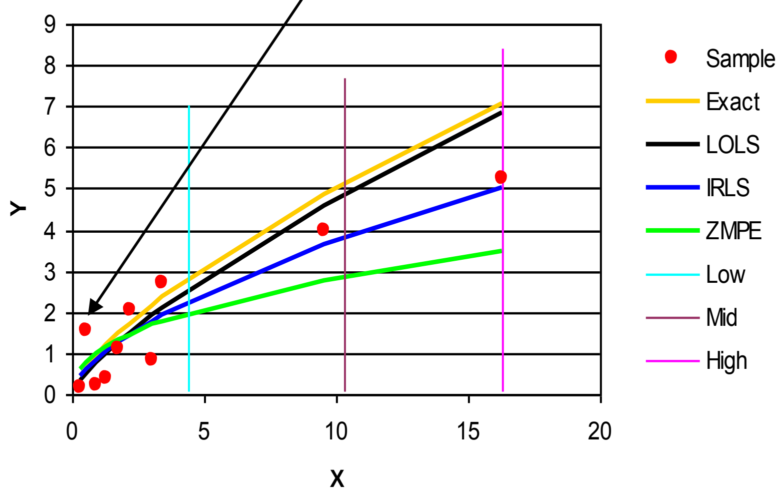
(Expressed as a percentage, negative if downward movement)



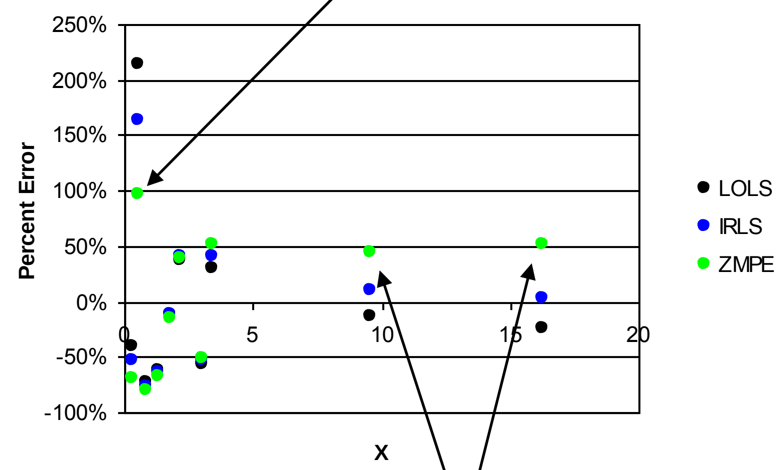
Typical ZMPE Behavior – Low End Pull

- ZMPE exponent is “pulled down” when data points in the low end of the X range with high Y values are present – and mid-range and high-end data points provide a “pivot” and “anchor”, forcing lower exponent.

Data point pulls ZMPE curve up -- and exponent down



Data point percent error substantially reduced by pull



Mid-range and high end data points provide pivot and anchor

Note: MUPE is also referred to as “IRLS” – Iteratively Reweighted Least Squares

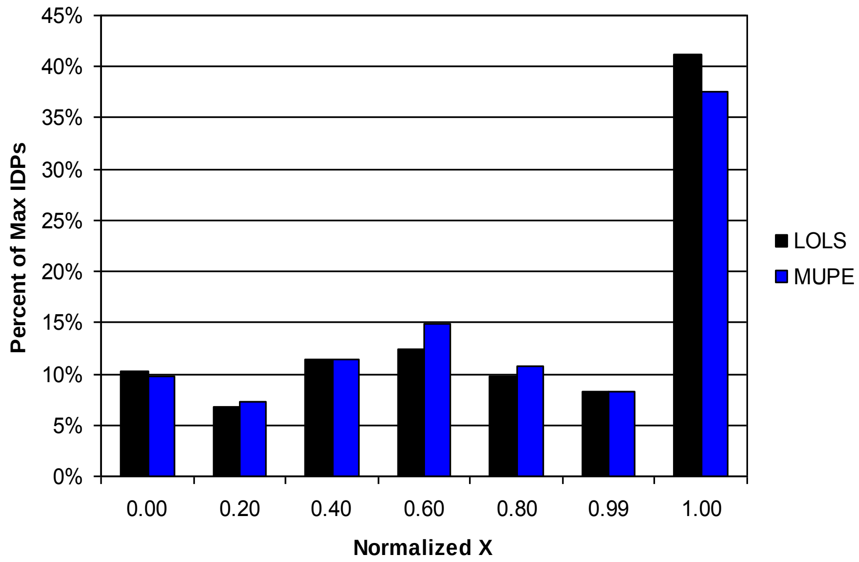
Summary – IDP Impact Study Results

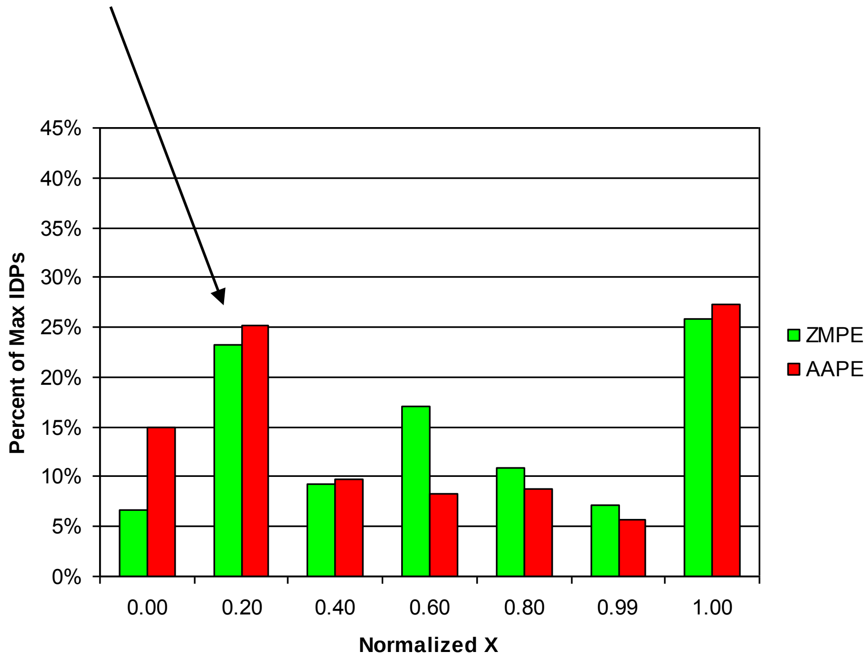
- LOLS and MUPE have about the same average IDP impact
- LOLS and MUPE are less sensitive to IDPs than ZMPE and AAPE
  - ZMPE impacts average 38% higher than LOLS and MUPE over 26 analysis cases (17% min, 78% max)
  - AAPE impacts average 55% higher than LOLS and MUPE
- Impacts decrease dramatically with increasing number of data points
- Impacts increase moderately with SPE
- Impacts are not sensitive to exponent B
- LOLS and MUPE have the same IDP 60-80% of the time
- All methods have the same IDP 15-30% of the time

Distribution of IDPs vs. Normalized X

Normalized X = X / Maximum X

Due to low-end pull





ZMPE and AAPE are sensitive to low-end data points with large positive percent errors

Topics

- CERAT Overview
- Background
- IDP Analysis Process
- CER Development Aids
- Summary

IDP Influence Analysis Process

- First, the analyst selects a target data point in the CER data set
  - Usually the data point with the highest estimated cost
- Data points are removed one at a time, and
- Impacts on the baseline estimates for the target data point are determined
- The 1st, 2nd and 3rd most influential data points are identified for each method
- For each data point removal, the maximum impact over all other data point estimates (besides the target data point) is also determined
- IDP impact assessment tools:
  - CER regression constants for each method and data point removal
  - Graphs of Adjusted Y* vs. each continuous variable
  - Graphs show CER equation without IDP – for 1st, 2nd & 3rd IDPs
  - Likely 1st, 2nd and 3rd IDP impact percentiles – for target data point
  - Cook's Distance (modified for proportional errors) -- for each data point
  - Graphs of Maximum Impacts vs. Cook's Distance

* Y is adjusted by "projecting" data points onto the plane of the graph using the CER equation

Types of CERs Handled By CERAT

- CERs have the form $Y=AX^BY^CD^Z\ldots$
- A term such as D^Z may be used for stratification
  - Z is a binary stratifying variable and D is a factor determined by regression
- The following apply only to ZMPE and AAPE methods
  - Estimating bias (average percent error) can be constrained to zero for any stratum (data subgroup)
  - The CER equation may have more than one term
  - Exponents may have a compound form:

    $B = S_B * B'$ where

    S_B is a binary stratifier variable, and

    B' is the exponent for the data points with $S_B = 1$
  - Compound exponents allow for different exponents for the same variable, depending on the data subgroup
  - Fixed factors may be applied to data: $Y_i=(AX_i^BY_i^CD_i^Z)*F_i$

Cook's Distance Definitions

Standard OLS Definition

$$D_i = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p \, \mathrm{MSE}}.$$

$\hat{Y}_j$ is the prediction from the full regression model for observation *j;*

$\hat{Y}_{j(i)}$ is the prediction for observation *j* from a refitted regression model in which observation *i* has been omitted;

MSE is the mean square error of the regression model; and

p is the number of fitted parameters in the model

Modified Definition for Constant Percent Error Models

$$MCDi = \frac{\sum_{j=1}^{n} \left[\left(\hat{Y}_j - \hat{Y}_{j(i)} \right) / \hat{Y}_j \right]^2}{pMSPE}$$

MSPE is the mean square percentage error of the regression model

IDP Analysis Primary Statistics Part A

	Influential Data Points (IDPs) and Impacts (DYs) on Estimate for Selected Data Point No. 25, Data Point 25							Results Over All Data Point Removals	
	B/L Y Est	1st IDP	1st DY	2nd IDP	2nd DY	3rd IDP	3rd DY	Min SPE	Max MCD
ZMPE	4.881	6	9.3%	4	8.9%	12	8.3%	37.3%	0.544
MUPE	4.880	12	12.3%	14	7.1%	14	6.3%	38.0%	0.384
LOLS	4.970	12	13.1%	14	7.5%	5	6.2%	39.1%	0.540
AAPE	5.291	6	12.7%	25	8.3%	1	6.3%	38.6%	0.839

Baseline estimates for target data point

1st influential data points

Impacts of 1st IDPs

Minimum SPEs over all data point removals

Maximum Modified Cook's Distances

IDP Analysis Primary Statistics Part B

Normalized IDP Impacts (NDYs) and NDY Percentiles (NDPs)

	1st NDY	1st NDP	2nd NDY	2nd NDP	3rd NDY	3rd NDP	Max SPE	Max GRSQ
ZMPE	0.23	17.9	0.22	15.2	0.20	12.0	42.2%	0.903
MUPE	0.29	54.3	0.17	14.0	0.15	10.5	44.1%	0.866
LOLS	0.30	60.8	0.17	10.2	0.14	1.9	46.4%	0.902
AAPE	0.31	40.3	0.20	18.7	0.15	7.8	45.1%	0.903

1st IDP normalized impacts

1st NDY percentiles

Maximum SPEs over all data point removals

Maximum Generalized R Squared values

Example ZMPE IDP Impacts

ZMPE IDP Analysis

CER Data Set		Impacts on Selected Data Point Estimate			Max % Impact Over All Data Pts	
Data Pt.	Description	New Est Y	New Est - B/L Est	% Diff	Max Y % Diff	Data Point
	Baseline Values	4.881				
1	Data Point 1	4.951	0.070	1.4%	22.3%	1
2	Data Point 2	4.916	0.035	0.7%	-23.3%	12
3	Data Point 3	4.860	-0.021	-0.4%	-1.2%	1
4	Data Point 4	5.313	0.432	8.9%	-30.7%	1
5	Data Point 5	4.831	-0.050	-1.0%	8.2%	5
6	Data Point 6	4.427	-0.454	-9.3%	-9.3%	25
7	Data Point 7	4.831	-0.050	-1.0%	7.1%	10
8	Data Point 8	4.837	-0.044	-0.9%	-1.7%	1
9	Data Point 9	4.815	-0.066	-1.4%	2.6%	9
10	Data Point 10	4.784	-0.097	-2.0%	5.0%	10
11	Data Point 11	4.917	0.037	0.7%	-3.8%	11
12	Data Point 12	5.287	0.406	8.3%	16.8%	12
13	Data Point 13	4.862	-0.019	-0.4%	-0.8%	11
14	Data Point 14	5.170	0.289	5.9%	9.7%	1

Est Y for D.P. 12 moves the most (-23.3%) when D.P. 2 is removed

2nd IDP

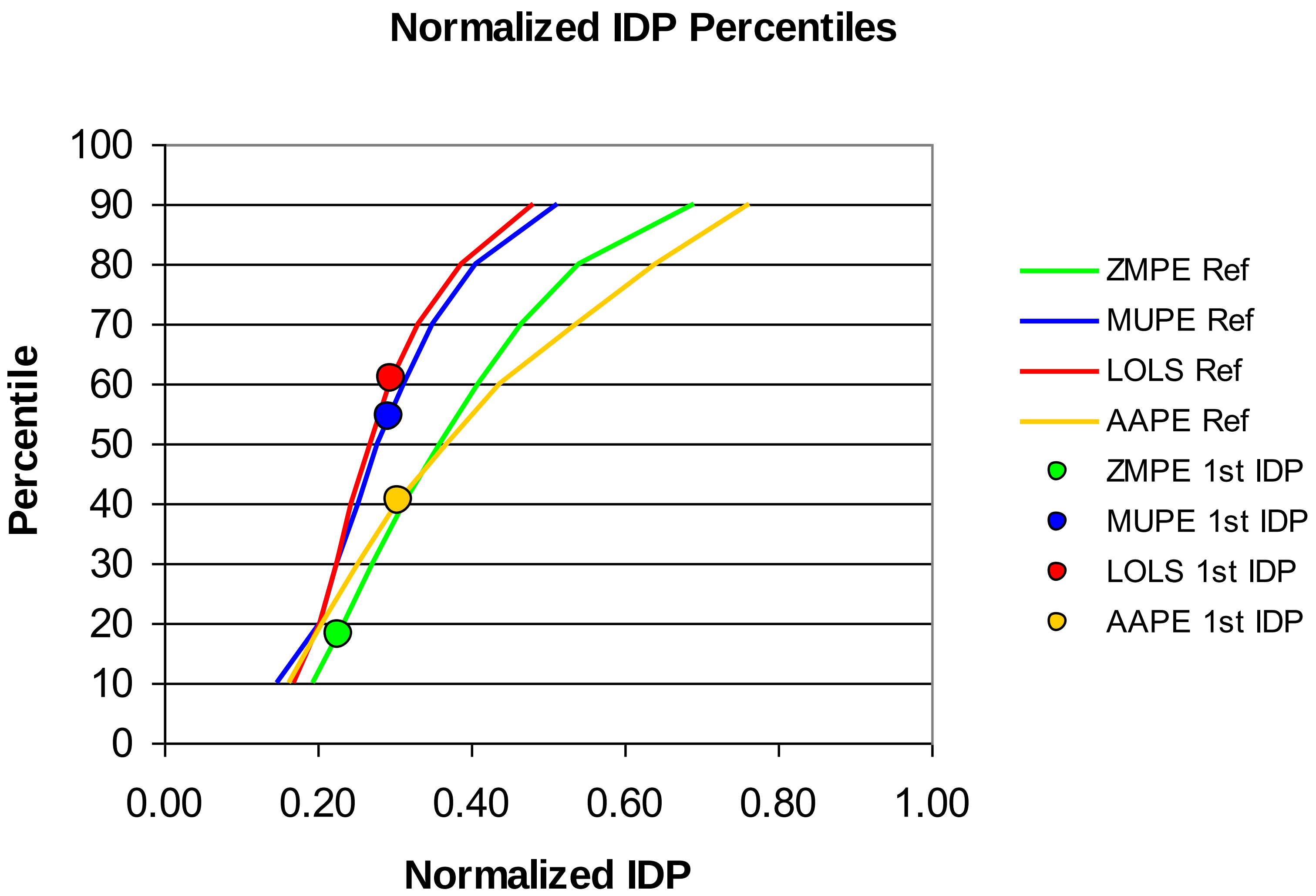
1st IDP

3rd IDP

Normalized IDP Percentiles

- Normalized IDP Percentiles (NDPs) are based on the Monte Carlo simulation results of the IDP study
- The CER equation form in the study was: $Y = AX^B$
- Although a typical CER can have more variables, the normalized IDP percentiles from the study are assumed to be appropriate for estimating underlying normalized IDP percentiles
  - Hence, “Likely” NDPs
- Approximate percentiles for IDPs, corresponding to the number of data points N and the SPE, are calculated from IDP study results
- The percentiles are calculated by interpolation
  - Or extrapolation below the 10^{th} percentile and above the 90^{th} percentile
  - For stratified regressions, the number of data points N is artificially reduced to better represent CER behavior
  - The percentiles represent a way of assessing how “unusual” a given IDP impact is

Normalized IDP Percentile Graph



ZMPE and AAPE percentiles are typically lower than LOLS and MUPE

ZMPE Regression Equation Constants

		ZMPE IDP Statistics						
Data Pt.	Description	A	B	C	D	E	SPE	MCD
	Baseline Values	1.036	0.250	0.365	0.926	1.016	41.0%	
1	Data Point 1	1.163	0.276	0.273	0.980	1.000	40.1%	0.171
2	Data Point 2	0.753	0.191	0.508	1.057	1.099	37.3%	0.382
3	Data Point 3	1.023	0.251	0.367	0.929	1.017	42.2%	0.002
4	Data Point 4	0.938	0.271	0.477	0.789	0.843	39.8%	0.544
5	Data Point 5	1.079	0.247	0.360	0.900	1.061	39.0%	0.058
6	Data Point 6	1.048	0.246	0.352	0.868	1.098	39.9%	0.144
7	Data Point 7	1.007	0.193	0.416	0.971	1.008	40.9%	0.043
8	Data Point 8	1.028	0.251	0.366	0.920	1.021	42.2%	0.003
9	Data Point 9	1.040	0.249	0.365	0.912	1.038	42.1%	0.008
10	Data Point 10	0.986	0.193	0.427	0.957	1.014	41.9%	0.022
11	Data Point 11	1.052	0.283	0.321	0.937	1.017	42.1%	0.012
12	Data Point 12	1.226	0.362	0.222	0.887	0.983	38.0%	0.125
13	Data Point 13	1.029	0.251	0.364	0.929	1.027	42.1%	0.001
14	Data Point 14	1.133	0.301	0.287	0.944	0.988	38.8%	0.067
15	Data Point 15	1.203	0.285	0.261	0.909	0.914	40.5%	0.239
16	Data Point 16	0.992	0.202	0.424	0.957	1.004	41.0%	0.028
17	Data Point 17	1.029	0.250	0.364	0.927	1.017	42.2%	0.001
18	Data Point 18	1.006	0.217	0.423	0.885	1.070	40.5%	0.087

Lowest SPE over all D.P. removals

Highest Modified Cook's Distance

Highest SPE

Data Point 2 has the largest SPE impact with large changes in CER constants.

Topics

- CERAT Overview
- Background
- IDP Analysis Process
- CER Development Aids
- Summary

CER Development Aids

- Advanced X-Y graphics
- Modified Cook’s Distance
- Residuals plotted vs. continuous IVs (linear & log)
- Residual histograms (linear and log residuals)
- Correlation matrices
- Variable Swing Factors
- Skew and R^2 graphs

Y Values Adjusted for Graphing

- Adjustment moves the data point along the regression surface to a specific value for each out-of-plane variable
  - CERAT uses the variable averages for continuous variables, and
  - The user specifies the values of the stratifiers (0 or 1)
- This enables a direct comparison between the actual data points and the baseline regression line based on the same set of out-of-plane variables
- Example: Y vs. weight graph, baseline ZMPE regression:

$$Y = A \cdot Wt^{B} \cdot Freq^{C} \cdot D^{TypeA} \cdot E^{TypeB}$$

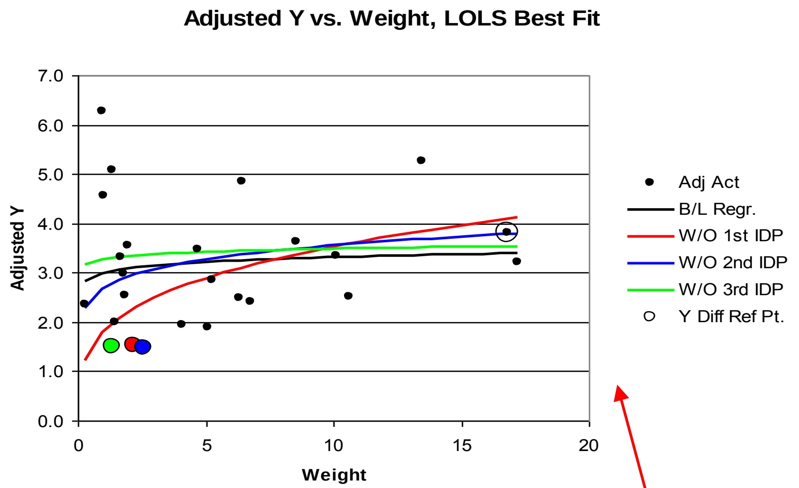
- A = 1.036   B = 0.250   C = 0.365   D = 0.926   E = 1.016
- Out-of-plane graphing values: Freq = 6.880   TypeA = 1   TypeB = 1
- Adjustment equation for Data Point 1 (Freq = 0.50, TypeA = 1; TypeB = 0):

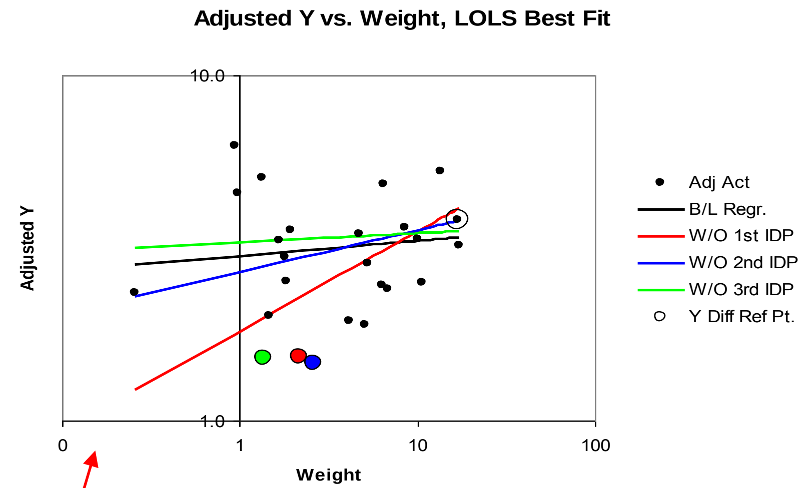
$$\text{Adj } Y = \text{Act } Y \cdot (6.880/0.50)^{0.365} \cdot (0.926^{1}/0/926^{1}) \cdot (1.016^{1}/1.016^{0})$$

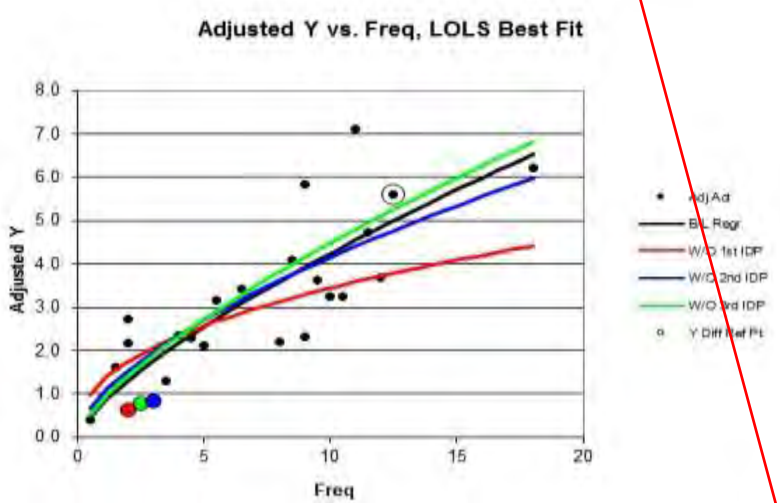
$$\text{Adj } Y = 0.329 \cdot 2.602 \cdot 1 \cdot 1.016 = 0.870$$

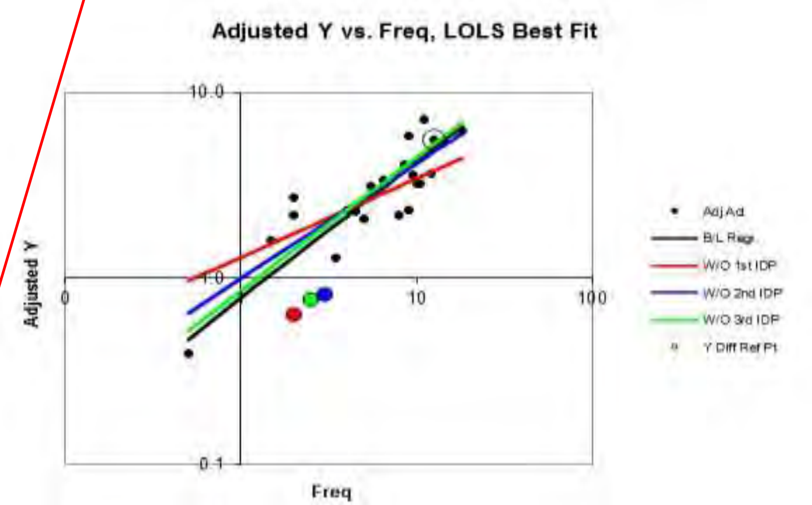
Adj. Act Y, ZMPE Charts		Adj. Act Y, LOLS Charts		Adj. Act Y, AAPE Charts		Adj. Act Y, MUPE Charts	
Weight	Freq	Weight	Freq	Weight	Freq	Weight	Freq
0.870	0.715	2.354	0.394	0.732	0.790	1.736	0.467
3.494	3.432	6.264	2.732	2.774	3.160	5.164	2.867

Adjusted Y vs. Weight & Frequency Graphs





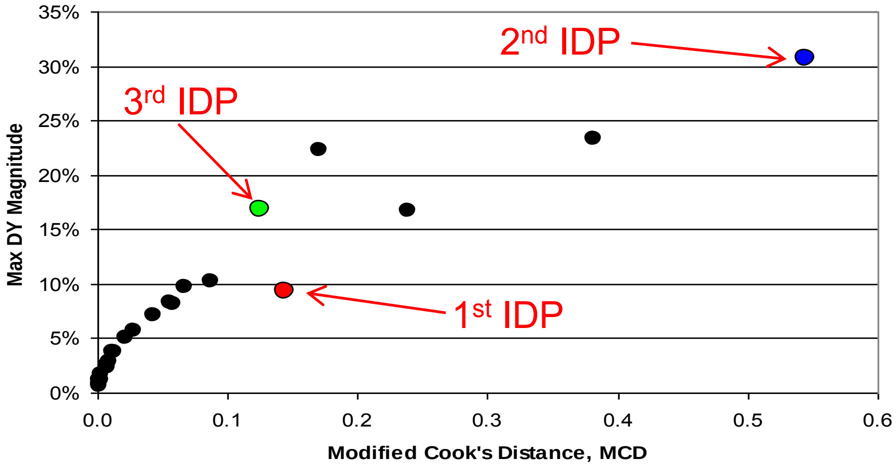




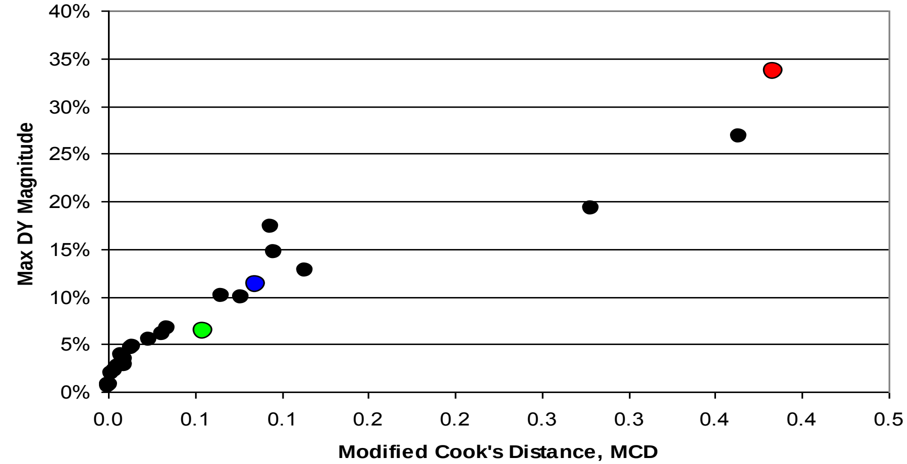
Individual graphs may seem counter-intuitive with more than one independent variable

Maximum DY vs. MCD Graphs

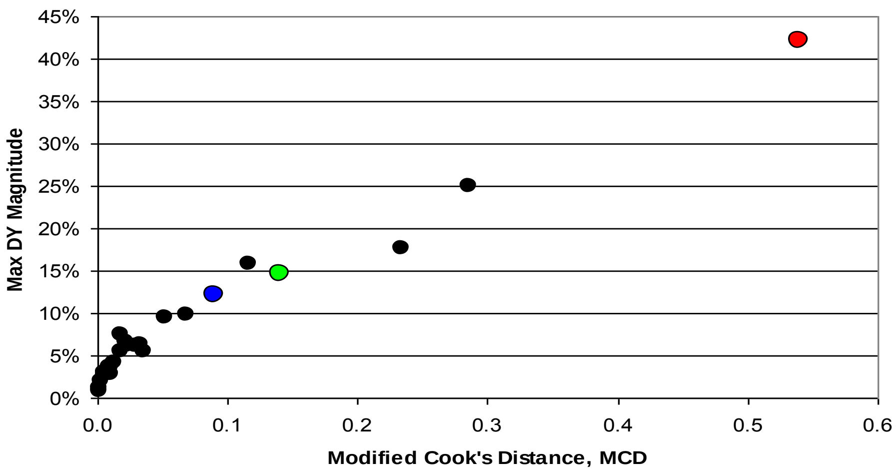
ZMPE Maximum DY Magnitude vs. MCD



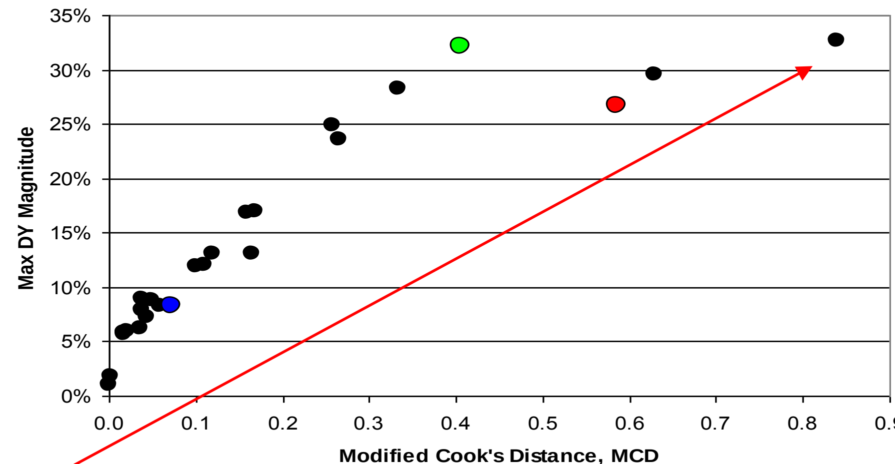
MUPE Maximum DY Magnitude vs. MCD



LOLS Maximum DY Magnitude vs. MCD

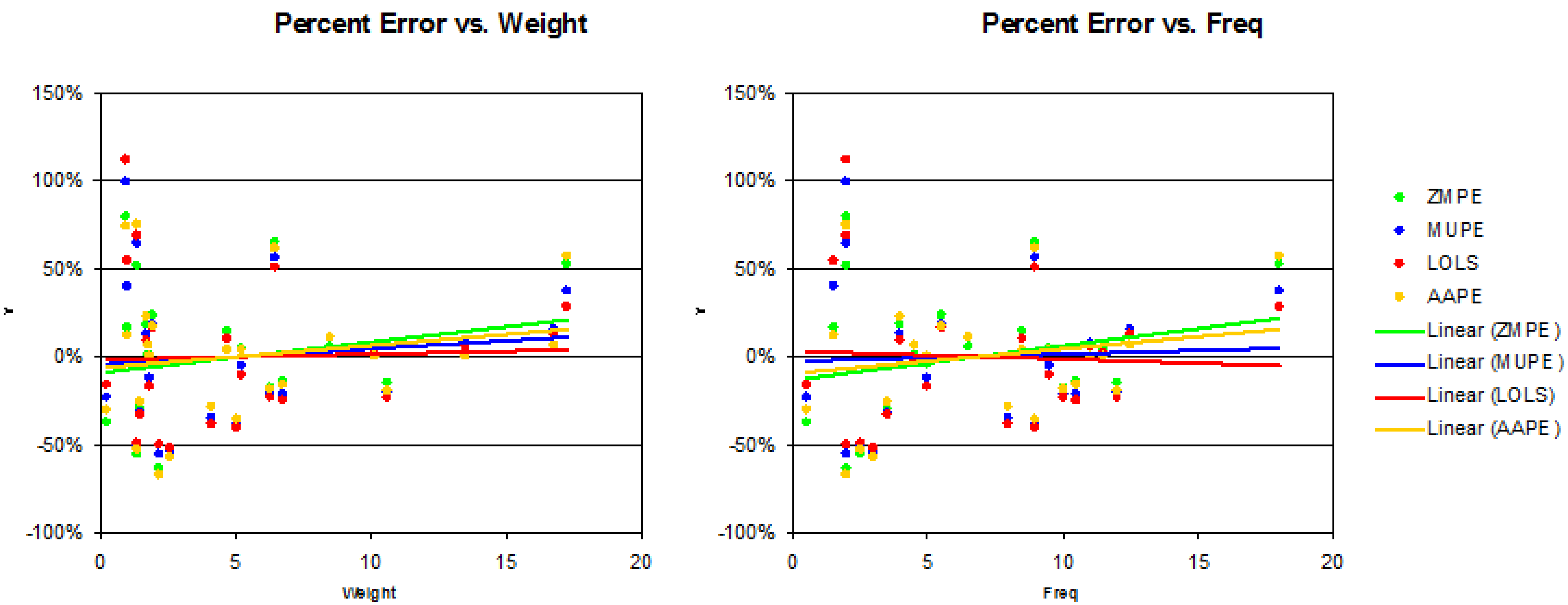


AAPE Maximum DY Magnitude vs. MCD



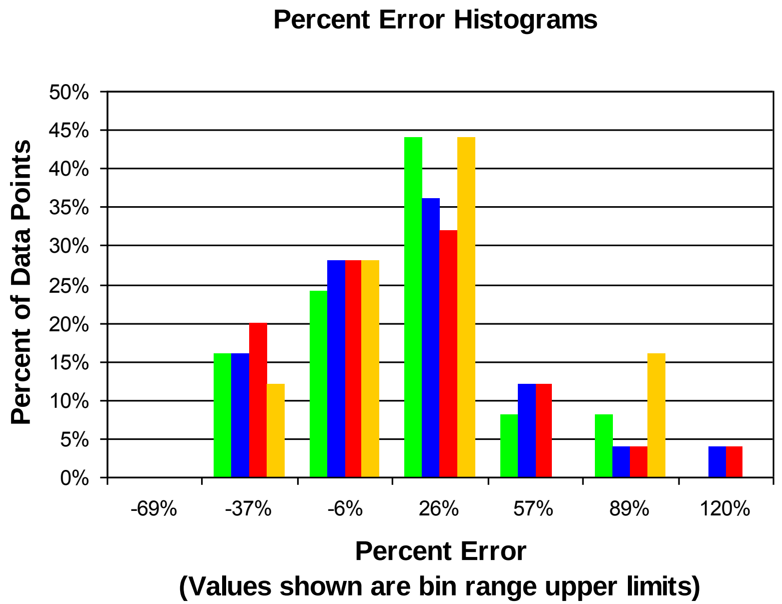
The data point with the maximum MCD may not be one of the top three IDPs

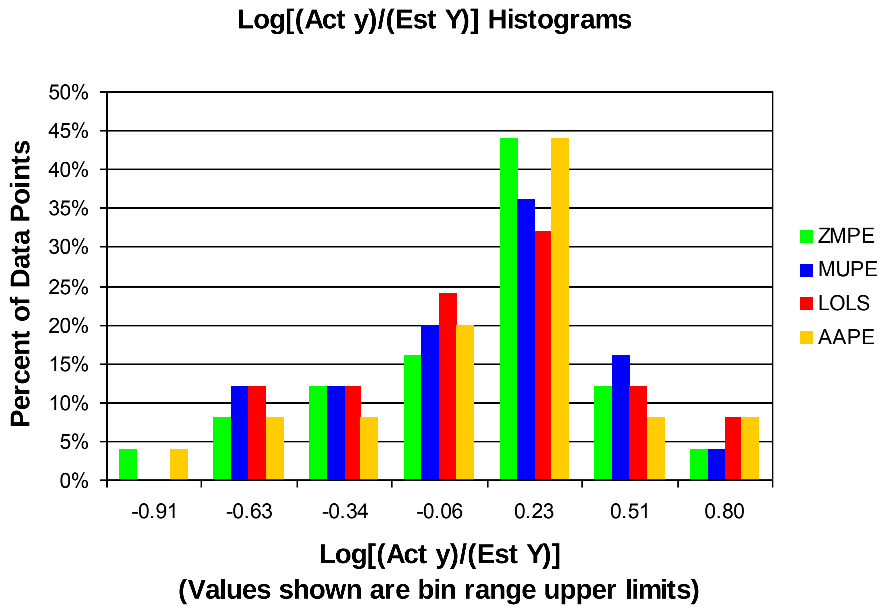
Percent Errors vs. Continuous Variables



LOLS and MUPE have flatter residual trend lines the ZMPE and AAPE
Log versions of graphs not shown

Error Distributions





The linear version shows right skew.
The log version suggests a truncated distribution

IV Swing Factor Definition

- A Swing Factor (SWF) defines how much IV values in the CER data set can impact (swing) a CER estimate
- The SWF is the ratio of the largest estimate to the smallest estimate resulting from:
  - Movement of the IV from its minimum to its maximum
- The SWF* for a continuous variable X is:

  (1) $SWF_X = (X_{max}/X_{min})$^B or (2) $SWF_X = (X_{min}/X_{max})$^B

  where B is the exponent for X

  SWF form 1 if B is positive, form 2 if B is negative
- For a stratifier variable Z, the SWF is:

  (1) D_Z if $D_Z > 1.0$ or (2) 1/ D_Z if $D_Z < 1.0$

  Where D_Z is the stratification factor for variable Z
- SWFs are a common-sense alternative to traditional T-stat tests

* SWF for a one-term CER. The SWF values for CERs with multiple terms are more complex

Correlation Matrix and Swing Factors

Variable	Y	Weight	Freq	Type A	Type B	Factor
Y	1.000					
Weight	0.870	1.000				
Freq	0.844	0.896	1.000			
Type A	0.234	0.306	0.223	1.000		
Type B	0.083	0.159	0.177	-0.315	1.000	
Factor	0.234	0.082	-0.060	0.073	-0.254	1.000
ZMPE SWFs	8.552	2.867	3.695	1.080	1.016	2.000
MUPE SWFs	9.536	1.545	9.254	1.022	1.036	2.000
LOLS SWFs	10.206	1.203	13.841	1.023	1.047	2.000
AAPE SWFs	8.294	3.801	3.366	1.187	1.092	2.000

Exponent = 0.103

Exponent = 0.044

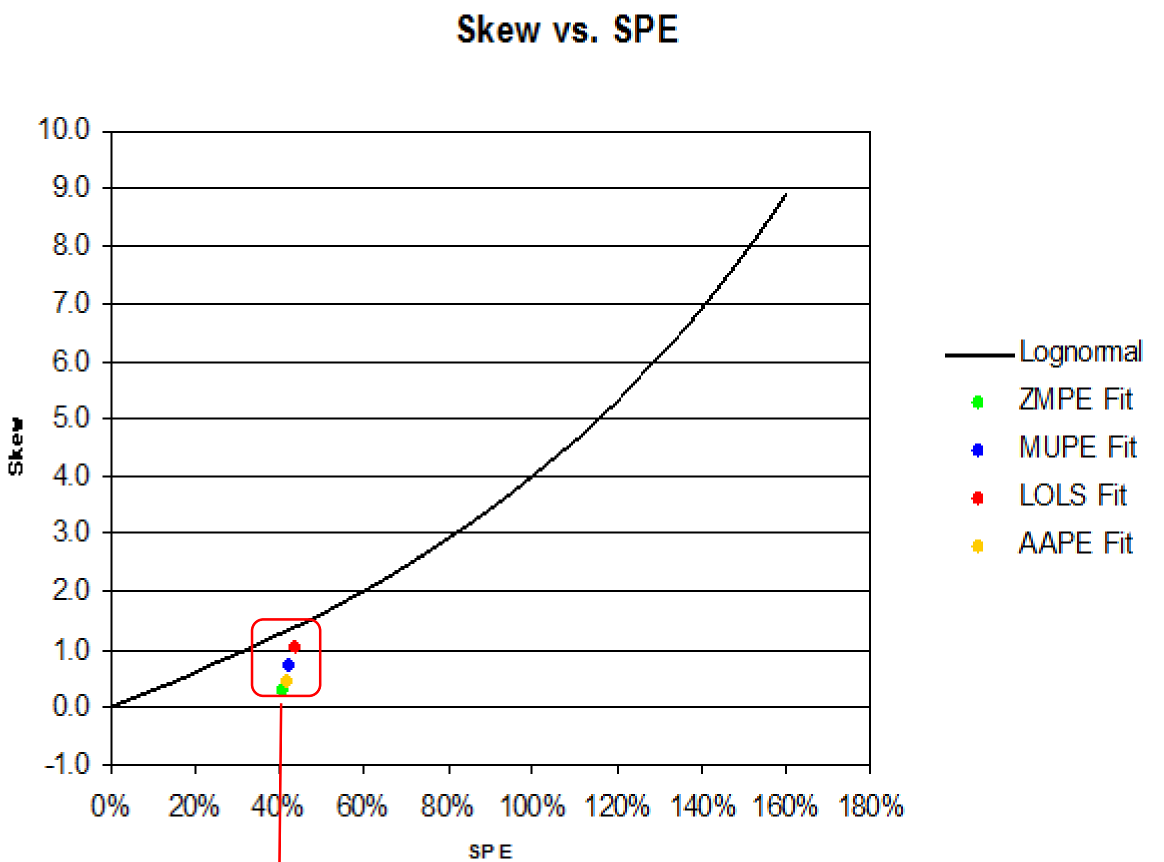
Note: Y values for SWFs are estimates with each variable at its maximum value.

Swing factors indicate variable is not significant

Marginally significant variables

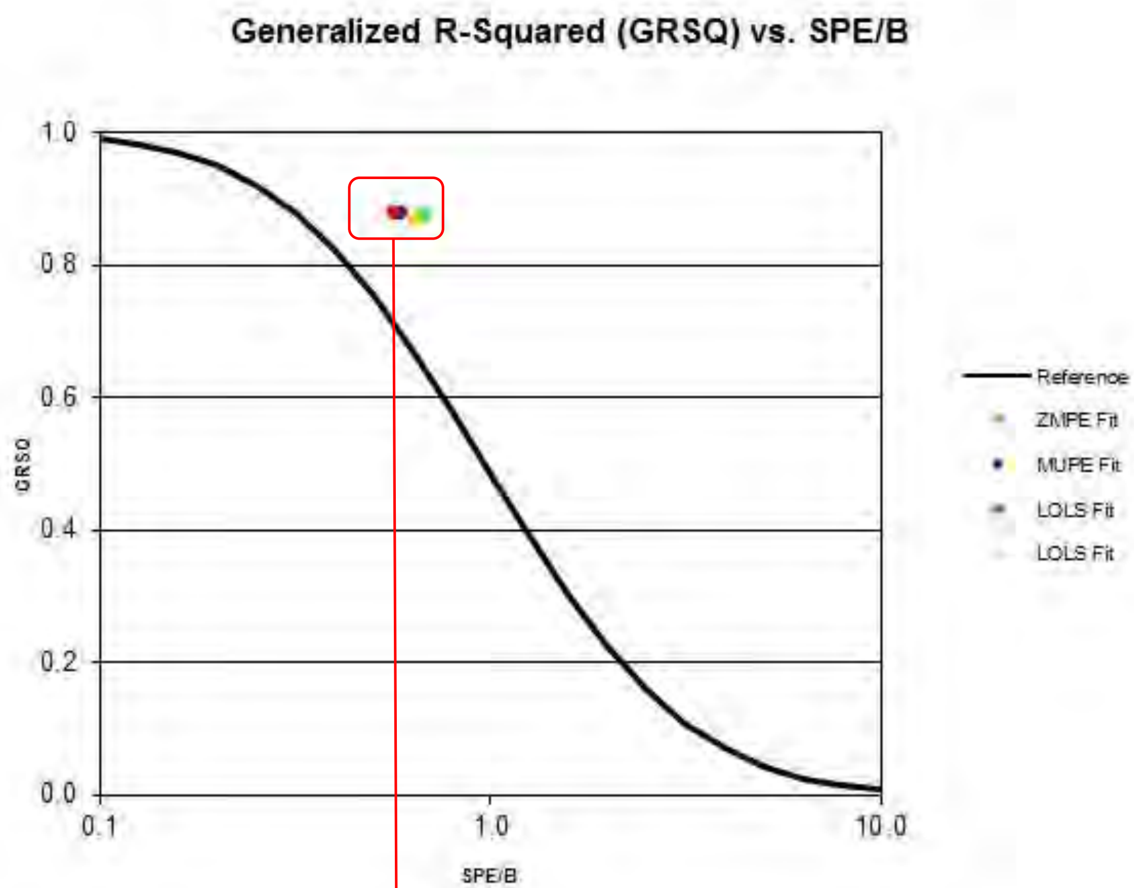
Variables with low exponents may have robust Swing Factors

Skew vs. Standard Percent Error



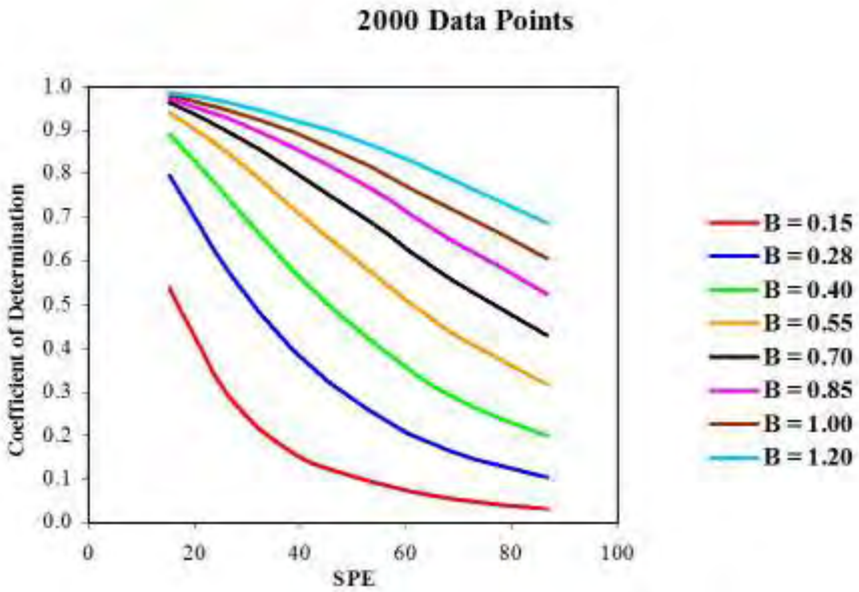
Skew is positive but less than exact lognormal (typical)

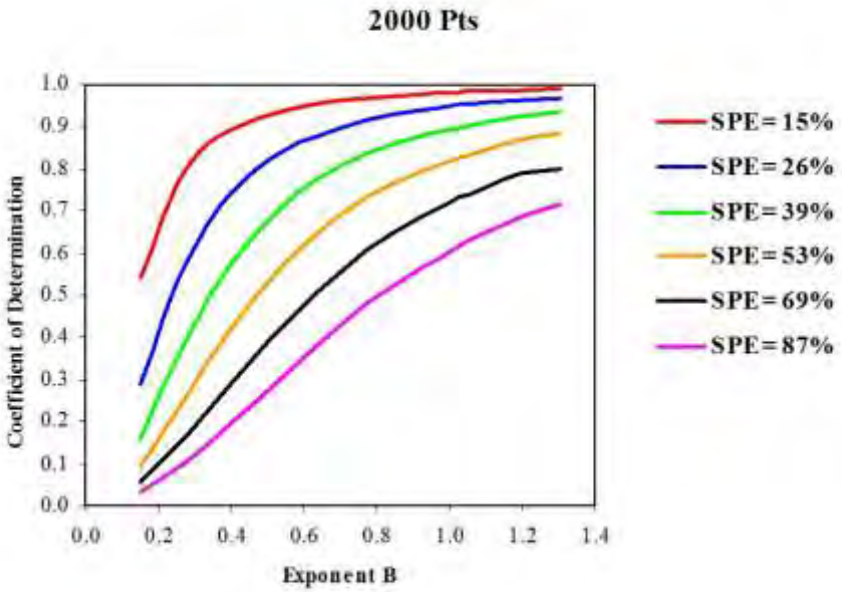
Generalized R-Squared and Skew



All methods are above one-variable average (good)

R^2 vs. SPE and Exponent B





These curves collapse into the single curve shown on the previous slide!

Topics

- CERAT Overview
- Background
- IDP Analysis Process
- CER Development Aids
- Summary

MC MacKenzie Consulting, Inc.

CAAG IDP Analysis Statistics

Description	Best-Fit Method	Box NR CERs	Box R CERS	S/W CERs	SEIT/PM CERs	All CERs
Number of CERs	**All**	**8**	**21**	**3**	**5**	**37**
Average Degrees of Freedom, DOF	**All**	**73**	**44**	**14**	**36**	**47**
Standard Percent Error, SPE	ZMPE	61%	50%	28%	52%	51%
	LOLS	65%	53%	28%	56%	55%
	AAPE	64%	52%	29%	52%	52%
Average 1st IDP NDY Magnitude	ZMPE	0.31	0.24	0.42	0.28	0.28
	LOLS	0.29	0.21	0.46	0.24	0.25
	AAPE	0.32	0.25	0.52	0.36	0.30
Average 2nd IDP NDY Magnitude	ZMPE	0.08	0.14	0.26	0.17	0.14
	LOLS	0.13	0.15	0.35	0.15	0.16
	AAPE	0.15	0.19	0.39	0.34	0.22
Average 1st IDP NDY Percentile	ZMPE	50	27	54	39	36
	LOLS	71	38	69	60	51
	AAPE	60	32	38	35	35
Average 1st IDP NDY Magnitude	Selected CER (ZMPE or LOLS)	0.28	0.22	0.42	0.25	0.25
Average 2nd IDP NDY Magnitude	Selected CER	0.11	0.15	0.26	0.18	0.15
Average 1st IDP NDY Percentile	Selected CER	56	33	54	37	41

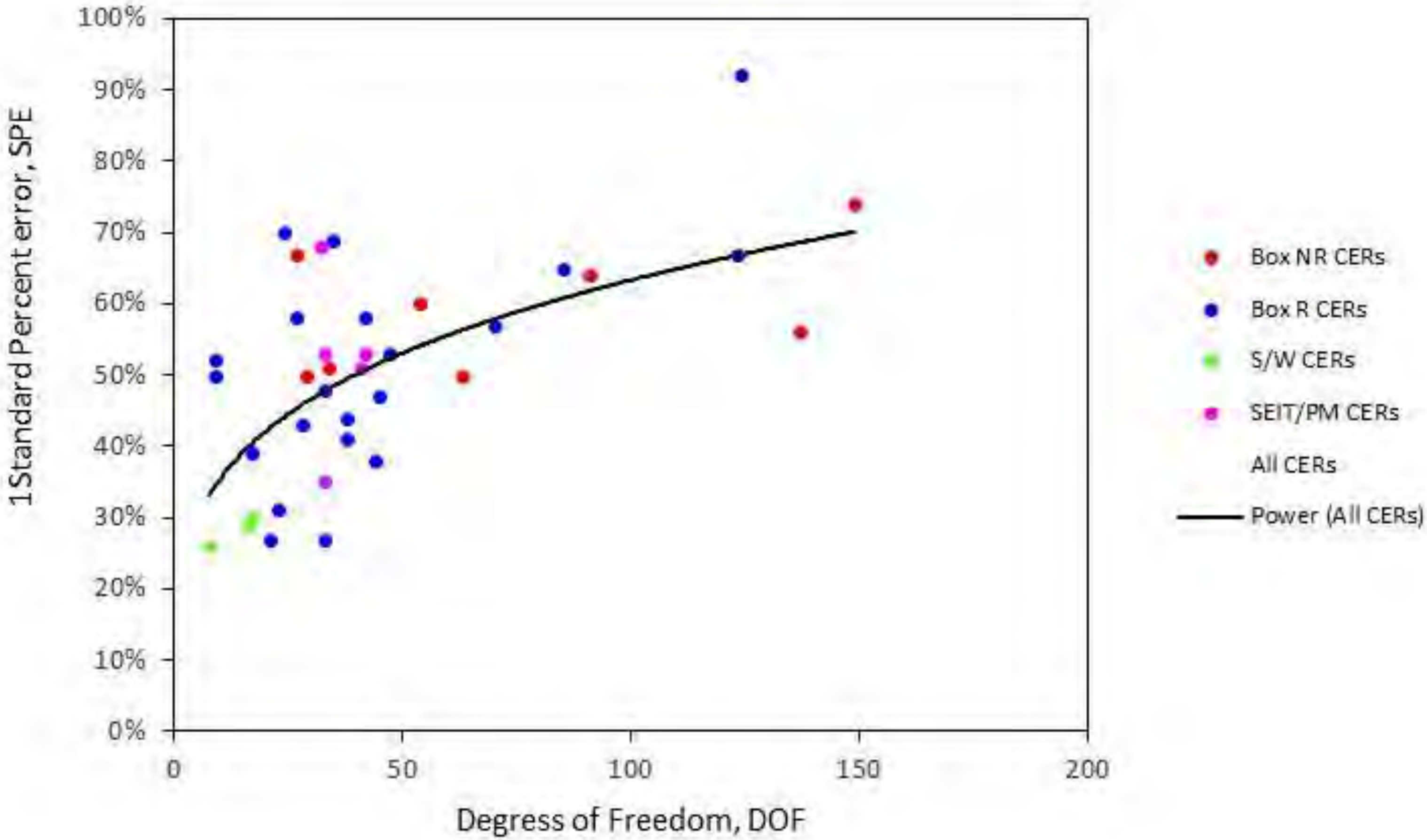
Caused by small data set

LOLS is lowest

LOLS is highest

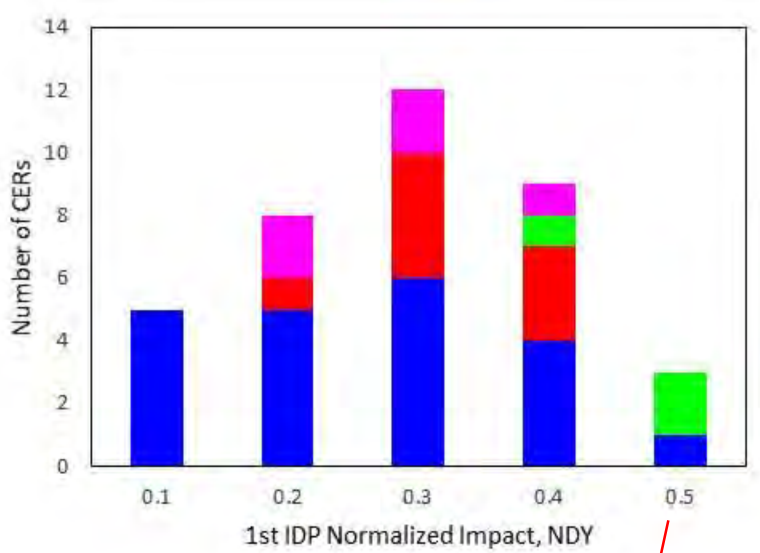
Close to 50th

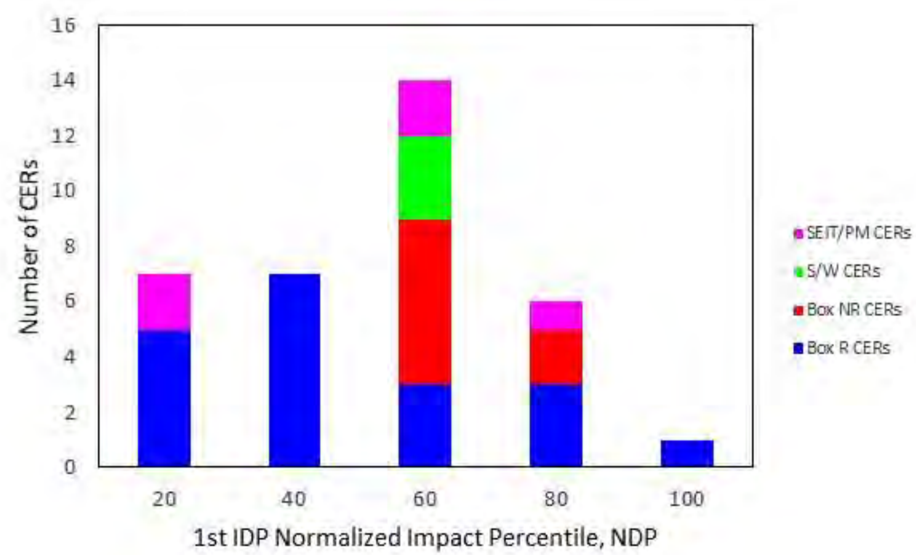
CER SPE vs. DOF



CERs with small DOF typically have low SPEs

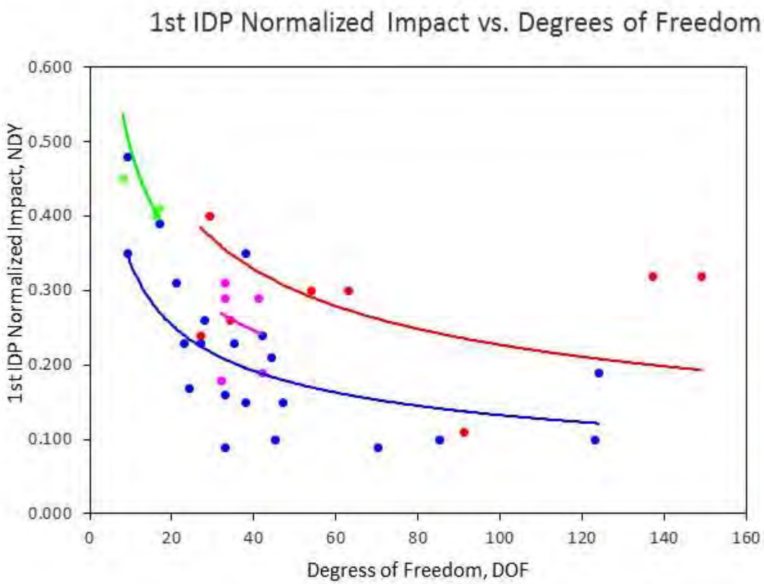
Histograms – 1st IDP NDY and NDP

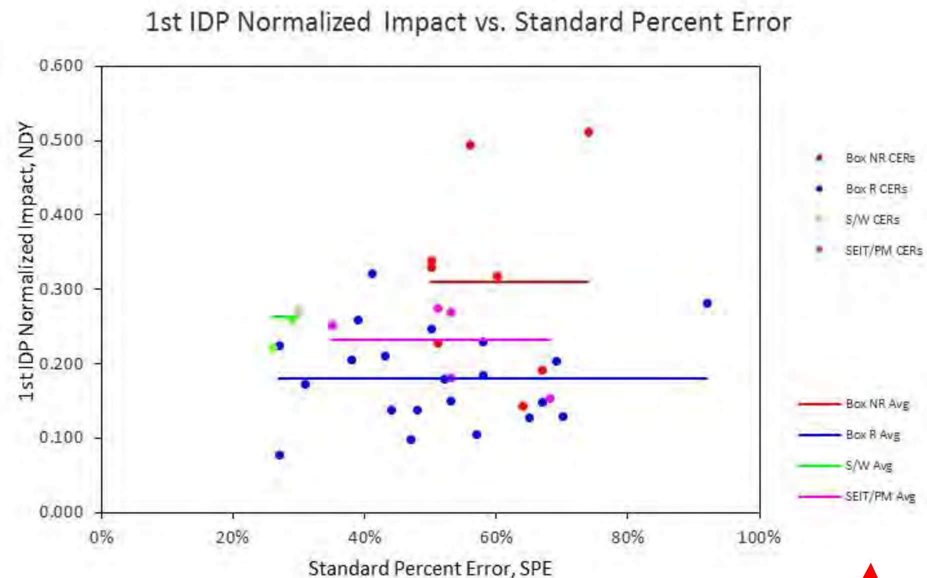


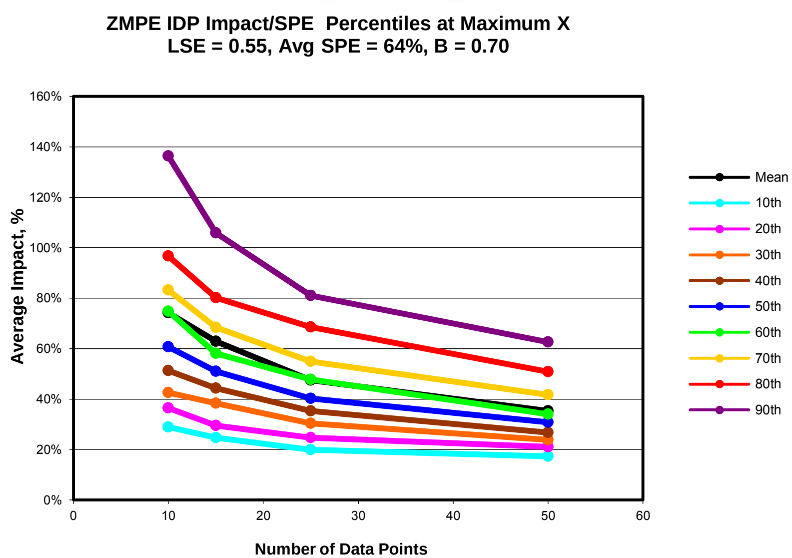


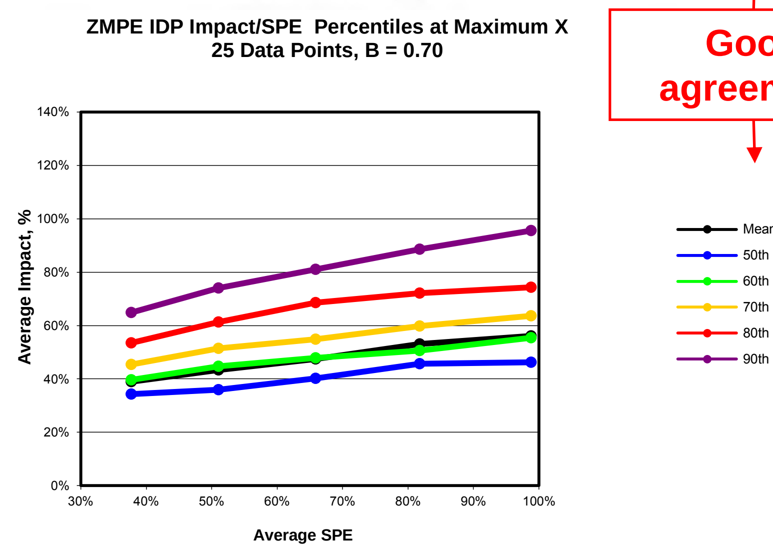
Max NDY = 0.48

Graphs – NDY vs. DOF & SPE





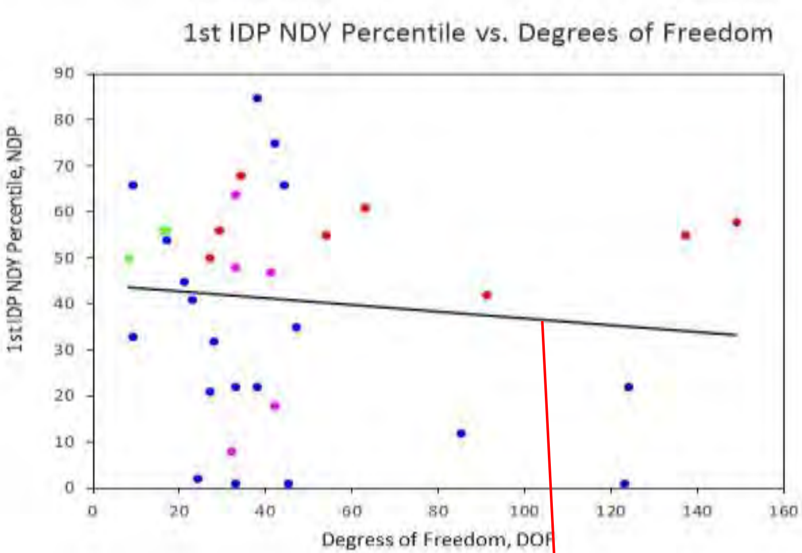


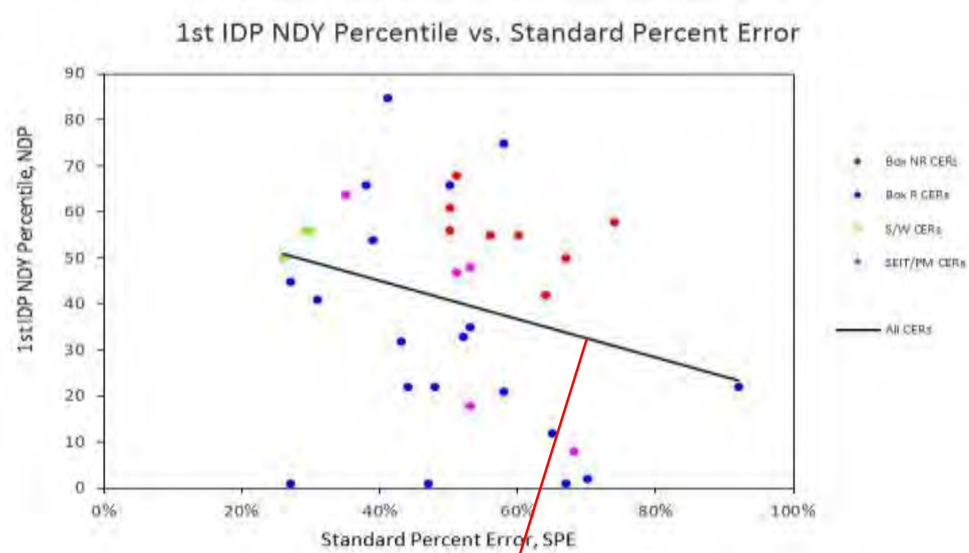


Good agreement

MacKenzie Consulting, Inc.

Graph – NDP vs. DOF & SPE





Good

Not so good

CER Development Recommendations

- Compare normalized DY magnitudes with those for other CERs with:
  - About the same DOF and SPE
  - About the same number of variables
- Pay attention to NDY percentiles for single-IV CERs
- Treat CERs with small data sets with care
  - Primarily by reviewing CER constants with/without an IDP
- Look at impacts other than target data point IDPs
  - CER constants, Max DY, MCD
- Review SWFs before removing a variable with a low exponent magnitude
- Review residuals histograms, skew graphics and GRSQ graphics for unusual data sets

CERAT – Experience and Benefits

- Used on 37 CER developments
- Results are in fair agreement with “theory”
- Seems to be helping analysis phase significantly
- Has changed the outcome in several cases
- Besides identifying the most influential data points it has:
  - Identified surprise impacts by less influential data points
  - Shown how the CER constants change with removal of each data point (one at a time)
- Not only aids decisions about removing data points
- But also helps assess the stability of LOLS and ZMPE solutions

Backup Charts

This research was jointly sponsored by MacKenzie Consulting, Inc. and the National Reconnaissance Office Cost and Acquisition Assessment Group (NRO CAAG). However, the views expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of the NRO CAAG or any other organization of the U.S. government.

Example ZMPE CER Worksheet

CER Example 3A -- Four Variables, 25 Data Points

ZMPE Best Fit

Leading Constant	**A**	**1.036**	**Bias**	**-5.0E-07**
Weight Exponent	**B**	**0.250**	**SPE**	**41.0%**
Frequency Exponent	**C**	**0.365**	**N**	**25**
Type A Stratum Multiplier	**D**	**0.926**	**K**	**5**
Type B Stratum Multiplier	**E**	**1.016**		

Stratum Bias >>> 3.2E-01 6.9E-02

Data Point Descriptions

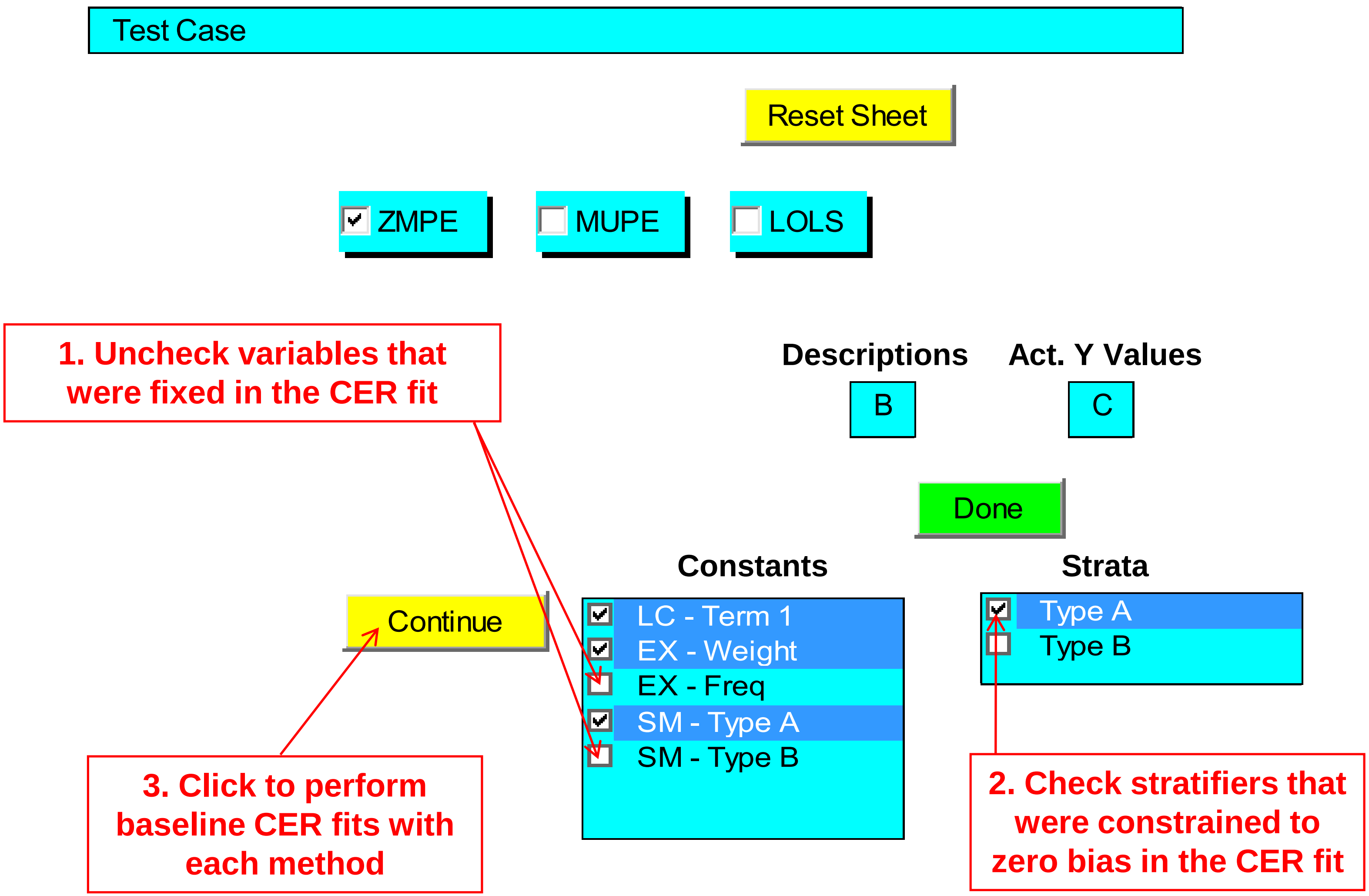
Actual Y Values

Select Y estimate for 1st data point

N	Data Point Description	Actual Y	Weight	Freq	Type A	Type B	Factor	Est Y	% Error
1	Data Point 1	0.329	0.256	0.5	1	0	1.000	0.530	-37.9%
2	Data Point 2	2.366	0.947	2.0	0	0	1.000	1.316	79.8%
3	Data Point 3	1.395	0.981	1.5	0	0	1.000	1.195	16.7%
4	Data Point 4	2.050	1.344	2.0	1	1	1.000	1.352	51.7%
5	Data Point 5	0.705	1.352	2.5	0	1	1.000	1.585	-55.5%
6	Data Point 6	5.636	6.427	9.0	1	0	1.000	3.406	65.5%
7	Data Point 7	1.170	1.473	3.5	1	0	1.000	1.670	-29.9%
8	Data Point 8	2.133	1.678	4.0	1	0	1.000	1.811	17.8%
9	Data Point 9	2.138	1.788	4.5	0	1	1.000	2.107	1.5%

CER Sheet / Controls / Baseline Regressions / IDP Analysis / ZMPE Fit / MUPE Fit / LOLS Fit / AAPE Fit

Defining Fixed Variables and Bias Constraints



Baseline Regression Statistics

Baseline Best-Fit Regressions -- Test Case 2/8/12 1:06 PM

From CER Sheet

Variables fixed in regressions

Constant	A	B	C	D	E
Original ZMPE Best Fit	1.052	0.263	0.350	0.925	1.000
CER Worksheet Column	N/A	D	E	F	G
CER Worksheet Header	LC	Weight	Freq	Type A	Type B
Used in Original Best-Fit?	Yes	Yes	No	Yes	No

B/L Regressions Constants

	A	B	C	D	E
ZMPE Baseline Fit	1.052	0.263	0.350	0.925	1.000
MUPE Baseline Fit	0.976	0.314	0.350	0.942	1.000
LOLS Baseline Fit	-0.068	0.343	0.350	-0.011	0.000
LOLS Linear Values	0.922	0.343	0.350	0.976	1.000
AAPE Baseline Fit	1.002	0.313	0.350	0.903	1.000

B/L Regressions Statistics

Max Cycle Change

	Bias	SPE	GRSQ	Skew	AAPE	MCC		
							MUPE Reweighting Cycles =	9
ZMPE Baseline Fit	-6E-07	39.1%	0.877	0.260	28.9%	1E-10	MUPE Max Bias Change =	1E-04
MUPE Baseline Fit	-3E-07	39.7%	0.884	0.507	28.6%	1E-05	LOLS Log Standard Error =	0.185
LOLS Baseline Fit	3E-17	40.8%	0.888	0.734	29.4%	1E-06	LOLS Zero Bias Factor =	1.079
AAPE Baseline Fit	8E-08	39.5%	0.881	0.450	28.4%	6E-08	LOLS Log Const. With ZBF =	-0.035

Variable Means	14.610			5.344	6.880	0.600	0.320	1.010		
ZMPE & AAPE Strata Bias			ZMPE:			2E-02	2E-02		AAPE:	2E-02	3E-04

Data Pt.	Description	Original Regression Data and ZMPE Baseline Fit								MUPE Baseline Fit			
		Act Y	Est Y	% Error	Weight	Freq	Type A	Type B	Factor	Est Y	% Error	Prev Y	Prev 2 Y

MacKenzie Consulting, Inc.

IDP Target Data Point Selection

CER Example 3A -- Four Variables, 25 Data Points

ZMPE Best Fit

Leading Constant	**A**	**1.036**	**Bias**	**-4.0E-09**
Weight Exponent	**B**	**0.250**	**SPE**	**41.0%**
Frequency Exponent	**C**	**0.365**	**N**	**25**
Type A Stratum Multiplier	**D**	**0.926**	**K**	**5**
Type B Stratum Multiplier	**E**	**1.016**		

Stratum Bias >>> **3.2E-01** **6.9E-02**

N	Data Point Description	Actual Y	Weight	Freq	Type A	Type B	Factor	Est Y	% Error
20	Data Point 20	3.188	6.291	10.0	0	1	1.000	3.861	-17.4%
21	Data Point 21	3.134	6.758	10.5	1	0	1.000	3.648	-14.1%
22	Data Point 22	7.090	13.480	11.0	1	0	1.500	6.615	7.2%
23	Data Point 23	4.766	10.130	11.5	0	1	1.000	4.577	4.1%
24	Data Point 24	3.620	10.605	12.0	1	0	1.000	4.287	-15.6%
25	Data Point 25	5.627	16.788	12.5	1	0	1.000	4.881	15.3%

Data Point 25 selected as target (any cell in row can be selected)

Notes | **CER Sheet** | Controls | Baseline Regressions | IDP Analysis | ZMPE Fit | MUPE Fit | LOLS Fit | AAPE Fit

Maximum DY vs. MCD vs. IDP

- $Abs(MaxDY)_j$ = The magnitude of the maximum percent impact over all data point estimates when data point j is removed
- Modified Cook's Distance is likely to be correlated with Abs(MaxDY)
- The data point with the largest Abs(MaxDY) is not necessarily the 1st IDP for a target data point
- The data point with the largest MCD is not necessarily the 1st IDP for a target data point
- The data point with the largest Abs(MaxDY) is usually (but maybe not always) the data point with the largest MCD