CRITICAL THINKING.
SOLUTIONS DELIVERED.

Covarus
Business · Math · Science

A Novel Non-Recurring Production CER Methodology

ICEAA Workshop – New Orleans, LA
June 18-21, 2013

Lisa Hackbarth
Associate – MCR LLC
lhackbarth@mcri.com

Raymond Covert
President – Covarus LLC
rpcovert@covarus.com

Outline

- Introduction
- Data
- Non-Recurring Estimate Methods
  - Factor Method
  - CER Method
  - Logistic Regression
  - Logistic-Regression-Enhanced CER
- Comparison
- Conclusion
- Works Cited

Introduction (1 of 3)

- In the process of developing a weapon system cost model, we regressed Recurring (R) Production CERs
  - Against various technical parameters (weight, range, etc.)
  - Using Zero Percent Bias, Minimum Percent Error (ZMPE) Unit-as-an-Independent Variable (UAIV) regression technique (Covert, R., & Wright, N., 2012)
- For completeness, we needed to include Non-Recurring (NR) Production costs as well
  - Typically done with a NR/R factor, but is there a better way?

Introduction (2 of 3)

- But there are problems with estimating NR Production costs as opposed to recurring production costs, such as…
  - Inconsistencies in the ways contractors breakout NR and R production costs
    - Some don't break it out at all
    - They have differing definitions of what constitutes NR
  - Not every lot has NR costs
  - Typically a decrease in NR costs in later lots
  - Unknown and unreported cost drivers
    - NR costs not necessarily driven by same variables as R Production costs

Introduction (3 of 3)

- The problem we faced was how to create an estimating method with a binary data set
  - In some cases the NR production cost data showed recognizable trends with quantity, and
  - In other cases data were absent or had zero values
- This creates a situation in which CER development is hindered by very poor fit statistics and a lack of recognizable cost drivers
- We will present the data and demonstrate how we applied a novel NR Production CER methodology

Outline

- Introduction
- Data
- Non-Recurring Estimate Methods
  - Factor Method
  - CER Method
  - Logistic Regression
  - Logistic-Regression-Enhanced CER
- Comparison
- Conclusion
- Works Cited

Data

- The data consist of the NR production costs of low rate initial production (LRIP) and full scale production (FSP) lots (in FY13$K)
  - We also include our estimate of the lot recurring production costs based upon previously developed CERs
- We gathered the first and last production units for each of these lots
- We then calculated an assumed lot midpoint (LMPA) based on the pooled learning curve slope derived from all of the production lots

Data Table (1 of 3)

Task	First	Last	LMP_A	NR 2013$K	REC^
LRIP 1 of 1	1	15	6	$ 238	$ 40,273
Lot #1	1	15	6	$ -	$ 19,726
LRIP 1 of 2	1	66	22	$ 24	$ 80,370
Lot #1	1	67	22	$ -	$ 37,229
Lot #1	1	80	27	$ 56,680	$ 78,912
Lot #1	1	132	43	$ -	$ 266,255
Lot #1	16	85	44	$ 14,095	$ 50,188
Lot #1	16	90	46	$ 408	$ 108,438
Lot #2	31	138	76	$ 3,294	$ 183,361
Lot #1	1	240	76	$ 3,196	$ 370,935
LRIP 1 of 2	1	352	111	$ 16,083	$ 130,457
LRIP 2 of 2	67	170	113	$ 4,768	$ 77,785
Lot #1	1	390	123	$ 37,046	$ 206,763
Lot #2	68	251	147	$ 5,623	$ 58,365
Lot #2	86	295	177	$ 29,704	$ 99,756
Lot #2	91	290	178	$ 1,280	$ 193,524
Lot #2	81	316	182	$ 24,056	$ 131,098
LRIP 1 of 3	1	735	229	$ 29,126	$ 52,885
Lot #3	85	434	230	$ -	$ 153,683
LRIP 1 of 1	1	800	249	$ 29,635	$ 143,951
Lot #3	139	403	256	$ -	$ 313,396
Lot #1	171	488	312	$ 16,264	$ 175,895
Lot #3	241	524	371	$ -	$ 273,870
Lot #3	252	687	446	$ 5,079	$ 99,310
Lot #3	317	609	453	$ 24,969	$ 124,004
Lot #3	291	662	460	$ 23,333	$ 271,337
Lot #4	404	668	529	$ -	$ 252,485
Lot #2	489	788	631	$ 2,655	$ 134,516
Lot #5	525	974	735	$ 5,872	$ 353,932
Lot #4	435	1178	767	$ 163	$ 228,237

LMP_A= Assumed Lot Midpoint

REC^ = Estimated Recurring Production Cost

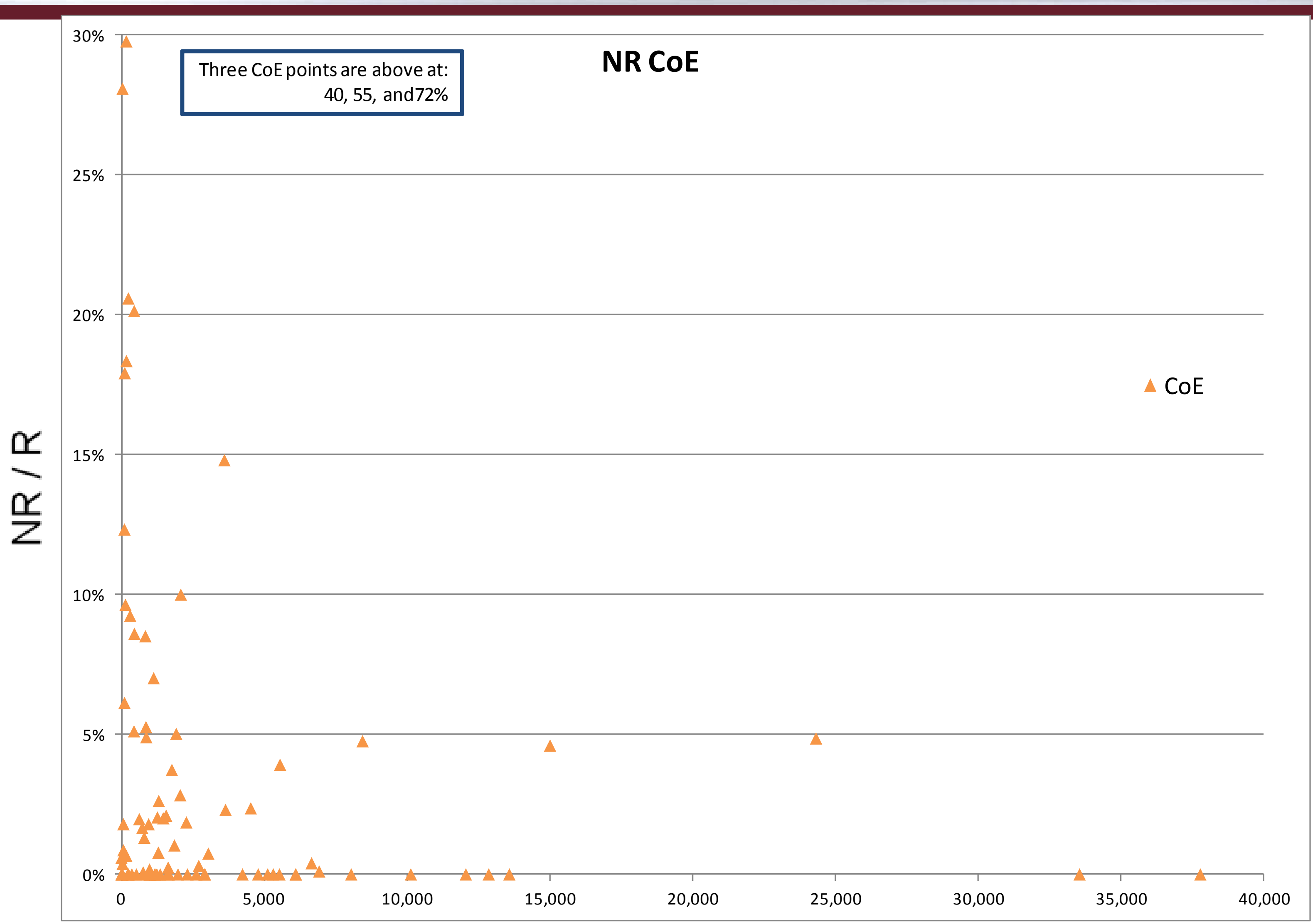
Data Table (2 of 3)

Task	First	Last	LMP_A	NR 2013$K	REC^
Lot #5	669	933	796	$ -	$ 223,493
LRIP 2 of 2	353	1408	805	$ 2,837	$ 216,673
Lot #4	506	1255	844	$ 19,042	$ 223,610
Lot #4	688	1061	866	$ 3,679	$ 69,924
Lot #4	663	1112	875	$ 13,286	$ 271,024
Lot #4	610	1259	909	$ 89,796	$ 223,468
Lot #3	789	1139	957	$ 2,493	$ 138,995
LRIP 2 of 3	874	1074	972	$ -	$ 36,498
Lot #1	1	3218	990	$ 657	$ 371,030
Lot #6	934	1086	1009	$ -	$ 120,251
Lot #2	391	1880	1016	$ -	$ 420,420
LRIP 2 of 3	736	1604	1134	$ 2,719	$ 38,786
Lot #7	1087	1266	1175	$ -	$ 135,183
Lot #6	975	1406	1182	$ -	$ 294,867
Lot #5	1113	1382	1244	$ -	$ 146,412
Lot #4	1140	1397	1266	$ 1,913	$ 94,002
Lot #5	1062	1555	1298	$ 638	$ 81,853
LRIP 2 of 3	1075	1572	1313	$ 2,170	$ 82,665
Lot #8	1267	1461	1362	$ -	$ 140,127
Lot #5	1398	1545	1471	$ 1,033	$ 51,564
Lot #6	1546	1595	1570	$ 359	$ 17,083
Lot #7	1407	1836	1615	$ -	$ 267,412
Lot #6	1383	1922	1643	$ 658	$ 269,552
Lot #9	1462	1881	1666	$ -	$ 284,266
Lot #6	1556	1996	1770	$ 2,488	$ 66,624
Lot #5	1256	2565	1861	$ 3,192	$ 308,631
Lot #1	1409	2508	1925	$ 8,744	$ 174,059
Lot #8	1837	2136	1984	$ -	$ 175,477
LRIP 3 of 3	1573	2610	2063	$ 4,264	$ 150,598
LRIP 3 of 3	1605	2620	2086	$ 3,781	$ 37,814

Data Table (3 of 3)

Task	First	Last	LMP_A	NR 2013$K	REC^
Lot #7	1997	2574	2278	$ 1,505	$ 81,001
Lot #9	2257	2376	2316	$ -	$ 67,027
Lot #3	1881	3473	2625	$ -	$ 338,724
Lot #6	2566	2865	2714	$ 191	$ 63,158
Lot #8	2537	3305	2910	$ -	$ 323,726
LRIP 3 of 3	2611	3264	2930	$ -	$ 85,471
Lot #2	2061	4197	3048	$ 1,353	$ 182,132
Lot #6	2954	4323	3610	$ 46,228	$ 312,283
Lot #1	3265	4052	3649	$ 2,224	$ 96,457
Lot #4	3474	5080	4244	$ -	$ 296,102
Lot #1	4053	5036	4533	$ 2,666	$ 112,912
Lot #12	4363	5239	4792	$ -	$ 318,187
Lot #3	4198	6127	5123	$ -	$ 140,901
Lot #2	3219	7861	5320	$ -	$ 324,151
Lot #13	5240	5837	5535	$ -	$ 207,843
Lot #1	2621	9395	5558	$ 7,380	$ 188,288
Lot #14	5838	6372	6102	$ -	$ 180,615
Lot #5	5081	7227	6113	$ -	$ 354,833
Lot #15	6373	6954	6661	$ 760	$ 191,423
Lot #2	2509	13131	6928	$ 1,190	$ 1,147,575
Lot #5	7228	8906	8048	$ -	$ 255,653
Lot #3	7041	9959	8446	$ 13,250	$ 278,251
Lot #6	8907	11433	10136	$ -	$ 359,212
Lot #7	11434	12700	12060	$ -	$ 171,012
Lot #8	12701	13022	12861	$ -	$ 42,636
Lot #4	11046	16343	13583	$ -	$ 464,430
Lot #2	9396	21714	15007	$ 11,719	$ 254,638
Lot #3	21715	27048	24318	$ 4,642	$ 95,480
Lot #5	30834	36347	33542	$ -	$ 222,375
Lot #6	36348	39210	37767	$ -	$ 111,450

NR Production Scatter Plot



LMP_A

Outline

- Introduction
- Data
- Non-Recurring Estimate Methods
  - Factor Method
  - CER Method
  - Logistic Regression
  - Logistic-Regression-Enhanced CER
- Comparison
- Conclusion
- Works Cited

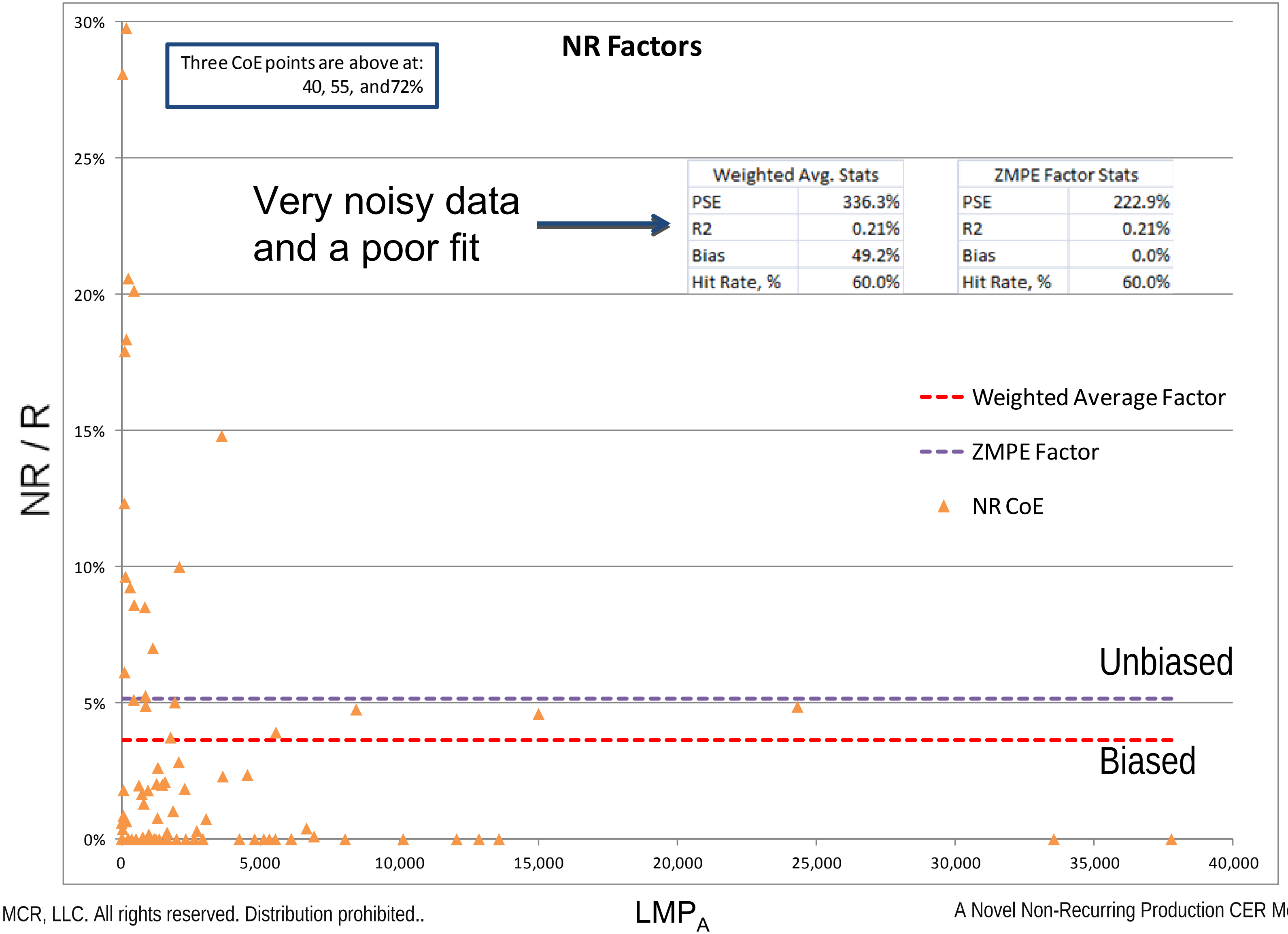
Non-Recurring Estimate Methods

- Initially, we used two methods of creating estimates of NR production costs
  - Factor Regression, $y=a*\varepsilon$
  - CER Regression, Log-Unit CER, $y= a + b * \ln(Unit) *\varepsilon$
- We then tried to treat the binary nature of the data using Logistic Regression
  - The goal was to be able to predict which NR production costs *should* be zero and which ones *should not*
- Finally, we combined the Logistic Regression and CER Regression to develop the estimates

Factor Method

- Cost-on-Estimate (CoE) factor, y=a*e
  – The coefficient "a" is the ratio of NR/R production costs
  – These factors were developed using the weighted average and ZMPE Method (which removes the bias)
- Drawbacks
  – Predicts NR costs for every lot (whether there were actual NR costs for that lot or not)
  – The statistics for the weighted average can be biased when compared to actual NR production lot costs, ZMPE method provides an unbiased factor

Factor Method



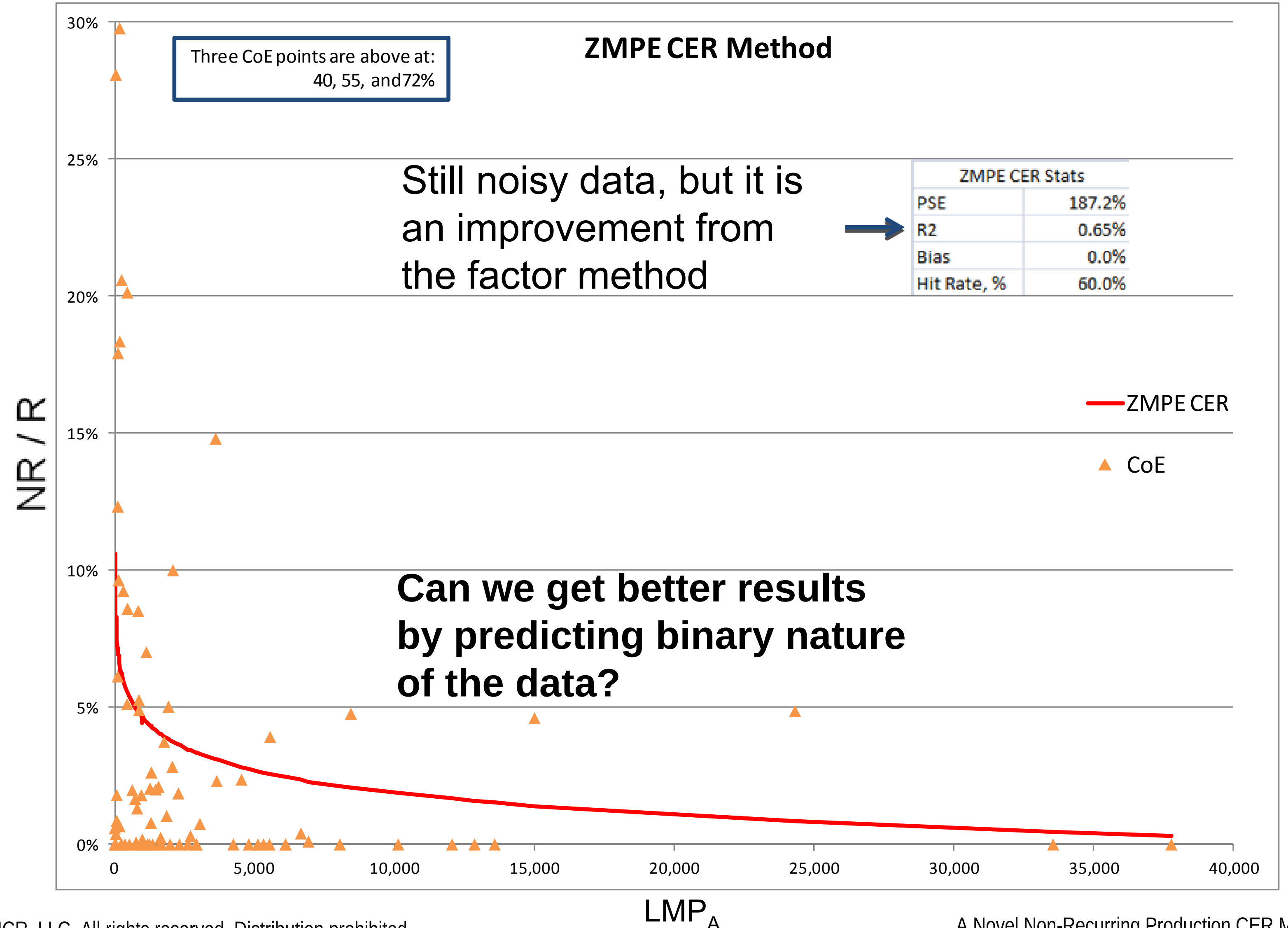
Weighted Avg. Stats	
PSE	336.3%
R2	0.21%
Bias	49.2%
Hit Rate, %	60.0%

ZMPE Factor Stats	
PSE	222.9%
R2	0.21%
Bias	0.0%
Hit Rate, %	60.0%

CER Method

- Regressed NR/R costs as a function of quantity
  – Used ZMPE regression technique
  – Scatter of data indicated a natural log (ln) equation would be the best fit: NR/R = a + b * ln(Unit)*ε
- Drawbacks
  – Predicts NR costs for every lot (whether there were actual NR costs for that lot or not)
  – Poor statistics partially due to zero-cost lots
  – When “unit” is large, can result in negative NR cost

CER Method



ZMPE CER Method
Three CoE points are above at: 40, 55, and72%
Still noisy data, but it is an improvement from the factor method
ZMPE CER Stats
PSE 187.2%
R2 0.65%
Bias 0.0%
Hit Rate, % 60.0%
Can we get better results by predicting binary nature of the data?
ZMPE CER
CoE
NR / R
30%
25%
20%
15%
10%
5%
0%
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
LMPA

Logistic Regression (1 of 3)

- Logistic regression used to find relationships between an independent variable and a series of dependent variables
- Commonly used in social sciences, but very rarely used in cost analysis
  - Used in cost growth studies (Lucas & White, 2009) (White, Sipple, & Greiner, 2004)
  - Used in this paper to model binomial (i.e., zero and non-zero) behavior of NR production cost data

Logistic Regression (2 of 3)

- Logistic regression uses the logistic function, π(x), to relate the dependent variables, X_i , to a range of values between zero and one (i.e. [0,1])

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}; 0 \leq \pi(x) \leq 1$$

where g(x) is the logit function (typically, but not necessarily, a linear relationship)

- The function g(x) is called the logit function which represents the log-odds of an event taking place

$$g(x) = a + bx_1 + cx_2 + dx_3 + ex_4 + \varepsilon$$

Logistic Regression (3 of 3)

- Several a priori variables and combinations were used until the highest 'hit rate' was found
  - 'Hit rate' is the percentage that the logit function correctly predicted the presence or lack of NR costs
  - Solved using ZMPE with additive error
- Best logistic regression found, g(x):

$g(x) = a+bx_1+cx_2+dx_3+\varepsilon$, where

x_1 = Natural log of the LMP_A, using the pooled learning curve slope

x_2 = Lot number, integer value for FSP lots and a fraction for each LRIP lot based on total program LRIP lots (e.g. for a program with 2 LRIP lots: LRIP 1 = 0.33, LRIP 2 = 0.66)

x_3 = First unit in the lot divided by 1000

a, b, c, and d are coefficients of the logistic regression

ε = The error of the regression

Logit Function Statistics

- Coefficients and fit statistics

g(x)	
a	2.760369895
b	-0.12299699
c	-0.348271119
d	-0.04408676
Measurement Data	
Observations	90
Coefficents	4
DoF	86
Error Statistics	
PSE	52.8%
Bias	(0.00)
Hit Rate, %	73.3%

- $f_{\pi(x)} = \begin{cases} 1 & if\ \pi(x) > 0.5 \\ 0 & otherwise \end{cases}$

  Interpretation: If π(x) > 0.5, the NR is deemed present for that lot

Logistic-Regression-Enhanced CER

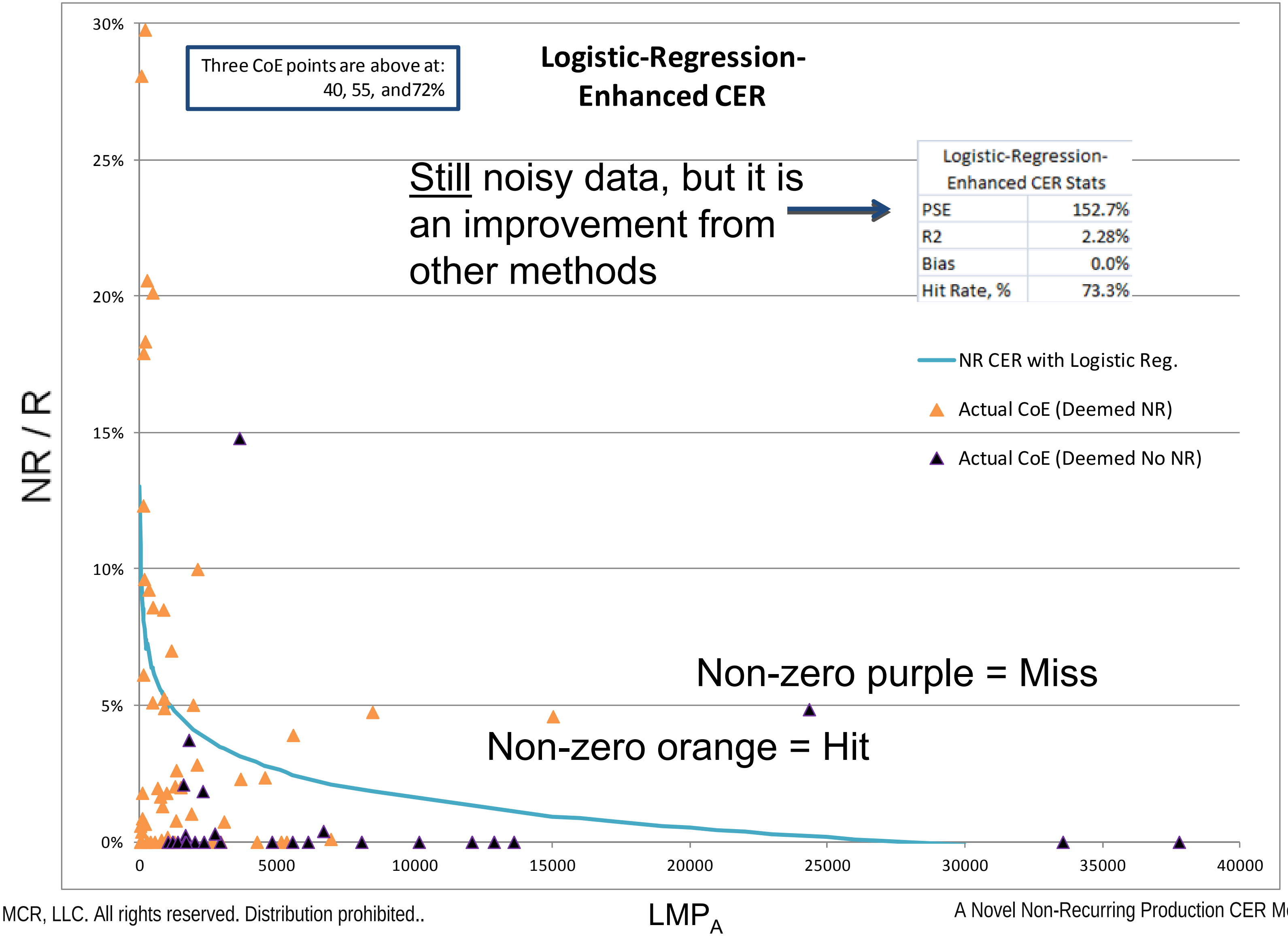
- Combined the logistic regression and CER regression in one step to find the coefficients of the equation

$$Y = f_{\pi(x)}(a + b * \ln(\text{Unit}))*\varepsilon$$

  - Regressed all lots (with zero and non-zero NR production costs) using the ZMPE method
- Goal is to find logistic-regression-enhanced CER with lowest PSE
  - If a hit and no actual NR, PSE for that lot is 0; if a miss and actual NR, PSE is -1; otherwise PSE calculated as:

$$\text{PSE} = 100\% * \sqrt{\frac{1}{n - m}\sum_{i=1}^{n}\left[\frac{y_i - f(x_i)}{f(x_i)}\right]^2}$$

Logistic-Regression-Enhanced CER



Logistic-Regression-Enhanced CER Stats	
PSE	152.7%
R2	2.28%
Bias	0.0%
Hit Rate, %	73.3%

Outline

- Introduction
- Data
- Non-Recurring Estimate Methods
  - Factor Method
  - CER Method
  - Logistic Regression
  - Logistic-Regression-Enhanced CER
- Comparison
- Conclusion
- Works Cited

Comparison

- The error statistics using the three methods are shown below

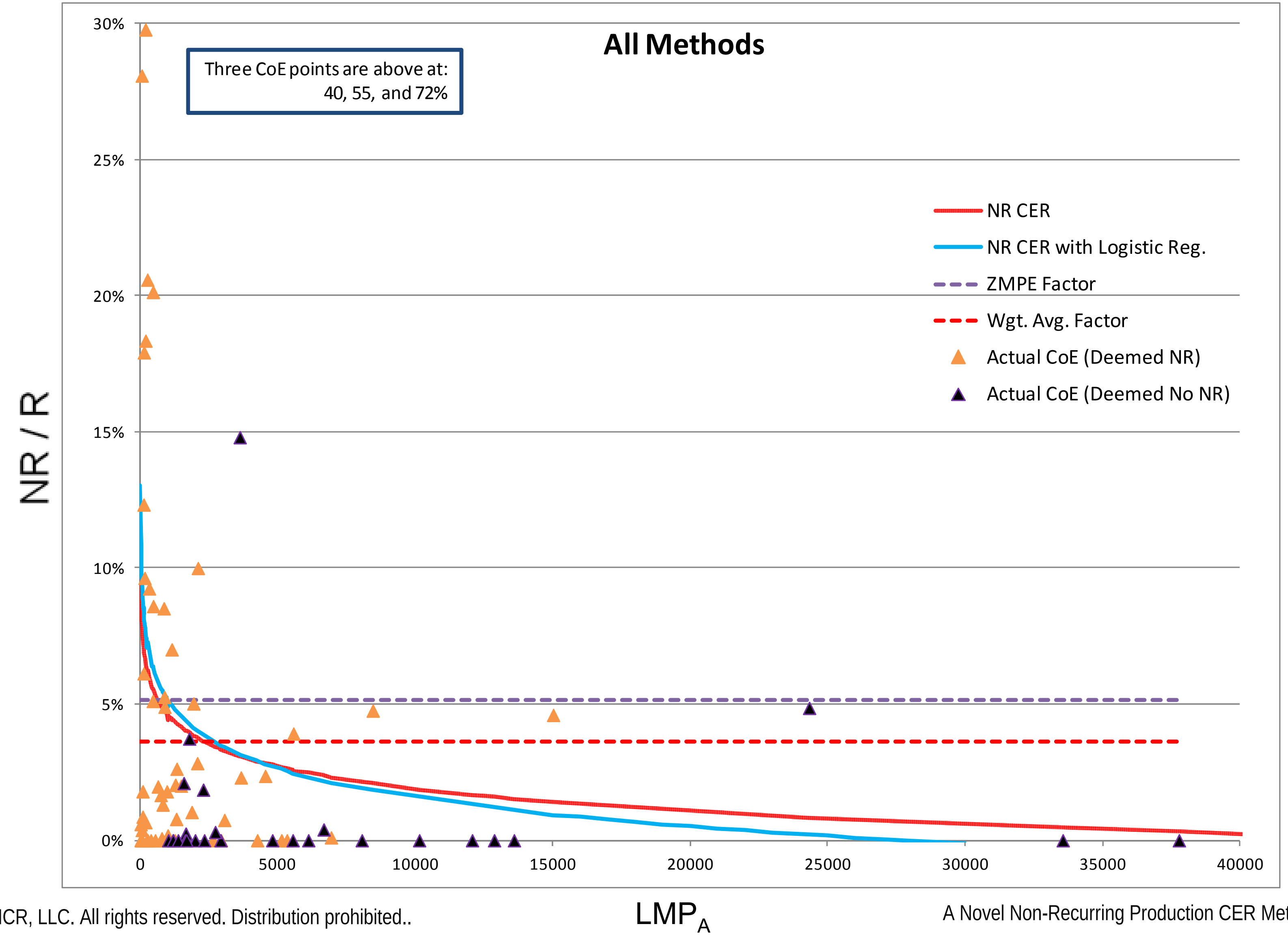
ZMPE Factor	
PSE	223%
R^2	0.2%
Pct Bias	0.0%
Hit Rate, %	60.0%

CER Method	
PSE	187%
R^2	0.7%
Pct Bias	0.0%
Hit Rate, %	60.0%

Logistic Reg + CER	
PSE	153%
R^2	2.3%
Pct Bias	0.0%
Hit Rate, %	73.3%

- While the PSE and R^2 statistics are still poor, we experienced successive improvements
  - Unit-driven CER is an improvement over a simple factor
  - Logistic-Regression-Enhanced CER is an improvement over the unit-driven CER

Comparison


All Methods
Three CoE points are above at: 40, 55, and 72%
NR / R
30%
25%
20%
15%
10%
5%
0%
0
5000
10000
15000
20000
25000
30000
35000
40000
NR CER
NR CER with Logistic Reg.
ZMPE Factor
Wgt. Avg. Factor
Actual CoE (Deemed NR)
Actual CoE (Deemed No NR)
LMP_A

Outline

- Introduction
- Data
- Non-Recurring Estimate Methods
  - Factor Method
  - CER Method
  - Logistic Regression
  - Logistic-Regression-Enhanced CER
- Comparison
- Conclusion
- Works Cited

Conclusion (1 of 2)

- NR production costs are difficult to estimate due to:
  - Binary nature of the data
  - Amount of scatter of the data
  - Lack of established NR production cost drivers with which to perform regressions
- This forced us to create NR factor-type CERs
- To improve the ability to estimate these costs, we combined logistic regression with CER development techniques

Conclusion (2 of 2)

- Logistic regression provides ability to predict binary nature of the data, and improve CER statistics
- Logistic-regression-enhanced CER improved the error statistics of the regression
  – Noise is reduced - PSE decreased 32% from the Factor Method and decreased 19% from the CER Method
  – Better modeling of binary nature of data - hit rate improved 22% over the other methods
  – Inherent correlation between actuals and estimates is starting to improve (through increased R^2), although still quite low, it is much greater than the other methods

Outline

- Introduction
- Data
- Non-Recurring Estimate Methods
  - Factor Method
  - CER Method
  - Logistic Regression
  - Logistic-Regression-Enhanced CER
- Comparison
- Conclusion
- Works Cited

Works Cited

- Covert, R. & Wright, N. (2012). Estimating Relationship Development Spreadsheet and Unit-as-an-Independent Variable Regressions. 2012 ISPA/SCEA Conference, Orlando, FL.
- Lucas, B. & White, E. (2009). Macro Approach to Estimate Engineering and Manufacturing Development Cost Growth. Cost Engineering, 51(6), 30-34.
- White, E., Sipple, V. & Greiner, M. (2004). Using Logistic and Multiple Regression to Estimate Engineering Cost Risk. Journal of Cost Analysis and Management, Summer, 67-79.

Questions?

MCR

CRITICAL THINKING.
SOLUTIONS DELIVERED.

Covarus

Business · Math · Science