COCOMO 2.0
Post-Architecture Calibration

Bradford K. Clark

10/10/96

Presentation Outline:

⇒ ***Motivation***

- Data collection activity
- COCOMO 2.0 Post-Architecture Model
- COCOMO calibration model
- Results to date
- Conclusions and future work

COCOMO Model Motivation:

- Need to address future software practices
- Update existing algorithmic cost model
- Model based on software engineering knowledge and observations
- Model values based on collected data from Industrial Affiliates to the Center for Software Engineering

Data Collection:

- Define the data needed.
- Collect data with a paper form or a computer software tool
- Affiliate Organizations providing majority of data.
  - Historical - whole project
- Site visits or phone interviews to record data
- Enter in data into the repository
  - Data is labeled with generic id
  - Stored in locked room
  - Limited access by researchers

Post-Architecture Model:

- Non-linear model:

$$PM = A \cdot (Size)^{B} \cdot \prod_{i=1}^{17} EM_i$$

- B consists of 5 scale factors:

$$B = 1.01 + 0.01 \cdot \sum_{j=1}^{5} SF_j$$

COCOMO Calibration Model:

- Need linear model for regression:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_p X_p$$

- COCOMO 2.0 Post-Architecture is non-linear

$$Y = B_0 X^{B_1}$$

- What should we do?
  - Expand COCOMO model
  - Transform products with logarithms to produce sums

Expanded COCOMO:

- Distribute the Scale Factors
- Results in 23 factors

$$PM_{est} = A \cdot (Size)^{1.01} \cdot (Size)^{SF_1} \cdot (Size)^{SF_2} \cdots EM_1 \cdots EM_{17}$$

Log Transformed COCOMO:

$$\ln(PM_{est}) = \ln(A) + 1.01\ln(Size) + SF_1 \ln(Size) + \cdots + \ln(EM_{17})$$

- Regression analysis will derive the coefficients, $\boldsymbol{B}_i$, for each factor
- ln(A) is dropped

Presentation Outline:

- Motivation
- Data collection activity
- COCOMO 2.0 Post-Architecture Model
- COCOMO calibration model

⇒ ***Results to date***

- Conclusions and future work

Results:

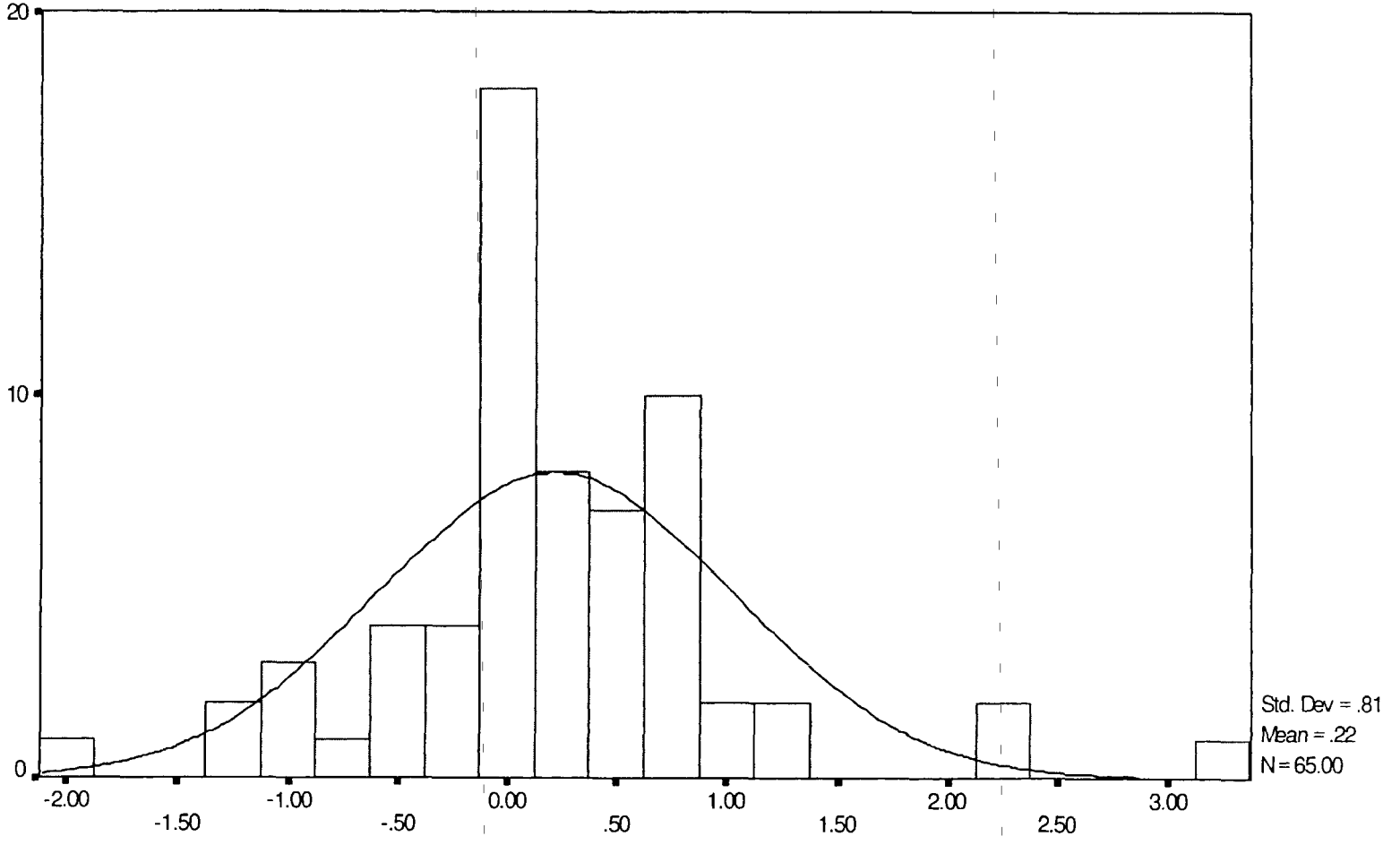
- 65 Observations from different Industrial categories:

Commercial:	2
Aerospace:	4
FFRDC:	2

- Results improved with stratification of data by organization
- Forecast accuracy measured with proportional error:

$$PE = \begin{cases} \left[PM_{est} \div PM_{act}\right] - 1, (PM_{est} - PM_{act}) \geq 0 \\ -\left[PM_{act} \div PM_{est}\right] + 1, (PM_{est} - PM_{act}) < 0 \end{cases}$$

PE Before Regression



Regression
(without stratification):

Adjusted R Square .94570
Standard Error .38872

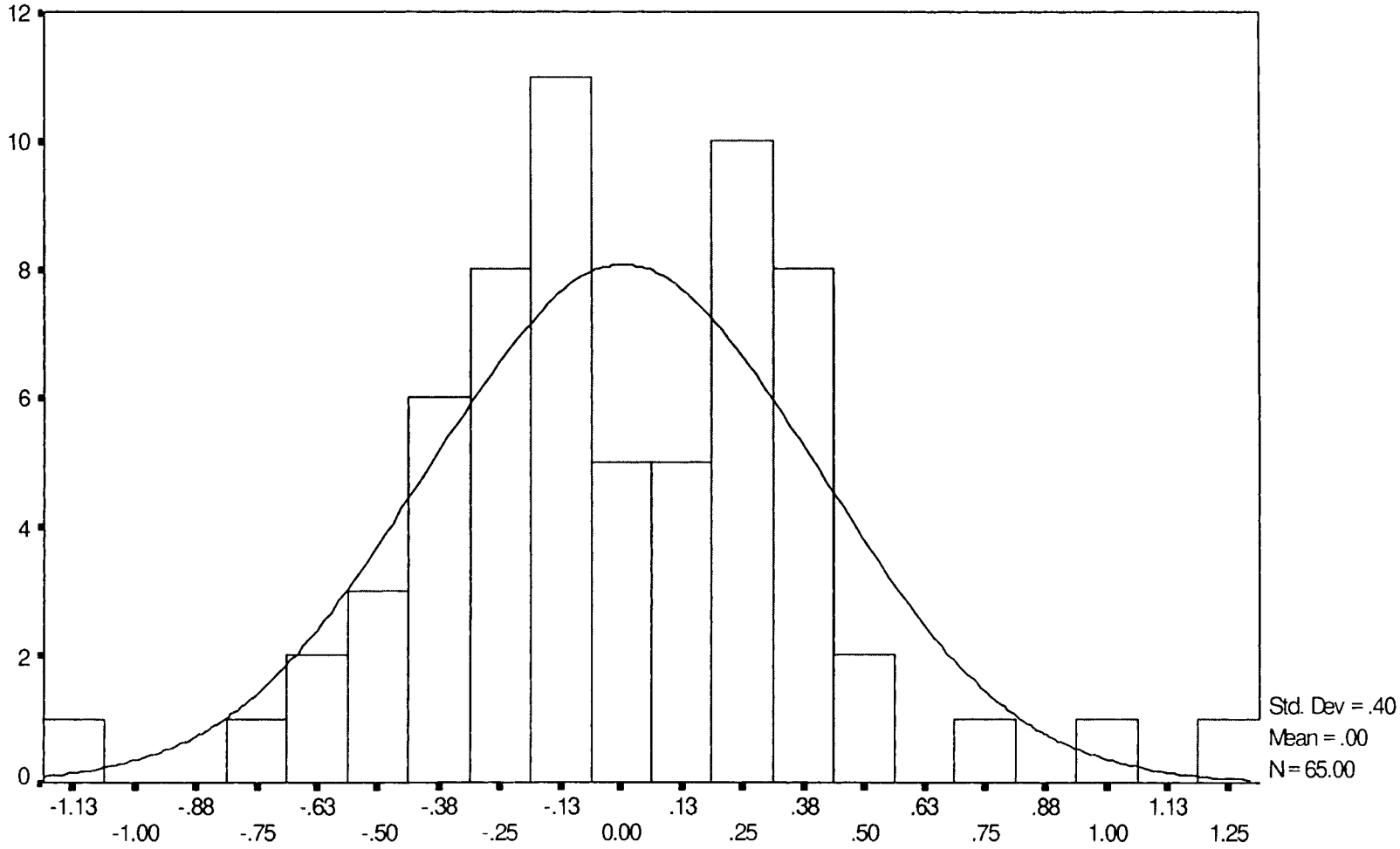
Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	23	171.88624	7.47331
Residual	41	6.19527	.15110

F = 49.45804

Variable	B	SE B	T
LN_ACAP	0.364606	0.695296	0.524
LN_AEXP	0.231304	0.602883	0.384
LN_CPLX	1.497551	0.608187	2.462
LN_DATA	2.321239	0.999881	2.322
LN_DOCU	-0.153751	0.929772	-0.165
LN_LTEX	0.124795	0.848733	0.147
LN_PCAP	1.236443	0.844417	1.464
LN_PCON	0.831002	1.073967	0.774
LN_PEXP	0.402171	0.595131	0.676
LN_PVOL	-0.045234	0.676307	-0.067
LN_RELY	0.584032	0.574250	1.017
LN_RUSE	-0.799948	0.609753	-1.312
LN_SCED	2.698342	1.243080	2.171
LN_SITE	-0.947197	1.157316	-0.818
LN_STOR	2.003848	0.849830	2.358
LN_TIME	1.209235	0.770581	1.569
LN_TOOL	3.058214	1.065924	2.869
LNSIZ101	1.107444	0.123345	8.978
LNS_FLEX	0.696658	1.275185	0.546
LNS_PMAT	1.084491	1.788497	0.606
LNS_PREC	2.413984	0.960125	2.514
LNS_RESL	-2.269417	1.956745	-1.160
LNS_TEAM	3.715327	1.995218	1.862
(Constant)	-0.560723	0.425930	-1.316

Regression PE:



Regression
(with stratification)

Adjusted R Square	.95331
Standard Error	.36043

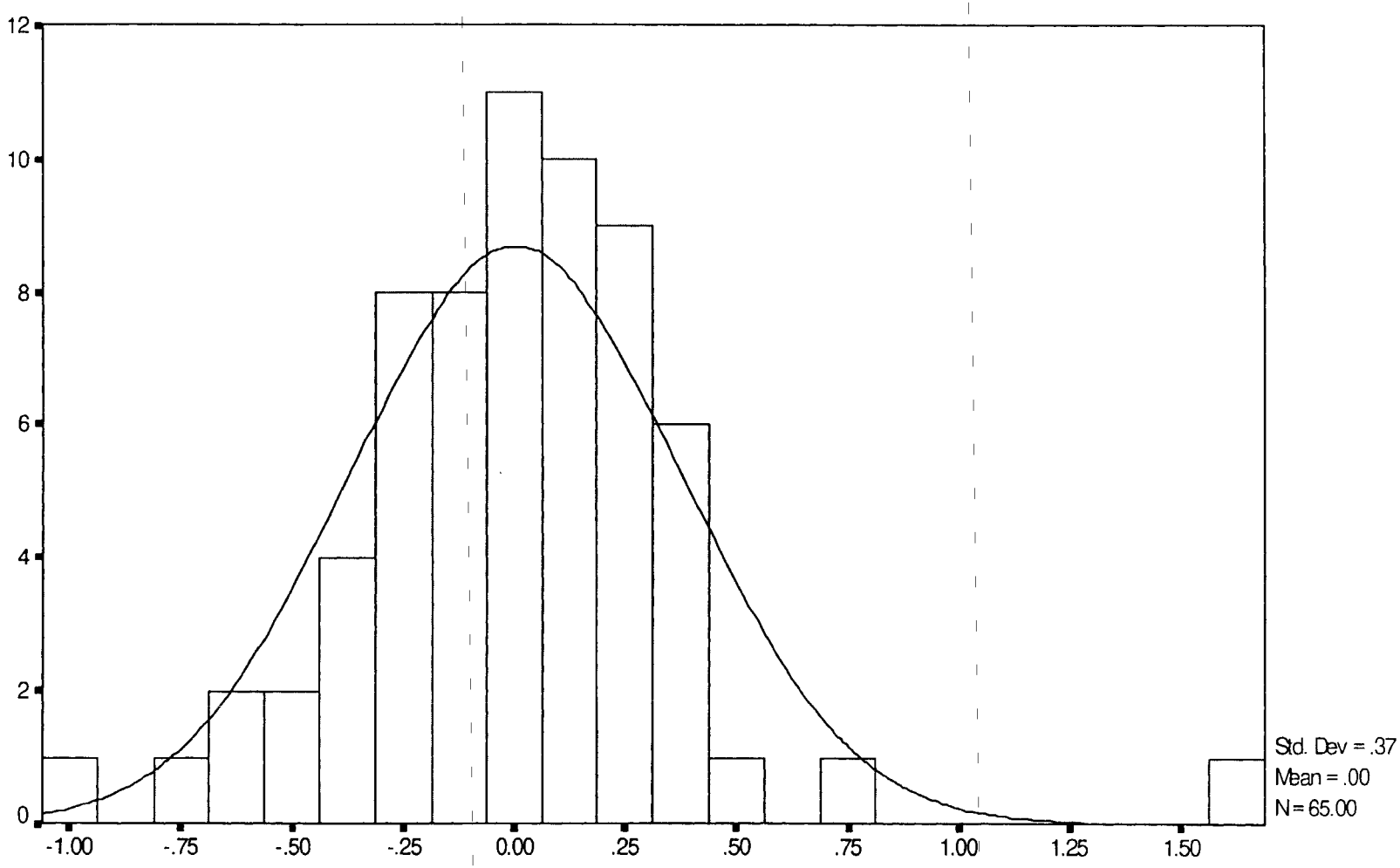
Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	26	173.14497	6.65942
Residual	38	4.93654	.12991

F = 51.26224

Variable	B	SE B	T
LN_ACAP	0.219985	0.664825	0.331
LN_AEXP	0.222100	0.561130	0.396
LN_CPLX	1.843231	0.599430	3.075
LN_DATA	2.852544	0.952041	2.996
LN_DOCU	0.649755	1.011005	0.643
LN_LTEX	0.423312	0.850311	0.498
LN_PCAP	1.659577	.841160	1.973
LN_PCON	0.470698	1.017761	0.462
LN_PEXP	0.586376	0.559265	1.048
LN_PVOL	1.225317	0.778112	1.575
LN_RELY	0.804534	0.538962	1.493
LN_RUSE	-0.289814	0.613725	-.472
LN_SCED	2.237578	1.237291	1.808
LN_SITE	-1.030253	1.098791	-.938
LN_STOR	0.739371	0.887684	.833
LN_TIME	1.277919	0.729614	1.752
LN_TOOL	2.366555	1.030659	2.296
LNSIZ101	1.024621	0.154880	6.616
LNS_FLEX	0.964983	1.321752	.730
LNS_PMAT	4.139000	2.761260	1.499
LNS_PREC	1.767164	1.141516	1.548
LNS_RESL	-1.807623	1.963190	-.921
LNS_TEAM	1.959708	2.128285	.921
ORG093	-1.038590	0.464163	-2.238
ORG587	-0.669031	0.366404	-1.826
ORG586	-0.167731	0.274594	-.611
(Constant)	-0.047728	0.539532	-.088

Regression PE:



Process Maturity Investigation with COCOMO 2.0:

- Assess effect of Process Maturity[1] on Software Development Effort within context of other influencing factors.
- Data collected on either CMM level of KPA Goals.
- These results from stratified analysis show a generally positive influence.
- Data needs to be inspected to determine cause of variation.
- More data points are needed.

1. As defined by SEI's Capability Maturity Model.

Conclusions:

- Regression technique can be used to calibrate COCOMO locally
- COCOMO calibrated to local organization is more accruate
- Qualify your data - inspect it, decide before-hand what an outlier looks like.

Future Work:

- Negative coefficients do not make sense in the model (check correlation's of parameter inputs)
- Schedule equation needs to be calibrated
- Calibration of COCOMO Early Design model