

Calibration Approach and Results of the COCOMO II Post- Architecture Model

**Sunita Chulani*, Brad Clark, Barry Boehm
(USC-Center for Software Engineering)**

**Bert Steece
(USC-Marshall School of Business)**

ISPA '98

Copyright USC-CSE 1998

***Email: sdevnani@sunset.usc.edu**

Outline

- ➔ **Brief History of COCOMO**
- ◆ **COCOMO II.1997**
 - Process
 - Prediction Accuracies
 - Comparison with COCOMO '81
- ◆ **Updates and Plans**
 - Plans for Improving Prediction Accuracies
 - COCOMO II Research Aim
 - Effects of Process Maturity on Effort
- ◆ **Acknowledgements and Information Sources**

COnstructive **CO**st **MO**del **(COCOMO)**

- ◆ **COCOMO published since 1981**
- ◆ **Commercial implementations of COCOMO**
 - ❖ **CoCoPro, CB COCOMO, COCOMOID, COSTMODL, GECOMO Plus, SECOMO, etc.**
- ◆ **Other models based on COCOMO**
 - ❖ **REVIC, Gulezian**
- ◆ **COCOMO II**
 - ❖ **Research effort started in 1994 to develop a 1990's-2000's software cost model**
 - ❖ **Address new processes and practices**
 - ❖ **COCOMO II.199Y/200Y**

Outline

- ◆ **Brief History of COCOMO**
- ➔ **COCOMO II.1997**
 - **Process**
 - **Prediction Accuracies**
 - **Comparison with COCOMO '81**
- ◆ **Updates and Plans**
 - **Plans for Improving Prediction Accuracies**
 - **COCOMO II Research Aim**
 - **Effects of Process Maturity on Effort**
- ◆ **Acknowledgements and Information Sources**

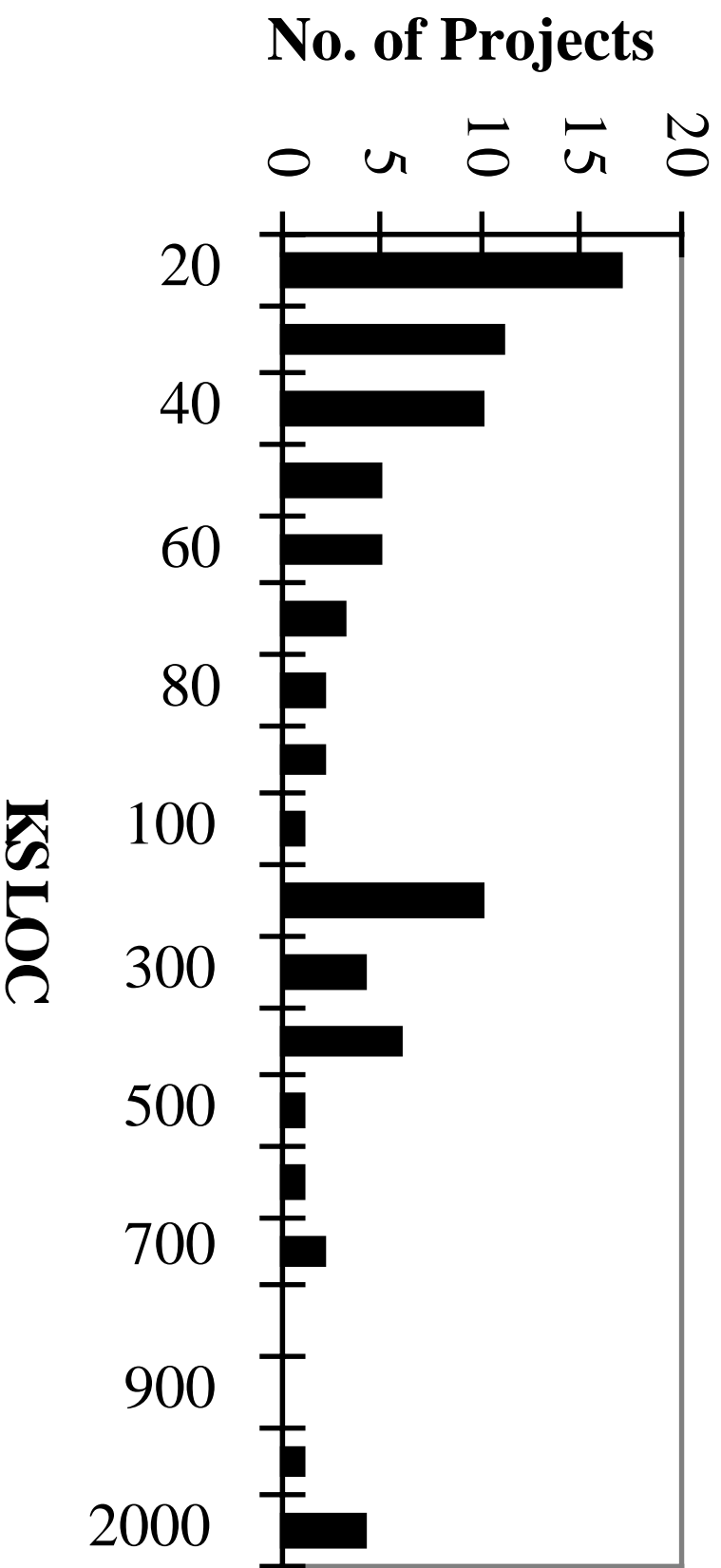
COCOMO II.1997 Calibration

Copyright USC-CSE 1998

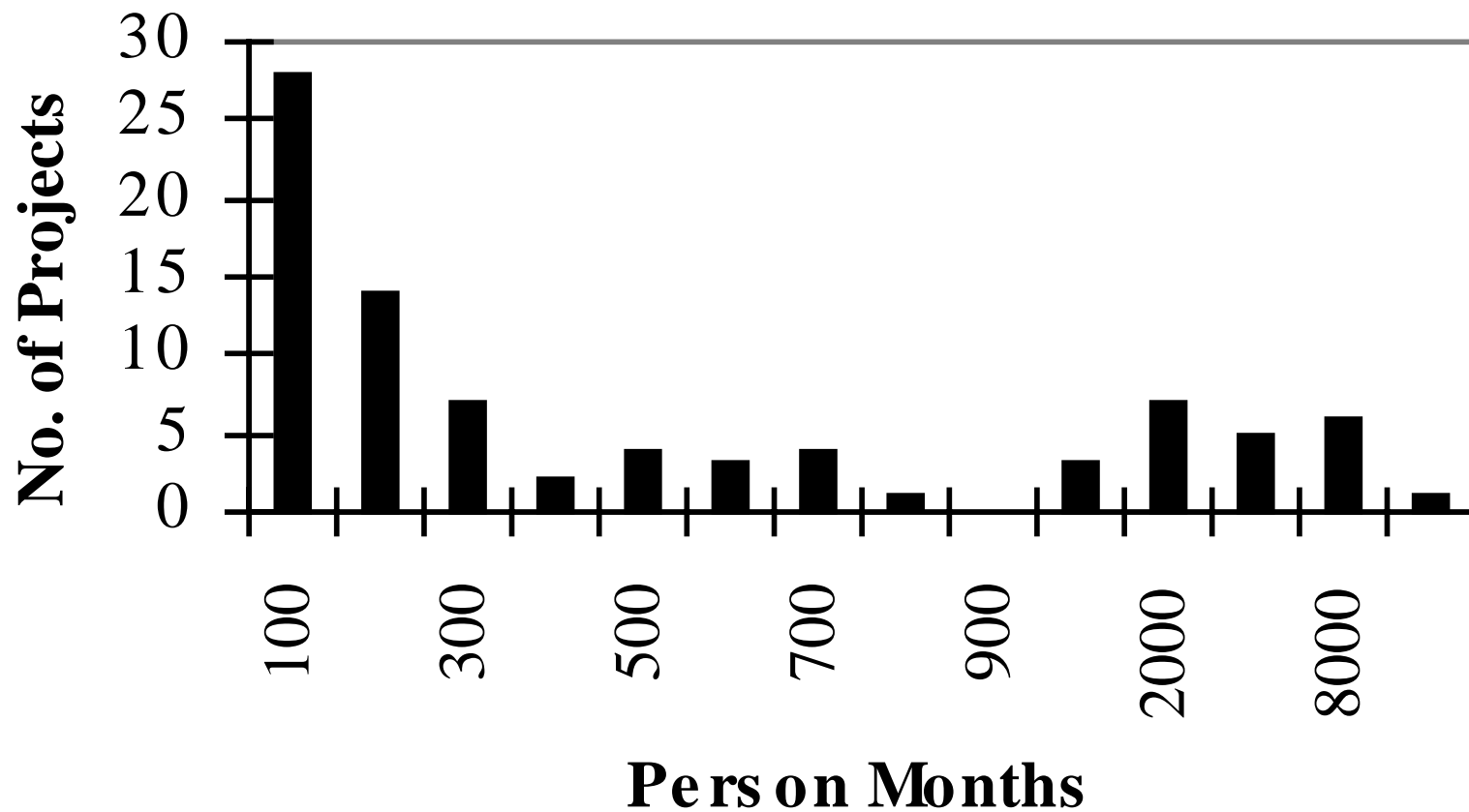
Process

- ◆ **Began with expert-determined a-priori model parameters**
 - Iterated with Affiliates (Result => Original Post Architecture Model)
- ◆ **Collected Data**
- ◆ **Identified and consolidated highly correlated model parameters**
- ◆ **Statistically determined estimates of consolidated model parameters from data**
 - Using logarithms to linearize regression
- ◆ **Used data determined model parameters to adjust a-priori model parameters**
 - Experimented with weighting factors

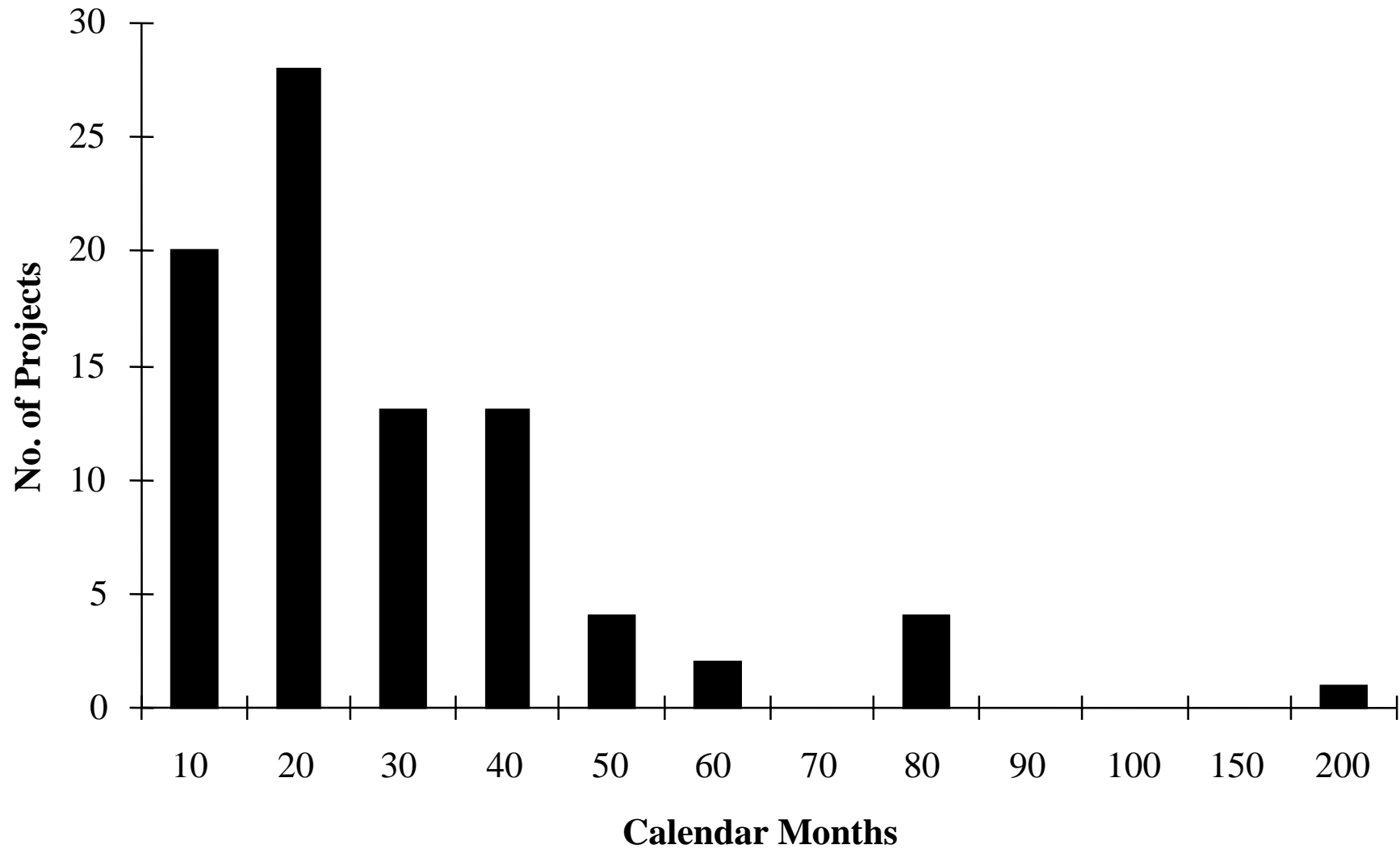
Distribution of Size



Distribution of Effort



Distribution of Schedule



Consolidated Highly Correlated Parameters

TIME	1.0000	0.6860	-0.2855	-0.2015
STOR	0.6860	1.0000	-0.0769	-0.0027
ACAP	-0.2855	-0.0769	1.0000	0.7339
PCAP	-0.2015	-0.0027	0.7339	1.0000
TIME	STOR	ACAP	PCAP	

- What do we do? \Rightarrow Combine :

TIME & STOR to give RCON (Resource Constraints)

ACAP & PCAP to give PERS (Personnel Factors)

Thus, 15 effort multipliers instead of 17 for calibration

COCOMO II Calibration Model

- ◆ Needed linear model for regression:

$$Y = B_0 + B_1X_1 + B_2X_2 + \Lambda + B_pX_p$$

- ◆ COCOMO II Post-Architecture is non-linear

$$Y = B_0X^{B_1}$$

- ◆ What did we do?
 - Expanded COCOMO II model
 - Transformed products with logarithms to produce sums

Expanded COCOMO II

- ◆ Distributed the Scale Factors
- ◆ Resulted in 21 predictor variables i.e. 15 Effort Multipliers + 5 Scale Factors + (Size)^{1.01}

$$PM_{est} = A \cdot (Size)^{1.01} \cdot (Size)^{SF_1} \cdot (Size)^{SF_2} \cdot \Lambda \cdot EM_1 \cdot \Lambda \cdot EM_{15}$$

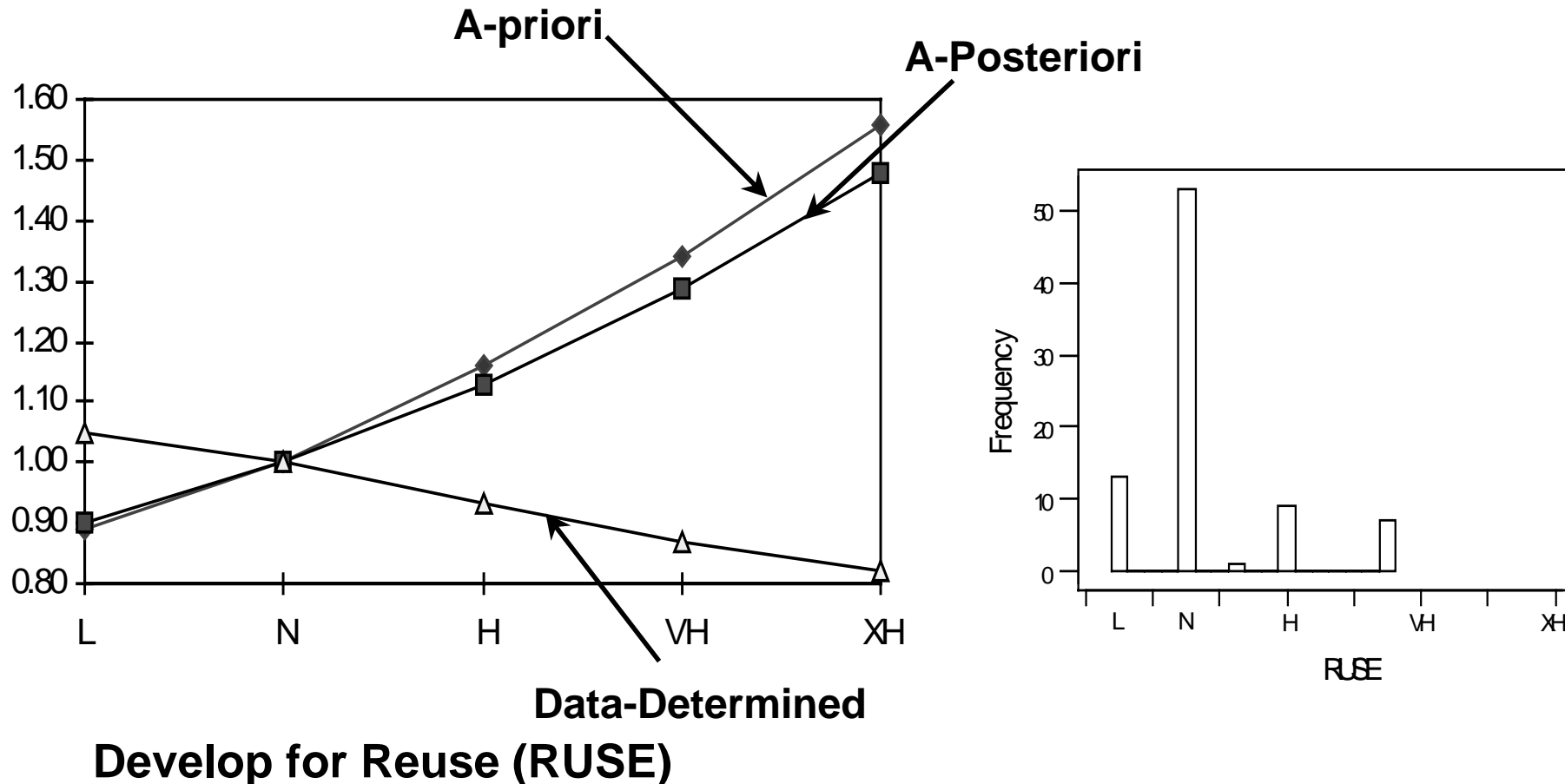
Log Transformed COCOMO:

$$\ln(PM_{est}) - \ln(Size)^{1.01} = \ln(A) + SF_1 \ln(Size) + \Lambda + \ln(EM_{15})$$

- ◆ Regression analysis derived the coefficients, B_i , for each factor

COCOMO II.1997 Calibration

- ◆ 83 projects
- ◆ Multiple Linear Regression
 - 10% weighted average between a-priori values and data-determined values



COCOMO II.1997 Accuracy Results

Effort Prediction	Before Stratification By Organization	After Stratification By Organization
PRED(.20)	46%	49%
PRED(.25)	49%	55%
PRED(.30)	52%	64%

Schedule Prediction	Before Stratification By Organization	After Stratification By Organization
PRED(.20)	48%	52%
PRED(.25)	54%	61%
PRED(.30)	61%	62%

COCOMO '81 & COCOMO II.1997

- ◆ **COCOMO '81: Pred (.20) = 68%**
- ◆ **COCOMO II.1997: Pred (.20) = 46%**
 - **Challenges faced in calibrating COCOMO II**
 - ❖ **GUI builders, COTS, 4GL's, reuse, requirements breakage**
 - **Need to rethink size metrics**
 - ❖ **Distributed interactive applications**
 - **Web-based, object-oriented, event-based**
 - **Middleware effects**
 - ❖ **New process models (evolutionary, incremental, spiral)**
 - **Phases overlap**
 - **Where are cost measurement endpoints?**
 - ❖ **Lack of good data**
 - **not enough data (i.e. very little degrees of freedom)**
 - **lack of dispersion**
 - **heteroskedasticity**

Outline

- ◆ **Brief History of COCOMO**
- ◆ **COCOMO II.1997**
 - **Process**
 - **Prediction Accuracies**
 - **Comparison with COCOMO '81**
- ➔ **Updates and Plans**
 - **Plans for Improving Prediction Accuracies**
 - **COCOMO II Research Aim**
 - **Effects of Process Maturity on Effort**
- ◆ **Acknowledgements and Information Sources**

Plans to Improve Accuracy

- ◆ **Gather more data**
- ◆ **Bayesian Regression Analysis**
- ◆ **Stratify data based on Language Level and Application Type**
- ◆ **Effort distribution based on activities**
- ◆ **Enhancement of COCOMO II database to continuously update the model**

Successive versions of Copyright USC-CSE 1998 COCOMO II

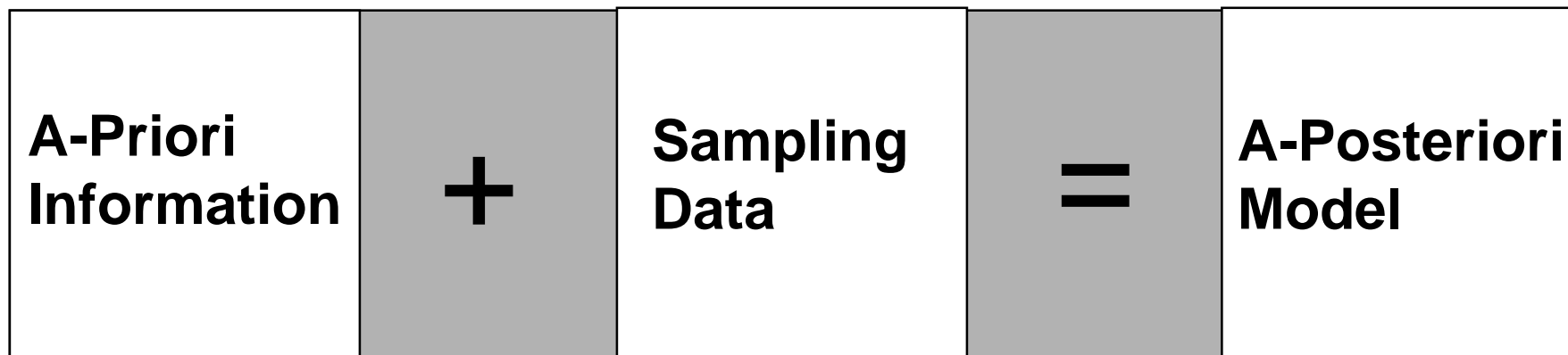
- ◆ **The 1997 version**
 - **Multivariate Linear Regression with 10% weighted average of expert-determined and data-determined**
- ◆ **The 1998 version**
 - **Bayesian Regression Analysis**
 - ❖ **Weighted average**
 - ❖ **Separate weights for each parameter based on significance**
 - **Model more Data-Determined**
- ◆ **The 19??/20?? version**
 - **100% Data-Determined**

Bayes Theorem

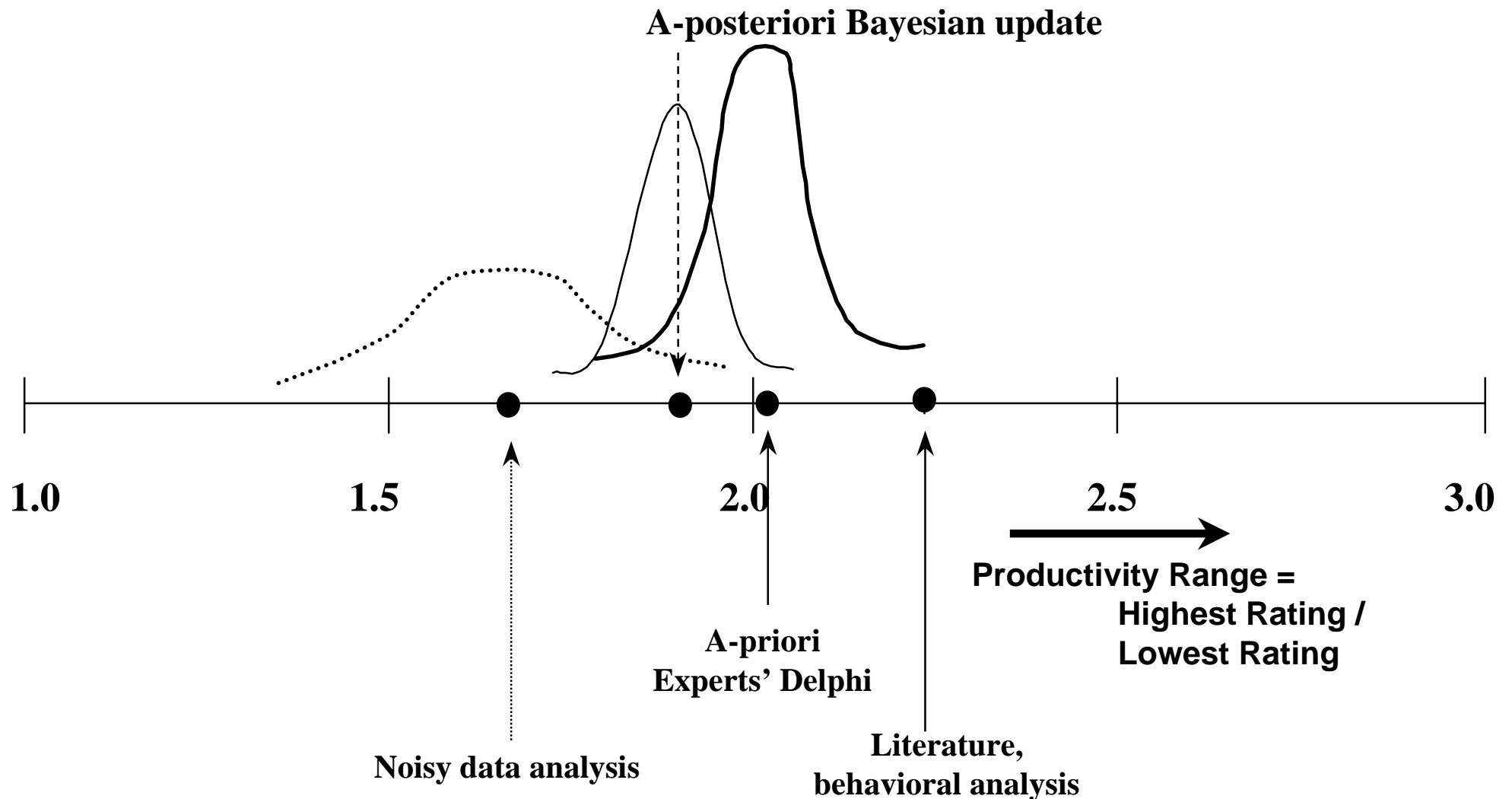
$$g(\beta / y) = \frac{f(y / \beta) g(\beta)}{f(y)}$$

$$g(\beta / y) \propto l(\beta / y) g(\beta)$$

posterior information \propto sample information X prior information



The Bayesian Approach for COCOMO II



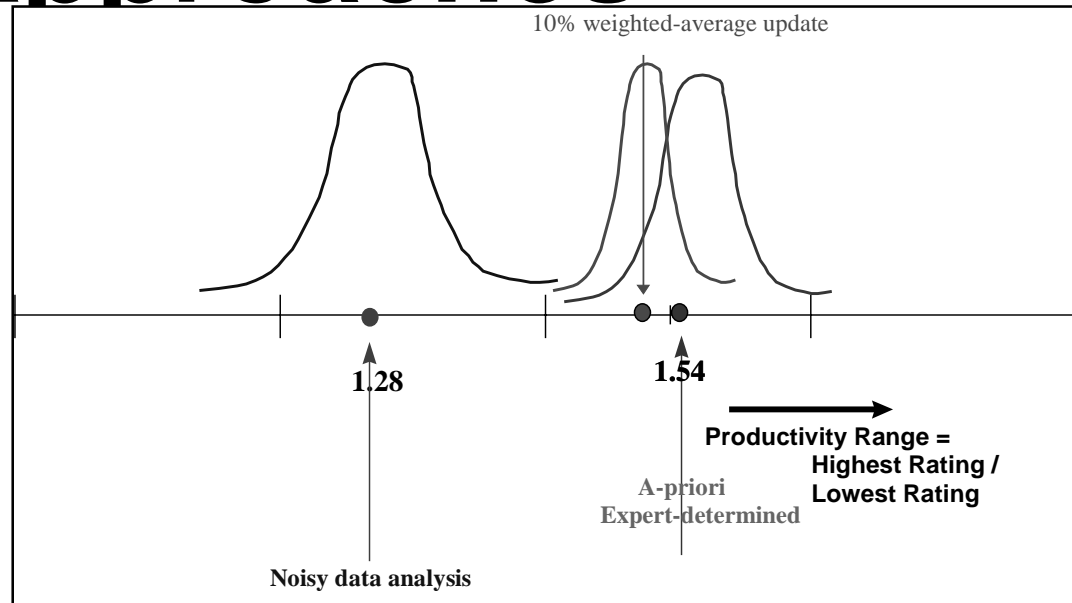
Prediction Accuracies*

Effort Prediction	Bayesian Approach 159 observations	COCOMO II.1997 83 observations
PRED(.20)	49%	46%
PRED(.25)	57%	49%
PRED(.30)	64%	52%

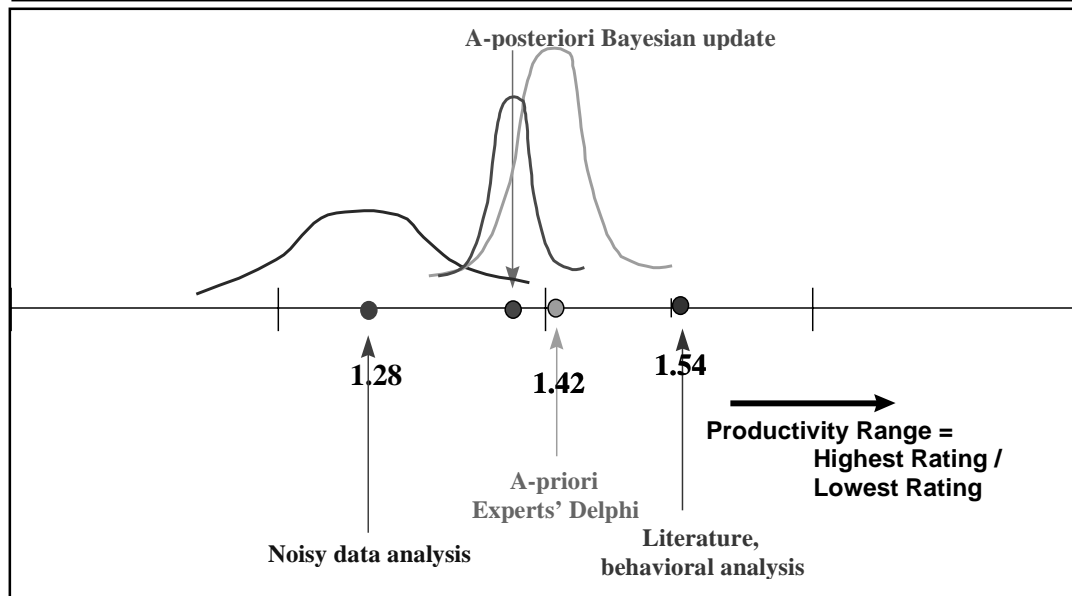
* Please note that the COCOMO II.1998 calibration is not yet complete. Hence these results may be different for the COCOMO II.1998 final calibration

COCOMO II Calibration Approaches

**COCOMO II.1997
10% weighed-
average approach**

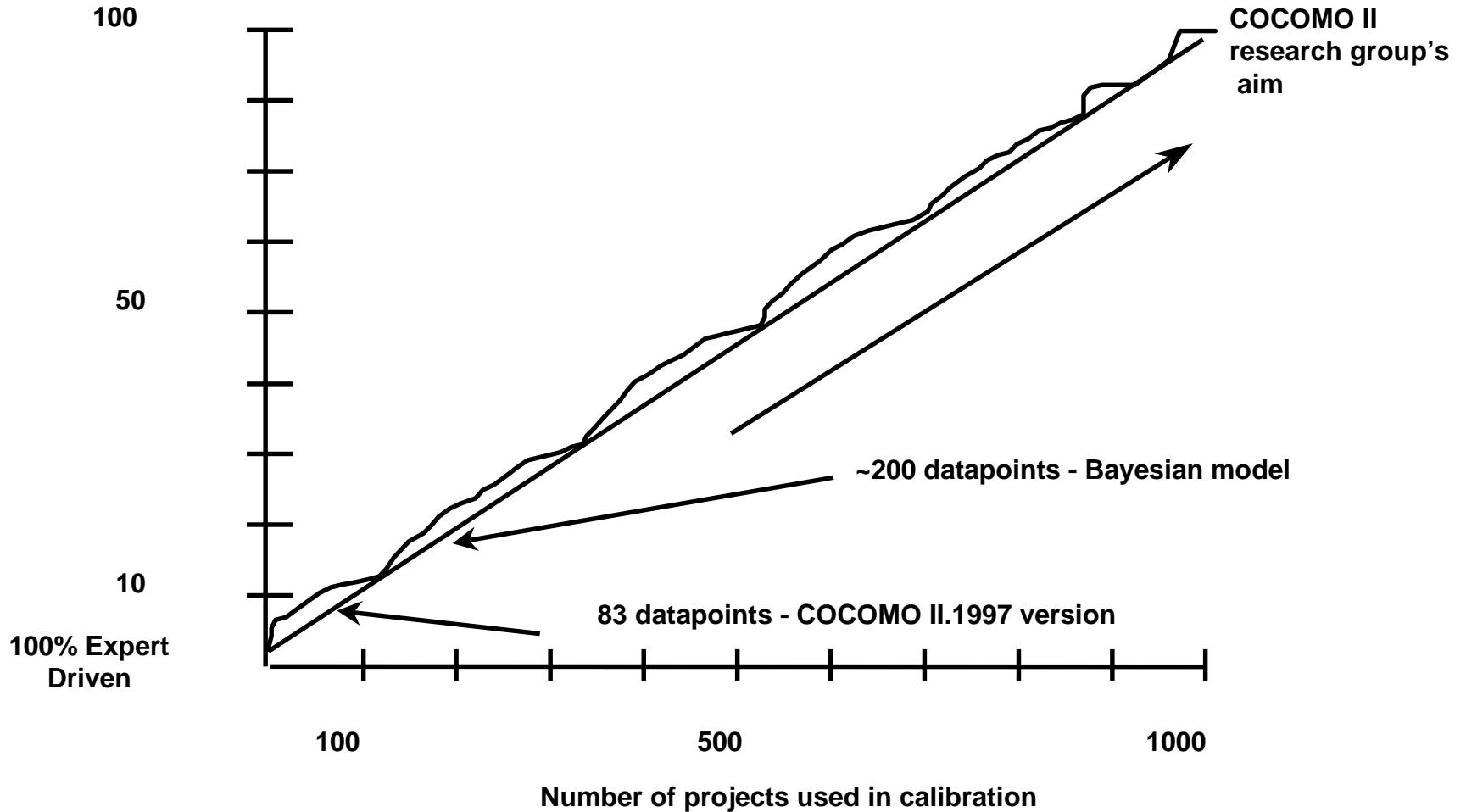


**Bayesian approach -
weight determined by
data and prior
significance**



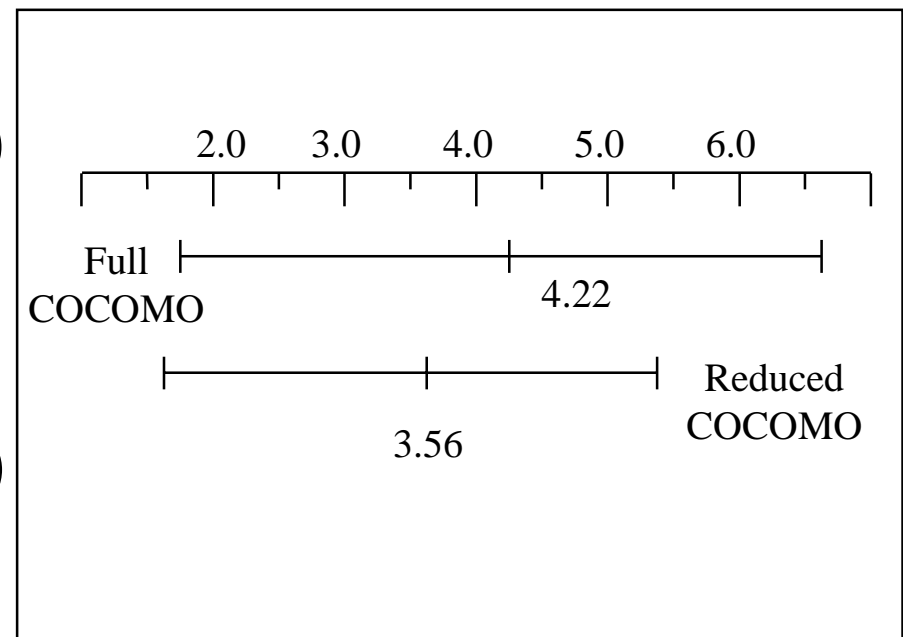
COCOMO II Research Group's Aim

100% Data Driven



COCOMO II PMAT Results

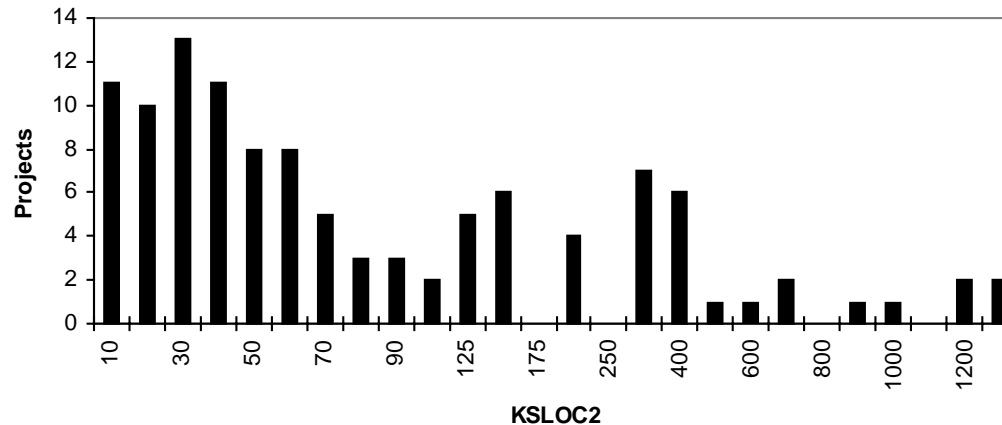
- ◆ Used COCOMO II Provisional Values
- ◆ PMAT is a scale factor in COCOMO II
- ◆ PMAT's influence varies with SIZE:
 $(\text{Size})^{\text{PMAT}}$
- ◆ Full COCOMO
 - b_{PMAT} : 4.22 (t-value = 3.05)
 - interval: $1.9 < b_{\text{PMAT}} < 6.6$
- ◆ Reduced COCOMO
 - b_{PMAT} : 3.64 (t-value =2.19)
 - interval: $1.5 < b_{\text{PMAT}} < 5.4$



Interpreting COCOMO II PMAT Results

Copyright USC-CSE 1998

KSLOC Frequency Distribution



- ◆ For example, let's say the Scale Factors yield an exponent of 1.10, all Effort Multipliers are 1.0. The effect of PMAT on effort is:
 - PMAT unchanged: $(30)^{1.11} / (30)^{1.10} = 1.035$ change in effort
 - PMAT adjusted by 2.0: 1 PMAT increment $\Rightarrow 2(1.035) = 1.07$ change in effort
 - PMAT adjusted by 6.0: 1 PMAT increment $\Rightarrow 6(1.035) = 1.21$ change in effort
 - Decrease effort 7% for 2.0 adjustment and 21% for 6.0 adjustment

Some Observations on Effects of Process Maturity on Effort

- ◆ **Process Maturity does effect effort. A one increment change in PMAT (Level 1 Upper to Level 2, Level 2 to Level 3, etc.) results in a 7% to 21% reduction in effort for a 30 KDSI project.**
 - Effect is larger for larger products**
- ◆ **Using it as a scale factor appears to provide a stronger influence on effort than as a multiplicative factor.**
- ◆ **Its influence is less than the personnel capability of the team, about the same as product complexity (CPLX), and higher than other COCOMO cost drivers.**
- ◆ **Process Maturity should be in all Software Cost Estimation Models; it is well defined and measurable.**

Outline

- ◆ **Brief History of COCOMO**
- ◆ **COCOMO II.1997**
 - **Process**
 - **Prediction Accuracies**
 - **Comparison with COCOMO '81**
- ➔ **Updates and Plans**
 - **Plans for Improving Prediction Accuracies**
 - **COCOMO II Research Aim**
 - **Effects of Process Maturity on Effort**
- ➔ **Acknowledgements and Information Sources**

Acknowledgements and Information Sources

◆ Website - <http://sunset.usc.edu/COCOMOII/Cocomo.html>

- Affiliate Prospectus
- Model Definition Manual
- Data Collection Form
- USC COCOMO Software and User's Manual
- Java COCOMO
- Little Expert COCOMO Calculator

◆ Acknowledgements

–AFRL Contract F30602-96-C-0274,
“KBSA Life Cycle Evaluation”

–COCOMO II Affiliates

Aerospace, Air Force Cost Analysis Agency, Allied Signal, AT&T, Bellcore, EDS, Raytheon E-Systems, GDE Systems, Hughes, IDA, JPL, Litton, Lockheed Martin, Loral, MCC, MDAC, Motorola, Northrop Grumman, Rational, Rockwell, SAIC, SEI, SPC, Sun, TI, TRW, USAF Rome Lab, US Army Research Labs, Xerox.