Calibration Results of COCOMO II.1997

Sunita Devnani-Chulani

USC-CSE

22nd Annual

Software Engineering Workshop

December 3, 1997

Presentation Outline

→ ***COCOMO calibration***

Calibration process

Results to date

- **Plans to Improve Accuracy**
- **Information Sources**

COCOMO II Calibration Process

- Began with expert-determined a-priori model parameters
  - Iterated with Affiliates (Result => A-Priori 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

Consolidated Highly Correlated Parameters

	TIME	STOR	ACAP	PCAP
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? $\Longrightarrow$ Combine :**
  TIME & STOR to give RCON (Resource Constraints)
  ACAP & PCAP to give PERS (Personnel Factors)

Thus, 15 effort multipliers instead of 17 for calibration

Statistical Data Analysis

Variable	Minimum	Maximum	Ratio (Max/Min)
EFFORT	6	11400	1900
SIZE	2.6	1292.8	497

Thus, we took log transforms to normalize the response variable.
Also, we took log transforms to linearize the parametrized model.

Expanded COCOMO

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

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

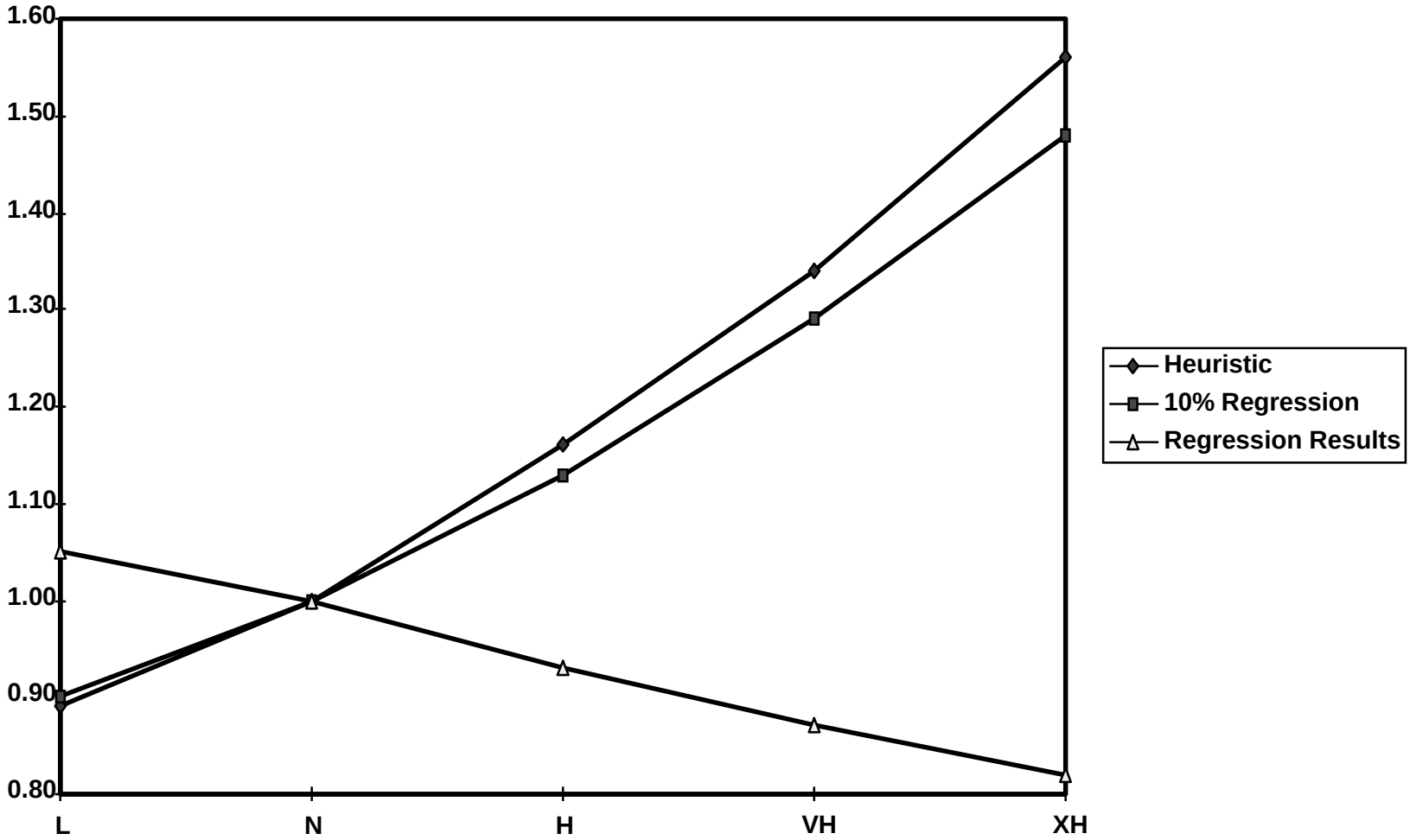
Log Transformed COCOMO:

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

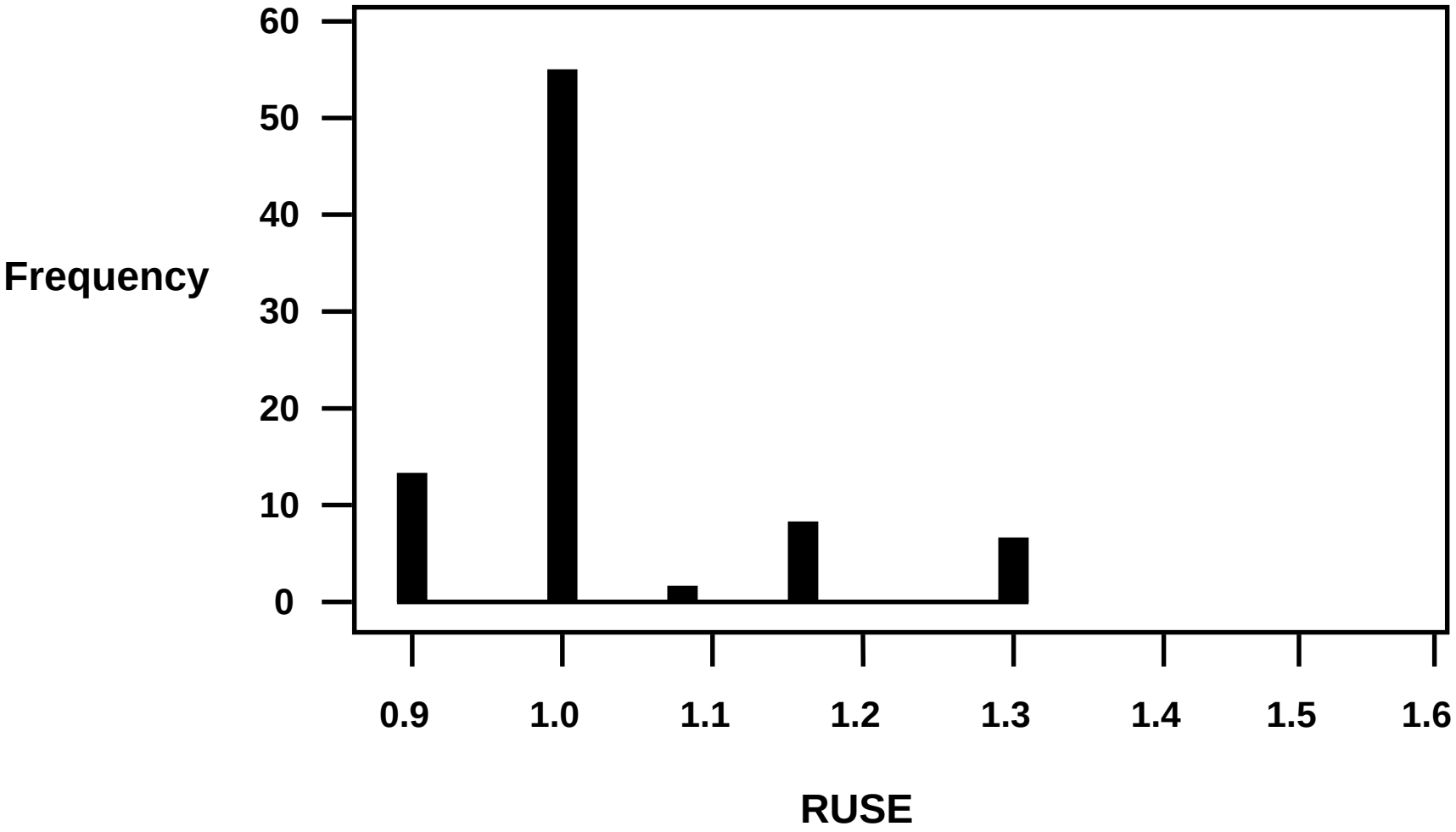
- **Regression analysis derived the coefficients, B_i , for each factor**

RUSE Effort Multiplier

- **Example of the effect of a negative coefficient**



Distribution of RUSE



Overview

- 83 Observations from different Industrial categories including Commercial, Aerospace, FFRDC
- Log transformations of Original Post Architecture Model to achieve linearity for linear regression analysis
- 21 predictor variables i.e. 15 Effort Multipliers +5 Scale Factors + Coefficient A
- Forecast accuracy measured with proportional error:

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

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%

Conclusions: Calibration Results

- Regression technique can be used to calibrate COCOMO locally using completed project data
- New cost drivers can be added and calibrated without destroying the structure of the COCOMO model
- COCOMO calibrated to local organization is more accurate then using generic COCOMO II model
- More project data is required to facilitate better calibration of generic COCOMO II model
- 1990's software data presents more challenges
  - Non-sequential processes: where are end-points?
  - Incremental development: how to separate the increments?
  - COTS, reuse, breakage, mixed language levels: what is size?

Presentation Outline

- COCOMO calibration
  - Calibration process
  - Results to date
- → Plans to Improve Accuracy
- Information Sources

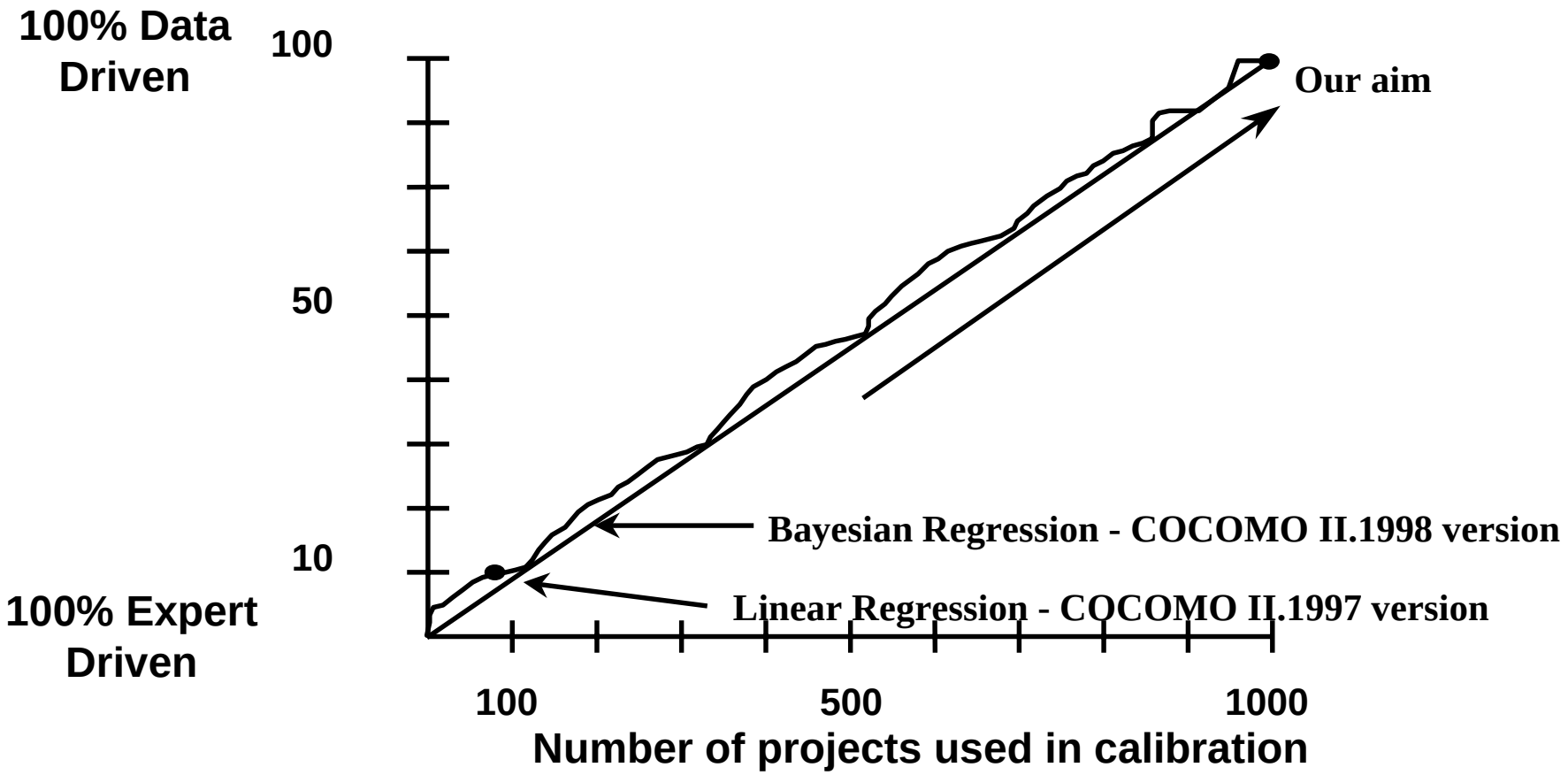
Plans to Improve Accuracy

- 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

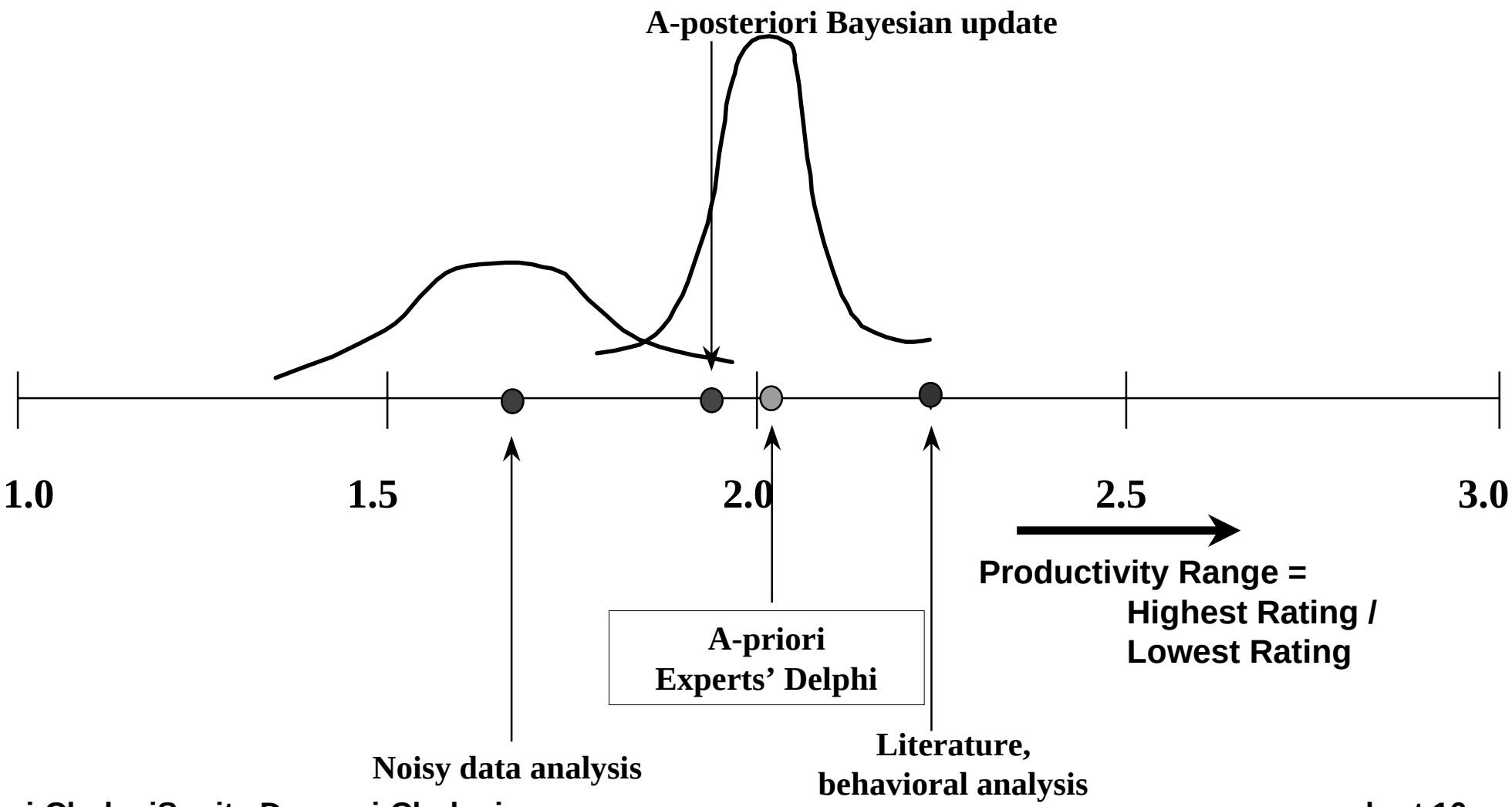
Successive versions of COCOMO II

- 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

Evolving Model Values



Bayesian Approach



Presentation Outline

- **COCOMO calibration**
  - **Calibration process**
  - **Results to date**
- **Plans to Improve Accuracy**
- → • ***Information Sources***

Information Sources

- **Phone: (213) 740-6470**
- **Email: cocomo-info@sunset.usc.edu**
- **Web site: http://sunset.usc.edu/COCOMOII/Cocomo.html**
  - **Affiliate Prospectus**
  - **Model Definition Manual (ver. 1.4)**
  - **Data Collection Form (ver. 1.6)**
  - **USC COCOMO Software and User's Manual**
  - **Java COCOMO**
  - **Little Expert COCOMO Calculator**