A STRATEGY FOR A CREDIBLE & AUDITABLE ESTIMATION PROCESS USING ISBSG REPOSITORY

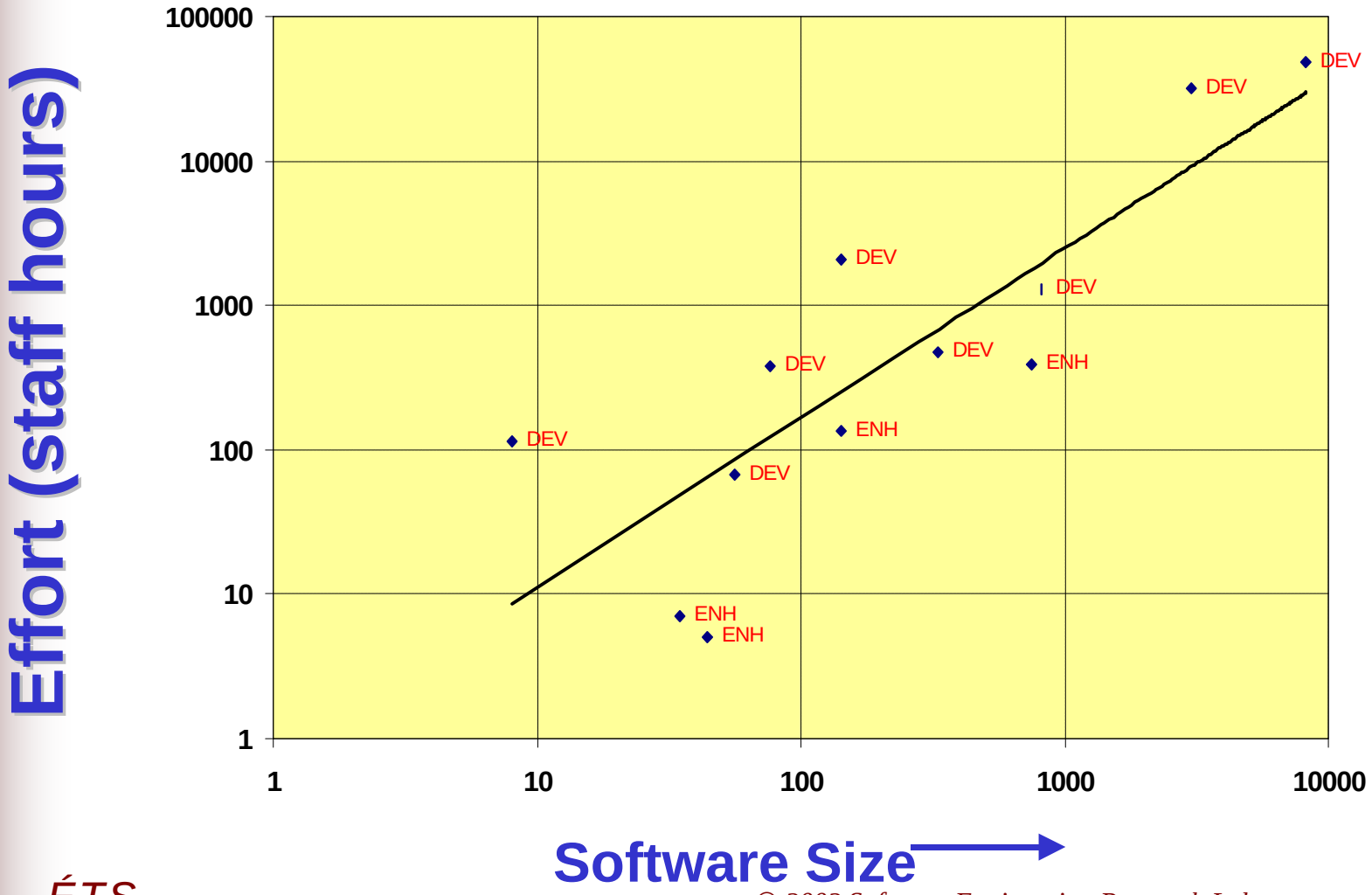
Alain Abran, Reiner Dumke,
JM Desharnais, I. Ndiaye, C. Kolbe

International Workshop on Software Measurements
October 7-9, 2002, Magdeburg (Germany)

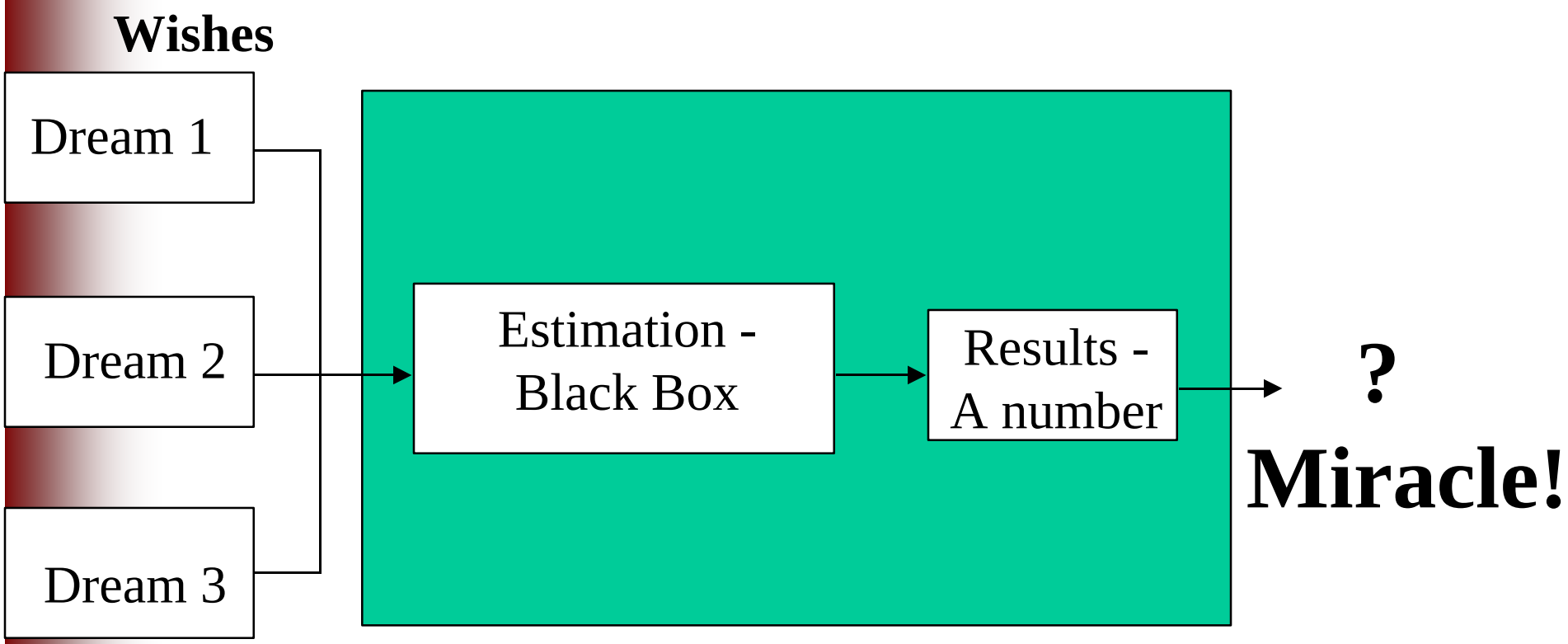
Presentation Plan

- Estimation Process
- Estimation using ISBSG
- Business Context

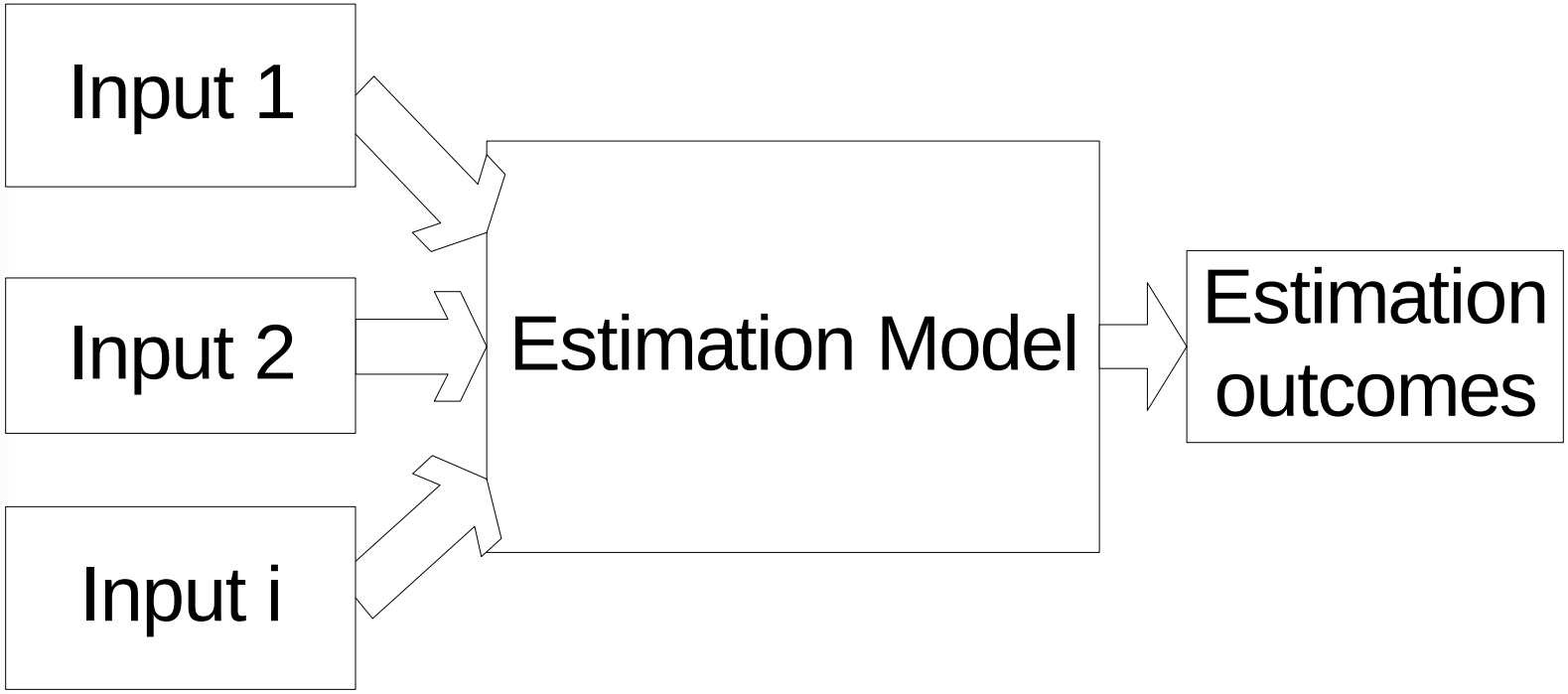
Estimation: IF project size is this BIG, then how much will it cost?



Most common used estimation method: Expert Judgment (…or Black Magic ?)



High-level model of an estimation process



Presentation Plan

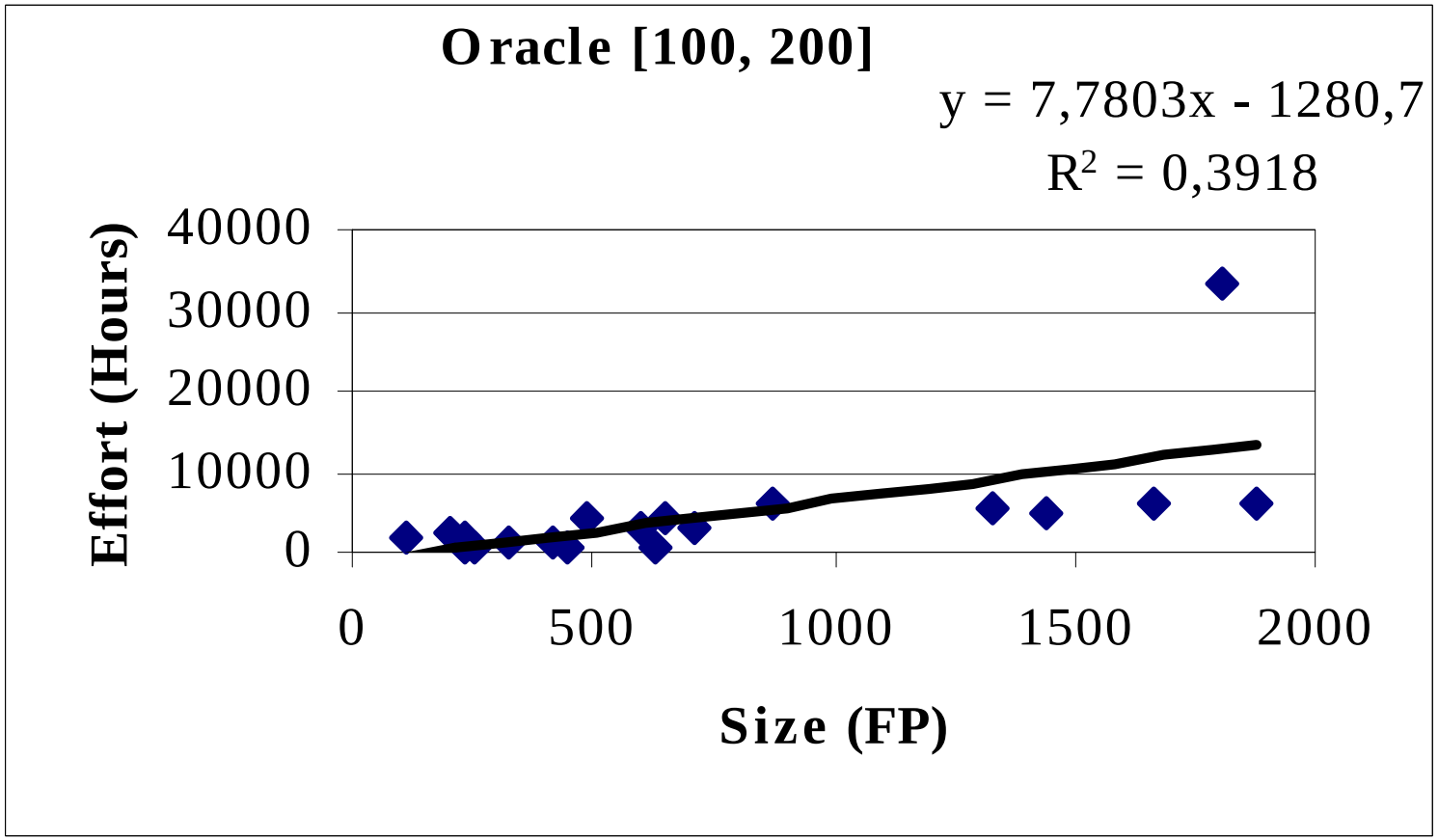
- Estimation Process
- Estimation using ISBSG
- Business Context

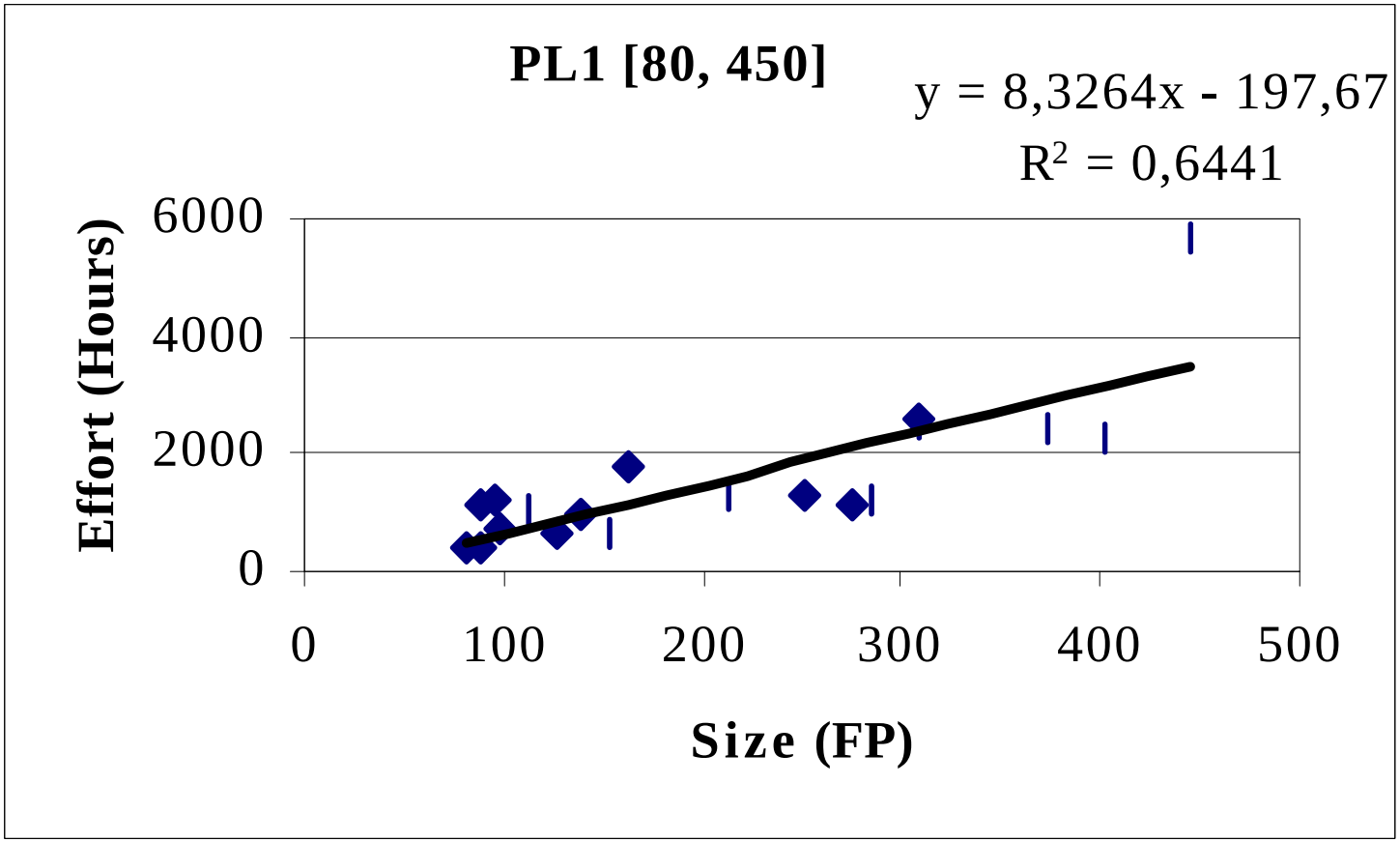
Avalaible Repositories

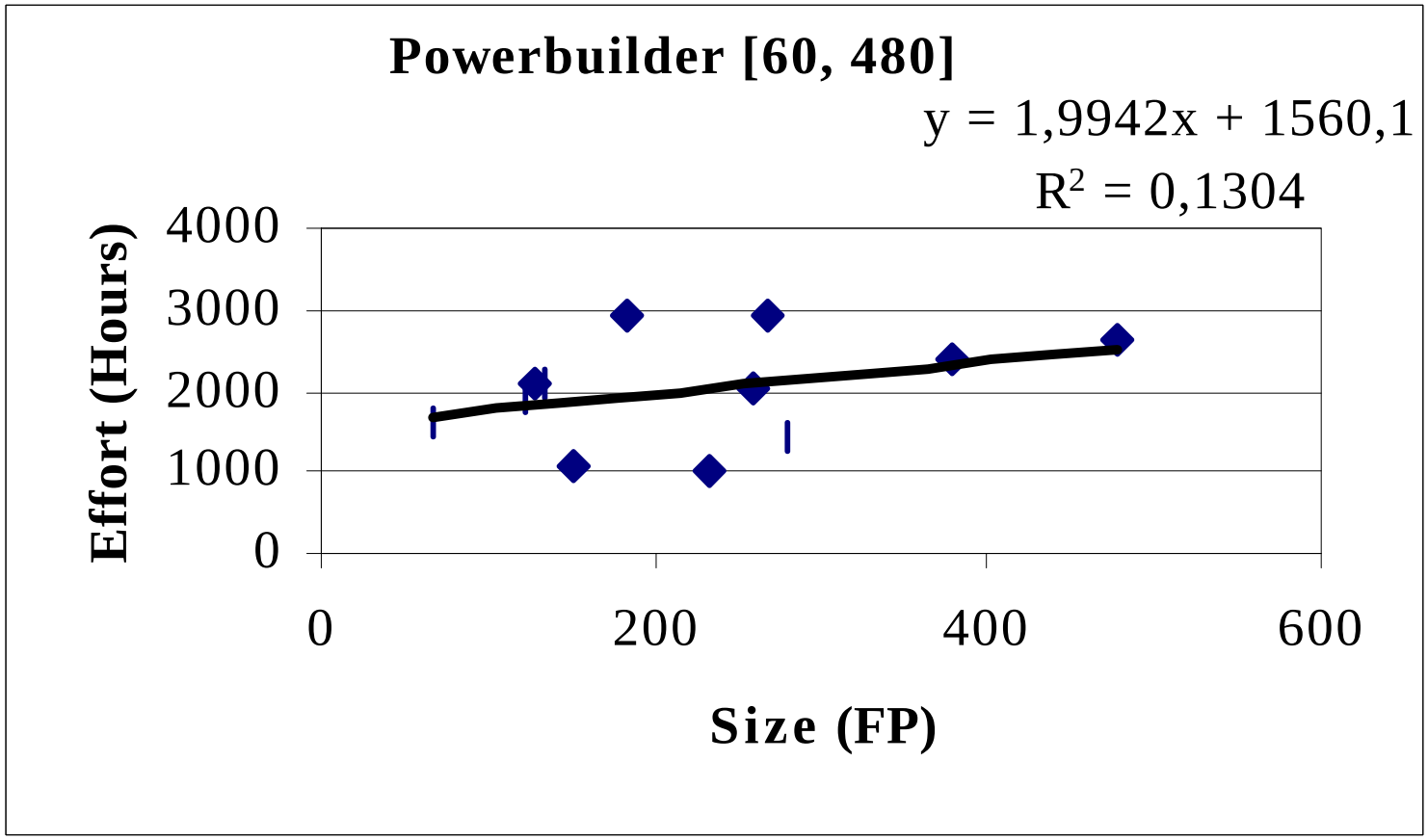
- Alternatives:
  - In-house historical data sets
- - Estimation tools = black boxes (in general)
- - Publicly available repositories = ISBSG (International Software Benchmarking Standards Group)

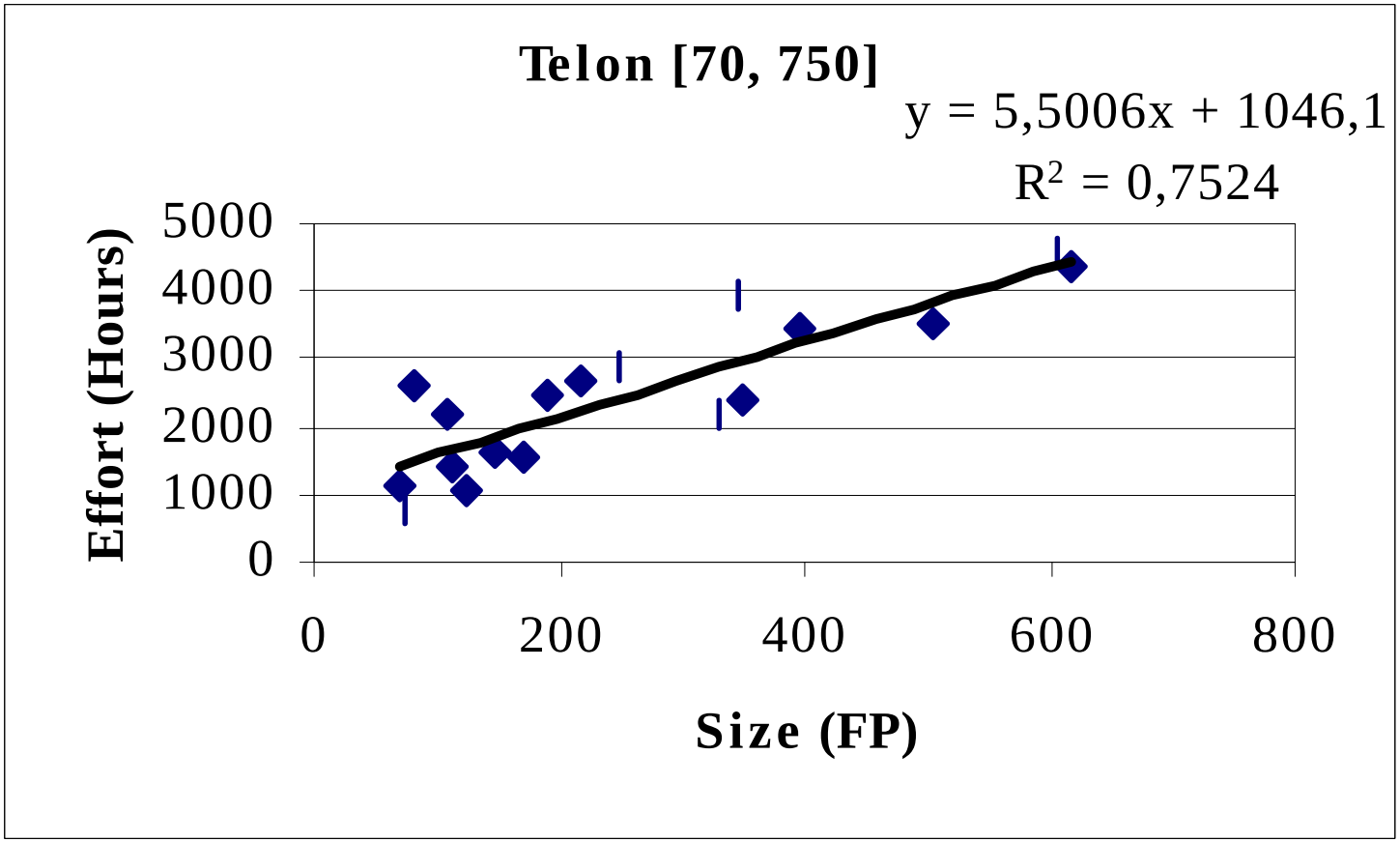
ISBSG 1999 Data set : 749 projects

- Selection criteria:
  - - with effort
  - - with functional size
  - - with programming languages
  - - with ISBSG quality criteria
- Data set used:
  - with required info = 479 projects
  - Sub-samples by programming languages = 371









Set of 12 samples - with all data points		Samples without outliers, and with sub-size intervals			
Programming language *	**N**	**Functional size interval**	**N**	**Functional size interval**	**No. outliers excluded**
Access	17	200-1500	11	200-800	6
C	15	40-2500	9	200-800	6
C++	21	70-1500	12 5	70-500 750-1250	4
Cobol	106	0-5000	60 32	60-400 401-3500	14
Cobol II	21	80-2000	9 6	80-180 181-500	6
Natural	41	20-3500	30 9	20-620 621-3500	2
Oracle	26	110-4300	19	100-2000	7
PL/1	29	80-2600	19 5	80-450 451-2550	5
Powerbuilder	18	60-900	12	60-400	6
SQL	20	280-4400	11 8	280-800 801-4500	1
Telon	23	70-1100	18	70-650	5
Visual Basic	34	30-2300	24	30-600	10
Total	**371**		**299**		**72**

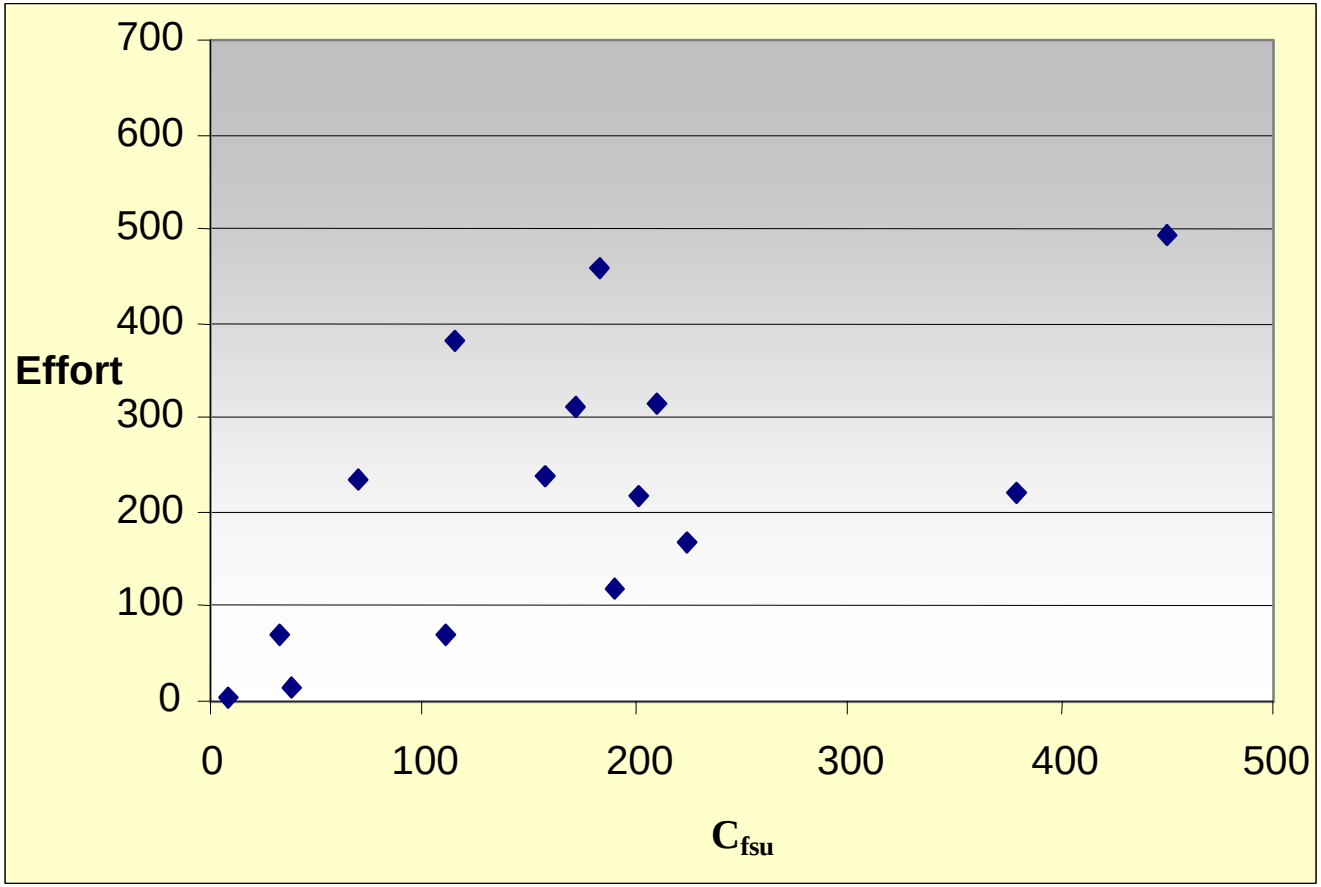
* Programming language as recorded in ISBSG repository

	Samples with outliers				Samples without outliers, and within size intervals			
Language	**N= 377**	**Size Interval**	**Linear Regression Equation (where x = FP units)**	**R^2**	**N= 302**	**Functional Size Interval**	**Linear Regression Equation (where x = FP units)**	**R^2**
Access	17	200-1500	Y=-0.10x+840.8	0.01	11	200-800	Y= 0.30x + 623.5	0.19
C	15	40-2500	Y= 4.05x + 4288	0.19	9	200-800	Y= 2.34x + 2951.7	0.29
C++	21	70-1500	Y=13.43x + 1346.4	0.62	12 5	70-500 750-1250	Y= 11.53x + 1197.1 Y= -6.57x + 23003	0.11 0.06
Cobol	106	0-5000	Y=4.94x + 5269.3	0.36	60 32	60-400 401-3500	Y= 10.83x + 299.1 Y= 12.32x – 14.1	0.44 0.64
Cobol II	21	80-2000	Y= 27.80x – 3593	0.96	9 6	80-180 180-500	Y= 16.39x – 92.3 Y= 26.73x – 3340.8	0.45 0.61
Natural	41	20-3500	Y=10.05x – 648.9	0.85	30 9	20-620 620-3500	Y= 6.13x + 264.9 Y= 10.53x – 1404.9	0.47 0.74
Oracle	26	110-4300	Y=6.20x + 509.7	0.42	19	100-2000	Y= 7.78x – 1280.7	0.39

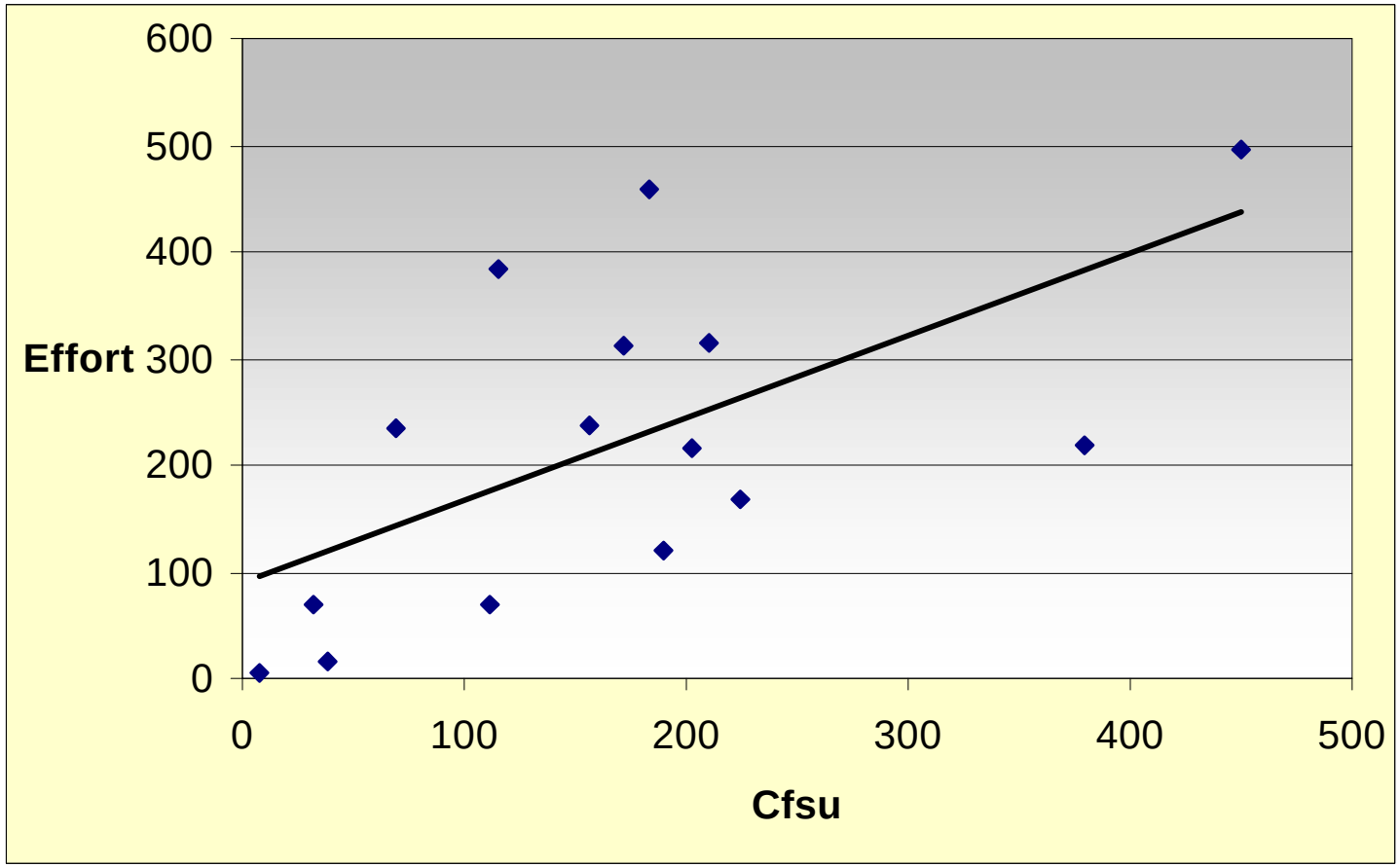
Table 4: Performance of ISBSG regression models on samples without outliers

Programming languages and size range	RRMS(%)	PRED(0.25)
Access [200,800]	15	91
C [200, 800]	50	22
C++ [70, 500]	86	25
C++ [750, 1250]	24	60
Cobol [60, 400]	42	43
Cobol [401, 3500]	51	35
Cobol [80, 180]	29	78
Cobol II [181, 500]	46	33
Natural [20, 620]	50	27
Natural [621, 3500]	35	33
Oracle [100, 2000]	120	21
PL1 [80, 450]	45	42
PL1 [451, 2550]	21	60
Powerbuilder [60, 480]	29	58
SQL [280, 800]	81	27
SQL [801, 4500]	45	25
Telon [70, 650]	22	56

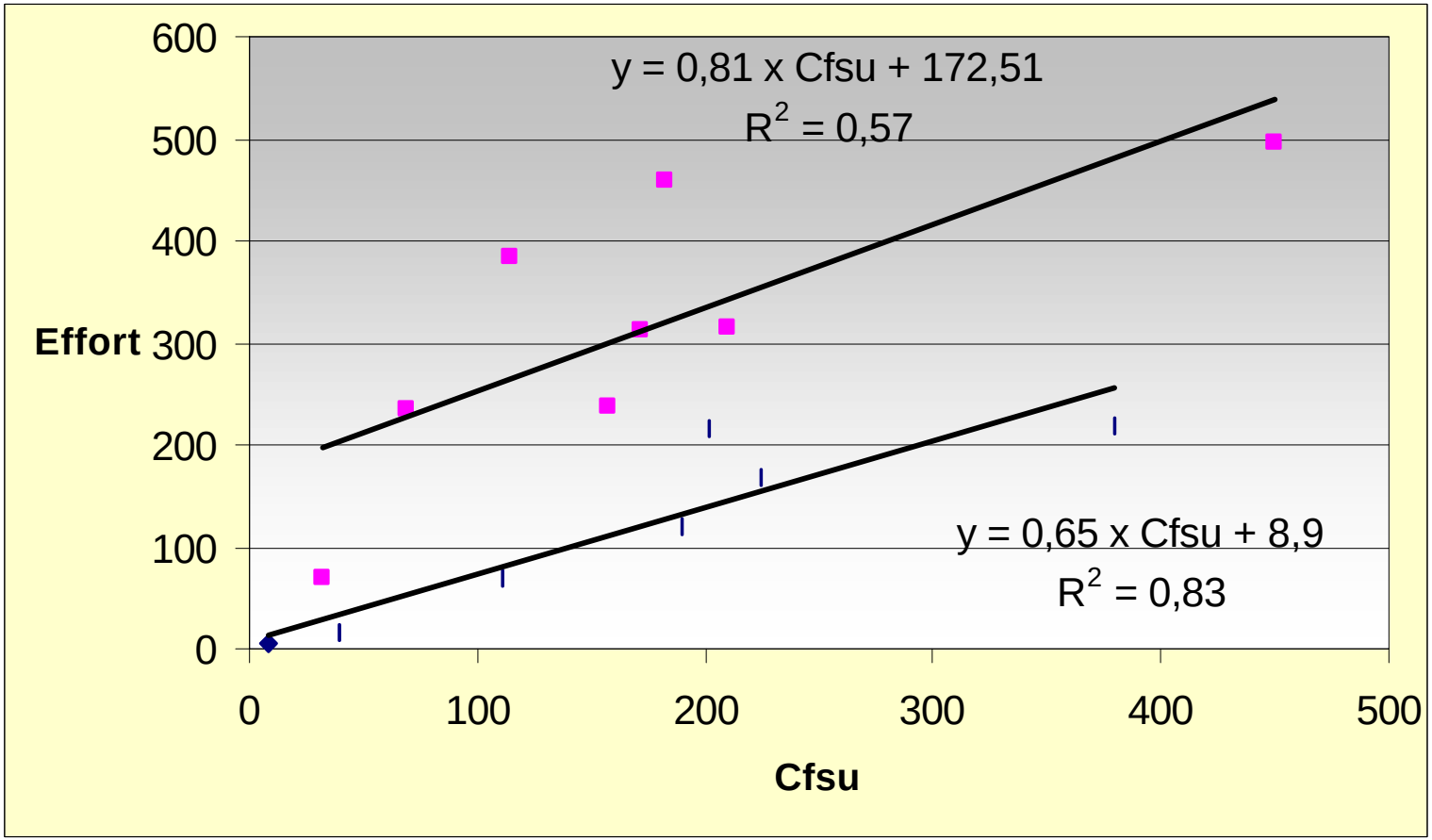
Web-based COSMIC Data in next ISBSG Release



Web-based COSMIC Data (next ISBSG Release) Regression Model – 1 Variable



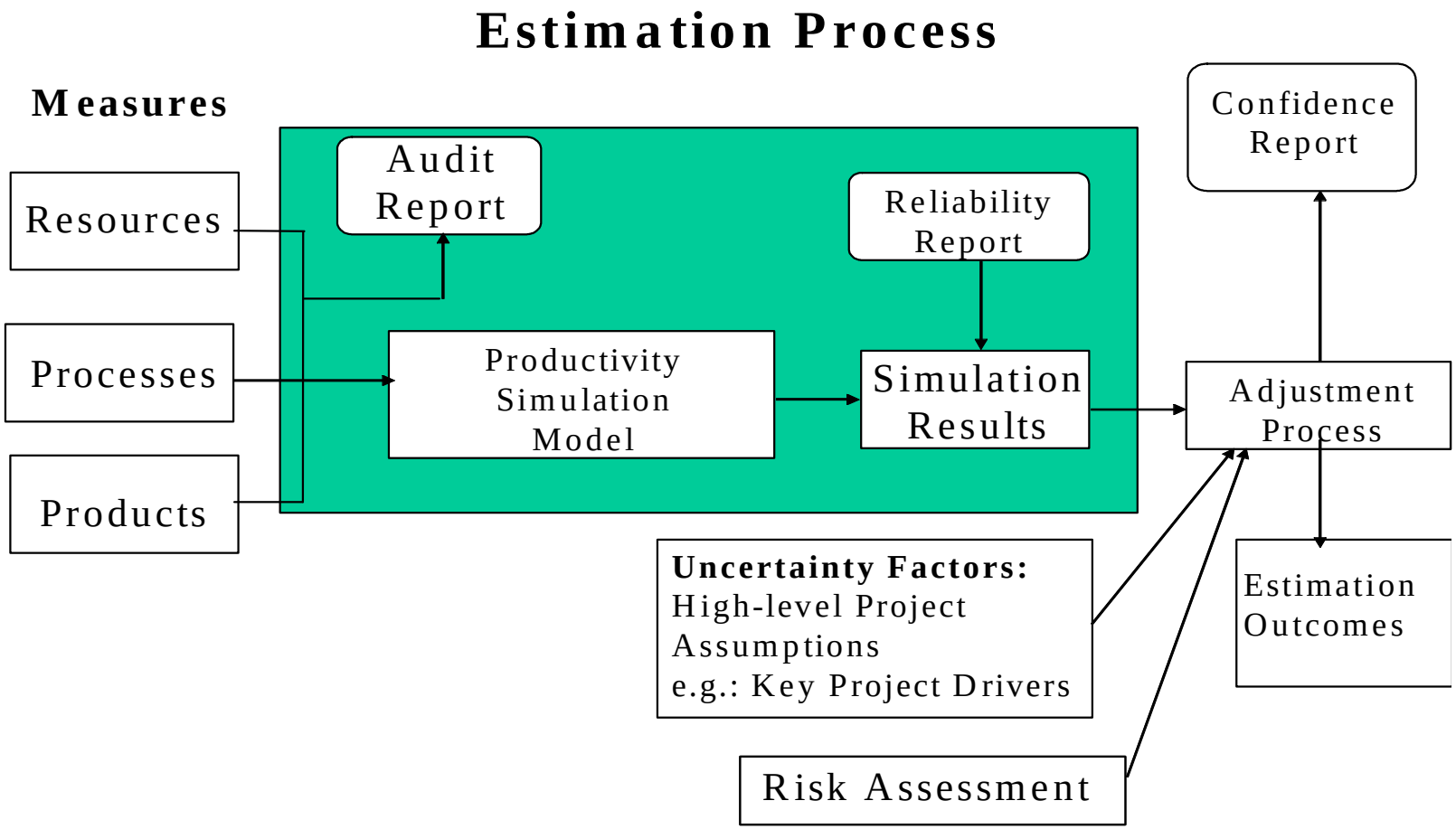
Web-based COSMIC Data *(next ISBSG Release)*
Regression Model – 2 Variables



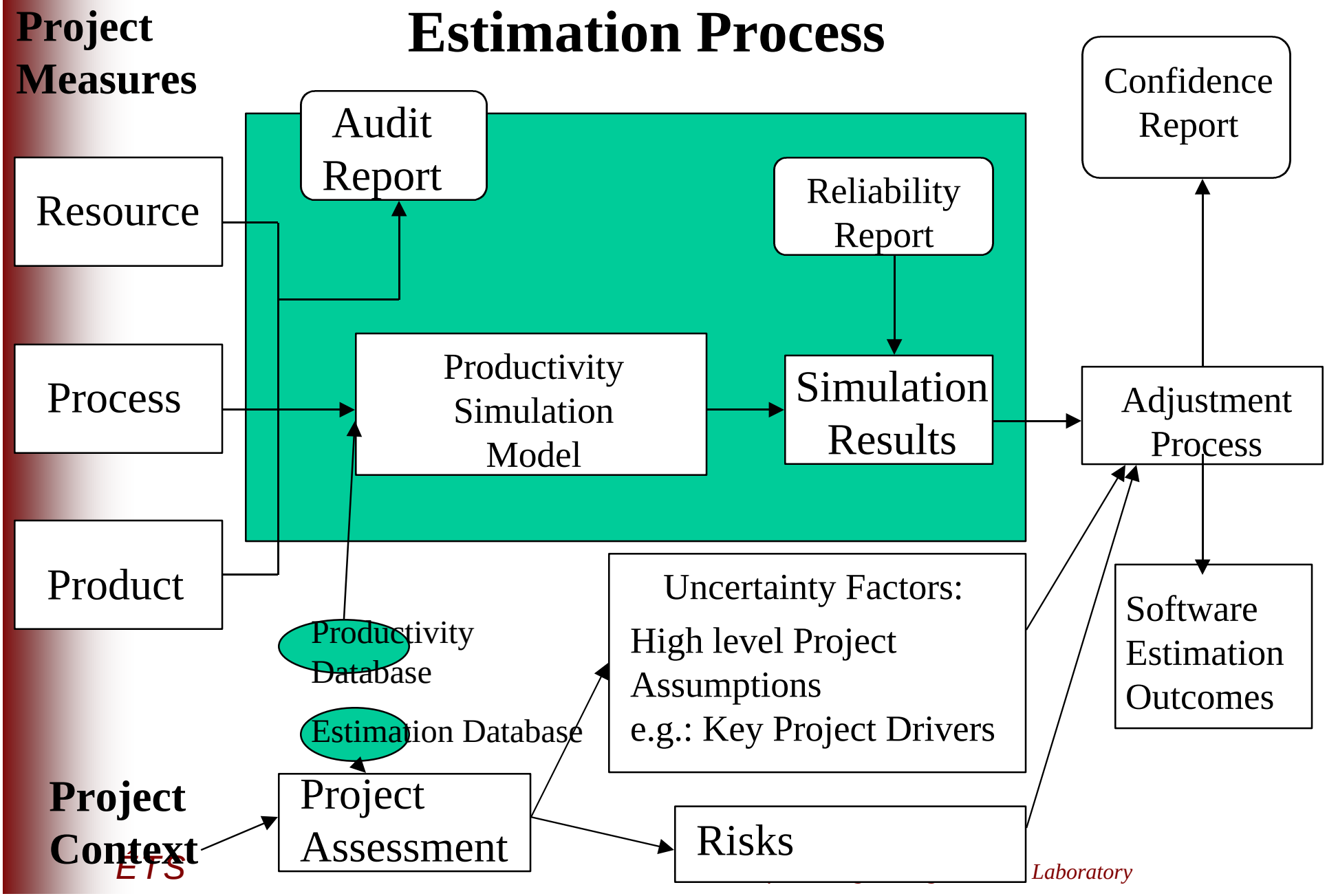
Presentation Plan

- Estimation Process
- Estimation using ISBSG
- **Business Context**

Key components of an estimation process



Estimation Process



Business Context

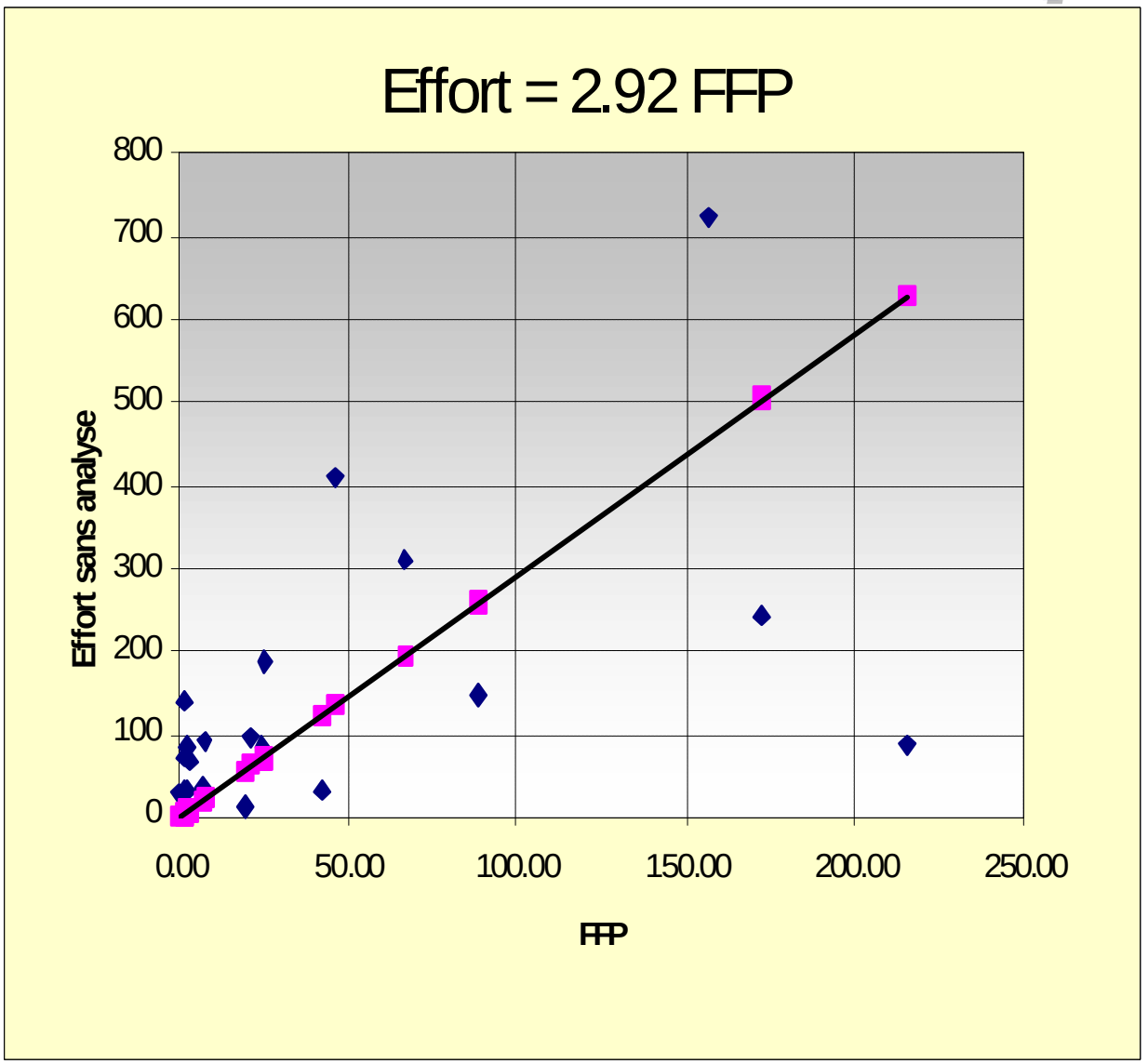
- The technical estimate based on:
  - From the estimation process with its statistical model data repository and (+ audit reports)
- The final estimate = a **business** estimate to
  - Make money
  - Minimize the risks
  - Win the business

Distinct Roles

- Technical responsibilities
  - related techniques & auditable requirements
- Business responsibilities
  - Related risks taken & documented assumptions on control variables & mitigation of risks and uncertainty
- Credibility requires both auditability and traceability

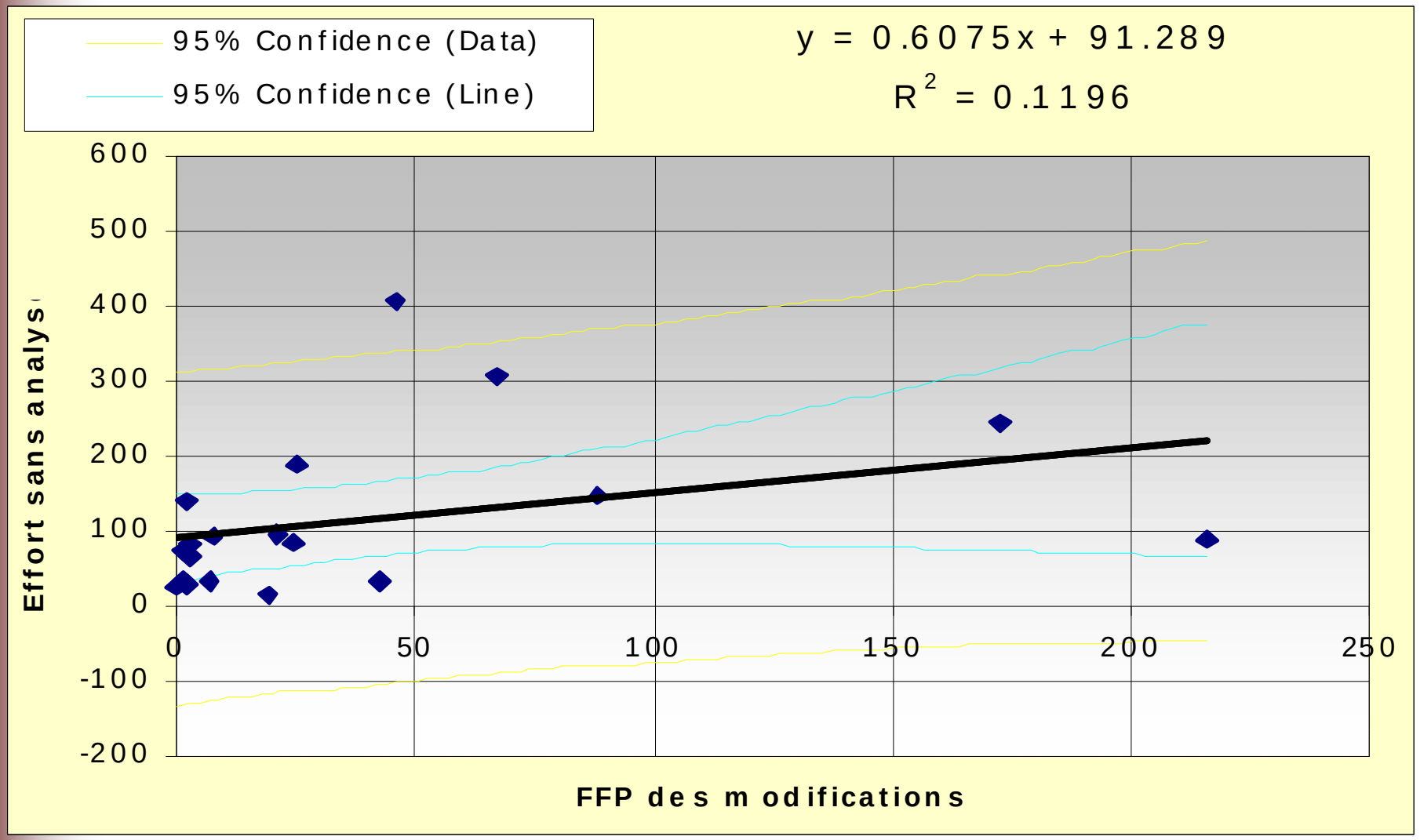
Thank you

Set C: Maintenance Requests



Silva & Abran, 2000

Set C: Maintenance Requests



Silva & Abran, 2000

Set C: Maintenance Requests

		N	A	B	R	RSquare
Y=A+B*X	Linear	19	91.289	0.607	0.35	0.12
Y=A*X^B	Power	19	43.808	0.245	0.50	0.245
Y=A*e^(B*X)	Exponential	19	63.067	0.006	0.39	0.15
Y=A+B*ln(X)	Logarithmic	19	44.121	29.29	0.51	0.26
Y=A+B/X	Hyperbolic 1	19	132.463	-48.330	0.32	0.10
Y=1/(A+B*X)	Hyperbolic 2	19	0.022	-8.8E-05	0.31	0.09
Y=X/(A+B*X)	Hyperbolic 3	19	0.007	0.016	0.28	0.08

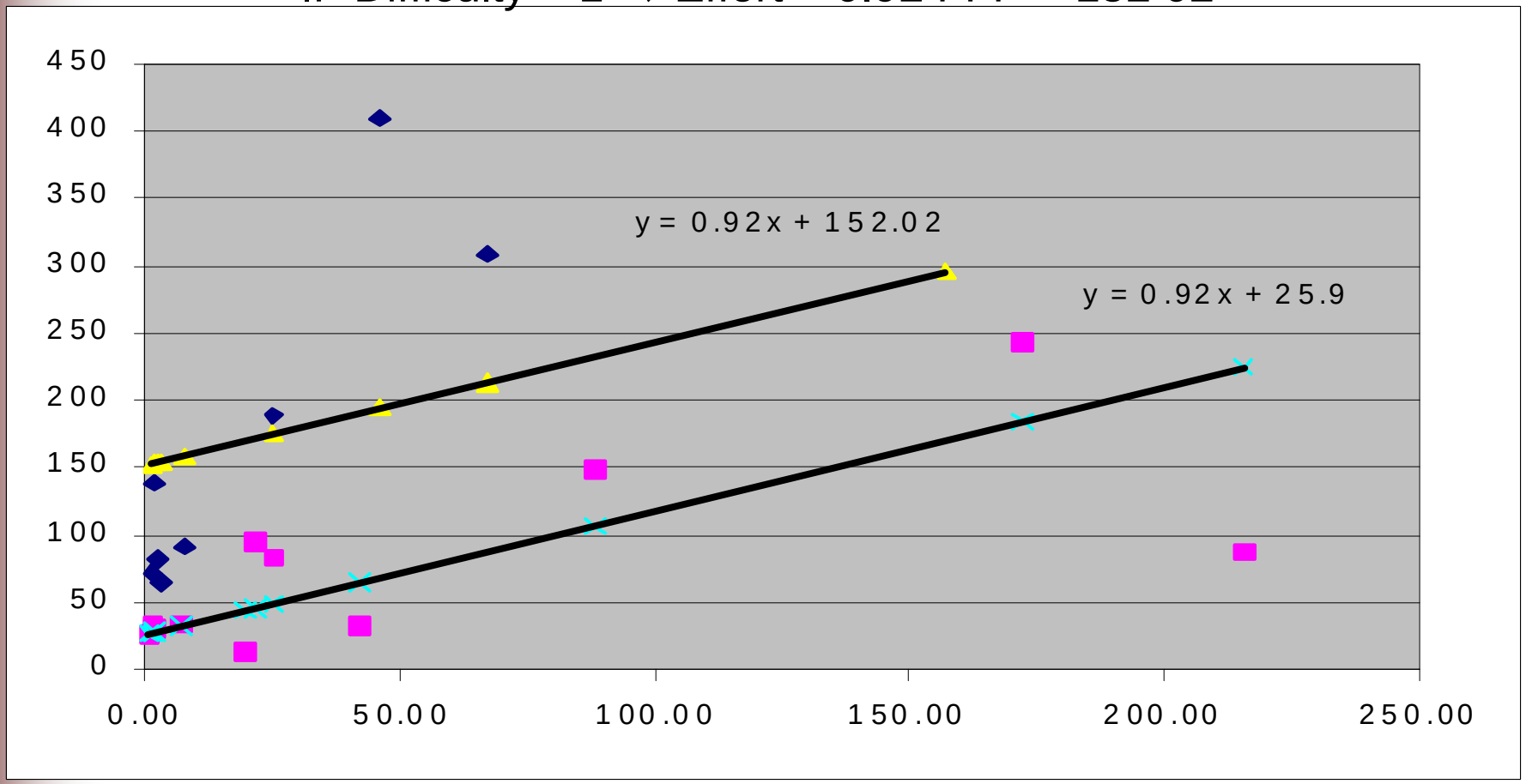
Silva & Abran, 2000

Set C: Maintenance Requests

Effort = 0.92 FFP + 126.12 Difficulty +25.9 and R^2 : 0.46 and n = 19

If Difficuty = 0 → Effort = 0.92 FFP + 25.9

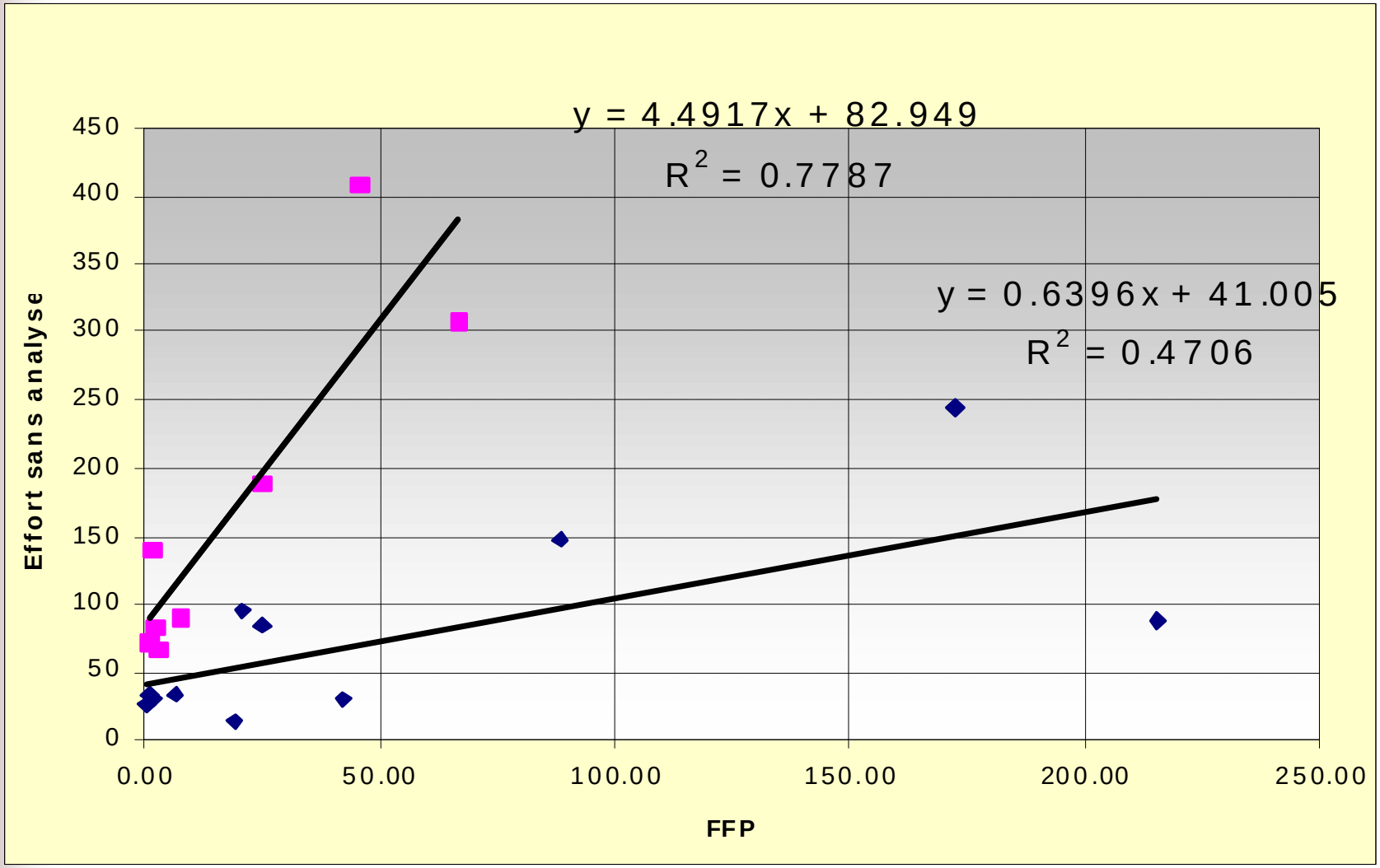
IF Difficulty = 1 → Effort = 0.92 FFP + 152 02



Silva & Abran, 2000

Set C: Maintenance Requests

Multiplicative model with size and complexity



Silva & Abran, 2000