Lessons Learnt from 30 years of Cost Estimation

Professor Barbara Kitchenham

Agenda

- Cost estimation over the past three decades
  - 70s
  - Early 80's
  - Late 80's
  - 90's
- Current Research
- Future trends

Industry Practice – 70s

- Top down
  - Expert opinion
    - Informal analogy
- Bottom up
  - Look-up tables for components
    - Average effort
      - For simple, average, complex modules
      - For new & enhanced modules
  - Expert opinion for individual tasks

70s – Problem Recognition

- Data collection started in major companies particularly for DoD projects
  - Wolverton (1974)
    - Cost per object instruction curves for
      - New v. Old
      - Function type
      - Difficulty
  - IBM Regression studies (Walston & Felix, 1977)
  - SLIM (Putnam, 1978)
    - $Size = C \times Effort^{1/3} Duration^{4/3}$
    - Relates elapsed time to effort during development
  - Function point concepts (Albrecht, 1979)

Early 80's – Dissemination

- Software Engineering Economics (Barry Boehm, 1981)
  - Constructive Cost Model (COCOMO)
    - General purpose model
      - Expert-opinion
      - Data used to validate model
  - COCOMO Effort=a×Size^b
    - Multiplicative parameter "a"
      - Effort per unit size
    - Exponential parameter "b"
      - >1 Diseconomy of scale
      - Large project are more difficult
    - 15 Cost drivers affect effort
      - Complexity, Timing constraints, Team Experience etc.

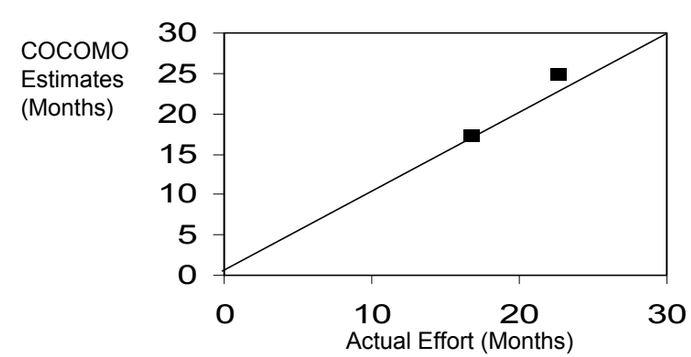
Early 80s - Dissemination

- Controlling Software Projects (DeMarco, 1982)
  - Building Company Specific Models
    - Use statistical regression analysis
  - Early life-cycle metrics from
    - Data Flow Diagrams (DFDs)
    - Entity Relationship (E-R) Diagrams
  - Different models/metrics at different stages in the lifecycle

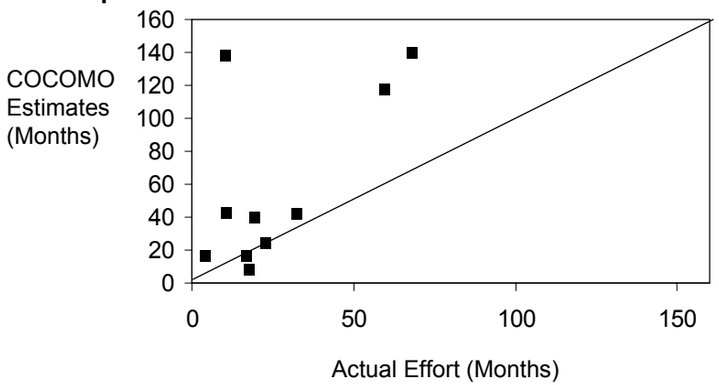
Late 80's - Evaluation

- Evaluation studies
- Accuracy of models on independent data sets
  - Commercial applications (Kemerer, 1987)
  - Real time and Operating Systems (Kitchenham & Taylor, 1984)

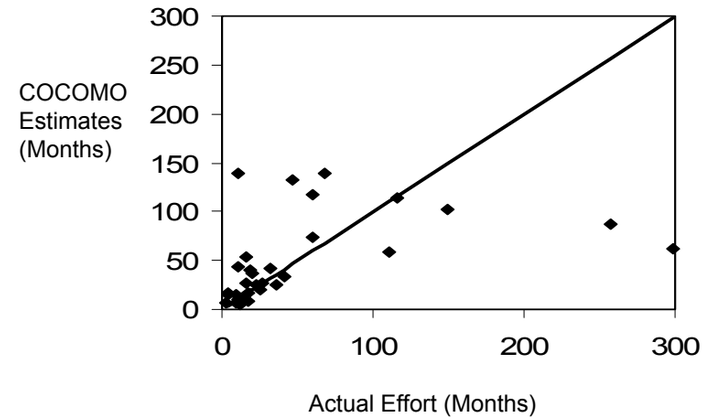
First 2 Project Results



Ten Project Results



33 Projects (ICL & BT)



Other results

- Kitchenham & Taylor
  - Couldn't confirm SLIM effort-duration relationship
    - Neither could Walkerden & Jeffery (1997)
- Kemerer
  - SLIM & COCOMO deliver poor estimates
  - COCOMO improves if re-calibrated

Lessons Learnt

- Don't get too excited about initial results
  - Wait for a long term trend
- General purpose cost models
  - May not fit other people's circumstances
  - Organization-specific or calibrated models are often more accurate
- Cost estimation models may be particularly poor for large projects
  - Work worse just when you need them most!

Industry Practice – 80s

- Top down
  - Expert opinion
    - Informal analogy
- Bottom up
  - Look-up tables for components
    - Average effort
      - For simple, average, complex modules
      - For new & enhanced modules
  - Expert opinion for individual tasks

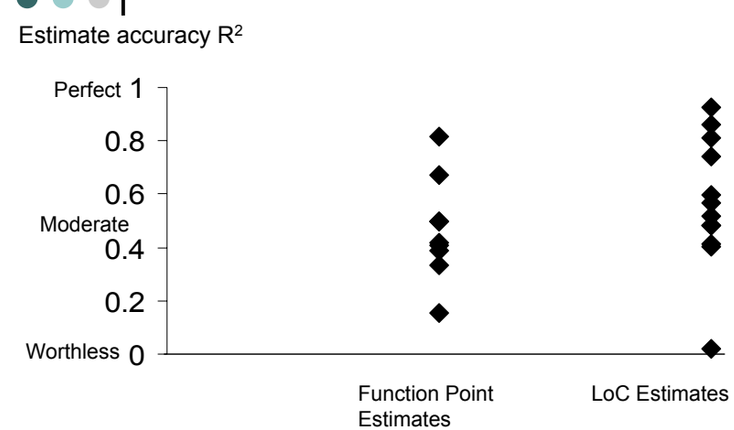
The 90's – Refinement

- Model development
  - More emphasis on model calibration & company specific model
    - Regression models
- Function point size model
  - Size evaluation from early lifecycle documents
- Economic arguments about functional form of cost model
  - Economies or diseconomies of scale
- Other data intensive methods
  - Analogy (Euclidean distance)
  - Neural nets, genetic algorithms etc.

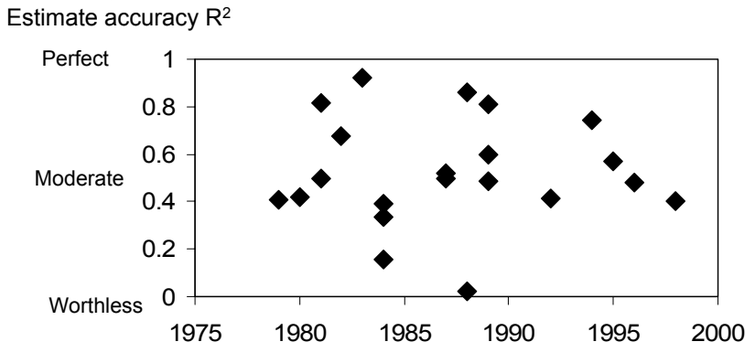
Model evaluation

- Collections of data sets
  - Bankar & Kemerer (1989, 1994)
  - Kitchenham (1992)
- Looking for commonalities
  - Economies or diseconomies of scale
  - Whether FPs give better predictions

Are FPs better than LoC?



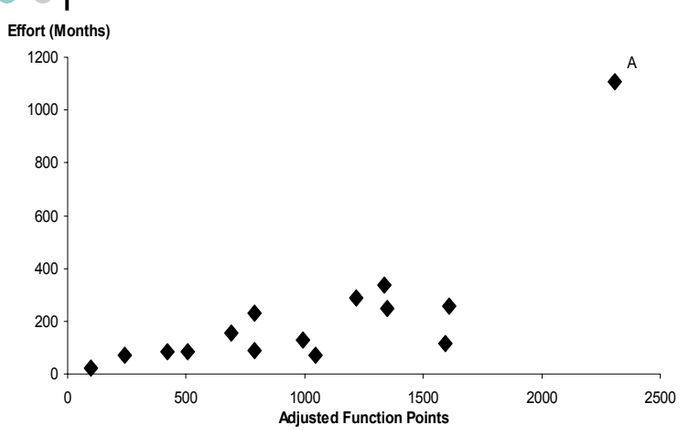
Are models improving?



Diseconomies of Scale

- Kitchenham (No) v. Bankar & Kemerer (Yes)
- Reasons (Kitchenham, 2002)
  - Different hypothesis testing procedures
    - Alpha levels (0.05 or 0.01)
    - One-sided v. Two sided
  - Different selection of data sets
  - Different treatment of data sets
    - Minimum homogenous subsets v. full data set
  - Treatment of outliers

Treatment of Outliers



Lessons Learnt

- Function point models are not much better than LoC
  - Although measurable early in the life cycle
- The models are not always very good
  - Account for about 50% of effort variation
  - Not good enough for industry requirements
- Models are not improving
  - New model development methods not much improvement
    - Genetic algorithms not much improvement on linear regression (Dolado, 2001)
- Don't let statistical algorithms decide things without checking what they are doing

Late 90s - Consensus

- Academics, Consultants, Tool Vendors & Practitioner agree
- Good estimates require
  - Stable development processes
  - Data from similar projects
  - Used to
    - Construct new models
    - Calibrate exiting models
- But practitioners cant use the models

Practical Problem

- Limitation of studies
  - Ignoring industry realities
    - Changing processes
    - Small numbers of comparable projects
  - Post-hoc analysis of static data sets
    - Accuracy insufficient for industry
- Implications
  - Need methods that cope with
    - Estimate uncertainty
    - Small data sets

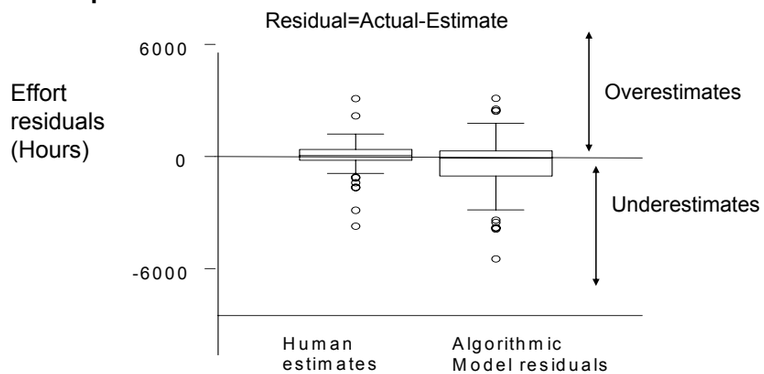
Industry Practice – 90s

- Top down
  - Expert opinion
    - Analogy
      - Backed by past project data
- Bottom up
  - Look-up tables for components
    - Average effort
      - For simple, average, complex objects
      - For new & enhanced objects
  - Expert opinion for individual tasks

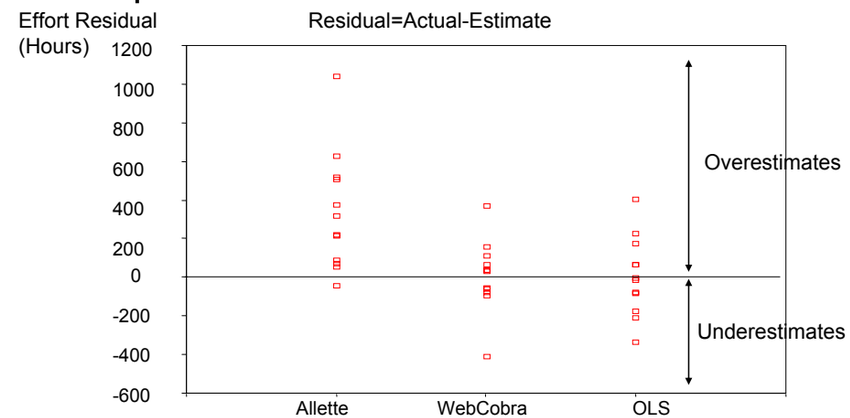
Recent Research

- Large Cross-Company Databases (ISBSG, ESA, Finnish, Tukutuku)
  - Cross-company models (Briand et al., 1999; Maxwell and Forselius, 2000; Jeffery et al., 2001, Kitchenham and Mendes, 2004)
    - Address the problem of small data sets
    - Less able to address
      - Accuracy
      - Rapid process changes
- Recognition of human expertise
  - Humans using estimation models better than either individually (Stensrud & Myrtveit, 1998)
  - Humans can outperform FP models (Kitchenham et al., 2002))
  - Composite models combining expert knowledge and data outperform regression models and humans (Briand et al., 1998, Ruhe et al., 2003)

Human v. Model estimates



Company specific cost model



Cost Models – Still Not Used

- Even good technical results may not be adopted by industry (Keung et al., 2004)
  - WebCobra Model gives excellent accuracy but not used by company
    - Use cost drivers
  - Possible reasons
    - Tool is difficult to use
    - Only a few projects are suitable
      - 124 projects (1998-2001)
      - Only 12 projects suitable for tool
    - Couldn't measure web object metric needed as input

Future?

- More emphasis on industry requirements
  - What academic results are really useful
    - Do results match industry needs
  - How useful results can be incorporated into industry practice
    - Academics shouldn't expect industry to change their methods to fit academic models
    - Need consider how estimation supports industry processes
      - Bidding
      - Budgeting
      - Planning
      - Portfolio management
    - Need to manage estimate uncertainty
      - Risk Management

Was I wasting my time?

- Winston Churchill
  - "Success is going from failure to failure without losing enthusiasm."

References

- Albrecht, A.J. Measuring Application Development Productivity, SHARE-GUIDE, 1979, pp 83-92.
- Banker, R. D. and C.F. Kemerer. (1989) Scale economies in New Software Development, *IEEE Transactions on Software Engineering*, SE-15(10), pp. 1199-1205.
- Banker, R. D., H. Chang and C. F. Kemerer. (1994) Evidence on economies of scale in software development, *Information and Software Technology* 36(5), pp. 275-282.
- Boehm, B.W. Software Engineering Economics. Prentice-Hall, 1981.
- Briand, L.C. El-Emam, K. and Bomarius, F. COBRA: A hybrid method fpr Software Cost estimation, Benchmarking & Risk Assessment, Proc 20th International Conf. On SE, 1998, pp390-399.
- L.C. Briand, K. El-Emam, K. Maxwell, D. Surmann, I. Wieczorek. An assessment and comparison of common cost estimation models. Proceedings of the 21st International Conference on Software Engineering, ICSE 99, 1999, pp 313-322
- DeMarco, T. Controlling Software Projects, Prentice-Hall, 1982

References

- R.Jeffery, M. Ruhe and I. Wieczorek. Using Public domain metrics to estimate software development, Proc, 7th International Symposium on Software Metrics 2001, pp. 16-27.
- Kemerer, C.F. An empirical validation of software cost estimation models. Communications ACM, 30, 5, 1987
- Kitchenham, B.A. and Taylor, N.R. (1985) Software project development cost estimation models, *Journal of Systems and Software* 5, pp267-278.
- Kitchenham, B.A. Empirical studies of assumptions that underlie software cost estimation models, *Information and Software Technology*, 34(4), 1992, pp. 304-310.
- Kitchenham, B.A. The question of scale economies in software – why cannot researchers agree? IST, 44, 2002, pp. 13-24.
- Kitchenham, B.A. and Mendes, E. A Comparison of Cross-company and Within-company Effort Estimation Models for Web Applications, 2004.
- Kitchenham, B., Pfleeger, S.L., McColl, B and Eagan, S. An Empirical Study of Maintenance and Development Estimation Accuracy, JSS, 64, 2002, pp57-77.

References

- K. Maxwell and P. Forselius. Benchmarking Software Development Productivity. IEEE Software 17(1), 2000, pp80-88.
- Putnum, L.H. A general empirical solution to the Macro sizing and estimating problem, IEEE Trans SE, 1978, pp345-361.
- Ruhe, M., Jeffery, R., Wieczorek, I. Using Web Objects for estimating software development for Web applications. Proc. 9th International Software Metrics Symposium, Sydney, IEEE Computer Spciety Press, 2003, pp. 30-37.
- Stensrud, E. and I. Myrtveit. (1998) Human performance estimating with analogy and regression models: An empirical validation, Proceedings 5th International Software Metrics Symposium (METRICS '98) IEEE Computer Society Press, Los Alamitos, CA
- Walston, C.E. and Felix, C.P. A method of programming measurement and estimation, IBM Sys. J., 16(1), 1977, pp. 54-73.
- Walkerden, F. and Jeffery, D.R. Software Cost estimation: A review of models, process and practice. Advances in Computers, pp 59-125, 1997.
- Wolverton, R.W. The cost of developing large-scale software, IEEE Trans SE, 1974, pp615-636