

An Investigation of Software Development Productivity in China

Mei He, Mingshu Li, Qing Wang, Ye Yang, and Kai Ye

*Laboratory for Internet Software Technologies
Institute of Software , Chinese Academy of Sciences*

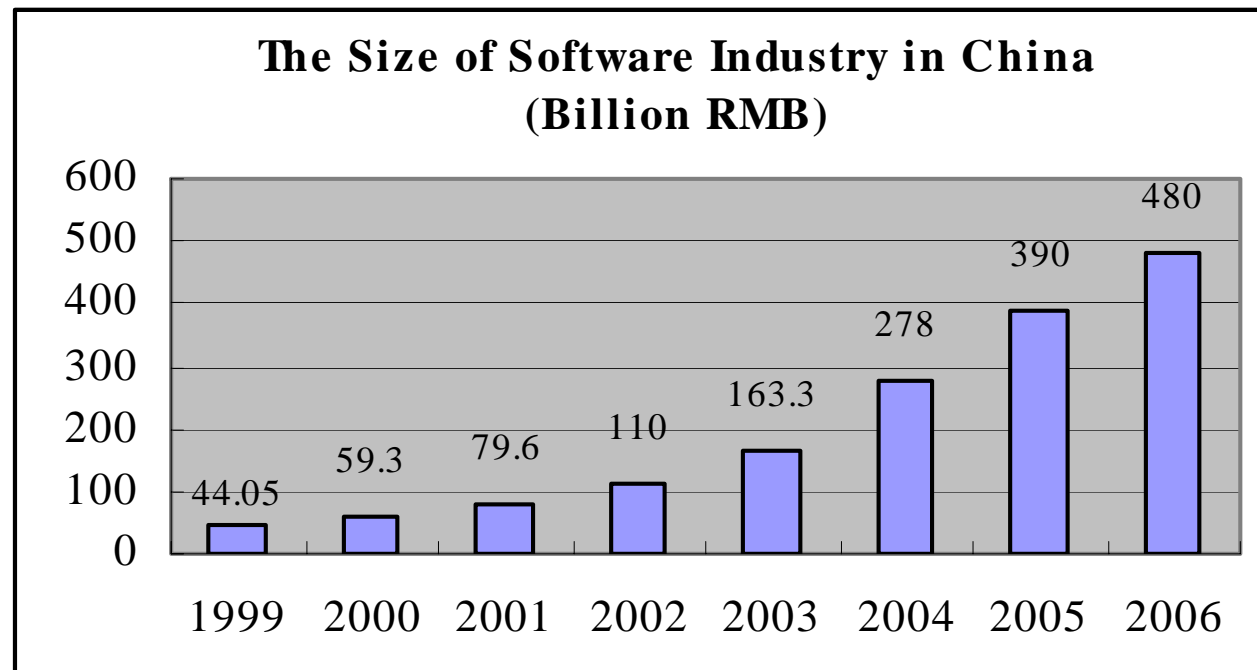
Agenda

- Introduction
- Motivation and Goals of the Investigation
- Analysis Results
- Conclusions

Agenda

- *Introduction*
- Motivation and Goals of the Investigation
- Analysis Results
- Conclusions

Introduction



Growth Trend

Introduction (con.)

- China grows into one of the major outsourcing service bases.
- However, there is lack of comprehensive studies on Chinese software industry.
- **Our study**
 - An investigation on software productivity in Chinese software industry using 999 software project data points

Agenda

- Introduction
- *Motivation and Goals of the Investigation*
- Analysis Results
- Conclusions

Software Productivity

- Definition:

$$\text{Productivity} = \text{Size} / \text{Effort}$$

- Different methods

- Line-of-code productivity
- Function Point Productivity

Related Work

Critical factors influencing productivity in previous studies

Data source	Year	N	Factors discussed
European Space Agency data (from 8 countries)	1996	99	Company, Language, Category, Environment, Team size, Duration, KLOC & a set of COCOMO drivers
Software Technology Transfer Finland (old)	2000	206	Company, Business sector, Operating system, DBMS tools, Hardware platform etc.
STTF (new)	2004 2005	622	Company, Process model, Business sector, Year, Hardware
ISBSG (Release 8)	2003	1238	organization type, business area, development type, primary programming language etc.
ISBSG (Release 10)	2007	>3000	Team size, Language, Dev Type, Platform, Dev techniques, Dev methodology, CASE tool

Summary of CSBSG Database

Summary of the CSBSG project data

Variable	N	Mean	Median	Min	Max
Project size (LOCs)	999	666341	36109	771	2339728
Effort (person hours)	999	6207	3152	132	134840
Team size	998	9	6	1	110
Development Region	From 9 primary regions including Beijing, Shanghai, Jiangsu, Tianjin, Shandong, Liaoning, Zhejiang, Anhui, Chongqing (#Obs.> 10)				
Business Area	Across 7 primary business areas including Public admin, Energy, Manufacturing, General, Retail & Inventory, Finance, Telecom (#Obs.> 20)				
Language type	Consisting of 7 primary language types: C#, Java, C++, ASP, VB, C, Cobol (#Obs.> 10)				
Development Type	New development, Enhancement, Re-development				

CSBSG: Chinese Software Benchmarking Standard Group (2006 - present)

Results from GLM analysis

Variance of productivity explained by each variable

Variables	Type	Variance explained
Organization	Class	91.32%
Development type	Class	25.66%
Business area	Class	21.36%
Region	Class	21.07%
Language	Class	16.33%
Project size	Nonclass	14.84%
Team size	Nonclass	2.63%

Research Questions

- **RQ1:** Do organizations differ a lot from each other in productivity?
- **RQ2:** Does new development always cost more effort than enhancement?
- **RQ3:** What business area corresponds to a higher productivity rate? What corresponds to a lower one?
- **RQ4:** Is there a relationship between software productivity and the region/city where the organization is from?
- **RQ5:** Does programming language matter in terms of affecting productivity?
- **RQ6:** Will larger team size reduce the productivity?
- **RQ7:** Will larger project size reduce the productivity?

Agenda

- Introduction
- Motivation and Goals of the Investigation
- *Analysis Results*
- Conclusions

Factor 1: Organization

- **Finding:** Organizations in the CSBSG database display great difference in productivity.
- Common intuition
- Consistent with some analysis on data from:
 - European Space Agency data (1996)
 - Software Technology Transfer Finland data (2000, 2004)

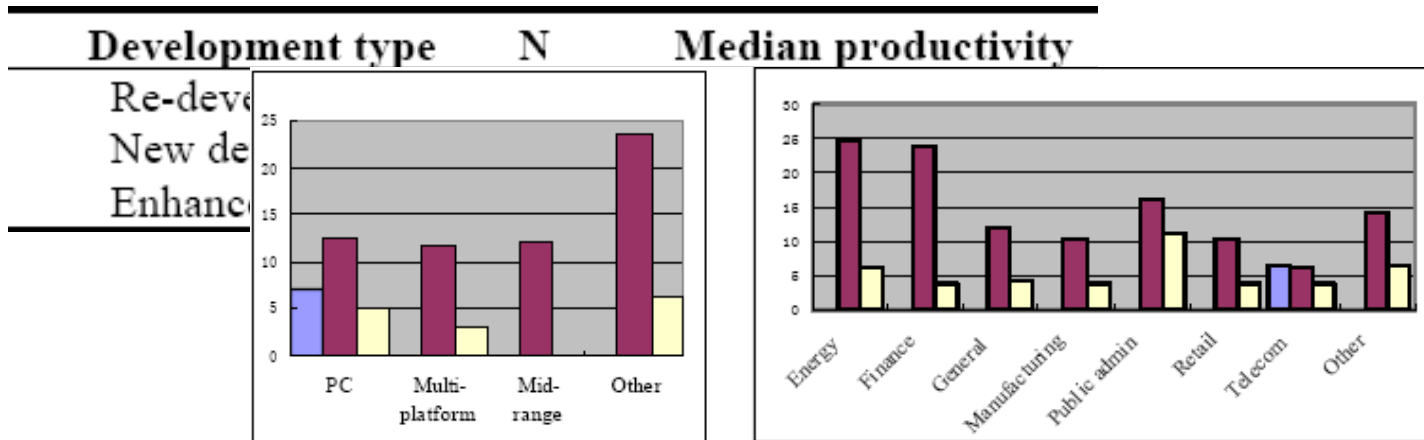
Factor 1: Organization (con.)

- ***Recommendation for Improvement:***
 - be cautious while using cross-organization data
 - better to build their own database

Factor 2: Development Type

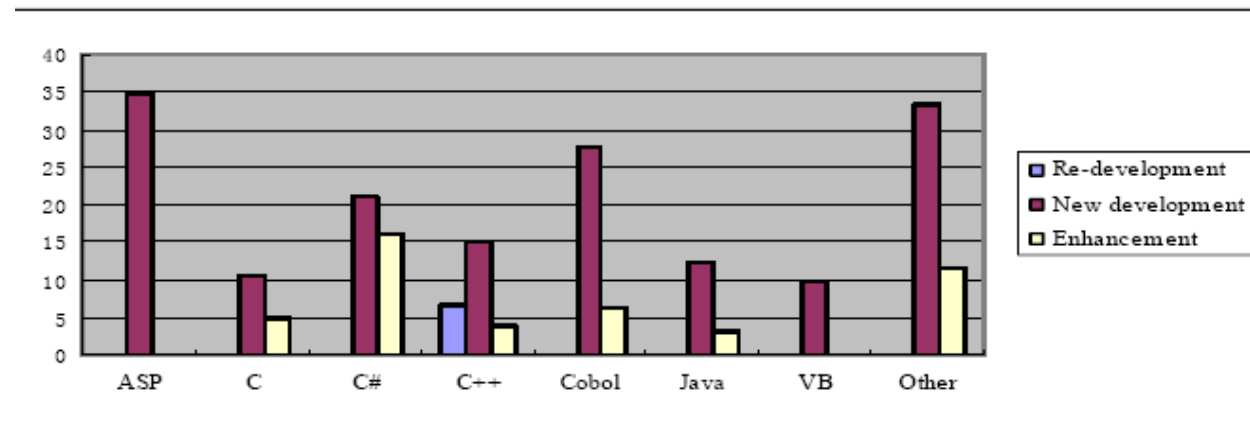
- **Finding:** Productivity in new development is found to be higher than in enhancement in the CSBSG database.

Productivity by Development Type



a) Different development platforms

b) Different business areas



c) Different languages

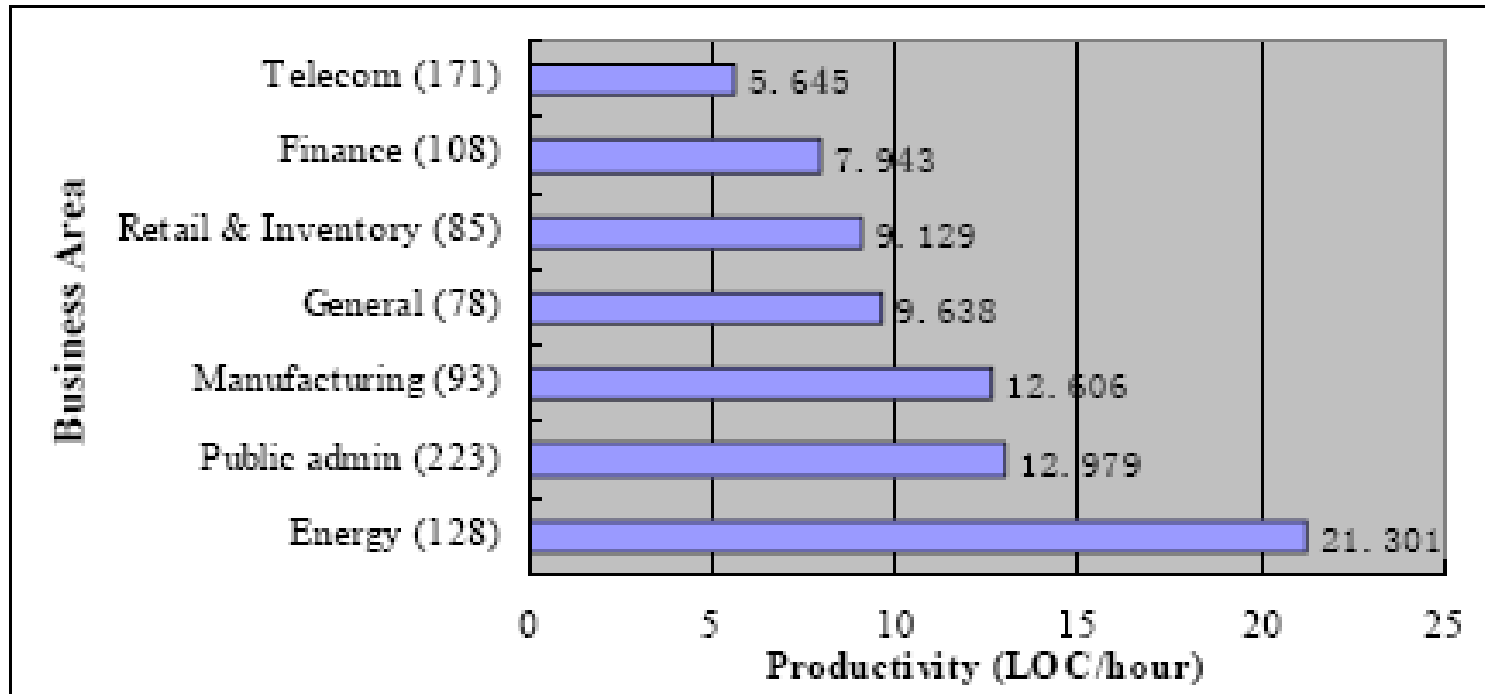
Factor 2: Development Type (con.)

- ***Recommendation for Improvement:***
 - Continue to use the same development team
 - Disciplined documentation in new development

Factor 3: Business Area

- **Findings:** Business areas of
 - Low productivity are
 - Telecom and Finance
 - High productivity are
 - Public Admin, Manufacturing and Energy

Factor 3: Business Area (con.)



Productivity by business areas

Factor 3: Business Area (con.)

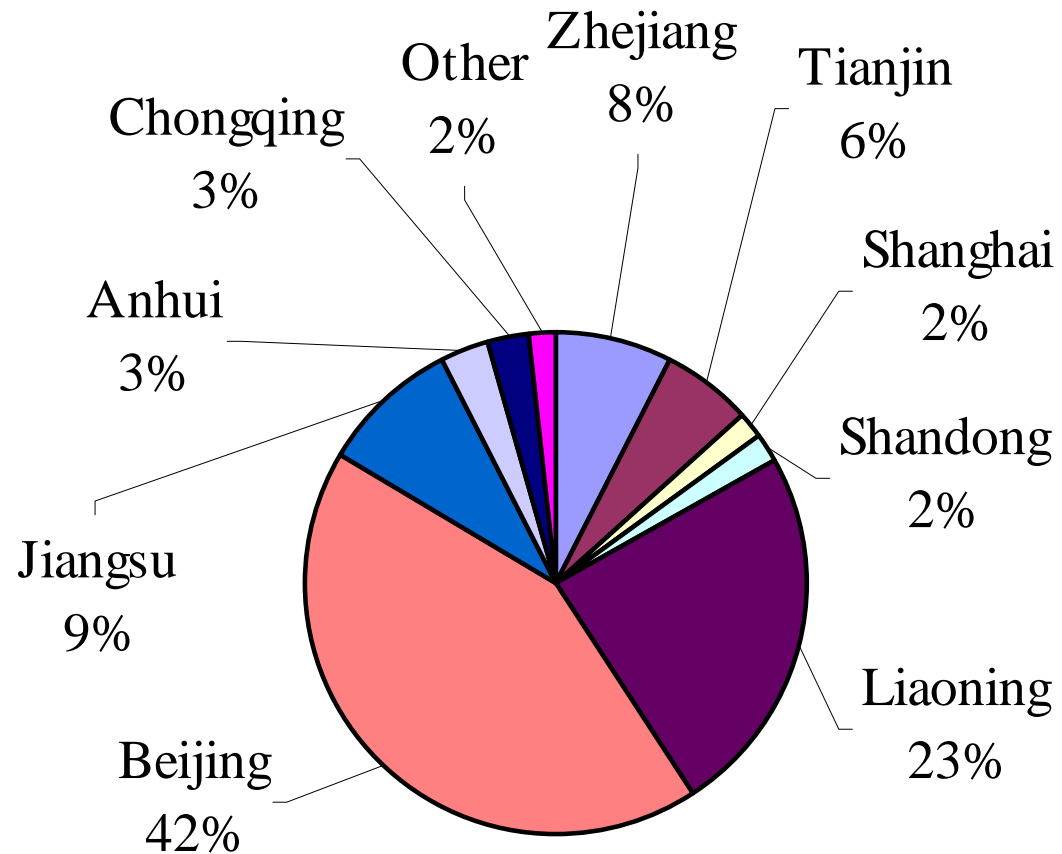
- ***Recommendation for Improvement :***
 - Pay attention to
 - Degree of requirement stability
 - Speed and complexity of technology advances
 - Degree of system heterogeneity

Factor 4: Region

- **Finding:**

- In comparatively more developed regions
 - more high-productivity organizations
 - high-productivity organizations appear in each business area.

Factor 4: Region (con.)



Projects Distribution in Main Regions

Factor 4: Region (con.)

- *Recommendation for Improvement :*
 - Factors such as high quality and availability of labor resource, preferential policies, and large numbers of excellent enterprises etc. contribute to software industry in some regions
 - Concentrate on few particular domain of software in less developed regions

Factor 5: Language

- **Finding:** High level programming languages are found to be more productive.

Factor 5: Language (con.)

Productivity by Language

Language	N	Median productivity	Average LOCs per FP
ASP	14	34.68	14
C#	205	18.68	25
VB	48	9.94	n/a
Java	362	9.58	53
C++	217	7.52	53
C	95	6.41	128
Cobol	24	6.27	107

Factor 5: Language (con.)

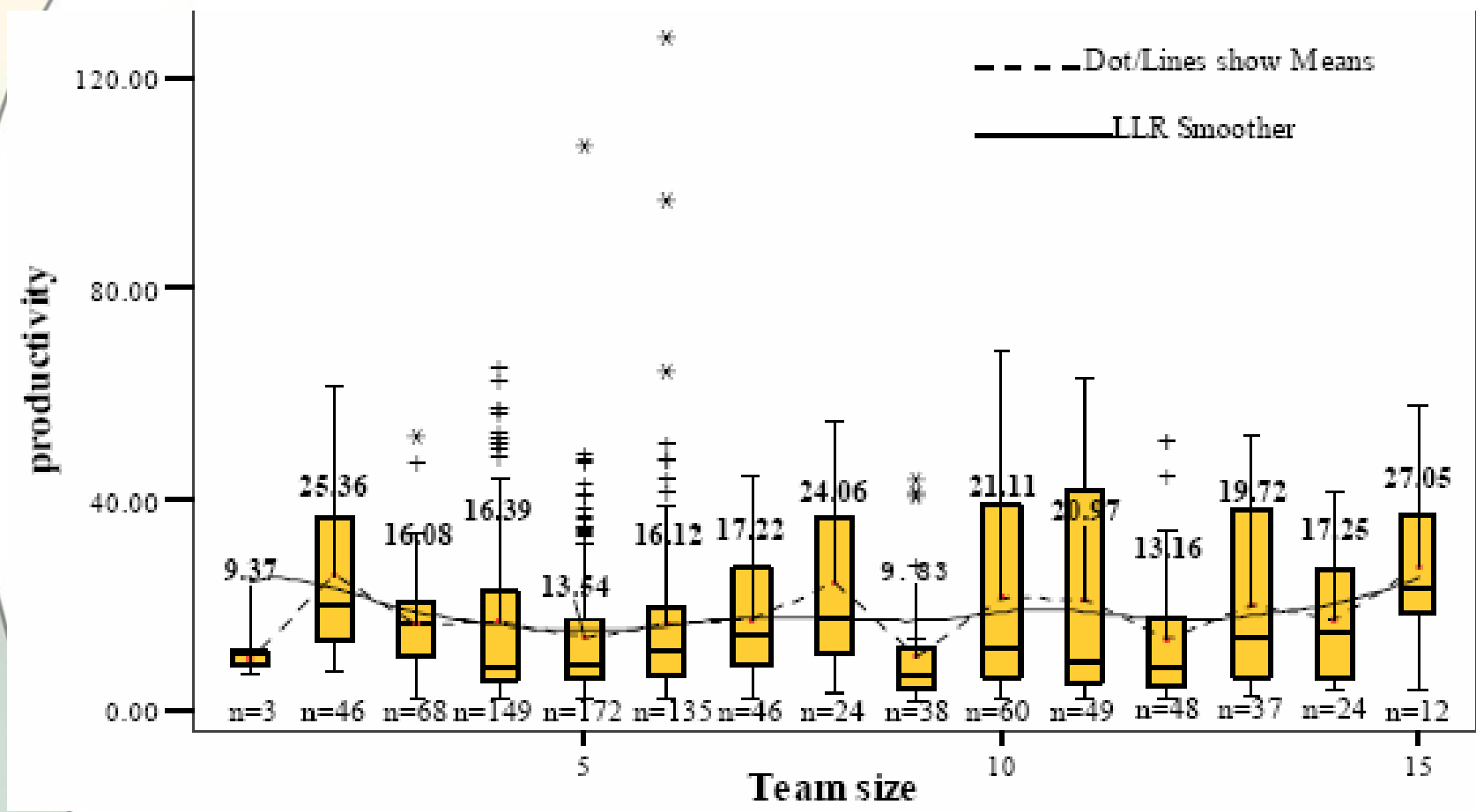
- ***Recommendation for Improvement :***
 - Pay attention to high level language
 - While selecting languages, better to also consider the required software performance, such as scalability, security etc.

Factor 6: Team Size

- **Finding:**

- Increasing team size is found to have a negative influence on productivity.

Factor 6: Team Size (con.)



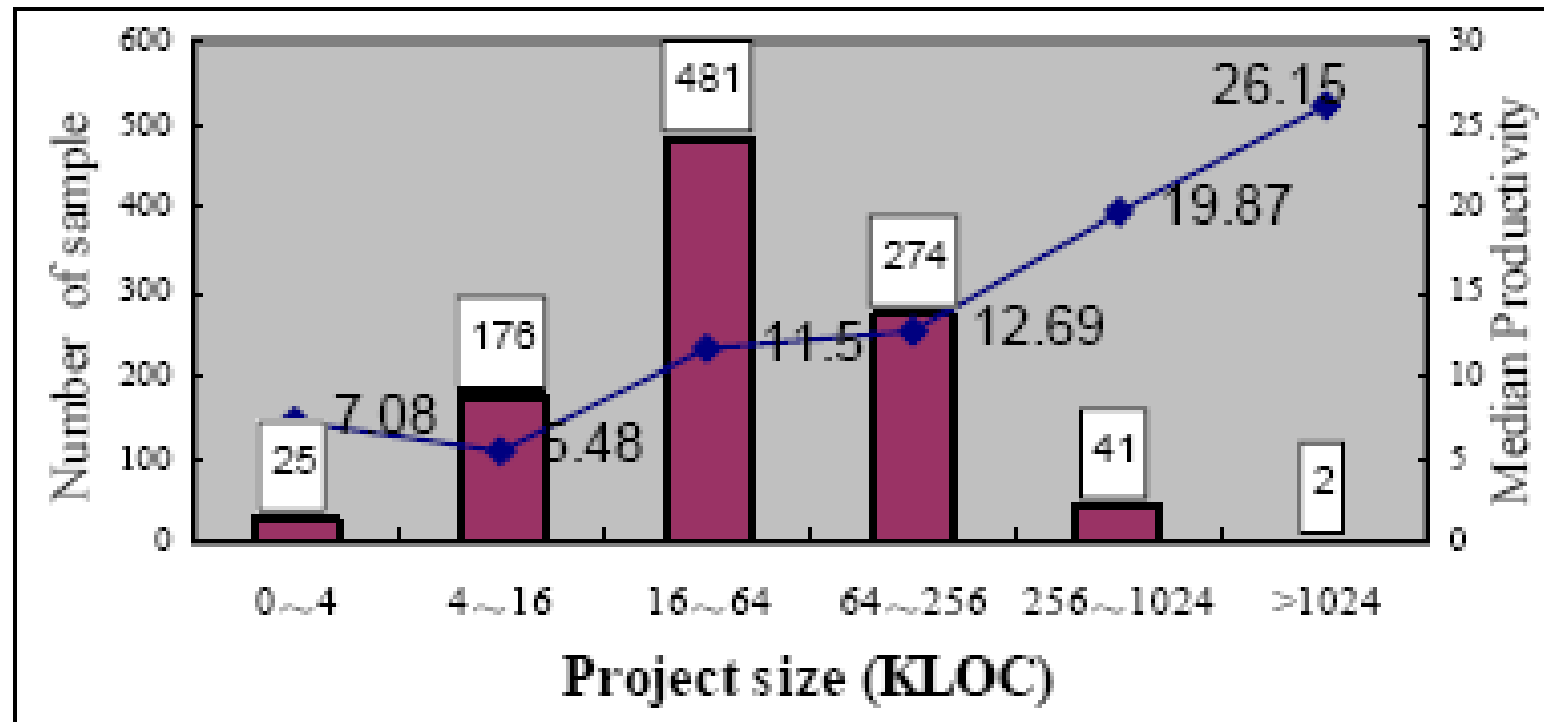
Factor 6: Team Size (con.)

- *Recommendation for Improvement :*
 - pay attention to personnel coordination and resource harmony while the team size extends largely
 - Be more cautious while team size grows to more than 15-person

Factor 7: Software Size

- **Finding 7 :** Productivity is found to increase with increasing software size (Lines of code).

Factor 7: Software Size (con.)



Agenda

- Introduction
- Motivation and Goals of the Investigation
- Analysis Results
- *Conclusions*

Conclusions

- In the CSBSG data, the most significant factors impacting software productivity are:
 - Organization
 - Region
 - Business area
 - Development type
 - Language
 - Team size
 - Software size

Conclusions (con.)

- **Limitation**
 - cross influence: interrelation among factors
- **Ongoing study**
 - An multi-variable model has been established to predict software project effort based on the CSBSG data
 - The modeling result confirms most of the findings presented in this paper

Thank You!