An Empirical Analysis of Software Productivity

Martin Shepperd, Carolyn Mair and Pekka Forselius

Abstract

The aim of our research is to discover what factors impact software project productivity (measured as function points per hour) using real world data. Within this overall goal we also compare productivity between different business sectors and project types. We analysed a data set of almost 700 projects that have been collected by STTF from a number of Finnish companies since 1978. These projects are quite diverse type (new and maintenance projects), in terms of size (6 to over 5000 function points), effort (55 to over 60000 person hours), application domain and implementation technology. There are three main findings. First productivity varies enormously between projects. Second, project type has limited influence on productivity. Third, application domain or business area has a major impact upon productivity. Because this data set is not a random sample generalisation is somewhat problematic, we hope that it contributes to an overall body of knowledge about software productivity and thereby facilitates the construction of a bigger picture.

Keywords*: project management, projects, software productivity, empirical analysis.*

1. Introduction

Defining productivity is problematic for a number of reasons [1]. However, it remains an important aspect of software project management and a precursor to effective cost prediction. For the purposes of this research we define productivity using Function Points (FP) as a measure of output and person hours of effort as the input.

Our goal is to investigate some of the more important factors which impact software project productivity using real world data from Finnish software projects. This work is motivated by previous work of Maxwell and Forselius [2]) who found that the particular company and the business sector were the most important variables in explaining productivity. More recently, Premraj and Shepperd [1] found that the most significant factors in explaining productivity, were in decreasing order of importance, Company, Business Sector, Year, and lastly Hardware.

The remainder of the paper is organised as follows. Section 2 describes the data and outlines briefly rules for removing and classifying the data prior to inclusion in the data base. In Section 3 we analyse the data set in terms of delivery rate (FP/hour). We conclude in Section 4 with a discussion of the analysis and comment on the factors which impact software project productivity.

2. The "Finnish" data set

The data set used for analysis in this paper is derived from the 2006 release of the Experience data set, also known as the "Finnish" data set. This data set is a commercial initiative by Software Technology Transfer Finland (STTF) to provide support for software development organisations for both project cost estimation and productivity analyses. An annual subscription allows organizations to access the data by means of Experience Pro[1]. This tool is also used to submit data and ensures standardization of variables. Before being added to the data base, the

[1] Information about the Experience Pro tool is available at www.sttf.fi/ExperiencePro.htm

project data are carefully assessed by experts at STTF. The database comprises new development or enhancement projects derived from a wide range of business sectors, platforms and development technologies. Project data include size (FPs), effort and a range of factors that characterize the type of project, the development circumstances, development and target technology. The possibility exists to submit data on 102 variables, although some variables are difficult to analyse because of the significant proportion of missing values. A fuller description of the data set may be found in Maxwell and Forselius [2, 3] and further information about Experience Function Points.

The data set comprised 737 projects completed between 1978 and 2005. The organisation responsible for collecting the data then removed 54 projects on the grounds of data quality concerns. We then removed a further 22 projects on the grounds of implausible delivery rate. This is discussed further in the next section.

Table 1: Data cleaning of the Finnish data set

Data set	Count	Percentage
Complete data set	737	100.0%
"Low quality" data removed	683	92.7%
Implausible data removed	661	89.7%

3. Analysis

3.1 Delivery Rate stats

Whilst the main theme of the paper is productivity we chose to use its inverse i.e. 'Delivery rate' (hour / Function Points) as the indicator of productivity since we believe this is easier to visualise. Essentially productivity generally leads to small fractions of an hour whereas delivery rate yields values greater than unity. However, one can be converted to the other merely by taking the reciprocal

As previously discussed because of the large range (more than three orders of magnitude) in delivery rates we removed outliers, as suggested in Premraj and Shepperd [1]. To do this we, in consultation with STTF staff, considered values with extremely high or low delivery rate to be implausible as they represent unreported or misreported factors. The values which are considered 'extreme' are <1 FP/hour and >30 FP/hour. There were 22 cases with implausible data for delivery rate, these were removed. Summary statistics for the remaining 661 cases are given in Table 2.

Table 2: Summary statistics for 661 projects (projects with outliers data have been removed)

Variable	Mean	Median	Min	Max
Delivery rate (FP/hr)	7.487	6.28	1.09	29.825

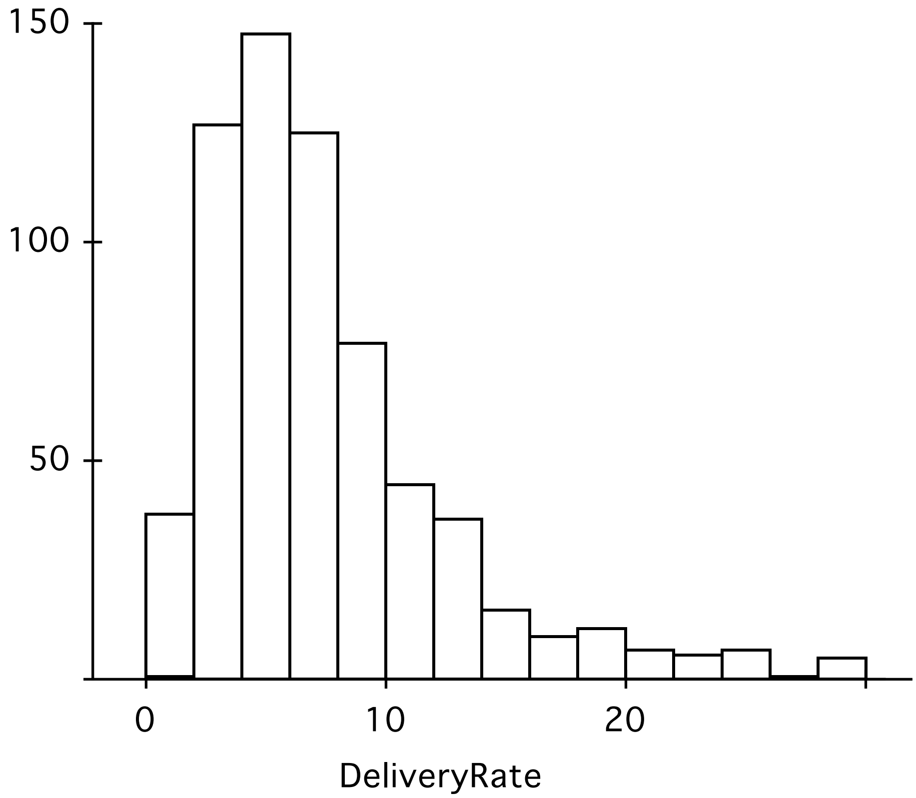

Figure 1: Delivery rate for 661 projects (projects with outliers data have been removed)

3.2 Naïve model of size and effort

The first step was to construct a simplistic regression model using just size in function points as the independent variable (Size). From the scatterplot (Figure 2) it can be seen that there is a positive relationship such that as size increases there is a tendency for effort to also increase. However, it can also be observed that there is a substantial amount of scatter about the regression line. The coefficient of determination (R^2) is 63.5% (see Table 3) which indicates that the naïve model can predict less than two thirds of the overall variation in effort. The other interesting feature of this model is that the intercept is not significantly different from zero, i.e. the 95% standard error (s.e.) when combined with the coefficient encompasses zero.

Table 3:Regression model statistics

	Coefficient	s.e. of Coeff	prob
Intercept	-22.3582	186.9	0.9048
Size	7.39491	0.2182	≤ 0.0001

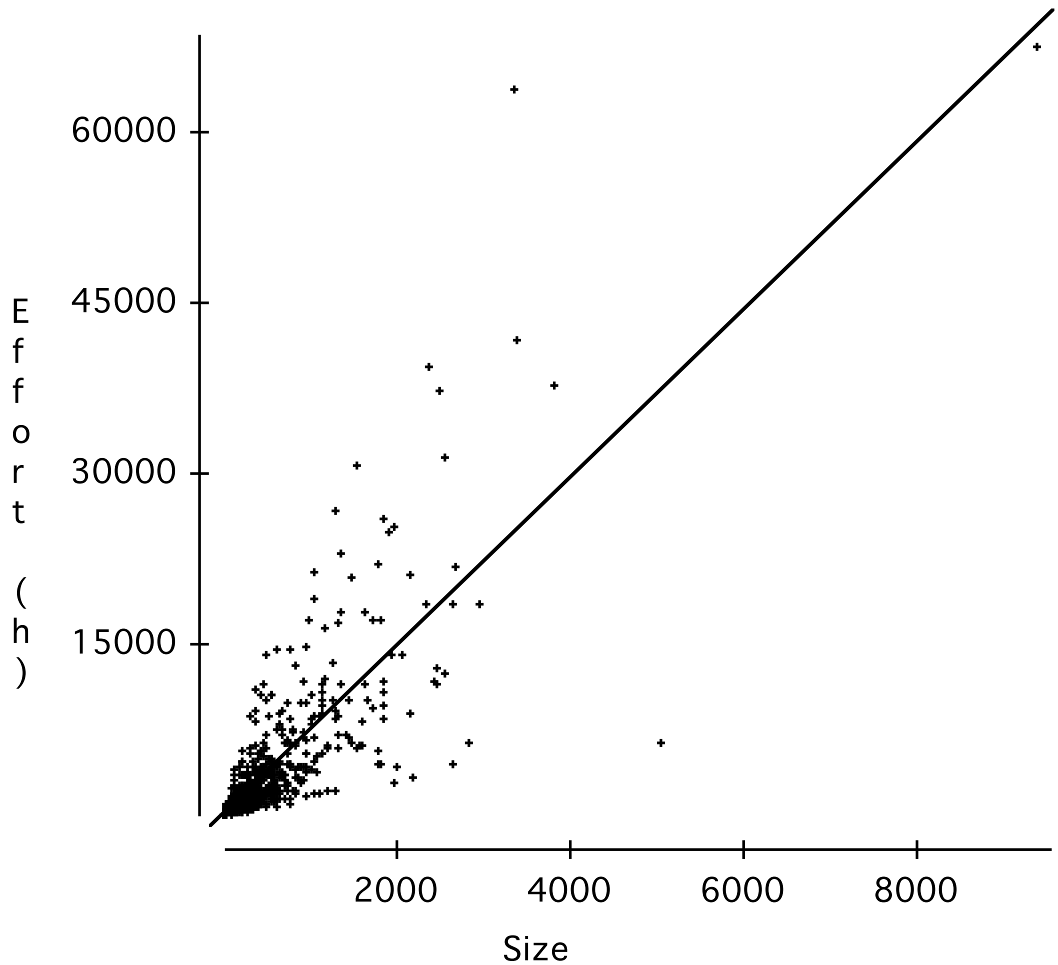

Figure 2: Scatterplot of size and effort

Since the naïve model only has a moderate fit to the data we next consider the impact of project type upon delivery rate.

3.3 Analysis by project type

Originally the projects in the data set were classified into 8 types (Table 4). However, for ease of analysis, these were regrouped into 4 types: Enhancement, Maintenance, New Development (incorporating both product and tailored) and Other (incorporating the remaining types) (refer to Fig. 3 for a pie chart describing the relative proportions of each category). Combining the projects into 4 types resulted in new development projects accounting for 76% of the data set, maintenance projects 13%, enhancement projects 9% and other 2%. The group 'Other' project type is composed of projects which do not fit into mainstream categories.

Table 4: Project type categories and frequencies

Project type	Count	Merged project types	Revised count	%
Annual maintenance	1	Other		2.6
Integration	2	Other		
Conversion	6	Other		
Other	8	Other	17	
New development product	12	New development	478	72.3
Enhancement	70	Enhancement	70	10.6
Maintenance	96	Maintenance	96	14.5
New development tailored	466	New development		

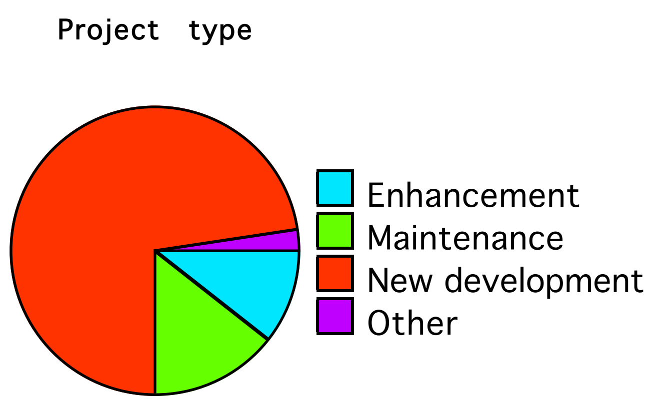

Figure 3: Project type (combined categories)

Table 5: Summary statistics for delivery rate for each category of project type

Project type	Count	Mean	Median	Lower quartile	Upper quartile
Enhancement	70	7.913	6.551	4.381	9.325
Maintenance	96	6.545	5.815	4.243	8.692
New development	478	7.720	6.506	3.965	9.567
Other	17	4.499	3.967	2.138	5.988

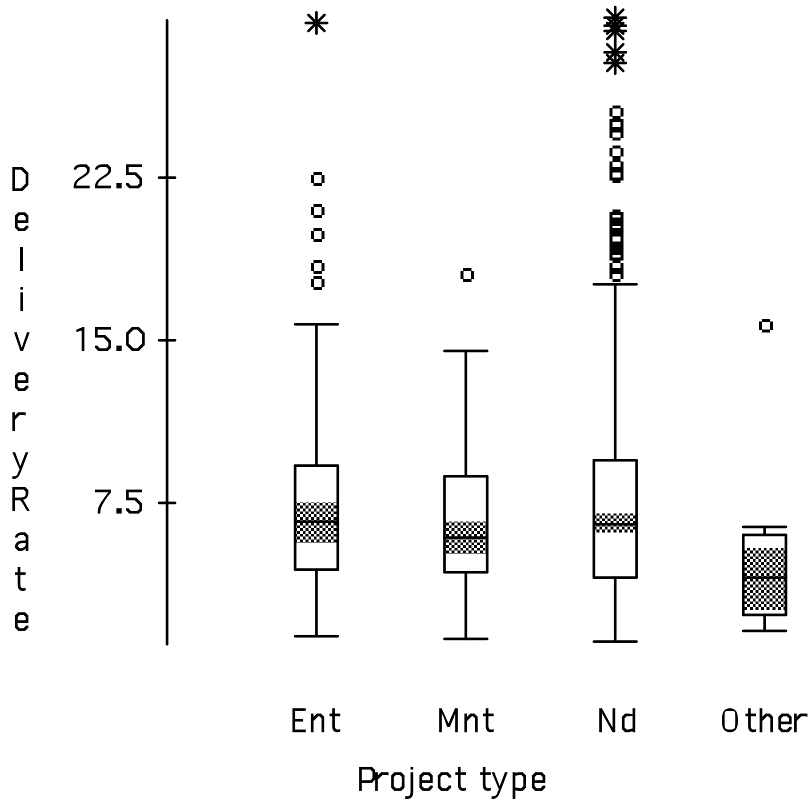

Figure 5: Side by side boxplots for Project type delivery rates

[4]

Next we conducted an analysis of variance for the different Delivery Rates for each project type to see if any of the differences revealed by the boxplots (Figure 5) were actually significant. The computed F-ratio is 3.472 with a tabled probability of 0.0159. This suggests that Project type has a significant impact upon delivery rate. However, we need to determine more specifically when this was the case so we conduct pairwise Mann-Whitney tests to compare each type. From this we find that the only significant differences involve the group type Other (New development p=0.004; Maintenance p=0.006; Enhancement p=0.002)

There is no significant difference between the delivery rate for enhancement and new development projects. However, there is a significant difference between projects grouped into 'Other' projects and the other three types. One might expect some difference in delivery rate to be present among the project types. However, this was detected only between 'Other' projects and the remaining three project type groups. The implications are that it is very difficult to have an effective universal productivity model and that the full diversity of project types must be taken into account.

3.4 Analysis by business type

Having established that there is no significant difference between project types, we now consider the impact of business type on delivery rate. Other studies have reported this factor to be highly influential [4]. The number and proportion for each business type is shown in Figure 6. Business type 'Other' now includes Construction.

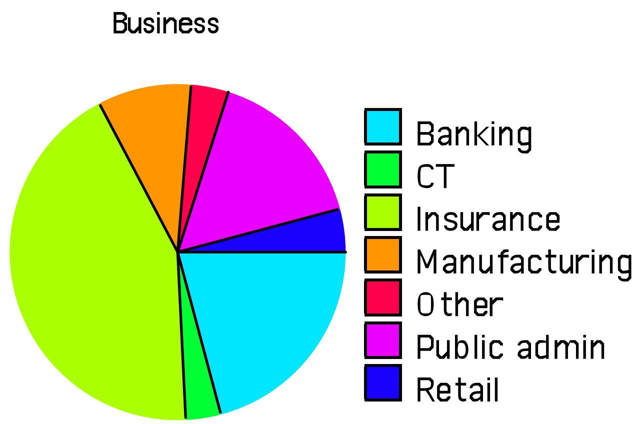

Figure 6: Pie chart of Business Type proportions

Table 6: Business type categories and frequencies

Business type	**Count**	**%**
Banking	137	20.726
CT	23	3.480
Insurance	285	43.116
Manufacturing	60	9.077
Other	23	3.480
Public admin	105	15.885
Retail	28	4.236

Table 7: Summary statistics for delivery rate for each category of Business Type

Business type	**Count**	**Mean**	**Median**	**Lower quartile**	**Upper quartile**
Banking	137	9.657	7.651	5.002	12.262
CT	23	9.805	6.573	4.284	15.139
Insurance	285	8.010	7.018	4.762	10.185
Manufacturing	60	4.574	3.989	2.564	5.932
Other	23	5.333	4.275	2.811	5.210
Public admin	105	5.439	4.649	3.023	6.679
Retail	28	5.317	4.600	3.198	7.457

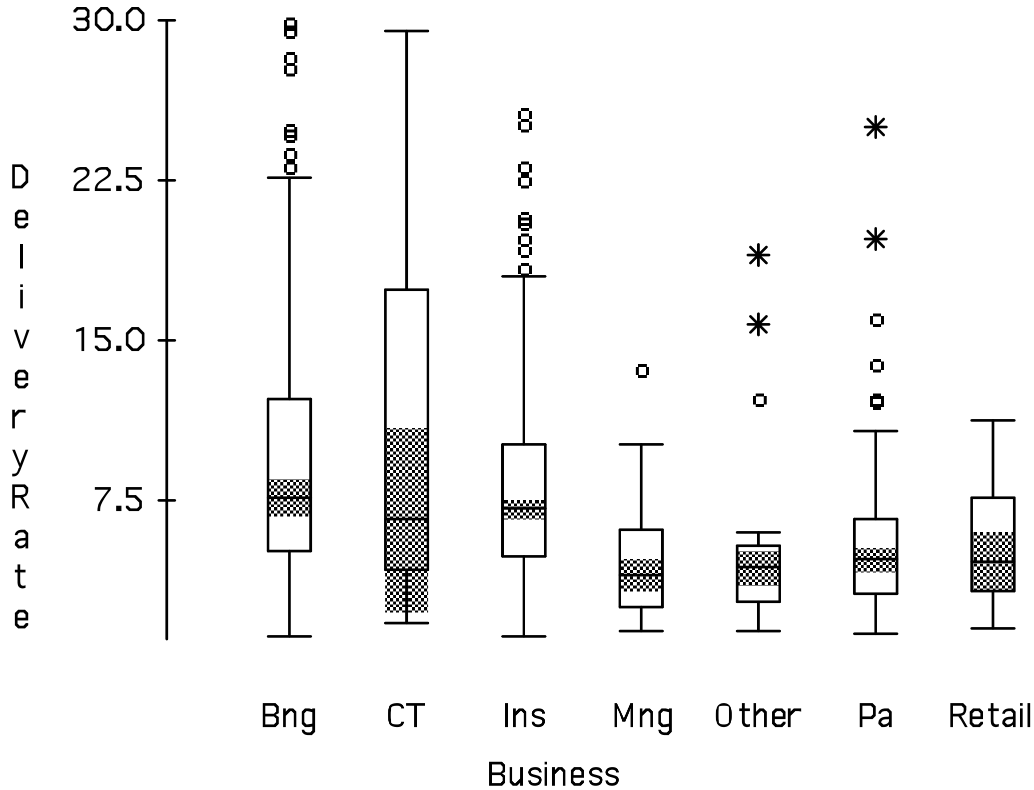

Figure 7: Side by side boxplots for Project type delivery rates

Next we conducted an analysis of variance for the different Delivery Rates for each business type to see if any of the differences revealed by the boxplots (Figure 4) were actually significant. The computed F-ratio is 14.201 with a tabled probability of <0.0001. This suggests that business type has a highly significant impact upon delivery rate. Again, we need to determine more specifically when this the case so we conduct pairwise Mann-Whitney tests to compare each type. From this we find that the business types seem to fall into two groups (Banking, CT and Insurance) which significantly more productive than the remaining business types (Manufacturing, Other, Public Admin and Retail). The implication here is that it is better to tailor a specific cost model to the individual business type rather than adopt a "one size fits all" strategy.

4. Discussion

In this paper we have described a preliminary analysis of software productivity (delivery rate). The principle findings are that even after considerable efforts to remove rogue data there is still a great deal of variability in productivity. Consequently, naïve models that seek to find a linear relationship between size (measured in function points) and effort only exhibit moderate (although highly significant) fit with the data. Next we observed that whilst to our surprise there were limited differences between maintenance/enhancement projects and new development there were a number of Other projects that did not fit into this classification that displayed very different productivity behaviour. On the other hand we found that business type or application domain had a major impact upon productivity. For this reason we believe some caution must be exercised when using "one size fits all" style of cost model.

These results show some similarity with other studies, particularly the large variability in productivity and the impact of different business types (see for example Maxwell and Forselius [2]). Clearly it would be interesting to make a more systematic comparison particularly with other well known repositories such as the ISBSG data set.

Lastly, is the question so what or how can we generalise from these results? A clear difficulty is that this is a non-random sample from a rather ill-defined population of software projects. It is likely that the projects come from companies who are interested in measurement and process improvement. This would likely make them better than level one in terms of the CMM model [5]. In other words we are probably dealing with the top quartile of projects from non-safety critical, non-embedded system type applications. However, we hope that as more studies such as this are published it enables a larger picture to be established and greater confidence to be placed in the findings.

References

[1] R. Premraj, M. Shepperd, B. Kitchenham, and P. Forselius, "An Empirical Analysis of Software Productivity Over Time," presented at 11th IEEE Intl. Softw. Metrics Symp. (Metrics05), Como, Italy, 2005.

[2] K. D. Maxwell and P. Forselius, "Benchmarking software development productivity," *IEEE Software*, vol. 17, pp. 80-88, 2000.

[3] K. Maxwell, L. Van Wassenhove, and S. Dutta, "Performance evaluation of general and company specific models in software development effort estimation," *Management Science*, vol. 45, pp. 787-803, 1999.

[4] K. Maxwell, L. Van Wassenhove, and S. Dutta, "Software development productivity of European space, military and industrial applications," *IEEE TSE*, vol. 22, pp706-718, 1996.

[5] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V. Webber, "Capability maturity model, version 1.1," *IEEE Software*, vol. 10, pp. 18-27, 1993.