

## **Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - A Case Study**

*Luigi Buglione<sup>1</sup>, Filomena Ferrucci<sup>2</sup>, Cigdem Gencel<sup>2</sup>, Carmine Gravino<sup>2</sup>,  
Federica Sarro<sup>2</sup>*

<sup>1</sup>*École de Technologie Supérieure (ETS) - Montréal (Canada); Engineering.IT –  
Rome (Italy)*

<sup>2</sup> *University of Salerno (Italy)*

<sup>3</sup> *Blekinge Institute of Technology, Karlskrona (Sweden)*

luigi.buglione@eng.it, {fferrucci, gravino, fsarro}@unisa.it, cigdem.gencel@bth.se

### ***Abstract:***

*Estimation is still a challenging process for planning and managing software projects. Often, estimates are being done on experiential or analogous basis or using effort estimation models. Mostly, these approaches take software size (e.g., Lines of Code, Function Points, Object points) and other cost factors as the main inputs to estimation. This study focuses on functional size based effort estimation for Web application development and investigates the significance of the functional sizes of each of the COSMIC Base Functional Component (BFC) types in explaining the variation in the development effort. A case study was conducted collecting data on 25 Web projects from a software organization. The results show that the size of only one of the BFC Types can explain the variation in the effort nearly as good as the total functional size.*

### ***Keywords***

*Functional Size Measurement, COSMIC, Effort Estimation, Base Functional Component Types, Web Applications, Regression Analysis, Case Study.*

## **1 Introduction**

Software managers still experience challenges especially in estimating, planning and managing large-scale software projects even though considerable amount of effort has been put forth on improving the estimation accuracy in the last 30 years.

A very large number of effort estimation methods, often with their associated tool supports, have been developed. Moreover, software engineering community has identified the need to develop benchmark datasets at organizational, national

and global levels and significant attempts have been done. However, there is still no software estimation method that has wide-scale acceptance and the software community has a tendency not to trust existing or newly developed methods [26].

Most of the time, the software effort estimation models or methods take software product size as the main predictor. Among those size measures, software Functional Size Measurement (FSM) methods have evolved quite a bit since Function Point Analysis (FPA) was introduced by Albrecht in 1979 [4]. Many variations have been developed since then [24]. These improvements made FSM methods more mature in terms of their conceptual basis and principles.

However, functional size based effort estimation still needs further investigation. Taking the total functional size as the main input, most studies aim to identify the cost factors in developing estimation models. In the last years, a few studies were conducted (see Section 2.2) to investigate whether considering the functional sizes of each of the Base Functional Component (BFC) types would better explain the variation in the effort than the total functional size. This study is a further investigation of these ideas through an empirical study utilizing a rather homogeneous Web application project dataset collected from a software organization.

The paper is organized as follows: Section 2 summarizes the related work. Section 3 presents the case study and Section 4 the results. Finally, the conclusions are given in Section 5.

## **2 Related work**

There is a plenty of previous work discussing the current state of the art for the effort and cost estimation models and their categorizations (e.g., [1][6][8][26][39]). Considerable amount of empirical studies, which take software size as the main predictor for software development effort (e.g., [1][5][31][32][34]), were performed to evaluate the reliability of the proposed models. In the following, we only recall the studies that took functional size measured by the Common Software Metrics International Consortium (COSMIC) method [13] as the main input in estimating Web application development effort.

### **2.1 Sizing and Estimating Web Applications with COSMIC**

The difficulties of applying the International Function Point Users Group (IFPUG) Function Point Analysis (FPA) to size an Internet bank system motivated first Rollo [43] to use COSMIC in the context of Web applications. However, he did not present any empirical results supporting his thesis. Subsequently, some studies have investigated the significance of COSMIC

## Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

---

functional size in estimating the development effort for Web applications ([16][18][19][38][43]). In particular, Mendes *et al.* applied COSMIC method to measure the functional size of Web sites, without server-side elaborations [38]. Using data from 37 Web systems developed by academic students, an effort estimation model was built applying Ordinary Least Squares Regression (OLSR). Unfortunately, this model did not provide good estimations and replications of the empirical study were highly recommended to find possible biases in the collection of the data and/or in the application of the method.

Subsequently, the observation that dynamic Web applications are mainly characterized by data movements (from a Web server to the client browser and vice-versa) suggested to apply the principles of the COSMIC method to size this type of Web applications [16]. An empirical study based on 44 Web applications developed by students in academia, was performed to assess the COSMIC approach [16]. The effort estimation model obtained by employing the OLSR provided encouraging results.

Recently, Ferrucci *et al.* conducted a case study aiming at investigating the significance of COSMIC functional size in Web application development effort estimation by exploiting a single-company dataset (obtained from a set of 15 Web applications developed by an Italian software company) [18]. The Web Objects size measure, proposed by Reifer for the Web [41], was also applied. Web Objects are characterized by the introduction of four new Web-related components together with the five Base Functional Components (BFC) of the IFPUG FPA method, namely Multimedia Files, Web Building Blocks, Scripts, and Links. The estimation models were built by applying OLSR and were validated by using a hold-validation approach. In particular, the performance of the obtained models was evaluated using a dataset of further 4 Web applications developed by the same software company some time after the first 15 Web applications. The results revealed that both COSMIC and Web Objects were good indicators of the development effort.

By exploiting the same dataset of 15 Web applications Ferrucci *et al.* also assessed the effectiveness of COSMIC when used in combination with WebCOBRA [19]. WebCOBRA is an extension for the Web of the COBRA method proposed by Briand *et al.* [6]. This can be considered as a composite method according to the taxonomy defined by Briand and Wiczorek [6] since it exploits expert's opinions collected in a controlled fashion together with other cost drivers within an algorithmic approach. The performed empirical analysis confirmed the positive results of a previous study that employed WebCOBRA in combination with Web Objects [44].

## 2.2 The Significance of Base Functional Components of Functional Size in Effort Estimation

In [2], Abran *et al.* defined software ‘functional profile’ as the relative distribution of the BFC Types for any particular project. They investigated whether or not the size-effort relationship was stronger if a project had a functional profile that was close to the average for the sample studied. For each sample, it was noted that there was one function type that had a stronger relationship with project effort.

Later, Abran and Panteliuc [3] studied the impact of the functional profile on the development effort for the projects measured by COSMIC. They concluded that the identification of the functional profile of a project and its comparison with the profiles of their own samples can help in selecting the best estimation models relevant to its own functional profile.

In [22], Gencel proposed using COSMIC functional size as a vector of measures when being used as an input to effort estimation models. By conducting case studies, it was explored whether the productivity values for developing different functionality types deviate significantly from a total average productivity value computed from total functional size and effort figures. The study concluded that the deviations are significant.

Gencel and Buglione [11][23] investigated whether effort estimation models based on the sizes of BFC Types in COSMIC rather than those based on a single total functional size value would improve the strength of the relationship between functional size and effort. They performed multiple-regression analyses on different sub-datasets formed considering different cost factors in the International Benchmarking Standards Group (ISBSG)<sup>1</sup> dataset. Both of the studies showed significant improvements in modeling the size-effort relationship. However, statistical analyses provided neither consistent weights for the BFC Types nor which of the BFC Types are significant in explaining the effort. The authors stated that the reason might have been the high heterogeneity of the projects in the ISBSG dataset.

In [20], Ferrucci *et al.* investigated the idea proposed in these two previous studies utilizing a rather homogeneous projects dataset consisting of 15 Web applications collected from a software organization.

The empirical study we present in this paper is a continuation work of the one described in [20]. In particular, other 10 projects were considered and analyzed together with the previous 15 projects in order to further assess whether findings

---

<sup>1</sup> URL: <http://www.isbsg.org>

## Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

highlighted in the previous studies were still valid and also answer some more research questions.

### 3 Case Study

Our research questions (RQs) for this empirical study were the following:

RQ1: Which BFC types are significant in explaining the variation in development effort?

RQ2: Do the functional sizes of each of the BFC types, when used as the main inputs instead of the total functional size, increase the reliability of the estimations?

RQ3: Is there a correlation between the contribution of the functional sizes of the BFC Types to the total functional size and the significance of each in estimating the development effort?

The details of the study are discussed in the following sub-sections.

#### 3.1 The Case Organization

The case organization is an Italian software company whose core business is the development of enterprise information systems, mainly for local and central government. Among its clients, there are health organizations, research centres, industry, and other public institutions. The company has about 50 employees and is specialized in the design, development, and management of solutions for Web portals, enterprise intranet/extranet applications (such as Content Management Systems, e-commerce, work-flow management systems) and geographical information systems. It is ISO 9001:2000 certified and also a certified partner of Microsoft, Oracle, and ESRI.

#### 3.2 Data Collection

We collected data on 25 Web application projects, including e-Government, e-Banking, Web portals, and Intranet applications. Web-oriented technologies such as J2EE, ASP.NET were used in development. Oracle was the commonly adopted Data Base Management System (DBMS). SQL Server, Access and MySQL were also used in some applications.

For data collection, timesheets were used. Each team member entered his/her development effort daily. At the end of each week, the project managers collected the total effort information and recorded this figure as the sum for the team. Here, the development effort refers to the total sum of all the efforts each

member spent for the requirements analysis, design, implementation, and testing phases.

The functional sizes of all 25 Web applications were measured by the project managers using COSMIC v2.2 [14]<sup>2</sup>. A template was developed by one of the authors of this study to collect the measurement details. COSMIC method requires four kinds of BFC Types; Entry, Exit, Read, and Write, to be reported when documenting the measurement details as well as the total functional size of the software in COSMIC Function Points (CFP) [13]. The project managers were trained on how to use this form. The provided information were reviewed by one of the authors, cross-checking the filled templates, the requirements specification and design documents.

Table 1 reports the descriptive statistics for the collected projects. Unfortunately, the measurement details could not be provided due to confidentiality reasons. The development effort (EFH) is given in person-hours and the COSMIC functional size in CFP.

**Table 1: Descriptive statistics of the whole dataset**

VAR	OBS	MIN	MAX	MEAN	MEDIAN	STD.DEV
EFH (person-hrs)	25	782	4537	2577	2686	988.136
Functional Size (CFP)	25	163	1090	602	611	268.473
<b>Base Functional Component (BFC) Types – COSMIC</b>						
Number of Entries ( <b>E</b> )	25	31	227	121.7	122	57.071
Number of Exits ( <b>X</b> )	25	27	316	122.3	110	71.985
Number of Reads ( <b>R</b> )	25	90	607	328.8	351	136.039
Number of Writes ( <b>W</b> )	25	0	120	29.2	20	31.859

### 3.3 Data Analysis

In order to investigate the variation in the contribution of different functionality types represented by BFC types with the total development effort, we first analyzed the type of the applications to form as homogeneous subsets of projects as possible before the statistical analysis.

Together with the project managers of the case projects, we classified the applications using the CHAR Method described in ISO/IEC TR 14143-5 [29]. 14 applications were categorized into “Information System” domain (in the following referred as *Subset1*). The other 11 projects were categorized into “Data Processing System” domain (in the following referred as *Subset2*).

<sup>2</sup> Current version of the Measurement Manual is v3.0.1, released on May 2009 [13].

## Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

*Table 2* shows the descriptive statistics for the projects in Subset1 and Subset2.

Our main research question for this study was whether effort estimation models based on the sizes of COSMIC BFC Types rather than those based on a single total functional size value would improve effort estimation. To this end, we first investigated the strength of the relationship between the total functional size and the total development effort. Then, we investigated the strength of the relationship between the sizes of the BFC Types and the total development effort. In particular, we used Multiple Linear Regression (MLR) [40] and Manual Stepwise Regression (MWSR) [33] which allows us to compute linear regression analysis in steps.

**Table 2: Descriptive statistics of Subset1 and Susbset2<sup>3</sup>**

VAR	Obs	Min	Max	Mean	median	Std.dev
<b>Subset1</b>						
EFH (person-hrs)	14	1176	4002	2526	2645	925.441
Functional Size (CFP)	14	264	986	614.4	617	235.673
<b>Base Functional Components (BFC) Types</b>						
Number of Entries ( <b>E</b> )	14	31	227	118.6	121.5	56.662
Number of Exits ( <b>X</b> )	14	29	225	123.36	112	63.672
Number of Reads ( <b>R</b> )	14	169	535	346.7	361.5	117.066
Number of Writes ( <b>W</b> )	14	0	105	25.79	18	32.418
<b>Subset2</b>						
EFH (person-hrs)	11	782	4537	2642	2686	1105.277
Functional Size (CFP)	11	163	1090	586.3	515	316.692
<b>Base Functional Components (BFC) Types</b>						
Number of Entries ( <b>E</b> )	11	37	224	125.6	122	60.104
Number of Exits ( <b>X</b> )	11	27	316	121	103	84.633
Number of Reads ( <b>R</b> )	11	90	607	306.1	265	159.948
Number of Writes ( <b>W</b> )	11	9	120	33.55	22	32.135

MLR is one of the most commonly used statistical techniques for exploring the relationship between a dependent variable and one or more independent variables, providing a prediction model described by an equation [36]:

---

<sup>3</sup> For confidentiality reasons, it is not possible to provide details about the functional sizes of the projects and their efforts.

$$(1) \quad y = b_1x_1 + b_2x_2 + \dots + b_nx_n + c$$

where  $y$  is the dependent variable (the effort),  $x_1, x_2, \dots, x_n$  are the independent variables (the cost drivers) with coefficient  $b_i$ , and  $c$  is the intercept. We used MLR for linear regression analysis. EFH (representing the total development effort in person-hours) was the dependent variable and E, X, R, and W (denoting the functional sizes of the BFC types) were the independent variables. Then, we made Linear Regression analysis considering the total functional size of the software as the independent variable.

When making Stepwise Regression (SWR), the estimation model is obtained by adding, at each step, the independent variable with the highest association to the dependent variable, taking into account all the variables currently in the model. SWR aims to find the set of independent variables that better explain the variation in the dependent variable. To select the variables to be added in the model a Manual SWR (MSWR) was applied, using the technique proposed by Kitchenham in [33]. The idea underlying this procedure is to select the important independent variables, and then to use linear regression to obtain the final model.

To evaluate the goodness of fit of a regression model, several indicators were considered. Among them, the square of the linear correlation coefficient,  $R^2$ , shows the amount of the variance of the dependent variable explained by the model related to the independent variable. Other useful indicators are the F value and the corresponding p-value (denoted by Sign F). Their high and low values, respectively, denote high degree of confidence for the prediction. We also considered the p-values and t-values for the corresponding coefficients and the intercept. The p-values give an insight into the accuracy of the coefficients and the intercept, whereas their t-values allow evaluating their importance for the generated model. In particular, p-values less than 0.05 are considered an acceptable threshold, meaning that the variables are significant predictors with a confidence of 5%. As for the t-value, a variable is significant if its corresponding value is greater than 1.5.

Whenever variables are highly skewed they should be transformed before being used in the MLR and MSWR procedure [36]. The residuals should be independent and normally distributed and the relationship between the dependent and the independent variables should be linear. A widely used transformation is the one considering the natural log ( $Ln$ ), which makes larger values smaller and brings the data values closer to each other [36]. In addition, whenever a variable needed to be transformed but had zero values, the natural logarithmic transformation was applied to the variables value after adding 1 as done in [37].



## Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

Moreover, the stability of each model built using MLR and MSWR should be verified. In order to accomplish this step, we used a residual plot showing residuals vs. fitted values to investigate if the residuals are randomly and normally distributed [36]. Then, Cook's distance values were used to identify the influential data points. As suggested in [37] each observation having distances higher than  $3 \times (4/n)$ , where  $n$  represents the total number of projects, are immediately removed from the data analysis. On the other hand, the observations having distances higher than  $4/n$  but smaller than  $(3 \times (4/n))$  are removed and the stability of model is tested by analyzing the effect of their removal on the model. If the goodness of fit improves (i.e., a higher  $R^2$  is obtained and the coefficients of the model remain stable) then the influential observations are not excluded from the analysis.

Moreover, in order to validate the obtained effort estimation models (i.e., to verify whether or not the predicted efforts were useful estimations of the actual development efforts) we exploited a leave-1-out cross validation. It is widely used in empirical studies when dealing with small datasets [6]. In particular, to apply the technique, the original dataset is divided into  $N$  different subsets (where  $N$  is the size of the original dataset) of training and validation sets, where each validation set has one observation. Then,  $N$  steps are performed and at each step, the training set is used to determine the estimation model and the validation set to assess the obtained estimates.

To evaluate the accuracy of the obtained estimations, we used some summary measures such as MMRE, MdMRE and Pred(0.25) [15], which have been widely used to assess the accuracy of software estimation models in empirical studies (see e.g., [8][16][18][19][31][34][37][38][44]). In the following, we briefly recall the main concepts underlying MMRE and Pred(0.25). The Magnitude of Relative Error [15] is defined as:

$$(2) \quad \text{MRE} = |\text{EFH}_{\text{real}} - \text{EFH}_{\text{pred}}| / \text{EFH}_{\text{real}}$$

where  $\text{EFH}_{\text{real}}$  and  $\text{EFH}_{\text{pred}}$  are the actual and the predicted efforts, respectively. MRE has to be calculated for each observation in the dataset. All the MRE values are aggregated across all the observations using the mean and the median, giving rise to the Mean of MRE (MMRE) and the Median MRE (MdMRE).

The prediction at level 0.25 [15]<sup>4</sup>, defined as:

$$(3) \quad \text{Pred}(0.25) = k / N$$

where  $k$  is the number of observations whose MRE is less than or equal to 0.25, and  $N$  is the total number of observations.  $\text{Pred}(0.25)$  is a quantification of the percentage of predictions whose error is less than 25%. According to [15], a good effort estimation model should have a  $\text{MMRE} \leq 0.25$  and  $\text{Pred}(0.25) \geq 0.75$ , that is, the mean estimation error should be less than 25%, and at least 75% of the estimated values should fall within 25% of their actual values.

Moreover, we tested the statistical significance of the obtained results by using absolute residuals in order to establish if one estimation model provided better results than others [35]. In particular, we performed statistical tests (the T-Test and the Wilcoxon test) to verify the following null hypothesis “the two considered population have identical distributions”. If the null hypothesis is true, then the number of positive and negative differences should be approximately the same.

## 4 Results

### 4.1 The Relationship between EFH and COSMIC Functional Size

In order to apply the Linear Regression analysis we verified the following assumptions for each training set: *linearity* (i.e., the existence of a linear relationship between the independent variable and the dependent variable); *homoscedasticity* (i.e., the constant variance of the error terms for all the values of the independent variable) and *residual normality* (i.e., the normal distribution of the error terms).

Table 3 presents the results of the linear regression analysis with statistics on useful indicators to verify the quality of the obtained models.

---

<sup>4</sup> A particular attention must be paid to the 25% threshold value proposed in [15]. That book – often referenced – was written in 1986 and 25 years have been passed by the date of writing. Thus, nowadays error estimation thresholds in ICT projects have been lowered and are more challenging, but there is not a standard figure. Therefore, the suggestion is to apply the same concepts and procedure presented with the typical thresholds suggested/applied in your own organization using your historical data.

Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

**Table 3:** The results of the Linear Regression analysis using the total functional size

*Whole Dataset*

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
CFP	3.429	0.279	12.302	1.34e-11	0.862	366.6	151.3	1.341e-11
Intercept	512.430	183.137	2.798	0.0102				

*Subset1*

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
CFP	3.679	0.397	9.273	8.04e-07	0.867	337.1	85.99	8.045e-07
Intercept	265.466	259.856	1.022	0.327				

*Subset2*

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
CFP	3.287	0.391	8.397	1.5e-05	0.874	392	70.51	1.500e-05
Intercept	715.413	258.108	2.772	0.0217				

We observed that the linear regression analysis was successfully applied to the whole dataset. For the three considered datasets the obtained models are characterized by a high R<sup>2</sup> value. Indeed, c.a. 86% is the amount of the variance of the dependent variable EFH that is explained by the model related to the variable CFP for all the three considered sets. Furthermore, for the whole dataset a high F value (151.3) and a low p-value (1.341e-11) are obtained, indicating that the prediction is available with a high degree of confidence.

The t-values and p-values for the corresponding coefficient and the intercept present values greater than 1.5 and less than 0.05, respectively. Therefore, the variable is a significant predictor with a confidence of 5%. As for the Subset1 and Subset2, the results of the performed analysis denote a high degree of confidence for the prediction. Indeed, the two obtained prediction models are characterized by high F values (85.99 and 70.51) and low p-values (8.045e-07 and 1.500e-05).

Concerning t-statistic for Subset2, the analysis showed that the intercept is characterized by a p-value less than 0.05 and a t-value greater than 1.5. On the other hand, the analysis for Subset1 revealed that the intercept is characterized by a p-value greater than 0.05 and a t-value less than 1.5. As for the comparison between the obtained prediction models, we can also note that the R<sup>2</sup> obtained for the three sets are very close.

To evaluate the accuracy of the estimates we employed the summary measures MMRE, MdmRE, and Pred(0.25). Table 4 presents the results, which highly fit

the acceptable threshold defined in [15] since MMRE (and MdMRE) values are less than 0.25 and Pred(0.25) values are greater than 0.75. Thus, we conclude that the total functional size obtained with COSMIC measurement method can be reliably used to estimate development effort for the Web applications included in our dataset.

**Table 4:** Accuracy of estimates obtained with the total functional size (CFP)

Dataset	MMRE	MdMRE	Pred(25)
Whole dataset	0.13	0.07	0.88
Subset1	0.11	0.08	0.93
Subset2	0.19	0.11	0.82

#### 4.2 The Relationship between EFH and the sizes of COSMIC BFC types

Table 5 and Table 6 show the results of the MLR and MSWR using the sizes of the BFC Types as independent variables and the total development effort as the dependent variable.

It is worth noting that the independent variable W (Write) was highly skewed as revealed by the Shapiro test (p-value = 7.503e-05 for the whole dataset, p-value = 0.002 for Subset1, and p-value = 4.403e-4 for Subset2). Thus, the variable was transformed to comply with the assumptions underlying Linear Regression [36], by applying the natural log and considering the new variable LnWrite (LnW).

From the results of the MLR analysis, we observed that all the three obtained models are characterized by high  $R^2$  (see Table 5). Indeed, c.a. 84%, 84%, and 87% is the amount of the variance of the dependent variable EFH that is explained by the model related to the variables R, LnW, X, and E for the whole dataset, Subset1, and Subset2, respectively.

However, for the first two models (i.e., those obtained from the whole dataset and Subset1) only the independent variable R (Read) is characterized by a p-value less than 0.05 and a t-value greater than 1.5. Thus, the other variables cannot be considered significant predictors. Regarding Subset2, only the independent variable X (eXit) can be considered a significant predictor since it has a t-value greater than 1.5 and a p-value very close to 0.05.

We want to highlight that the analysis we performed (and the results we obtained) with MLR on Subset1 and Subset2 has to be considered with caution since we built the two models on a small number of observations which is very sensitive to relatively small variations in the observations used when building

Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

---

the models. This was a further reason motivating us to also apply MSWR in our empirical study.

**Table 5:** The results of the MLR analysis using single BFC types of COSMIC (i.e., E, X, R, W)

a) Whole dataset

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
Intercept	381.693	244.811	1.559	0.135	0.842	393	32.93	1.533e-08
<b>Read (R)</b>	<b>3.279</b>	<b>1.093</b>	<b>3.000</b>	<b>0.007</b>				
LnWrite (lnW)	97.453	67.071	1.453	0.162				
Entry (E)	2.283	3.177	0.719	0.481				
Exit (X)	4.687	2.373	1.975	0.062				

b) Subset1

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
Intercept	53.971	325.337	0.161	0.876	0.839	371.7	17.89	2.604e-4
<b>Read (R)</b>	<b>5.464</b>	<b>1.536</b>	<b>3.557</b>	<b>0.006</b>				
LnWrite (lnW)	36.746	80.421	0.457	0.659				
Entry (E)	-0.118	4.605	-0.026	0.980				
Exit (X)	4.108	3.806	7.079	0.308				

c) Subset2

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
Intercept	1909.236	1021.887	1.868	0.111	0.869	400.3	17.56	0.002
Read (R)	0.543	2.391	0.227	0.828				
LnWrite (lnW)	-512.407	439.242	-1.167	0.288				
Entry (E)	7.015	6.821	1.029	0.343				
<b>Exit (X)</b>	<b>11.052</b>	<b>4.601</b>	<b>2.402</b>	<b>0.053</b>				

The results of the MSWR procedure revealed that the best fitting model for two of the three datasets (the whole dataset and Subset1) identifies R (Read) as the preeminent effort predictor. This suggests that most of the total development effort is devoted to R (Read) data movements.

The R<sup>2</sup> values show that c.a. 77% and 85% is the amount of the variance of the dependent variable EFH that is explained by the model related to the variables R (Read) for the whole dataset and Subset1, respectively. For both models the

independent variable R (Read) can be considered a significant predictor as revealed by the corresponding p-value and t-value.

However, the intercept of the model obtained with Subset1 is characterized by p-value > 0.05 and a p-value < 1.5. As for Subset2, the best fitting model was obtained by employing only the independent variable X (eXit), suggesting that most of the total development effort is devoted to X (eXit) data movements. In this case, the R<sup>2</sup> value showed that c.a. 85% is the amount of the variance of the dependent variable EFH that is explained by the model related to the variable X (eXit).

Furthermore, the t-statistic revealed that both the independent variable X (eXit) and the intercept can be considered significant as revealed by the corresponding p-value and t-value. It is worth noting that the rule of thumb<sup>5</sup> recalled above is satisfied since the obtained models employ only one variable and they are built using a number of observations greater than 5.

**Table 6:** The results of the MSWR analysis using single BFC types of COSMIC

*a) Whole dataset*

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
R	6.380	0.724	8.864	7.83e-09	0.772	482.4	77.68	7.828e-09
Intercept	478.874	256.860	1.864	0.075				

*b) Subset 1*

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
R	7.283	0.887	8.209	2.89e-06	0.849	374.5	67.38	2.885e-06
Intercept	0.400	323.499	0.001	1				

*c) Subset2*

	Value	Std. Err	t-value	p-value	R <sup>2</sup>	Std Err	F	Sign F
X	12.028	1.696	7.093	5.71e-05	0.848	453.8	50.32	5.708e-05
Intercept	1186.876	246.621	4.813	0.001				

Table 7 shows the results of the performed accuracy evaluation. We observed that MMRE and MmMRE highly fit the acceptable threshold as defined in [15] since the corresponding values are less than 0.25 for all the three sets. As for Pred(25) the results suggest that the estimates obtained for Subset1 fit the

---

<sup>5</sup> “A rule of thumb in regression analysis is that 5 to 10 records are required for every variable in the model” [39].

## Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

acceptable threshold defined in [15] since the corresponding value is greater than 0.75.

On the other hand, the estimates obtained with Subset2 are characterized by a Pred(25) value less than 0.75, while the Pred(25) value obtained for the whole dataset is 0.72 (very close to 0.75). Thus, we conclude that the estimation models based on the independent variables R (Read) can provide reliable estimates according to the thresholds provided by Conte *et al.* [15].

**Table 7: Accuracy of estimates obtained with MSWR and single BFC types of COSMIC**

Dataset	Employed predictors	MMRE	MdMRE	Pred(25)
Whole dataset	Read (R)	0.17	0.11	0.72
Subset1	Read (R)	0.14	0.12	0.79
Subset2	X (eXit)	0.24	0.11	0.64

### 4.3 Discussion

Table 8 summarizes the results we obtained by applying MLR and MSWR, when using COSMIC as size measure. The results show that the  $R^2$  achieved employing the total functional size (i.e., CFP) are slightly higher than those obtained using single BFC types (i.e., E, X, R, and LnW). Furthermore, we observe that even if the  $R^2$  obtained applying MLR on E, X, R, and LnW are very close to the ones obtained with CFP, three of the predictors (i.e., E, X, and LnW for the whole dataset and Subset1, and E, R, LnW for Subset2) are not significant.

**Table 8: Comparison of the results using COSMIC as size measure**

Dataset	# of obs	Predictors	$R^2$	Significance of predictors	Accuracy (satisfy Conte <i>et al.</i> threshold)
Whole dataset	25	CFP (Total Functional Size)	0.862	Yes	Yes
	25	R, LnW, X, E,	0.842	No	-
	25	R	0.772	Yes	Yes
Subset1	14	CFP (Total Functional Size)	0.867	Yes	Yes
	14	R, LnW, X, E	0.839	No	-
	14	R	0.849	Yes	Yes
Subset2	11	CFP (Total Functional Size)	0.874	Yes	Yes
	11	R, LnW, X, E	0.869	No	-
	11	X	0.848	Yes	Yes (except Pred(25))

The interesting result was the application of MSWR, which suggested that R (X for Subset2, respectively) can be considered a significant predictor for the whole dataset and Subset1 (for Subset2, respectively). Moreover, for Subset1 the model employing only R (Read) is characterized by a  $R^2$  value close to the one of the model employing the total functional size (CFP) and for Subset2, the model employing only X (eXit) shows similar results.

Regarding the estimation accuracy evaluation, the comparison of results in terms of MMRE, MdmRE, and Pred(25) reported in Table 4 and Table 7 suggest that the total functional size obtained by COSMIC measurement method (i.e., CFP) allows us to obtain better estimates than those obtained using single BFC types (i.e., Read and eXit). Moreover, as suggested in [35], we also tested the statistical significance of the obtained results by using absolute residuals. To this end, we performed the Wilcoxon test. The analysis revealed that the absolute residuals obtained with CFP were significantly better than those obtained using R (Read) in the case of the whole dataset (p-value=0.016). On the other hand, for Subset1 there was no significant difference between the absolute residuals obtained using CFP and those achieved using only Read (p-value=0.196). A similar result was obtained for the estimation model employing only the variable Exit in the case of Subset2.

Interestingly, these results are in line with our previous findings for the whole dataset and two subsets [20]. The same BFC Types were found to be significant in explaining the variation in the effort. These might suggest that for these kinds of Web applications in this case company, using the size of only one of the COSMIC BFC Types to estimate the development effort might be a promising option when a quick estimate is required. However, this requires further investigation by collecting more projects data.

A final observation is the possible correlation between the contribution of BFCs to the total functional size and the BFCs, which are found to be significant in estimating development effort. Table 9 shows the distribution of the BFCs with respect to the three datasets considered in our empirical analysis when using COSMIC.

**Table 9:** *The distribution of BFC types for COSMIC*

Dataset	Entry (E)	Exit (X)	Read (R)	Write (W)
Whole dataset	20.32%	19.40%	55.77%	4.50%
Subset1	18.92%	19.42%	57.99%	3.67%
Subset2	22.10%	19.38%	52.95%	5.56%



## Which COSMIC Base Functional Components are Significant in Estimating Web Application Development? - An Empirical Study

The relative distributions of the BFC Types in all three datasets are very similar. The contribution of R (Read) to the total functional size is the greatest for all the three datasets (in particular, R provides more than 52% of the total functional size), while W (Write) provides the smallest contribution. Thus, in contrast to the results obtained in [11][23] with the ISBSG dataset [27], we note a correlation between the amount of contribution of BFCs to the total functional size and the BFCs significant in estimating development effort for the Whole Dataset and Subset1. However, for Subset2, although it has a similar percentage distribution of R (Read), the significant BFC Type for effort estimation was found to be X (eXit).

### 5 Conclusions & Prospects

This paper investigates whether considering the COSMIC BFC Types rather than the total functional size improves effort estimation accuracy. Using a sample of 25 Web-based projects collected from an Italian company, we found that the size of one of the BFC Types is as good as using the total functional size as the main input in explaining the variation in effort. This suggests that for Web applications, using the size of only one of the COSMIC BFC Types to estimate the development effort might be a promising option when an early and quick estimate is required. However, this requires further investigation through more empirical studies.

As future work, we plan to investigate the hypothesis of this study using different datasets, considering different cost drivers, in particular the application and organizational domains as well as the primary programming language.

### 6 References

- [1] Abran, A., Ndiaye, I., Bourque, P., "Contribution of Software Size in Effort Estimation", research Lab. In Software Engineering, Ecole de technologies Superieure, Canada, (2003)
- [2] Abran A., Gil B., Lefebvre E.: Estimation Models Based on Functional Profiles. International Workshop on Software Measurement -- IWSM/MetriKon, Kronisburg (Germany), Shaker Verlag, (2004), 195-211, URL : <http://publicationslist.org/data/a.abran/ref-2258/831.pdf>
- [3] Abran, A., Panteliuc, A., Estimation Models Based on Functional Profiles. III Taller Internacional de Calidad en Tecnologias de Information et de Communications, Cuba, February 15-16 (2007), URL : <http://publicationslist.org/data/a.abran/ref-2079/1047.pdf>
- [4] Albrecht, A. J.: Measuring application development productivity. In: Proceedings of the IBM Applications Development Symposium, pp. 83--92. Monterey, California (1979)

- [5] Boehm, B.W., E. Horowitz, R. Madachy, D. Reifer, B. K.. Clark, B. Steece, A. W. Brown, S. Chulani, C. Abts, “Software Cost Estimation with Cocomo II”, Prentice-Hall, 2000, ISBN 9780130266927.
- [6] Briand L.C., Emam K.E., Bomarius F., “COBRA: a hybrid method for software cost estimation, benchmarking, and risk assessment”, in Proceedings of the international conference on Software engineering, IEEE Computer Society, 1998, pp. 390–399.
- [7] Briand, L., Wieczorek, I., “Software Resource Estimation”, Encyclopedia of Software Engineering (2002), Volume 2. P-Z (2nd ed.), Marciniak, John J. (ed.) New York: John Wiley & Sons pp. 1160–1196.
- [8] Briand, L., T. Langley, I. Wieczorek, “A Replicated Assessment and Comparison of Common Software Cost Modeling Techniques”, *Proceedings of International Conference on Software Engineering*, IEEE press, 2000, pp. 377–386.
- [9] Buglione L., Some thoughts on Productivity in ICT projects, WP-2010-01, White Paper, version 1.3, August 2010, URL: [www.semq.eu/pdf/fsm-prod.pdf](http://www.semq.eu/pdf/fsm-prod.pdf)
- [10] Buglione L. & Abran A., “A Model for Performance Management & Estimation”, *Proceedings of 11<sup>th</sup> IEEE International Software Metrics Symposium*, p.9. ISBN 0-7695-2371-4
- [11] Buglione, L., Gencel, C., “Impact of Base Functional Component Types on Software Functional Size Based Effort Estimation”, *Proceedings of PROFES*, 2008, pp. 75-89
- [12] Buglione L., Gencel C., “The Significance of IFPUG Base Functionality Types in Effort Estimation: An Empirical Study”, IFPUG ISMA5 (5<sup>th</sup> International Software Measurement and Analysis) Conference, Sao Paulo (Brazil), 13-15 September 2010.
- [13] COSMIC, COSMIC v.3.0.1, Measurement Manual, 2009, Common Software Measurement International Consortium, URL: [www.cosmicon.com](http://www.cosmicon.com)
- [14] COSMIC, COSMIC Full Function Points v.2.2, Measurement Manual, 2004, Common Software Measurement International Consortium, URL: [www.cosmicon.com](http://www.cosmicon.com)
- [15] Conte D., Dunsmore H.E., Shen V.Y., “Software engineering metrics and models”, The Benjamin/Cummings Publishing Company, Inc., 1986.
- [16] Costagliola G., Di Martino S., Ferrucci F., Gravino C., Tortora G., Vitiello G., “A COSMIC-FFP: Approach to Predict Web Application Development Effort”, *J. of Web Engineering* 5(2), 2006, pp. 93-120.
- [17] Déry, Abran A., “Investigation of the effort data consistency in the ISBSG repository”, *Proceedings of the 15th Intern. Workshop on Software Measurement (IWSM 2005)*, 2005, pp. 123-136, Shaker Verlag, ISBN 3-8322-4405-0.
- [18] Ferrucci F., Gravino C., Di Martino S., “A Case Study Using Web Objects and COSMIC for Effort Estimation of Web Applications”, *Proceedings of the 34<sup>th</sup> Euromicro Conference / Software Engineering and Advanced Applications (SEEA 2008)*, 2008, pp. 441-448. ISBN 978-0-7695-3276-9.
- [19] Ferrucci, F., Gravino, C., Di Martino, S., “Estimating Web Application Development Effort Using Web-COBRA and COSMIC: An Empirical Study”,

Which COSMIC Base Functional Components are Significant in Estimating  
Web Application Development? - An Empirical Study

*Proceedings of the 35<sup>th</sup> Euromicro Conference / Software Engineering and Advanced Applications* (SEEA 2009), pp. 306-312.

- [20] Ferrucci, F., Gravino, C., Buglione, L., Estimating Web Application Development Effort using COSMIC: Impact of the Base Functional Component Types, Proceedings of the 7<sup>th</sup> Software Measurement European Forum (SMEF 2010), Rome (Italy) June 10-11 2010, ISBN 978-88-6301-033-6, pp. 103-116, URL: [www.dpo.it/smef2010](http://www.dpo.it/smef2010)
- [21] Gencel, C., and Buglione, L.: Do Different Functionality Types Affect the Relationship between Software Functional Size and Effort?, J.J. Cuadrado-Gallego et al. (Eds.): MENSURA 2007, LNCS 4895, Springer-Verlag Berlin Heidelberg (2008), 72–85.
- [22] Gencel, C., “How to Use COSMIC Functional Size in Effort Estimation Models?, IWSM / MetriKon / Mensura 2008, LNCS 5338, 2008, pp. 205–216.
- [23] Gencel, C., Buglione, L., “Do Base Functional Component Types Affect the Relationship between Software Functional Size and Effort?”, *Proceedings of IWSM/Mensura 2007*, pp. 72-85.
- [24] Gencel, C., Demirors, O.: Functional Size Measurement Revisited. In: ACM Trans. on Software Eng. and Meth. (TOSEM), Vol.17, No.3, pp. 71-106. (2008)
- [25] IFPUG, Function Point CPM, Release. 4.2, IFPUG, Westerville, OH, 1999. [www.ifpug.org](http://www.ifpug.org)
- [26] Jørgensen, M., Shepperd, M., “A Systematic Review of Software Development Cost Estimation Studies”, IEEE Transactions on Software Engineering, Vol. 33, Issue 1, Jan. 2007, 33-53.
- [27] ISBSG, *ISBSG Data Repository r10*, January 2007, URL: [www.isbsg.org](http://www.isbsg.org)
- [28] ISO/IEC, IS 14143-1:2007, Information Technology – Software Measurement – Functional Size Measurement – Part 1: Definition of Concepts, February 2007
- [29] ISO/IEC TR 14143-5: Information Technology- Software Measurement - FSM - Part 5: Determination of Functional Domains for Use with Functional Size Measurement, 2004.
- [30] IT Performance Committee, Non-Functional Size Measure, Presentation, IFPUG ISMA3 Conference, September 2008, URL: <http://www.ifpug.org/Webforum/discus/board-auth.cgi?lm=1226921110&file=/10145/10151.html>
- [31] Jeffery, D. R., M. Ruhe, I. Wiczorek, ”A comparative study of two software development cost modeling techniques using multi-organizational and company-specific data”, *Information & Software Technology* 42(14), pp. 1009-1016 (2000)
- [32] Jørgensen, M., K. Moløkken-Østfold, “Reasons for Software Effort Estimation Error: Impact of Respondent Role, Information Collection Approach, and Data Analysis Method”, *IEEE Trans. Software Eng.* 30(12), pp. 993-1007 (2004).
- [33] Kitchenham, B. A., “A Procedure for Analyzing Unbalanced Datasets”, *IEEE Trans. Software Eng.*, 24(4), 1998, 278-301.
- [34] Kitchenham, B. A., E. Mendes, “Software Productivity Measurement Using Multiple Size Measures”, *IEEE Trans. Software Eng.* 30(12), pp. 1023-1035 (2004)
- [35] Kitchenham B.A., Pickard L.M., MacDonell S.G., Shepperd M.J., “What accuracy statistics really measure”, *IEE Proc. – Software*, 148(3), 2001, pp. 81-85.

- [36] Maxwell, K. *Applied Statistics for Software Managers*. Software Quality Institute Series, Prentice Hall, 2002, ISBN 0130417890
- [37] Mendes E., Kitchenham B., “Further Comparison of Cross-company and Within-company Effort Estimation Models for Web Applications”, *Proceedings of International Software Metrics Symposium (METRICS’04)*, 2004, pp. 348-357.
- [38] Mendes E., Counsell S., Mosley N., Triggs C., Watson I., “A Comparative Study of Cost Estimation Models for Web Hypermedia Applications”, *Empirical Software Engineering* 8(2), 2003, pp. 163-196.
- [39] Menzies, T., Chen, Z., Hihn, J., Lum, K., “Selecting Best Practices for Effort Estimation”, *IEEE Trans. Software Eng.*, 32(11), November 2006, pp. 883-895.
- [40] Montgomery D., Peck E., Vining, “Introduction to Linear Regression Analysis”, John Wiley and Sons, Inc., 1986.
- [41] Reifer D., “Web-Development: Estimating Quick-Time-to-Market Software”, *IEEE Software*, 17(8), 2000, pp. 57-64
- [42] Reifer D., *Web Objects Counting Conventions*, Reifer Consultants, Mar. 2001, URL <http://www.reifer.com/download.html>
- [43] Rollo T., “Sizing E-Commerce”, *Proceedings of the ACOSM 2000 - Australian Conference on Software Measurement*, 2000, URL: [http://www.gifpa.co.uk/library/Papers/Rollo/sizing\\_e-com/v2a.pdf](http://www.gifpa.co.uk/library/Papers/Rollo/sizing_e-com/v2a.pdf)
- [44] Ruhe, M., Jeffery, R., Wiczorek, I., “Cost estimation for Web applications”, *Proceedings of International Conference on Software Engineering (ICSE’03)*, 2003, pp. 285–294.