

A case study on the conversion of Function Points into COSMIC

Filomena Ferrucci, Carmine Gravino, Federica Sarro
School of Computer Science
University of Salerno
Via Ponte Don Melillo, I-84084, Fisciano (SA), Italy
{fferrucci, gravino, fsarro}@unisa.it

Abstract— Some empirical studies have been carried out so far to identify a function able to convert IFPUG functional size units to the COSMIC ones. In this paper, we report on a replication of those studies using a dataset of 25 Web applications. As for the estimation technique, we employed linear regression analysis; while k-fold cross validation was exploited to validate the estimation models. The results show that there is a significant correlation between the size expressed in terms of COSMIC and IFPUG, but differently from previous studies the conversion factor is not so close to 1.

Keywords-Functional size measurement; COSMIC; IFPUG

I. INTRODUCTION

Among the measurement approaches proposed in the literature Functional Size Measurement (FSM) is widely employed to measure the size of Users Functional Requirements [13]. It is an ISO standardized technique [16], largely adopted since it is independent from the non-functional technical and quality requirements [1]. Furthermore, several organizations work around the world to improve and promote the application of FSM approaches and tools (e.g., IFPUG [15] and COSMIC [5]). However, so far only a few studies have been performed to investigate the effectiveness of FSM methods for Web applications and in particular for development effort estimation. These studies have revealed that 1st generation of FSM methods (i.e., IFPUG Function Points Analysis [15]) may fail to capture some specific features affecting the effort required for Web applications (see e.g., [19]). About, 2nd generation methods, and in particular COSMIC, the preliminary studies conducted with Web applications seem to highlight interesting results (see e.g., [9][10]). Nevertheless, many companies have employed for years IFPUG Function Points Analysis (FPA) and have their IFPUG-based historical data for estimating and planning their software projects [7]. As a consequence, even if they could be interested in applying COSMIC they are prevented to this due to the additional effort and costs. Indeed, to apply data of past projects, these should be resized using COSMIC with costs that might be not sustainable. In order to facilitate this migration, some researchers have proposed to exploit in the first phase of this process the size estimates available in terms of FPA to automatically obtain the size estimates in terms of COSMIC. To this aim, some studies have been carried out to identify conversion equations [1][7][8][11][14]. Following this research direction, we have performed a case study on 25 Web applications of an Italian software company in order to verify

whether it is possible to obtain accurate size estimates in terms of COSMIC exploiting the available information on the size in terms of FPA. As for the estimation technique, we employed linear regression analysis since it has been applied in previous studies; while k-fold cross validation was exploited to validate the estimation models. Widely used summary measures (e.g., MMRE and Pred(25)) were employed as evaluation criteria.

The remainder of the paper is organized as follows. In Section II we present the related work. In Section II we describe the experimental method we exploited to perform the case study, while the results of the empirical analysis are reported and discussed in Section IV. Section V concludes the paper giving some final remarks and future work.

II. RELATED WORK

Few case studies have been conducted to identify conversion equations between FPA and COSMIC and the most relevant are summarized and discussed in [7]. In particular, Cuadrado-Gallego *et al.* performed a review of previous investigations that mainly exploited linear regression analysis to identify a sound mathematical basis for converting FPA measurement in COSMIC measurement, i.e., by constructing an equation as:

$$(1) \text{ CFP} = a + b \cdot \text{FP},$$

where the independent variable CFP represents the COSMIC measure and the independent variable FP represents the FPA measure. Furthermore, in their investigation they considered other three datasets and enriched the analysis by studying the interval variation of the parameters at 95% of confidence level. The results they obtained are reported in Table II, which shows for each case study the obtained parameters, the confidence interval for the coefficient of the obtained equation, and the coefficient of determination R^2 . Cuadrado-Gallego *et al.* observed that the analysis conducted on the further three datasets (i.e., the last three rows of Table I) suggested that in general a coefficient very close to 1 was obtained and the intersection of most probable intervals was (0.7 - 0.8). On the other hand, the previous studies (the first six rows of Table I) were also characterized by a coefficient around 1 but with a most probable interval of (1.1 - 1.2). As for R^2 , Table II suggests that a high value was obtained in all the studies (except for the first case study in [7]).

Starting from the observation that an important aspect when proving a conversion approach is that the equation must

consider the origin of coordinates, Cuadrado-Gallego *et al.* also analyzed a non linear relationship between FPA and COSMIC by exploiting the log transformation of the variables CFP and FP in the application of linear regression analysis. Thus, the equation obtained is of this form:

$$(2) \text{Log(CFP)} = \text{Log}(a) + b \cdot \text{Log}(FP)$$

which, when transformed back to the original raw data scale, gives the equation:

$$(3) \text{CFP} = a \cdot \text{FP}^b$$

The results obtained in [7] are reported in Table II and suggest that the equations have a similar behavior in the exponent (around 1) [7]) with a probable interval (0.9 – 1.1).

TABLE I. PARAMTERS OBTAINED USING LINEAR EQUATIONS

Study	b	a	Confidence Interval b	R ²
[11]	1.1	-6.2	(0.8 – 1.4)	0.98
[8]	1.1	-7.6	(0.8 – 1.4)	0.97
[14]	1	6.5	(0.6 – 1.3)	0.98
[1]	1.2	-86.8	(1.1 – 1.3)	0.99
[1]	0.8	18	(0.8 – 1.2)	0.91
[8]	1	-3.2	(0.8 – 1.2)	0.93
[7]-1	0.8	-36.6	(0.6 – 1.1)	0.7
[7]-2	0.9	0.2	(0.7 – 1.1)	0.86
[7]-3	0.7	-4.5	(0.7 – 1.8)	0.9

TABLE II. PARAMTERS OBTAINED USING NON LINEAR EQUATIONS

Study	b	a	Confidence Interval b	R ²
[11]	1.1	0.6	(0.9 – 1.4)	0.99
[8]	1.1	0.6	(0.5 – 1.7)	0.97
[14]	1.1	0.7	(0.9 – 1.4)	0.99
[1]	1.2	0.3	(0.9 – 1.4)	0.94
[1]	1.1	1.1	(0.5 – 1.4)	0.88
[8]	1.1	0.7	(0.9 – 1.2)	0.95
[7]-1	1.2	0.3	(0.9 – 1.4)	0.82
[7]-2	1	0.8	(0.6 – 1.4)	0.73
[7]-3	1	1	(0.9 – 1.1)	0.99

III. CASE STUDY PLANNING

This section presents the design of our empirical study.

A. Dataset

The empirical study is based on a dataset coming from a medium-sized software company operating in Italy, whose core business is the development of enterprise information systems, mainly for local and central government. The company is specialized in the development and management of solutions for Web portals, enterprise intranet/extranet applications (such as Content Management Systems, e-commerce, work-flow managers, etc...), and Geographical Information Systems. It has about fifty employees and a turnover of about 5M €. It is certified ISO 9001:2000, and it is also a certified partner of Microsoft, Oracle, and ESRI.

Data used in the study are related to a set of 25 Web applications, including e-government, e-banking, Web portals,

and Intranet applications, developed with different Web-oriented technologies (e.g., J2EE, ASP.NET). Oracle has been the most commonly adopted DBMS, but also SQL Server, Access and MySQL were employed in some applications. These applications were developed between 2003 and 2008 and the information on them has been obtained during a long term investigation aimed at verifying the effectiveness of COSMIC, FPA, and Web Objects in estimating Web applications development effort (see [9][10]). Table III provides some descriptive statistics, where CFP denotes the size expressed in terms of COSMIC and FP in terms of FPA.

TABLE III. DESCRIPTIVE STATISTICS

Variable	Obs	Min	Max	Median	Mean	Std. Dev.
CFP	25	163	1,090	611	602	268.5
FP	25	110	973	336	400	216.4

B. Employed technique

Ordinary Least-Squares Regression (OLSR) is a statistical technique that explores the relationship between a dependent variable and one or more independent variables [17]. In our empirical study we exploited simple linear regression to obtain a model as the one shown in equation (1), where CFP is the dependent variable and FP is the independent variable. To evaluate the goodness of fit of a regression model, several indicators have to be considered. Among them, the square of the linear correlation coefficient, R², 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), which high and low values, respectively, denote a high degree of confidence for the prediction. We have 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 us to evaluate 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.

Moreover, observe that as highlighted in [7] an important aspect when proving a conversion approach is that if the size measured in terms of FP is zero then also the size in terms of CFP have to be zero since this means no functional size. Thus, similarly to the study presented in [7], we also carried out an analysis by first exploiting a log transformation of the original data and then applying linear regression analysis. This strategy is also usually employed whenever variables are highly skewed and they are transformed before applying linear regression analysis. This is done in order to comply with the assumptions underlying linear regression [17] (i.e., residuals should be independent and normally distributed; relationship between dependent and independent variables should be linear).

C. Validation method and evaluation criteria

In our analysis, we carried out a validation to verify whether or not the predicted size obtained by exploiting the FP based model is a useful estimation of the observed size

measured in terms of CFP. To this end, we applied a cross-validation splitting a dataset into training and validation sets. Training sets are used to build models with OLSR and validation sets are used to validate the obtained models. In particular, we exploited a leave-one-out cross validation, which means that the original dataset is divided into $n=25$ different subsets (25 is the size of the original dataset) of training and validation sets, where each validation set has one project. This validation method is widely used in empirical studies when dealing with small datasets [3].

Regarding the evaluation criteria, we used three widely used summary measures, namely MMRE, MdmRE, and Pred(25) (see definitions in [4]), to assess the CFP estimations achieved using the FP based model. According to [4], a good prediction model should have an $MMRE \leq 0.25$, to denote that the mean estimation error should be less than 25%, and a $Pred(25) \leq 0.75$, meaning that at least 75% of the predicted values should fall within 25% of their actual values. Since we applied a $k(=25)$ fold cross validation we also considered the k -fold regression coefficient Q^2 that resulted in optimum number of components and lowest standard error of prediction [7]:

$$(4) Q^2 = 1 - (\sum(\text{Size}_{\text{predicted}} - \text{Size}_{\text{observed}})^2) / (\sum(\text{Size}_{\text{observed}} - \text{Size}_{\text{mean}})^2)$$

where $\text{Size}_{\text{mean}}$ is the mean value of the size in terms of COSMIC, respectively.

IV. RESULTS AND DISCUSSION

The following subsections present the results of the empirical study and discuss case study validity.

A. Results with FP and CFP

In order to apply OLSR we first verified the OLSR assumptions, whose analysis is reported in the following. Note that the results are intended as statistically significant at $\alpha=0.05$ (i.e., 95% confidence level).

Linearity. The linear relationship between CFP and FP was revealed by the Pearson's correlation test (statistic = 7.36 with $p\text{-value} \leq 0.01$) [12]. This result suggested that there was a significant correlation between size expressed in terms of CFP and FP, also encouraging to find a mathematical conversation among the two functional size units¹.

Homoscedasticity. We investigated this assumption by performing a Breush-Pagan Test [2], with the homoscedasticity of the error terms as null hypothesis. The assumption can be considered to be verified since the $p\text{-value}$ (0.18) was greater than 0.05 and thus the null hypothesis cannot be rejected.

Normality. To test this assumption we used the Shapiro-Wilk Test [20], by considering as null hypothesis the normality of error terms. The results revealed that the assumption can be considered to be verified since the $p\text{-value}$ (0.24) of the statistic (0.95) was greater than 0.05 and thus the null hypothesis cannot be rejected.

Table IV shows some statistics about the model obtained with OLSR employing CFP and FP. The model is

¹ Functional size unit (FSU) is the the widely used ISO term for the unit of measure of any FSM method [7]

characterized by an R^2 value of 0.70 and an F value of 54.12, and a low $p\text{-value}$ (<0.01), indicating that the prediction is indeed possible with a high degree of confidence. The $t\text{-values}$ and $p\text{-values}$ for the corresponding coefficient and the intercept present values greater than 1.5 and less than 0.05, respectively. Moreover, the coefficient 1.01 and the intercept 207 are significant at level 0.05.

We can observe that a coefficient very close to 1 was obtained with an interval of confidence of (0.8-1.4) which is very close to the ones obtained in [11][14][1][8] and quite different from the intersection of most probable intervals (i.e., 0.7-0.8) obtained in [7] where different previous studies were reviewed and updated (see Table I). It is interesting to note that also the previous studies (discussed and summarized in [7]) highlighted a similar behavior in the slope of the straight line, i.e., around 1. On the other hand, differently from the previous studies (see Table I), we have obtained a less R^2 value and a greater value for the intercept. Thus, focusing on these observations it seems that the conversion equation for our dataset of 25 Web applications is less accurate than those obtained in previous studies for the employed software.

TABLE IV. THE RESULTS OF OLSR WITH CFP AND FP

Parameters	Value	Std.Err	t-value	p-value	Confidence interval
Intercept (a)	207	61.5	3.36	<0.01	(79.9-333.4)
Coefficient (b)	1.01	0.15	7.36	<0.01	(0.8-1.4)
Statistical indicators of the model	R^2		Std.Err	F	Sign F
	0.70		149.8	54.1	<0.01

In order to satisfy the condition that if the size measured in terms of FP is zero then also the size in terms of CFP is zero (see Section III.B), we applied OLSR on the variables obtained by exploiting a log transformation on CFP and FP. The results are reported in Table V. The model is characterized by R^2 and F value less than those obtained with the linear equation of Table IV. These results are quite expected since the analysis of OLSR assumption revealed that there was a significant (linear) correlation between CFP and FP and transformation could be considered not necessary. However, we also investigated the accuracy of the model obtained by log transforming the variables since we want to also consider the property that the equation must verify the origin of the coordinates [7]. The $t\text{-values}$ and $p\text{-values}$ for the corresponding coefficient and the intercept present values greater than 1.5 and less than 0.05, respectively. Moreover, the coefficient 0.71 and the intercept 2.19 are significant at level 0.05.

We can observe that, differently from the results achieved in [7], the exponent is not very close to 1 and the confidence interval of the coefficient is quite different from those obtained in the previous studies (see Table II). Thus, we cannot converge to a proposal of a conversation equation based on the assumption of $CFP \approx FP$ as done in [7].

As stated before, to evaluate the prediction accuracy of the models obtained with OLSR, we performed a leave-one-out cross validation, whose results are reported in Table VI. We can observe that these results do not satisfy the thresholds

provided in [4] since MMRE values are higher than 0.25 and Pred(25) values are less than 0.75. Furthermore, the Q^2 values are not so high. However, we can highlight that the MMRE and Pred(25) values obtained with linear model (i.e., not transforming the variables FP and CFP) are very close to the thresholds. Thus, the linear model (i.e., without transformation of the variables) produces more accurate size estimates

TABLE V. THE RESULTS OF OLSR WITH LNCFP AND LNFP

Parameters	Value	Std.Err	t-value	p-value
Intercept ($Ln a$)	2.19	0.71	3.10	<0.01
Coefficient (b)	0.71	0.12	5.82	<0.01
Statistical indicators of the model	R^2	Std.Err	F	Sign F
	0.60	0.34	33.92	<0.01
Parameters of non linear equation	b	Confidence Interval b	a	Confidence Interval a
	0.71	(0.5-1.0)	8.92	(7.5-10.4)

TABLE VI. PREDICTION ACCURACY INDICATORS FOR CFP AND FP

Study	MMRE	MdMMRE	Pred(25)	Q^2
CFP and FP	0.30	0.17	0.72	0.64
LnCFP and LnFP	0.30	0.20	0.68	0.71

B. Case Study Validity

It is widely recognized that several factors can bias the construct, internal, external, and conclusion validity of empirical studies. In order to mitigate the threats to construct validity we defined a template to be filled in by the project managers to calculate the values of the size measures. They took into account the COSMIC Measurement Manual (version 2.2.) to calculate the number of Cosmic Functional Points [6] and the counting conventions of the FPA method [15] to calculate Function Points. We supervised the collection procedure and two of the authors analyzed the filled templates and the analysis and design documents in order to cross-check the provided information. As for the internal validity, no initial selection of the subjects was carried out, so no bias has been apparently introduced. Moreover, the Web applications were developed with technologies and methods that subjects had experienced. Consequently, confounding effects from the employed methods and tools can be excluded. Regarding the conclusion validity, we carefully applied the estimation methods and the statistical tests, verifying all the required assumptions. With regard to the external validity, the applications involved in our empirical analysis are representative samples of modern Web applications, taking into account their type, functionalities, target platforms, and complexity. Moreover, the results of our empirical analysis highlighted that the obtained coefficient of the conversion equation is quite close to the ones obtained in other case studies with different types of software.

V. CONCLUSION

The results have shown that there is a significant correlation between the size expressed in terms of COSMIC and FP, but differently from the previous studies the

conversion factor is not so close to 1 [7]. Furthermore, the considered statistical indicators of the equation are quite good and the values of the measures used to assess the accuracy of the obtained estimates (i.e., MMRE, Pred(25), and Q^2) are very close to the thresholds provided in [4].

REFERENCES

- [1] A. Abran, J.M. Desharnais, F. Azziz, "Measurement convertibility: from function points to COSMIC", in Proceedings of the International Workshop on Software Measurement (IWSM 2005), Shaker-Verlag, Aachen, 2005, pp. 227-240.
- [2] T. Breush, A. Pagan, "A simple test for heteroscedasticity and random coefficient variation", *Econometrica* 47 (1992), pp. 1287-1294.
- [3] L. Briand, J. Wurst, "Modeling Development Effort in Object-Oriented Systems Using Design Properties", *IEEE Transaction on Software Engineering* 27 (11) (2001), pp. 963-986.
- [4] D. Conte, H.E. Dunsmore, V.Y. Shen, "Software engineering metrics and models", The Benjamin/Cummings Pub. Company, Inc., 1986.
- [5] COSMIC Web site (2007) [cited 2010] URL: <http://www.cosmicon.com>, 2007.
- [6] COSMIC Measurement manuals and documents (2007) [cited 2010], URL: <http://www.cosmicon.com>, 2007.
- [7] J. Cuadrado-Gallego, L. Buglione, M. Domínguez-Alda, M. Fernández de Sevilla, J. Gutiérrez de Mesa, O. Demirörs, "An experimental study on the conversion between IFPUG and COSMIC functional size measurement units", *Information & Software Technology* 52(3), (2010), pp. 347-357.
- [8] J. Desharnais, A. Abran, J.J. Cuadrado-Gallego, Convertibility of function points to COSMIC: identification and analysis of functional outliers, in Proc. of International Workshop on Software Measurement (IWSM 2007), Shaker-Verlag, Aachen, 2007, pp. 130-146.
- [9] S. Di Martino, F. Ferrucci, C. Gravino, "Estimating Web Application Development Effort Using Web-COBRA and COSMIC: An Empirical Study", in Proc. of EuroMicro Conference on Software Engineering and Advanced Applications (SEAA'09), ACM press, 2009, pp. 306-312.
- [10] F. Ferrucci, C. Gravino, S. DiMartino, "A Case Study Using Web Objects and COSMIC for Effort Estimation of Web Applications", in Proc. of EuroMicro Conference on Software Engineering and Advanced Applications (SEAA'08), IEEE press, 2008, pp. 441-448.
- [11] T. Fetcke, "The Warehouse Software Portfolio, A Case Study in Functional Size Measurement", TR No. 1999-20, Département d'informatique, Université du Québec à Montréal, Canada, 1999, <<http://www.gelog.etsmtl.ca/cosmic/casestudies/fetcke1999b.pdf>>.
- [12] J. Freund, *Mathematical Statistics*, Prentice-Hall, U. S. River, NJ, 1992.
- [13] Gencel, Ç., Demirörs, O.: Functional size measurement revisited. *ACM Trans. Softw. Eng. Methodol.* 17(3), 2008.
- [14] V. Ho, A. Abran, T. Fetcke, "A Comparative Study Case of COSMIC, Full Function Point and IFPUG Methods", Département d'informatique, Université du Québec à Montréal, Canada, 1999, <<http://www.lrgl.uqam.ca/publications/pdf/599.pdf>>.
- [15] International Function Point Users Group 2004. Function point counting practices manual, release 4.2.1.
- [16] ISO/IEC Functional Size Measurement standards 14143 Parts 1 to 6.
- [17] K. Maxwell, "Applied Statistics for Software Managers". Software Quality Institute Series, Prentice Hall, 2002.
- [18] E. Mendes, S. Counsell, N. Mosley, C. Triggs, I. Watson, "A Comparative Study of Cost Estimation Models for Web Hypermedia Applications", *Empirical Software Engineering* 8(2), 2003, pp. 163-196.
- [19] T. Rollo, "Sizing E-Commerce", in Proc. of Australian Conference on Software Measurement, IEEE press, 2000.
- [20] P. Royston, "An extension of Shapiro and Wilk's W test for normality to large samples", *Applied Statistics* 31 (2) (1982), pp. 115-124.
- [21] Total Metrics: How to Decide which Method to Use. Method for Software Sizing, Copyright Total Metrics URL: http://www.totalmetrics.com/function-point-resources/downloads/R185_Why-use-Function-Points.pdf