Functional versus Design Measures for Model-Driven Web Applications:
A Case Study in the Context of Web Effort Estimation

Lucia De Marco[1], Filomena Ferrucci[1], Carmine Gravino[1], Federica Sarro[1], Silvia Abrahao[2], Jaime Gomez[3]

[1]*University of Salerno, Fisciano (SA), Italy*
{ldemarco, fferrucci, gravino, fsarro}@unisa.it
[2] *Universidad Politécnica de Valencia, Valencia, Spain*
sabrahao@dsic.upv.es
[3] *Universidad de Alicante, Alicante, Spain*
jaime@dlsi.ua.es

Abstract—In the literature we can identify two main approaches for sizing model-driven Web applications: one based on design measures and another based on functional measures. Design measures take into account the modeling primitives characterizing the models of the specific model-driven approach. On the other hand, the functional measures are obtained by applying functional size measurement procedures specifically conceived to map the modeling primitives of the model-driven approach into concepts of a functional size measurement method. In this paper, we focus our attention on the Object-Oriented Hypermedia (OO-H) method, a model-driven approach to design and develop Web applications. We report on the results of an empirical study carried out to compare the ability of some design measures and OO-HFP (a model-driven functional size measurement procedure) to predict the development effort of Web applications. To this aim, we exploited a dataset with 30 Web projects developed using OO-H. The analysis highlighted that each design measure was positively correlated with the Web application development effort. However, the best estimation model obtained by exploiting the Manual Stepwise Regression employed only the measure Internal Links (IL). Furthermore, the study highlighted that the estimates obtained with the IL based prediction model were significantly better than those achieved using the OO-HFP based prediction model. These results seem to confirm previous investigations suggesting that Function Point Analysis can fail to capture some specific features of Web applications.

Keywords-Model-driven development; Web applications; Effort estimation; Design measures; Functional size measurement.

I. INTRODUCTION

A survey on Web development practices [8] showed that practitioners are not using development processes proposed in the literature (e.g., RUP). They found that the employed processes are in general 'ad-hoc', although some organizations are starting to look into the use of agile methods. This trend is therefore changing as we are moving towards model-driven development processes whose goal is the development of software at a higher level of abstraction based on models and model transformations. In this context, several model-driven approaches have been proposed to design and develop Web applications (e.g., WebML [10], W2000 [6], OO-H [15]). These approaches support the construction of different views (i.e., models) of a Web application comprising at least a structural model, a navigation model, and a presentation model. Moreover they provide a tool able to automatically generate Web applications from these models. Also for this type of software applications it is important to have suitable measures that can support project managers to allocate resources, control costs, and schedule, leading to projects that are finished on time, within budget, and satisfying quality requirements.

In the literature we can identify two main approaches for sizing model-driven Web applications – one based on design measures and another based on functional measures. Design measures take into account the modeling primitives characterizing the models of the specific model-driven approach. Examples of these measures are the ones defined for the W2000 approach (e.g., Components, Clusters, and Slots) in [7], where the authors investigated the ability of those measures to predict the design effort of Web applications developed using W2000 [6]. On the other hand, functional measures are obtained by applying a functional size measurement procedure specifically conceived to map the modeling primitives of the model-driven approach into concepts of a functional size measurement method (e.g., Function Point Analysis [16], COSMIC [12]). An example of this approach is OO-HFP (Object-Oriented Hypermedia Function Points) [1], a model-driven functional size measurement procedure conceived for Web applications developed using the OO-H (Object-Oriented Hypermedia) [15] method and based on the Function Point Analysis.

At the best of our knowledge no study has been carried out to compare these two kinds of measures. Therefore, the contribution of this paper is to address this issue focusing our attention on the OO-H method. In particular, we identify and investigate design measures obtained from OO-H models and compare them to OO-HFP. Both size measures were compared using an industrial dataset containing 30 Web projects developed using OO-H and verifying their effectiveness in estimating development effort.

The remainder of this paper is organized as follows: Section II presents an overview about OO-H, the OO-H

design measures, and OO-HFP. Section III presents the case study design. Section IV presents an analysis and discussion of the results. Section V discusses related work. Finally, Section VI presents our conclusion and future work.

II. BACKGROUND

A. OO-H Method

OO-H (Object-Oriented Hypermedia) is a model-driven Web application development method based on the object-oriented paradigm [15]. The OO-H method includes a design process, a pattern catalog, the definition of the navigation access diagram, the definition of the abstract presentation diagram, and a CASE tool that supports and automates the development process. The design process begins by using two kinds of models: the UML Class Diagram (UCD) and the Navigational Access Diagrams (NADs). The UCD represents the domain information structure, i.e., the static part of the Web application, while each NAD provides the navigational view of the Web application for a specific type of user, representing information, services, and navigation paths required for the execution of a single navigation. Example of a NAD is shown in Fig. 1. The NADs are composed of a set of modeling primitives: Navigational Classes (NCs), Navigational Targets (NTs), Navigational Links (NLs), and Collections (Cs). Starting from the NADs and following a set of mapping rules, a set of Abstract Page Diagrams (APDs) can be generated. The definition of the APDs is based on a set of XML DTDs. Both the NADs and the APDs capture the interface related design information with the aid of a set of patterns, defined in an interface pattern catalog, which is integrated in the OO-H proposal; it contains a set of constructs to facilitate the reuse of design experiences and the consistency both among the different interface modules and among application interfaces. Once the APDs are refined, a Web application can be generated for the desired environment (e.g., HTML, ASP's, JSP's), with complete independence from final implementation issues. The development process and the transformation rules have been automated in the CASE tool VisualWADE [25].

B. OO-H Design Measures

Starting from the OO-H primitives we identified the design measures shown in Table I that also specifies the model where they are used. As for UCD, we identified the number of classes (CL), the number of associations (AS), the number of aggregations (AG), the number of compositions (CO), and the generalizations (GE). As for NAD seven measures were identified. An NT groups elements of the model that collaborate in the coverage of a user navigational requirement. An NC represents a view over a set of attributes and operations of a class from the class diagram. A Cs is a hierarchical structure that groups a set of navigational links. A navigational link defines the path that a user can follow through the interface. Each navigational link has an origin node and a corresponding destination node. There are six types of navigational links: I-Link (Internal Link), defining the path inside the same navigational target; T-Link (Traversal Link), defining the path between navigational classes belonging to different navigational targets; R-Link (Requirement Link), representing points at the starting navigation point of each NT; X-Link (eXit Link), representing points outside the boundary of the application; couple of S-Link and R-Link (Service and Response Links), representing the services and the view after response.

C. OO-HFP Measure

The OO-HFP (Object-Oriented Hypermedia Function Points) measurement procedure was proposed to size Web applications, developed with the OO-H method, in terms of the IFPUG Function Points (FP, for short) [1]. FP represents the version of the Function Point Analysis (FPA) managed by the International Function Point Users Group (IFPUG). The FPA method was the first FSM method proposed in the literature; it was introduced by Albrecht of IBM in 1975 to measure a software product size in terms of number of functions provided to the end user in the early development phases, making the size measure no more related to the technology choices adopted. The original formulation was extended several times and in 2003 it was certified by ISO as an international standard.
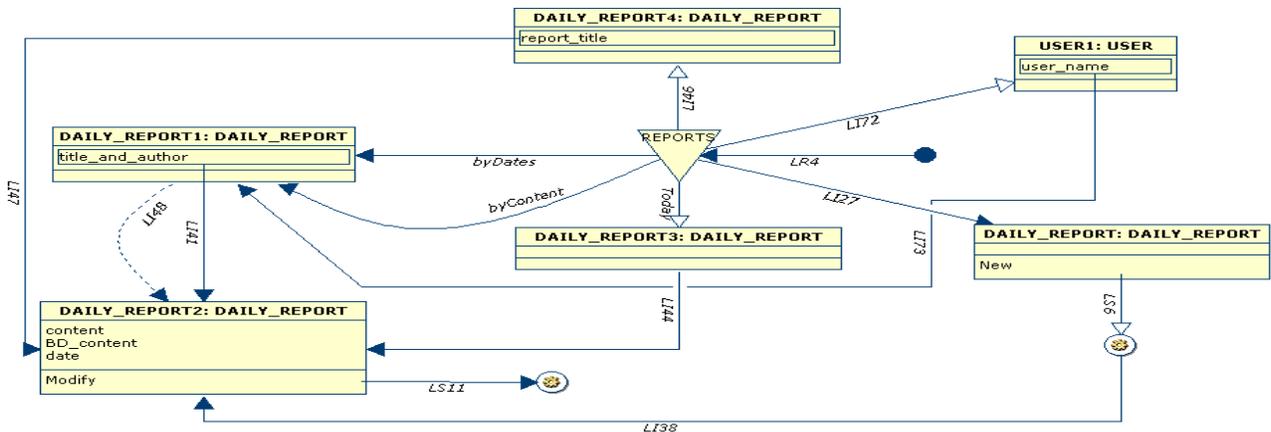
Figure 1. NAD for Task Manager application [1]

TABLE I. OO-H DESIGN MEASURES

Model	Measure	Description	Scale
UML Class Diagram (UCD)	CL	Number of classes	ratio
	AS	Number of Associations	ratio
	AG	Number Aggregations	ratio
	CO	Number of Compositions	ratio
	GE	Number of Generalizations	ratio
Navigational Access Diagram (NAD)	NT	Navigational Targets	ratio
	NC	Navigational Classes	ratio
	IL	Number of I-Links	ratio
	TL	Number of T-Links	ratio
	XL	Number of X-Link	ratio
	RL	Number of R-Links	ratio
	SL	Number of S-Links	ratio
	Cs	Number of Collections	ratio

The FP method identifies and classifies each function as External Input (EI), External Output (EO), External Inquiry (EQ), Internal Logical File (ILF), and External Interface File (EIF). Each function is then weighted depending on its type and level of complexity in agreement with standard values as specified in the Counting Practices Manual. The level of complexity is determined using the number of Data Element Types (DETs), File Types Referenced (FTRs), and Record Element Types (RETs).

The OO-HFP counting is based on the design model of an OO-H Web application (consisting of an UCD and a set of NADs) and exploits a specifically conceived set of mapping rules and of numerical assignment rules [1]. The mapping rules provide a mapping between the FPA concepts, as the counting scope and the boundary of the Web application, the data (ILF and EIF) and transactional (EI, EO, and EQ) functions, and the OO-H modeling primitives. Moreover, a set of measurement rules is used to identify: 1) the DETs and RETs for each class in the UCD exploited to determine the complexity of a class or of a class hierarchy; 2) the DETs and the FTRs employed to determine the complexity of each method in the UCD; 3) the DETs and RETs used to determine the complexity of a NT. Once DETs, RETs, and FTRs have been counted for the application to be measured, the FPA counting rules are applied to classify the function complexity (low, average, high), to assign weights to the functions, and to aggregate the assigned values into an overall functional size value for the Web application.

III. CASE STUDY DESIGN

The performed empirical study aimed at investigating whether the identified OO-H design measures can be exploited to size model-driven Web applications and estimate the effort needed to develop them. Moreover, we were also interested in comparing the accuracy in estimating development effort of OO-H design measures with the one of the OO-HFP proposed in [1]. Thus, the research questions we addressed in our study are:

RQ1. Are the identified OO-H design measures and the OO-HFP measure correlated with the effort?

RQ2. Are there significant differences in the accuracy of the effort estimates built with the OO-H design measures and the OO-HFP measure?

A. Dataset

The data employed in our empirical analysis was obtained from 30 Web applications developed using the OO-H approach between 2003 and 2006 by a Spanish Web software company. They were new developments and included typical Web applications, e.g., applications for cinema management, hospital management, Web-based content management, and intranets. These applications were generated automatically in a range of programming languages including Javascript (DHTML/DOM), J2EE (JSP, Servlet, EJB), PHP, and SQL. The staff involved in the development of these applications consisted of a small team ranging from one to five analysts. The values of the OO-H design measures and OO-HFP measure were obtained by using the VisualWADE tool [25] that has been extended with plug-ins to automate the counting of these measures from the OO-H conceptual models. The VisualWADE tool was developed using the Python language, which gives powerful capabilities for introspection that is the ability to query and manipulate the objects and metaobjects at run-time. These correspond to the OO-H models and its corresponding metamodel which are defined within the tool as data. In addition, its modular architecture is easily extensible by means of plug-ins that allows to dynamically add new features, such as the OO-HFP and OO-H design measure plug-ins that gather information about the OO-H model and its corresponding OO-H metamodel to perform the necessary calculations to produce the measurement reports.

Table II shows the descriptive statistics of the considered size measures and of the variable Effort, denoting the actual effort in person/hours. It is worth noting that we excluded from our analysis the variables having more than 50% zero values (i.e., AG, CO, GE, and XL).

TABLE II. DESCRIPTIVE STATISTICS OF THE EMPLOYED MEASURES

Measure	Min	Max	Mean	Median	Std. Dev
CL	1	71	18	10	21
AS	0	65	14	8	19
NC	2	408	59	20	100
NT	1	64	11	5	15
IL	4	671	88	26	148
TL	0	432	41	9	87
RL	1	38	8	5	10
SL	0	411	31	8	76
Col	1	50	13	7	14
OO-HFP	30	2193	391	203	512
Effort	16	3644	623	123	1123

B. Data Analysis

In order to assess the relationship between each considered measure and the variable denoting the development effort, we applied two nonparametric association statistics: Kendall's tau [17] and Spearman's rho [14], as done in [7]. These statistics range from +1 to -1, where +1 indicates perfect correlation and -1 indicates a perfect inverse correlation. No correlation is indicated by 0. Even if Spearman's rho is widely employed in the literature, we decided to apply also Kendall's tau because it has several advantages from statistical point of view, e.g., a near normal distribution of the observed function for small number of observations in the dataset [17]. Since we are also interested in quantifying the relationship between the dependent variable Effort and each of the independent variables shown in Table II we also exploited a parametric technique like Ordinary Least Square Regression (OLSR). Indeed, OLSR explores the relationship between a dependent variable and one or more independent variables, providing a model described by an equation:

$$y = b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + c \qquad (1)$$

where y is the dependent variable, x_1, x_2, ..., x_n are the independent variables, b_i is the coefficient that represents the amount variable y changes when variables x_i changes 1 unit, and c is the intercept.

In particular, to assess the correlation between Effort and the identified considered size measures, we applied the univariate OLSR, i.e., by considering at each time Effort as y and one of the size measures as x (i.e., CL). This technique is widely used to find whether there is statistical significant correlation between two variables. To evaluate the goodness of fit of the obtained regression models, several indicators can be considered. Among them, we exploited the square of the linear correlation coefficient, (adjusted) R^2, that shows the amount of variance of the dependent variable explained by the model related to the independent variable. 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 us to evaluate their importance for the generated model. In particular, p-values lower 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, we also assessed the stability of each effort estimation model by following the procedure proposed in [22], which consists of the following steps:

 a. The use of a residual plot showing residual vs. fitted values. The goal is to determine whether the residuals are random or normally distributed.

 b. The use of Cook's distance values for all observations to identify influential data points. With this procedure, any observation having distance greater than $3\times(4/n)$ (where n represents

the total number of observations) is removed from the data analysis. While, those observations that have distances greater than 4/n but smaller than $(3\times(4/n))$ are removed to test the model stability, which is done by observing the effect of their removal on the model. If the model coefficients remain stable and (adjusted) R^2 improves, the influential points are retained in the data analysis.

Observe that we have exploited both nonparametric and parametric techniques also to have more checks to highlight and confirm the empirical evidence that we can have with just one technique [7]. OLSR also allowed us to assess and compare the accuracy of the effort estimates obtained by employing the identified OO-H design measures with the accuracy of the effort estimates achieved using OO-HFP. Indeed, we built the linear regression model exploiting the OO-H design measures as independent variables and the model employing only OO-HFP as independent variable. Once a prediction model was constructed, the effort estimation for a Web application was obtained by using in the prediction model the application size in terms of the chosen size measures (e.g., OO-HFP or OO-H measures).

To verify whether or not the predicted effort is a useful estimation of the actual effort, we applied a cross-validation, splitting the original dataset into training and validation sets. Training sets are used to build models and validation (or test) sets are used to validate the obtained models. In particular, we exploited a leave-one-out cross validation, where the original dataset is divided into n (i.e., n is the size of the original dataset) different subsets of training and validation sets, with each validation set containing one observation [9].

As evaluation criteria we exploited absolute residuals (| Effortactual – Effortpredicted|), by employing boxplots and statistical tests to assess them. Boxplots are widely employed in exploratory data analysis since they provide a quick visual representation to summarize the data, using five values: median, upper and lower quartiles, minimum and maximum values, and outliers [19]. On the other hand, statistical tests are used to establish whether one estimation model provided significantly better estimates than another. In particular, we performed the Wilcoxon signed rank test (or the unpaired version, i.e., Mann-Whitney test) to verify the following null hypothesis: "the two considered population of absolute residuals have identical distributions". This kind of test is used to verify the hypothesis that the mean of the differences in the pairs is zero. In all the performed statistical tests we decided to accept a probability of 5% of committing a Type-I-Error [27]. Thus, we reject the "null hypothesis" if the p-value is less than 0.05, where the p-value denotes the statistical significance of the test.

IV. RESULTS AND DISCUSSION

In the following, we report and discuss the achieved results taking into account the defined research questions.

A. Research Question RQ1

Table III shows the results achieved by applying the Spearman's rho and Kendall's tau tests. In particular, we reported the statistics and the p-values for both the tests and each employed independent variables. The dependent variable is Effort.

We can observe that all the employed size measures are positively associated with Effort. Furthermore, the results of the tests are statistical significant as the p-values of the statistics reveal. This means that when the value of a size measure (e.g., CL) increases, the value of the variable Effort increases as well. We can also note that the independent variable characterized by the highest association with dependent variable Effort is IL (i.e., Internal Link). However, other independent variables are characterized by a high association (i.e., NC, NT, TL, RL, and OO-HFP) since their Spearman's rho statistics are greater than or very close to 0.9. The variables NC and NT also achieved statistics greater than 0.8 as IL, when Kendall's tau test was applied.

The above results are confirmed by the application of the univariate OLSR reported in Table IV where for each built estimation model it is reported the employed independent variable (e.g., CL), the value of the coefficient and the intercept, their t-value and p-value, and the adjusted R^2. Indeed, the model characterized by the highest adjusted R^2 is the one employing IL as independent variable (0.94). The models employing NC also achieved an adjusted R^2 value greater than (or equal to) 0.9, which can be considered high. On the other hand, the models based on variables NT, RL, TL, and OO-HFP were characterized by an adjusted R^2 value greater than 0.8 but less than 0.9.

Thus, the above results suggested that we can positively answer the first research question, i.e., the identified OO-H design measures and the OO-HFP measure were significantly (and positively) correlated with the variable denoting the development effort.

B. Research Question RQ2

To answer the second research question, we have considered the accuracy of the estimates obtained by using the effort estimation model based on OO-HFP (i.e., the one reported in the last row of Table IV) and the effort estimation model employing OO-H design measures. To this end, we built a model applying Manual StepWise Regression (MSWR) and employing all the considered OO-H design measures. In particular, we exploited the technique proposed in [18], which allows us to compute linear regression analysis in steps. The estimation model is built by adding, at each step, the independent variable with the highest correlation to the dependent variable, taking into account all the variables currently in the model. The idea underlying this procedure is to select the best fitting model, and then to use OLSR to obtain the final model. The results of the performed analysis (that we cannot report for sake of space) suggested that the best fitting model obtained employing OO-H design measures was the one employing only IL and TL as independent variables.

To evaluate and compare the accuracy of the effort estimates obtained using this model with the accuracy of the estimates achieved using the OO-HFP based model, we applied a leave-one-out cross validation. The analysis of the results in terms of absolute residuals suggests that the estimates achieved with the IL and TL based model are better than those achieved with the OO-HFP model (see Table V). This result is confirmed by the analysis of the boxplots of residuals depicted in Fig. 2. Indeed, the median of the IL and TL boxplot is more close to zero than the one of the OO-HFP boxplot. The box length and tails of the IL and TL boxplot are less skewed than the ones of OO-HFP boxplot. Moreover, the results of the Wilcoxon test suggest that the absolute residuals achieved with the IL and TL based model are significant less than those obtained by using the OO-HFP model (p-value = 0.022).

TABLE III. RESULTS OF ASSOCIATION ANALYSIS

Ind. variable	Spearman's rho		Kendal's tau	
	statistic	p-value	statistic	p-value
CL	0.814	<0.01	0.636	<0.01
AS	0.808	<0.01	0.638	<0.01
NC	0.941	<0.01	0.816	<0.01
NT	0.939	<0.01	0.820	<0.01
IL	0.938	<0.01	0.823	<0.01
TL	0.916	<0.01	0.716	<0.01
RL	0.907	<0.01	0.786	<0.01
SL	0.675	<0.01	0.535	<0.01
COL	0.896	<0.01	0.726	<0.01
OO-HFP	0.919	<0.01	0.766	<0.01

TABLE IV. RESULTS OF UNIVARIATE OLSR

Ind. Var.	coefficient			Intercept			Adj. R^2
	value	t-value	p-value	value	t-value	p-value	
CL	1.14	9.08	<0.01	2.54	8.01	<0.01	0.74
AS	1.04	8.45	<0.01	3.12	11.09	<0.01	0.71
NC	1.06	19.25	<0.01	1.88	10.21	<0.01	0.93
NT	1.13	14.15	<0.01	3.28	19.97	<0.01	0.87
IL	1.12	21.07	<0.01	1.16	5.79	<0.01	0.94
TL	0.75	12.29	<0.01	3.59	21.15	<0.01	0.84
RL	1.23	11.72	<0.01	3.30	16.92	<0.01	0.82
SL	0.64	5.08	<0.01	3.82	11.60	<0.01	0.46
COL	1.17	8.93	<0.01	2.80	9.44	<0.01	0.73
OO-HFP	1.28	14.83	<0.01	-1.71	-3.64	<0.01	0.88

a. After eliminating influential observations according to the analysis of the stability of the model

TABLE V. DESCRIPTIVE STATISTICS OF ABSOLUTE RESIDUALS

| Estimation Model based on	Min	Max	Mean	Median	Std. Dev.
IL,TL	0.79	1264.14	136.04	18.19	302.20
OO-HFP	1.32	2003.64	229.79	46.13	514.93
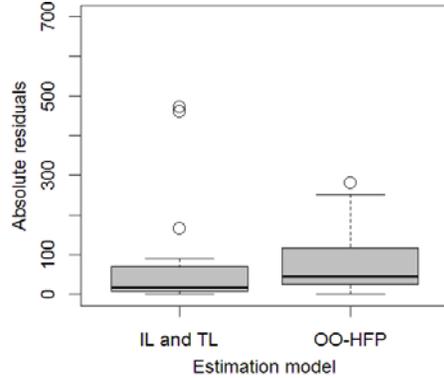
Figure 2. Boxplots of absolute residuals.

Thus, the analysis allowed us to positively answer the second research question, i.e., there was a significant difference in the accuracy of the effort estimates achieved in terms of the OO-H design measures and the accuracy of the estimates obtained with the OO-HFP measure. In particular, the IL and TL based model provided significant better estimates. Let us recall that both IL and TL concern with navigational aspects. Thus, this confirms suggestions of other studies [6], highlighting that aspects concerning Web pages navigation can play a crucial role in the estimation of Web application development effort. Indeed, navigational features are considered a specific characteristic of Web applications [11].

C. Threats to Validity of the Case Study

Several factors can bias the construct, internal, external, and conclusion validity of empirical studies. As for the construct validity, the collection of the information related both to the size measures and the actual effort represents a crucial aspect. We calculated the employed measures using the plug-ins realized for the VisualWADE tool [25]. This obviously reduces the possibility of manual error and rule misunderstanding. To mitigate the threat of an erroneously implementation, a careful testing of the plug-ins was carried out. Furthermore, one of the authors supervised the procedure employed by the involved company to collect in a controlled and uniform fashion the information used for the empirical analysis. In order to maximize accuracy, an interview format, in addition to the self-report format was used. The procedure consisted in using an excel file for collecting a range of variables for the 30 projects. The project manager in the company filled in the excel file and one of the authors performed interviews to ensure that the data collected was accurate. As for the internal validity, the developers involved in the study were professionals who worked in the software company. No initial selection of the developers was carried out, so no bias has been apparently introduced. The Web applications employed in the case study were designed and developed with the OO-H method and VisualWADE tool that developers had experienced. Consequently, confounding effects from the employed methods and tools can be excluded. As for the reliability of

the data and lack of standardization, the size measures were obtained automatically. Furthermore, the actual effort was collected using the same questionnaire for all the Web applications and the developers were instructed on how to use it, to correctly provide the required information. To mitigate possible conclusion validity threats, we carefully applied the estimation techniques and the statistical tests, verifying all the required assumptions. As for the external validity, let us observe that the applications employed in our empirical analysis were developed by a single software company. It is recognized that the results obtained in an industrial context might not hold in other contexts [9].

V. RELATED WORK

Related work is represented by studies investigating design measures or Functional Size Measurement methods for model-driven Web applications.

A. Studies on Design Measures

Baresi and Morasca [7] presented three empirical studies to investigate some goals concerning the effort needed to design Web Applications using W2000 as model-driven approach [6]. One of the goals concerns a proposal of metrics to measure internal attributes of W2000 artifacts and the assessment of the impact of those metrics on the total effort needed to design Web applications. In particular, they proposed eleven size measures for presentation, navigation, and content, which are the three main layers around which W2000 models are organized. The empirical results highlighted that different studies correlate different measures with the actual effort, thus not providing a clear indication on which of the analyzed measures is most suitable as effort predictor.

B. Studies on Functional Size Measurement Methods

In [3] OO-HFP a dataset of 12 applications was exploited to assess the effectiveness OO-HFP for effort estimation with respect to other set of measures, namely the ones defined by Mendes et al. for the Tukutuku database [22]. The results revealed that the obtained estimates were comparable with those achieved using the Tukutuku measures. The approach was further assessed in [1] using the same set of industrial Web applications employed in our study and by comparing its effectiveness as effort predictor with respect to the use of standard IFPUG FPA. The results showed that the effort estimates obtained using OO-HFP were more accurate than those obtained using IFPUG FPA.

Another model-driven approach for sizing Web applications in terms of IFPUG FPA was provided in [13] for applications built using WebML and the tool WebRatio [10]. An initial analysis of the approach was performed by comparing the size automatically obtained in terms of FPA with the size determined manually by two skilled analysts [13]. They employed data coming from four projects developed by different companies and the results showed that the automatically obtained sizes differed from the manual counting for a maximum amount of about 11%.

In [2] a COSMIC based procedure for OO-H Web applications is proposed, namely OO-HCFP. COSMIC [12] represents an FSM of second generation and was introduced to overcome some issues of FPA. However, the usefulness of OO-HCFP for estimating development effort of OO-H Web applications has not yet been investigated.

VI. CONCLUSIONS AND FUTURE WORK

We performed a case study to assess the effectiveness of design measures obtained from OO-H conceptual models as indicators of the development effort employed to design and develop model-driven Web applications. Furthermore, we compared the identified OO-H design measures, representing internal attributes of OO-H models, with the OO-HFP measure obtained by applying a functional size measurement procedure based on Function Points Analysis [1]. The data employed in our empirical analysis was obtained from 30 Web applications developed using the OO-H model-driven approach [15]. The results have highlighted that each design measure is positively correlated with the Web application development effort. However, the best estimation model obtained by exploiting MSWR employed only the measures IL and TL. Furthermore, the study has revealed that the estimates obtained with this model are significantly better than those achieved using the OO-HFP based model. These results can help the software company that provided the data for our study to perform better effort estimations since they provide evidence about the effectiveness of some design measures obtained from OO-H conceptual models as effort predictors. These results also confirmed previous investigations suggesting that FPA can fail to capture some specific features of Web applications [24]. Nevertheless, functional size measures present several advantages with respect to design measures, first of all the fact that they do not depend from the specific model-driven approach. This motivates the investigation of others functional size measurement procedures for OO-H (and possibly for other model-driven approaches) that are able to capture the aspects that are more correlate with effort. Thus, we are planning to analyze the effectiveness of a second generation functional size measurement method, i.e., COSMIC [2].

REFERENCES

[1] S. Abrahão, J. Gómez, E. Insfran, "Validating a size measure for effort estimation in model-driven Web development", Information Sciences, vol. 180, n. 20, 2010, pp. 3932-3954.

[2] S. Abrahão, L. De Marco, F., Ferrucci, C., Gravino, F. Sarro, "A COSMIC Measurement Procedure for Sizing Web Applications Developed using the OO-H", in Procs. of the International Workshop on Advances in Functional Size Measurement and Effort Estimation, co-located with the European Conference on Object-Oriented Programming (ECOOP 2010).

[3] S. Abrahão, E. Mendes, J. Gomez, E. Insfran, "A Model-Driven Measurement Procedure for Sizing Web Applications: Design, Automation and Validation", in Procs. of ACM/IEEE International Conference on Model Driven Engineering Languages and Systems (MoDELS 07). 2007, pp. 467–481.

[4] S. Abrahão, O. Pastor, "Measuring the functional size of Web applications", International Journal of Web Engineering and Technology, vol. 1, n. 1, 2003, pp. 5–16.

[5] S. Abrahão, G. Poels, "A family of experiments to evaluate a functional size measurement procedure for Web applications", Journal of Systems and Software vol. 82, n. 2, 2009, pp. 253-269.

[6] L. Baresi, F. Garzotto, P. Paolini, "From Web Sites to Web Applications: New Issues for Conceptual Modeling", in Procs of the International Workshop on The World Wide Web and Conceptual Modeling, co-located with the International Conference on Conceptual Modeling, 2000.

[7] L. Baresi, S. Morasca, "Three Empirical Studies on Estimating the Design Effort of Web Applications", ACM Trans. on Softw. Engineer. Method., vol. 16, n. 4, 2007.

[8] C. Barry, M. Lang "A survey of Multimedia and Web Development Techniques and Methodology Use", IEEE Multimedia 8, 2001, 52-60.

[9] L. Briand, J. Wurst, "Modeling Development Effort in Object-Oriented Systems Using Design Properties". IEEE Trans. on Softw. Engineer., vol. 27, n. 11, 2001, pp. 963-986.

[10] S. Ceri, P. Fraternali, A. Bongio, "Web Modeling Language (WebML): a modeling language for designing Web sites", in Computer Networks vol. 33, n. 1-6, 2000, pp. 137–157.

[11] J. Conallen, Building Web Applications with UML, Addison-Wesley Object Technology Series, 1999.

[12] COSMICON. 2009. The COSMIC Functional Size Measurement Method, Version 3.0.1, Measurement Manual. In http://www.cosmicon.com/portal/public/COSMIC%20Method%20 v3.0.1%20Measureme nt%20Manual.pdf

[13] P. Fraternali, M. Tisi, A. Bongio, "Automating function point analysis with model driven development", in Procs.of the conference of the Center for Advanced Studies on Collaborative research,(CASCON 06). ACM, 2006, pp. 233-247.

[14] J.D. Gibbons, Nonparametric statistical inference (2nd ed.), New York, Marcel Dekker.

[15] J. Gomez, C. Cachero, "OO-H Method: Extending UML to Model Web Interfaces", IDEA Group Publishing, 2003, pp. 144-173.

[16] IFPUG. 1998. Hints to Counting Web Sites. In IFPUG White Paper.

[17] M.G. Kendall, Rank correlation methods (4th ed.), London, Griffin.

[18] B. Kitchenham, "A Procedure for Analyzing Unbalanced Datasets", IEEE Transac. on Softw. Engineer., vol. 24, n. 4, 1998, pp. 278-301.

[19] B. Kitchenham, L. Pickard, S. MacDonell, M. Shepperd, "What accuracy statistics really measure", IEEE Proceedings – Software, vol. 148, n.3, 2001, pp. 81-85.

[20] B. Marín, O. Pastor, A. Abran, "Towards an accurate functional size measurement procedure for conceptual models in an MDA environment", Data & Know. Engin. vol. 69, 2010, pp. 472–490.

[21] Maxwell, K. Applied Statistics for Software Managers. Software Quality Institute Series, Prentice Hall, 2002.

[22] E. Mendes, B. Kitchenham, "Further Comparison of Cross-Company and Within-Company Effort Estimation Models for Web Applications", in Procs. of International Metrics Symposium (METRICS 04), IEEE Press, 2004, pp.348-357.

[23] N. Moreno, A. Vallecillo, "Towards interoperable Web engineering methods", Journal of the American Society for Information Science and Technology, vol. 59, n. 7, 2008, pp. 1073–1092.

[24] D. Reifer, "Estimating Web Development Costs: There are Differences", in Crosstalk, 2002, pp. 13- 17.

[25] VISUALWADE. 2004. http://www.visualwade.com/

[26] C. Wohlin, P. Runeson, M. Host, M. Ohlsson, B. Regnell, and A. Wesslen. Experimentation in Software Engineering - An Introduction. Kluwer, 2000.