STATISTICAL TESTS ON METRICS FOR THE DESIGN OF BUSINESS COMPONENTS WITH LOW COUPLING AND HIGH COHESION

Dencho N. Batanov[1], Gerrit K. Janssens[2]

[1] Department of Computer Science and Engineering
Frederick University Cyprus, Nicosia, Cyprus
e-mail: com.bd@fit.ac.cy
[2] Transportation Research Institute (IMOB)
Hasselt University, Hasselt, Belgium
e-mail: gerrit.janssens@uhasselt.be

Abstract: In order to satisfy the requirements for business components with low coupling and high cohesion, the identified business activities and business components need to be evaluated on the basis of formal measuring the coupling and cohesion characteristics. A metric is an instrument to devise appropriate measures for software characteristics. In this research, the behaviour of some object-oriented metrics used in measuring the coupling and cohesion of business components is investigated through simulation and statistical analysis.

Key words: business component, component identification, web information system architecture

1. INTRODUCTION

Web-based information systems are widely recognized as one of the most challenging for the development and implementation of modern systems [1]. The challenge stems from the fact that the Web is a completely different computing environment compared to conventional computer-based environments. Basically two types of approaches to Web-based information system development exist. The first is extended and/or modified Web site development as shown in [2]. The approach emphasizes Web site management. Since Web-based information systems are database-intensive systems, such approach focuses not only on presentation, but also on business logic, including databases as well. The second type of approach is based on the premise that a Web-based information system is a product of software engineering (e.g. [3]). These approaches extend or modify existing software engineering approaches for Web-based information systems [4]. While the results from the analysis and design phases for both software engineering and Web engineering of a particular business system are very similar, the software implementation model does not fit the Web implementation model.

The object-oriented paradigm makes use of objects as building blocks but they do not represent directly functional requirements and, therefore, such approach needs further enhancement in order to master dynamic systems. Snoeck et al. [5] states that a business process is divided into function-oriented (activities) and process-oriented parts (relationship between activities). The part, which is function-oriented, represents object behaviour while the process-oriented part acts as a collaboration of these activities. Thus, emphasizing structural aspects is mainly a result of modelling individual object behaviour but neglects identifying functional aspects of the system as a whole. The most crucial issues in modern industrial information systems are flexibility, adaptability, and maintainability. To obtain these qualities, a component-based software development approach implementing business processes with software components is required. Whenever a business process changes, the corresponding software component is also adapted appropriately. Furthermore, any existing software component can be reused/rewired for different enterprises to reduce development time.

A business process-based methodology (BPBM) for developing Web information systems constructed using components has been developed in Arch-int and Batanov [6]. The approach blends advantages of the structured and object-oriented paradigms for identifying and designing components called business components. The central idea of business component modelling is reusability of elementary units, which are business activities. An elementary unit that represents an atomic changeable business process can be implemented with a portable set of Web-based software components. The BPBM should be able to answer the following two research questions: (1) which methodologies are the most suitable for information systems based on a Web environment?, and (2) how do these methodologies provide mechanisms in supporting data interchange? The paper by Arch-int and Batanov provides guidelines for implementing the business component model following the three-tier Web-based information system architecture based on HTTP and XML infrastructure.

A B2B environment focuses on interchanging data among enterprises instead of communication between distributed objects. These enterprises may implement their systems with various kinds of applications, databases, and platforms. An interaction between two distinguished systems based on HTTP can be accomplished using XML documents as distributed objects for interfacing such systems. As the rendering and displaying of an XML document for client machines request no additional software package, universal client devices can access XML-based Web information systems. The Simple Object Access Protocol [7] is an example of the protocol based on XML-enabled Web servers that supports EDI using only HTTP and XML technology.

2. DEFINITIONS AND METRICS SUITES

2.1. Definitions

A metrics suite is conducted to measure the coupling and cohesion of business components at two levels: (1) business activity level and (2) business component level. A result of the metrics represents the degree of coupling or cohesion as guidance for building and generating reusable business activities and business components.

A *business object* (BO) [8] is an object with well-defined boundaries and an identity that encapsulates a business state and behaviour. A business state is a *structural property* represented by *attributes or instance variables* while a behaviour is a *behavioural property* represented by *methods* that operate on the attributes. Business Objects (BOs) represent business resources in a business domain such as employee, product, payment, information, process, etc.

Let *BO* be a business object composed of *m* attributes, $A=\{a_1, a_2, \ldots, a_m\}$, and *n* methods, $M=\{M_1, M_2, \ldots, M_n\}$. In terms of its overall behaviour, a *BO* can be defined as

$$BO \equiv \{M_1(A_1), M_2(A_2), \ldots, M_n(A_n)\} \qquad (1)$$

such that $A_i \subseteq A$, for all $i=1, 2, \ldots, n$. M_i operates on a set of attributes, A_i. Note that A_i and A_j, may contain one and the same elements but are processed by different methods, M_i and M_j, respectively.

A number of business objects participate cooperatively in a particular business process. This cooperation can be represented by sharing interactions between the associated business objects [9]. More specifically, each business object of a particular business object pair of an interaction must provide a public method for interfacing. Such interactions determine business object dependencies.

Traditionally, a software component is defined as *"a self-contained piece of software with well-defined interface or set of interfaces."* [10]. A larger-grained component called a ***business component***, as defined by Herzum and Sims [11], focuses on a business concept as *"the software implementation of an autonomous business concept or business process. It consists of all software artifacts necessary to represent, implement, and deploy a given business concept as autonomous, reusable element of a larger distributed information system."*

No matter how a component is called (*component, distributed component, or business component*), it has the following basic characteristics [10]:

- A component is a self-contained software construct: a self-contained piece of software unit is one that can be independently (autonomously) delivered, installed, and deployed.
- A component socket (plug-out for other components to plug-in) provides a well-defined and well-known run-time interface. This means that it can be easily combined with other components to provide useful functionality.

- A component is built for composition and collaboration with other components.

Coupling Metrics

Briand and Morasca [12] define coupling of a software part in terms of *import* and *export* coupling. Import coupling measures a number of interactions in which a software part interacts with (depends on) the rest of the software system. In contrast, export coupling measures a number of the rest of the software system, which depends on the software part. This definition is similar to that of coupling between two objects defined by Chidamber and Kemerer [13]. *Coupling* refers to the degree of dependence between items, for example, the degree of dependence of a business component on other business components. In our context, coupling of a business activity is identified by the degree of interaction between a business activity and other business activities.

Cohesion Metrics

Cohesion is an important concept in OO-programming. It indicates whether a class represents a single abstraction or multiple abstractions. Chidamber and Kemerer [13] define cohesion of two things to be the internal consistency within parts of design. Cohesion refers to the internal consistency within an item (business component or business activity). In this research, cohesion in a business activity is measured in terms of proportion of internal interactions and external interactions. An internal interaction counts the number of interactions between business objects within the business activity. In contrast, an external interaction counts the number of interactions between business objects within the business activity and business objects of other business activities.

At present, there are two basic approaches, *extensible Web site development* and *extensible software development*, for developing Web-based applications. Each of them has different shortcomings for supporting Web application development. We propose a business component-based approach for such development, which relies on business-oriented, object-oriented, component-oriented, and coupling and cohesion measurements.

The analysis model is a set of conceptual business activities and business components, which are represented using the interaction-based method. The conceptual business activities are then identified using business activity-level metrics into identified business activities. The identified business activities are classified and grouped to form elements of particular business components. It is expected that these basic elements will satisfy both low coupling and high cohesion criteria. Several metrics have been described in detail in [14].

2.2 A Metrics Suite for Business Activity Identification

An associated business object must be identified explicitly by the business activities to which it belongs. An identified business activity is viewed as a modular unit with respect to an atomic software component. More specifically, business objects belonging to each atomic component, implementing a respective business activity should be identified and represented as working elements.

Coupling and cohesion are essential properties in modelling an efficient software module. Here, we consider that the functionality of respective business activities should be in correspondence with requirements for high cohesion within the business component and low coupling with the rest of the components of the system. Metrics to measure degrees of interaction among associated business objects of business components in terms of numerical values in order to gain the essential properties are proposed.

Two suitable sets of metrics for identifying and measuring business activities (Metrics 1-4) and business components (Metrics 5-7) are defined.

Metric 1: *Coupling between Business Objects—Business Object Interaction (CBO-BOI) is the number of interactions between two BO's, which operate on their business state with various kinds of operations.*

Metric 2: *Coupling between Business Activity—Interaction-based (CBA-I) of an identified BA is the proportion of the number of internal interactions (II) and the number of external interactions (EI).*

Metric 2e (Extensible Metric 2): *Extensible Coupling between Business Activity—Interaction-based (eCBA-I) of a BA is the ratio of the number of internal interactions (II) to the number of external interactions (EI). Each interaction is measured by weight assigned to operations.*

Metric 3: *Cohesion in Business Activity in terms of involving Attributes (CBA-A) is the average of the number of involving attributes of each business object of a business activity.*

Metric 3e (Extensible Metric 3): *Extensible Cohesion in Business Activity in terms of all involving Attributes (eCBA-AA) is the maximum of the number of all involving attributes of business objects of a business activity.*

Metric 4: *Lack of Cohesion of Business Activity (LCBA).*

2.3 A Metrics Suite for Business Component Identification

A business component is composed of identified business activities that satisfy both coupling and cohesion measurements. The first step of the technique is to combine a set of identified business activities with *high coupling* into a single business component with the objective of reducing coupling between identified business components. Then, the identified business component is evaluated for a possible split into more business components based on the cohesion measurement.

From the underlying object-oriented perspective, we can define coupling and cohesion measurements of business components as they are defined for business activities.

Metric 5: *Coupling between Business Activity—Business Object Interaction (CBA-BOI) is equal to the number of interactions between BOs of the two business activities.*

Metric 6: *Coupling between Business Component—Interaction-based (CBC-I) of an identified business component (BC) is equal to the proportion of the number of internal interactions (II) and the number of external interactions (EI).*

Metric 7: *Cohesion in a Business Component in terms of involving Attributes (CBC-A) is equal to the average of the number of involving attributes of each business object of a business component.*

Metric 8: *Lack of Cohesion of Business Component (LCBC).*

2.4 The Metric for Business Component Design: Layered Architecture

Metric 9: *The Service of a Business Component (SBC) is equal to the number of services that a considered business component provides for others. These services are represented using external interactions with ordered pair interactions, R-C, R-U, R-R, and U-U.*

3. SIMULATION AND STATISTICAL ANALYSIS

The aim of this analysis is to examine whether the metrics, defined in Section 2, are the proper measurements. This means to prove that they can measure coupling and cohesion as basic characteristics of business activities and business components and that each of them is an independent metric.

For this purpose, five sets of data are generated for simulating five business sample systems by using random numbers. The observed data (sample) for this analysis are simulated data based on a set of rules (criteria and constraints).

3.1 Design of the Simulation Experiment

Assumptions:

Let

- $\{BO_i\}$ be the set of business objects in a given business system *i*,
- $\{A_i\}$ be the set of attributes in a given business object *i*,
- $\{BDU_i\}$ be the set of business data units in a given business object *i,*
- $\{M_i\}$ be the set of methods in a given business object *i,*
- $\{BA_i\}$ be the set of business activities in a given business system *i*, and
- $\{I_i\}$ be the set of interactions in a given business activity *i*.

These sets are finite sets with limited numbers of samples for this analysis. Statistically, a number of samples should not be less than 25 in order to be able to be representative for the population. A set of these sets listed above is a population and an element of the set is a sample of the population. Hence, the number of elements in each set should statistically not be less than 25. However, without restrictions such as data gathering (gathered by simulation) and computer resources (e.g., hard disk capacity/CPU speed), our data are simulated by allowing the number of samples to be more or less as real as the information system. Regarding the purpose, each set can have a maximum number of up to five hundred elements, which is adequate to be statistical representative for the population and to simulate a real information system.

Since these elements are generated by independent uniform random number streams, BO_i, BA_i, I_i, M_i, BDU_i, and A_i are discrete *independent and identically distributed* (iid) random variables. This suggests that the number of attributes, methods, and BDUs of each business object follow a probability distribution based on the respective set of generated random numbers. The same is true for the number of business objects, business activities, and interactions of the system, including the number of interactions for each business activity. Further, the observer cannot predict the attributes or methods, of one business object based on the knowledge of the attributes on methods of another business object in the system.

As the discrete random numbers will be represented by numbers of systems with different sizes, a uniform random number needs to be transformed by weighing it with the maximum size of the respective variable. For example, the maximum number of business objects (*max_bo*), which has been limited to MAX_BO (in our case 500), is generated using the uniform random number (U_I) multiplied with MAX_BO. That is, $max\_bo = U_I$ * MAX_BO, where U_I is a uniform random number in the interval [0,1].

These uniform random numbers are used to generate values using the following steps.

1. Generate the maximum number of business objects, *max_bo,* and generate information for all business objects within a system [{BO_i}, i=1,2,.., *max_bo*].
   1.1 Generate the maximum number of attributes [{A_i}]
   1.2 Generate the maximum number of BDU [{BDU_i}]
   1.3 Generate number of attributes and assign for each BDU
   1.4 Generate the maximum number of methods [{M_i}]
   1.5 Generate the BDU number and assign the BDU for each method

2. Generate the maximum number of interactions, *max_interact,* and generate information for all interactions within a system [{I_i}, i=1,2,.., *max_interact*].

2.1 Generate data for the first argument of an interaction (information from a first business object).

2.2 Generate data for the second argument of the interaction (information from a second business object).

3. Generate the maximum business activities, *max_ba,* and generate a set of interactions for each BA[{BA_i}, i=1, 2, .., *max_ba*].

However, the simulation of data should follow the following restrictions: (1) based on the BPBM approach, there are three kinds of operations, C, U, and R operations, and (2) every business object must have at least one C operation operating on it business state.

3.2 Hypothesis formulation and testing approach

Logistic regression is the most suitable for situations in which one wants to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous (binary values).

In general, a multivariate logistic regression model with a single dichotomous dependent variable *Y* and *n* independent variables is formulated as below

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_n X_n \qquad (2)$$

where

- Y is the *dependent variable* (binary values)
- X_i is an independent variable, *covariates* of the model (*i= 1..n*)
- b_0 is a constant
- b_i is the regression coefficient respective the independent variable X_i .

Hypotheses:

It is hypothesised that the degree of interactions according to the degree of operations can be used to generate business activities and business components with providing low coupling and high cohesion. Recall, the low coupling and high cohesion characteristics are restricted to the meaning as defined in section 2. Thus, the null hypothesis (H_0) is that the method proposed by the BPBM is able to identify business activities and business components.

Recall also that the objective of generating business activities (identified BAs) is to build business components with low coupling and high cohesion. Thus, it should have low coupling between business activities and high cohesion within the business activity itself. Therefore, both of them should be tested with respect to these requirements.

- ***Hypothesis 1: Business Activity Identification***

  H_0: The business activity identification (*Metrics 2 & 3*) *can* be used to generate business activities with low coupling and high cohesion.

  H_1: The business activity identification (*Metrics 2 & 3*) *cannot* be used to generate business activities with low coupling and high cohesion.

The analytical model for the logistic regression model of the hypothesis is given as:

$$bi\_BA = b_0 + b_1 * CBA\_I + b_2 * CBA\_A \quad (3)$$

where

- bi_BA is the *dependent variable* (binary)
- CBA_I is an independent variable representing the coupling between BAs
- CBA_A is an independent variable representing the cohesion within a BA.

- ***Hypothesis 2: Business Component Identification***

  H_0: The business component identification (*Metrics 6 & 7*) *can* be used to generate business components with low coupling and high cohesion.

  H_1: The business component identification (*Metrics 6 & 7*) *cannot* be used to generate business components with low coupling and high cohesion.

The analytical model for the logistic regression model of this hypothesis is given as:

$$bi\_BC = b_0 + b_1 * CBC\_I + b_2 * CBC\_A \quad (4)$$

where

- bi_BC is the *dependent variable* (1 or 0/Yes or No)
- CBC_I is an independent variable representing the coupling between business components
- CBC_A is an independent variable representing the cohesion within a business component.

Data Classification

As the dependent variable takes binary values only, every sample (business activity or business component) should be classified, depending on values of independent variables, into two populations, $\prod_1$: *can identify* and $\prod_2$: *cannot identify.*

The proposed classification is based on the method called *Classification and Regression Trees (CART)* [15] as illustrated in Figure 2.

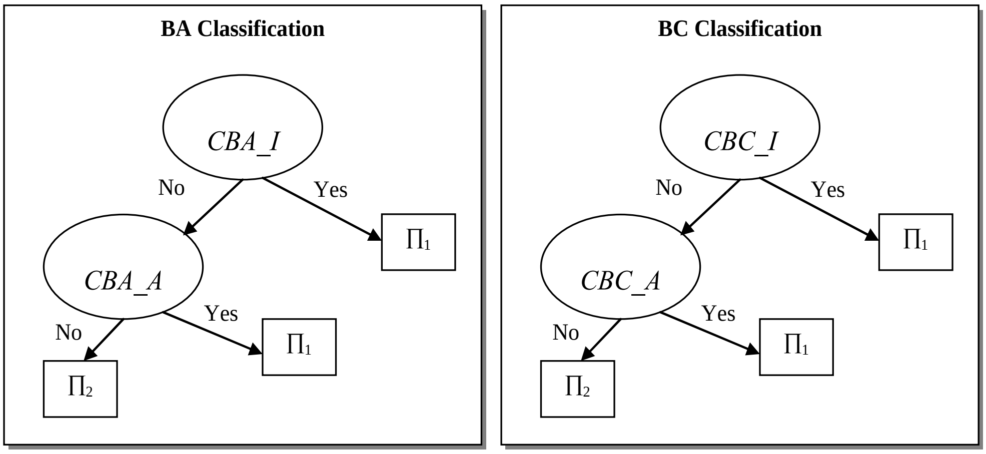

Figure 2: The Data Classification using CART Method

Based on the CART method, two dummy variables, *bi_BA* and *bi_BC*, are created to store the group number of each sample for business activity and business component, respectively, as illustrated in Table 1.

Table 1: The Criteria of Data Classification

Criteria	bi_BA
1. CBA_I ≤ *0.50*	1
2. CBA_A ≥ *AVG*	1
3. Otherwise ❑ CBA_A < *AVG* ❑ CBA_I > *0.50*	0

(a) Business Activity

Criteria	bi_BC
1. CBC_I ≤ *0.50*	1
2. CBC_A ≥ *AVG*	1
3. Otherwise ❑ CBC_A < *AVG* ❑ CBC_I > *0.50*	0

(b) Business Component

3.3 Results of statistical analysis

3.3.1 Descriptive statistics

The basic data and descriptive statistics of simulated data of five systems (A,B,C,D,E) are shown in Table 2 with columns *No. BO* (number of business units), *Attr* (average number of attributes), *Method* (average number of methods), *BDU* (average number of business data units), *No. Interactions* (number of interactions) and *No. BA* (number of business activities).

Table 2: Simulated Data and Their Descriptive Statistics

System	**Business Object**				**No. Interactions**	**No. BA**
	No. BO	Attr[a]	Method[a]	BDU[a]		
A	311	197	191	207	382	236
B	194	190	202	188	359	175
C	178	213	197	203	220	157
D	258	188	190	209	289	185
E	129	215	187	207	372	87

[a] *: The average value per business object*

Statistics on the generated business activities and business components are shown in Table 3, with, in the part 'Identified BA', *BA* (number of business activities), *Int* (average number of interactions), *BO* (average number of business objects), *CBA_I* (average coupling BA-I), *CBA_A* (average cohesion BA-A), *CBA_AA* (average cohesion BA-all A), and, in the part 'Identified BC', *BC* (number of business components), *BO* (average number of business objects), *BA* (average number of business activities), *CBC_I* (average coupling BC-I), *CBC_A* (average cohesion BC-A), *SBC* (average service of BC).

Table 3(a): The Descriptive Statistics of Identified BAs

System	**Identified BA**					
	BA	Int [a]	BO [a]	CBA_I [a]	CBA_A [a]	CBA_AA [a]
A	236	2	1	0.59	0.35	0.73
B	175	2	1	0.52	0.33	0.74
C	168	2	1	0.58	0.34	0.71
D	185	2	1	0.54	0.31	0.69
E	87	4	3	0.71	0.44	0.96
Avg.	**170**	**2**	**1**	**0.59**	**0.35**	**0.77**

Table 3(b): The Descriptive Statistics of Identified BCs

System	**Identified BC**					
	BC	BO [a]	BA [a]	CBC_I [a]	CBC_A [a]	SBC [a]
A	92	2	2	0.86	0.73	0.96
B	58	3	3	0.79	0.67	0.91
C	60	3	3	0.87	0.79	0.99
D	76	2	2	0.79	0.66	0.94
E	19	9	4	0.75	0.65	0.84
Avg.	**61**	**4**	**3**	**0.81**	**0.70**	**0.93**

Regarding a business activity (BA) and a business component (BC) the following rules should be satisfied:

- an identified BA may consist of $\{I_i\}$, $i \geq 1$, and $\{BO_j\}$, $j \geq 0$
- an identified BC may consist of $\{BA_k\}$, $k \geq 1$, and $\{BO_l\}$, $l \geq 1$
- a business object can belong to *several BAs* but must belong to *one and only one BC.*

3.3.2 Correlation

Since the two variables, CBA_I and CBA_A, represent coupling and cohesion, respectively, both explaining variables should not be linearly dependent, which should mean that one of both measures is redundant. It means that low coupling between BAs does not imply that the cohesion is high. The same reasoning is true for CBC_I and CBC_A of business component measurements. To test the independence assumption, their relationship is measured by the Pearson correlation coefficient (*r*). The correlations are shown in the third column of Table 4.

Table 4: Correlation between the two Independent Variables

Metric	**System**	**CBA_A**	**Metric**	**System**	**CBC_A**
BA: CBA_I	A	0.52**	***BC:*** CBC_I	A	0.002
	B	0.49**		B	0.11
	C	0.65**		C	0.09
	D	0.47**		D	-0.17
	E	0.40**		E	-0.10

** *Correlation significant at 99% confidence level*

Results:

- ***BA:*** The relationship between both independent variables at the business activity level of all systems is significant at the 99% confidence level, which is in conflict with the assumption. The main reason is that a business object may *belong to several* business activities so the characteristics (coupling and cohesion) of these business activities do interfere.

- ***BC:*** Most of the correlations in Table 4 are weak. The two metrics, CBC_I and CBC_A, are likely not to be redundant from a predictive point of view.

3.3.3 Univariate Regression Analysis

The goal of this analysis is to test if only a single independent variable can determine the dependent variable or not. More specifically, the method of identification (BAs and BCs) can provide low coupling regardless of cohesion (or can provide cohesion regardless of coupling). The general model of the analysis is formulated below. Table 5 presents the results of this analysis.

Table 5: Results of the Univariate Analysis

Metrics	***System***	b_0	***Regression Coefficient (b_1)***		
			Value	***Significance***	***Exp(b_1)***
CBA-I(BA)	*A*	*6.122*	*-5.711*	*0.002*	*0.003*
	B	*6.486*	*-5.828*	***0.005***	*0.000*
	C	*5.112*	*-4.313*	*0.002*	*0.013*
	D	*5.786*	*-5.404*	*0.002*	*0.004*
	E	*4.652*	*-4.386*	*0.002*	*0.012*
CBA-A(BA)	*A*	*0.414*	*2.895*	*0.000*	*18.082*
	B	*0.897*	*2.759*	*0.002*	*10.831*
	C	*1.035*	*1.888*	*0.002*	*6.609*
	D	*0.592*	*2.845*	*0.000*	*17.194*
	E	*-0.594*	*4.119*	*0.001*	*61.743*
CBC-I(BC)	*A*	*5.104*	*-5.587*	*0.001*	*0.004*
	B	*5.779*	*-6.004*	*0.008*	*0.002*
	C	*3.470*	*-3.609*	***0.025***	*0.027*
	D	*6.829*	*-7.219*	*0.002*	*0.001*
	E	*4.449*	*-5.729*	***0.072***	*0.003*
CBC-A(BC)	*A*	*-6.014*	*9.160*	*0.000*	*9508.437*
	B	*-3.056*	*6.105*	*0.000*	*447.995*
	C	*-4.996*	*7.033*	*0.000*	*1133.767*
	D	*-4.994*	*9.698*	*0.000*	*16277.882*
	E	*-3.626*	*5.680*	***0.049***	*292.997*

Results:

- ***BA*****:** Most systems of the analysis are significant at 0.002 level (except the cases in **bold**). It can be concluded that the BA identification method provides low coupling or high cohesion at 99.98% confidence level. The degrees of significance between coupling and cohesion are not likely to be different for the same reason as the results of correlation testing (see Table 4).
- ***BC*****:** The results are significant at 0.01 level (except the case in **bold**). It can be observed that the degree of significance of cohesion is higher than the degree of significance of coupling since all BOs, which belong to several Bas, are selected to be a single BC. Thus, most attributes (business data) requested by the BC are available.

3.3.4 Multivariate Regression Analysis

The multivariate analysis model for this study is actually a bivariate analysis model of the type

$$Y = b_0 + b_1 X_1 + b_2 X_2 \qquad (5)$$

where X_1 and X_2 are *covariates* of the model. X_1 can be either CBA_I or CBA_A while X_2 can be either CBC_I or CBC_A.

The goal of the multivariate analysis is to determine whether these two metrics (CBA_I and CBA_A or CBC_I and CBC_A) provide low coupling and high cohesion *simultaneously*. The results of the statistical analysis are shown in Table 6.

Results:

- Most systems for both BA and BC multivariate analysis are still significant but they are less significant compared with the univariate analysis. This concludes that the method in identifying provides low coupling between components (BAs or BCs) and at the same time provides high cohesion within a single component. See column with heading *p* in Table 6.
- ***BA*****:** Most systems of this analysis are significant at 0.05 level (except the case in **bold**). It can be concluded that the BA identification method provides low coupling and high cohesion with 95 % confidence.
- ***BC*****:** The results show very high significance at the 0.01 level (except the case in **bold**). It can be concluded that the BC identification method provides low coupling and high cohesion with 99 % confidence.

Table 6: Results of the Covariate Analysis

Metric Level	***System***	***Coefficient*** b_1	***Coefficient*** p	***Metric Level***	***System***	***Coefficient*** b_1	***Coefficient*** P
Business Activity (CBA_I, CBA_A)	*A*	*1.643*	*0.006*	*Business Component (CBC_I, CBC_A)*	*A*	*4.062*	*0.000*
	B	*1.137*	***0.171***		*B*	*2.438*	*0.011*
	C	*1.408*	*0.062*		*C*	*3.303*	*0.001*
	D	*2.088*	*0.013*		*D*	*8.696*	*0.020*
	E	*2.043*	*0.062*		*E*	*2.369*	***0.296***

4. CONCLUSION

A metrics suite for automatically identifying and generating business activities and business components in order to provide low coupling and high cohesion properties is proposed. The results of statistical analysis have shown that the proposed metrics can be used to identify and generate business activities and business components providing low coupling and high cohesion with high confidence.

However, to guarantee that these metrics behave similarly across various application domains, the data simulation procedure and the constraints or rules should be appropriate with the particular domain.

REFERENCES

[1] Huang G.Q., Mak K.L., Issues in the development and implementation of Web applications for product design and manufacture, *International Journal Computer Integrated Manufacturing* **14** (1), 2001, pp. 125–135.

[2] German D.M., Cowan D.D., Towards a unified catalog of hypermedia design patterns. *Proceedings of the Annual 33rd Hawaii International Conference on System Science,* 2000, pp. 6067-6073, Washington DC, IEEE Computer Society.

[3] Gellersen H.W., Gredke M., Object-oriented Web application development, *IEEE Internet Computing* **3** (1), 1999, pp. 60–68.

[4] Chen J.Q., Heath R.D., Building Web applications challenges, architectures, and methods, *Information Systems Management* **18** (1), 2001, pp. 68–79.

[5] Snoeck M., Poelmans S., Dedene G., An architecture for bridging OO and business process modelling. *Proceedings of the IEEE International Conference on Technology of Object-Oriented Languages and Systems (TOOLS'00),* 2000, p. 132.

[6] Arch-int S., Batanov D.N., Development of industrial information systems on the Web using business components, *Computers in Industry*, **50**, 2003, pp. 231-250.

[7] World Wide Web Consortium (W3C), Simple Object Access Protocol (SOAP) 1.1, a submission to the W3C on 8 May 2000 is accessible at URL http://www.w3.org/TR/2000/NOTE-SOAP-20000508/.

[8] OMG-Business Object Domain Task Force, *Business Object Concepts*, White Paper, January 1999, OMG-document bom/99-01-01.

[9] Hartel, P., Specifying business processes over objects, entity-relationship approach: business modeling and re-engineering*, Proceedings of the 13th International Conference on the Entity-Relationship Approach,* Manchester, U.K., December 13-16, 1994.

[10] Brown A.W., Wallnau K.C., The Current State of CBSE*, IEEE Software*, **15** (5), 1998, pp. 37-46.

[11] Herzum P., Sims O., *Business Component factory: A Comprehensive Overview of Component-Based Development for the Enterprise,* Wiley & Sons, 2000.

[12] Briand L.C., Morasca S., Defining and Validating Measures for Object-Based High-Level Design, *IEEE Transactions on Software Engineering*, **25** (5), 1999, pp. 722-743.

[13] Chidamber S.R., Kemerer C.F., A Metric Suite for Object Oriented Design, *IEEE Transactions on Software Engineering*, **20** (6), 1994, pp. 476-493.

[14] Batanov D.N. , Janssens G.K., Metrics for the design of business components with low coupling and high cohesion, *Journal of Research in Engineering and Technology (Thailand)*, **6**(2), pp. 127-150.

[15] Breiman L., Friedman J., Olshen R., Stone C., *Classification and Regression Trees.* Belmont, CA: Wadsworth, 1984.

Information about authors:

Dencho N. Batanov received his PhD degree from the Technical University, Sofia, Bulgaria. He has been a professor in the Program of Computer Science and Information Management at the Asian Institute of Technology, Bangkok, Thailand. Currently he is professor at the Dept. of Computer Science and Engineering at the Frederick University in Nicosia, Cyprus. His research interests are application of information technologies in education, management, industrial engineering and manufacturing, knowledge-based and expert systems, object-oriented software engineering, distributed systems and technologies, component and framework-based software development, and e-business. He has published more than 100 papers in journals and conference proceedings, and several textbooks and manuals.

Gerrit K. Janssens holds a Ph.D. in Computer Science from the Free University of Brussels (VUB). Currently he is Professor of Operations Management and Logistics at the Hasselt University, Belgium within the Faculty of Business Administration. He also holds the CPIM certificate of the American Production and Inventory Control Society (APICS). During the last eighteen years he has been several times visiting faculty in universities in South-East Asia and in Southern Africa. His main research interests include the development and application of operations research models in production and distribution logistics.