SCORING OF CORPORATE WEBSITES USING PARTIAL LEAST SQUARES PATH MODELING

PAVOL KRÁĽ, PETER LACO
Matej Bel University, Faculty of Economics,
Department of Quantitative Methods and Information Systems,
Tajovského 10, Banská Bystrica, Slovakia
e-mails: pavol.kral@umb.sk, peter.laco@umb.sk

Abstract

World Wide Web, especially corporate websites are used by businesses to publish contacts, billing and service information, and product offerings. Businesses can also engage in interactive communication through questionnaires, discussion forums, and on-line news to help them react very flexibly and provide feedback to their customers. In practice, however, not all businesses use these options effectively and the level of quality on business websites varies wildly. Consequently, a flexible and easy applicable methodology providing companies with a possibility to evaluate their own presence would be beneficial. There are many mathematical, statistical and decision-making methods which could form its core. In our contribution, we create such a methodology for scoring of corporate websites based on partial least squares path modeling. The proposed methodology is intended for application in the conditions of Slovak SMEs.

Keywords*: internet, corporate website, partial least squares path modeling*

JEL Codes*: D83, M15, O33*

1. Introduction

Small and medium-sized enterprises (SMEs) are currently facing various challenges and the Internet is one of them. Its ability to spread information about the enterprise, its products and services is virtually unlimited. Today's technologies enable both current and potential customers to view, order, or even purchase these products using their computers directly from home and in recent years using their phones virtually anywhere. Thanks to the generally accepted internet standards, business processes are simplified and there is an increase in the automation of supplier-customer relationships. It also improves in-house communication and the sharing of information and knowledge throughout the enterprise.

There are also risks associated with the use of this phenomena especially security of information, the risk of its misuse, and the dissemination of false or misleading information damaging to a business's reputation. The risk of greater competition stems from the very nature of the internet - a free open network that can be used by virtually any business, regardless of its size or geographic location. As stated by Lesáková (2008, p. 607) "The future success of small and medium enterprises in the new world of global economics will be determined by … implementation of innovation, information and communication technologies by SMEs".

Most businesses use internet services primarily for the World Wide Web in the form of a business websites where businesses can publish contacts, billing and service information, and product offers. Businesses can also engage in interactive communication through questionnaires, discussion forums, and on-line news to help them react very flexibly and provide feedback to their customers. In practice, however, not all businesses use these options effectively and the level of quality on business websites varies wildly.

Consequently, a flexible and easy applicable methodology providing companies with a possibility to evaluate their own web presence would be beneficial. There are various mathematical, statistical and decision-making methods that could be used. Therefore, our main goal is to design a way of evaluating corporate websites with regard to the specific needs of small and medium-sized enterprises. Our intention is to create such a methodology for scoring of corporate websites based on partial least squares path modeling. The proposed methodology is intended for application in the conditions of Slovak SMEs. Our partial goal is to select a sample of Slovak SMEs and apply proposed methodology on sample data.

During recent years, different website evaluation approaches have been introduced. These deal mostly with website usability (Alva, Martinez, Cueva, 2003; Tezza, Bornia, Andrade, 2011), but also content (Robbins, Stylianou, 2003), website accessibility (Leporini, Paterno, 2003), design (Palmer, 2002), functionality, or technical aspects of websites. The complex review of the literature from 1995–2006 was presented by Chiou, Lin, Perng (2010), who identified that in 41% of studies technology-oriented factors were used to evaluate websites, such as navigability, accessibility, usability, or information quality. In 12% of studies marketing factors were used, like advertising, customer service, promotion, order confirmation, or online transaction. In the most of studies (47%) a combine framework as a mixture of both information system and marketing approaches was used. Chiou, Lin, Perng (2010) also investigated used research methodologies, that were survey (42%), experiment (23%), content analysis (17%), case study (10%), automatic evaluation (5%) and concept development (3%).

In our previous research (Kollár, Kráľ, Laco, 2014; Kollár, Kráľ, Laco, 2015; Laco, Kallová, 2017) we focused on identification of criteria that are crucial for websites evaluation. We identified these criteria and we verified their importance for web users using questionnaires. In this article we introduce a new methodology that was not applied to this field yet and we use it on data of Slovak SMEs websites.

2. Methodology

Our scoring is based on Partial Least Squares Path Modeling (Sanchez, 2013) which is an alternative approach to Structural Equation Modeling. It allows us to generate scores corresponding to latent (not directly measurable) variables representing some general concepts using manifest (directly measurable) variables and therefore to create scoring based on some rather abstract concepts, e.g. quality. A model in PLSPM consist of two parts – an inner one and an outer one. The inner part describes relationship among assumed latent variables, the outer part connects latent variables and the corresponding manifest variables. In the paper, we restrict ourselves to reflexive manifest variables, i.e. we assume that directly measurable variables simple reflects (are caused by) latent variables. One set of manifest variables corresponds to each of latent variables. More formally (Sanchez, 2013), we assume a data set X consisting of n observations and p variables which can be divided in J mutually exclusive blocks X_1, X_2,…, X_J and each of these blocks is associated with a latent variable LV_j. We denote estimation of this latent variable by Y_j. For the inner model we assume:

$$LV_j = \beta_0 + \sum_{i \to j} \beta_{ji} LV_i + \varepsilon_j, \tag{1}$$

where latent variables LV_i predict LV_j, β_{ji} are path coefficients and $\operatorname{cov}\left(LV_j, \varepsilon_j\right) = 0$.

For the outer model it holds:

$$X_{jk} = \lambda_{0\,jk} + \lambda_{jk} LV_j + \varepsilon_k, \tag{2}$$

where λ_{jk} are called loadings and X_{jk}, denotes the *k*-th manifest variable in the *j*-th block. Scores Y_j corresponding to a latent variable LV_j are then estimated as follows:

$$Y_j = \sum_k w_{jk} X_{jk}, \tag{3}$$

where w_{jk} denotes an outer weight corresponding to a manifest variable X_{jk}.
The outer weights w_{jk} are determined as follows. First, we assign to each variable X_{jk} a weight equals one. Then these weights are transformed to $\tilde{w}_{jk}$ such that the resulting scores have a unit variance. Then it holds

$$Y_j = \pm f_j \sum_k \tilde{w}_{jk} X_{jk}, \tag{4}$$

where f_j a number standardizing Y_j and we choose the sign for each weight such that the majority of X_{jk} is positively correlated with the resulting Y_j.

From the computational point of view, PLSPM model fitting is an iterative process which consists of three stages:

1. determination of weights and corresponding scores of latent variables,
2. estimation of path coefficients,
3. estimation of loadings.

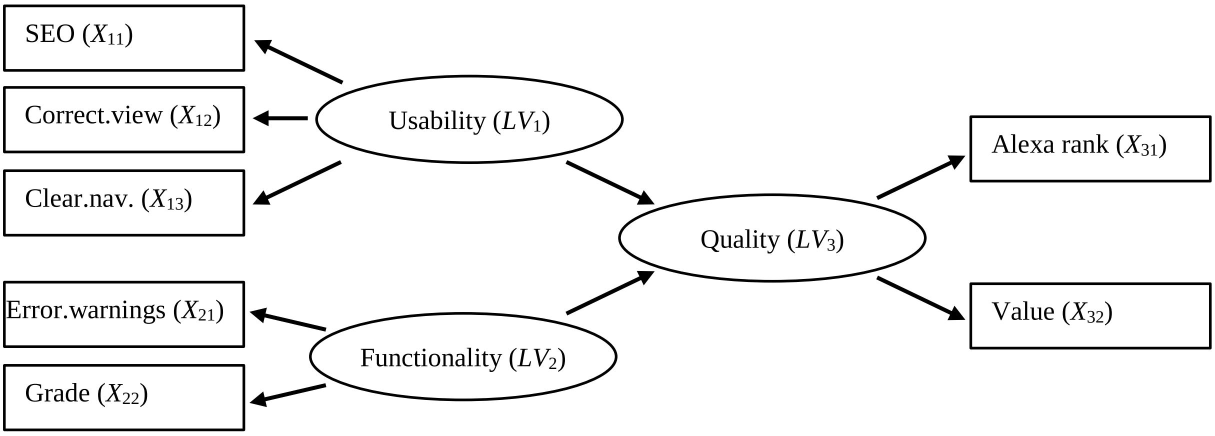
3. Case study of websites' evaluation

We decided to apply described methodology on websites of Slovak SMEs of chosen sector and geographical region. We have chosen websites that sell electronics in Banska Bystrica region and they are accessible for visitors via Google search. Websites have been included in our sample based on their appearance in Google search results, keywords: elektro predaj (electronics sell) realized from computer located in Banská Bystrica. Due to these conditions, websites that sell electronics in Slovakia (mostly from Banská Bystrica) appeared as search results. Google AdWords results were omitted, and only organic search results were taken into consideration. We investigated first 35 results what corresponds to first 3 and half page of standard Google search results. We consider this number of results appropriate due to the fact that click-through rate fall rapidly after first ten results (www.advancedwebranking.com/cloud/ctrstudy), so probability that visitors see website that is below position 35 is rather low (aprox 5%). Some results (10) had to be excluded due to irrelevant content. Another 4 websites were excluded due to lack of data, so at the end we gathered data describing 21 corporate websites.

3.1 Data

Our model is depicted on path diagram that can be seen on Figure 1. Manifest variables are represented in a rectangular form, latent variables are represented in an elliptical form and relationships between variables are represented with straight arrows.

Figure 1: Websites evaluation model.



Source: The author's work.

The inner model consists of three latent variables that cannot be measured directly and relationships among them. We assume that overall quality of website is based on usability and functionality.

The outer model consists of relationships between individual latent variables and manifest variables they depend on. Based on theoretical approaches to website evaluation listed above, we assume that usability depends on following factors.

Search Engine Optimization (in our model as SEO) is crucial for users to be able to find the corporate website on internet. SEO is a marketing discipline focused on growing visibility in organic (non-paid) search. According to Statcounter GlobalStat (gs.statcounter.com, 10.5.2018) the most searches Worldwide were 90% to 92% of all searches during last 12 months (in Slovakia it was 96% to 97%). That is why we measured level of SEO as a number of position of specific website in search results of Google search from May 9, 2018, keywords: elektro predaj (electronics sell) realized from computer located in Banská Bystrica.

Correct view of webpages (in our model as Correct.view) is another important aspect for web users. Especially today corporate websites are displayed on various devices with totally heterogeneous screen resolutions. That is why it is very important for corporate websites to be displayed correctly at various screen resolutions. In our model we evaluated this factor from scale 0 to 3. Value 0 (worst) was assigned for websites that were displayed correctly only at traditional desktop screen resolution and 3 (best) was assigned for websites that were displayed correctly at all screen resolutions, that are used according to Statcounter GlobalStat (gs.statcounter.com, 10.5.2018). Value 2 was assigned to websites with some minor errors at some resolutions and value 1 was assigned to websites with major errors at most resolutions.

Clear navigation (in our model as Clear.nav.) is important for users that came to corporate website. In our model we evaluated this factor from scale 0 (worst) to 3 (best). We considered if there is a main navigation available (at visible part of website in browser), if it is clear where in the structure of a website is a specific webpage, if there is a site search available.

As for the functionality, we assume that it depends on the fact whether web standards are followed and speed of website. We measured the level of web standards followed by website as a number of errors and warnings (in our model as Error.warnings.) of HTML validator (validator.w3.org). The speed of website was measured by Pingdom Website Speed Test (tools.pingdom.com). To achieve objectivity, we used the Google PageSpeed performance grade (in our model as Grade), which may range from 0 (worst) to 100 (best). It is described in more details at developers.google.com/speed.

The overall quality of website was measured by Alexa Traffic Rank and estimated value of the website. Alexa Traffic Rank (in our model as Alexa rank) is an estimate of the site's popularity. It is calculated using a combination of average daily visitors and pageviews of specific website over the past 3 months. We gained data for our research from www.alexa.com/siteinfo. The estimated value of the websites was gained from various servers that provide this kind of service: siteprice.com, siteprice.org, yourwebsitevalue.com, worthofweb.com, websiteoutlook.com. During the research it turned out that not all servers were able to estimate the value of the investigated website, so in the final phase we decided for worthofweb.com that was able to calculate estimation for all website's value.

The complete dataset is available at www.ef.umb.sk/amse2018. The corresponding correlation matrix containing Pearson correlation coefficients is listed in Table 1.

Table 1: Correlation matrix of manifest variables

	SEO	Error.warnings	Grade	Correct.view	Clear.nav.	Alexa rank	Value
SEO	1.00	0.44	0.44	0.58	0.22	0.37	0.21
Error.warnings	0.43	1.00	0.11	0.35	0.06	0.30	0.09
Grade	0.44	0.11	1.00	0.20	0.27	0.38	-0.17
Correct.view	0.58	0.35	0.20	1.00	0.44	0.51	0.13
Clear.nav.	0.22	0.06	0.27	0.44	1.00	0.51	0.12
Alexa rank	0.37	0.30	0.38	0.51	0.51	1.00	0.14
Value	0.21	0.09	-0.17	0.13	0.12	0.14	1.00

Source: The author's work.

3.2 Software

We used statistical system R 3.4.1 (R Core Team, 2017) and R package `plspm` (Sanchez et. al., 2017) for fitting models.

3.3 Results

Our model is summarized in Tables 2 – 7.

Table 2: Total effects of latent variables

Relationships	Direct	Indirect	Total
usability -> functionality	0.000	0	0.000
usability -> quality	0.564	0	0.564
functionality -> quality	0.450	0	0.450

Source: The author's work, R output of the model.

In Table 2, relationships between latent variables (usability, functionality, quality) is listed. We can see that there is a medium positive relationship between usability and quality of a web page and between functionality and quality of a web page, respectively. The observed effects correspond to our expectations. Nevertheless, relationship between functionality and quality needs further explanation. Our model does not take into account complexity of websites and we can expect that more complex websites are more valuable but at the same time include more errors and departures from W3C standards.

Table 3: Unidimensionality of model blocks

Block	Mode	MVs	C.alpha	DG.rho	Eig.1st	Eig.2nd
Usability	A	3	0.677	0.824	1.84	0.791
Functionality	A	2	0.203	0.715	1.11	0.887
Quality	A	2	0.242	0.725	1.14	0.862

Source: The author's work, R output of the model.

Values of Cronbach's reliability coefficient much smaller than 0.7 indicate that latent variables Functionality and Quality could not be reliably measured. However, values of Dillon-Goldstein's rho statistics are greater than 0.7 for each of the blocks, i.e. each block is a good representation of it corresponding latent variable. "This index is considered to be a better indicator than the Cronbach's alpha because it takes into account to which extent the latent variable explains its block of indicators" (Sanchez, 2013). Having the first eigenvalues significantly bigger than one and the second smaller than one for usability is consistent with unidimensionality. On the other hand, for Functionality and Quality these values are quite close to one indicating that the corresponding manifest variables are too weakly correlated to assume unidimensionality.

Table 4: Outer model

Block	Weight	λ_{jk}	Communality	Redundancy
Usability				
SEO	0.532	0.900	0.810	-
Correct.view	0.395	0.879	0.773	-
Clear.nav.	0.290	0.600	0.360	-
Functionality				
Error.warnings	0.501	0.684	0.467	-
Grade	0.752	0.874	0.763	-
Quality				
Alexa rank	0.937	0.984	0.968	0.816
Value	0.184	0.426	0.181	0.153

Source: The author's work, R output of the model.

Values of loadings (λ_{jk}) are satisfactory, i.e. greater or equal to 0.7, for the variables SEO, Correct.view, Grade and Alexa rank indicating a close relationship between individual variables and their corresponding latent variables. Less satisfactory are loadings of variables Clear.nav. and Error.warnings. Comunality (λ_{jk}^2) shows the worst representation of the variable Value as only 18 % of its variability can be explained by variability of the latent variable quality. It indicates that our model may be omitting some key factors crucial for determining the financial value of a website.
Redundancy (computed as $\lambda_{jk}^2 \cdot R^2$) measures fraction of the variance of an indicator in an endogenous block that is predicted by the independent latent variables associated to that block, i.e. functionality and usability predict about 82 % variability of Alexa rank, but only about 15 % of Value. By averaging redundancies of indicators, we get a mean redundancy which represents fraction of overall variability of indicators explained by independent latent variables (see Table 7).

Table 5: Loading and cross-loadings

Block	Usability	Functionality	Quality
Usability			
SEO	0.900	0.669	0.864
Correct.view	0.879	0.429	0.641
Clear.nav.	0.600	0.377	0.472
Functionality			
Error.warnings	0.417	0.684	0.496
Grade	0.566	0.874	0.744
Quality			
Alexa rank	0.851	0.854	0.984
Value	0.288	0.05	0.426

Source: The author's work, R output of the model.

Table 5 allows us to compare loadings of indicators and their cross-loadings, i.e. correlations with latent variables not corresponding to them. It indicates that all our exogenous variables are correctly assigned to blocks corresponding to respective latent variables and none of them should be reassigned to other block of variables.

Table 6: Coefficients of the inner model

Parameter	Estimate ($\widehat{\beta_{jl}}$)	SE	t	p-value
Usability	0.564	0.1208	4.67	0.000191
Functionality	0.450	0.1208	3.73	0.001539

Source: The author's work, R output of the model.

As we can see in Table 6Table 5 allows us to compare loadings of indicators and their cross-loadings, i.e. correlations with latent variables not corresponding to them. It indicates that all our exogenous variables are correctly assigned to blocks corresponding to respective latent variables and none of them should be reassigned to other block of variables.

Table 6, both Usability and Functionality have statistically significant coefficients. The pseudo goodness-of -fit statistics takes an acceptable value 0.722. Overall (see Table 7), 84 % of variability of the latent variable Quality was explained by our model, 57.5 % of variability of the corresponding block are explained by the latent variable Quality and 48.5 % of variability of indicators in an endogenous block is explained by usability and functionality. These results indicate that our model is too simplistic.

Table 7: Summary of inner model

Variable	Type	R^2	Block communality	Mean redundancy
Usability	Exogenous	-	0.648	-
Functionality	Exogenous	-	0.615	-
Quality	Endogenous	0.843	0.575	0.485

Source: The author's work, R output of the model.

Using bootstrap with 200 resamples of our sample of SME's websites, we got the following bootstrap confidence intervals for model parameters presented in the above tables.

Table 8: Validation via 95 % two-sided bootstrap confidence intervals

Measure	Original	Mean.Boot	LB	UB
Weights				
SEO*	0.532	0.536	0.394	0.746
Correct.view	0.395	0.332	-0.159	0.465
Clear.nav.	0.290	0.134	-0.516	0.415
Error.warnings	0.501	0.503	-0.441	0.811
Grade	0.752	0.631	-0.581	0.917
Alexa rank*	0.937	0.809	0.520	0.979
Value*	0.184	0.288	0.081	0.492
Loadings				
SEO*	0.900	0.926	0.818	0.990
Correct.view	0.879	0.723	-0.438	0.972
Clear.nav.	0.600	0.367	-0.921	0.902
Error.warnings	0.684	0.704	-0.640	0.947
Grade	0.874	0.811	-0.747	0.964
Alexa rank*	0.984	0.974	0.945	0.997
Value*	0.426	0.640	0.292	0.992
Paths				
Usability →Quality*	0.564	0.630	0.133	0.984
Functionality→Quality	0.450	0.346	-0.221	0.830

Note: * denotes significant weights, loadings or paths.
Source: The author's work, R output of the model.

4. Conclusion

Importance of corporate websites is indisputable. We explained their pros and cons as well as approaches to their evaluation in the introduction part. In methodology we introduced the Partial Least Squares Path Modeling that was later used for our analysis. The crucial part of article is case study of evaluation of chosen sector and geographical region Slovak SMEs. We described the model that we proposed, data we used, the way how they were measured, used software and the results.

Using PLS Path modelling we fitted a simple model aimed to generate a scoring of corporate websites. The model consists of three latent variables, namely Usability, Functionality and Quality of a website and seven reflective (exogenous) directly measurable variables (SEO, Correct.view, Clear.nav., Error.warnings, Grade, Alexa rank, Value). We eliminated the variable Language version from our final model as it was not statistically significant, and it had very small loading and weight. We examined importance of particular variables in the corporate website evaluation process via loadings and weights. Moreover, as scores corresponding to latent variables are generated as a part of a model fitting, we can, in theory, rank corporate websites based on their usability, functionality and overall quality. Although we applied bootstrapping here, the presented results are rather preliminary and descriptive, without possibility to make inferential statements outside the sample at hand. Nevertheless, our analysis revealed that the ability of the selected variables to determine a value of a web-page was rather limited and the main part of the proposed model was falsified. Consequently, future research regarding identification of the proper variables and alternative formulations of the valuation model is vital.

Our analysis also indicates some potential problems in our future research. First, the reviewed literature recommends more complex set of factors that should be taken into the consideration, but they are difficult to measure. The second problem is data. It was time consuming to gather individual values for each factor for each corporate website, so we plan to apply for a grant to find resources to realize more complex research.

Acknowledgements

The support of the grant scheme KEGA 018UMB-4/2018 Couching approach as a new form of critical thinking development of students in higher education is gladly acknowledged.

References

[1] Advanced Web Ranking. 2018. Click-Through Rates Study [cit. 10-06-2018] https://www.advancedwebranking.com/cloud/ctrstudy/.

[2] Alva, M., Martinez, A., Cueva, J. 2003. Comparison of methods and existing tools for the measurement of usability in the web. In: Web Engineering ICWE 2003, pp. 386-389. ISSN 0302-9743.

[3] Statcounter GlobalStat. 2018. Search Engine Market Share Worldwide. Screen Resolution Stats Worldwide [cit. 10-05-2018] http://gs.statcounter.com.

[4] Chiou, Cw., Lin, Cc., Perng, C. 2010. A strategic framework for website evaluation based on a review of the literature from 1995–2006. Information & Management, vol. 47, iss. 5-6, pp. 282-290. DOI: 10.1016/ j.im.2010.06.002.

[5] Kollár, I., Kráľ, P., Laco, P. 2014. Methodology for Assessing Website Improvement in Corporate Environment In: Conference Proceedings of 17th International Scientific Conference "Applications of Mathematics and Statistics in Economics", Wrocław: Wroclaw University of Economics, pp. 131-140. ISBN 978-83-7695-421-9.

[6] Kollár, I., Kráľ, P., Laco, P. 2015. Measuring the Level of Corporate Internet Presentation in Relation to Economic Performance. In: Conference Proceedings of 18th International Scientific Conference "Applications of Mathematics and Statistics in Economics", Prague: University of Economics. ISBN 978-80-245-2099-5.

[7] Laco, P., Kallová, M. 2017. Factors influencing corporate websites' value. In: Conference Proceedings of 20th International Scientific Conference "Applications of Mathematics and Statistics in Economics", Wrocław: Wroclaw University of Economics, pp. 275-284. ISBN 978-83-7695-693-0.

[8] Lesáková, Ľ. 2008. Influence of globalisation on small and medium enterprises. Ekonomický časopis. vol. 56, iss. 6, pp. 607-621.

[9] Leporini, B., Paterno, F. 2003. Criteria for usability of accessible web sites. In: Universal Access: Theoretical Perspectives, Practice, and Experience, pp. 43-55. ISSN: 0302-9743.

[10] Palmer, J. W. 2002. Web site usability, design, and performance metrics. Information Systems Research. pp. 151-167. DOI: 10.1287/isre.13.2.151.88.

[11] R Core Team. 2017: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, http://www.r-project.org/.

[12] Robbins, S., Stylianou, A. 2003. Global corporate web sites: an empirical investigation of content and design. Information & Management, vol. 40, iss. 3, pp. 205-212. DOI: 10.1016/S0378-7206(02)00002-2.

[13] Sanchez, G. 2013. PLS Path Modeling with R Trowchez Editions. Berkeley. 222 pp. http://www.gastonsanchez.com/PLS Path Modeling with R.pdf.

[14] Sanchez, G., Trinchera, L., Russolillor, G. 2017. plspm: Tools for Partial Least Squares Path Modeling (PLS-PM). R package, version 0.4.9, http://cran.r-project.org/web/packages/plspm/index.html.

[15] Tezza, R., Bornia, A., De Andrade, D. 2011. Measuring web usability using item response theory: Principles, features and opportunities. Interacting With Computers. pp. 167-175. DOI: 10.1016/j.intcom.2011.02.004.