ISSN 1748-7595 (Online)

Kent Business School

Working Paper Series

A Structural Model for Revenues in E-Commerce

Cecilio Mar Molinero, Kent Business School
C Serrano-Cinca, University of Saragossa, Spain
Y Fuertes Callen, University of Saragossa, Spain

KBS
UNIVERSITY OF KENT

Working Paper No.118
April 2006

A structural model for revenues in e-commerce

By:

*C. Serrano-Cinca

Department of Accounting and Finance
University of Saragossa, Spain.
serrano@unizar.es

Y. Fuertes Callén

Department of Accounting and Finance
University of Saragossa, Spain.
yfuertes@unizar.es

C. Mar Molinero

Kent Business School
University of Kent, UK.
C.Mar-Molinero@kent.ac.uk

January 2006

**Address for correspondence: C. Serrano-Cinca:* Department of Accounting and Finance, Fac CC Económicas y Empresariales, Univ. Zaragoza, Gran Vía 2, Zaragoza (50.005) SPAIN serrano@unizar.es

A structural model for revenues in e-commerce

ABSTRACT

We suggest an explanatory model for the financial results of firms that operate in the Internet. The model has been built within a structural modeling framework. Non-financial information has been used to identify several intangible constructs that are relevant in order to explain financial results: "potential customers", "relevance in search engines", "World Wide Web popularity", and "renown in on-line media and social networks". The model has been estimated with a sample of USA e-tailers, and it has been possible to test a variety of hypothesis. The results show that there is a significant relationship between the intangible constructs and the financial results, and that this relationship is stronger when the firm operates mainly in the Internet.

KEYWORDS

Electronic commerce, dot com, non-financial indicators, intangible assets, structural equations, Partial Least Squares (PLS).

A structural model for revenues in e-commerce

1. INTRODUCTION

In this paper we study the electronic commerce (e-commerce) industry and the factors that affect sales and profits. This we do by estimating a Structural Equation Model (SEM) estimated by means of Partial Least Squares (PLS). When building structural equation models it is common practice to use surveys as the source of information. In our case, the data set consists of a set of published observed variables taken from virtual shops. The success of e-commerce is measured through items derived from financial statements.

E-commerce has undergone a series of steps from the high expectations of its initial stages, through its youth with high growth but no profits, to a crisis, finally arriving at a maturity level when standard analytical methods are becoming relevant. The information that has dominated in the analysis of an electronic commerce firm (e-tailer, dot.com) has changed in each one of such stages.

The origins of e-commerce were characterized by a madness of youth. There were great expectations but no profits. Little or no information was available in order to assess the worth of a firm; all that one could go on was the name of the firm and a statement of its mission. A mere change in the name of the firm could produce abnormal returns; Cooper et al (2001). Performance in the markets was not matched with actual profit results. Using data from the year 2000, Serrano-Cinca et al (2005) observed a paradoxical Pearson correlation coefficient of -0.836 between sales and profits, suggesting that the higher the sales in e-commerce, the higher the losses. The explanation may be that firms were positioning themselves in the market, aiming at obtaining market share rather than profits. Indeed, they were prepared to sustain losses in order become household names. Such behavior cannot be long sustained and, after the crisis, using data from 2001, the same correlation had become much smaller (-0.100), not significantly different from zero.

In the absence of profits, analysts had to take into account a different item in the accounts: sales. But sales can be massaged using methods that, having no impact on net profits, give a positive image of the company. Examples are reciprocal transactions between

firms aimed at fattening at the same time sales and expenditure; Alpert (1999), Bowen et al. (2002) y Davis (2002).

Aware of the limitations, or absence, or accounting information, analysts became understandably interested in non-financial information. Livnat and Zarowin (1990), Black (1998), and Jorion and Talmor (2001) argued that in the first stages of the life of a company, it is important to focus on variables that capture growth potential, such as investment in R&D. As firms grow, they spend on publicity, plant, services and logistics, while generating income at the same time. As firms mature, the value of fixed assets increases. They conclude that, in the early stages of a firm, non-financial information is of great relevance. Such non-financial information, in the case of dot com firms, takes the form of web-metrics. Web-metrics are indicators of activity such as the number of visitors to the firm's web page, the number of pages visited, and the amount of time that the visitor spends on the page. The expectation is that web-metrics are indicators of future sales, profits, and market prices; Trueman et al (2000, 2001, and 2003), Demers and Lev (2001), Hand (2001), Rajgopal et al (2003), and Nikolaeva (2005). It can be argued that the importance of web metrics was exaggerated, and that the astronomical prices paid for some firms on the basis of web page visits were not justified, as visitors did not always become customers. The result was a crash in the sector in the year 2000.

After the crash, analysts became again interested in financial information. The study of the relationship between financial and non-financial information continued, but the results were not conclusive. Rajgopal et al (2003), using a sample of e-commerce firms, show that web metrics complement traditional accounting information in the explanation of market prices, but Keating and al (2003) argue that, in the Internet sector, traditional financial information market prices are better explained by accounting information than by non-financial information.

In this paper we look behind the relationship between web metrics and profits or market prices, by exploring the causes of such visits. This we do by building a structural model that contains a series of constructs such as the positioning of the firm in search engines, its renown in on-line social networks and in electronic media, and popularity in the World Wide Web. The constructs are non-financial latent variables (intangibles), calculated from a selection of non-financial indicators such as the number of web pages that link to a virtual shop, or the number of posts that appeared in blogs and that make reference to the company. In the model we propose a firm has to obtain visitors to its virtual shop- potential customers-

in order to sell in the Internet. This is achieved by becoming visible in the World Wide Web, in the blogsphere, in search engines, and in the on-line media.

The model has been estimated using data from 400 leading USA e-tailer firms. The sample has been divided into two sub-samples, the first sub-sample contains dot com firms, and the second sub-sample contains "*brick and click*" firms. We use Structural Equation Models estimated using the technique of Partial Least Squares; Bollen (1989), Chin (2001).

The next section describes the model and formulates the hypotheses. The selection of appropriate indicators for each construct is discussed next. This is followed by an analysis of the results obtained in the measurement model. A subsequent section concentrates on the structural model and the results of the hypotheses tests. The paper ends with a concluding section.

2. MODEL ENTERTAINED AND HYPOTHESES.

Many papers are concerned with the success and failure of electronic commerce. Quaddus and Achjari (2005) give a literature review of 23 studies in this area, of which 12 are concerned with theoretical considerations, 8 are surveys, and the rest are case studies, and interviews. Quantitative information is most of the times analyzed by means of SEM, an example being the paper by Torkzadeh and Dhillon (2002)- further developed by Chang et al (2004). This paper takes the work of Keeney (1999) to propose a set of intangible constructs- such as "trust"- and a set of measurable indicators- derived from survey questions aimed at either consumers or managers- from which the constructs are derived. In our case, we also employ SEM as an analytical tool, but the constructs are derived from a set of observable indicators relating to virtual shops that are taken from a variety of sources. We use financial information taken from annual accounts to measure the success of e-commerce: sales and profits. Unlike other approaches to the study of e-commerce- such as those that derive from Marketing or Sociology- there is little research into the selection and validation of the indicators that are contained in the constructs employed in this research.

The proposed model is summarized in Figure 1. It relates a variety of constructs on the relevance of the firm in the Internet, their impact on potential customers, and their final influence on the financial results of the firm.

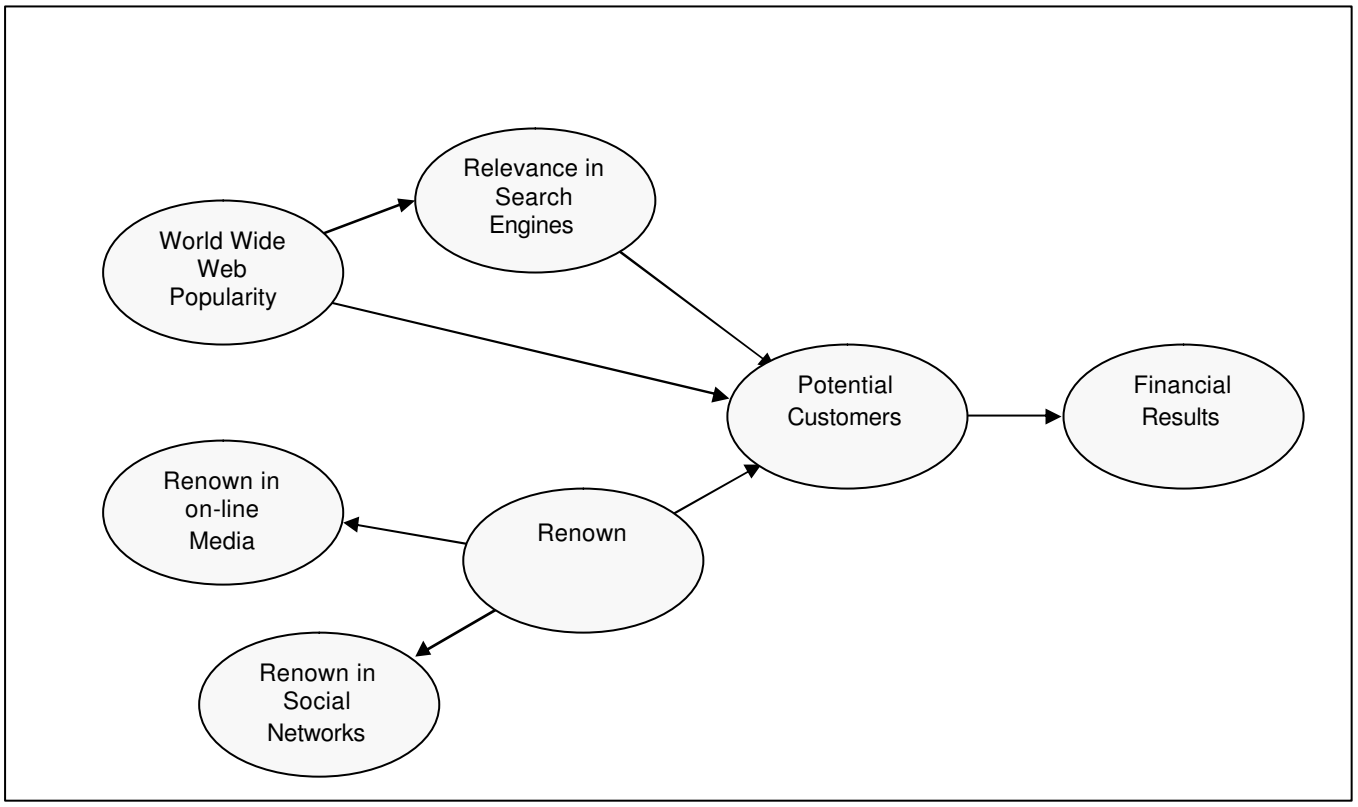

Figure 1. The structural model.

We now proceed to describe the constructs that make up the model in Figure 1.

The construct "World Wide Web popularity" attempts to capture the importance of a website in the net. The World Wide Web is the most important tool in the Internet, whose principal characteristic is a hypertext system based on hyperlinks. To be popular, a web has to be well connected, and it has to be possible to reach it from many other web pages.

When given various search criteria, alternative search engines use different algorithms to rank web pages. A good search engine ranks in the top positions of the search the most relevant pages. As Internet searches are one of the activities to which users devote most time, it is important for firms that their pages be well positioned by search engines. The construct "relevance in search engines" tries to reflect such importance. The importance that firms place in being well ranked by search engines has given rise to a new activity: Search Engine Optimization (SEO). According to Zhang and Dimitroff (2005), SEO is the process of identifying what factors affect the accessibility to the search engine of a web page, in order to

incorporate in the web page all the elements that may enhance the visibility of the page in a search.

It is important for any firm that deals in consumer products, particularly in consumer products aimed at final consumers, to achieve good renown. Renown gives access to many sources of information, such as portals and Internet media, which tend to be continuously brought up to date. The construct "Renown in online media" attempts to capture such renown or fame.

Viral marketing techniques, aimed at spreading information the way an epidemic spreads, are prevalent in the Internet. They attempt to make use of the navigators' natural way of communicating with each other in order to promote a web page, so that the number of visitors and the number of possible transactions increase. In contrast with other media, such as television, radio, or the press, there is active user participation. What users comment in chats, blogs, forums, discussion groups, or e-mail groups matters. It is, therefore, important for firms to achieve "Renown in social networks", the name we have given the construct.

We include in our model a second-order construct which we name just "Renown", and includes both "Renown in online media" and "Renown in social networks". Second-order constructs are constructs of constructs and do not have direct links to measurable indicators.

The construct "Potential customers" makes reference to the success that the firm has in attracting visitors to its web page, considering that such visitors are potential consumers. This is an intangible that is normally approached through web metrics.

The remaining construct in the model is described as "Financial results", and captures the success that has been had in making sales and obtaining profits.

We will now describe the hypotheses that underlie the model.

2.1 The effect of potential customers on financial results

The introduction made reference to various papers that find a positive relationship between webmetrics- such as number of pages visited and number of visitors- and the sales of the firms that operate in the Internet, or that find a relationship between webmetrics and the market values. For example, Trueman et al (2001), using the three webmetrics "unique visitors", "pages visited", and "time spent", finds a significant relationship between income

growth and webmetrics growth. This increase is more prominent in virtual shops than in Internet portals. Serrano-Cinca et al (2005), studying the situation before and after the dot com crisis, find that Pearson's correlation coefficient between visitors and sales changed from negative before the crisis (although not significantly different from zero) to positive (although low) after the crisis. This indicates that the sector has been maturing and we put forward the following hypotheses:

Hypothesis 1a:

"Potential customers" have a positive effect on "Financial results".

Hypothesis 1b:

The positive effect of "Potential customers" is higher in sales originating in the Internet than in other items in the accounts, such as total sales or profits, as some sales may not originate in the Internet.

Hypothesis 1c:

The positive effect of "Potential customers" on "Financial results" is higher in dot com firms that in *"brick and click"* firms.

2.2 On "World Wide Web popularity", "relevance in search engines", and "potential customers".

Nielsen Netratings (2004), found in a study of USA Internet users, that 76% of them use search engines, and that search engines continue to be the most popular tool in web navigation. The study remarks that on line customers, besides finding virtual shops in which make their purchases, use search engines to shop around. A joint study by ComScore Networks and DoubleClick (2005) monitors the information habits of customers 12 weeks before their on line purchases. They found that more than 50% of purchase decisions start in search engines, stressing the high number of queries that take place several weeks before the purchase.

Drèze and Zufryden (2004) argue that visibility in the Internet- measured through positioning in search engines, positioning in directories, and others- precedes web traffic. They construct a visibility index and test several hypotheses, using questionnaire data from 100 Internet firms. They relate online visibility with web traffic and are able to even predict

web traffic. They show that on line visibility has a higher and more significant impact than expenditure in advertising, or brand awareness.

In line with the view that search engines are an important source of web traffic for a firm, those firms that sell products in the Internet want to appear high up in search engines. In other words, they want to appear high up in the list when a potential customer uses the relevant keyword in the search engine. For example, it is important for a bookshop to appear in the first positions of the list when a customer inputs the name of a book in the search engine.

The following hypothesis can be put forward.

Hypothesis 2a:

The "Relevance in search engines" has a positive effect on "Potential customers".

Search engines use algorithms in order to determine the position that a firm takes in the search. An example is PageRank^{Tm}, used by Google; Brin and Page (1998). Such algorithms are responsible for the "Relevance in the search engine" construct.

The more links a web page receives, the higher its rank in search engines, but this is only one of pieces of information taken into account by algorithms such as Google's PageRank^{Tm}. This system resembles the methods long used in the assessment of academic papers: the more references a paper receives, the higher its standing in academic circles.

Important as they are the number of links received, they are not all equivalent, and search engines do not treat all of them in the same way. Search engines discard, for example, link farms- web pages whose sole purpose is to place links to specific web pages- and penalise links associated with spam. They may also attach higher importance to links originating in relevant web pages.

Although algorithms such as PageRank^{Tm} provide a global measure of the importance of a web page from the point of view of the search engine, the position of the page in the ranking depends on more factors. Of prime importance is the relevance in terms of key words used in the search. As an example, consider the web pages of a University. These tend to have high values of PageRank^{Tm}, but this does not ensure that they appear top of the list when one is looking for non-academic products.

Following this discussion, we put forward the following hypothesis.

Hypothesis 2b:

"World Wide Web popularity" has a positive effect on "Relevance in search engines".

Besides improving the positioning in search engines, links can be a source of visitors, or even direct sales. This was studied by Thelwall (2001) who found that about two thirds of the pages of commercial firms include links to external sites, and that in 31% of the cases these are links to other firms with which the firm that owns the web page has trade. Sometimes, redirecting a visitor to another site is a source of income for the firm that redirects the visitor. This has become known as "on-line affiliated programs".

A study carried out by Ennew et al (2005) using a sample of 500 e-commerce firms, found that the number of links that a virtual shop receives explain about 60% of the variation of site traffic. It follows that receiving links ought to be a crucial strategy for any firm that operates in the dot com sector.

We, therefore, suggest the following hypothesis.

Hypothesis 2c:

"World Wide Web popularity" has a positive impact on "Potential customers".

2.3 The relationship between "Renown in on-line media and social networks" and "Potential customers".

The Internet is more than users who passively navigate in the World Wide Web. From its very beginnings it has included other applications that allowed users to be active. We are thinking, in particular, in discussion groups, forums, or e-mail lists. New collaborative tools, such as blogs, wikis, P2P networks, and the open source, have joined these possibilities. All these alternative uses of the Internet result in "social networks" and a more participative net. The importance of social networks, such as blogs, has been studied in the areas of Politics-Adamic and Glance (2005)-, Journalism- Persing (2004), and Haas (2005)-, and Sociology-Barton (2005).

The first point in Locke et al (2001) *Cluetrain Manifesto* states that "markets are conversations". Further on they defend the view that "people in networked markets have figured out that they get far better information and support from one another than from vendors". These postulates are implemented by the leader in Internet auctions, eBay, as stated by one of its managers, William C. Cobb, who in an interview published in The Economist

(2005), said: "generally in a company employees know most about the business. In eBay, our community knows more than we do". A market study by the consulting firm BIGresearch (2005), on the basis of a sample of 14000 participants, revealed important differences on the influence of blogs in purchase decisions. Such differences depend on the age of the person interviewed and on the product, but one can argue that between 1% and 10.5% of the people are influenced by blogs. Drèze y Zufryden (2004), when building an index of "visibility in the Internet" for e-commerce firms, include information about impact on forums, chats and news groups.

Other authors take a more critical attitude towards such social networks. They acknowledge their increasing role in the creation of currents of opinion, while, at the same time, raising their concern. Barton (2005) analyses the role of blogs, discussion groups and wikis, framing the discussion on Habermas (1998) concept of "critical public sphere". This critical public sphere refers to a network of opinions that is today under the influence of the media. According to Barton (2005), the Internet is increasingly being dominated by the mass media. He warns against the passivity that the mass media, in the interest of multinational corporations, demands from the readers of newspapers or the viewers of television, and argues in favor of retaining Internet's democratic structure. Haas (2005), expanding this point further, argues that blogs are not a change with respect to other media; in his opinion, blogs reproduce and imitate the discourse of newspapers and other communication media.

Following this discussion, we put forward the following hypotheses:

Hypothesis 3a:

"Renown" has a positive influence on "Potential customers".

Hypothesis 3b:

There are significant differences between "Renown in on-line media" and "Renown in social networks". These form a second order construct that we describe as "Renown".

3. EMPIRICAL STUDY. MEASUREMENT MODEL.

Structural equations consist of two parts: the measurement model and the structural model. The measurement model contains the relationships between each construct and its indicators. This step attempts to identify a set of measurable indicators that define the construct. The indicators must satisfy a series of conditions such as unidimensionality and reliability.

The structural model addresses the relationships that exist between constructs. This model is estimated using the technique of Partial Least Squares (PLS). The software employed is PLS-Graph, 3.0 version; Chin (2001). The traditional way of structural equation modeling is associated with LISREL. PLS has been compared with LISREL by Chin (1998). Rangarajan et al (2005) argue that PLS is robust to departures from normality, and works better with small and medium-sized samples. These assumptions are common in financial contexts, as discussed in Deakin (1976) and Ezzamel et al (1990).

The next section discusses the sample of firms and the selection of indicators associated with each construct. We analyze the unidimensionality, reliability, and validity of the indicators selected.

3.1 The data.

We started with the top 400 firms in InternetRetailer' Guide, 2005 edition. This contains the largest firms in the USA that sell through the Internet. Total sales and net profits were taken from Yahoo Finance in the case of firms that are quoted in the stock exchange. When the firm was not quoted, the data was taken from Duns & Bradstreet. In total, sales data was obtained for 360 firms. Profits data was available for 196 firms. All data corresponds to 2004.

Some firms are "dot com" and others are *"brick and click"*; i.e., they sell through the Internet but also through traditional retail outlets. InternetRetailer gives the percentage of sales that originate in the Internet, which is obtained from the notes in the accounts, or directly asking the firms concerned. However, InternetRetailer sometimes estimates this information. We were unhappy with the estimation method employed, and we treated estimated data as missing information. This reduced the number of firms for which on-line sales information was available to 139. Following Demers and Lev (2001), Jorion and Talmor (2001), Hand (2001), Rajgopal et al (2002), Keating et al (2003), and Kotha et al (2004) we describe as "dot com" a firm if more than 50% of its sales originate in the Internet. Under this definition, 63 of the firms in the final sample are dot com. If the percentage of sales through the Internet is

lower, we classify the firm as *"brick and click"*, ending up with 64 such firms. In 12 further cases we know the amount of on-line sales but not the total value of sales.

The study will make use of four different samples. The main sample used in the study consists of the 139 with full information on Internet sales. Two smaller samples have been extracted from this: a sample of 63 "dot com" firms, and a sample of 64 *"brick and click"*. These are small samples and we must be cautious in the extrapolation of the conclusions. The fourth sample in the analysis includes the 360 firms for which there is information about total sales. PLS is appropriate in the case of small sample sizes, as each construct is the object of a separate estimation and was chosen for this reason.

3.2. Selection of indicators for each construct.

FASB(2001) points out that many firms do not make appropriate use of non-financial indicators. It claims that such indicators are introduced without relation to the context of the firm, or the sector of activity; that the measurement and presentation of the indicators is not consistent in time; and that often non-financial indicators are only vaguely defined. In our opinion, non-financial indicators must satisfy the same qualitative conditions that have to be met by financial information: understability, relevance, reliability, and comparability; FASB (1980).

Table 1 shows the indicators selected for each construct. The data was collected in June 2005.

Constructs and indicators		Factor loading			
		Full sample (N=360)	*Reduced sample (N=139)*	*Dot com sample (N=63)*	*Brick&click sample (N=64)*
World Wide Web popularity	- Number of links to the web. Source: Yahoo.com	0.935	0.962	0.968	0.834
	- Number of links to the web. Source: MSN.com	0.802	0.746	0.740	0.857
	- Number of links to the web. Source: Altavista.com	0.965	0.942	0.943	0.910
	- Number of links to the web. Source: Alltheweb.com	0.940	0.897	0.896	0.893
Relevance in search engines	- Google's PageRank	0.863	0.924	0.958	0.860
	- Yahoo's WebRank	0.892	0.917	0.965	0.807
	- Positioning of 5 generic keywords	0.748	0.797	0.794	0.814

Renown in on-line media	- Number of news. Source: News.google.com	0.942	0.996	1.000	0.897
	- Number of news. Source: News.yahoo.com	0.983	0.997	0.999	0.932
	- Number of news. Source: MSN.com	0.983	0.996	0.999	0.626
Renown in social networks	- Number of posts. Source: Google Groups	0.989	0.997	0.999	0.838
	- Number of posts. Source: Blogpulse	0.994	0.997	0.998	0.866
	- Number of posts. Source: Blogsearch	0.993	0.998	0.998	0.866
	- Number of posts. Source: Technorati	0.970	0.998	0.999	0.743
Potential customers	- Number of pages visited. Source: Comscore	0.993	0.994	0.998	0.998
	- Number of unique visitors. Source: Comscore	0.993	0.994	0.998	0.998

Table 1. Results of the exploratory analysis of indicators.

3.2.a) Indicators for the construct "*World Wide Web Popularity*".

The measurement of this construct should take into account the total number of incoming Internet links for each of the virtual shops in the sample. An estimate of this number can be obtained using a special command available in some search engines; this requires writing in the search "link:" followed by the address of the web page. This approach was followed, for example, by Brock and Zhou (2005). This option is available in the main search engines. It has been used within Yahoo, MSN, Teoma, Alltheweb, Altavista and Wisenut. Google was not used because this option has not been updated since November 2000.

3.2.b) Indicators for the construct *"Relevance in Search Engines"* .

Each search engine employs a different algorithm. In the case of Google, a good indicator of relevance is the value of PageRank. PageRank has a heavy weight in determining the rank that a web page will achieve in the search. Brin and Page (1998) argue that PageRank is an objective measurement that is well related to the subjective concept of "importance". This value is available in Google Toolbar, or in Google Directory. As PageRank is related to the logarithm of the actual number, we attempted to transform it back to the original number by calculating e^{PR} where PR is the page rank and e is the basis of neperian logarithms.

A similar procedure has been followed in the case of the search engine Yahoo, using what they call WebRank. WebRank can be obtained from Yahoo Toolbar.

No equivalent to WebRank or PageRank appears to exist for the other search engines.

PageRank and WebRank are not the only indicators of relevance in a search engine. It is also important to be well positioned with the keywords used in the search. Following this idea, the third indicator in this construct is related to keywords. The importance that keywords have in determining the visibility of a web page in a search engine has been studied by Zhang and Dimitroff (2005); they explored the impact on visibility of the position of keywords in web pages and their frequency. In our case, as we are dealing with Internet sales, we chose to search for five generic words: "shop", "shopping", "shop online", "store", and "buy". These are generic keywords. It would have been better to specify for each web site keywords associated with the product being sold; for example, in the case of an on-line bookshop, one would have been tempted to use the title of a book as keyword but, apart from the fact that this was impracticable, a study by ComScore Networks and DoubleClick (2005) on user's behaviour reveals that most searches employ generic terms, and that product specific keywords only amount to between 18.1% and 25% of all searches.

A separate search was made in Yahoo for each of the five keywords. Yahoo has the restriction of only showing the first 1000 results in any search. The position of each firm in each search was noted. The number 1000 was allocated to firms that did not appear amongst the first 1000. It was decided that a high number should be associated with a high position in the ranking. In order to achieve this, the number obtained for every firm and every search was deducted from 5000. Take, for example, Amazon; it ranks as number 3 under keyword "shop", number 3 under "shopping", number 1 under "shopping on-line", number 4 under "store", and number 3 under "buy". Using our method, Amazon's indicator is 4986. Conversely, a firm that does not appear amongst the first 1000 in any of the five keywords obtains a zero value for this indicator.

3.2.c) Indicators for the construct *"Renown in on-line media"*.

The study of news and announcements and their impact on financial results has long interested researchers in Accounting and Finance. This has generated a rich literature on which we have drawn in order to select the indicators associated with this construct. Example are Trueman et al. (2003), and Benbunam-Fich and Fich (2004) who study the reaction of the markets to the publication of a variety of news. Rajgopal et al (2003) use the database Lexis/nexis in order to analyze the visibility of Internet firms in the mass media.

We have used news searchers/gatherers, such a Google's News, Yahoo's News, and MSN's News. These services access hundreds of information sources, such as newspapers and specialized web pages, and are continuously updated. The number of times each firm appeared in each search was counted. The search always included in the name the dot com, in order to avoid generic names such as "Cooking"- in this case the search included "cooking.com".

3.2.d) Indicators for the construct *"Renown in social networks"*.

In order to select indicators that are relevant to this construct, we have employed search engines that are specific to discussion groups and blogs: Google Groups, Technorati, Daypop, Blogdigger, Blogpulse, and Blogsearch. As in the case of the previous construct, the name of the firm was always followed by dot com, and the number of posts that mentioned each firm was counted.

3.2.e) Indicators for the construct *"Potential customers"*.

Three web metrics are used as indicators for this construct: number of visitors, number of pages visited, and reach per million. Visitors (unique visitors) provides a non-duplicated measure of the number of different visitors who access a web site during a particular period of time, in this case one month.

Number of pages visited (visits), is the total number of web pages visited by users in one month.

Reach per million (reach) gives, for each million internauts, the number who visited a particular web page.

The information on visitors and visits was taken from InternetRetailer. Reach per million was taken from Alexa.com.

3.2.f) Indicators of financial results.

Three such indicators were taken: total sales, Internet sales, and net profit. Each one of them was taken in turn, and not all of them at the same time. This means that, strictly

speaking, we should be talking about constructs with a single indicator. In other words, we constructed three structural models, one for each indicator.

3.3 Unidimensionality analysis.

The objective of this part of the analysis is to test that the indicators that compose each construct are unidimensional. To check this, five Principal Component Analysis (PCA) exercises were conducted, one for each construct. Input to each PCA were the indicators in the construct. We tested that only one component was associated with an eigenvalue higher than 1.0, using Kaiser's (1960) criterion. The percentage of total variance explained by the first component was also noted. This was followed by a Factor Analysis where the factors were estimated from PCA. We checked that the factor loadings associated with the first factor took values greater than 0.5 for each indicator. Following these steps, the following indicators were discarded: the number of links in search engines Teoma and Wisenut in the construct "World Wide Web popularity", and the number of posts in blog search engines Daypop and Blogdigger in the construct "Renown in social networks".

Table 1 shows the factor loadings of the indicators, having discarded the ones that did not meet the unidimensionality criterion (and the reliability criterion which will be discussed in the next section). Table 2 shows the percentage of variance explained by the first component in each construct.

	Full sample (N=360)		*Reduced sample (N=139)*		*Dot com sample (N=63)*		*Brick&click sample (N=64)*	
	σ (%)	α	σ (%)	α	σ (%)	α	σ (%)	α
World Wide Web popularity	83.26	0.873	79.37	0.883	79.41	0.884	76.37	0.851
Relevance in search engines	69.98	0.752	77.68	0.817	82.69	0.860	68.46	0.711
Renown in on-line media	94.01	0.699	99.64	0.790	99.85	0.792	68.88	0.364
Renown in social networks	97.37	0.766	99.47	0.779	99.67	0.781	68.86	0.293
Potential customers	98.56	0.819	98.74	0.813	99.51	0.800	97.65	0.897

σ *(%) is the percentage of variance explained*

α *is Cronbach's alpha*

Table 2. Results of exploratory analysis for constructs

3.4 Reliability analysis.

The study of reliability requires assessing the internal consistency of the indicators that form part of and each construct. We have calculated Cronbach's (1970) alpha, a statistic that varies between 0 (no homogeneity) and 1 (highest homogeneity). It is customary to require a reliable data set to contain alpha values in excess of 0.7. The values are shown in Table 2.

Having performed this reliability analysis, we discard the indicator "Alexa reach" from the construct "potential customers". In general, constructs satisfy the reliability criterion except in the sample of 64 "brick and click" firms, where the constructs "renown in on-line media" and "renown in social networks" fall short of the minimum value required. The highest reliability value for these constructs is obtained in the sample that contains only "dot com" firms. We interpret these results as implying that "renown" is a construct that can be reliably used if we concern ourselves with virtual shops, but may not be appropriate as defined in other contexts.

Two further statistics were calculated: the index of composite reliability- Jöreskog (1971)-, and the analysis of variance extracted (AVE)- Fornell and Larcker (1981). Table 3 shows both values. In all cases, the values obtained exceed the minimum recommended for acceptance (0.6 and 0.5); Baggozi and Yi (1998).

	Full sample (N=360)		*Reduced sample (N=139)*		*Dot com sample (N=63)*		*Brick&click sample (N=64)*	
	ρ_ε	AVE	ρ_ε	AVE	ρ_ε	AVE	ρ_ε	AVE
World Wide Web popularity	0.947	0.817	0.928	0.765	0.927	0.762	0.923	0.749
Relevance in search engines	0.860	0.673	0.912	0.776	0.934	0.827	0.862	0.676
Renown in on-line media	0.979	0.940	0.999	0.996	1.000	0.999	0.866	0.688
Renown in social networks	0.990	0.962	0.998	0.993	0.998	0.993	0.898	0.688
Potential customers	0.991	0.983	0.990	0.980	0.994	0.987	0.984	0.969

ρ_ε *is the value of the composite reliability index*

AVE is the average variance extracted

Table 3. Reliability measures for constructs

3.5 Analysis of construct validity.

We used the criteria suggested by Jöreskog y Sörbom (1993) in order to assess the convergent validity of the indicators. The weak criterion of convergence checks if the factorial regression coefficients of each indicator are significant, testing at the same time that their t-value exceeds 2.58 (p= 0.01). The strong convergence criterion removes indicators whose standardized factorial regression coefficients are smaller than 0.5. The indicators that contribute the least to the explanation in the model are removed, taking as a threshold rule that the value of R square be less than 0.5. In our case, all the criteria are satisfied.

Turning now to the criterion of discriminant validity, if the theory says that two constructs are unrelated, they should also be significantly different. Bagozzi (1994) translates this condition requiring that the correlation between the constructs that are included in the model should not be higher than 0.8. Chin's (1998) criterion is that the square root of AVE should be greater than the correlation between this construct and all others. Table 4 shows all the correlation coefficients between constructs. The main diagonal of the table contains the square roots of AVE instead of the unity. Note that there four tables, because four models have been estimated, one for each sample.

Full sample (N=360)	PopW	RelSE	RenOM	PotC	RenSN	FinR
PopW	**[0.904]**					
RelSE	0.664	**[0.820]**				
RenOM	0.541	0.662	**[0.970]**			
PotC	0.494	0.731	0.753	**[0.991]**		
RenSM	0.585	0.725	0.943	0.733	**[0.980]**	
FinR (O)	0.646	0.856	0.778	0.828	0.780	**[1]**
FinR (S)	0.075	0.248	0.055	0.420	0.046	**[1]**
FinR (P)	0.123	0.220	0.084	0.435	0.056	**[1]**

Reduced sample (N=139)	PopW	RelSE	RenOM	PotC	RenSN	FinR
PopW	**[0.875]**					
RelSE	0.744	**[0.881]**				
RenOM	0.743	0.797	**[0.998]**			
PotC	0.739	0.838	0.886	**[0.990]**		
RenSM	0.746	0.796	0.996	0.881	**[0.996]**	
FinR (O)	0.667	0.853	0.778	0.828	0.780	**[1]**
FinR (S)	0.171	0.397	0.085	0.312	0.090	**[1]**
FinR (P)	0.356	0.420	0.166	0.373	0.174	**[1]**

Dot com sample (N=63)	PopW	RelSE	RenOM	PotC	RenSN	FinR
PopW	**[0.873]**					
RelSE	0.816	**[0.909]**				
RenOM	0.755	0.945	**[0.999]**			
PotC	0.763	0.941	0.956	**[0.993]**		
RenSM	0.757	0.944	0.997	0.950	**[0.996]**	

Brick&click sample (N=64)	PopW	RelSE	RenOM	PotC	RenSN	FinR
PopW	**[0.865]**					
RelSE	0.751	**[0.822]**				
RenOM	0.604	0.611	**[0.829]**			
PotC	0.614	0.550	0.501	**[0.984]**		
RenSM	0.909	0.816	0.628	0.618	**[0.829]**	

FinR (O)	0.773	0.958	0.984	0.962	0.983	**[1]**
FinR (S)	0.770	0.958	0.974	0.954	0.974	**[1]**
FinR (P)	0.973	0.946	0.981	0.944	0.986	**[1]**

FinR (O)	0.543	0.641	0.390	0.562	0.665	**[1]**
FinR (S)	0.554	0.575	0.356	0.635	0.655	**[1]**
FinR (P)	0.698	0.552	0.446	0.674	0.779	**[1]**

PopW (World Wide Web Popularity); RelSE (Relevance in search engines); RenOM (Renown in on-line media); PotC (Potential customers); RenSM (Renown in social networks); FinR (Financial results).
Numbers in the main diagonal indicate the square root of the AVE (average variance extracted).

Table 4 Correlation between latent constructs

We observe in Table 4 the presence of high correlations between constructs, in some cases very near the values that suggest rejection of discriminant validity. Consider the sample that contains 139 firms; in it, all the constructs satisfy the discriminant validity criterion except when "renown in social networks" and "renown in on-line media" are involved. The correlation between these constructs is very high. This suggests that we are dealing with two different ways of looking at one single construct. To this effect, the modeling process was repeated with a single construct (renown). This new construct contained seven indicators-three associated with "renown in on-line media" and four associated with "renown in social networks" and was found to be one-dimensional, and satisfies the Crombach's alpha reliability criterion. This supports the views of Haas (2005) who argues that blogs are not a departure from other mass media. The high correlation coefficient between "renown in social networks" and "renown in on-line media" is consistent with the view that bloggers discuss the same subjects that are discussed by on-line newspapers. One is tempted to conclude that blogs are resonance boxes for standard news, although a full analysis would require a deeper analysis of what is being discussed and how this is done. In summary, the data reject hypothesis 3b: no differences have been found between "renown in social networks" and "renown in on-line media".

4 STRUCTURAL MODEL RESULTS

This section discusses the results in the context of the maintained hypotheses.

Table 5 shows parameter estimates and t-values for the estimated model. R square values are shown in Table 6.

Relationships between constructs	*Full sample (N=360)*		*Reduced sample (N=139)*		*Dot com sample (N=63)*		*Brick&click sample (N=64)*	
	Beta	***t*-value**	**Beta**	***t*-value**	**Beta**	***t*-value**	**Beta**	***t*-value**
Potential customers ? FinR (Online sales)	0.746	4.605*	0.777	5.823*	0.933	8.057*	0.650	3.532*
Potential customers ? FinR (Total sales)	0.463	3.716*	0.438	1.997	0.915	6.310*	0.718	4.372*
Potential customers ? FinR (Profit)	0.488	3.065*	0.480	2.477	0.846	3.029*	0.720	3.446*
WWW popularity ? Relevance in search engines	0.654	7.873*	0.737	6.708*	0.851	9.088*	0.794	19.58*
WWW popularity ? Potential customers	-0.118	0.558	0.136	0.581	0.037	0.099	0.278	0.887
Relevance in search engines ? Potential customers	0.426	4.878*	0.314	3.379*	0.280	2.723*	0.023	0.309
Renown ? Potential customers	0.485	4.722*	0.481	4.214*	0.650	5.165*	0.440	1.024
Renown ? Renown in on-line media	0.944	16.600*	0.975	27.518*	0.992	46.535*	0.909	13.078*
Renown ? Renown in social networks	0.965	26.877*	0.988	37.076*	0.993	58.459*	0.954	38.105*

* significant at 0.01 level

Table 5. Relationships between constructs

	Full sample (N=360)	*Reduced sample (N=139)*	*Dot com sample (N=63)*	*Brick & click sample (N=64)*
Relevance in search engines	0.441	0.553	0.666	0.564
Renown in on-line media	0.963	0.998	0.999	0.746
Renown in social networks	0.979	0.999	1	0.873
Potential customers	0.648	0.832	0.924	0.419
Financial results (online sales)	0.685	0.686	0.926	0.316
Financial results (total sales)	0.176	0.098	0.911	0.404
Financial results (profits)	0.189	0.139	0.891	0.454

Table 6. *R*-square values for latent constructs

Hypothesis 1

Hypothesis 1a stated that "potential customers" have a positive impact on "financial results". If we consider as "financial results" the amount of on-line sales, the hypothesis is supported by the data in all four samples. The value of the path parameter ranging from 0.65

for “brick and click” to 0.93 for “dot com”, being the parameters significantly different from zero in all the cases. This is the way we would expect the sector to behave. If we describe “financial results” through total sales or net profit, the positive impact of “potential customers” is not as clear as in the case of Internet sales. This hypothesis is even rejected in some of the samples. We conclude that Hypothesis 1b, that the effect of “potential customers” is higher in Internet sales, is supported by the data.

Hypothesis 1c argued that the positive impact of “potential customers” on “financial results” is higher in “dot com” firms than in “brick and click” firms. A look to path parameters, t-values, and R square results confirms that the hypothesis is supported by the data. The combination of hypotheses 1b and 1c suggests that the model is particularly valuable in order to explain on-line sales particularly in the case of firms that conduct most of their business in the Internet.

Hypothesis 2

The data support hypothesis 2a, that the “relevance in search engines” has a positive impact on “potential customers”, except in the sample of “brick and click” firms. Path coefficients and t values are significant in all other cases. The R square of the construct “potential customers” ranges from a minimum of 0.419 in the case of the 64 “brick and click” firms and a maximum of 0.924 in the case of the 64 “brick and click” firms.

Hypothesis 2b, that “World Wide Web popularity” has a positive impact on “relevance in search engines”, is supported by the data in all the samples. In other words, the more the number of incoming links to a web page, the higher the relevance in search engines.

The effect between “World Wide Web popularity” and “potential customers” presumed in hypothesis 2c does exist, as one can see examining the correlation between both constructs, shown in Table 4. This effect appears to be an indirect one, through “relevance in search engines”. Path coefficients and t-values do not support the view that there is a direct effect. It is worth recalling that the indicators of the construct “relevance in search engines”- PageRank, WebRank, and keywords positioning- go beyond than the mere sum of links, the basis of the construct “World Wide Web popularity”. The search engine optimization requires more than obtaining lots of links to our website.

Hypothesis 3

The data support hypothesis 3a, that "renown" has a positive effect on "potential customers". We find, once again, that "brick and click" firms are an exception to the rule. With this proviso, we find high path coefficients and significant t-values. It follows that firms, in particular those firms that conduct most of their business in the Internet, must pay attention to their presence in on-line information media, in social networks, in forums, and in blogs.

5 CONCLUSIONS

In this paper we have developed an explanatory model for the sales of the firms that conduct their business in the Internet. It has been argued that in order to sell in the Internet, firms need to receive visitors to their virtual shops- potential customers-, and in order to obtain such visits, they need to be popular in the World Wide Web, they need to be well positioned in the top places of search engines, and they need to achieve renown in both social networks and on-line media. The model has been validated using a sample of "dot com" and "brick and click" e-tailers.

Internal information users, particularly managers, must pay attention to webmetrics. This information is available in the file "access.log", which also contains visitor's precedence, making it possible to find out if the visitor has been referred from another web page, or has arrived as a consequence of a search, and in this last case gives the keywords that have been used. A good strategy aimed at promoting Internet sales would require enhancing the visibility of the firm in the Internet. This would imply increasing the number of incoming links to the firm's web page, improving their positioning in search engines through the use of appropriate keywords, and maintaining a presence in on-line media and social networks- such as blogs and forums. How to do this requires a detailed marketing plan, something that goes beyond the scope of the present paper.

Turning now to external users of information, we can say that financial analysts can and should pay attention to the intangibles described in the constructs. The constructs have been built on the basis of information available from public sources, using special facilities available in search engines. These facilities, such as SEO tools, news gatherers from on-line media, blog post searches are available to all.

The analysis has been carried out using Structural Equation Modeling (SEM). Although the use of SEM is gaining in popularity, particularly in marketing studies, very little work appears to have been carried out in financial analysis area using SEM. Furthermore, the data used tends to proceed from consumer questionnaires, and in this study we have shown that there is no reason not to use public statistical information. In fact, given the characteristics of the data, no other statistical technique would have capture the richness of the situation in the way in which SEM does.

REFERENCES

Adamic, L. and Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. WWW2005, Japan

Alpert, B. (1999). How Internet Ad Barter Generates Over Half of Some Firms' Revenues. Barron's, 79 (22), 45.

Bagozzi, R. (1994). Structural Equation Model in Marketing Research, Principles of Marketing Research. Blackwell Publishers, Oxford.

Bagozzi, R. and Yi, Y. (1988). On the evaluation of structural equation models, Academy of Marketing Science, 16 (1), 74-94.

Barton, M.D. (2005). The future of rational-critical debate in online public spheres. Computers and Composition, 22, 177–190

Benbunan-Fich, R. and Fich, E. (2004). Effects of web traffic announcements on firm value. International Journal of Electronic Commerce, 8 (4), 161-181.

BIGresearch (2005). Simultaneous Media Usage Survey (SIMM VI). BIGresearch Management Worthington, Ohio. http://www.bigresearch.com/news/big062205.htm

Black, E.L. (1998). Life-cycle impacts on the incremental value-relevance of earnings and cash-flow measures. Journal of Financial Statement Analysis, 4 (1), 40-56.

Bollen, K. A. (1989). Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics. New York: Wiley

Bowen, R.; Davis, A. and Rajgopal, S. (2002). Determinants of revenue reporting practices for Internet firms. Contemporary Accounting Research, 19 (4), 523-562.

Brin, S. and Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 30 (1-7), 107-117

Brock, J.K.U. and Zhou, Y. (2005). Organizational use of the Internet: Scale development and validation. Internet Research, 15 (1), 67-87.

Chang, J., Torkzadeh, G. and Dhillon, G. (2004). Reexamining the Measurement Models of Success for Internet Commerce. Information & Management, 41 (5), 577-584.

Chin, W. W. (1998). The Partial Least Squares Approach for Structural Equation Modeling. In: Modern methods for business research, 295-336. (Marcoulides, G. A. Ed.), Lawrence Erlbaum Associates, Hillsdale.

Chin, W. W. (2001). PLS-Graph User's Guide Version 3.0. Soft Modeling Inc.

ComScore Networks and DoubleClick (2005). Search Before the Purchase: Understanding Buyer Search Activity as it Builds to Online Purchase, Febr. 2005, http://doubleclick.com/us/knowledge_central/documents/RESEARCH/searchpurchase_0502.pdf.

Cooper, M.; Dimitrov, O. and Rau, R. (2001). A rose.com by any other name. The Journal of Finance, 56 (6), 2371-2388.

Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row.

Davis, A. (2002). The value relevance of revenue for Internet firms: Does reporting grossed-up or barter revenue make a difference? Journal of Accounting Research, 40 (2), 445-477.

Deakin, E.B. (1976). Distributions of financial accounting ratios: some empirical evidence, The Accounting Review, 51, 90–96.

Demers, E. and Lev, B. (2001). A rude awakening: Internet shakeout in 2000. Review of Accounting Studies, 6, 331-360.

Drèze X. and Zufryden, F. (2004). Measurement of online visibility and its impact on Internet traffic. Journal of Interactive Marketing, 18 (1), 20-37

Economist, The (2005). Meg and the power of many. 11/6/05, 63-66.

Ennew, C., Lockett, A., Blackman, I. and Holland, C. (2005). Competition in Internet Retail Markets: The Impact of Links on Web Site Traffic. Long Range Planning, 38, 359-372.

Ezzamel, M., Brodie, J., and Mar-Molinero, C. (1990). The distributional properties of financial ratios in UK manufacturing companies, Journal of Business Finance and Accounting 17 (1), 1-29.

Financial Accounting Standards Board (FASB) (1980). Statement of Financial Accounting Concepts n.2, Cualitative Characteristics of Accounting Information. FASB, May 1980, Stamford, Connecticut.

Financial Accounting Standards Board. (FASB) (2001). Business and Financial Reporting: Challenges from the New Economy. New York: FASB.

Fornell, C. and Larcker, D. (1981). Structural Equation Models with Unobserved Variables and Measurement Error. Journal of Marketing Research, 36 (3), 39-50.

Haas, T (2005). From Public Journalism to the Public's Journalism? Rhetoric and reality in the discourse on weblogs, Journalism Studies, 6 (3), 387-396.

Habermas, J. (1998). The structural transformation of the public sphere (Thomas Burger, Trans.). Cambridge. MIT Press.

Hand, J.R. (2001). The Role of Book Income, Web Traffic, and Supply and Demand in the Pricing of U.S. Internet Stocks. European Finance Review, 5 (3), 295-317.

Jöreskog, K. (1971). Statistical analysis of sets of congeneric tests. Psychometrika, 36, 109-133.

Jöreskog, K.G. and Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ: Lawrence Erlbaum

Jorion, P. and Talmor, E. (2001). Value Relevance of Financial and Non Financial Information in Emerging Industries. The Changing Role of Web Traffic Data. Annual Meeting American Accounting Association, Atlanta

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Keating, E.; T. Lys and Magee, R. (2003). The Internet Downturn. Finding Valuation Factors in Spring 2000. Journal of Accounting and Economics. 34, 189-236.

Keeney, R.L. (1999). The Value of Internet Commerce to the Customer, Management Science, 45 (4), 533-542.

Kotha, S., S. Rajgopal, M and Venkatachalam (2004). The Role of Online Buying Experience as a Competitive Advantage: Evidence from Third-Party Ratings for E-Commerce Firms. Journal of Business, 77 (2), 100-134

Livnat J. and Zarowin, P. (1990). The incremental informationt content of cash-flow components. Journal of Accounting and Economics. 13, 25-46.

Locke, C; Levine, R; Searls, D. and Weinberger, D. (2001). The Cluetrain Manifesto. The End of Business as Usual. Perseus Books.

Nielsen/Netratings (2004). Internet Users Turn to Search Engines for Online Shopping, Nielsen Netratings. http.//www.netratings.com/pr/pr_040223_us.pdf

Nikolaeva, R. (2005). Strategic Determinants of Web Site Traffic in On-Line Retailing. International Journal of Electronic Commerce, 9 (4), 113-132.

Persing, B (2004). Politics Makes Strange Webfellows. Serials Review 30 (4), 334-342.

Quaddus, M. and Achjari, D. (2005). A model for electronic commerce success. Telecommunications Policy, 29, 127–152.

Rajgopal, S., Kotha, S. and Venkatachalam, M. (2002). Managerial Actions, stock returns and earnings. The case of business-to-business internet firms. Journal of Accounting Research, 40 (2), 529-556.

Rajgopal, S.; Kotha, S. and Venkatachalam, M. (2003). The Value-Relevance of Network Advantages. The Case of E-Commerce Firms. Journal of Accounting Research, 40 (4), 135-162.

Rangarajan, D., Jones, E and Chin, W.W. (2005), Impact of sales force automation on technology-related stress, effort, and technology usage among salespeople. Industrial Marketing Management, 34 (4), 345-354.

Serrano-Cinca, C.; Fuertes-Callén, Y. and Mar-Molinero, C. (2005). Measuring DEA Efficiency in Internet companies. Decision Support Systems, 38 (4), 557-573.

Thelwall, M. (2001). Commercial Web Site Links. Internet Research, 11 (2), 114–124.

Torkzadeh, G., and Dhillon, G. (2002). Measuring factors that influence the success of Internet commerce. Information Systems Research, 13 (2), 187-204.

Trueman, B., Wong M.H.F. and Zhang, X.J. (2001). Back to basics. Forecasting the revenues of Internet firms. Review of Accounting Studies, 6 (2-3), 305-329.

Trueman, B., Wong M.H.F. and Zhang, X.J. (2000). The eyeballs have it. Searching for the value in internet stocks. Journal of Accounting Research, 38, 137-162.

Trueman, B., Wong M.H.F. and Zhang, X.J. (2003). Anomalous stock returns around internet ?rms' earnings announcements. Journal of Accounting and Economics, 34, 249–271

Zhang, J. and Dimitroff, A. (2005). The impact of webpage content characteristics on webpage visibility in search engine results. Information Processing and Management, 41, 665–690.