Using Principal Component Analysis (PCA) to Rank Countries on their Readiness for E-Tail

Soumitra Sharma

Institute of Technology, Banaras Hindu University
_____

Keywords

E-Tail, Principal Component Analysis, E-Tail Readiness Index

Abstract

E-Tail, or Internet based retailing, has emerged as an innovative channel for retailers to reach target consumers in the comfort of their homes. E-Tailing has exhibited an upward trend across the world in recent years, though it's more prevalent in certain regions such as Europe and North America. With varying levels of Internet penetration, telecom infrastructure, the business and legal environment, e-tailing is at varying stages of maturity in countries across the globe. In order to objectively compare the actual preparedness of various countries to exploit their potential for e-tail, an analytical model was built using Principal Component Analysis (PCA). This model returns a numeric value, coined as the e-Tail Readiness Index. A set of countries were subsequently ranked on the basis of this index. As expected, while the more developed countries of Europe and North America occupied the top slots, the emerging economies of Eastern Europe and Latin America occupied the middle slots. Surprisingly, India and China figured right at the bottom of the heap.

_____

Introduction to e-Tail

E-Retailing, or e-Tailing, has emerged as a great opportunity in the world of business. As compared to the traditional brick-and-mortar retailing, e-tailing seems to be a more attractive alternative. E-Tailing is a relatively new concept as it has made its presence felt only a few years ago. It can be better explained as Internet retailing, wherein various technologies and media are combined to sell products online. E-Tail has quickly grabbed the attention of many entrepreneurs world-wide. Thanks to the growing popularity of Internet all over the world, people have started shopping online in a big way.

Differences between Retail and e-Tail

Anything involving a direct sale to a consumer at any point of time can be termed as retailing. This retailing can be done in a shopping mall, a departmental store or in a basic mom-and-pop store. This imposes geographical constraints on the retailer. On the other hand, an e-tailer has global reach due to the Internet acting as a medium of retail.

A conventional retailer has to identify a good location for his operations, and attract consumers using advertising and promotional schemes. However, an e-tailer has

to attract consumers 'virtually' to his site and offer services. Thus, while the target consumer remains the same in both forms of retail, the business model is entirely different. In conventional retailing, as much as 50% of the initial investment could go towards acquiring real estate. This is not the case with e-tail.

Consumer retention is also very different in both the scenarios. Retaining a consumer on the Internet is not simple as his/her behavior is hard to predict. While conventional retailers succeed in forming a bond with the consumer through personalized attention via face-to-face interaction, convenient geographical location etc., such loyalty is difficult to obtain in virtual space. A consumer may simply switch from one shopping website to another in case it offers a better deal.

Current Trends in e-Tail

E-Tail is most prevalent in Europe and North America. Specifically, Germany, Austria and the UK come at the top, with almost 95% Internet users shopping online in each country. South Korea and Taiwan come at the top in Asia-Pacific, with close to 90% of all Internet users indulging in e-tail (ACNielsen News Release, 2005). On the whole, e-tailing is exhibiting an upward trend across the world, with the emerging economies showing more vibrant growth as compared to the developed countries.

Consumer behavior depends heavily on both the buying channel (conventional retail v/s the Internet) and the category of products being bought. The most popular product categories purchased online across the world include books, followed by videos/DVDs/games, airline tickets/reservations and clothing/accessories/shoes (ACNielsen News Release, 2005). Chinese and South Koreans are one of the biggest buyers of books on the Internet, and therefore, e-tailers should make sure that their websites support the corresponding native languages. Japan, on the other hand, is a huge opportunity for online book sellers and grocery sites.

Airline tickets/reservations are the most prevalent in Malaysia, New Zealand, Singapore and Australia. Tours and hotel reservations are the most prevalent in Finland and Spain (ACNielsen News Release, 2005). Online purchase of videos/DVDs/games is most preferred in UK and France. On the other hand, electronics are the most preferred category for online purchase in Latin America. Overall, credit cards are the most preferred mode of online payment across the world. In most regions, including North America and Latin America, credit cards are followed by PayPal in popularity.

E-Tail Readiness Index

With varying levels of Internet penetration, telecom infrastructure, the business and legal environment, e-tailing is at varying stages of maturity in countries across the globe. Although e-tail's potential to capture a significant chunk of the traditional retail market has now been widely accepted, especially in developed countries, it is interesting to compare the actual preparedness of various countries to exploit this potential. In order to do so objectively, an analytical model was built that returns a numeric value for the set of countries under study. This numeric value has been coined as the *e-Tail Readiness Index*, and serves as a yardstick to measure the readiness of a country's e-tailing market to take-off. Due to constraints of data availability, this index was calculated for a set of thirty countries, with data for all parameters collected for the year 2005.

Constituent Parameters

The *e-Tail Readiness Index* has been defined as a composite index of seven individual parameters. These include:

- Number of personal computer owners per 1000 people
- Number of Internet users per 100 people
- Number of telephone connections (mobile + fixed) per 100 people
- Per-capita GDP
- Business and Regulatory Environment Index
- Consumer and Business Adoption Index
- Supporting E-services Index

The first three parameters reflect the status of ICT (Information and Communication Technology) infrastructure in the country (data taken from EIU database). The fourth parameter, the per-capita GDP of the country, is a proxy for the average buying power of the citizens in the country.

The 'Business and Regulatory Environment Index', is a measure of the general business climate of the country, along with the country's legal and regulatory framework from the perspective of Internet usage (EIU White Paper, 2005). It is calculated as the mean of the 'business environment' score and the 'legal and policy environment' score, developed by the Economist Intelligence Unit. The 'Consumer and Business Adoption Index' reflects the citizen's acceptance of Internet based services (EIU White Paper, 2005). The values for this parameter have been directly obtained from the scores developed by the Economist Intelligence Unit.

The 'Supporting E-services Index' is a measure of various ancillary services present to support Internet based businesses (EIU White Paper, 2005). From the perspective of e-tailing, these could include web designers, ISP infrastructure providers, software professionals, and IT consulting services. The values for this parameter too have been directly obtained from the scores developed by the Economist Intelligence Unit. The weight assigned to each parameter was arrived at using the Principal Component Analysis method. This is discussed in the next section.

Principal Component Analysis

The e-Tail Readiness Index (C_j) was developed as a linear composite index of the previously discussed seven parameters. This is represented in equation (1) below.

$$C_j = W_1 x_{1j} + W_2 x_{2j} + W_3 x_{3j} + W_4 x_{4j} + W_5 x_{5j} + W_6 x_{6j} + W_7 x_{7j} \quad (1)$$

Where C_j refers to the *e-Tail Readiness Index* for the jth country, x_{ij} refers to the standardized value of the ith parameter for the jth country, and W_i the corresponding weight. Instead of subjectively assigning weights to each parameter, the Principal Component Analysis (PCA) method was adopted to calculate weights objectively. This PCA method is based on the principle that "*when different characteristics are observed about a set of events, the characteristic with more variation explains more of the variation in the dependent variable compared to a variable with lesser variation in it*" (Annexure on Principal Component Analysis – Government of India, 2003). This technique assigns weights in such a way that they maximize the 'sum of the squares of correlation' of the indicators

with the composite index. The various steps involved in the process are discussed in detail below.

Standardization of Data

As all short-listed parameters are measured in different units, they are not directly additive. Therefore, it's necessary to convert them into standard comparable units, to eliminate any bias of the initial scale. Each data point was standardized according to equation (2).

$$x_{ij} = (X_{ij} - X_m)/\sigma \quad (2)$$
$$i = 1, 2 \ldots 7 \text{ (Parameter No.)}$$
$$j = 1, 2 \ldots \ldots 30 \text{ (Country No.)}$$

X_{ij} is the original value of the ith parameter for the jth country. X_m is the mean, and σ is the standard deviation, of the series formed by values of the ith parameter for all 30 countries.

Construction of a Correlation Matrix

A correlation matrix was constructed, as given in Table 1, by calculating the simple correlation coefficients of each corresponding pair of indicators. The complete matrix was constructed via the following steps.

	x1	x2	x3	x4	x5	x6	x7
x1	rx1x1	rx1x2	--	--	--	--	rx1x7
x2	--	--	--	--	--	--	--
x3	--	--	--	--	--	--	--
x4	--	--	--	--	--	--	--
x5	--	--	--	--	--	--	--
x6	--	--	--	--	--	--	--
x7	rx7x1		--	--	--	--	rx7x7
Column Sum of Coefficients	Σrxix1 (i=1…7)	Σrxix2 (i=1…7)	--	--	--	--	Σrxix7 (i=1…7)
Grand Column Sum	S=Σ (Σrxix1+ Σrxix2+…………+ Σrxix7)						

Table 1: Correlation Matrix

Step1: The simple correlation coefficients of each corresponding pair of indicators were calculated. For example, rx1x2 is the correlation coefficient of two arrays. The first array is the series of all 30 values for the parameter x1, and the second is the similar series for x2. The elements of the diagonal of the matrix would be unity as they are self-correlated (rxixi=1). The matrix is also symmetrical as the elements of each row are identical to the elements of the corresponding column (rxixj= rxjxi).

Step 2: The 'column-sum' of all coefficients in each particular column was calculated. Seven such 'column-sums' were obtained.

Step 3: The above seven column-sums were added to obtain a 'grand column-sum' denoted by S.

Calculation of Weights

The weight W1 (for parameter x1) was calculated according to equation (3).

$$W1 = (\Sigma rxix1) / \sqrt{S} \quad (3)$$

Similarly, all weights W2…W7 were calculated. These are given in Table 2.

Parameters →	PC owners per 1000 people	Internet users per 100 people	Number of mobile and fixed line connections per 100 people	Per-capita GDP	Business and Regulatory Environment Index	Consu-mer and Business Adoption Index	Supporting E-services Index
Weight	0.94	0.95	0.91	0.95	0.96	0.92	0.96

Table 2: Weights Obtained from PCA

These weights were put in equation (1) to calculate the e-Tail Readiness Index (Cj) for each of the 30 countries. The final composite index values for the countries were normalized to represent them on a scale of 1 to 10.
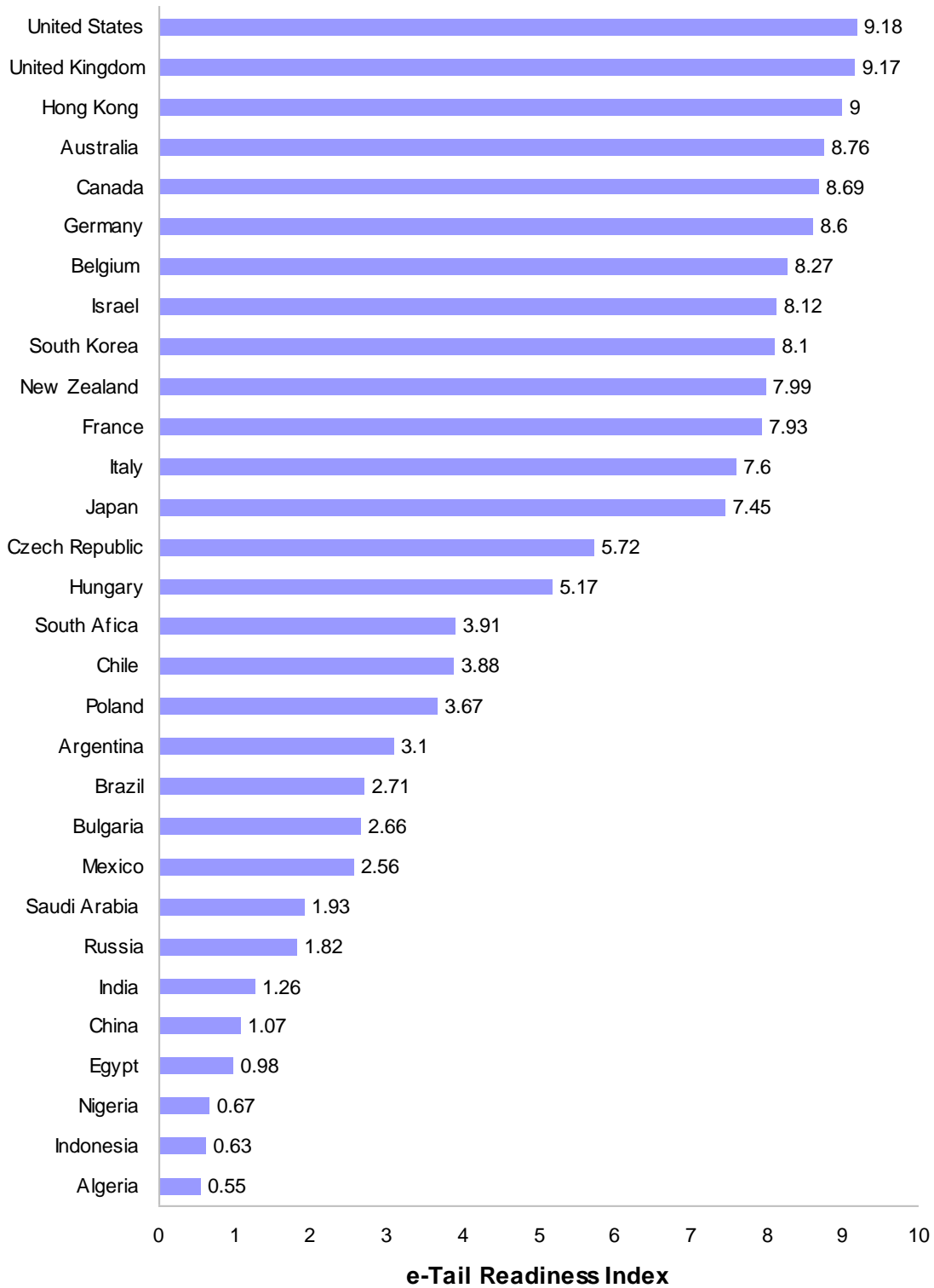
Figure 1: Comparison of e-Tail Readiness across Countries

Results and Discussion

The Figure 1 represents the results of the analytical model developed earlier.
As depicted in the bar-graph given above, the United States leads the pack, scoring 9.18 (on a scale of 10), followed by the United Kingdom, Hong Kong, Australia and Canada. It is interesting to note that even though the US lags behind countries such as the UK and Germany, in terms of Internet penetration and telecom density, it scored the highest in the composite index. This was primarily due to its high scores in the Consumer and Business Adoption Index, Supporting E-services Index and per-capita GDP. The Supporting E-services Index for the US was one of the highest among all 30 countries, thus reflecting the presence of excellent IT-support services. On the other hand, Canada ranked fourth on the index, despite excellent values of ICT infrastructure parameters. This was due to a low Consumer and Business Adoption Index. This highlights that, despite excellent ICT penetration, Canadians are not psychologically ready as yet, to utilize these ICT channels for online purchases. Rather, they are more likely to consider the web as a marketing, customer service, and pre-purchase research tool. It is validated by the fact that only 17% of Canadians made online purchases in 2001, as compared to 27% of Americans (Canadian IT Law Newsletter, 2002).

Unlike the US, the online retail market in Western Europe is quite heterogeneous, consisting of local markets at different levels of development. Internet penetration varies significantly across Europe, with the Nordic countries at the highest end of the spectrum, and countries such as Italy, Spain, France and Portugal at the lower end. Similarly, online purchasing habits vary considerably across the region, with the UK, France and Italy being heavy users of credit cards, while Germany and the Nordic countries are more reluctant to make credit card payments. In Germany and the Nordics, less than 20 % of online transactions are paid by credit cards, whereas the number stands at around 90 % for the UK. Against this backdrop, the UK was the only European country at par with the US in terms of e-tail readiness, scoring 9.17 on a scale of 10. Apart from the UK, Germany also scored well – a decent 8.6. France and Italy were down the list, and were ranked 11th and 12th respectively.

The BRIC countries were part of the lowest ranked countries in the group. Brazil led the BRIC group, followed by Russia, India and finally China. Although, Russia scored better than Brazil in ICT infrastructure parameters, it lagged behind considerably in the other three parameters. Same was the case with China and India, with China lagging behind India by a considerable margin in all parameters other than the ICT infrastructure related ones.

Conclusion

As discussed in previous sections, the Principal Component Analysis is a handy mathematical tool to rank countries on their readiness for e-tail. The e-Tail Readiness Index developed earlier can be used by e-tailers to do a gap analysis of various countries, vis-à-vis benchmark countries that have the highest scores (such as the US and UK). The index can also be used as a barometer of a country's maturity with respect to e-tail, and the potential it holds for the future. While the more developed countries of Europe and North America occupied the top slots (as expected), it is the emerging economies of

Eastern Europe and Latin America that occupy the middle slots and will be of most interest to retailers. As indicated by the index, these markets are ripe enough for sustainable e-tail, but immature enough to exhibit bright growth prospects for the future. Surprisingly, the 'in-vogue' growth engines of the world, India and China, figure right at the bottom of the heap. Therefore, going strictly by numbers, both the global giants are still not viable enough for e-tailers. Although, strategically speaking, it could be a totally different story.

References

(2005). 'ACNielsen Announces One-Tenth of the World's Population Shopping Online: 627 Million People Have, including 325 Million in the Last Month'. ACNielsen News Release Retrieved from http://www2.acnielsen.com/news/20051019.shtml on December 22, 2007 (2002). Canadian IT Law Newsletter. Canadian IT Law association

Retrieved from http://www.it-can.ca/newsletters/040402.pdf on December 22, 2007 (2003). Annexure on Principal Component Analysis. Department of Information Technology, Ministry of Communications and Information Technology, Government of India Retrieved from http://www.mit.gov.in/download/ANNEX_242-254.PDF on December 22, 2007 (2005). the 2005 E-Readiness Rankings. The Economist Intelligence Unit and the IBM Institute for Business Value – White Paper Retrieved from http://graphics.eiu.com/files/ad_pdfs/2005Ereadiness_Ranking_WP.pdf on December 22, 2007

The Economist Intelligence Unit (EIU) Database – http://www.eiu.com