Martínez de Ibarreta, Carlos; Gijón, Covadonga

Conference Paper

Risk behaviour, fraud and e-trust of individual consumers in Spain

26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015

Provided in Cooperation with:
International Telecommunications Society (ITS)

Suggested Citation: Martínez de Ibarreta, Carlos; Gijón, Covadonga (2015) : Risk behaviour, fraud and e-trust of individual consumers in Spain, 26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015, International Telecommunications Society (ITS), Calgary

This Version is available at:
https://hdl.handle.net/10419/127162

Risk behaviour, fraud and e-trust of individual consumers in Spain

Carlos Martínez de Ibarreta,
Comillas Pontifical University, ICADE, Spain
charlie@cee.upcomillas.es

Covadonga Gijón,
IMDEA Software Institute, Spain
covadonga.gijon@imdea.org

1. Introduction

Nowadays a big and increasing proportion of world's population uses internet in their day life and web connection via mobile is having a growing importance. At the same time number and typology of cyber-attacks is growing ever faster. They include all types of malware, such as viruses, trojans, adware, worms, heuristics, rogue ware, and several forms of online fraud, such as phising, stealing of passwords or personal information, ...

Online commerce is growing, and as Buttler (2014) said, the security of this kind of commerce is very important for the organisations, consumers and governments. Educate individuals for using security software and better passwords is one of the most important things in this kind of commerce.

In the other hand, Corritore, Kracher and Wiedenbeck (2003) said that trust will be a key determinant of online markets. If the Internet user not trust in the security of the webpage, s/he will not go inside that webpage.

User behaviour has grown significantly. Jin, Chen, Wang, Hui and Vasilakos (2013) focus users' behaviour in online social networks, and they analyse the behaviour in four different perspectives: connection and interaction, traffic activity, mobile social behaviour, and malicious behaviour.

There are lots of protection tools against cyber-attacks. There are some whose operation that can be automated, such as antivirus or firewall, whereas there are others that requires an active behaviour from the user, such as deleting cookies or doing backups of crucial files.

Furthermore, day by day there is more information about the risks and dangers of a misuse of internet, which may involve all sorts of negative consequences, from computer malfunction to economical or personal damages.

Despite all that, as data shows, a 40% of users acknowledge to have done consciously any kind of behaviour that could have been put their computer at risk to be harmed by any malware or cyber-attack, and consequently, a significant fraction of users has been victim of some kind of cyber-attacks.

This paper tries to give empirical answer about which are the factors that influence the likelihood of an internaut does such a risky behavior.

These factors can be grouped into several categories as follows. For each of one some empirically testable hypothesis are proposed.

A. Psychological factors related with age and gender. Psychology and neurosciences show that there is a general law that says that risk preferences decline with age and they tend to be lower in females than in males. Therefore it can be expected that this law also comply in the internet behaviour.

B. Education and knowledge about the internet and its risks. The global hypothesis that it is made is that the more knowledge level one individual has, the lesser the likelihood to engage in risky behaviour that could harm the computer.

This educational dimension can be detailed in several sub dimensions:

B1) Education in a broad sense. It is supposed that if one individual has achieved a greater education level and he or she also lives in an environment where there are more availability of technological information and resources, such as in a big city, it is less likely that she or he engages into risky behavior.

> *H1a: There is an inverse relationship between education level and risky behaviour.*
>
> *H1b: There is an inverse relationship between hábitat size and risk behaviour*

B2) Knowledge about the Internet reached by means of experience. This experience can be achieved by tenure or by being familiar with services and applications well known by the user. It is supposed that when an individual is using a new application or service is easier that he or she makes a risky behavior, because he or she is not familiar with it and it is easier to make a mistake. Conversely, that is harder when one has enough expertise in using some application.

> *H2a: There is an inverse relationship between internet experience level and likelihood of risky behavior*
>
> *H2b: The greater the number and diversity of services and applications used, the greater the likelihood of making risky behavior.*

B3) Computer protection level. It is supposed that an user that has installed more protection tools, such as antivirus, firewall or their updates, or an user that takes intentional security actions such as delete cookies or delete temporary files, has greater knowledge level about the internet and its risks. Therefore there will be lower probability of he or she make risky behaviours.

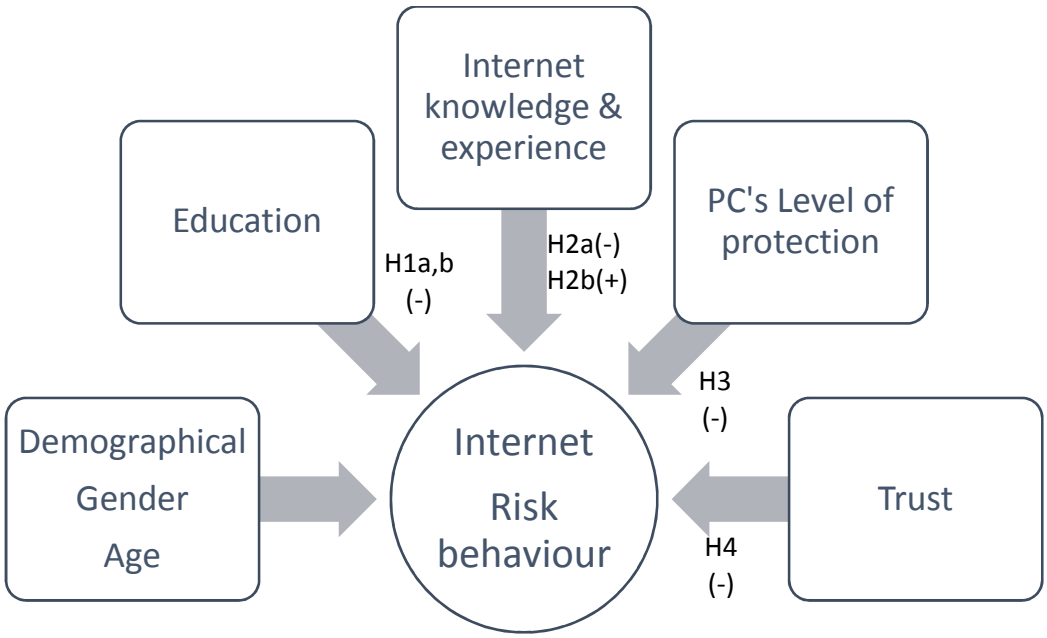
> *H3 The more cyber-security level (both active and pasive) an individual has, the lower his or her probability to make risky behavior.*

B4) Trust level on operations with Banks, public agencies or shops. Mainly it is referred with operations related with payments or with giving personal information, both online or in the real world (offline). It is supposed that people with more technological knowledge, will have greater trust levels, because they are more rational about which activities have low risk and which not. Consequently, those people will be less likely to make risky behaviour when they are using the computer or surfing the Internet.

> *H4 The greater the trust level (either on online or in the real world) on technological operations, the lower the likelihood to behave risky on the internet.*

Figure 1 outlines a conceptual model with the relationships among these constructs as well the hypothesis established

Figure 1. Conceptual model and hypothesis



The rest of paper is organized as follows, in section 2 there is information about the data, the variables and the models used for the empirical research. Section 3 presents the main empirical results. Finally section 4 concludes.

1. Data, variables, models

Data

The sample consists of a survey with data on 3,010 households of Spanish Internet users: Estudio sobre Ciberseguridad y confianza en los hogares españoles (*Study on Cybersecurity and Trust in Spanish households*). The data was collected from December of 2013 to January 2014, by Instituto Nacional de Tecnologías de la Comunicación (INTECO) and Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información (ONTSI), which are government body that performs, among other functions, the collection and analysis of socio-economic data. The survey is about households that have Internet connection and includes questions about socio-demographics, different kinds of security, Internet use, behaviour on Internet use, security incidences, phishing, etc.

The sample is representative by gender. It can be seen on Table 1 the demographic profile of the respondents to the survey. The survey is about the household but the profile is the person that answered the survey. It can be seen that the majority of respondents are older than 35, with more than half of them being between 35 and 54 years old.

Table 1. Demographic profile of respondents.

		Frequency	*Percent*
Gender	Male	1517	50.40
	Female	1493	49.60
Age	15-24	173	5.75
	25-34	617	20.50
	35-44	1083	35.98
	45-54	715	23.75
	>55	422	14.02
Habitat size	<10k	385	12.79
	10-50k	635	21.10
	50-100k	284	9.44
	>100k	340	11.30
	Capital <500k	568	18.87
	Capital >500k	798	26.51
Education level	Primary	53	1.76
	High School	1434	47.64
	College	1523	50.60
Web Browser	Microsoft Internet Explorer	578	19.20
	Opera	23	0.76
	Apple Safari	73	2.43
	Google Chrome	1628	54.09
	Mozilla Firefox	703	23.36
	Other	2	0.07

More than half of the sample lives in capital cities or cities with a population over 100.000 habitants. Also more than 50% of the respondents have education at a graduate level.

On the other hand, it can be seen than the main web browser used by the respondents is Google Chrome (54.09%), followed by Mozilla Firefox (23.36%), Microsoft Internet Explorer (19.20%).

Models

To test the hypothesis proposed in section a logit model has been specified, where the dependent variable, labelled as "*risky behaviour*" stands for if the individual has been or not knowingly engaged in any conduct risk when using his or her computer or when surfing by the internet.

The explanatory variables, following the conceptual model are grouped into the following four vectors of variables: socio-demographics (SD), Internet use patterns (I), cyber-security user behaviour (CS), and levels of trust and e-trust (T).

Therefore the model can be expressed as is shown in equation [1] :

$$P(risky\_behaviour_i = 1) = \frac{e^{z_i}}{1+e^{z_i}} \qquad [1]$$
$$z_i = \alpha + \mathbf{SD}\boldsymbol{\beta} + \mathbf{I}\boldsymbol{\gamma} + \mathbf{CS}\boldsymbol{\delta} + \mathbf{T}\boldsymbol{\theta} + \varepsilon_i$$

In order to prevent inference errors derived from heteroskedasticity, robust standard errors have been used to prevent inference errors derived from that problem.

Variables definition

This subsection is devoted to explain the definition of the sets of variables mentioned above, that is, the Internet use patterns (I), cyber-security user behaviour (CS), and levels of trust and e-trust (T).

Each of the variables labelled as "indexes" are built as the scores of the first principal component resulting from a principal components analysis PCA. In each of the PCA performed, it has been employed as original variables those indicators (mainly of a binary nature) which appear in the questionnaire and related with the corresponding construct. Table 2 details the indicators that form each of the indexes, as well the percentage of users in each of the indicators with value one.

A) Internet use patterns

- *Internet Intensity Use*. It has been computed as a recency – frequency measure, multiplying the users answers to two items in the questionnaire: *how long are you using the Internet?* (recency) and *how often do you connect to the Internet at home?* (frequency). This kind of measure can be found among others in Jennings, M. K., & Zeitner, V. (2003). ts values go from a minimum value of 1, when the individual has internet connexion at home for less than one year and he or she connects to the Internet lesser than once at month, up to a maximum value of 12, when the user has internet connexion at home for more than five years and he or she connects to the Internet at least once a day.
- *Diversity Internet Use Index*. The higher its value, the greater the number of internet applications, programs and services used by the individual in the last three months. From Table 2 it can be highlighted than the services more often used are e-mail (94%), social networks (84%) and e-banking (79%).

B) Cyber-security user behaviour

- *PC Protection Tools Index*. The greater its value, the more the number of computer protection tools, such antivirus or firewall, that the respondent has used in the last three months. The most often used tools are antivirus (82%), antivirus updates (68%) and operative system updates (57%).
- *PC Security Measures Index*. The higher its value the greater the number of security measures than the individual has performed during the last three months. These measures, such as using passwords or deleting cookies, generally require an active behaviour from user because they cannot be automated, unlike protection tools, that are of more passive nature and can be automated. The most often employed security measures are using passwords (57%), deleting cookies and temporary files (55%) and doing backups (39%).
- *Security programs update frequency.* This ordinal variable takes values from 1 (“*I do not know*) to 7 (“*my computer does it automatically”*)
- *PC scan frequency*. This other ordinal variable takes values from 1 (“*never”)* to 6 (“*my antivirus does it automatically*”)

C) Internet use patterns

- *Internet Intensity Use*. It has been computed as a recency – frequency measure, multiplying the users answers to two items in the questionnaire: *how long are you using the Internet?* (recency) and *how often do you connect to the Internet at home?* (frequency). This kind of measure can be found among others in Jennings, M. K., & Zeitner, V. (2003).

- *Diversity Internet Use Index*. The higher its value, the greater the number of internet applications, programs and services used by the individual in the last three months. From Table 2 it can be highlighted than the services more often used are e-mail (94%), social networks (84%) and e-banking (79%).

D) Cyber-security user behaviour

- *PC Protection Tools Index*. The greater its value, the more the number of computer protection tools, such antivirus or firewall, that the respondent has used in the last three months. The most often used tools are antivirus (82%), antivirus updates (68%) and operative system updates (57%).
- *PC Security Measures Index*. The higher its value the greater the number of security measures than the individual has performed during the last three months. These measures, such as using passwords or deleting cookies, generally require an active behaviour from user because they cannot be automated, unlike protection tools, that are of more passive nature and can be automated. The most often employed security measures are using passwords (57%), deleting cookies and temporary files (55%) and doing backups (39%).
- *Security programs update frequency.* This ordinal variable takes values from 1 ("*I do not know*) to 7 ("*my computer does it automatically"*)
- *PC scan frequency*. This other ordinal variable takes values from 1 ("*never")* to 6 ("*my antivirus does it automatically*")

E) Levels of trust and e-trust

- *Offline Trust Index*. The greater its value the higher the trust of individuals to give personal information or to do payments with credit or debit cards in real world (banks, shops, public agencies). Table 3 shows that (in a scale from 1 to 5, where 3 means "*medium trust level*") the activities that give more trust are make banking operations at the bank office (3,7) and give personal information in a public agency for a procedure (3,4) whereas the minimum trust corresponds with giving personal information in a private entity for a procedure (3,0).
- *Online Trust Index*. It is defined in a similar way to the Offline Trust Index. From Table 3 it can be seen that, broadly speaking, online activities generate less trust than their offline equivalents. Those that give more trust are give personal information in a public agency's webpage (3,3) and making online purchases without using credit/debit cards (3,3), whereas giving personal information via e-mail or instant messenger is the activity that produces less trust (2,7).

Table 2. Summary of indexes and indicators

Construct	% explained variance (KMO)	Indicators	%
Diversity internet use	20,9 (0,82)		
		e-mail	94
		social networks	84
		e-banking	79
		music or video streaming	71
		e-commerce	58
		online payments	54
		e-administration	45
		P2P	42
		internet forums, blogs	38
		online calls or video calls	30
		chat	29
		downloading files from server	27
		online courses	25
		online games	23
		online casinos	9
PC Protection Tools	37,9 (0,81)		
		Antivirus	82
		Antivirus updates	68
		Operating system updates	57
		Firewall	42
		Configurations to block pop-ups	35
		Anti spam filters	32
		Anti spy programs	27
		Security plug-ins for internet browser	25
		Extensions to block online advertisement	19
PC Security Measures	24,6 (0,70)		
		Passwords	57
		Delete cookies and temporary files	55
		Files backup	39
		Have a partition of hard disk only for data	24
		Use of digital certificates for identification/sign	23
		User profile with restricted rights	16
		Use of electronic DNI	14
		Use of encryption tools	7

Notes: KMO stands for Kaiser-Meyer-Olkin measure of sampling adequacy for performing PCA

Table 3. Summary of trust indexes

Construct	% of variance explained (KMO)	Indicators 1- no trust 5- very much trust	mean
Online Trust index	55 (0,83)		
		give personal information in a public agency's webpage	3,3
		make online purchases without using credit/debit cards	3,3
		e-banking	3,2
		give personal information for join in some service	3,1
		making online purchases using credit/debit card	3,1
		give personal information via e-mail or instant messenger	2,7
Offline Trust index	66 (0,81)		
		make banking operations at the bank office	3,7
		give personal information in a public agency for a procedure	3,4
		make banking operations at an ATM	3,4
		pay with credit/debit card in a shop	3,3
		give personal information in a private entity for a procedure	3,0

In Table 4 it can be seen the main descriptive statistics of variables, other than socio-demographics. It can be noted that all of them which are synthetic indexes have zero mean.

Table 4. Descriptive statistics

	Obs.	*Mean*	*Std. Dev.*	*Min.*	*Max.*
Risky Behaviour	3010	0.40	0.49	0	1
Internet Intensity Use	3010	11.55	1.45	1	12
Diversity Internet Use Index	3010	0.00	1.77	-4.45	4.60
PC Protection Tools Index	3010	0.00	1.85	-2.79	3.85
PC Security measure Index	3010	0.00	1.40	-1.92	5.06
Security program update frequency	3010	6.13	1.66	1	7
PC Scan frequency	2483	4.67	1.71	1	6
Offline Trust Index	3010	0.00	1.82	-3.77	5.64
Online Trust Index	3010	0.00	0.49	-4.77	5.47

Table 5 contains the correlation matrix between the variables and the significance of each correlation. The coefficients are below .3 in all cases but the correlation between Online and Offline Trust Indexes, which is close to 0.8

Table 5. Correlation matrix

	Internet Risky behaviour	Male	Age	Educ.	Habitat Size	Internet Intensity Use	Diversity Internet Use I	PC Protect. Tools I	PC Sec measure I	Sec. Prog. Update Freq.	PC Scan Freq.	Offline Trust I	Online Trust I
Male	0.070*												
Age	-0.101*	0.132*											
Education	0.028*	-0.050*	-0.064*										
Habitat Size	-0.052*	0.033*	0.044*	0.086*									
Internet Intensity Use	-0.016*	0.048*	-0.027	0.083*	-0.002								
Diversity Internet Use I	0.164*	0.096*	-0.126*	0.074*	0.013	0.156*							
PC Protect. Tools I	0.086*	0.194*	0.094*	0.006*	0.040*	0.112*	0.370*						
PC Sec measure I	0.039*	0.111*	0.058*	0.089	0.028*	0.086*	0.306*	0.460*					
Sec. Prog. Update Freq.	-0.028	0.010*	0.137*	-0.064	0.006	0.102*	0.082*	0.231*	0.172*				
PC Scan Freq.	-0.074*	-0.046*	0.045	-0.045*	-0.055*	0.031	-0.012	0.049*	0.096*	0.325*			
Offline Trust I	0.003	0.078*	-0.034*	0.091*	0.027	0.031*	0.188*	0.095*	0.086*	0.073*	-0.006		
Online Trust I	0.001	0.080*	-0.066*	0.077*	0.010	0.029*	0.218*	0.063*	0.060*	0.051*	0.027	**0.796***	
G. Chrome	0.053*	0.005	-0.089*	-0.037*	0.009	-0.020	0.020	-0.021	-0.047	-0.023	-0.012	0.008	0.021

Notes: * Significant at 5%. Sample size: 3010 internet users.

3. Results

This section is devoted to present and discuss the results of the estimates of the risk behaviour model. This allows to test which of the hypothesis have empirical support from the data and which ones do not.

Table 6 shows the main estimation results of the binary logit model, using robust standard errors to avoid inference problems due to the presence of heteroskedasticity.

Table 6. Logit model for determinants of Internet risky behaviour

Dependent variable:	Internet Risk Behaviour (yes/no)
Male	-.262 (.265)
Age	-.259*** (.059)
Male * Age	.166** (.078)
Education	.131 (.082)
Habitat size	-.072*** (.023)
Internet Intensity Use	-.071** (.034)
Diversity Internet use Index	.184*** (.029)
PC Protection Tools Index	.070** (.030)
PC Security measures Index	-.028 (.034)
Security Program Update Frequency	-.027 (.035)
PC Scan frequency	-.080*** (.026)
Offline trust Index	-.020 (.039)
Online trust Index	-.027 (.039)
Google Chrome Browser	.190** (.085)
Constant	.678 (.496)
Wald χ^2 (p-value)	132.51 (0.0000)
White test, χ^2 (p-value)	186.28 (0.0000)
Degrees of freedom White	100
Pseudo R^2	0.0424
n	2483
Mean VIF/Max VIF	3.07/12.21

Notes: In parenthesis robust std. errors. * Significant at 10%, ** significant at 5% and *** significant at 1%. Heteroskedasticity consistent covariance matrix estimates (Eicker-White) is used.

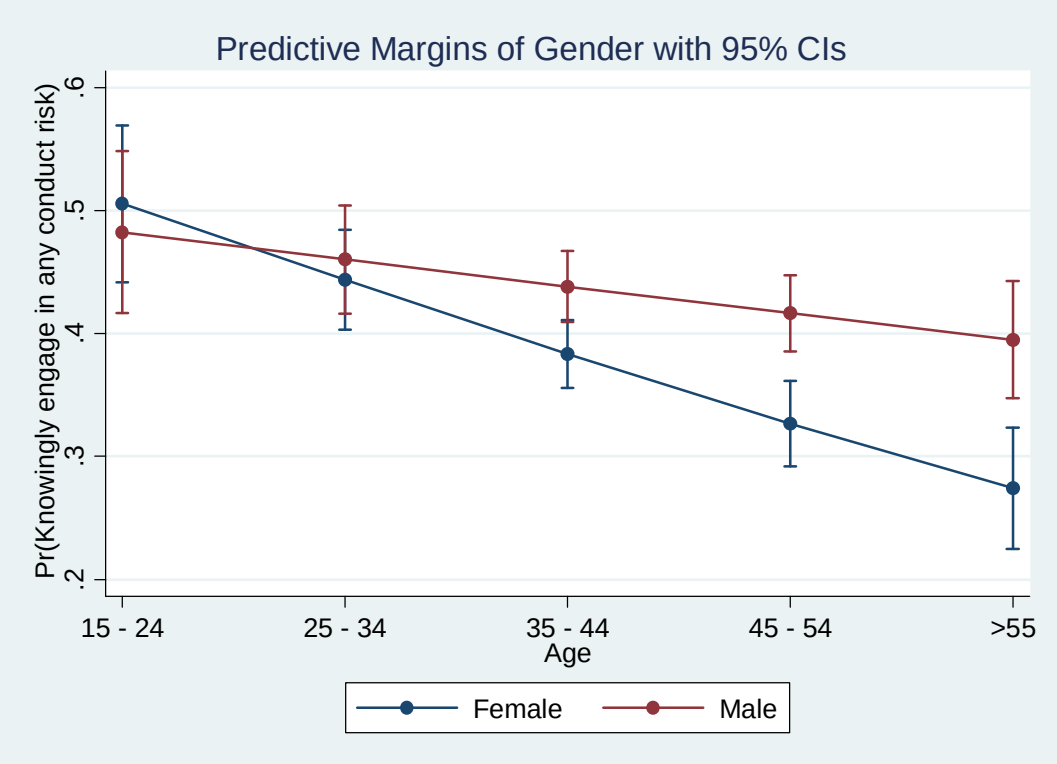
Psychological factors

In order to capture how age can be modifying the impact of gender on the likelihood of a risky behaviour, an interaction term between both variables has been included in the final model specification.

It can be seen that coefficient of age is negative and significant at 1% but the coefficient of the interaction term between being male and age is significant and positive. Thus, probability of doing risky behaviour while surfing internet decreases with age, but in a steeper way for women.

Figure 2 shows graphically this findings. At the youngest age interval, likelihood of risky behaviour is high and there are no gender differences. However, at the eldest interval, both probabilities are lower but much more lower for women than for males.

Figure 2. Interaction effects between Age and Gender on Risky Internet Behaviour Probability



Educational factors.

The level of formal education is not a significant variable in the model. Therefore there is not enough empirical support for hypothesis H1a, which proposed an inverse relationship between education level and risky behavior.

Conversely, habitat size is a significant variable at 1% level, and its coefficient is negative. Thus, hypothesis H1b is strongly supported by data, and there is evidence that living in a big city reduces the probability of doing a risky behaviour.

Knowledge about the Internet reached by means of experience.

The coefficient for the internet intensity use variable is significant and negative at a 5% level. Therefore the more intense is the use of Internet (in recency-frequency terms) the lower the probability of being engaged in risky actions when surfing the internet. Consequently, hypothesis H2a has empirical support.

Furthermore, as it is stated by hypothesis H2b, the coefficient on the diversity internet use index variable is also significant and it has positive sign. That means that as the number of applications, services and programs used by the individual grows, the probability of doing some risky action also grows. Hence hypothesis H2b is also supported by data.

It is worth mention that the positive relationship found can be due to other reason besides lack of enough expertise. It is well known that curious people (with higher levels of openness to experience, if terms of "Big Five" jargon) use to be more risky. Perhaps those people tend to use a wider palette of services or applications when they are surfing the Internet.

Computer protection level.

The coefficient of PC Scan frequency variable is positive and significant at1% level as expected. However neither PC Security Measures index nor Security Program Updating Frequency are significant at any commonly used level.

More surprisingly, the coefficient of PC protection Tools index is significant at 5% but of positive sign. That is, the opposite sign that expected and proposed by H3.

Therefore there is mixed evidence in favor of H3.

The unexpected sign of PC protection tools index variable can be justified if one thinks about the "*false confidence*" phenomenon. Perhaps if the user knows that he has enough protection tools, he can feels protected and then he or she can engage in risky actions that probably he or she would not do if they had less protection level.

*Trust level*s.

Contrary to expectations, neither online trust index nor offline trust index are significant variables in the model. Thus there is not empirical support for H4,

that hypothesized a negative relationship between trust levels and risky behavior.
Perhaps this result can be caused by the confluence of two opposite mechanisms; on the one hand, lack of trust in online or in the real world operations that imply giving personal information or making payments can cause to be more conservative and not involve in online risky behaviours; on the other hand, having more knowledge about the internet can cause not only having more trust on these kind of operations (payments) in the safe places but also to be less risky.

4. Conclusion

This paper gives some empirical findings about the determinants that make an internet user to behave or not in a risky way when surfing the internet.

This has been made by estimating a logit model using data from a survey on 3,010 households of Spanish Internet users.

Results show that the more elder you are, the lesser likely you behave in a risky way in the Internet, and this result is more intense if you are woman.

On the other hand, the more knowledge and experience you have acquired about internet and their dangers, less likely you engage in risky behaviour. However having a greater general education level has not influence in such probability.

A surprising result is that having lots of protection tools can cause "false confidence" feelings, and therefore, make you behave with more risk, conversely to what was supposed.

Finally, trust and e-trust levels on technological transactions of money or information do not seem to affect the likelihood of behaving in a risky way.

5. References

- Butler, M. J. (2014). Towards online security: key drivers of poor user behaviour and recommendations for appropriate interventions. **South African Journal of Business Management**, 45(4), 21-32.
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. **International Journal of Human-Computer Studies**, 58(6), 737-758.
- Jin, L., Chen, Y., Wang, T., Hui, P., & Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey. **Communications Magazine**, IEEE, 51(9), 144-150.

- Kridel, D. J.; Rappoport, P. N. y L. D. Taylor (2002), IntraLATA long-distance demand: carrier choice, usage demand and price elasticities, **International Journal of Forecasting**, Elsevier, vol. 18(4), pages 545-559.
- Taylor, L. D: (1994), Telecommunications demand in theory and practice. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Jennings, M. K., & Zeitner, V. (2003). Internet use and civic engagement: A longitudinal analysis. Public Opinion Quarterly, 67(3), 311-334.