unibz
Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 103/ 2024

Phishing Attacks: An Analysis of the Victims' Characteristics Based on Administrative Data

Alessandro Fedele, Mirco Tonin,
Matteo Valerio

Phishing Attacks: An Analysis of the Victims' Characteristics Based on Administrative Data[1]

Alessandro Fedele[a] Mirco Tonin[a,b] Matteo Valerio[a]

[a] Free University of Bolzano [b] IRVAPP - Bruno Kessler Foundation; IZA; CESIfo

February 2024

Abstract

Using administrative data on phishing attacks targeting almost 150,000 Italian- and German-speaking customers of an Italian bank in 2022-23, we investigate how individual characteristics are associated to the likelihood of victimization. We find that younger customers and Italian speakers are more likely to be victims of phishing, while we find no differences in terms of gender or size of the place of residence.

JEL Code: L86. **Keywords:** phishing attacks; administrative data; victims' characteristics.

1 Introduction

Phishing is a cybercriminal's attempt to get sensitive information, such as personally identifiable information, or banking and credit card details. Phishing is typically carried out by email, telephone, or SMS, in which attackers impersonate a legitimate entity, such as a bank.

Phishing is a widespread and still growing phenomenon. An estimated 3.4 billion phishing emails a day are sent by cybercriminals (Smith, 2023). The report by Zscaler ThreatLabz (2023) reveals a worldwide 47.2% increase in phishing attacks in 2022 compared to 2021. Proofpoint (2023) reports that 84% of the organizations surveyed faced successful phishing attack(s) in 2022 and that the resulting financial losses increased by 76% compared to 2021.

In this paper, we use administrative data on phishing attacks aimed at stealing money from the customers of an Italian bank. We explore the link between the probability of victimization and individual characteristics. A growing body of literature on phishing, mostly in computer science and psychology, explore factors affecting individuals' susceptibility to phishing, such as analytical reasoning (Kelley et al., 2023), age (Sheng et al., 2010), and

[1] This study was carried out within the project "Cyber resilience: markets, investments and regulation", funded by European Union, Next Generation EU within the PRIN 2022 PNRR program (D.D.1409 del 14/09/2022 Ministero dell'Università e della Ricerca). This manuscript reflects only the authors' views and opinions and the Ministry cannot be considered responsible for them.

gender (Sheng et al., 2010; Iuga et al., 2016), among others. To our knowledge, the economics/finance literature on financial fraud has not yet considered phishing. In finance, the interest is in frauds committed by financial advisers (Adeabah et al., 2022). The only economics paper we are aware of is Bian et al. (2023). They exploit Apple's policy that bounded the sharing of personal information on iOS. Using U.S. data on complaints that (individuals who believe they have been) victims of financial fraud have the option to submit, they find that more people disallowing tracking reduce the number of complaints.

The above papers either rely on experimental setups where participants are asked, e.g., to compare real and fake versions of websites, or on survey data. While lab settings are useful in identifying causal relationships, they have issues in terms of external validity. Conversely, survey data have the merit of considering actual frauds, but may suffer from recall or self-selection biases that affect the reliability and/or generalizability of results. To our knowledge, our paper is the first attempt to overcome these potential problems by using administrative data, where phishing scams are recorded through the bank administrative systems, rather than being self-reported.

We have data on the bank customers' gender, language of communication with the bank (Italian or German, given that the German-speaking province of South Tyrol in Italy is part of the bank geography), age, and place of residence. The existing evidence on gender and age is mixed. The well-known overconfidence of males (Barber and Odean, 2001) could lead to them being more prone to victimization: however, Sheng et al. (2010) and Iuga et al. (2016) find opposite evidence. Regarding age, older people are believed to be more likely victims of financial fraud (Kieffer and Mottola, 2017; Dadalt, 2016); conversely, Sheng et al. (2010) observe that younger individuals are more susceptible to phishing.

We study the likelihood of victimization for almost 150,000 bank customers in 2022-2023. (i) We find that Italian-speaking customers are more likely to be victims than German-speaking ones. The most likely reason is that phishing messages are in Italian: customers who communicate with the bank in German are more suspicious due to this incongruency, and thus less likely to fall into the trap. (ii) We also show that age is negatively correlated with the likelihood of victimization; this could be due to younger generations being more familiar with online banking and, therefore, less cautious with its use. (iii) We find no gender difference, nor differences based on the size of the place of residence. Unfortunately, we do not observe potentially interesting characteristics like education or wealth. Nevertheless,

relying on a unique dataset of actual frauds, we can contribute to the understanding of a growing and understudied phenomenon like phishing.

The paper is organized as follows. Section 2 describes the institutional background and the dataset. Section 3 provides the results. Section 4 concludes.

2 Institutional Background and Data

We consider a brick-and-mortar bank, also providing online banking; the bank operates in Northeast Italy and caters to retail, private and corporate customers. The typical cyberattack under consideration is as follows: customers receive a SMS warning about unauthorized activity and asking to verify account details. The message includes a link which redirects to a bank's fake login page. If a customer attempts to login into this page, she/he reveals her/his personal credentials and information.

Due to the prevalence of Multi-Factor Authentication, personal credentials are not enough to execute transfers into the fraudsters' accounts. Therefore, fraudsters contact the customer via phone, impersonating the bank contact center. Because of the customer attempt to login into a fake page, fraudsters can access her/his internet banking and are thus able to provide information (e.g., recurring payments) that makes it credible the caller is indeed the bank. With the customer's cooperation, fraudsters can then activate the Multi-Factor Authentication and, eventually, get banking transactions authorized. Upon observing abnormal operations on the customer's account, however, the bank fraud team can often block the transfer.

We rely on the master database of all the customers. We excluded deceased customers, minors, companies, and customers who are also employees; we also excluded those residing abroad because foreign telephone numbers are not targeted by fraudsters. The total number of observations is 147,751. We recorded 276 successful attacks in the period January 1, 2022–December 31, 2023. An attack is considered successful when the customer discloses credentials and enables fraudsters to enter her/his bank account and attempt to transfer funds, regardless of whether money is finally lost.

3 Results

Table 1 shows the summary statistics by comparing the whole sample to the subset of victims. The gender breakdown is proportionate: 47% of both clients and victims are females. The percentage of Italian- (German-) speaking individuals is 74% (26%) among customers

and 83% (17%) among victims. The median age of victims, 43, is lower than the median age of customers, 47. 34% of the victims are in the first quartile of the age distribution, 18-34 years old, and the percentage monotonically decreases when considering the other three quartiles. Finally, slightly more victims live in large municipalities.

Table 1: Descriptives

Variables		Whole Sample	Victims
Female		47%	47%
Language – Italian		74%	83%
Age (Median)		47	43
Age	18 - 34	25%	34%
	35 - 47	25%	28%
	48 - 58	25%	23%
	59 - 101	25%	15%
Settlement (population)	125 - 4,871	25%	24%
	4,871 - 12,820	25%	22%
	12,820 - 35,522	25%	26%
	35,522 - 2,770,226	25%	28%
Number of observations		147751	276

Age and *Settlement* are split into quartiles calculated for the whole sample.

Next, we estimate the following logit regression:

$$\Pr(Fraud_i = 1|X) = F\left(\beta_0 + \beta_1 Female_i + \beta_2 ItalianLanguage_i + \beta_3 Age_i + \sum_{h=2}^{4} \gamma_h Settlement_h\right)$$

where $Fraud_i$ is a dummy variable taking the value of 1 if customer i fell prey to phishing and 0 otherwise. $Female_i$ equals 1 if the customer is a woman. $Italian\ Language_i$ equals 1 if the customer's language of communication with the bank is Italian, 0 if German. Age_i is the age of customer i in years; in an alternative specification that relaxes the linearity assumption, we replace this variable with dummies capturing the age quartiles for the whole population of customers. $Settlement_h$ is a set of dummies based on the quartiles in terms of population size of the customer's residence place, as defined in Table 1.

In Table 2, models (1) and (2), we report the marginal effects, with age expressed in years. In model (2), we keep only the variables that turned out significant in model (1). In models (3) and (4), we use instead age quartiles.

Table 2: Results

Dependent variable: Fraud=1 if victim of phishing	(1)	(2)	(3)	(4)
Female	-0.0000100 (0.0002250)	-	-0.0000133 (0.0002245)	-
Italian Language	0.0009524** (0.0003169)	0.0009565** (0.0003015)	0.0009603** (0.0003171)	0.0009636** (0.0003017)
Age	-0.0000301*** (0.0000066)	-0.0000298*** (0.0000067)	-	-
Age_q2	-	-	-0.0002619 (0.0002878)	-0.0002728 (0.0002873)
Age_q3	-	-	-0.0007339* (0.0003526)	-0.0007366* (0.0003104)
Age_q4	-	-	-0.0014064*** (0.0003526)	-0.0013934*** (0.0003523)
Settlement_q2	-0.0003276 (0.0003388)	-	-0.0003305 (0.0003391)	-
Settlement_q3	-0.0000878 (0.0003360)	-	-0.0000845 (0.0003362)	-
Settlement_q4	0.0001176 (0.0003223)	-	0.0001243 (0.0003228)	-
N	147751	147751	147751	147751
PseudoR2	0.0076144	0.0070971	0.0081739	0.0076372

Estimation method: Logit. Results are expressed as marginal effects on the probability of victimization. ***, ** and * denote statistical significance at the 1, 5 and 10 percent level. The standard errors reported in parentheses are corrected for heteroskedasticity.

The coefficient for gender is small and non-significant. The coefficient for language reveals that Italian-speaking customers are 0.1 percentage points (pps) more likely to be victims compared to German-speaking customers. This is a strong effect, given a 0.19% baseline

fraud rate (276/147751). The coefficient for age is negative and highly significant: being 10 years older reduces the probability of victimization by 0.03 pps (or 16%). Relaxing the assumption of linearity for age in models (3) and (4), we see that compared to the first quartile (18-34 years old), the likelihood of victimization is lower but not significant for the second quartile, and lower and significant for the third and fourth. Customers in the age range 48-58 (above 59) are 0.07 (0.14) pps less likely to fall for phishing: these are non-trivial effects considering that the victimization rate is 0.19%.

4 Conclusion

The growth of digitalization implies that more and more activities will migrate online, including fraud attempts. In the banking sector in particular, an increasing share of services are offered online. To protect customers, it is important to understand the characteristics of those more prone to victimization. Knowing this, financial institutions and regulators can better target information campaigns aimed at raising awareness of the risks related to phishing attacks. While there is a lot of discussion related to the protection of elderly people (DaDalt, 2016), our findings suggest that more attention is warranted vis-a-vis younger generations.

Bibliography

Adeabah, Andoh, Asongu, Gemegah (2023). Reputational risks in banks: A review of research themes, frameworks, methods, and future research directions. *Journal of Economic Surveys* 37, 321-350.

Barber, Odean (2001) Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics* 116(1), 261-92.

Bian, Pagel, Tang (2023). Consumer Surveillance and Financial Fraud. *NBER Working Paper* 31692.

DaDalt (2016). Older adults and fraud: Suggestions for policy and practice. *Journal of Economic & Financial Studies* 4(03), 38-44.

Iuga, Nurse, Erola (2016). Baiting the hook: factors impacting susceptibility to phishing attacks. *Human-centric Computing and Information Science* 6(8)

Kelley, Hurley-Wallace, Warner, Hanoch (2023). Analytical reasoning reduces internet fraud susceptibility. *Computers in Human Behavior* 142 107648.

Kieffer, Mottola (2017). Understanding and Combating Investment Fraud. In Olivia S. Mitchell, P. Brett Hammond, and Stephen P. Utkus (eds), *Financial Decision Making and Retirement Security in an Aging World*, Pension Research Council Series, 185-212.

Müller, Wood, Hanoch, Huang, Reed (2020). Older and wiser: Age differences in susceptibility to investment fraud: The protective role of emotional intelligence. *Journal of Elder Abuse & Neglect* 32(2), 152–172.

Proofpoint (2023). 2023 state of the phish. https://www.proofpoint.com/us/resources/threat-reports/state-of-phish

Sheng, Holbrook, Kumaraguru, Cranor, Downs (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, New York, 373–382.

Smith (2023). Top Phishing Statistics for 2023: Latest Figures and Trends, October 23, 2023. *StationX* https://www.stationx.net/phishing-statistics/

Zscaler ThreatLabz (2023). Zscaler ThreatLabz 2023 Phishing Report. Available at: https://info.zscaler.com/resources-industry-reports-threatlabz-phishing-report.