POLITECNICO DI TORINO

Master's Degree
Engineering and Management

Software License Compliance
A quantitative study on software piracy in Italy

Supervisor

Prof. L. Benfratello

Candidate

Francesco Di Francesco

April 2019

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT 1

1. THE SOFTWARE INDUSTRY & BUSINESS MODEL 2

1.1 THE SOFTWARE INDUSTRY 2

1.2 FROM DEVELOPMENT TO MARKET 2

1.2.1 SOFTWARE DEVELOPMENT 3

1.2.2 SOFTWARE MAINTENANCE 5

1.2.3 SOFTWARE PUBLICATION 6

1.3 BUSINESS MODELS 6

1.3.1 DISTRIBUTION APPROACH 7

1.3.2 SOURCE CODE LICENSING 9

1.3.3 REVENUE STREAMS 10

1.3.4 BUSINESS MODEL INTERACTION 11

2. CAD SOFTWARE INDUSTRY 13

3. SOFTWARE COMPLIANCE 15

3.1 INTELLECTUAL PROPERTY AND LICENSE AGREEMENT 15

3.2 COMPLIANCE ACTIVITIES 18

3.2.1 ADVISORY 19

3.2.2 ASSESMENT 20

3.2.3 ANTIPIRACY 21

3.3 COMPLIANCE WORLDWIDE 23

3.3.1 WORLDIWIDE COMPLIANCE IN NUMBERS 26

4. ECONOMIC STUDIES COVERING SOFTWARE PIRACY 28

4.1 PIRACY - GDP 28

4.2 PIRACY – IT DEVELOPMENT 29

4.3 PIRACY – REGULATORY FACTORS 29

4.4 PIRACY – LEVEL OF LITERACY 30

4.5 DATA COLLECTION 30

5. EMPIRICAL ANALYSIS OF A CAD SOFTWARE PIRACY IN ITALY 32

5.1 PARAMETRIC TECHNOLOGY CORPORATION HISTORY 32

5.2 DATA AND METHODOLOGY 34

5.3 RESULTS 35

5.3.1 GEOGRAPHICAL POSITION 35

5.3.2 COMPANY SIZE 40

5.3.3 COMPANY ECONOMIC AND FINANCIAL STATUS 42
5.3.4 LOCATION-SIZE-WEALTH JOINT ANALYSIS 44
5.3.5 ECONOMIC ACTIVITY 47
5.3.6 ECONOMIC ACTIVITY JOINT ANALYSIS 50
5.3.7 PIRACY AND INDUSTRY HEALTH 56
5.3.8 DIFFERENTIATION AMONG PIRATES – PIRACY RATE 61
5.3.9 ROBUSTNESS TEST 66
CONCLUSIONS 68
REFERENCES 70
BIBLIOGRAPHY 70
WEB REFERENCES 70

LIST OF FIGURES

Figure 1: Recursive process for software development. 3
Figure 2: Main CAD software used by 750 of 3D Hubs customers. 14
Figure 3: Working Environment of Creo CAD Software developed by PTC Inc. 14
Figure 4: A software protected by copyright alone; its algorithm can be emulated using a different coding language. 15
Figure 5: A patent protect the algorithm too if it solves a specific technical problem. 16
Figure 6:Source BSA Global Software Survey June 2018. National rate of unlicensed software against national malware encounter rate. 25
Figure 7: Source Revulytics. Top 20 countries for Software license misuse. 26
Figure 8: Source A.R.- Andrés. R.K. Goel / Journal of Policy Modelling 34 (2012). Convex curve describing the relation between GDP and rate of Piracy. 29
Figure 9: Current logo of Paramentric Technology Corporation. 33
Figure 10: Map of the Italian country divided into its 20 regions. 35
Figure 11: Map of the Italian country divided into 3 Macro-areas. 38
Figure 12: Example of an Ateco 2007 code for cereals cultivation. 47
Figure 13:C sector, Industrial production index 2007-2019. 57
Figure 14:F sector, Climate of confidence 2006-2019. 57
Figure 15: F sector, Industrial production index 2005-2018. 57
Figure 16:G sector, Climate of confidence 2009-2019. 58
Figure 17: G sector, Turnover Index 2005-2018. 58
Figure 18: J sector, Climate of confidence 2009-2019. 58
Figure 19: J sector, Turnover Index 2005-2018. 59
Figure 20: L sector, Climate of confidence 2009-2019. 59
Figure 21: M sector, Climate of confidence 2008-2018. 59
Figure 22: M sector, Turnover Index 2010-2017. 60

LIST OF TABLES

Table 1: Source BSA Global Software Survey June 2018. Rates and Values ($M) of unlicensed software installed worldwide, in Europe and in Italy. 27
Table 2:Software pirates and customers for each Italian region. 36
Table 3: Chi2 Test - dependence of Piracy and Italian regions. 37
Table 4: Software pirates and customers for each Italian region excluding Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta. 37
Table 5: Chi2 Test - dependence of Piracy and Italian regions excluding Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta. 38
Table 6: Software pirates and customers for each Macro-Area. 39
Table 7:Chi2 Test - dependence of Piracy and Macro-Areas. 39
Table 8: Chi2 Test - dependence of Piracy and Macro-areas excluding Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta. 39
Table 9: Logistic Regression - Correlation among Piracy and the two Macro-areas NORD and CENTRO. 40
Table 10: Software pirates and customers for the 4 categories of SIZE defined. 41
Table 11: Chi2 Test - dependence of Piracy and Size. 42
Table 12: Logistic Regression - Correlation between Piracy and Redditiv_1 wealth variable. 43
Table 13: Logistic Regression - Correlation between Piracy and Redditiv_2 wealth variable. 43
Table 14:Logistic Regression - Correlation between Piracy and Redditiv_3 wealth variable. 43
Table 15: Logistic Regression - Correlation among Piracy, Size varibles, NORD and CENTRO Macro-areas. 44
Table 16: Logistic Regression - Correlation among Piracy, NORD, CENTRO and Redditiv_1. 44
Table 17: Logistic Regression - Correlation among Piracy, NORD, CENTRO and Redditiv_2. 45
Table 18: Logistic Regression - Correlation among Piracy, NORD, CENTRO and Redditiv_3. 45
Table 19: Logistic Regression - Correlation among Piracy, Size varibles, NORD, CENTRO and Redditiv_1. 45
Table 20: Logistic Regression - Correlation among Piracy, Size varibles, NORD, CENTRO and Redditiv_2. 46
Table 21: Logistic Regression - Correlation among Piracy, Size varibles, NORD, CENTRO and Redditiv_3. 46
Table 22:Description of the 23 Ateco 2007 code analysed and relative number of identified companies (frequency) 48
Table 23: Pirates companies and customers for the identified ateco 2007 codes. 49
Table 24: Chi2 Test - dependence of Piracy and Economic Activity (ateco 2007 code). 50
Table 25: Pirates companies and customers for the identified 9 economic activities aggregates. 51

Table 26: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and dummies for the 9 economic activities aggregate 51
Table 27: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and Manufacturing activities. 52
Table 28: Pirates companies and custoemrs for the identified 8 ateco macro sector (ateco_letter). 53
Table 29: Economic activities in ateco "C" (Manufacturing). 53
Table 30: Economic activities in ateco "F" (Construction). 54
Table 31: Economic activities in ateco "G" (Wholesale and retail trade; repair of motor vehicles and motorcycles). 54
Table 32: Economic activities in ateco "J" (Information and communication services). 54
Table 33: Economic activities in ateco "K" (Financial and insurance activities). 54
Table 34: Economic activities in ateco "L" (Real Estate). 54
Table 35: Economic activities in ateco "M" (Professional, scientific and technical activities). 54
Table 36: Economic activities in ateco "N" (Rental, travel agencies, business support services). 55
Table 37: Chi2 Test - dependence of Piracy and Economic Activity (ateco_letter). 55
Table 38: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and dummies for the 8 ateco_letter economic activities. 55
Table 39: Years of recession for the Economic Activities (ateco_letter). 60
Table 40: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and Recession variable. 60
Table 41: Magnitude of Piracy in the 3 Italian Macro-areas. 62
Table 42: Chi2 Test - dependence of Piracy_rate and Macro-areas. 62
Table 43: Magnitude of Piracy for the 4 different companies'size. 62
Table 44: Chi2 Test - dependence of Piracy_rate and Companies'size. 63
Table 45: Chi2 Test - dependence of Piracy_rate and Economic activities for all the ateco 2007 code analysed. 63
Table 46: Magnitude of Piracy for the 23 ateco 2007 code with at least 5 observations. 63
Table 47:Chi2 - Dependence of Piracy rate and the 23 economic activities with at least 5 observations. 64
Table 48: Linear Regression - Correlation among piracy_rate and wealth variable Redditiv_1. 65
Table 49: Linear Regression - Correlation among piracy_rate and wealth variable Redditiv_2. 65
Table 50: Linear Regression - Correlation among piracy_rate and wealth variable Redditiv_3. 65
Table 51: Robustness test of previous analysis with "Serial" pirate sample. 66

ABSTRACT

Living in the era of technology and digital transformation implies to deal with software every day. What is commonly known is that a software runs on a machine and makes it possible to perform some activities; a lot of individuals and companies make effort to develop software as the market, composed by computer users and firms, is hungry for new applications.

Software is a commercial good, a non-physical one and what people purchase is the "right to use the software" or in other terms its "License"; what unfortunately happen nowadays is that software licenses are often stolen.

This bad behaviour goes by the name of "Software Piracy" and the counter action that more and more software companies are taking to this problem is to invest in "License Compliance".

This master's thesis introduces the reader into the world of Software Compliance describing first what a Software is and what a software business may comprehend and defining later what are the activities that a software company can adopt to prevent or address an illegal act of software piracy.

Additionally, a quantitative study on Software Piracy is conducted thanks to precious data gathered during a period of work in the software company "Parametric Technology Corporation (PTC)", this study analyses the Software Piracy behaviour in the Italian country considering companies that use the software developed by PTC for Computer Aided Design (CAD).

Statistical analysis is performed considering geographical location of the companies, their size and wealth and the economic activity they perform.

Findings are that location of the company in the Italian country seems to be a relevant discriminant for piracy while financial and economic variables are less critical in defining this behaviour. The analysis may be improved matching the dataset information with more data covering uncovered aspects in this master's thesis.

1. THE SOFTWARE INDUSTRY & BUSINESS MODEL

Working in the Software industries implies to face a business built around a non-physical asset, from a legal point of view it means that the company which produces software is the owner of the relative intellectual property and has to manage it.

1.1 THE SOFTWARE INDUSTRY

The word “software” began to be used during the 1950s, computers were already in the market and, until that years, all the programs used by companies to work were mostly built internally and not shared with other companies or possible customers. In 1955 the commercial exploitation of software began, and the company accredited to be the first to develop software independently is the “Computer Usage Company” located in New York City.

During 1960s the demand for software grew rapidly because of the need coming from governments, universities and business customers and with the advent of personal computers in 1970s the market grew even more and expanded not only in educational and business software but also in games, applications and utilities.

Further improvement in computer hardware in the following years was followed by a similar trend in software industry where the demand increased in a heterogeneous way with the need for platform different from computers (embedded software for electronic devices, telephones, machineries and so on).

This leads to the industry as we see nowadays that is generally composed by software producers, distributors and a crowded consumer base.

1.2 FROM DEVELOPMENT TO MARKET

The software industry includes several businesses, mainly they are **development, maintenance and publication of software.**

1.2.1 SOFTWARE DEVELOPMENT

It is a broad activity that includes all that is involved between the conception of the desired software and the final working one.

Following the Software Development Handbook (Geoff, Gill, Texas Instruments October 1981) the development requires some combination of the following steps (some changes in further steps may require changes in previous ones so it is a recursive process) [picture pg43 – the software development process]:

- Functional Specification
- System Design
- Software Design
- Programming
- Translation of the source
- Configuration and linking
- Debugging
- Integration Software and testing of hardware
- Evaluation of the final system

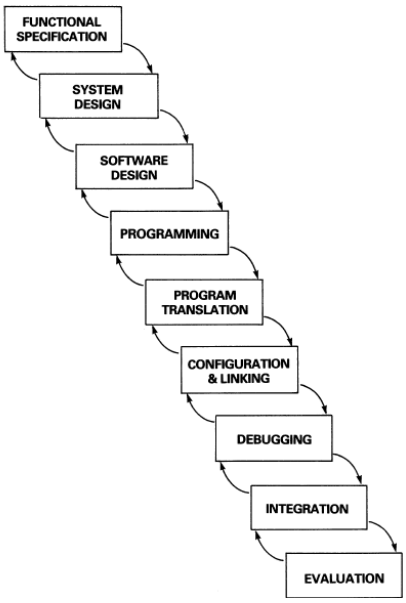


Figure 1: Recursive process for software development.

Functional Specification

The development of a software commonly requires a problem to be solved, in this case the producer will find a problem internally (by the idea of an employee for example) or looking outside through a market research. In other case the company is looking at ways to improve existing software or when it comes to disruptive ideas the new software doesn't solve a problem but tries to create a new need for customers.

Once the objective to reach is clear the companies requires to collect specification needed to its software to perform its objective in the best possible way

System Design

This task is common when the software developer is also in charge of hardware development or changes to an existing one. The purpose is to derive from the specification an implementation strategy, that is basically how to integrate hardware and software for example a special interface may be required.

Software Design

This activity is not yet programming but is the translation of function the software should perform into software algorithms and data structures.

Programming

Programming involves turning a software design into source program code, following the syntax rules of a particular programming language, during this activity software engineers actually code the software (implementation) and they test it gradually to avoid any bug, mistake and deliver a final working product.

Translation of the source

The software is coded in an high-level language, that is a language really close to the human way of speaking (usually a coding language use some English word and abbreviation to give instruction to the computer) , while a machine language is a binary code composed of 0 and 1, that is why software tools called “compilers” are used to translate the high level language into the machine binary code.

Configuration and linking of the software

The software code is mostly written as several smaller packages which are easier to manage, thus it is required to link all these pieces in one program, a task that requires first a configuration or selection of pieces to be welded together and finally the exploitation of a linker, a software that automatically creates the links required.

Debugging

Through several test and simulations, the software is analysed, and all the bug pointed out must be fixed, this is a preliminary activity that allows the creation of a final working software, but maintenance will be required even after sale as the software may evolve or some problems weren’t faced before the final publication.

Integration Software and testing of hardware

Following the simulation of the software in working condition the software producer should also perform an Emulation of the software, that is to emulate condition in which the software runs on different hardware for which it was conceived.

Evaluation

The software is finally tested and evaluated by experts external to the development to verify its functionalities and discover some issues it could still have.

After the software passes the final evaluation is ready to be Published and go to market.

1.2.2 SOFTWARE MAINTENANCE

Although maintenance is commonly perceived as an activity related to bug-fixing the software industry consider it as a "modification of a software product after delivery to correct faults, to improve performance or other attributes", in other ways just a little percentage of this activity is related to issue solving, most of the time software developers try to improve their product and add new features. (updates)

According to the ISO/IEC 14764 standard, 4 types of software maintenance are recognized:

- **Corrective maintenance**: Reactive modification of a software product performed after delivery to correct discovered problems.
- **Adaptive maintenance:** Modification of a software product performed after delivery to keep a software product usable in a changing environment.
- **Perfective maintenance:** Modification of a software product after delivery to improve performance.
- **Preventive maintenance:** Modification of a software product after delivery to detect and correct latent faults in the software product before they become effective faults.

1.2.3 SOFTWARE PUBLICATION

There exists an intermediary between the software developer and the distributor, it is the “Software Publisher”. In some companies, two or all three of these roles may be combined.

Software publishers often license software from the developers in exchange of a royalty payment, then they usually help them to reach larger or foreign markets and share some of the profit, usually the publisher is the one who will bear most of the cost of entering these markets.

The duties of the publisher can vary greatly depending on the agreement reached between the parties. Duties can include:

- Translating the product into the local language
- Building demand in the local market by advertising campaigns
- Produce and Design boxed software products
- Give technical support of the product locally

In some cases, Publishers find a market need and pay a software company to develop the final product.

When a software is released and published there are other activities performed by the companies, for example a lot of services may be offered to the customer from support and maintenance to extra features. What is missing now is the business model selected by the company to market their software.

1.3 BUSINESS MODELS

A business model describes how an organization creates and delivers value to customers. It characterizes products a company provides and the way it is compensated for them through a revenue model.

Along the software history there were changes not only from the technical point of view, but also on the business side, according to the Altexsoft website (Altexsoft is a company of software engineering and R&D), business model of a company involves a combination of the following characteristics:

- **Distribution approach**. How the company provides services or create products for customers. We distinguish three main distribution approaches that can be used: on-premise, cloud, and hybrid.
- **Source code licensing**. Considering the source code, the company creates, the software may be proprietary or open-source.
- **Revenue streams**. A software revenue stream defines the way company is paid for its products and services, it can include ad-revenue, sales, subscriptions, and their combinations.
- **Business model interaction**. It can be a one-to-many or many-to-many. The former is traditional value delivery, when a provider directly solves customer problems (Microsoft, Adobe, etc.). The latter implies that a company creates a platform where both end-providers and customers meet (Airbnb, Uber, CarToGo, etc.).
- **B2C or B2B market**. A company must choose the target audience, B2B means that company sells services to other businesses and B2C means selling a product or service directly to a consumer.

1.3.1 DISTRIBUTION APPROACH

On-premises software distribution approach

The software is installed and runs within a client's in-house infrastructure (a computer or a server). The traditional distribution approach has been used for many years by such companies as SAP, Oracle, and Microsoft.

This approach allows to customize the product in a way it satisfies client' needs and puts all the infrastructural expenses to the client side. On the other hand, it requires a lot of time for implementation and requires that the client has its own IT support staff and in-house server hardware (in B2B market). Problems may arise with upgrades (difficult to implement with customization) and the approach requires a hig up-front investment by the customer that may be considered too risky.

Cloud-based software distribution approach

The software runs at a hosting provider or in the cloud service; it is also called software as a service (SaaS), which is a cloud-based distribution approach in which a provider hosts its applications and makes them available to customers on-line, so that

businesses customers and individuals don't need to install any software on their machines. SaaS is typically delivered through a subscription model.

Cloud products and services have faster implementation for customers and are accessible remotely from anywhere at any time given an internet connection.

The SaaS approach doesn't require huge up-front investments from users. The customers just need to subscribe and log into their account to get full access to the app and its updates, furthermore, SaaS approach gives the opportunity to provide the same software version for all customers leaving to the owner just a single version to maintain, upgrade, debug, and provide support for. Finally, revenue comes from subscriptions on an ongoing basis.

Some issues can arise between cloud solutions and integration with existing on-premise enterprise applications and the software producer is also responsible for a reliable infrastructure (It requires a big investment since a company may need to store data from customers, make it accessible and protect it from hackers) Examples are Amazon Web Services, Dropbox, Netflix etc.

Hybrid software distribution approach

An approach that combines a SaaS solution with an on-premise software application, so that a cloud-driven technology complements an on-premise one. It can happen that a software company combines on-premise products with additional cloud services like libraries, add-on, etc. In this way on-premise software gets updated regularly and without asking the client to purchase a separate license for each new version of a product.

The Hybrid approach allows a client the flexibility to move information between on-premises data centres and cloud services and discriminate what kind of data should be moved, for example medical records (a sensitive data) can stay on-premise, while operations that don't involve sensitive data can be conducted in the cloud. This approach has some weakness, it requires more effort from developers since they provide both an on-premise and a web accessible software, it can be hard to manage upgrades, especially if some on-premise versions are customized.

1.3.2 SOURCE CODE LICENSING

Software companies can create proprietary or open source software.

Proprietary software

Generally, it is a copyrighted source code and doesn't let the user access, change, or reuse it, sometimes even if a program is free to use, it may have proprietary code that users and third-parties can't change. Most companies make their software products proprietary to protect it from copying, changing, or emulating.

In this way the company creates a business based on the sales of the proprietary software and choose to sell it on-premise or through a cloud-based service.

Some drawbacks are that the proprietary software may be unaffordable to some of potential clients leading to some revenue loss and the owner will be the only one responsible for finding and fixing code vulnerabilities. This means that proprietary software is more likely to be vulnerable to malware and attacks than open-source code where possible issues can be better detected as a community effort.

Open source software

A software producer may guarantee free access to its products, in this case the vendor may charge for customization, support, and maintenance or for a more complete version of the software.

An example is the Jet Brains company which creates tools for developers that are used all over the world in different well-known companies such as Wikipedia, Salesforce, and Pinterest. It also developed PyCharm, an integrated development environment (IDE) used in computer programming released for free and in a professional version under License.

The Open Source environment gives the possibility to the producer to show its software to a large audience with no initial cost, it allows clients to customize a product for their needs and as mentioned above, it has less chance of having vulnerabilities as they can be detected by a community.

It can be more difficult to enter the B2B market with an open source software as corporate clients will struggle to approve an open source code and it will be more difficult to make the business sustainable for the software company as open source products require finding additional revenue streams.

1.3.3 REVENUE STREAMS

Software companies have a hybrid mix of revenue streams to meet different market needs, some of them are:

Paid license

The concept of a license-based revenue stream entails a provider charging a customer money once for installing software. This is probably the most common and simple monetization approach widely used with on-premise products.

In a traditional on-premises distribution approach, a one-time upfront, perpetual license fee is used, which gives customers the right to use the software indefinitely. However, technical support and the right to product updates can be priced separately in some cases, especially for professional software.

Free, with in-app advertising

In-app advertising allows to make money by selling advertising space in the application software.

Free, with in-app purchases

The product provides physical or virtual goods sales. Everything can be sold from clothes and food to an in-game currency. It works especially in gaming environment, where players play for free and pay for in-game features (Fortnite, League of Legend, etc.)

Subscriptions

Based on a subscription license. It means that customers pay a per-user fee, either monthly or annually, which allows them to use the software during the subscription term.

The subscription payment includes not only software licenses, but also support services and new versions of the software as they are released. There are several possible options for the subscription model.

- **Standard subscription**. Users pays a recurrent fee.
- **Free trial**. Users can try the product for a specified time and then decide whether they should pay for a subscription.

- **Freemium**. It is a combination of free and paid (premium) versions. It's a free service with the option to get access to a paid, premium, software with additional features.

Usage-based license

A usage-based license is often employed in B2B products. It means that clients subscribe to "a pay-as-you-go" license based upon some measure of consumption, paying for only what they use. That metric could be tied to different usage aspects such as registrations, enrolments, purchases, course completions, certificate completions, franchise locations, and even logging onto the system.

Transaction fee

The transaction fee system is one in which the company charges a commission When there is a transaction between customers. The amount of the transaction fee can be both a percentage or a flat fee. This kind of revenue streams requires the presence of a two-sided marketplace that access the platform provided by the software company. This revenue stream requires an engaged audience using the marketplace or service because the goods or services become more valuable when more people use them. Some Examples are Airbnb, Uber, eBay.

Extra charge for services

This is a revenue stream that involves getting an extra fee for providing special services for corporate clients. Such companies have special requirements for software products that need to be customized only for them. Famous open source software companies Red Hat and Hortonworks leverage support and consulting services as their main monetization streams.

1.3.4 BUSINESS MODEL INTERACTION

There are two types: one-to-many business model interaction and the many-to-many business model interaction.

In the One-to-many business model interaction companies create products and services and sell them to customers, while in a Many-to-many business model interaction, one group of users (producers) create and earns value on the platform for the other group of users (consumers) that consumes.

In the one to many approach producers focus on the end consumer only, in the many-to-many value model they focus on both value producers and consumers. For example, YouTube requires building tools for producers (video hosting) and for consumers (video viewing).

This is the way the software industry works most of the time, from a new idea to the actual building of the code and finally to the definition of the business built around it.

Looking above, the starting point was that a software company runs a business founded on an Intellectual Property (IP) and manages it, that is why companies exploits several tools to do so, from **license agreements**, to entire business departments like the **software compliance** one.

The following chapter will give a more precise view of the CAD software industry as it is the one analysed to conduct the empirical research of this thesis work, while cap xxx will introduce the software compliance business and the antipiracy activity.

2. CAD SOFTWARE INDUSTRY

CAD stands for Computer Aided Design and it is a software used to design, develop and optimize products, especially tools and equipment for manufacturing processes and for construction. CAD is used to develop preliminary design and layouts, making calculations, creating models and drawings, as well as interfacing with analysis, for example manufacturing analysis (several tests can be performed), sales and marketing ones and many others.

CAD produces data and information that can be universally interpreted by trained personnel, it allows to draw either two-dimensional or three-dimensional diagrams and view an object from any angle, even from the inside. The main advantages of a CAD drawing are that the editing is faster and cheaper with respect to manual method, it also reduces design time by allowing precise simulation rather than build and test physical prototypes.

The global CAD software market was valued 8.325 M$ in 2017 according to Prescient & Strategic (P&S) a company that offers analysis of different market to support businesses. This kind of software is mostly used in industries like manufacturing, aerospace/defence, automotive, healthcare, arts, media, entertainment and others (consumer products, ship building, garments and interior design). It is interesting to discover that a great size of the market is composed of mid-level designers who works for not critical project, mostly because the CAD market is shifting to cloud-based application and younger mid-level workers trust more the new cloud technology and its lower price, in fact they account for 45% revenue of 2017 CAD market.

Still according to P&S analysis the latest trend in the worldwide CAD industry are:

- A shift from license-based business model to a subscription-based one (cloud);
- A transition from 2D CAD to 3D CAD;
- Acquisition of smaller companies/startups by leading company to enhance their offering and increase their customer base;
- Introduction of mobile application designed to make the CAD usage more flexible and not only linked to Personal Computers.

The market now is composed by several companies and among them key players are PTC Inc., Autodesk Inc., Dassault Systemes, Bentley Systems Inc., Hexagon AB and GStarsoft.

According to a research performed by 3D Hubs (a large network of manufacturing companies), among 750 of its customers the most common CAD software used by engineers are Solidworks (by Dassaults Systems) and AutoCAD, while Creo (developed by PTC Inc.) represent the 4% of the user base.

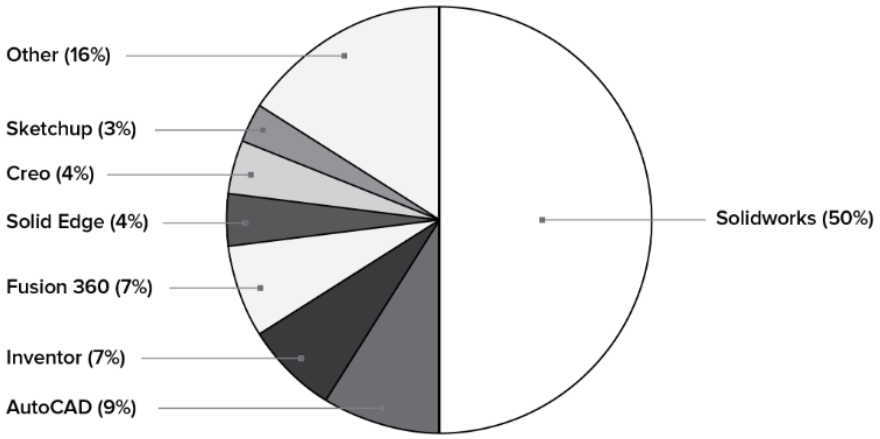


Figure 2: Main CAD software used by 750 of 3D Hubs customers.

The Market for Creo software in Italy is the one that will be considered for the research of this master's thesis. Detailed information about the Italian industries considered can be found in chapter xx as a deeper explanation is given to understand what the tendency to piracy in Italy is.

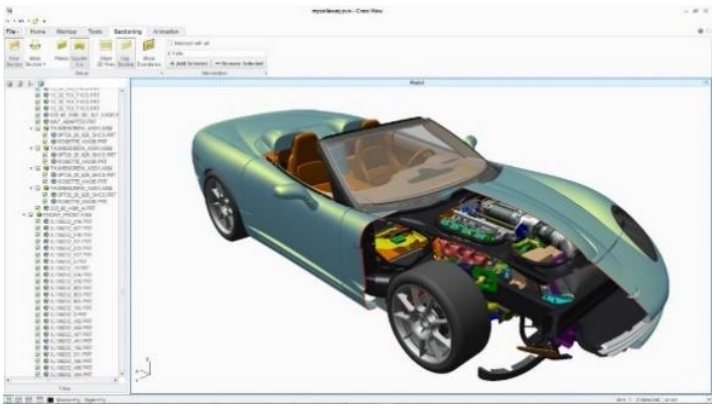


Figure 3: Working Environment of Creo CAD Software developed by PTC Inc.

3. SOFTWARE COMPLIANCE

3.1 INTELLECTUAL PROPERTY AND LICENSE AGREEMENT

The term Intellectual property refers to creations of the mind such as inventions, literary and artistic works, designs, and symbols, names and images used in commerce and other non-tangible assets like a software.

IP is protected in law by patents, copyright and trademarks, which enable people to earn recognition or financial benefit from what they invent or create.

In the software case the instrument used can be the copyright and the patent; to understand their application is important to decompose the software into:

- **CODE**: The code is a series of instructions that allows the machine to execute a function.
- **ALGORITHM**: Is a list of steps that solves a problem. (each step is elementary in the sense that it cannot be decomposed in other sub-steps)

So, starting from an algorithm we can use different code or programming languages to solve the same problem. From a legal point of view the copyright protects the code (creation of the mind), while for the algorithm there is no protection.

There are cases in the European Union in which some software is patented, generally it is possible when they offer a solution to a specific technical problem while a patent will not be granted to a software that solves more general problems (like calculations or management of systems), the drawback is that patenting a software can be too costly and risky as competitors can use reverse engineering to find solution to the same problem avoiding the patent protection.

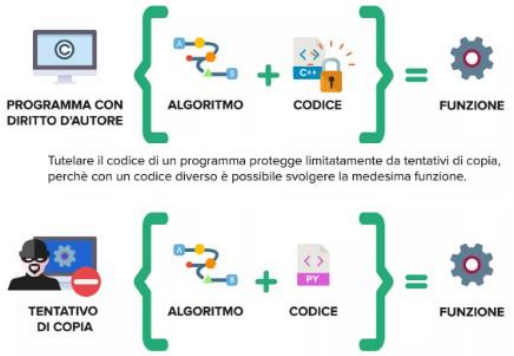


Figure 4: A software protected by copyright alone; its algorithm can be emulated using a different coding language.

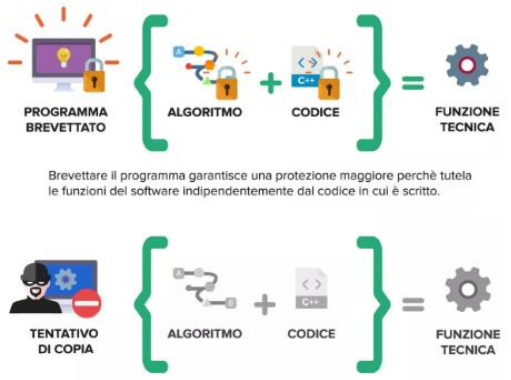


Figure 5: A patent protect the algorithm too if it solves a specific technical problem.

As the current state of the art suggests, **copyright** is the only recognized legal protection for software and is sufficiently efficient to protect the IP.

A copyrighted software protects the producer from:

- Reproduction/copies of the software.
- Creation of derivative works based on the original code (to alter or build upon the existing software).
- Distribution of software copies.
- Publication of the software.

Law allows the Copyright owner to transfer each of these rights to others through **licensing.** By a license Agreement the software producer (licensor) allows a customer (licensee) to use the software after purchase under some constraints listed in the **End User License Agreement (EULA)** that the customer must read and accept to lawfully get the authorization to use the copyrighted software.

Knowing what the rights of a software producer are is useful to understand what the software compliance is and how it works.

Software Compliance is a branch of a broader discipline called Software Asset Management (SAM) which is a series of tools and activities that companies enact to: control and optimize the use of software across the organization, control costs as well as business and legal risks, optimize software licensing investments, and align IT investments with business needs.

By its definition Compliance means a state of being in accordance with established guidelines or specifications. From a customer point of view, it means that users of applications must follow what is stated on the license agreement to be compliant.

The guidelines stated on the EULA may cover different areas, generally it's important that a customer knows:

- What has been purchased
- What is installed and where it is installed
- When to phase out what is installed
- What the software usage policies are

Being compliant to the software specification is not only the correct way to avoid any unlawful usage of the IP and the consequent legal issues that may arise, it is a great way to monitor and optimize the software inventory of a company giving space to some benefits like: reduction in Software need and support cost, ensuring a secure environment (software not licensed may hide malware and cause data loss and stealing from hackers) and reduce cost to cover legal enforcements.

Giving a look at the On-premise software PTC license agreement is possible to understand what main restrictions the company imposes; these restrictions are common all over the software industry, so knowing them is a good way to have a general view of the main clauses used to protect the software IP.

The agreement consists of 4 major elements:

1) Introduction: explains how to correctly enter the contract (…individual accepts this agreement… either clicks the "I ACCEPT" button or installs, accesses, or uses any software or documentation from PTC…).
2) Copyright clauses: A list of clauses that the customer must follows to lawfully use the software. They comprehend the main Copyright clauses (as written above like prohibition to copy, public, modify the product etc.), and other restrictions linked to:
    - Installation Location: the software may be installed only in the country of purchase while for global usage the company may provide different agreement with a higher cost. This clause may also imply that the company code (Italian Partita IVA) for which the license is provided should not change, if there are any organization policy change in the company (change of the name, type of organization etc.) the company must notify the software house to avoid legal issues.

- User Location: this mainly refers to concurrent usage by more users, the software may be used only by one or a few concurrent users.
- Utilization level: a clause that may depend on the type of customer and define a limit of usage of the license linked to some company indicator (for example a license may entitle the customer to use the software until its Inventory in the balance sheet reach some level, for example $ 100k, if the level is higher the customer may purchase other licenses or reduce its inventory level to avoid any breach of the contract clause).
- Other clauses: they may be introduced considering the customer and its need.

3) Term: The license agreement always states a term of use that the customer must respect, the term usually refers to support and other services that the company performs until the license agreement is valid. After the term the customer has to purchase a new license to use the software and get access to the company support. The term can also refer to payment, for example the customer must respect a time limit to complete the purchase otherwise he will be in breach of the contract.
4) Company Liabilities: as the points above states what are the company rights, there is an ending part of the document that describes the company liabilities (customers right), for example the warranty the producer must provide, the quality standard and all the consequences that producer may face if he doesn't respect these clauses.

3.2 COMPLIANCE ACTIVITIES

The duty of the producer is to monitor its product usage and be efficient in detect and solve any incorrect/unlawful exploitation of the software, so considering the license agreement the company deploys its compliance activities.

Three major activities performed are:

- **ADVISORY**: A service offered to customer to clarify the license agreement, to teach how to correctly use the software avoiding breaching any clause and maximize the license usage to improve their business.

- **ASSESSMENT**: Is a customer-oriented process that is used to monitor the usage of licenses by companies and to understand if they are compliant with the license agreement or some changes are need to the contract or to the customer behaviour to make them compliant.
- **ANTIPIRACY**: Is a preventive action that the producer performs to avoid the illegal use of unlicensed copies of its software.

Any of these tasks may be accomplished in several ways from different companies; the following description is mostly based on the PTC Compliance department and its way of manging the software compliance, additionally more importance will be given to the Antipiracy activity as it is the fundamental activity that allowed to perform the empirical research of this thesis work.

3.2.1 ADVISORY

A service designed for educate users on the license agreement, the aim is to solve any doubt a customer may have on the way he should use the software correctly.

The advisory activity can be both a preventive or a corrective action, it depends on the customer ability to understand that he may have misunderstood the license terms and actively communicate with the software company to solve its doubts and act correctly. If the customer is not managing its installed base (installed licensed software) in a meticulous way, he may be in the breach of the license agreement and the advisory activity will find a corrective action to solve the breach.

The main activities performed by the name of advisory services are:

- Evaluation of user contractual license entitlements.
- Analysis and evaluation of the software usage with respect to the license agreement.
- Guidance for an optimal License allocation.
- Evaluation of further License deployment.

While the first two tasks may be useful to find unlawful usage of the software and enact corrective actions, the latest two are designed to improve the user efficiency optimizing its way to use the license (a customer may be over-licensing the software when he can actually save money, or he may need more licensed software to boost its own business).

Corrective actions are needed when the user is breaching the license agreement, it may happen when:

- There are Cross Border/Territory violation: the license is used in a different country from the one entitled in the license agreement.
- A major change has occurred in the Customer's organization: An existing license is still used when the company changes its legal name, or it went bankruptcy.
- Improper end usage by a third party: The license is used not to support the customer business but a third party one (It may be a supplier or a consultancy agent etc.) or the same license is used by a third party which is not entitled to do so.
- Under licensing/Over usage: There exist a gap between the actual usage by the user and the limit imposed by the license agreement.
- Change of Hardware: A customer changes hardware and use the same license on both new and old machines.
- License transfer without notification: The license is transferred to another company without the permission form the producer.
- Home Use: The license owned by the company is used to an employee house, this is an issue as the license may not be used off site.

These are only a few occasions in which the producer of the software has to actively inform the user on the unlawful usage of the license and correct its behaviour.

3.2.2 ASSESMENT

According to Almunawar, Mohammad, Nabil, Susanto, Heru, Assessment identifies a problem and describes how much of a problem it is. Assessment sets the results of an analysis on the install base of the organization and assigns an urgency level to issues may arise from this activity. In a more general way, an assessment uses a structured approach, is repeatable, and describes the level of a problem.

As part of an Assessment, Audit is the activity performed to evaluate the level of Compliance a company possess. It compares a given situation with a standard, in this case the license agreement. The results of an audit explain how much reality deviates from an expected or required situation.

An Audit allows the software producer to identify the unlawful usage of the software by the client, the difference with the advisory service is that the producer performs an audit when some problems are already founded in the customer network, so the producer decides to analyse all the install base of the customer and points out all the problem. The output is the gap between what is stated in the license agreement and the current use, so the client bears the responsibility to adapt its behaviour to the agreement and pay back the producer for the wrong usage in order to avoid a law suit.

Is important for a customer to manage the software possessed and its usage to avoid an audit as it may create some trouble for the company business:

- An Audit requires to stop the ongoing business to perform the analysis on the install base (the customer may create delay in its own projects and bear more costs because of a sudden stop of all activities)
- Being “audited” is not a good advertise for a company as the audit activities always come to solve a problematic situation.
- The audited company may ask for a second counter-audit by a third party and this will be another cost for the company as it is a paid service and it will cause a longer break of the company business.

3.2.3 ANTIPIRACY

A user may be using the software produced without any authorization to do so, it means that the user hasn’t signed (accepted) any license agreement and it is using the software unlawfully.

In this case the software company needs to protect its rights and exploit a system that allows to identify the use of unlicensed software also known as cracked software.

A crack is basically another program that is installed on the same computer that allows to use the original software without purchasing it by skipping the authorization phase, in other words let’s say that during a software installation there is a phase “A” that asks for an authorization (license key code for example) that is needed to go to a phase “B”, completion of the installation for instance; the crack will allow the user to jump to phase “B” without demonstrating to have the rights to use the software by skipping phase “A”.

When the user installs the software thanks to a crack, he has access to the software and its function and generally avoids any kind of support or help by the company (he hides himself from the producer), this is an act of stealing and breaching the copyright of the IP owner, an act punishable by law.

Italian law states that this kind of breach is punishable:

- By criminal law: ex art. 171 bis substituted by L. 18 august 2000, n.248, imprisonment from 6 months to 3 years and fine from € 2582 to € 15493. According to the art. 171 bis of law n.633 22/04/1941 and modified by L.18 august 2000 n.248, is punishable by law not only the one who reproduces the illegal software copy (crack) but also who uses the unlicensed software within his organization.
- By administrative law: ex art. 174 bis – fine equal to the double of the software market price for each copy of illegal software used.
- By civil law: ex art. 156-167:
    - Assessment of the injured and inhibitory right
    - Destruction of illegal products
    - Seizure or inhibitory suitable to prevent the continuation of the offense
    - Publication of the sentence on one or more newspapers at the expense of the losing party.

Furthermore an user of unlicensed software is not only risking legal enforcement as stated above; the use of a cracked product can be really risky from a cyber-security point of view, in fact the cracked software is a programme that comes from hackers, people who are able not only to reproduce the original software and allow its usage without any license, they mostly are expert computer users (software developers) and capable of access other computers and steal data from them, so they are really dangerous especially for a company security.

A hacker may hide a malware in its illegal copy of the software and use it to steal data from the one using the crack, as the data importance raised dramatically during last years it is easily clear how much can be dangerous the activity of such people.

Through monitoring technologies a software company is able to detect the unlawful usage of its products and act in order to stop the illegal use of software and recover damage before going to court; one way is to contact the user who possess the

unlicensed copy and consider the purchase of a licensed copy as a recovery damage, in this way the software producer is able not only to stop the illegal use but also to build a sustainable business relationship with an user who is actually interested in the usage of the original software.

3.3 COMPLIANCE WORLDWIDE

Software Compliance is an activity which increased in popularity in the latest decades as more and more company have understood the importance of protecting their intellectual property and how it can be not only a preservative activity but also a great business. Monitoring customers behaviour and their level of compliance make it possible to exploit the gap between their legal entitlements (license agreements) and their current conduct that in many cases is not correctly following what is stated in the EULA. This means that the company can easily recover the intellectual property damage by boosting software sales and make customers compliant without reaching the court and start a law suit, as both the customer and the producer have more interest to avoid legal expenses and bad advertising.

To picture the current worldwide situation in the software industry it is helpful to consider the BSA global software survey (June 2018) conducted across 110 national region and economies to quantify the volume and value of unlicensed software installed on personal computers.

BSA The Software Alliance is a company that, among its activities, conducts compliance and enforcement programs to protect intellectual property rights for software companies; members of such alliance are several well-known companies such as Adobe, PTC, Salesforce, Autodesk, IBM and many others.

As the company description tells "… team works … to improve market conditions for software companies that commercialize intellectual property, which they create by investing heavily in research and product development. BSA's compliance and enforcement team, meanwhile, ensures our member companies' innovations are fully licensed and end users understand the benefits of legitimate software".

According to the 2018 software survey by BSA:

Use of unlicensed software, while down slightly, is still widespread: The survey registered a reduction in the worldwide unlicensed software; still the 37 percent of software installed on personal computers is not licensed. Drawbacks are that this bad habit creates a delay in the local economies and in the thriving technology use, impedes growth in a company business and induces higher security risks.

CIOs are finding unlicensed software is increasingly risky and expensive:

Installing a non-licensed copy of a software has a 1/3 chance of containing malware; a malware, according to the definition by Malwarebytes (a company that develops application to protect software users) is "… any malicious program or code that is harmful to systems …. malware seeks to invade, damage, or disable computers, computer systems, networks, tablets, and mobile devices, often by taking partial control over a device's operations … is all about making money illicitly … it can steal, encrypt, or delete data, alter or hijack core computer functions, and spy on a computer activity without your knowledge or permission".

Each malware attack can cost a company $2.4 million on average as it requires to spend money on:

- Recovery of the damage (possibly hiring an expert company to solve the issue).
- Stop of business operation, a malware recovery can take up to 50 day to be completed.
- Malware may lead to loss of business data and can also affect the company's brand and reputation.

Also, the cost for dealing with malware associated with unlicensed software is growing too. BSA found that It can cost a company more than $10000 per infected computer, leading to a worldwide expense of almost $359 billion a year.

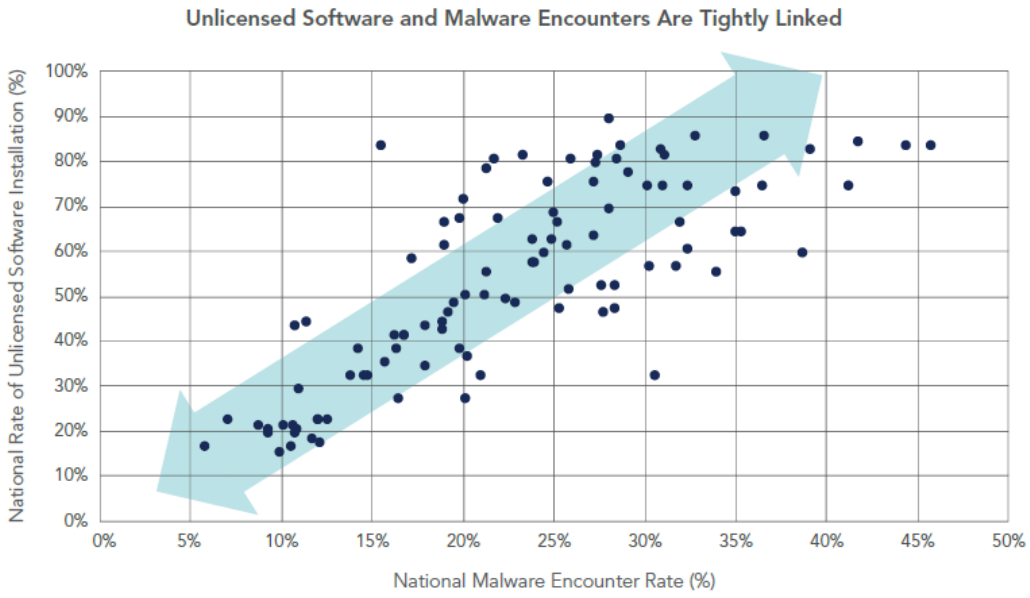


Figure 6:Source BSA Global Software Survey June 2018. National rate of unlicensed software against national malware encounter rate.

Improving software compliance is now an economic enabler and security imperative.

Because of the increasing cost from malware, companies' leaders are turning to fully licensed software that is the best possible defence against malware incursions, data breaches, and other security risks.

Another key finding of the BSA survey is that a correct software asset management is boosting company business by an average of 11 percent.

Organizations can take meaningful steps today to improve software management and achieve important gains.

To improve the efficiency of the software asset management organizations can implement proven software and best practices to get more out of their technology. SAM software helps the company to ensure that software running on their network is legitimate and fully licensed thus decreasing cyber-risks and improve productivity, reduce downtime, centralize license management, and finally reduce costs up to 30 percent savings in annual software costs.

Another key finding by the BSA 2018 survey is that organizations taking proactive steps to protect their intellectual property can improve, from a 20 percent increase in software compliance, their profits by 11 percent, in other words a boost of more than half a million dollars for a medium company in the survey.

3.3.1 WORLDIWIDE COMPLIANCE IN NUMBERS

Further data about compliance performance is gathered thanks to Revulytics, a company that provides embedded usage analytics to the world's leading software vendors. The company was founded in 2006 to help software vendors address the challenges of unlicensed software, both software piracy and overuse by existing customers.

The latest data available for the year 2018 tells what the most uncompliant countries are when it comes to license compliance, looking at the graph Italy is in a bad situation as it is 7th.

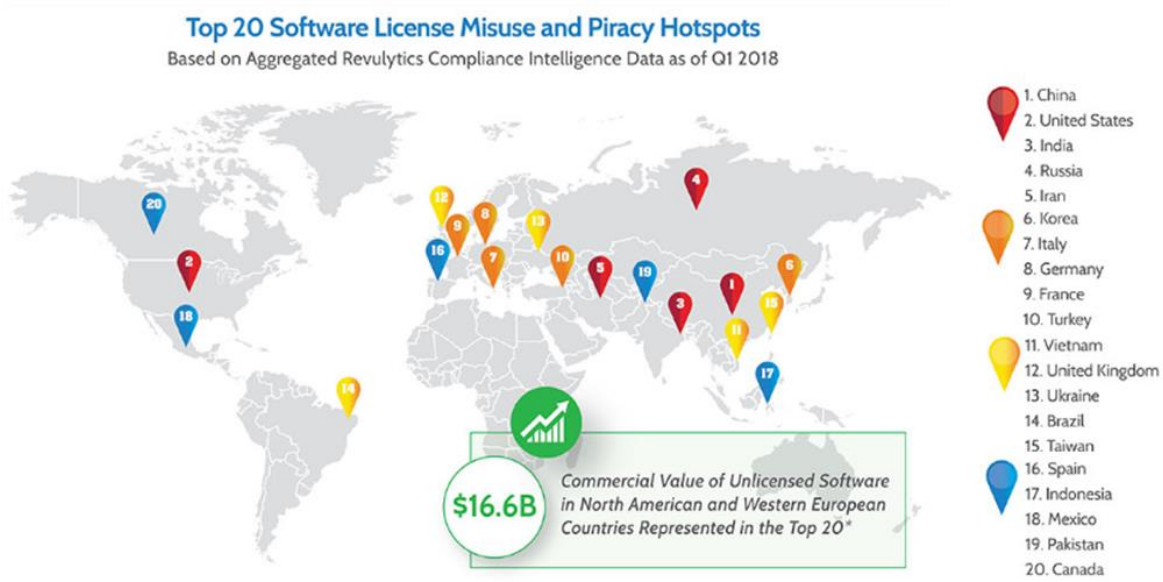


Figure 7: Source Revulytics. Top 20 countries for Software license misuse.

Another interesting insight from Revulytics is a research based on The V.i. Labs Data Service, a federated database available to their customers. It contains details on more than one million infringing organizations using unlicensed versions of Revulytics customers' software. The research revealed that unlicensed software is used in:

- 100 % of the top 50 engineering schools
- 95 % of the top 100 colleges/universities
- 78 % of the top 100 software companies
- 51 % of Fortune 100 companies

To understand the piracy trend and the value of unlicensed software (worldwide) in the last seven years (from 2011 to 2017) and in Italy the BSA Report gives a comprehensive overview.

The rates of unlicensed software worldwide decreased from a 42% in 2011 to a 37% in 2017, considering the commercial value of the software it is possible to tell that almost 20k M$ of software that was previously unlicensed is now correctly purchased, giving to the software industry and regional technological development a great boost.

The same trend is followed by the European Union (reduction of 5% in unlicensed software) and Italy too where it is possible to see almost 600 M$ saving for the software industry.

Table 1: Source BSA Global Software Survey June 2018. Rates and Values ($M) of unlicensed software installed worldwide, in Europe and in Italy.

RATES OF UNLICENSED SOFTWARE INSTALLATION				
	2017	2015	2013	2011
TOTAL WORLDWIDE	37%	39%	43%	42%
European Union	28%	29%	31%	33%
Italy	43%	45%	47%	48%

COMMERCIAL VALUE OF UNLICENSED SOFTWARE ($M)				
	2017	2015	2013	2011
TOTAL WORLDWIDE	$46,302	$52,242	$62,709	$63,456
European Union	$9,982	$11,060	$13,486	$14,433
Italy	$1,278	$1,341	$1,747	$1,945

Software piracy is an illegal activity widespread around the world as data from expert companies confirms so, before taking a step into this thesis research, it is fundamental to consider previous studies on the same topic to get a more precise idea on what are the main variables that may be linked to such behaviour, that is why next chapter is dedicated to the Economical researches which cover the software antipiracy subject.

4. ECONOMIC STUDIES COVERING SOFTWARE PIRACY

This master's thesis is not the first study performed on the software piracy, it actually finds its place among several different researches that cover the same topic and face it from different perspective looking for the causes and effect of such event.

Key findings from previous research are really effective to point out evidence of the existence of some kind of linkage among software piracy and real world macro variables in fact, while as a common taught one may expect software piracy to be an unpredictable event just linked to the avoidance of the payment of software price it is surprising how studies revealed interesting explanation to the theft of the software IP.

4.1 PIRACY - GDP

One macro variable extensively studied against the software piracy level is the Gross Domestic Product; according to K. Bagchi, P. Kirs, R. Cerveny (2006) and R. K. Goel Michael, A. Nelson (2009) GDP is inversely related to software piracy, it means that richer countries have lower need for piracy, in these countries individuals are better able to afford legal software and they also face an higher opportunity cost of breaking the law.

From a "cause" perspective, GDP has a positive effect on piracy (the higher GDP the better), while considering the effect of this variable on piracy and vice versa is interesting to consider A. R. Andrés, R. K. Goel (2011) study which explain the existence of a nonlinear relationship between piracy and GDP.

Depending on the level of software piracy in a country there is a two-sided effect:

- A negative effect confirmed also by other studies (Bezmen & Depken, 2004)
- A positive relation between piracy and economic growth as a greater piracy lowers economic growth, albeit at a diminishing rate.

This means that there is a convex relationship and a plausible explanation for this convex relation is that modest levels of piracy lower growth by reducing investment, but as piracy rates increase beyond some threshold level, the rate of decline in growth goes down due to positive network externalities from piracy.

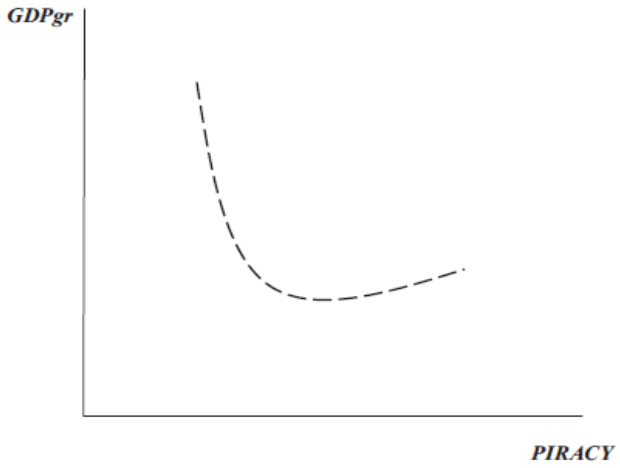


Figure 8: Source A.R.- Andrés. R.K. Goel / Journal of Policy Modelling 34 (2012). Convex curve describing the relation between GDP and rate of Piracy.

4.2 PIRACY – IT DEVELOPMENT

IT development of a country is also another factor that has two different effect on software piracy.

According to K. Bagchi (2006) the IT development present a similar effect of that of GDP, in other words increased availability of legal software over the Internet, especially given the decrease in software prices, lessens the need for piracy while considering K. Goel (2009) the diffusion of Internet and computer technologies and accessibility to information networks in a country enable both potential pirates and protectors of intellectual properties to do their jobs more effectively.

4.3 PIRACY – REGULATORY FACTORS

Despite the richness of a country and the IT development other variable were studied, in particular the regulatory factors (K. Bagchi 2006), by imposing higher tariffs on high tech goods piracy increases so a less regulated market allows for lower level of piracy. Looking at the legislation a not diffused copyright policy which is not enforced encourage higher level of piracy as a pirate doesn't consider the legal risk incurred while breaching the law, the thesis is also in accordance with A. G. Peace, D. F. Galletta, J. Y.L. Thong (2014) a study that points out how, within software-using organizations,

increasing employee awareness of the potential severity and certainty of punishment can lead to decreases in intentions to pirate software.

4.4 PIRACY – LEVEL OF LITERACY

Other insights are given by the level of literacy in a country (Goel & Nelson 2009) as greater literacy enhances the piracy of software. A possible explanation for this result is that a more educated population might be better prepared to produce, and demand pirated software and may be better able to circumvent the government enforcement mechanisms. Alternately, a low educated population might not use computers widely and thus have relatively little use for pirated software.

4.5 DATA COLLECTION

Numerous studies are performed on the software piracy subject and is clear how they show evidence of relation with several variables, now the focus goes on how piracy is defined in these studies.

Gathering data on software piracy is troublesome as it is difficult to get information on a software crack even for a software company.

The cited studies and others, collect information on piracy trough different channel:

- BSA data on annual piracy collected by the partners companies (Goel & Nelson 2009 - Goel 2011)
- Piracy rate is the difference between software programs installed and software applications legally license (A. Rodríguez 2010)
- “By assuming that for each new personal computer sold there will be a set of accompanying software sales, the difference between expected demand and supply (in the form of sales) is attributed to software piracy” (Trevor T. Moores 2003)
- Piracy level extracted thanks to a survey performed on 201 respondents among MBA students, a sample of working adults taking evening classes in the part-time MBA program of a mid-Atlantic U.S. university (A. Graham, D. Galletta, Y.L. Thong 2014).

The current research will be based on a different set of data, the sample collected thanks to the collaboration of a Software company is composed by companies in which real piracy happened and is analysed against another sample of the same size with similar companies (customers of the same software company) which never pirated the software.

Starting from this dataset a deeper analysis could be performed to search the causes of piracy not only in macro variables and country level factors but also on the economical wealth and other available data of companies using that software.

5. EMPIRICAL ANALYSIS OF A CAD SOFTWARE PIRACY IN ITALY

This study is based on data gathered in the software compliance office of Parametric Technology Corporation (PTC) and refers to piracy in the Italian market of the CAD software developed by the company.

5.1 PARAMETRIC TECHNOLOGY CORPORATION HISTORY

It is a Software Company founded by Samuel Geisberg in the 1985 in Boston, Massachusetts. As primary Business the company developed parametric, computer-aided design (CAD) modelling software in 1988 called Pro/ENGINEER.

In 1998 Steve Walske was named CEO and the company shipped Pro/ENGINEER and was considered first to market with parametric modelling design software. This positioned PTC as a leader in the CAD industry until the mid-1990s when a new generation of low-cost competitors entered the market.

The company developed also other software with different functions:

- 1986 was the year of Mathcad, a software primarily intended for the verification, validation, documentation and re-use of engineering calculations. It was the first to introduce live editing of typeset mathematical notation, combined with its automatic computations.
- In 1998 an Internet-based software for product lifecycle management (PLM) called Windchill positioned the company as first to market with internet-based solutions for Product Lifecycle Management.
- PTC Integrity Lifecycle Manage is an application lifecycle management (ALM) platform developed by MKS Inc. and was first released in 2001. The software is client/server, with both desktop and web client interfaces. It provides software development organizations with a collaborative environment in which they can manage the end-to-end processes of development, from requirements management, engineering change management, revision control, and build management to test management and software deployment as well as associated reports & metrics.

- During 2002 the company released Pro/ENGINEER Wildfire considered the first CAD system to support web-based services.
- In 2010 James E. Heppelmann was announced as CEO (It maintains currently that role). Company renamed Pro/ENGINEER to PTC Creo and promised a design software scalable, open, and easy-to-use.
- Servigistics came out in August 2012 and had a broad and deep set of capabilities for service parts optimization in the industry.
- ThingWorx and Axeda during 2014. The former is a platform for the rapid development of applications designed for smart, connected sensors, devices, and products – or the Internet of Things, the latter is a cloud-based service and software for managing connected products and machines and implementing Machine-to-Machine (M2M) and Internet of Things (IoT) applications.
- Vuforia and Kepware in 2015. Vuforia is an Augmented Reality Software Development Kit (SDK) for mobile devices that enables the creation of Augmented Reality applications. Kepware is a connectivity platform for a diverse set of industrial automation endpoints. The software enables users to connect, manage, monitor, and control heterogeneous devices in the Industrial Internet of Things.

Nowadays PTC has a global presence worldwide in 30 different countries with 6000 employees; in 2017 registered 1.16 B$ Revenue and it is a leader company in the 6 segments in which operates:

- CAD (Computer Aided Design)
- PLM (Product Lifecycle management)
- Industry 4.0
- SLM (Service Lifecycle Management)
- IoT (Internet of Things)
- AR (Augmented Reality)

Figure 9: Current logo of Paramentric Technology Corporation.

5.2 DATA AND METHODOLOGY

Data covering events of piracy was collected thanks to the compliance office of PTC Company during a six months internship in Software Anti-piracy in year 2018. This data refers to piracy events happened in Italy starting from 2012. It is important to understand that this data finds its focus during years 2016-2017 as these years reflect a higher intensity in the compliance office work.

The original dataset is composed of 6040 compliers and 926 non compliers. All non-compliers have been searched in the AIDA database, provided by Bureau van Dijk. Bureau van Dijk is a company that provides private company information offering a range of company databases, AIDA stands for "Analisi Informatizzata delle Aziende Italiane" and is a database which contains information about Italian companies. For the sake of this study, information about companies' balance sheet and income statement were collected.

The use of this database was allowed by the Politecnico di Torino university which provided the key to access AIDA.

Companies were matched by using company name and other identification variables (address, name of the owner) in order to retrieve financial and other firm level characteristics. A total of 395 non compliers have been identified and data retrieved. Out of the 6040 compliers, a random sample of 1200 companies have been selected. These firms have been searched in AIDA leading to a final sample of compliers of 545 units.

In this way the analysis can be structured in a comparison between "bad" companies and "good" companies which do not crack the software. The comparison works out as the two samples are composed by companies that exploit the same software and operates in a similar industry/market or the exact same one.

Finally, the two samples were imported in Stata, a statistical software for data analysis, and merged in one dataset containing all the companies (pirates and customers) and their economic data.

The following statistics and tests performed on the dataset were executed on Stata IC 14 version, which license was provided by the Politecnico di Torino university.

The primary statistics exploited are independence tests for categorical variables (Pearson Chi2) and regression analysis. As in many cases the dependent variable is a binary one, the logistic regression was largely adopted. The significance level considered is the standard of the Stata software, with a 5% of significance level (α), common values being 1% or 5% generally.

5.3 RESULTS

5.3.1 GEOGRAPHICAL POSITION

The first variable studied on our dataset is the geographical position in the country of the firm, we searched the existence of dependence between piracy, a binary variable 0 if the company doesn't crack the software and 1 if the company does crack the software, and the region of the company (20 variables for the different Italian regions).

Figure 10: Map of the Italian country divided into its 20 regions.

We tested the existence of independence of the two variables with a chi2 test where the null hypothesis to test is "Are the two variables independent".

Table 2:Software pirates and customers for each Italian region.

REGIONE	PIRATA		TOTAL
	NO	YES	
Abruzzo	8	5	13
	61.54	38.46	100.00
Basilicata	1	4	5
	20.00	80.00	100.00
Calabria	1	12	13
	7.69	92.31	100.00
Campania	13	18	31
	41.94	58.06	100.00
Emilia Romagna	81	76	157
	51.59	48.41	100.00
Friuli-Venezia Giulia	11	11	22
	50.00	50.00	100.00
Lazio	19	20	39
	48.72	51.28	100.00
Liguria	9	11	20
	45.00	55.00	100.00
Lombardia	184	91	275
	66.91	33.09	100.00
Marche	12	21	33
	36.36	63.64	100.00
Molise	1	0	1
	100.00	0.00	100.00
Piemonte	65	36	101
	64.36	35.64	100.00
Puglia	8	18	26
	30.77	69.23	100.00
Sardegna	1	0	1
	100.00	0.00	100.00
Sicilia	3	2	5
	60.00	40.00	100.00
Toscana	29	25	54
	53.70	46.30	100.00
Trentino-Alto Adige	9	7	16
	56.25	43.75	100.00
Umbria	5	6	11
	45.45	54.55	100.00
Valle D'Aosta	2	0	2
	100.00	0.00	100.00
Veneto	83	32	115
	72.17	27.83	100.00
Total	**545**	**395**	**940**
	57.98	**42.02**	**100.00**

Table 3: Chi2 Test - dependence of Piracy and Italian regions.

PEARSON CHI2	d.o.f.	p-value
64,224	19	0.000

As the p-value is lower than the 5% significance level, we cannot accept the null hypothesis. Therefore, we can say that, statistically, there exists a certain dependence between the two variables, in other words a software is more likely to be cracked in some Italian regions with respect to others.

As a robustness check, we excluded from the regional analysis the regions with 5 or less observation as the good practice is to not consider variables with less than 5 observations for the chi2 test which could distort the analysis. The regions excluded are Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta.

The new result still confirms the existence of some relation between piracy and regions:

Table 4: Software pirates and customers for each Italian region excluding Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta.

REGIONE	PIRATA		TOTAL
	NO	YES	
Abruzzo	8	5	13
	61.54	38.46	100.00
Calabria	1	12	13
	7.69	92.31	100.00
Campania	13	18	31
	41.94	58.06	100.00
Emilia Romagna	81	76	157
	51.59	48.41	100.00
Friuli-Venezia Giulia	11	11	22
	50.00	50.00	100.00
Lazio	19	20	39
	48.72	51.28	100.00
Liguria	9	11	20
	45.00	55.00	100.00
Lombardia	184	91	275
	66.91	33.09	100.00
Marche	12	21	33
	36.36	63.64	100.00
Piemonte	65	36	101
	64.36	35.64	100.00
Puglia	8	18	26

	30.77	69.23	100.00
Toscana	29	25	54
	53.70	46.30	100.00
Trentino-Alto Adige	9	7	16
	56.25	43.75	100.00
Umbria	5	6	11
	45.45	54.55	100.00
Veneto	83	32	115
	72.17	27.83	100.00
Total	537	389	926
	57.99	42.01	100.00

Table 5: Chi2 Test - dependence of Piracy and Italian regions excluding Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta.

PEARSON CHI2	d.o.f.	p-value
58,361	14	0.000

As we can see the p-value is still lower than 5% and looking at the total value for pirate (1) and customers (0) the value are really close as we see a shift of just 0.01 % in the results with respect to the analysis with all the regions.

Then we created a "Macro-Area" variable and considered 3 different areas:

- NORD: Friuli-Venezia Giulia, Liguria, Lombardia, Emilia-Romagna, Piemonte, Veneto, Trentino-Alto Adige and Valle D'Aosta.
- CENTRO: Lazio, Abruzzo, Toscana, Umbria, Marche, Molise.
- SUD: Basilicata, Calabria, Campania, Puglia, Sardegna, Sicilia.

Figure 11: Map of the Italian country divided into 3 Macro-areas.

The chi2 test was performed now with the two variables piracy and macroarea. Results are:

Table 6: Software pirates and customers for each Macro-Area.

MACROAREA	PIRATA		TOTAL
	NO	YES	
CENTRO	74	77	151
	49.01	50.99	100.00
NORD	444	264	708
	62.71	37.29	100.00
SUD	27	54	81
	33.33	66.67	100.00
Total	545	395	940
	57.98	42.02	100.00

Table 7:Chi2 Test - dependence of Piracy and Macro-Areas.

PEARSON CHI2	d.o.f.	p-value
31,693	2	0.000

The existence of some relation between the new variable macroarea and piracy is still confirmed, what is interesting from this aggregate is that we are now able to tell that not only there may exist a certain relation between the variable but also that the Macro-Areas behave in a different way:

- SUD: composed by the southern regions; in this macroarea the effect on piracy is higher than the other.
- CENTRO: composed by central regions; the pirate and customers are balanced but there is some tendency to pirate software.
- NORD: composed by northern regions; the amount of piracy is critically lower with respect to the other areas.

The same stands using a chi2 and excluding the 5 regions which may distort the results:

Table 8: Chi2 Test - dependence of Piracy and Macro-areas excluding Sardegna, Sicilia, Basilicata, Molise and Valle D'Aosta.

PEARSON CHI2	d.o.f.	p-value
31,800	2	0.000

Now we can consider SUD as a reference of cracking macroarea and test the behaviour of NORD and CENTRO to find out how strong is the correlation among these variables and piracy. To do so we use a regress function, a logistic one, to understand the

relation among the dependent binary variable “pirate” and the independent variable that is the macroarea, more specifically we want to understand how much a company is less likely to pirate a software with respect to the SUD area if it is located in the center or in the nord.

Table 9: Logistic Regression - Correlation among Piracy and the two Macro-areas NORD and CENTRO.

LOGISTIC REGRESSION		
	# Obs.	940
	LR chi2(2)	31.46
	Prob > chi2	0.0000
	Pseudo R2	0.0246

PIRATA	Coef.	Std. Err.	z	P>|z|	[95% Conf.]	
NORD	-1.213.023	.2481848	-4,89	0.000	-1.699.456	-.7265894
CENTRO	-.6534069	.2864543	-2,28	0.023	-1.214.847	-.0919667
_cons	.6931472	.2357023	2,94	0.003	.2311792	1.155.115

The p-value (in the P>|z| column) are lower than the 5% significance level for the 2 regressors Nord and Centro. Therefore, we cannot exclude the existence of a correlation with respect to piracy and firm localization, with firms in the South being more prone to piracy to firms located in the Centre and especially in the Nord. In other terms, a company which operates in one of these two macroareas is less likely to crack the software and the magnitude of such “good” behaviour is stronger in the nord as the coefficient is larger (in absolute value) than the coefficient for the central macroarea.

Next steps in the research are taken to understand if there is a variable bias in the statistical dependence of piracy and geographical position. We searched the existence of other variables able to explain the piracy behaviour to check if location alone was not crucial as the analysis pointed out.

5.3.2 COMPANY SIZE

The following step was to consider a “size” variable for each company and to detect the existence of a relation with size and piracy. Afterwards, we have done the same with a measure of profitability.

Our objective is to determine whether exists a relation among piracy and these measures and if it is stronger than the relation considered with the geographical

position of the companies, furthermore we can detect if correlation exists between geographical position and economic variables.

We define the size of a company considering its total operating revenue, that is the value that the company creates thanks to its primary activities (products or services) in a fiscal year considering not only sold products or services but also what remains as stock.

The dataset considered for this analysis contains 870 companies due to economic data availability, in this case for the operating revenue variable, the year considered is the 2016 as it allowed to use the largest possible sample. In this sample there are 375 pirate companies and 495 non-pirate customers.

4 categories of size were defined (so that each category has a similar concentration of companies, that was possible considering the four quartiles given by the operating revenue variable from year 2016)

- Dimensione_1: operating revenue < 1.114 k€
- Dimensione_2: 1.114 k€ <= operating revenue < 4.912 k€
- Dimensione_3: 4.912 k€ <= operating revenue < 21.665 k€
- Dimensione_4: operating revenue >= 21.665 k€

The size variable “Dimensione” was tested against the binary variable “pirata” and the results show that there is no such statistical dependence as it happened with the geographical variable:

Table 10: Software pirates and customers for the 4 categories of SIZE defined.

SIZE	PIRATA		TOTAL
	NO	YES	
1	116	101	217
	53.46	46.54	100.00
2	123	95	218
	56.42	43.58	100.00
3	130	87	217
	59.91	40.09	100.00
4	126	92	218
	57.80	42.20	100.00
Total	495	375	870
	56.90	43.10	100.00

Table 11: Chi2 Test - dependence of Piracy and Size.

PEARSON CHI2	d.o.f.	p-value
1,942	3	0,585

The p-value is larger than the 5% significance level so we refuse the null hypothesis, in other words, we cannot tell that there exists a dependence between these two variables. By the way data is still insightful, we can tell that for lower size (1-2) the number of pirates is higher than larger size (3-4). What we imagine is that in smaller companies the decision upon cracking the software is commonly taken by the entire firm or an entire office, while for larger companies the cracking behaviour may be the act of an individual not properly controlled.

A common thought is that there may be a certain correlation between size of a company and piracy rate (people expects smaller and poorer companies to crack software), but these statistics are turning this thought and another confirmation comes with an analysis of companies' wealth.

5.3.3 COMPANY ECONOMIC AND FINANCIAL STATUS

Another variable of economic "wealth" was defined: we considered 3 different ways to define a company's wealth starting from a more comprehensive factor of FINANCIAL HEALTH that is the value added to a stricter one that is earning before taxes, then each of these variables was divided by the total immobilization of the same fiscal year to standardize the measure.

The 3 measure of wealth are:

- Redditiv_1: (value added 2016)/ (tot. immobilization 2016)
- Redditiv_2: (operative revenue 2016)/ (tot. immobilization 2016)
- Redditiv_3: (Earnings before taxes 2016)/ (tot. immobilization 2016)

Again, the year 2016 gave us the largest sample for the analysis, in this case we have a sample of 837 companies.

First, we tested the existence of correlation between piracy and the defined "wealth" variables with logit regressions:

Redditiv_1:

Table 12: Logistic Regression - Correlation between Piracy and Redditiv_1 wealth variable.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	0.49
Prob > chi2	0.4835
Pseudo R2	0.0004

PIRATA	Coef.	Std. Err.	z	P>|z|	[95% Conf.]	
Redditiv_1	-.0048558	.0071337	-0.68	0.496	-.0188375	.009126
_cons	-.2579257	.0748011	-3.45	0.001	-.4045331	-.1113182

Redditiv_2:

Table 13: Logistic Regression - Correlation between Piracy and Redditiv_2 wealth variable.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	3.02
Prob > chi2	0.0823
Pseudo R2	0.0026

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
Redditiv_2	-.0335257	.0224477	-1.49	0.135	-.0775225	.010471
_cons	-.2468825	.0721157	-3.42	0.001	-.3882267	-.1055382

Redditiv_3:

Table 14:Logistic Regression - Correlation between Piracy and Redditiv_3 wealth variable.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	3.65
Prob > chi2	0.0561
Pseudo R2	0.0032

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
Redditiv_3	-.0367084	.0226557	-1.62	0.105	-.0811128	.007696
_cons	-.2446456	.0720165	-3.40	0.001	-.3857955	-.1034958

These three tests show that wealth variable seems to be heavier in defining piracy behaviour when the value considered is increasingly net, but no statistical correlation between wealth and piracy can be taken in consideration as p-values are higher than 0.05 like what happened with the size variable, this implies that, until now, the correlation with the geographical position is the only one reliable.

5.3.4 LOCATION-SIZE-WEALTH JOINT ANALYSIS

So far, it seems that wealth and size of the company are not crucial for the piracy behaviour. In order to exclude this dependence, we used these variables jointly as regressors to see if they may be relevant considering some level of correlation (among them) and with the geographical position.

The size variables considered are the first 3 as the last one is used to compare the effect on piracy given that a company has a lower size.

SIZE:

Table 15: Logistic Regression - Correlation among Piracy, Size varibles, NORD and CENTRO Macro-areas.

LOGISTIC REGRESSION		
	# Obs.	870
	LR chi2(2)	29.14
	Prob > chi2	0.0000
	Pseudo R2	0.0245

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.1071844	.1966676	0.55	0.586	-.278277	.4926459
dimensione_2	.0354028	.1964202	0.18	0.857	-.3495736	.4203793
dimensione_3	-.1377438	.1981026	-0.70	0.487	-.5260177	.2505302
NORD	-1.178424	.2626451	-4.49	0.000	-1.693199	-.6636494
CENTRO	-.622001	.3011196	-2.07	0.039	-1.212185	-.0318173
_cons	.7071564	.2833342	2.50	0.013	.1518316	1.262481

WEALTH:

With Redditiv_1:

Table 16: Logistic Regression - Correlation among Piracy, NORD, CENTRO and Redditiv_1.

Logistic regression		
	# obs	837
	LR chi2(3)	26.13
	Prob > chi2	0.0000
	Pseudo R2	0.0228

pirata	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
redditiv_1	-.0026946	.0070503	-0.38	0.702	-.016513	.0111238
nord	-1.193927	.27218	-4.39	0.000	-1.72739	-.6604641
centro	-.6594003	.311419	-2.12	0.034	-1.26977	-.0490303
_cons	.7427813	.2595786	2.86	0.004	.2340167	1.251546

With Redditiv_2:

Table 17: Logistic Regression - Correlation among Piracy, NORD, CENTRO and Redditiv_2.

Logistic regression

# obs	837
LR chi2(3)	28.15
Prob > chi2	0.0000
Pseudo R2	0.0246

pirata	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
redditiv_2	-.0284014	.0220581	-1.29	0.198	-.0716345	.0148317
nord	-1.186028	.2720413	-4.36	0.000	-1.719219	-.6528363
centro	-.6625601	.3113756	-2.13	0.033	-1.272845	-.0522751
_cons	.752178	.2596435	2.90	0.004	.2432861	1.26107

With Redditiv_3:

Table 18: Logistic Regression - Correlation among Piracy, NORD, CENTRO and Redditiv_3.

Logistic regression

# Obs	837
LR chi2(3)	28.64
Prob > chi2	0.0000
Pseudo R2	0.0250

pirata	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
redditiv_2	-.0313417	.022252	-1.41	0.159	-.0749549	.0122714
nord	-1.182875	.2720991	-4.35	0.000	-1.71618	-.6495709
centro	-.6604168	.3114032	-2.12	0.034	-1.270756	-.0500779
_cons	.7515666	.259608	2.90	0.004	.2427443	1.260389

Both wealth and size do not show any significant correlation with the geographical area as, testing the variables together, results are still the same (p-value for nord-centro are still significant while the ones of size and wealth variables are still larger than 5%).

Then we checked if there is an interaction among size and wealth depending on the size and considering the geographical area.

Considering redditiv_1:

Table 19: Logistic Regression - Correlation among Piracy, Size varibles, NORD, CENTRO and Redditiv_1.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	27.85
Prob > chi2	0.0001
Pseudo R2	0.0243

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.1283925	.2043205	0.63	0.530	-.2720683	.5288534
dimensione_2	.0071018	.1978174	0.04	0.971	-.3806132	.3948168
dimensione_3	-.1385943	.1989775	-0.70	0.486	-.528583	.2513944
NORD	-1.190395	.2730859	-4.36	0.000	-1.725634	-.6551569
CENTRO	-.6541893	.3121225	-2.10	0.036	-1.265938	-.0424404
redditiv_1	-.0027353	.0071481	-0.38	0.702	-.0167453	.0112747
_cons	.7437918	.2919424	2.55	0.011	.1715953	1.315988

Considering redditiv_2:

Table 20: Logistic Regression - Correlation among Piracy, Size varibles, NORD, CENTRO and Redditiv_2.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	29.21
Prob > chi2	0.0000
Pseudo R2	0.0261

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.1327295	.2043328	0.65	0.516	-.2677556	.5332145
dimensione_2	.0128053	.1977328	0.06	0.948	-.374744	.4003546
dimensione_3	-.1376444	.1988887	-0.69	0.489	-.527459	.2521703
NORD	-1.182217	.2728812	-4.33	0.000	-1.717055	-.6473799
CENTRO	-.6570285	.3120411	-2.11	0.035	-1.268618	-.0454391
redditiv_2	-.0283129	.0218443	-1.30	0.195	-.071127	.0145012
_cons	.7501046	.2922781	2.57	0.010	.1772501	1.322959

Considering redditiv_3:

Table 21: Logistic Regression - Correlation among Piracy, Size varibles, NORD, CENTRO and Redditiv_3.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	30.38
Prob > chi2	0.0000
Pseudo R2	0.0265

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.1332357	.2043778	0.65	0.514	-.2673374	.5338089
dimensione_2	.0114862	.1976878	0.06	0.954	-.3759747	.3989471
dimensione_3	-.1356272	.1989631	-0.68	0.495	-.5255877	.2543333
NORD	-1.179015	.2729391	-4.32	0.000	-1.713965	-.6440639
CENTRO	-.6548759	.3120739	-2.10	0.036	-1.266529	-.0432223
redditiv_3	-.0311265	.0221121	-1.41	0.159	-.0744655	.0122125
_cons	.7490497	.2923032	2.56	0.010	.176146	1.321953

All the tests show no interaction among the variables of size and wealth with respect to the geographical area. So far, we can tell that the geographical position alone is the only variable able to explain the piracy behaviour in Italy.

5.3.5 ECONOMIC ACTIVITY

Thanks to the data available on the Aida Database it was possible to gather information about the economic activity classification of each company described by the Ateco 2007 code adopted by ISTAT, The Italian National Institute of Statistics that is a public research organisation, main producer of official statistics in the service of citizens and policy-makers.

Starting in January 2008 Istat has adopted the new Ateco 2007 classification of economic activities, this is nonetheless the Italian national version of the European nomenclature, Nace Rev. 2, published in the Official Journal of 20 December 2006 (Regulation (EC) no 1893/2006 of the European Parliament and of the Council of 20 December 2006).

The nomenclature is composed of an alphanumeric code (letters and numbers) and their combination follows a scheme to give a detailed description of the economic activity and hence the industry/market in which a company operates.

While the initial letter gives the Macro-activity, the following number provide growing details and can be subdivided into:

- Sezioni (Section)
- Divisioni (Division)
- Gruppi (Group)
- Classi (Class)
- Categorie (Category)
- Sottocategorie (Sub-category)

01.11.10 - COLTIVAZIONE DI CEREALI (ESCLUSO IL RISO)

	Codice	Definizione
Sezione	A	AGRICOLTURA, SILVICOLTURA E PESCA
Divisione	01	COLTIVAZIONI AGRICOLE E PRODUZIONE DI PRODOTTI ANIMALI, CACCIA E SERVIZI CONNESSI
Gruppo	01.1	COLTIVAZIONE DI COLTURE AGRICOLE NON PERMANENTI
Classe	01.11	COLTIVAZIONE DI CEREALI (ESCLUSO IL RISO), LEGUMI DA GRANELLA E SEMI OLEOSI
Categoria	01.11.1	COLTIVAZIONE DI CEREALI
Sottocategoria	01.11.10	COLTIVAZIONE DI CEREALI (ESCLUSO IL RISO)

Figure 12: Example of an Ateco 2007 code for cereals cultivation.

The dataset available was composed of 256 different ateco 2007 codes, to get an insightful clue about the dependence of the companies activity with respect to software piracy we grouped these codes in macro-categories considering only the first two digits of the code, furthermore we neglected the economic activities code where we observed less than 5 companies to avoid any distortion in the tests. The dataset analysed was composed of 880 companies of which 371 pirate companies and 509 non-pirate ones, while the economic macro category analysed were 23 different ones. (the table describes the several economic areas).

Table 22:Description of the 23 Ateco 2007 code analysed and relative number of identified companies (frequency)

ATECO_2	FREQ.	ECONOMIC ACTIVITY
22	29	MANUFACTURE OF RUBBER AND PLASTIC PRODUCTS
24	14	METALLURGY
25	142	MANUFACTURE OF METAL PRODUCTS (EXCLUDING MACHINERY AND EQUIPMENT)
26	34	MANUFACTURE OF COMPUTERS, ELECTRONICS AND OPTICS; ELECTROMEDICAL EQUIP…
27	59	MANUFACTURE OF ELECTRICAL AND NON-ELECTRICAL HOUSEHOLD APPLIANCES
28	224	MANUFACTURE OF MACHINERY AND EQUIPMENT (NOT QUALIFIED BY OTHER CODES)
29	35	MANUFACTURE OF MOTOR VEHICLES, TRAILERS AND SEMI-TRAILERS
30	14	MANUFACTURE OF OTHER MEANS OF TRANSPORT
31	9	FURNITURE MANUFACTURING
32	15	OTHER MANUFACTURING INDUSTRIES
33	20	REPAIR, MAINTENANCE AND INSTALLATION OF MACHINERY AND EQUIPMENT
41	5	BUILDING CONSTRUCTION
43	13	SPECIALISED CONSTRUCTION WORK
46	52	WHOLESALE TRADE (EXCLUDING TRADE IN MOTOR VEHICLES AND MOTORCYCLES)
47	6	RETAIL TRADE (EXCEPT OF MOTOR VEHICLES AND MOTORCYCLES)
58	5	PUBLISHING ACTIVITIES (INCLUDING SOFTWARE)
62	24	SOFTWARE PRODUCTION, IT CONSULTANCY AND RELATED ACTIVITIES
64	16	FINANCIAL SERVICES ACTIVITIES (EXCEPT INSURANCE AND PENSION FUNDING)
68	25	REAL ESTATE
70	37	MANAGEMENT AND MANAGEMENT CONSULTANCY ACTIVITIES
71	35	ARCHITECTURAL AND ENGINEERING ACTIVITIES; TECHNICAL TESTING AND ANALYSIS
72	8	SCIENTIFIC RESEARCH AND DEVELOPMENT
74	59	OTHER PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES

This variable explains also what the market for CAD software is, the background of this research is now more clear as we can tell that the majority of cad users are manufacturer companies with a large number of machinery producers (industrial machinery) followed by metal products producers (components, dies for several uses and other metal products different from machineries), numerous are the manufacturer of electronical components, pc and household appliances, other companies are Architectural or Design companies, a huge part is composed by

manufacturer of motor vehicles and is also interesting the presence of several Consultancy companies of different kind.

The Chi2 test on this data reveals that, statistically, we cannot accept the null hypothesis that asks for the independence of the two variables, hence we can say that there exists a certain dependence between software piracy and economic activity (p-value is lower than 5%). Results are the following:

Table 23: Pirates companies and customers for the identified ateco 2007 codes.

ATECO_2	PIRATA		TOTAL
	NO	YES	
22	12	17	29
	41.38	58.62	100.00
24	5	9	14
	35.71	64.29	100.00
25	78	64	142
	54.93	45.07	100.00
26	19	15	34
	55.88	44.12	100.00
27	33	26	59
	55.93	44.07	100.00
28	136	88	224
	60.71	39.29	100.00
29	21	14	35
	60.00	40.00	100.00
30	7	7	14
	50.00	50.00	100.00
31	6	3	9
	66.67	33.33	100.00
32	6	9	15
	40.00	60.00	100.00
33	11	9	20
	55.00	45.00	100.00
41	2	3	5
	40.00	60.00	100.00
43	10	3	13
	76.92	23.08	100.00
46	32	20	52
	61.54	38.46	100.00
47	5	1	6
	83.33	16.67	100.00
58	3	2	5
	60.00	40.00	100.00
62	6	18	24

	25.00	75.00	100.00
64	15	1	16
	93.75	6.25	100.00
68	22	3	25
	88.00	12.00	100.00
70	31	6	37
	83.78	16.22	100.00
71	21	14	35
	60.00	40.00	100.00
72	3	5	8
	37.50	62.50	100.00
74	25	34	59
	42.37	57.63	100.00
Total	509	371	880
	57.84	42.16	100.00

Table 24: Chi2 Test - dependence of Piracy and Economic Activity (ateco 2007 code).

PEARSON CHI2	d.o.f.	p-value
60,478	22	0,000

Two elements are tested thanks to the ateco 2007 code:

- The industry sector more likely to crack given other variables in a joint analysis.
- The dependence of piracy and industry sector health.

5.3.6 ECONOMIC ACTIVITY JOINT ANALYSIS

Next test is deployed to define if a company working in some economic field is more likely to pirate the software with respect to others. Keeping constant wealth, having SUD location as benchmark and considering the different 4 level of size.

We introduced dummy variables to test this qualitative variable, belonging to one economic category or another. A dummy variable is a binary one and is helpful when used in a multivariate regression to catch the effect of a qualitative variable (ateco2007 code in this case) on the dependent variable (piracy).

To perform a robust test the ateco 2007 category were grouped considering the first digit of the code. We identified 9 categories of which 2 were not useful to the analysis due to low number of observation (categories 8 and 9), category 7 has the benchmark

role. The following tables show the economic activity involved and the result of the regression analysis performed with the most significant wealth variable (redditiv_3).

Table 25: Pirates companies and customers for the identified 9 economic activities aggregates.

ateco_1	Pirata		Total
	NO	YES	
1	11	8	19
	57.89	42.11	100.00
2	308	237	545
	56.51	43.49	100.00
3	33	29	62
	53.23	46.77	100.00
4	52	29	81
	64.20	35.80	100.00
5	3	2	5
	60.00	40.00	100.00
6	46	25	71
	64.79	35.21	100.00
7	82	62	144
	56.94	43.06	100.00
8	5	1	6
	83.33	16.67	100.00
9	1	0	1
	100.00	0.00	100.00
Total	541	393	934
	57.92	42.08	100.00

Table 26: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and dummies for the 9 economic activities aggregate

LOGISTIC REGRESSION

# Obs.	833
LR chi2(2)	35.10
Prob > chi2	0.0005
Pseudo R2	0.0308

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.2739796	.2328512	1.18	0.239	-.1824003	.7303596
dimensione_2	.081455	.2025263	0.40	0.688	-.3154892	.4783992
dimensione_3	-.132866	.2003228	-0.66	0.507	-.5254914	.2597594
nord	-1.1718	.2747795	-4.26	0.000	-1.710358	-.6332421
centro	-.6041505	.3147179	-1.92	0.055	-1.220986	.0126851
redditiv_3	-.0305084	.0222848	-1.37	0.171	-.0741857	.0131689
ateco_1dum1	.2422997	.5297356	0.46	0.647	-.795963	1.280562
ateco_1dum2	.1998842	.2332997	0.86	0.392	-.2573747	.6571432
ateco_1dum3	.1676402	.3532235	0.47	0.635	-.5246651	.8599455

ateco_1dum4	-.1741059	.3292349	-0.53	0.597	-.8193945	.4711828
ateco_1dum5	-.2097367	.9682803	-0.22	0.829	-2.107531	1.688058
ateco_1dum6	-.2664453	.3418107	-0.78	0.436	-.9363819	.4034912
ateco_1dum7	0	(omitted)				
ateco_1dum8	0	(omitted)				
ateco_1dum9	0	(omitted)				
_cons	.5871064	.3752794	1.56	0.118	-.1484276	1.32264

The regression tells us that it seems to be no significance in the aggregate ateco_1. What is interesting to note is that coefficients in the first 3 category are positive while in the 4-5 and 6 are negative. In this table a positive coefficient represents an activity where piracy is more likely (with respect to code 7, our benchmark). To find out if this is really significant, we grouped category 1-2-3 in the aggregate Manufacturing (man) and performed another regression.

Table 27: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and Manufacturing activities.

LOGISTIC REGRESSION

# Obs.	837
LR chi2(2)	34.43
Prob > chi2	0.0000
Pseudo R2	0.0301

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.3076028	.2229859	1.38	0.168	-.1294415	.7446472
dimensione_2	.0850137	.2012808	0.42	0.673	-.3094895	.4795169
dimensione_3	-.1251478	.1993005	-0.63	0.530	-.5157696	.265474
nord	-1.166755	.2738733	-4.26	0.000	-1.703537	-.6299736
centro	-.6297088	.3133547	-2.01	0.044	-1.243873	-.0155449
redditiv_3	-.0296712	.0222075	-1.34	0.182	-.0731971	.0138547
man	.3454097	.1727123	2.00	0.046	.0068997	.6839196
_cons	.4323075	.3322359	1.30	0.193	-.2188628	1.083478

From this detailed analysis we can tell that the manufacturing category is significant (p-value lower than 5% significance level) in defining the piracy behaviour and a company working in the manufacturing business is more likely to pirate software.

To get an even more robust and reliable test we decided to group each activity by the macro category defined by the ateco code first letter.

What we identified is that 8 category present more than 5 observation, the group identified are described by the tables below, chi2 test is performed.

Table 28: Pirates companies and customers for the identified 8 ateco macro sector (ateco_letter).

ateco_letter	Pirata		Total
	0	1	
C	349	273	622
	56.11	43.89	100.00
F	14	7	21
	66.67	33.33	100.00
G	38	22	60
	63.33	36.67	100.00
J	10	23	33
	30.30	69.70	100.00
K	17	1	18
	94.44	5.56	100.00
L	22	3	25
	88.00	12.00	100.00
M	82	61	143
	57.34	42.66	100.00
N	4	1	5
	80.00	20.00	100.00
Total	536		927
	57.82	42.18	

Table 29: Economic activities in ateco "C" (Manufacturing).

(C) ATTIVITÀ MANIFATTURIERE	
10	INDUSTRIE ALIMENTARI
14	CONFEZIONE DI ARTICOLI DI ABBIGLIAMENTO; CONFEZIONE DI ARTICOLI IN PELLE E PELLICCIA
15	FABBRICAZIONE DI ARTICOLI IN PELLE E SIMILI
16	INDUSTRIA DEL LEGNO E DEI PRODOTTI IN LEGNO E SUGHERO (ESCLUSI I MOBILI); FABBRICAZIONE DI ARTICOLI IN PAGLIA E MATERIALI DA INTRECCIO
17	FABBRICAZIONE DI CARTA E DI PRODOTTI DI CARTA
18	STAMPA E RIPRODUZIONE DI SUPPORTI REGISTRATI
19	FABBRICAZIONE DI COKE E PRODOTTI DERIVANTI DALLA RAFFINAZIONE DEL PETROLIO
20	FABBRICAZIONE DI PRODOTTI CHIMICI
21	FABBRICAZIONE DI PRODOTTI FARMACEUTICI DI BASE E DI PREPARATI FARMACEUTICI
22	FABBRICAZIONE DI ARTICOLI IN GOMMA E MATERIE PLASTICHE
23	FABBRICAZIONE DI ALTRI PRODOTTI DELLA LAVORAZIONE DI MINERALI NON METALLIFERI
24	METALLURGIA
25	FABBRICAZIONE DI PRODOTTI IN METALLO (ESCLUSI MACCHINARI E ATTREZZATURE)
26	FABBRICAZIONE DI COMPUTER E PRODOTTI DI ELETTRONICA E OTTICA; APPARECCHI ELETTRO MEDICALI, APPARECCHI DI MISURAZIONE E DI OROLOGI
27	FABBRICAZIONE DI APPARECCHIATURE ELETTRICHE ED APPARECCHIATURE PER USO DOMESTICO NON ELETTRICHE
28	FABBRICAZIONE DI MACCHINARI ED APPARECCHIATURE NCA

29	FABBRICAZIONE DI AUTOVEICOLI, RIMORCHI E SEMIRIMORCHI
30	FABBRICAZIONE DI ALTRI MEZZI DI TRASPORTO
31	FABBRICAZIONE DI MOBILI
33	RIPARAZIONE, MANUTENZIONE ED INSTALLAZIONE DI MACCHINE ED APPARECCHIATURE

Table 30: Economic activities in ateco "F" (Construction).

(F) COSTRUZIONI	
41	COSTRUZIONE DI EDIFICI
42	INGEGNERIA CIVILE
43	LAVORI DI COSTRUZIONE SPECIALIZZATI

Table 31: Economic activities in ateco "G" (Wholesale and retail trade; repair of motor vehicles and motorcycles).

(G) COMMERCIO ALL'INGROSSO E AL DETTAGLIO; RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI	
45	COMMERCIO ALL'INGROSSO E AL DETTAGLIO E RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI
46	COMMERCIO ALL'INGROSSO (ESCLUSO QUELLO DI AUTOVEICOLI E DI MOTOCICLI)
47	COMMERCIO AL DETTAGLIO (ESCLUSO QUELLO DI AUTOVEICOLI E DI MOTOCICLI)

Table 32: Economic activities in ateco "J" (Information and communication services).

(J) SERVIZI DI INFORMAZIONE E COMUNICAZIONE	
58	ATTIVITÀ EDITORIALI
62	PRODUZIONE DI SOFTWARE, CONSULENZA INFORMATICA E ATTIVITÀ CONNESSE
63	ATTIVITÀ DEI SERVIZI D'INFORMAZIONE E ALTRI SERVIZI INFORMATICI

Table 33: Economic activities in ateco "K" (Financial and insurance activities).

(K) ATTIVITÀ FINANZIARIE E ASSICURATIVE	
64	ATTIVITÀ DI SERVIZI FINANZIARI (ESCLUSE LE ASSICURAZIONI E I FONDI PENSIONE)
66	ATTIVITÀ AUSILIARIE DEI SERVIZI FINANZIARI E DELLE ATTIVITÀ ASSICURATIVE

Table 34: Economic activities in ateco "L" (Real Estate).

(L) ATTIVITA' IMMOBILIARI	
68	ATTIVITÀ IMMOBILIARI

Table 35: Economic activities in ateco "M" (Professional, scientific and technical activities).

(M) ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE	
70	ATTIVITÀ DI DIREZIONE AZIENDALE E DI CONSULENZA GESTIONALE
71	ATTIVITÀ DEGLI STUDI DI ARCHITETTURA E D'INGEGNERIA; COLLAUDI ED ANALISI TECNICHE
72	RICERCA SCIENTIFICA E SVILUPPO
74	ALTRE ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE

Table 36: Economic activities in ateco "N" (Rental, travel agencies, business support services).

(N) NOLEGGIO, AGENZIE DI VIAGGIO, SERVIZI DI SUPPORTO ALLE IMPRESE	
79	ATTIVITÀ DEI SERVIZI DELLE AGENZIE DI VIAGGIO, DEI TOUR OPERATOR E SERVIZI DI PRENOTAZIONE E ATTIVITÀ CONNESSE
82	ATTIVITÀ DI SUPPORTO PER LE FUNZIONI D'UFFICIO E ALTRI SERVIZI DI SUPPORTO ALLE IMPRESE

Table 37: Chi2 Test - dependence of Piracy and Economic Activity (ateco_letter).

PEARSON CHI2	d.o.f.	p-value
32,672	7	0,000

As we expect the chi2 test shows statistical dependence. Again, we performed a logit regression considering dummies variables for each one of the ateco_letter sector described above.

Table 38: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and dummies for the 8 ateco_letter economic activities.

LOGISTIC REGRESSION

# Obs.	829
LR chi2(2)	61.93
Prob > chi2	0.0000
Pseudo R2	0.0546

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.4682678	.2399862	1.95	0.051	-.0020965	.9386321
dimensione_2	.1079521	.2039338	0.53	0.597	-.2917508	.5076549
dimensione_3	-.1334123	.2004957	-0.67	0.506	-.5263767	.2595521
nord	-1.125617	.2831266	-3.98	0.000	-1.680535	-.5706989
centro	-.6157446	.3236239	-1.90	0.057	-1.250036	.0185465
redditiv_3	-.0390064	.0241969	-1.61	0.107	-.0864314	.0084187
(C) ateco_letter_dum1	.307352	.2341339	1.31	0.189	-.1515419	.766246
(F) ateco_letter_dum2	-.4098636	.5556926	-0.74	0.461	-1.499001	.679274
(G) ateco_letter_dum3	.0382908	.3617092	0.11	0.916	-.6706463	.7472279
(J) ateco_letter_dum4	1.242691	.4822522	2.58	0.010	.2974936	2.187887
(K) ateco_letter_dum5	-2.176368	1.065555	-2.04	0.041	-4.264817	-.0879192
(L) ateco_letter_dum6	-1.955326	.7755992	-2.52	0.012	-3.475472	-.4351793
(M) ateco_letter_dum7	0	(omitted)				
(N) ateco_letter_dum8	0	(omitted)				
_cons	.4317209	.3820544	1.13	0.258	-.317092	1.180534

From this more detailed analysis we can tell that manufacturing sector is still one in which people crack more but the p-value show no significance as it did before. The real outcome is that sector “J”, Information and Communication services, is now the one in which piracy behaviour is more concentrated. Statistical dependence is registered for sector K and L too where a negative sign in the coefficient defines two sectors in which a company is less likely to pirate.

Our thought is that piracy is widespread in the manufacturing industry and information and communication. In the first sector because the CAD software is largely used, in the second one because there are more and more employees using computers to work so the demand for software is higher and thus the piracy behaviour.

This analysis is enriched by further tests performed with piracy_rate variable defined in chapter 5.3.8.

5.3.7 PIRACY AND INDUSTRY HEALTH

In this case we used again the macro activity defined by the first letter of the ateco code and gathered information about industry health through istat databases.

Statistics and Graph about the economic trust of a given field/market and industry health (considering either production volume for manufacturing industries and revenue for service ones) are gathered from the Istat 2018 report of industries sector competitivity.

Index considered by Istat in the following graphs are:

- Clima di fiducia nel settore economico (climate of confidence in the economic sector), an index that for manufacturing sector is given by the arithmetic mean of 3 measures, level of orders, level of inventory and expected production volume which result is then indexed.
  For the construction sector is the indexed arithmetic mean of construction plans (orders) and expected level of employment.
  For the service industry is the indexed arithmetic mean of expected orders and economy trend.
- Indice della produzione industriale (Industrial production index) given by the variation of production volume along the time.
- Indice del fatturato (turnover index) which is the variation in turnover of service enterprises over time, expressed at current prices.

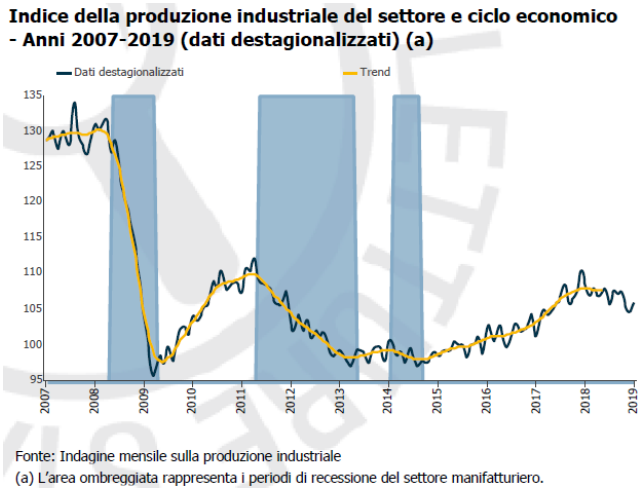


Figure 13:C sector, Industrial production index 2007-2019.

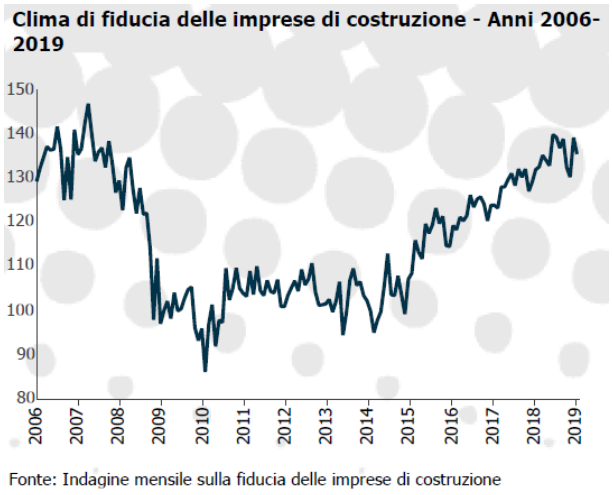
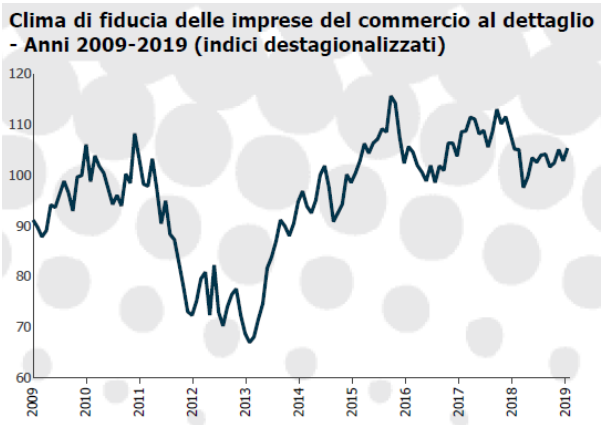


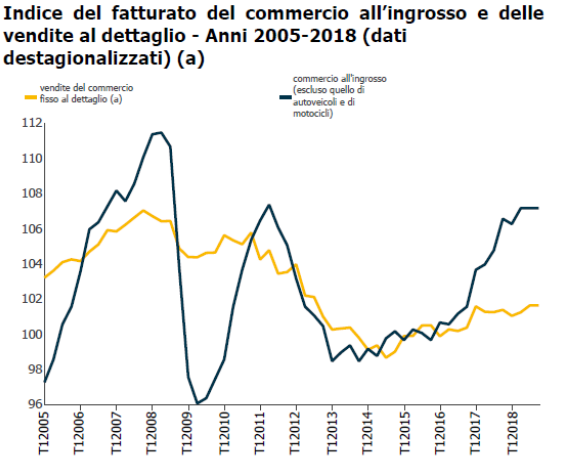
Figure 14:F sector, Climate of confidence 2006-2019.

Figure 15: F sector, Industrial production index 2005-2018.



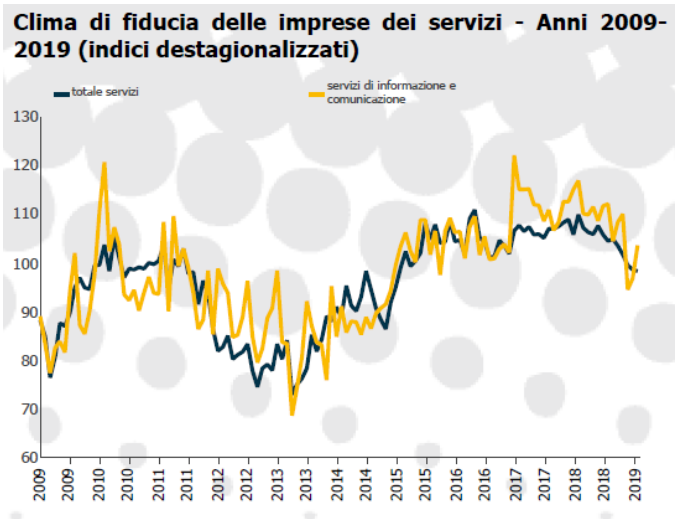
Fonte: Indagine mensile sulla fiducia delle imprese del commercio al dettaglio

Figure 16:G sector, Climate of confidence 2009-2019.



Fonti: Rilevazione mensile delle vendite al dettaglio e Rilevazione trimestrale del fatturato dei servizi
(a) Media aritmetica dei dati mensili destagionalizzati.

Figure 17: G sector, Turnover Index 2005-2018.



Fonte: Indagine mensile sulla fiducia delle imprese dei servizi

Figure 18: J sector, Climate of confidence 2009-2019.

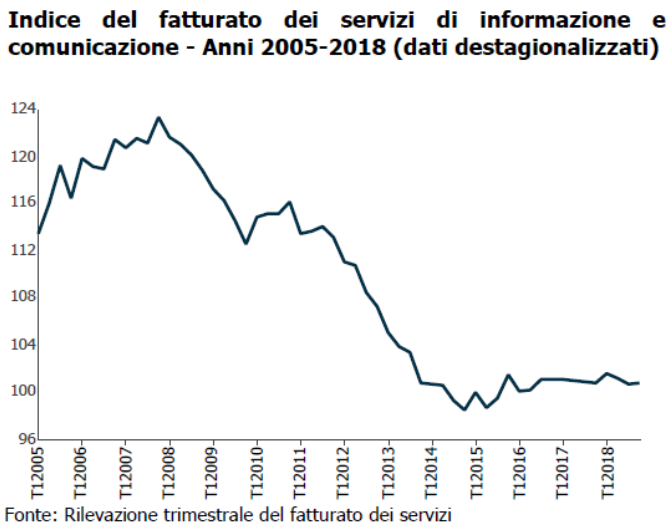


Figure 19: J sector, Turnover Index 2005-2018.

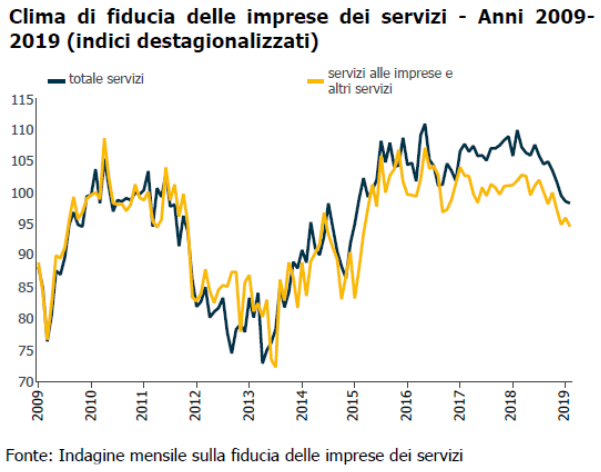


Figure 20: L sector, Climate of confidence 2009-2019.

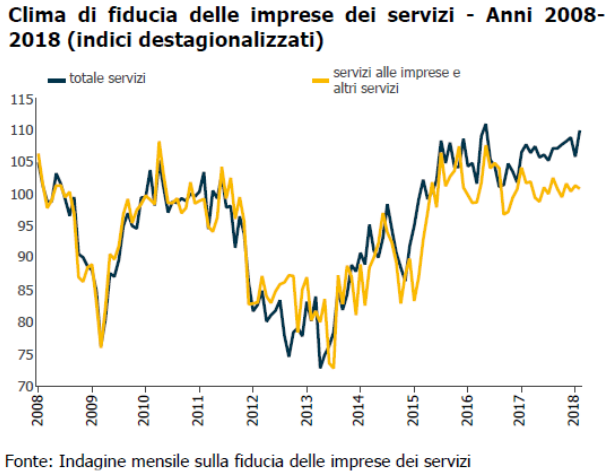


Figure 21: M sector, Climate of confidence 2008-2018.

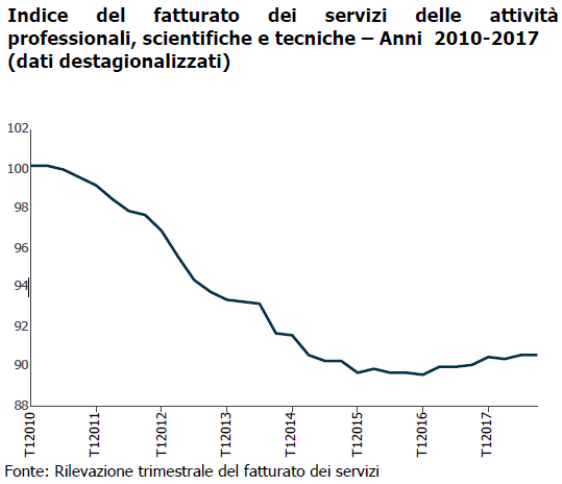


Figure 22: M sector, Turnover Index 2010-2017.

A new dataset was defined where we identified for each of the ateco macro categories the years of recession, signal of bad health in the industry sector.

Table 39: Years of recession for the Economic Activities (ateco_letter).

ateco_letter	Year of recession		
C	2012	2013	2014
F	2012	2013	2014
G	2012	2013	2014
J	2012	2013	2014
L	2012	2013	-
M	2012	2013	-

The dataset was imported in stata and merged with the existent one to perform a regression analysis with the new binary variable "recession". Recession gets a value of 1 if for the ateco_letter industry sector there is a year of recession, 0 otherwise.

The resulting logistic regression is:

Table 40: Logistic Regression - Correlation among Piracy, Size, NORD, CENTRO, Redditiv_3 and Recession variable.

LOGISTIC REGRESSION

# Obs.	817
LR chi2(2)	34.14
Prob > chi2	0.0000
Pseudo R2	0.0305

PIRATA	Coef.	Std.Err	z	P>|z|	[95% Conf.]	
dimensione_1	.2167913	.2090184	1.04	0.300	-.1928772	.6264598
dimensione_2	.033691	.1988774	0.17	0.865	-.3561015	.4234835
dimensione_3	-.1466627	.2001421	-0.73	0.464	-.538934	.2456085

nord	-1.192957	.2798052	-4.26	0.000	-1.741365	-.6445487
centro	-.59262	.3197984	-1.85	0.064	-1.219413	.0341734
redditiv_3	-.0316762	.0220847	-1.43	0.151	-.0749613	.011609
recession	-.8423569	.6205813	-1.36	0.175	-2.058674	.3739601
_cons	.779963	.2987107	2.61	0.009	.1945008	1.365425

The p-value for the recession variable is not lower than the 5% significance level so we cannot consider the existence of statistical dependence between industry health and piracy.

5.3.8 DIFFERENTIATION AMONG PIRATES – PIRACY RATE

Thanks to the data of pirate companies, information about the number of cracked licenses and events, was available. This made possible to define a new variable that describes the magnitude of piracy that a company perpetrated, the new variable called piracy_rate is defined as:

piracy_rate = events_count/machines_count

- event_count represents the number of uses of the cracked licenses, that means that every time the cracked software is opened an event is registered, it doesn't say anything about the duration of the session or the way of usage of the software. When the software is turned down and opened again, a new event is counted.
- machine_count is the number of different computers in which a cracked license is installed, in other words it is the number of cracked licenses.

The dataset is now composed of 355 pirate companies as for 40 of them (the total is 395) data covering machines and events was missing.

We defined four levels of piracy depending on the piracy_rate value, the higher the rate and so the level, the more a company was cracking the software:

- level_1 = piracy_rate < 6
- level_2 = 6 <= piracy_rate < 13.3
- level_3 = 13.3 <= piracy_rate < 33.4
- level_4 = piracy_rate >= 33.4

After defining this variable, we tested the dependence of the magnitude of software piracy with the previous defined variables of macroarea, size and economic activity (ateco2007).

PIRACY RATE vs MACROAREA

Table 41: Magnitude of Piracy in the 3 Italian Macro-areas.

MACROAREA	LEVEL				TOTAL
	1	2	3	4	
CENTRO	10	20	19	21	70
	14.29	28.57	27.14	30.00	100.00
NORD	67	56	54	59	236
	28.39	23.73	22.88	25.00	100.00
SUD	11	13	16	9	49
	22.45	26.53	32.65	18.37	100.00
Total	88	89	89	89	355
	24.79	25.07	25.07	25.07	100.00

Table 42: Chi2 Test - dependence of Piracy_rate and Macro-areas.

PEARSON CHI2	d.o.f.	p-value
8,263	6	0,219

There is no evidence of statistical dependence between piracy_rate and geographical position as all kind of pirates are equally distributed in each area (almost 25%) and the p-value is larger than the significance level of 0.05.

PIRACY RATE vs SIZE

Table 43: Magnitude of Piracy for the 4 different companies'size.

SIZE	LEVEL				TOTAL
	1	2	3	4	
1	17	19	31	24	91
	18.68	20.88	34.07	26.37	100.00
2	14	24	17	26	81
	17.28	29.63	20.99	32.10	100.00
3	22	20	19	22	83
	26.51	24.10	22.89	26.51	100.00
4	28	22	18	14	82
	34.15	26.83	21.95	17.07	100.00
Total	81	85	85	86	337
	24.04	25.22	25.22	25.52	100.00

Table 44: Chi2 Test - dependence of Piracy_rate and Companies'size.

PEARSON CHI2	d.o.f.	p-value
15,398	9	0,081

With companies size too, there is no statistical dependence, even if the p-value show more significance with respect to the macroarea variable, hence we cannot accept the null hypothesis given a p-value of 0.081 that is larger than our significance level.

By the way as we hypothesised in the size analysis alone, we can say that in smaller companies “Size 1” high level cracking “Level 4” is more common while in larger companies “Size 4” low level cracking “Level 1” is more common.

These numbers confirm our thought, in smaller companies software piracy is a behaviour adopted by a community (offices, entire firms) while in larger companies pirates are employees who act alone.

PIRACY RATE vs ECONOMIC ACTIVITY

The test is performed with the sample of all available economic activities and again with the 23 ateco 2007 codes for which we had at least 5 observations:

Table 45: Chi2 Test - dependence of Piracy_rate and Economic activities for all the ateco 2007 code analysed.

PEARSON CHI2	d.o.f.	p-value
112,874	108	0,355

Table 46: Magnitude of Piracy for the 23 ateco 2007 code with at least 5 observations.

ECONOMIC ACTIVIY	level				Total
	1	2	3	4	
22	3	7	1	4	15
	20.00	46.67	6.67	26.67	100.00
24	2	3	3	1	9
	22.22	33.33	33.33	11.11	100.00
25	16	11	12	15	54
	29.63	20.37	22.22	27.78	100.00
26	5	1	5	3	14
	35.71	7.14	35.71	21.43	100.00
27	5	5	5	7	22
	22.73	22.73	22.73	31.82	100.00
28	20	22	18	22	82
	24.39	26.83	21.95	26.83	100.00
29	3	2	3	5	13
	23.08	15.38	23.08	38.46	100.00

30	3	2	2	0	7
	42.86	28.57	28.57	0.00	100.00
31	1	0	2	0	3
	33.33	0.00	66.67	0.00	100.00
32	0	2	0	7	9
	0.00	22.22	0.00	77.78	100.00
33	4	1	0	2	7
	57.14	14.29	0.00	28.57	100.00
41	1	0	1	1	3
	33.33	0.00	33.33	33.33	100.00
43	0	1	1	1	3
	0.00	33.33	33.33	33.33	100.00
46	7	4	5	3	19
	36.84	21.05	26.32	15.79	100.00
47	0	1	0	0	1
	0.00	100.00	0.00	0.00	100.00
58	1	1	0	0	2
	50.00	50.00	0.00	0.00	100.00
62	4	4	7	0	15
	26.67	26.67	46.67	0.00	100.00
64	0	0	0	1	1
	0.00	0.00	0.00	100.00	100.00
68	0	1	1	1	3
	0.00	33.33	33.33	33.33	100.00
70	0	2	2	2	6
	0.00	33.33	33.33	33.33	100.00
71	2	2	1	4	9
	22.22	22.22	11.11	44.44	100.00
72	1	2	2	0	5
	20.00	40.00	40.00	0.00	100.00
74	4	9	11	6	30
	13.33	30.00	36.67	20.00	100.00
Total	82	83	82	85	332
	24.70	25.00	24.70	25.60	100.00

Table 47:Chi2 - Dependence of Piracy rate and the 23 economic activities with at least 5 observations.

PEARSON CHI2	d.o.f.	p-value
68,426	66	0,395

The same happen with the economic activity, there is no dependence with level of piracy, there are not peculiar industries in which pirates behave in a worst way with respect of other.

PIRACY RATE vs WEALTH

Another analysis is performed among piracy_rate and wealth variables defined above.

To test the existence of linear dependence we use a regression analysis, results are:

With Redditiv_1:

Table 48: Linear Regression - Correlation among piracy_rate and wealth variable Redditiv_1.

Source	SS	df	MS
Model	2142.91769	1	2142.91769
Residual	882110.859	325	2714.18726
Total	884253.777	326	2712.4349

# obs	327
F(1, 325)	0.79
Prob > F	0.3749
R-squared	0.0024
Adj R-squared	-0.0006
Root MSE	52.098

piracy_rate	Coef.	Std. Err.	t	P>|t|	[95% Conf.Interval]	
redditiv_1	-.3742008	.4211356	-0.89	0.375	-1.202697	.454295
_cons	32.88899	3.228756	10.19	0.000	26.5371	39.24089

With Redditiv_2:

Table 49: Linear Regression - Correlation among piracy_rate and wealth variable Redditiv_2.

Source	SS	df	MS
Model	103.46575	1	103.46575
Residual	884150.311	325	2720.4625
Total	884253.777	326	2712.4349

# obs	327
F(1, 325)	0.04
Prob > F	0.8455
R-squared	0.0001
Adj R-squared	-0.0030
Root MSE	52.158

piracy_rate	Coef.	Std. Err.	t	P>|t|	[95% Conf.Interval]	
redditiv_2	-.2525763	1.295137	-0.20	0.845	-2.800487	2.295334
_cons	31.75432	2.999426	10.59	0.000	25.85357	37.65506

With Redditiv_3:

Table 50: Linear Regression - Correlation among piracy_rate and wealth variable Redditiv_3.

Source	SS	df	MS
Model	89.4205508	1	89.4205508
Residual	884164.356	325	2720.50571
Total	884253.777	326	2712.4349

# obs	327
F(1, 325)	0.03
Prob > F	0.8562
R-squared	0.0001
Adj R-squared	-0.0030
Root MSE	52.158

piracy_rate	Coef.	Std. Err.	t	P>|t|	[95% Conf.Interval]	
redditiv_3	-.2376402	1.310769	-0.18	0.856	-2.816302	2.341022
_cons	31.73671	2.990074	10.61	0.000	25.85436	37.61905

As the tables show, no one of the defined measures for wealth is able to explain the piracy_rate, they show p-values larger than the 5% significance level (p-value is represented in the P>|t| column).

5.3.9 ROBUSTNESS TEST

The piracy_rate so defined was used to differentiate among low and high level of piracy. What we do now is to consider a cut-off level to distinguish lone pirate events from serial pirates. To discriminate we exploited the median value of piracy_rate equal to 13.3 so a company with a piracy_rate higher than this median value was considered a serial pirate.

We identified 178 serial pirates among the piracy sample, then we generated a new sample comprehensive of all the compliant companies and the serial pirate ones for a total of 723 observations.

We then performed again all the test to see how robust our findings were. What changes now is that we are considering serial pirate companies excluding the ones in which piracy was occasional (performed by lone pirates).

Results are gathered in the next table:

Table 51: Robustness test of previous analysis with "Serial" pirate sample.

TEST	RESULTS
Chi2 - Piracy in Macroarea	p-value = 0.000
Chi2 - Piracy w.r.t. Size	p-value = 0.091
Logit Reg. - Piracy w.r.t redditiv_1	p-value = 0.399
Logit Reg. - Piracy w.r.t redditiv_2	p-value = 0.116
Logit Reg. - Piracy w.r.t redditiv_3	p-value = 0.125
Logit Reg. - Joint analysis size-location-redditiv_3	Size 1, lowest level has a p-value of 0.022, it starts to be significant and affect piracy behaviour
Logit Reg. - Joint analysis size-location-redditiv_3 and economic activity	Considering aggregates (ateco_letter) previous finding are confirmed, industry that pirate more are Manifacturing, Information and Communciation services
Logit Reg. - Joint analysis size-location-redditiv_3 and Industry Health	Confirms previous findings

All previous findings are confirmed, and the test performed on the company size shows now that the smallest is the company the more it is likely to pirate the software, it

seems now that size matters more than the location, more precisely in the regression analysis size level 1 show more significance than variable centro.

CONCLUSIONS

The aim of this thesis work was to investigate for the first time the world of software piracy with concrete data to define the existence of determinants in the piracy behaviour.

We recall that our analysis just covered the market of a specific CAD software in the Italian country and key findings are:

- The location of a company in the country is a determinant in the piracy act, with southern regions of Italy being the ones in which companies are more likely to crack the software.
- There are some economic industries in which using pirated software is more likely than others. This is the case of Manufacturing sector, Information and Communication services.
- Part of piracy behaviour may be explained by the size of a company when we consider serial cracking companies. In this case we noticed that a small company considering its revenue is more likely to be a serial pirate with respect to larger companies.
- The size of a company, jointly analysed with piracy rate (magnitude of piracy), generates a distinction between two kind of behaviour. In small companies the decision upon cracking a software is a common one and applied all along the firm, for larger companies, the level of piracy is lower, meaning that pirates in these companies act alone without any supervision.

Recalling that this study differentiate itself from the available literature as existing papers covers software compliance worldwide (considering software aggregate of different kind), these findings are just the tip of the iceberg, it is clear that more exhaustive data could bring to new findings that may be addressable to all the aspects of software piracy.

The hope is that new data will be made available by different companies and their compliance offices to consent a deeper analysis that would be useful to define determinants of the piracy act.

However, it is not straightforward to guess how much interest companies will have in improve the compliance business and allow more studies as the software industries is facing a fast shift towards the cloud computing.

Being on a cloud platform has many advantages for a software producer making it easier to prevent illegal copies of their products as the software itself will run on a company server and will be streamed to the final customer.

The shift to cloud computing is bringing new challenges to the software compliance:

- New regulations are needed as working on cloud platforms implies to work in multiple territories considering that a customer may work in a country and the software runs in a server located in another one. Most legal problems associated with cloud computing, such as security, privacy, copyright, ownership and access to data must dealt with a cross-jurisdictional basis.
- New means of piracy may be developed to perpetrate the illegal use of software provided on cloud platforms. The compliance business must be up to date with software piracy and be able to address any threat coming from unlawful users in the cloud paradigm.

Furthermore, the possibility to collect data covering cloud software piracy and compare it to current piracy studies would be another great element to define the piracy behaviour.

REFERENCES

BIBLIOGRAPHY

1. Geoff V., Jim G., *Software Development Handbook,* 2nd edition, Texas Instruments, 1981
2. Heru Susanto, Mohammad Nabil Almunawar, *Information Security Management Systems: A Novel Framework and Software as a Tool for Compliance with Information Security Standard*, 1st edition, Apple Academic Press, 2018
3. K. Bagchi, R. Cerveny, *Global software piracy: can economic factors alone explain the trend? Searching for explanations to increasing global software piracy rates,* 2006.
4. A. R. Andrés, R. K. Goel, *Does software piracy affect economic growth? Evidence across countries*, 2011
5. T. T. Moores, *The Effect of National Culture and Economic Wealth on Global Software Piracy Rates*, 2003
6. R. K. Goel, M. A. Nelson, *Determinants of software piracy: economics, institutions, and technology*, 2009
7. A. R. Andrés, *Software piracy and income inequality*, in *Applied Economics Letters*, 2010
8. A. G. Peace, D. F. Galletta, J. Y.L. Thong, *Software Piracy in the Workplace: A Model and Empirical Test*, in *Journal of Management Information Systems*, 2014
9. S. Wang, *The cloud, online piracy and global copyright governance*, in *International Journal of Cultural Studies*, 2016

WEB REFERENCES

1. Altexsoft: https://www.altexsoft.com/blog/business/software-business-models-examples-revenue-streams-and-characteristics-for-products-services-and-platforms/
2. Prescient & Strategic Intelligence: https://www.psmarketresearch.com/market-analysis/cad-software-market
3. DesignTech CAD Academy: https://www.designtechcadacademy.com/knowledge-base/introduction-to-cad
4. 3D Hubs: https://www.3dhubs.com/knowledge-base/3d-modeling-cad-software#introduction

5. Harvard business review: https://hbr.org/sponsored/2018/03/software-asset-management-a-new-defense-against-cybersecurity-threats
6. Microsoft SAM: https://www.microsoft.com/en-us/sam
7. World intellectual property organization: http://www.wipo.int/portal/en/
8. UfficioBrevetti.it: https://www.ufficiobrevetti.it/software/
9. Duke, Fernandez, What is Software Compliance, September 2016 https://www.linkedin.com/pulse/what-software-compliance-duke-fernandez/
10. PTC Legal agreements: https://www.ptc.com/-/media/Files/PDFs/legal-agreements/fy18/Clickwrap-Agreement-Apr-18.ashx?la=en&hash=22277BEADD549CDC3D19EA497308883A
11. BSA Global Software Survey June 2018: https://gss.bsa.org/wp-content/uploads/2018/05/2018_BSA_GSS_Report_en.pdf
12. Malwarebytes: https://www.malwarebytes.com/malware/
13. Revulytics: https://www.revulytics.com/
14. Parametric Technology Corporation: https://www.ptc.com/en/why-ptc
15. Bureau van Dijk, AIDA Database: https://www.bvdinfo.com/en-gb/our-products/data/national/aida
16. Rapporto sulla competitività dei settori produttivi: https://www.istat.it/it/archivio/212438
17. CDM Cloud Computing Shift: http://www.gruppocdm.it/blog/infor-customer-italia-la-strategia-per-la-trasformazione-digitale/