



LUT School of Business and Management

Bachelor's thesis, Business Administration

Financial Management

**The impact of digital publications on the sales of print books in Finland
Digitaalisten julkaisujen vaikutus painettujen kirjojen myyntiin Suomessa**

09.05.2020

Author: Emma Lankinen

Supervisor: Maija Hujala

TIIVISTELMÄ

Tekijä:	Emma Lankinen
Tutkielman nimi:	Digitaalisten julkaisujen vaikutus painettujen kirjojen myyntiin Suomessa
Akateeminen yksikkö:	LUT-kauppakorkeakoulu
Koulutusohjelma:	Kauppätieteet, Talousjohtaminen
Ohjaaja:	Maija Hujala
Hakusanat:	Kirjat, digitaaliset julkaisut, innovaatio, innovaatioiden diffuusio

Digitaalisten julkaisujen myynti on lähtenyt viime vuosina nopeaan kasvuun, kun suomalaiset ovat alkaneet omaksua digitaalisia julkaisuja erilaisten uusien palveluiden kautta. Tämän tutkielman tarkoituksena on selvittää, ovatko digitaaliset julkaisut syrjäyttämässä painetut kirjat Suomen kirjamarkkinoilta. Aineistona käytetään Suomen Kustannusyhdistyksen aikasarjoja painetuista kirjoista, sähkökirjoista, fyysisistä tallenteista, äänikirjoista ja ammatillisista digikirjastoista vuosilta 2008-2018. Tutkielman teoriaosuudessa tarkastellaan innovaatioiden luokittelutapoja ja etenkin innovaatioiden diffuusiota, jolla digitaalisten julkaisujen myynnin nopeaa kasvua pyritään selittämään. Tutkimus toteutetaan kvantitatiivisin menetelmin lineaarista regressioanalyysia hyödyntäen.

Tutkimuksen tulokset osoittavat, että Suomessa digitaalisten julkaisujen myynnillä on selvä negatiivinen vaikutus painettujen kirjojen myyntiin. Digitaalisten julkaisujen kokonaisymyynnin ja painettujen kirjojen kokonaisymyynnin välinen korrelaatio on tulosten mukaan myös negatiivinen. Painettuja kirjoja myydään kuitenkin vielä moninkertaisesti digitaalisia julkaisuja enemmän, joten vaikka vaikutus havaittiin negatiiviseksi, niin digitaaliset julkaisut eivät ole ainakaan vielä lähivuosina syrjäyttämässä painettuja kirjoja Suomen kirjamarkkinoilta.

ABSTRACT

Author: Emma Lankinen
Title: The impact of digital publications on the sales of print books in Finland
School: School of Business and Management
Degree programme: Business Administration, Financial Management
Supervisor: Maija Hujala
Keywords: Books, digital publications, innovation, diffusion of innovations

In the past few years, the sales of digital publications have started to increase rapidly, as Finns have begun to adopt digital publications through various new services. The purpose of this thesis is to investigate whether digital publications are going to replace print books from the Finnish book market. The data used is time series data from The Finnish Book Publishers Association, which covers print books, e-books, physical recordings, audiobooks, and professional digital libraries from the years 2008-2018. In the theoretical part of the thesis, the ways to categorize innovations are examined, as well as the diffusion of innovations, which aims to explain the rapid increase in the sales of digital publications. The research is carried out using quantitative methods, utilizing linear regression analysis.

The results of the study show that the sales of digital publications have a clear negative effect on the sales of print books in Finland. According to the results, the correlation between the total sales of digital publications and the total sales of print books is also negative. However, print books are sold many times more than digital publications, and therefore, although the effect was found to be negative, digital publications will not be replacing print books from the Finnish book market, at least in the next few years.

TABLE OF CONTENTS

1. Introduction.....	1
1.1 A peek at previous research	1
1.2 Research purpose, questions and limitations.....	2
1.3 Research methodology	4
1.4 Structure of research.....	4
2. Concepts and theory	5
2.1 E-books.....	5
2.2 Physical recordings, audiobooks and digital libraries	6
2.3 Distribution of digital publications	6
2.4 Ways to categorize innovations	7
2.5 Diffusion of innovations by Rogers (2003)	8
2.5.1 Innovation-decision process	9
2.5.2 Characteristics of the innovation	11
2.5.3 Adopter categories	12
2.5.4 S-shaped curve and bell curve	14
2.5.5 Cracks in the bell curve	15
2.5.6 Previous research on diffusion of innovations	16
3. Methodology.....	18
3.1 Description of data	18
3.2 Stationarity	21
3.3 Unit root tests	22
3.3.1 Dickey-Fuller test.....	22
3.3.2 Augmented Dickey-Fuller test	22
3.4 Cointegration.....	23
3.5 Regression analysis	23
4. Research results.....	25
4.1 Assumptions of regression analysis.....	25
4.2 Augmented Dickey-Fuller test results.....	27
4.3 Linear regression analysis	29
5. Discussion and conclusions.....	30
REFERENCES	32

APPENDICES

APPENDIX 1. Shapiro-Wilk test for normality results.

APPENDIX 2. Component-plus-residual plot.

APPENDIX 3. White's test for heteroskedasticity results.

APPENDIX 4. Breusch-Godfrey LM test results.

APPENDIX 5. Time series graphs for variables and the first differences of variables.

LIST OF TABLES

Table 1	Framework for defining innovation (Adapted from Henderson & Clark 1990, 12).	8
Table 2	Variables and their meanings.....	19
Table 3	Variables and their descriptive statistics.....	21
Table 4	Assumptions of linear regression (Adapted from Brooks 2014, 91).	24
Table 5	Descriptive statistics.....	26
Table 6	Augmented Dickey-Fuller test results.....	27
Table 7	Augmented Dickey-Fuller test results for the first differences.	29
Table 8	OLS results.	30

LIST OF FIGURES

Figure 1	Total sales of general literature without VAT (1000€), excluding comics (The Finnish Book Publishers Association 2018).	3
Figure 2	The Innovation-Decision Process (Adapted from Rogers 2003, 170).	9
Figure 3	The S-shaped curve of adoption and adopter categories (Adapted from Rogers 2003, 273, 281).	14
Figure 4	Chasm and cracks in the bell curve (Adapted from Moore 1999, 17).	15
Figure 5	Scatter graphs for the sales of digital publications and print books.	20

1. Introduction

Digital publications, such as e-books and audiobooks, have become more common in the last ten years and their popularity has shown great growth. Print books no longer take up space in a suitcase when numerous books can conveniently be stored on a mobile phone, laptop or tablet. People can literally carry an entire library with them everywhere they go. There is no need to travel to a library or bookstore to get a physical book when obtaining a book is just a few clicks away. Digital publications can be borrowed from the library or they can be purchased from bookstores or online services. As people have become more aware of them, their advertising has increased and they can be encountered on various channels, such as radio or television. Thus, it is commonly questioned whether new digital forms of reading are slowly replacing print books now that people are starting to adopt their use.

Some new services for digital publications entered the Finnish book market in the mid-2010s, for example BookBeat and Storytel in 2016 (Bonnier Press Release Feb. 8, 2017, Tietoa Storytelistä 2020). Therefore, it is fair to say that digital publications, especially e-books, are a rather new phenomenon in Finland. Previous research papers are mostly based on a time when these services were not yet available, and when the growth was slower than nowadays. Hence, it is important to carry out up-to-date research on the subject, for results may now differ from previous research due to these fresh actors.

1.1 A peek at previous research

The relationship between e-books and print books has received extensive attention among researchers. Especially the question of whether e-books cannibalize the sales of print books has been widely discussed. Bounie, Eang, Sirbu and Waelbroeck (2013) conducted a study on the topic in question by comparing the sales of best-selling print books and e-books on Amazon. The study concludes that bestselling titles by popular authors are most likely to suffer the negative impact of e-books on the sales of print books. During the study, two new segments called digital outsiders and print-preferred

books were found. Digital outsiders are described as books, which are successful as e-books but not as print books, and, on the contrary, print-preferred books are books which are successful as print books but not as e-books. The study also argues that e-books have a positive relationship with the market profitability of old print books.

Van der Velde and Ernst (2009) studied the same subject. They claim that the sales of e-books do not have a negative impact on the sales of print books. In fact, the sales of e-books sometimes even led to an increase in the sales of print books. They found in their study, that the pros of e-books are their findability and ease of use, whereas the cons are limited to the difficulty of reading from the screen. It is concluded, that print books are not yet being abandoned and they continue to stay on top.

As e-books are feared to negatively impact on print book sales, one solution has been to delay e-book releases. Thus, the effects of delaying e-book releases have also been a subject of research. For example, Chen, Hu and Smith (2018) investigated the matter in their study. According to their study, delaying e-book releases actually has a major negative impact on digital sales, while it has no significant impact on print book sales. It could therefore be concluded that delaying e-book releases is not worthwhile in terms of sales figures.

However, these studies only examine the effects of e-books, leaving audiobooks and other digital forms of reading outside. Country-specific differences may also exist within the studies. As Wilson (2014) claims, the rise of e-books is different in various countries and cultures, and each country experiences them at their own pace.

1.2 Research purpose, questions and limitations

This research examines the relationship between the sales of digital publications and the sales of print books. According to the yearly statistics of The Finnish Book Publishers Association (2018), the total sales of print books without VAT have been decreasing in the field of general literature (Figure 1). Figure 1 displays years on the horizontal axis and euros on the vertical axis. The orange line represents print books, the gray line digital publications and the blue line the combined number of print books and digital

publications. In 2013, the total sales of print books without VAT, excluding comics, were over 141,7 million euros, whereas in 2018, the sales had dropped to 114,8 million euros. During the same period, the total sales of digital publications had doubled from 14 million euros to over 28 million euros.

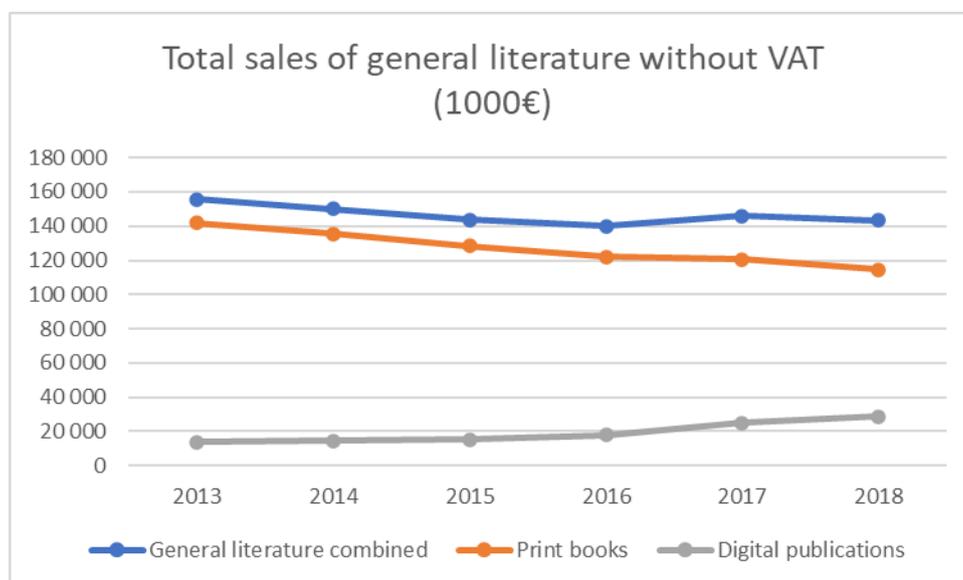


Figure 1 Total sales of general literature without VAT (1000€), excluding comics (The Finnish Book Publishers Association 2018).

The purpose of this thesis is to investigate whether or not digital publications are slowly replacing print books by examining their sales figures in Finland. Thus, the main research question reads as follows:

How do the sales of digital publications affect the sales of print books in Finland?

The main research question consists of sub-questions intended to provide support to the main research question. The sub-questions are:

1. *What kind of correlation exists between the sales of digital publications and the sales of print books?*

2. How much do the sales of digital publications affect the sales of print books in the field of fiction literature, children and youth literature, non-fiction literature and materials?

The research is geographically limited to Finland, for there is little research done in this area. Comics are excluded from the research due to missing data of digital publications. The focus of the research is on the sales of different types of books, and therefore loans from libraries are also excluded from the research. The period considered in the study is relatively short and therefore the number of samples is limited. This should be taken into account when evaluating the results of the study.

1.3 Research methodology

The data used is ready-made and collected from the yearly statistics of The Finnish Book Publishers Association. The research material consists of long term reports from the year 1994 to the year 2018, but since data from digital publications is not available from the outset, data that is more meaningful to the thesis is collected from recent years.

Due to the purpose of the study, the methods used are quantitative. The methods mainly focus on regression analysis in order to obtain answers to research questions. Unit root tests are also conducted, for the data used is time series data. Thus, the main methods used are Ordinary Least Squares method and the Augmented Dickey-Fuller test.

1.4 Structure of research

This study consists of five chapters. The study begins with an introduction that addresses previous research on the subject and outlines the research objectives as well as the research questions. Chapter 2 provides a theory for this research paper by explaining the necessary concepts along with some underlying factors behind the topic.

The theory mainly consists of diffusion of innovations, which also serves as a framework for this thesis. Chapter 3 covers the methodology and data used in the thesis. Research results are displayed and analyzed in Chapter 4. Summary, including conclusions of research results and recommendations for future research, is the last chapter of the thesis.

2. Concepts and theory

The concepts and theory relevant to the thesis are introduced in this chapter. Within this study, the term “digital publications” includes physical recordings, electronic audiobooks, electronic books and professional digital libraries. The term “print books” stands for physical books where the text is printed on paper. The diffusion of innovations is used as the framework for the thesis, for it gives a solid base for understanding the growth and adoption of digital publications. Rogers' (2003) theory of diffusion of innovations is widely accepted, and therefore most of the theory in this thesis is focused around it. The theory provides an idea of the elements that influence the diffusion process of innovations, in this case digital publications.

2.1 E-books

The concept of electronic books is not unambiguous. Electronic books, also known as e-books, digital books, eBooks and e-editions all carry the same meaning despite the different name. An electronic book or, in abbreviated form, an e-book, is an electronic text that is electronically accessible to all devices with a screen (Armstrong, Edwards & Lonsdale 2002). Another good definition is in Brown's (2001) article *Beyond print: reading digitally*, where e-book is described as “electronic version of a print book (published or soon to be published), which is downloaded for reading on portable devices, PDAs, PCs or laptops using proprietary "e-book reader" software”.

Christensen (2007, 22) divides technologies into two groups: sustaining and disruptive technologies. Sustaining technologies improve product performance and are therefore appreciated by mainstream customers. However, e-books can be seen as a disruptive technology (Luck, Lamp, Craig & Coldwell-Neilson 2016, Ronte 2001, Wilson 2014).

Disruptive technologies differ from sustaining technologies in providing lower performance in the short term. At first, the attributes of disruptive technologies may not reflect the values of the main customers, but their other attributes tend to attract brand new customers. These types of technologies are usually cheaper and smaller, and a common feature for describing them is ease of use, for they tend to be simpler. Disruptive technologies create new markets, and as a result from constant upgrading, they can also conquer the old markets. (Christensen 2007, 22-23) Christensen himself sees modular digital textbooks as a disruptive technology for ordinary textbooks (Christensen 2007, 33).

2.2 Physical recordings, audiobooks and digital libraries

Instead of reading books, it is possible to listen to them. A physical recording can be a multimedia publication, but it can also be an audiobook. The difference between an electronic audiobook and a physical recording is that electronic audiobooks are accessed via streaming services or downloaded from the internet, whereas physical recordings are only available in certain formats, such as CD or DVD. (The Finnish Book Publishers Association 2020) According to The Finnish Book Publishers Association (2020), a professional digital library is “a web-based entity, that includes substantial amount of other material and functions than access rights to single books or audiobooks”.

2.3 Distribution of digital publications

The distribution of digital publications is determined by their pricing methods. The Finnish Book Publishers Association (2020) categorizes these methods as time-based pricing, unit-based pricing and user number pricing. In time-based pricing, the consumer pays a certain amount to the publisher or the reseller and obtains the rights to use their services for a certain period of time. For example, if there is a monthly fee, the consumer will have access to the services for one month after the payment. In unit-based pricing, the reseller or the end customer pays for one piece of work at a time. Thus, the price is determined by the price of the piece being sold. In user number pricing, a

library or another party can buy a group license for their use. (The Finnish Book Publishers Association 2020)

2.4 Ways to categorize innovations

New ideas, practices and objects can be called innovations (Rogers 2003). Innovations can be divided into two groups according to how much change the consumer has to make when adopting them. If the consumer is required to change their current behavior, the innovation is discontinuous, whereas if no major change is needed, the innovation is continuous (Moore 1999). When it comes to discontinuous innovations, old knowledge and skills can no longer be utilized, but something new must be learned (Apilo, Taskinen & Salkari 2007, 25).

The division method where innovations are divided into incremental or radical, is the most common way to sort innovations. In this method, innovations are divided according to their novelty value. A change in a way of acting is not necessary for incremental innovations, unlike for radical innovations. Radical innovations are very close to the concept of breakthrough innovations in a situation where an innovation's technological development has been rapid. (Apilo et al. 2007, 23) Sometimes radical innovations are called revolutionary innovations, whereas continuous incremental innovations are called evolutionary innovations (Apilo et al. 2007, 23, 25). Rogers (2003, 426) views radical innovations as the cause of major changes and uncertainty. He also claims them to be synonyms for disruptive and discontinuous innovations. However, Christensen (2007, 21) argues, that previously mentioned division into sustaining and disruptive differs from incremental and radical. Thus, it can be assumed that the concepts are close to each other, but they are not quite synonymous.

One way of categorizing innovations is also the division into modular and architectural innovations. If the operating principles of the module can be changed in a way which allows the connections to the system remain unchanged, the innovation is modular. On the other hand, if the connections to the system change with the interfaces of the components, the innovation is architectural. (Apilo et al. 2007, 25-26) Henderson and Clark (1990) created a framework for defining innovations (Table 1). It divides the last

four innovations mentioned above into groups based on both changes in the core concepts and changes in the relationships between core concepts and components.

Table 1 Framework for defining innovation (Adapted from Henderson & Clark 1990, 12).

		Core concepts	
		Reinforced	Overtured
Linkages between Core Concepts and Components	Unchanged	Incremental Innovation	Modular Innovation
	Changed	Architectural Innovation	Radical Innovation

2.5 Diffusion of innovations by Rogers (2003)

The results of this thesis indicate whether digital publications, such as e-books, are slowly replacing print books. The process of adopting new innovations is essentially involved in the formation of sales figures. Of course, digital publications will not replace print books any time soon, if people are unaware of them or unable to adopt them for daily use. This is where the diffusion of innovations comes in. Rogers (2003, 11) defines diffusion in his book *Diffusion of Innovations* as “the process by which an innovation is communicated through certain channels over time among the members of a social system”. Therefore, the four elements of diffusion include the innovation itself, communication channels, time and a social system.

2.5.1 Innovation-decision process

Rogers (2003, 170) states, that the innovation-decision process proceeds in five steps; knowledge, persuasion, decision, implementation and confirmation (Figure 2). The figure describes how people first become aware of the innovation, such as e-books or audiobooks, and then they get better involved with it. After forming an attitude towards the innovation, they make the decision whether or not to adopt them. If the attitude and the decision are favorable, they implement the innovation and later seek confirmation that they did the right decision. Likewise, if the attitude and the decision are unfavorable, they reject it.

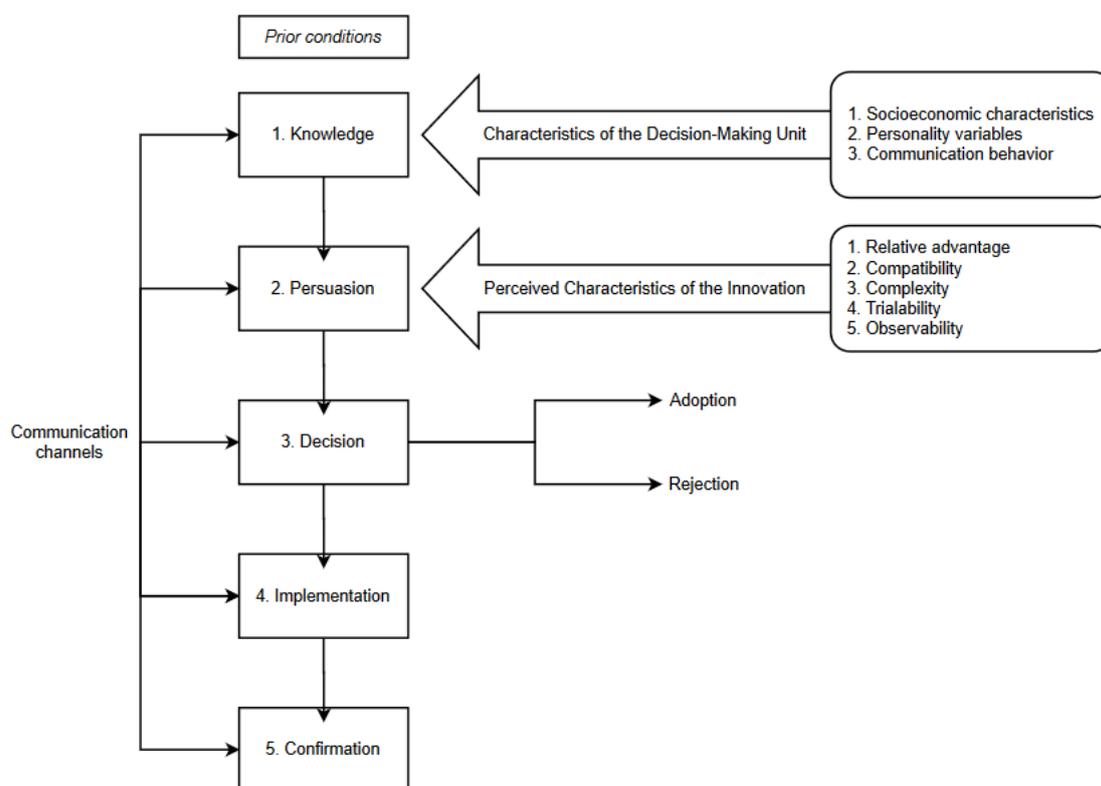


Figure 2 The Innovation-Decision Process (Adapted from Rogers 2003, 170).

The communication channels, which are different ways for information exchange, affect each step. Mass media and interpersonal channels are the main channels for the diffusion of innovations. Mass media includes channels such as radio and television,

whereas in interpersonal channels, information is transmitted face-to-face. Communication via the Internet has emerged alongside these two communication channels. (Rogers 2003, 18) Based on the research Suomi lukee (Taloustutkimus Ltd., The Booksellers' Association of Finland & The Finnish Book Publishers Association 2018), magazine reviews and the Internet were among the most important book buying stimuli for Finns in 2018.

According to Rogers (2003, 20-21), it takes time for diffusion to occur. Time can be seen in three ways as the element of diffusion. First, it is included in the innovation-decision process, where time is defined as the amount of time it takes for the individual to complete the whole process. Second, some members adopt the innovation at an early stage and others adopt it later. Therefore, it is the time of adoption compared to other members. Third, it is the rate of adoption; ergo, how many members of a group have accepted an innovation at a given time.

Rogers (2003, 23-27) states, that diffusion also occurs within a social system, where individuals, groups or even systems can be its units. These units share the same goal to solve a specific problem. In a social system, information on innovation is exchanged, which affects the pace of diffusion. In addition, social systems have a formal or informal structure. A social structure is made up of relationships between members, which may arise through a hierarchy. This is called a formal structure. Correspondingly, Rogers brings up an informal structure in which communication between members is the key. In addition to structures, individual roles, such as opinion leader and change agent, along with norms, which are accepted behavior patterns, influence how quickly innovations are adopted.

Rogers (2003, 28-29) classifies innovation-decisions into three types that occur within a social system. The first type he brings up is optional innovation-decision. This type reflects the decision a person makes without the opinions of others in the social system. Thus, decision-making is done independently. The second type that Rogers describes is collective innovation-decision. This decision-making is, as its name implies, collective. Ergo, the decision is made together with other members of the social system and therefore there is a consensus in the system. The third type Rogers mentions is

authority innovation-decision. In this type, a person in a high position makes the decision and the individual acts in accordance with that decision. This type is frequently the fastest way to adopt innovations.

Rogers (2003, 30-31) sees, that the consequences, ergo the changes that the individual experiences after adopting or rejecting an innovation, are also an essential part of how the social system affects diffusion. He mentions three ways of classifying consequences. The consequences may be desirable or undesirable, direct or indirect, or anticipated or unanticipated. Occasionally, the innovations that are expected to have desirable, direct and anticipated consequences are to be brought to light, rather than those that are undesirable, indirect or unanticipated.

2.5.2 Characteristics of the innovation

In addition to communication channels, time, and social system, there are five attributes that influence the speed at which an innovation is adopted (Rogers 2003, 15-16). These attributes influence especially in the second stage of the innovation-decision process (Figure 2). According to Rogers (2003, 221), relative advantage, compatibility, complexity, trialability and observability can explain more than half of the variance in the rate of adoption.

Rogers (2003, 230-258) describes the five attributes as follows. The first attribute is the greatness of the *relative advantage* of an innovation compared to the previous idea. For example, relative advantage can be economic when the price of the innovation is low. It can also be social if its adoption results in a better social status. The second attribute is the *compatibility* of an innovation with values, past experiences and needs. In the worst case, incompatibility with values can prevent the adoption of innovation. Past experiences reduce uncertainty, but occasionally adopters' experiences are negative, in which case past experiences may even lead to the rejection of innovation. The needs of adopters must also be taken into account. The better the needs are met, the faster the innovation will be adopted. The third attribute is the *complexity* of understanding and using an innovation. It can be generalized that the complexity of innovation slows down its rate of adoption and especially for completely new innovations,

complexity plays an important role in whether or not it is adopted. The fourth attribute is the *trialability* of an innovation. The trialability of innovations provides certainty and it also makes individuals adopt innovations faster. The last is the fifth attribute, the *observability* of the results of an innovation. As with trialability, innovations that are observable and easy to describe to others are being adopted more quickly.

The whole process ultimately leads the adopter to a decision. The innovation can either be adopted or rejected. It is possible, that the innovation is first adopted but rejected later due to, for example, the dissatisfaction of the adopter, in which case the phenomenon is called discontinuance. Of course, the situation can also be reversed, where innovation that has already been rejected is adopted later. (Rogers 2003, 21)

2.5.3 Adopter categories

Rogers (2003, 280-282) claims, that there are five adopter categories based on the time it takes for individuals to adopt innovations (Figure 3). These categories are innovators, early adopters, early majority, late majority and laggards. The five categories represent ideal types and therefore exceptions may and will occur.

First are the innovators. Innovators are the smallest group of the five, and according to Rogers (2003, 282) they are best described as being venturesome. Moore (1999, 30), however, calls innovators as technology enthusiasts. They love exploring new technology products, and innovations are a key part of their lives (Moore 1999, 12). The innovators bring new ideas to the social system and therefore their role in diffusion process is essential (Rogers 2003, 283).

Second, there are the early adopters who are also called visionaries (Moore 1999, 33). buy innovations based on intuition. Early adopters are mostly interested in gaining access to the benefits of the new product (Moore 1999, 12). Rogers (2003, 283) points out, that this group is known for speeding up the diffusion process. As the name suggests, early adopters acquire the innovation at an early stage and therefore later potential adopters will often seek advice from them concerning the innovation. Early

adopters' opinions are respected, and for that reason their acceptance of an innovation also reduces the uncertainty surrounding it.

Third, there are the early majority, which Moore (1999, 41) calls the pragmatists. He also claims, that together with the next two groups, the mainstream market is formed. Early majority are one of the major categories and they spend a great deal of time considering carefully before adopting an innovation. That feature makes them particularly practical about their buying decisions, unlike previous groups. (Moore 1999, 12-13) Rogers (2003, 283-284) sees early majority mainly as a link relevant to the diffusion process, for they form a link between the groups that are extremely early and somewhat late.

The fourth group is called the late majority and with the early majority, they are the other major category. The group is also seen as conservatives (Moore, 46). Moore (1999, 13) describes them as skeptical towards innovations for they are not sure of their innovation handling skills. According to Rogers (2003, 284), late majority adopts innovations only after they feel certain about them and this does not happen until majority of others have already adopted the innovation.

The last group is called laggards, or skeptics (Moore 1999, 54), for they are the last group to adopt innovations. Laggards even prefer avoiding new technologies (Moore 1999, 13). In the same way with the late majority, Rogers (2003, 284-285) also mentions laggards being skeptical towards innovations, for they frequently base their actions on past behavior. Laggards do not adopt an innovation until they are absolutely certain of its success. Therefore, the decision-making process for them tend to be longer than for other groups.

Rogers (2003, 298-299) claims, that the five adopter categories have differences in socioeconomic status, personality variables and communication behavior. For example, there are generalizations that earlier adopters' socioeconomic status tends to be higher and their attitudes more favorable. Earlier adopters are more active socially and they have more exposure to different channels. These differences naturally affect the innovation-decision process.

2.5.4 S-shaped curve and bell curve

Rogers (2003, 23) brings up the S-shaped curve, which presents the rate of adoption over time (Figure 3). The horizontal axis displays time and the vertical axis the number of adopters. If an innovation is rapidly gaining ground, the S-shaped curve may be steeper compared to an innovation that is slowly being adopted. The growth of digital publications has been enormous when the focus is on percentages, but the increase will likely adapt to follow the S-shaped curve. However, the total sales of digital publications are still minor compared to the sales of print books.

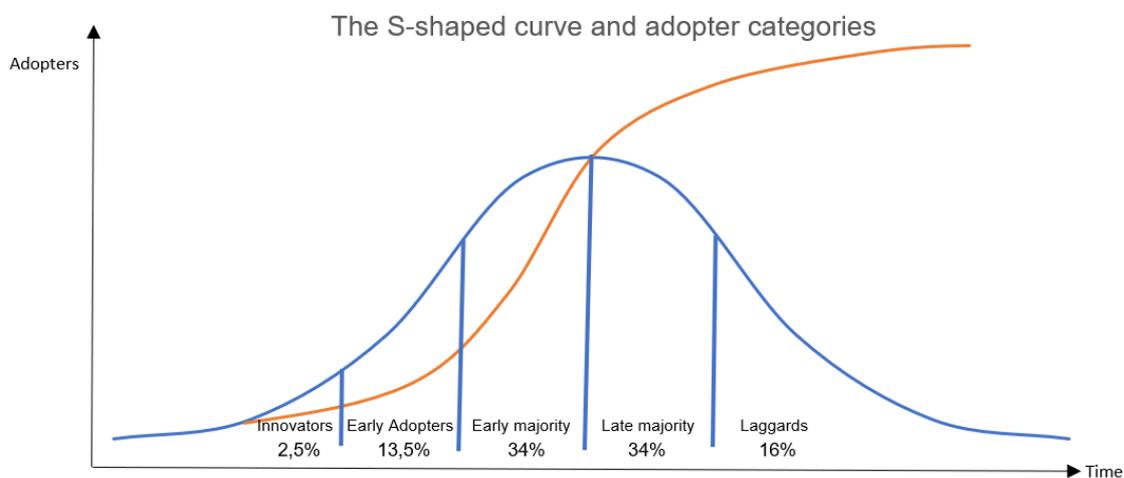


Figure 3 The S-shaped curve of adoption and adopter categories (Adapted from Rogers 2003, 273, 281).

The bell curve is formed by the five adopter categories. According to Rogers (2003, 280-281), the categories are divided by the mean time of adoption and the standard deviation. The vertical line between early majority and late majority represents the mean time of adoption, while the next vertical line on the right represents the mean plus one standard deviation. The next vertical line on the left, on the other hand, represents the mean minus one standard deviation and the next vertical line after that represents the mean minus two standard deviation (Figure 3). The innovators include only 2,5 percent of the entire system, early adopters include 13,5 percent, early majority 34 percent as well as late majority, and laggards 16 percent.

2.5.5 Cracks in the bell curve

Moore (1999, 16) challenges the original model of the bell curve and implies that there are cracks between the five adopter categories (Figure 4). The cracks reflect the differences between the categories concerning adoption of innovations. Due to these cracks, innovations should be sold to each category using different techniques. (Moore 1999, 16)

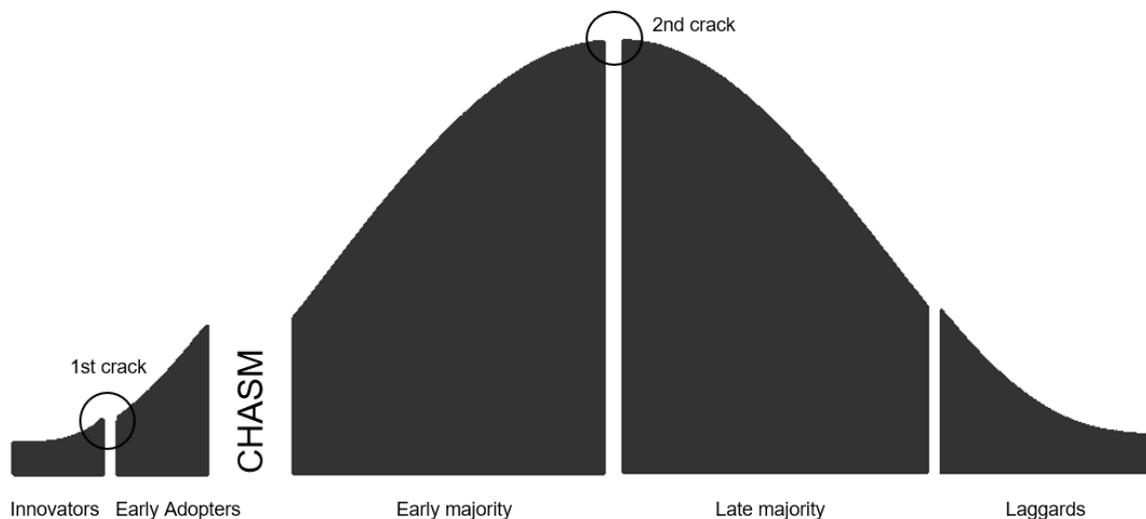


Figure 4 Chasm and cracks in the bell curve (Adapted from Moore 1999, 17).

Moore (1999, 17) claims, that the cracks between the innovators and the early majority, as well as between the early majority and late majority, are only minor cracks compared to a chasm between the early adopters and early majority. Innovators and early majority form the early market, whereas early majority, late majority and laggards form the mainstream market (Moore 1999, 30, 41). The chasm therefore exists between two markets. Moore (1999, 19-20) emphasizes, that for the continuation of diffusion, it is crucial to cross this major chasm. Thus, the main focus should be on it. The chasm is formed by different expectations of the two categories. Early adopters are expecting something completely different from the old, and they are ready to withstand even the cons of the innovation, such as bugs of early versions. Early majority, on the other

hand, are expecting something that improves productivity, but does not differ much from the old. Unlike the early adopters, they do not expect to experience the cons of the innovation. Thus, they want it to work properly right from the start.

Moore (1999) outlines four steps to cross the major chasm. First, he advises on segmenting the customers and focusing on a niche market (Moore 1999, 89). The next step is understanding the principles of the whole product concept and organizing the marketplace accordingly (Moore 1999, 108). The third step includes understanding competition as a whole, which essentially involves creating competition and positioning the company with the goal of making the product easier to buy (Moore 1999, 132, 148). The last step includes two important decisions. One is choosing the right channel for distribution and the other is pricing (Moore 1999, 163).

2.5.6 Previous research on diffusion of innovations

The diffusion of innovations is a well-researched topic in many different areas. In relation to e-books, Raynard (2017) has done a study on how Rogers' theory can be useful in marketing academic e-books and developing strategies to increase the number of users. According to the study, the early majority has begun to adopt academic e-books. The study recommends that the focus of marketing and developing strategies should be on how the two different formats can complement each other. Customers' demographic and psychographic characteristics should be studied first to increase the number of academic e-book users.

Martin and Quan-Haase (2013) conducted a study on the adoption of e-books by academic historians. They base their study on Rogers' innovation-decision process. According to the study, the confusion associated with defining e-books decreases the speed of diffusion process. Another particularly interesting finding in their study is that the two stages of the innovation-decision process, the knowledge and the persuasion stage, are cyclical. This means that instead of going straight from the first stage to the next, historians actually jump back and forth between the two; they find new information, they test it and they adopt it, until they find new information again. The study concludes that the use of e-books will not replace the use of print books and vice versa.

Many other researchers came to a similar conclusion as Martin and Quan-Haase (2013) about e-books not replacing print books (Walton 2014, Aharony & Bar-Ilan 2018, Bergström & Höglund 2018). Walton (2014) examined, through the diffusion of innovations, which factors lead undergraduate students to choose an e-book instead of a print book. The study shows that students chose an e-book when it was the only option left, but if a print book was available, students preferred to choose it instead. Walton argues that the adoption of e-books is positively influenced not only by their convenience, but also by leisure reading, conducting research and forced adoption. Aharony and Bar-Ilan (2018) studied reading preferences of Information Science students and how different variables affect them. In addition to the argument about the popularity of print materials, they argue that relative advantage and comprehension affect the adoption of electronic materials. Bergström and Höglund (2018), on the other hand, studied the use of e-books by the Swedish population. According to their study, the characteristics of e-book users seem to correspond to the group mentioned by Rogers (2003), early adopters. Their results show that attitudes have a significant impact on e-book adoption, although most still tend to prefer print books.

The diffusion of innovations is, of course, not only limited to books, but the subject spans many themes and fields. For example, Sääksjärvi and Hellén (2019) conducted a study regarding innovators and early adopters. They conclude that the two adopter categories can actually predict the reactions of the mainstream market, concerning new products and ideas. This information can be especially useful for companies to get their ideas implemented. Girardi and Chiagouris (2018), on the other hand, found in their study that the gap between early adopters and the mainstream market would be widening due to technological advancement. They claim that the diffusion of innovations is evolving over time and therefore, changes must also be made in communication plans in order to cross that gap. Jahanmir and Lages (2015) also studied one part of the adopter categories, laggards, for four years and developed a method called the Lag-User Method, which explores how laggards influence new product development. According to their study, when laggards are taught this method, they can actually become developers of new innovations. In addition, the study claims that using the lag-user method has a positive effect on laggards' perception. This positive effect means that laggards not only understand innovations better, but they are also able to see their

own skills in inventing new products and believe in those skills. Furthermore, they realize that everyone can learn to innovate, even them. It is concluded that using the right method, both the chasm mentioned by Moore (1999) and the obstacles related to the adoption of innovations can therefore be overcome.

3. Methodology

This section describes the methods used in the study. First, the data used in the study and the criteria for selecting the variables are introduced. The names and definitions of the variables are also reviewed. After describing the data, the essential theory behind the methods used is presented. The chapter mainly focuses on stationarity, unit root tests and linear regression analysis.

3.1 Description of data

The data used in the thesis consists of data collected from the annual statistics of The Finnish Book Publishers Association. The data used is from 2008-2018 and it covers the sales figures of fiction literature, non-fiction literature, children and youth literature, as well as the sales figures of materials such as textbooks. The data has been transferred from the annual spreadsheets to Excel in order to obtain a desired format for use.

The data had to be modified by completely excluding comics. Therefore, the definition of general literature used in the study refers only to the sum of fiction, non-fiction, and children and youth literature. The same goes, of course, for the definition of total sales used in the study, which includes not only general literature but also materials. All sales figures are presented in thousands of euros.

The data includes thirteen different variables and each variable has eleven observations. Therefore, the data set is particularly small, and it should be taken into account when evaluating the results of the study. Because the observations are measured from

consecutive times, the data used is time series data. Thus, the variable *Years* covers the years from 2008 to 2018 and it is selected as the time variable when performing tests with Stata.

The purpose of this study is to see the impact of the sales of digital publications on the sales of print books. Therefore, the dependent variable is the total sales of print books, which is called *PB_TOT*, whereas the independent variable is the total sales of digital publications, *DP_TOT*. A natural logarithm is taken from the variable *DP_TOT*, which is called *In_DP_TOT*. The prefix *DP* refers generally to digital publications, while the prefix *PB* refers to print books. Thus, *PB_Fic* stands for the sales figures of fiction literature, *PB_Child* stands for the sales figures of children and youth literature, *PB_Nonfic* stands for the sales figures of nonfiction literature and *PB_Mat* stands for the sales figures of materials. The suffixes are the same for digital books. *DP_GLTOT* and *PB_GLTOT* refer to the sales figures of general literature for digital publications and print books. The names and meanings of the variables used are summarized in Table 2.

Table 2 Variables and their meanings.

Variable	Meaning	Variable	Meaning
<i>Years</i>	Years from 2008-2018	<i>PB_Fic</i> , <i>DP_Fic</i>	Print sales of fiction literature, digital sales of fiction literature
<i>PB_TOT</i> , <i>DP_TOT</i>	Total sales of print books, total sales of digital publications	<i>PB_Child</i> , <i>DP_Child</i>	Print sales of children and youth literature, digital sales of children and youth literature
<i>In_DP_TOT</i>	Natural logarithm of total sales of digital publications	<i>PB_Nonfic</i> , <i>DP_Nonfic</i>	Print sales of nonfiction literature, digital sales of nonfiction literature
<i>PB_GLTOT</i> , <i>DP_GLTOT</i>	Total print sales of general literature, total digital sales of general literature	<i>PB_Mat</i> , <i>DP_Mat</i>	Print sales of materials, digital sales of materials

The first sub-question of the study is “*What kind of correlation exists between the sales of digital publications and the sales of print books?*”. The correlation between two variables indicates the degree of dependence between them (Hill, Griffiths & Lim 2012, 31). According to the research Suomi lukee (Taloustutkimus Ltd. et al. 2018), people who usually buy a lot of print books in Finland have replaced some of them with e-books, while people who have never bought print books have actually started reading

books in the form of e-books. The research therefore suggests that the correlation between digital publications and print books could be negative. The correlation between the variables is presented graphically in Figure 5 using scatter graphs.

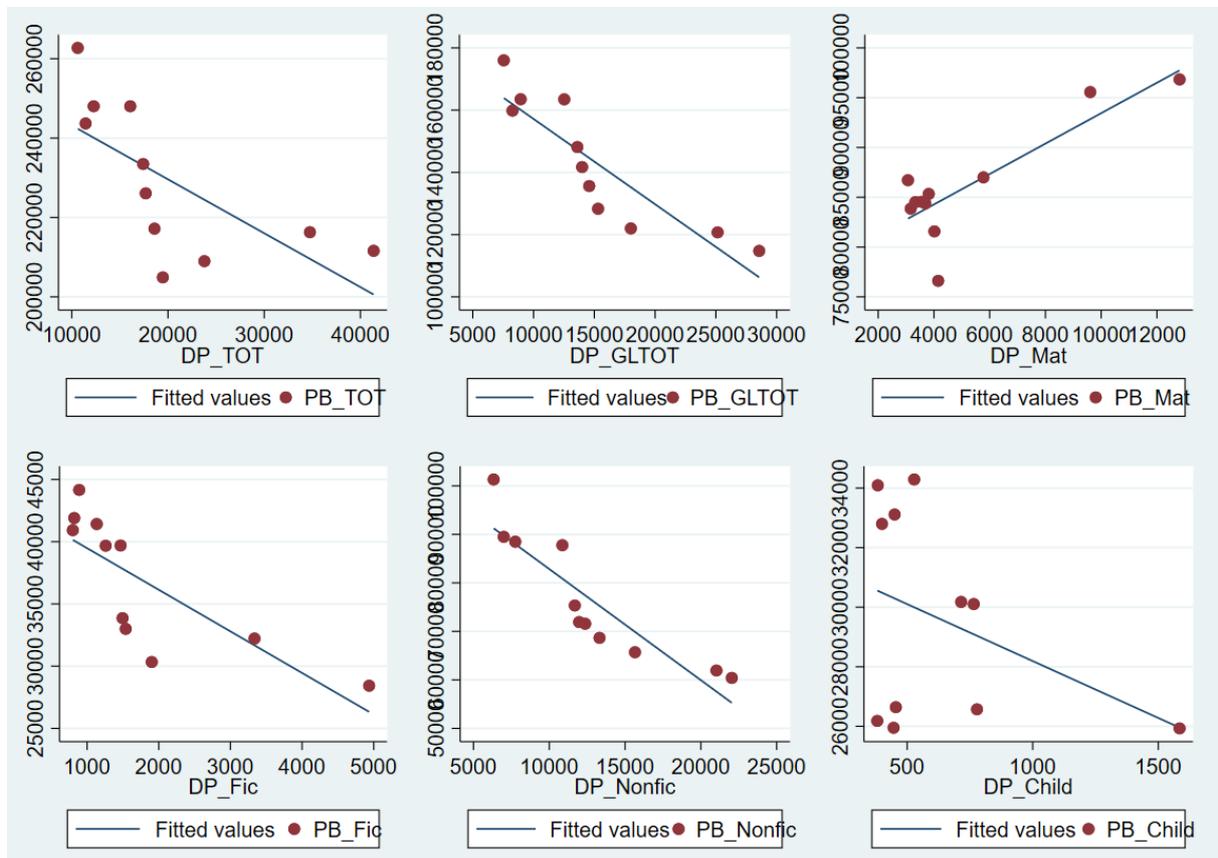


Figure 5 Scatter graphs for the sales of digital publications and print books.

When examining the scatter graphs, only the digital sales of materials seem to have a positive correlation with the print sales of materials, while for the other variables the correlation is negative. The most unclear correlation is between the digital sales of children and youth literature and the print sales of children and youth literature, where there is no obvious regularity, and therefore the relationship between them is weak. Sometimes outliers can cause a biased correlation, even if there were no actual correlation between the two variables (Metsämuuronen 2005, 351). In the case of children and youth literature, that is a possibility to be considered. The descriptive statistics of the variables are also displayed in Table 3.

Table 3 Variables and their descriptive statistics.

Variable	Mean	Std. Dev.	Min	Max	Obs.
DP_Fic	1780.64	1264.81	801	4937	11
PB_Fic	36874.18	5395.96	28431	44161	11
DP_Child	626.00	352.48	381	1585	11
PB_Child	29622.91	3488.51	25930	34289	11
DP_Nonfic	12729.09	5179.55	6330	22041	11
PB_Nonfic	76600.09	13231.28	60406	101314	11
DP_Mat	5182.46	3150.57	3069	12813	11
PB_Mat	86085.18	5745.06	76608	96834	11

3.2 Stationarity

Because the data used in the study is time series data, its stationarity must be tested to obtain reliable results. A time series can be said to be stationary when its mean (1) and variance (2) are constant and therefore they do not vary across the study period, and the covariance (3) between two observations does not depend on the actual time t , but on the length of time s between two observations (Hill et al. 2012, 476). Stationarity can affect the behavior of a time series over different time periods, and non-stationarity can lead to spurious regressions. This means that there may appear to be a strong relationship between two variables, even though the variables do not have any real impact on each other. (Brooks 2014, 353-354)

$$E(y_t) = \mu \quad (1)$$

$$var(y_t) = \sigma^2 \quad (2)$$

$$cov(y_t, y_{t+s}) = cov(y_t, y_{t-s}) = \gamma_s \quad (3)$$

3.3 Unit root tests

Unit root tests can confirm the stationarity of the time series (Hill et al. 2012, 484). They measure the number of differences needed to make a non-stationary time series stationary. $I(0)$ series represents a stationary time series while $I(1)$ series represents a non-stationary time series which has a single unit root, meaning, that stationarity is achieved by differencing the time series once. Differencing is common especially in economic time series. (Brooks 2014, 360) The best-known unit root test is the Dickey-Fuller test, from which the Augmented Dickey-Fuller test is derived.

3.3.1 Dickey-Fuller test

The properties of a time series must always be observed. Thus, there are three types of equations which differ by their constant and trend. The first Dickey-Fuller test (4) is a random walk model with no constant and no trend, where Δ is the first difference of y_t , γ is the coefficient and u_t is the error term. In the second test (5), a constant term α is added to the equation. The third test (6) includes both, a constant α and a trend λt . The null hypothesis tested is $\gamma=0$ and the alternate hypothesis is $\gamma<0$. If the null hypothesis is rejected, the time series can be said to be stationary. (Hill et al. 2012, 484-485)

$$\text{i. } \Delta y_t = \gamma y_{t-1} + u_t \quad (4)$$

$$\text{ii. } \Delta y_t = \alpha + \gamma y_{t-1} + u_t \quad (5)$$

$$\text{iii. } \Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + u_t \quad (6)$$

3.3.2 Augmented Dickey-Fuller test

Augmented Dickey-Fuller test is used in this study. The main idea behind it is to use p lags of the dependent variable and examine the existence of a unit root while taking autocorrelation into account (Brooks 2014, 363). If there were autocorrelation, the values preceding a variable could affect its value, which may lead to errors (Heikkilä 2014,

194). The test is described in the equation 7. The null hypothesis is the same as in Dickey-Fuller tests. If the null hypothesis is not rejected, the time series is non-stationary, and it has a unit root. If the null hypothesis is rejected, the time series is stationary. (Hill et al. 2012, 486)

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + u_t, \quad (7)$$

where

Δy_t = dependent variable

ψ = parameter tested

p = number of lag terms

u_t = error term

3.4 Cointegration

As said before, the use of non-stationary time series can lead to spurious regressions, and therefore their use should be avoided. However, if the time series turns out to be non-stationary and the variables are still to be used, there is a way to save the situation to avoid spurious regressions. This exception applies if the variables are cointegrated, ergo, the variables that are at least $I(1)$ do not drift apart in the long run (Brooks 2014, 374). Testing the stationarity of the errors determines whether the variables are cointegrated or not (Hill et al. 2012, 488). Brooks (2014, 377) mentions three different methods for testing cointegration. The first is a two-step Engle-Granger test, the second is a three-step Engle-Yoo test and the third is Johansen test.

3.5 Regression analysis

Regression analysis is one of the most widely used methods when trying to model reality in a mathematical form (Metsämuuronen 2005, 659). It is mostly used to study the relationship between two or more independent variables x_k and the dependent variable y . The classical linear regression model is displayed in equation 8.

$$y_t = \alpha + \beta x_t + u_t, \quad (8)$$

where

y_t = dependent variable

x_t = independent variable

α = intercept

β = slope coefficient

u_t = random error term

The most frequently used method in a regression analysis is the Ordinary Least Squares method, OLS. (Brooks 2014, 75-78) The idea behind OLS is to minimize the sum of squared residuals (Hill et al. 2012, 51). There are certain assumptions that must be met in order for the simple regression analysis to be performed reliably. These assumptions are listed in Table 4. However, if the assumptions required by OLS are not met, it is also possible to use other methods instead to get reliable results. For example, the Generalized Least Squares method, GLS, is an alternative way to minimize the weighted sum of squared residuals (Brooks 2014, 186).

Table 4 Assumptions of linear regression (Adapted from Brooks 2014, 91).

Assumption	Interpretation
1. $E(u_t) = 0$	The errors have zero mean
2. $var(u_t) = \sigma^2 < \infty$	The variance of the errors is constant and finite over all values of x_t
3. $cov(u_i, u_j) = 0$	The errors are linearly independent of one another
4. $cov(u_t, x_t) = 0$	There is no relationship between the error and corresponding x variate
5. $u_t \sim N(0, \sigma^2)$	The errors are normally distributed

4. Research results

In this section, the results of the study are viewed. First, the assumptions of the regression analysis are checked. Based on those results, the method for regression analysis is selected. Unit root tests are then performed and analyzed to determine possible cointegration between the variables. Lastly, the regression analysis is conducted, and the results obtained are analyzed.

4.1 Assumptions of regression analysis

The answer to the second sub-question “*How much do the sales of digital publications affect the sales of print books in the field of fiction literature, children and youth literature, non-fiction literature and materials?*” was investigated with a regression analysis. The total sales of print and digital books were chosen as variables in order to observe all aspects of the sub-question. Prior to the regression analysis itself, the fulfillment of the assumptions had to be checked in order to choose the right method for regression analysis. Table 5 shows the descriptive statistics for the total sales of print books and digital publications. In order to examine the normal distribution of the data, attention must be paid to skewness and kurtosis. Skewness and kurtosis both indicate the shape of the distribution. If the data is normally distributed, their values should be zero. If the skewness is negative, the data is distributed more to the right, and if it is positive, the data is distributed more to the left. A positive kurtosis, on the other hand, indicates a sharp peak and a negative kurtosis a flat peak. (Brooks 2014, 66-67) The table shows that the skewness of total sales of print books is close to zero and the kurtosis is also at an acceptable level. The values of the total sales of digital publications, however, could be better. Therefore, a natural logarithm was taken from the independent variable to correct the values. The skewness decreased from 1.17 to .54 and the kurtosis decreased from 3.25 to 2.43.

Table 5 Descriptive statistics.

Variable	Mean	Skewness	Kurtosis	Std. Dev.	Min	Max	Obs.
PB_TOT	229182.40	.33	1.79	19209.16	204909	262726	11
DP_TOT	20318.18	1.17	3.25	9686.55	10621	41376	11

The Shapiro-Wilk test for normality was conducted in order to assure the normal distribution of the series. With 5% significance level, the null hypothesis for normal distribution was not rejected for the total sales of print books and for the logarithmically modified total sales of digital publications. The Shapiro-Wilk test was also conducted to the residuals to determine whether they are normally distributed. As with the variables, the null hypothesis for normal distribution was not rejected with 5% significance level. More detailed test results can be found in Appendix 1.

The linearity of the variables was examined graphically using the component-plus-residual plot. The graph is shown in Appendix 2. Based on the graph, there appears to be a clear linear relationship between the independent variable and the dependent variable.

The variance of the errors should be constant. In other words, they must be homoscedastic. Heteroskedasticity is the opposite of homoskedasticity, ergo, the variance of the errors is not constant. Heteroskedasticity should not occur when using OLS, for it can lead to wrong standard errors. (Brooks 2014, 181, 185) The assumption of homoskedasticity was tested using White's test for heteroskedasticity. The results are displayed in Appendix 3. As can be seen from the results, the null hypothesis of homoskedasticity is not rejected at the 5% significance level. Thus, the assumption of homoskedasticity is also met.

One important assumption especially for a time series data is that there is no autocorrelation. This means that the errors must not correlate with each other. (Brooks 2014, 188) Possible autocorrelation was tested with the Breusch-Godfrey serial correlation LM test, where the null hypothesis is no serial correlation. The detailed results are displayed in Appendix 4. At the 5% significance level, the null hypothesis is not re-

jected. Thus, there is no autocorrelation. As with previous assumptions, this assumption also remains valid. Thus, all assumptions were met and therefore OLS was chosen as the method for regression analysis.

4.2 Augmented Dickey-Fuller test results

The Augmented Dickey-Fuller test was performed to ensure the stationarity of the time series. Before performing the actual test, however, the shape of the time series was examined using graphs. As can be noticed from the graphs (Appendix 5), a clear linear trend is visible in both time series. The horizontal axis represents years and the vertical axis represents the sales figures. Without the natural logarithm, the trend in the total sales of digital publications would have been exponential. However, with the natural logarithm, it was corrected to be linear.

Next, the number of lagged differences for both variables was examined. A modified Dickey-Fuller t test for a unit root was used to determine them. In both cases, the number of lags with a trend equation was found to be 1. Thus, the Augmented Dickey-Fuller test was performed with a trend equation and one lag. The results for both variables are displayed in Table 6 and they are rounded to two decimal places. As mentioned before, the null hypothesis for the Augmented Dickey-Fuller test is non-stationarity and that it has a unit root. For both variables, the null hypothesis can not be rejected at any level of significance. Therefore, it must be stated that the time series are non-stationary.

Table 6 Augmented Dickey-Fuller test results.

In_DP_TOT		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-2.47	.34
Test critical values:	1% level	-4.38	
	5% level	-3.60	
	10% level	-3.24	

PB_TOT		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-1.81	.70
Test critical values:	1% level	-4.38	
	5% level	-3.60	
	10% level	-3.24	

Even if the time series are non-stationary, their differences may be stationary. Thus, the first differences of the time series are examined next. The graphs of the first differences for both the total sales of print books and the total logarithmic sales of digital publications are displayed in Appendix 5. Now, a clear linear trend is no longer visible, with graphs varying on either side of either zero or some other constant.

As before, the number of lagged differences for both variables was examined using a modified Dickey-Fuller t test for a unit root. The number of lags with an equation of no trend was found to be 1. This time, two test equations were used. One was an equation without a constant and the other was an equation with a constant. This was due to the fact that for total logarithmic sales of digital publications, it is clear that the graph varies around a constant when taking the first difference, but for total sales of print books, the graph of the first difference varies around either zero or some other constant. Thus, the Augmented Dickey-Fuller test was performed with both equations and one lag. The results are displayed in Table 7. If an equation without a constant is used, the test statistics obtained are greater than any critical value. If, on the other hand, an equation with a constant is used, the p -values obtained are greater than the 5% significance level. Therefore, the null hypothesis for non-stationarity and a unit root can not be rejected for either variable, regardless of whether the equation used was with or without a constant. Time series remain non-stationary.

Table 7 Augmented Dickey-Fuller test results for the first differences.

D1.ln_DP_TOT	no constant	with constant	
	t-Statistic	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-.65	-2.07	.26
Test critical values: 1% level	-2.66	-3.75	
5% level	-1.95	-3.00	
10%level	-1.60	-2.63	

D1.PB_TOT	no constant	with constant	
	t-Statistic	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-1.52	-2.53	.11
Test critical values: 1% level	-2.66	-3.75	
5% level	-1.95	-3.00	
10%level	-1.60	-2.63	

4.3 Linear regression analysis

If the second differences had proved to be stationary, the study would have progressed to co-integration tests. Attempts were made to take the second differences from the time series, but due to the shortness of the time series, they could no longer be taken. Therefore, non-stationary time series must be used, although there is a risk of spurious regressions. This is good to keep in mind when interpreting the results of the regression analysis.

Ordinary Least Squares method was used instead of Generalized Least Squares method, for all of the assumptions of OLS were met. Table 8 shows the results of the linear regression. The regression model is statistically significant at the 5% significance level ($p < .01$) with an F-value of 14.42. The independent variable ln_DP_TOT is also statistically significant, for its p-value is lower than the used significance level of 5% ($p < .01$). The independent variable can thus be used to explain changes in the dependent variable. The table shows that as the total logarithmic sales of digital publications increase by one unit, the total sales of printed books decrease by 35108.93 units. R^2 value was relatively high, 62% rounded. This means that variations in the independent variable can explain up to 62% of the variations in the dependent variable. However,

when looking at the figures, it is good to recall that the time series are non-stationary and therefore do not necessarily give as reliable results as stationary time series would.

Table 8 OLS results.

DEPENDENT VARIABLE		PB_TOT	
Independent Variable	Parameter Estimate	t-value	p-value
In_DP_TOT	-35108.93***	-3.80	<.01
Model fit			
F-value (df 1, 9)	14.42***	R-Square	.62

5. Discussion and conclusions

The purpose of this thesis was to investigate whether digital publications are replacing print books by examining their sales figures in Finland. The study was conducted using quantitative methods, utilizing unit root tests and linear regression analysis. The data used was collected from the yearly statistics of The Finnish Book Publishers Association and it consisted of the years 2008-2018. Comics were excluded from the study.

When evaluating the results of the study, it is good to keep in mind that regression analysis had to be performed on non-stationary time series, possibly due to the short observation period. Thus, spurious regressions could occur, which negatively affects the reliability of the results. It is also important to note that foreign research tend to focus mainly on e-books, whereas this study included all forms of digital publications.

The main research question of the study was *"How do the sales of digital publications affect the sales of print books in Finland?"* The results showed that the sales of digital publications have a negative impact on the sales of print books in Finland. Regression analysis indicated that when the total logarithmic sales of digital publications increase by one unit, the total sales of printed books decrease by over 35108 units. It was also discovered that there is mainly a negative correlation between the sales of digital publications and the sales of print books. The results are in contrast to those obtained by

Van der Velde and Ernst (2009), whose study shows that e-books do not have a negative impact on print books. However, according to the annual statistics of The Finnish Book Publishers Association, print books are still clear leaders in the book market, just as Van der Velde and Ernst state in their study. Many other researchers have come to the same conclusion as well (Martin and Quan-Haase 2013, Walton 2014, Aharony & Bar-Ilan 2018, Bergström & Höglund 2018). Thus, although the sales of digital publications have increased rapidly especially in recent years, they will not immediately replace print books. This may be because most people have grown up with print books and therefore it is still difficult to fully adopt new technology. It is also good to reiterate Wilson's (2014) words that every country experiences the adoption of e-books at its own pace. This means that studies conducted in different countries can give quite different results. The sales of digital publications have hardly reached their peak yet, and therefore it is possible that somewhere in the future they will become even more popular than print books, especially if their growth continues along the model of Rogers' (2003) S-shaped curve. But as print books are by far the most popular book format, it is worth investing in their production, at least until when, as, and if the sales of digital publications meet the sales of print books.

As the Finnish book market has not been studied that much, there are certainly enough topics for further research. Further research could for example be conducted on the same subject again in a few years, for more data would be available on digital publications and, therefore, the results would be more reliable. Further research could also be conducted qualitatively on the correlations between digital publications and print books. It would be interesting to discover why the correlation between the total sales of digital materials and print materials is the only positive correlation among the sales of digital publications and the sales of print books.

REFERENCES

- Aharony, N. & Bar-Ilan, J. (2018) Students' academic reading preferences: An exploratory study. *Journal of Librarianship and Information Science* 50, 1, 3-13.
- Apilo, T., Taskinen, T. & Salkari, I. (2007) *Johda innovaatioita*. 1st ed. Helsinki: Talentum.
- Armstrong, C., Edwards, L. & Lonsdale, R. (2002) Virtually there? E-books in UK academic libraries. *Program* 36, 4, 216-227.
- Bergström, A. & Höglund, L. (2018) E-books: In the shadow of print. *Convergence*.
- Bonnier. (2017) Bonnier Press Release Feb. 8, 2017: 2016: Shifting With an Eye to Consumer Revenues [online document]. [Accessed 12.2.2020]. Available at <https://www.bonnier.com/en/pressreleases/bonnier-2016-shifting-with-an-eye-to-consumer-revenues>
- Bounie, D., Eang, B., Sirbu, M. & Waelbroeck, P. (2013) Superstars and outsiders in online markets: An empirical analysis of electronic books. *Electronic Commerce Research and Applications* 12, 1, 52-59.
- Brooks, C. (2014) *Introductory Econometrics for Finance*. 3th ed. Cambridge: Cambridge University Press.
- Brown, G. J. (2001) Beyond print: Reading digitally. *Library Hi Tech* 19, 4, 390-399.
- Chen, H., Hu, Y. & Smith, M. (2018) The Impact of E-book Distribution on Print Sales: Analysis of a Natural Experiment. *Management Science* 65, 1, 19-31.
- Christensen, C. (2007) *Edelläkävijän dilemma*. Translated by H. Nylund. Helsinki: Talentum.

Girardi, P. & Chiagouris, L. (2018) The Digital Marketplace: Early Adopters Have Changed. *Journal of Marketing Development and Competitiveness* 12, 1, 84-95.

Heikkilä, T. (2014) Tilastollinen tutkimus. 9th new ed. Helsinki: Edita.

Henderson, R. & Clark, K. (1990) ARCHITECTURAL INNOVATION - THE RECONFIGURATION OF EXISTING PRODUCT TECHNOLOGIES AND THE FAILURE OF ESTABLISHED FIRMS. *Administrative Science Quarterly* 35, 1, 9-30.

Hill, R. C., Griffiths, W. E. & Lim, G. C. (2012) Principles of econometrics. 4th ed. Hoboken, NJ: John Wiley & Sons.

Jahanmir, S. & Lages, L. (2015) The Lag-User Method: Using laggards as a source of innovative ideas. *Journal Of Engineering And Technology Management* 37, 65-77.

Luck, S. E., Lamp, J. W., Craig, A. & Coldwell-Neilson, J. (2016) The book: Production and participation. *Library Review* 65, 1/2, 2-19.

Martin, K. & Quan-Haase, A. (2013) Are e-books replacing print books? tradition, serendipity, and opportunity in the adoption and use of e-books for historical research and teaching. *Journal of the American Society for Information Science and Technology* 64, 5, 1016-1028.

Metsämuuronen, J. (2005) Tutkimuksen tekemisen perusteet ihmistieteissä. 3rd ed. Helsinki: International Methelp.

Moore, G. A. (1999) Crossing the chasm: Marketing and selling high-tech products to mainstream customers. Rev. ed. New York: HarperBusiness.

Raynard, M. (2017) Understanding Academic E-books Through the Diffusion of Innovations Theory as a Basis for Developing Effective Marketing and Educational Strategies. *The Journal of Academic Librarianship* 43, 1, 82-86.

Rogers, E. M. (2003) Diffusion of innovations. 5th ed. New York: Free Press.

Ronte, H. (2001) The impact of technology on publishing. *Publishing Research Quarterly* 16, 4, 11-22.

Storytel. (2020) Tietoa Storytelistä [online document]. [Accessed 12.2.2020]. Available at <https://www.storytel.com/fi/fi/tietoja-storytelista>

Sääksjärvi, M. & Hellén, K. (2019). Idea selection using innovators and early adopters. *European Journal of Innovation Management* 22, 4, 585-599.

Taloustutkimus Ltd., The Booksellers' Association of Finland & The Finnish Book Publishers Association. (2018) Suomi lukee [online document]. [Accessed 5.5.2020]. Available at <https://kirjakauppaliitto.fi/wp-content/uploads/2019/05/Suomi-lukee.pdf>

The Finnish Book Publishers Association. (2018) 2013-2018 Total sales of literature type without VAT, (1 000 €) [online document]. [Accessed 10.2.2020]. Available at <http://tilastointi.kustantajat.fi/PublicReporting/Yearly.aspx?reportName=SalesHistoryGenresPrintAndDigital-C.xml&language=ENG>

The Finnish Book Publishers Association. (2020) Reading instructions for statistics [online document]. [Accessed 14.2.2020]. Available at <http://tilastointi.kustantajat.fi/PublicReporting/Yearly.aspx?language=ENG>

Van der Velde, W. & Ernst, O. (2009) The future of eBooks? Will print disappear? An end-user perspective. *Library Hi Tech* 27, 4, 570-583.

Walton, E. W. (2014) Why undergraduate students choose to use e-books. *Journal of Librarianship and Information Science* 46, 4, 263-270.

Wilson, T. D. (2014) The E-Book Phenomenon: A Disruptive Technology. *Information Research: An International Electronic Journal* 19, 2, paper 621.

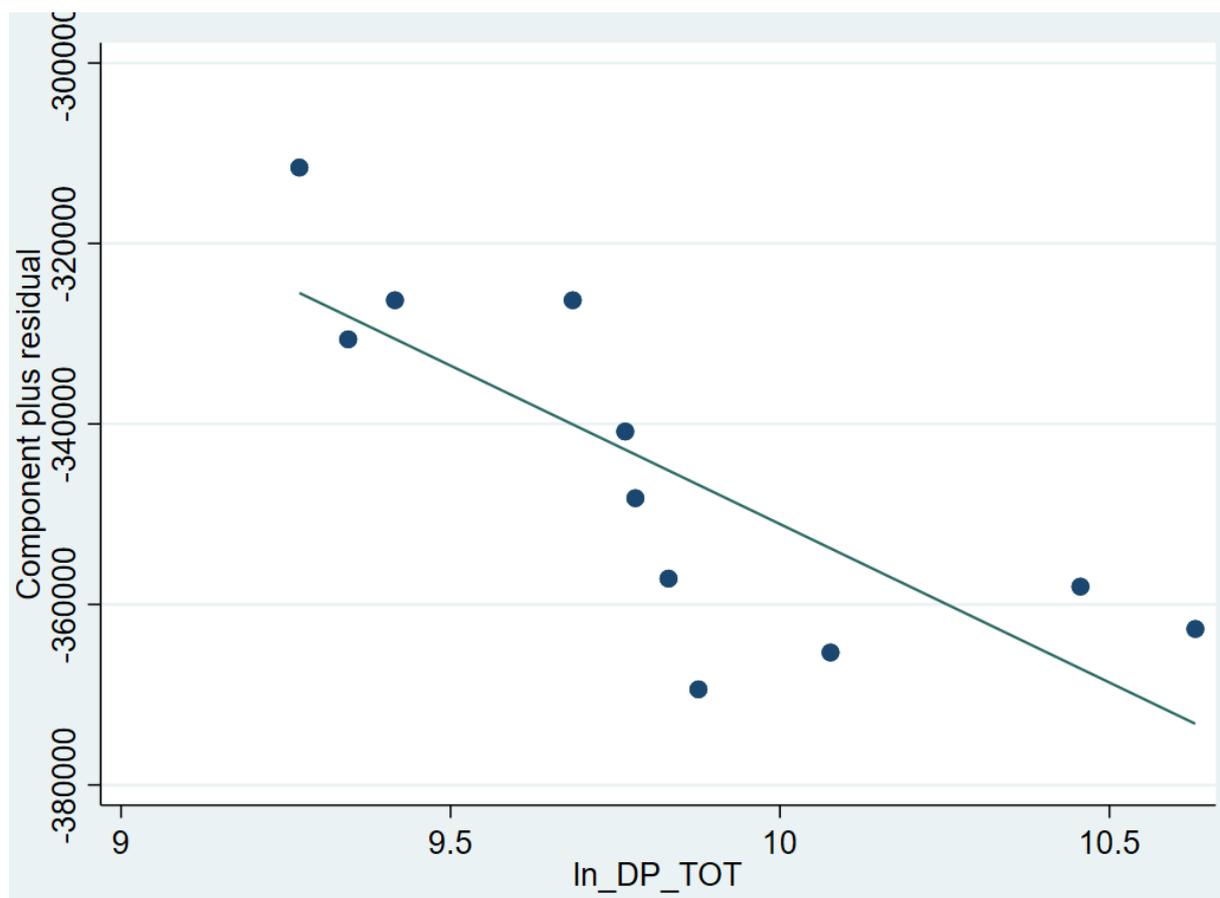
APPENDICES

APPENDIX 1. Shapiro-Wilk test for normality results.

Variable	W	V	Z	Prob (Z)
PB_TOT	.93	1.12	.20	.42
DP_TOT	.84	2.55	1.82	.03
ln_DP_TOT	.93	1.01	.15	.44

Variable	W	V	Z	Prob (Z)
r	.94	1.01	.02	.49

APPENDIX 2. Component-plus-residual plot.



APPENDIX 3. White's test for heteroskedasticity results.

Test of First and Second Moment Specification		
DF	Chi-Square	Pr>ChiSq
2	.73	.69

APPENDIX 4. Breusch-Godfrey LM test results.

Breusch-Godfrey LM test for autocorrelation			
Lags	DF	Chi-Square	Pr>ChiSq
1	1	2.43	.12

APPENDIX 5. Time series graphs for variables and the first differences of variables.

