No. 17-2024

Maximilian Maurice Gail and Phil-Adrian Klotz

Digital vs. Physical Goods: Evidence from the Book Market

Digital vs. Physical Goods: Evidence from the Book Market

Maximilian Maurice Gail[1*] and Phil-Adrian Klotz[2]

[1*]Chair for Industrial Organization, Regulation and Antitrust, Department of Economics, Justus-Liebig-University Giessen, Licher Strasse 62, Giessen, 35394, Germany.
[2*]Chair for Microeconomics, Department of Economics, University of Hohenheim, Schloss Osthof-West, Stuttgart, 70599, Germany.

*Corresponding author(s). E-mail(s): maximilian.m.gail@wirtschaft.uni-giessen.de; Contributing authors: philadrian.klotz@uni-hohenheim.de;

Abstract

The digitization has reshaped the multimedia industry in a tremendous way and particularly driven the transition from physical to digital goods. With respect to the book industry, Amazon reported that purchases of e-books had surpassed those of print books for the first time in 2011. The goal of this paper is to examine the substitutability between digital and print book formats in a country with fixed book prices (Germany) and a country without such regulated prices (United Kingdom). We use a unique cross-sectional data set of book prices for these two countries and exploit genre as well as publisher variation to estimate cross-format elasticities in an IV setting. We find that consumers basically consider e-books as (imperfect) substitutes for print books, even though there are country-, genre- and format-specific differences. Our results have important implications for the implementation of fixed book prices and the taxation of different book formats as well as for the release strategies of the publishers for the individual book formats.

1 Introduction

The multimedia industry (including books, music and movies) has seen a remarkable growth in the most recent past, mainly driven by the digitization. Thereby, many industries have been making or already made the transition from physical to digital goods. Today, most people do not visit brick-and-mortar stores to buy CDs or DVDs but use streaming services like Spotify, Deezer and Netflix to consume music or movies. Concerning the film industry, DVD sales in the United States have declined from \$16.3 billion in 2005 to \$1.56 billion in 2023,[1] whereas the number of Netflix subscribers worldwide has passed the 200 million barrier for the first time in 2020.[2]

The transition from print to digital books (e-books) is one of the most significant changes in the 21st century.[3] Rao (2005) even states that "the most important development in the world of literature after Gutenberg is the electronic book". In 2011, Amazon reported that purchases of e-books had surpassed those of print books. Rainie et al (2012) have found that 43% of Americans aged 16 and older read some form of digital content in 2011. At the same time, there had been a corresponding decrease in the percentage of Americans reading print books (from 78% in 2011 to 75% in 2012).

Digital books exhibit some basic properties that distinguish them from print books. In terms of supply, the marginal production costs of digital books are almost zero. Regarding the demand side, consumers can own and access thousands of e-books anytime with an e-book device, limited only by its memory capacity. Beyond, they do not need to pay shipping charges and wait for delivery when they buy e-books. In contrast, disadvantages of e-books are the requirement of an e-book reader advice and the necessity of setting up an e-book store account. Moreover, some people might feel uncomfortable about reading electronic texts. Hence, it is an open question whether print books and e-books are substitutes (Gilbert, 2015).

Therefore, the book format preferences of consumers are an interesting research area from an economic as well as a policymaker point of view. Economically, it is an open research question whether consumers consider the three different book formats (hardcover, paperback and e-book) as substitutes or rather as complements. Information on elasticities of different book formats are also important for policy makers because books are often considered as merit goods, which exhibit positive external effects, justifying a state intervention to internalize this positive externality.[4] In the European Economic Area (EEA), all countries except Bulgaria and Denmark have introduced a reduced value added tax (VAT) rate for print books to subsidize this product (Colbjørnsen, 2015). However, it is questionable whether the same cultural policy arguments and legal considerations, which are emphasized for the VAT rate reduction of print books, also apply for e-books.[5]

The main purpose of this article is to examine the substitutability between different book formats in a country with regulated prices (Germany) and a country without a governmental book price regulation (United Kingdom).[6] Therefore, we estimate cross-format elasticities between digital (e-books) and print books (paperback and hardcover) to investigate whether these book formats are substitutes or rather complements. Since there is reverse causality between book quantities and prices, we

---

[1]See https://www.cnbc.com/2019/11/08/the-death-of-the-dvd-why-sales-dropped-more-than-86percent-in-13-years.html and SVOD dara sourced from Omdia (technology.informa.com).

[2]See https://de.statista.com/statistik/daten/studie/196642/umfrage/abonnenten-von-netflix-quartalszahlen

[3]E-books have been around since the early 1970s, however, they only gained greater popularity when E Ink (electrophoretic ink) became available and offered a much better reader experience from 2006 onwards (OECD, 2012).

[4]The support of books is especially associated with cultural policy aims of supporting artistic quality, language, literacy and education. State intervention is also justifiable for goods exhibiting negative externalities like alcohol (see, e.g., Ogwang and Cho (2009)) or tobacco (see, e.g., Grace et al (2014)).

[5]Policymakers aim to promote the diversity and wide availability of books as well as their consumption and readership. Nevertheless, the arguments that link these policy objectives for books generally refer to . Germany has introduced a reduced VAT rate for digital books in 2019 (see https://www.globalvatcompliance.com/globalvatnews/germany-annual-tax-act-2019-vat-rate-on-ebooks/).

[6]Even though there is no fixed book price (FBP) system in the UK, e-books are partly sold under a so-called agency model on Amazon.co.uk. Thereby, retail prices are set by publishers and the retailer (=Amazon) merely acts as an agent for the publishers (see De los Santos and Wildenbeest (2017)). Thus, the agency model has similar effects as a FBP system, but is not imposed on a market level.

employ an instrumental variable (IV) approach using the number of pages of the print formats as instruments for the hardcover and paperback prices.

We find that consumers basically consider the print book formats (paperback and hardcover) as substitutes for digital books. The estimated cross-format elasticities are higher in the British sample (compared with the German one), which can very likely be attributed to the price regulation in Germany. We also find differences in the format preferences between British and German readers. Beyond, our results imply that the substitutability between digital and print book formats is stronger for fiction than for non-fiction books (independently of the the specific print format and country).

The rest of the paper is structured as follows. In Section 2, we describe the related literature. Facts on the different book formats, the construction of our data set and descriptive statistics for both countries are presented in Section 3. Section 4 presents our two main estimation strategies and the related results. We discuss the implications of our analysis in Section 5. Finally, we conclude and outline the contributions of our paper in Section 6.

2 Related Literature

Our article contributes to the literature of cross-format elasticities in the multimedia industry and the relationship between different book formats and sales channels.

First and foremost, our article contributes to studies estimating cross-price elasticities. As already stated in the introductory chapter, estimating cross-price elasticities is especially important for goods exhibiting positive or negative externalities. For instance, cross-price elasticities have been estimated for the demand of alcoholic beverages (Meng et al, 2014; Ogwang and Cho, 2009; Tomlinson and Branston, 2014) or the demand for tobacco (Grace et al, 2014).

Analyzing cross-format elasticities is also a major topic in empirical studies studying the multimedia industry (video, music or video games). With respect to the music industry, empirical papers analyze the relationship between buying music and music streaming (Dang Nguyen et al, 2012; Michel, 2006) and between different formats (LPs, CDs, etc.) (Stevans and Sessions, 2005). For the movie industry, several studies have shown that the different film channels are rather complements than substitutes (Gong et al, 2015; Hashim et al, 2019).

Second, our paper is related to studies investigating the relationship between the three different book formats e-book, paperback and hardcover. Thereby, the current state of research does not unambiguously answer the question whether digital and print books are substitutes or rather complements. This question is closely linked to the relationship between offline and online sales channels, while most of the literature states that the digital sales channel is rather a complement than a substitute for physical bookstores (Wang and Goldfarb, 2017; Goetz et al, 2020).

Most of the previous literature on different book formats has been focused on the relationship between paperback and hardcover books. Clerides (2002) finds that differences in prices between these two formats cannot be attributed solely to production costs, making this a case of quality discrimination. Barrot et al (2015) show that the price elasticity for hardcover books is substantially smaller (-3.7) than for paperback books (-4.3). Beyond, their estimation results exhibit positive and significant spillover effects of hardcover sales on paperback sales indicating a market expansion effect.

More recently, also the effect of digital books on print book formats and the offline sales channel has been analyzed by different authors. Li (2019) finds that e-books are closer substitutes to online print books than to offline print books. Lee and Lee (2013) figure out that the demand of e-books is getting less price elastic as time goes on and that the existence of a corresponding print book reduces consumers' price elasticity, which is slightly counterintuitive. Li (2013) has analyzed the degree of substitution between e-books and printed books. Two-thirds of e-book sales were found to be a result of the cannibalization of print book sales, with the remaining third stemming from market expansion. Using sales data from Amazon, Chen et al (2019) find that delaying e-book availability results in a 43.8% decrease in e-book sales but no increase

in print book sales. This finding suggests a limited substitutability between e-books and printed books.

Some other authors show that print books and e-books have unique attributes and serve distinct functions in meeting people's reading needs, which may vary by individual demographic, contextual, and situational factors. Only the 'technological adopters' substitute print books by e-books (Crosby, 2019). Another important finding is that individuals indicate to read e-books more for work or school (contextual factors) (Singer and Alexander, 2017). Moreover, people prefer print books over e-books in certain situations such as reading to a child or sharing books with other people. E-books are preferred when requiring a book quickly or reading a book while on the move (Zhang and Kudva, 2013).

3 Facts and Statistics on Book Formats

Our data set contains sales ranks and prices for the three different book formats e-book, paperback and hardcover. In this section, we will present our sample for the empirical analysis. Before, we will describe the characteristics of different book formats in Section 3.1. In Section 3.2, we will explain the construction of our data set and, finally, present the descriptive statistics in Section 3.3.

3.1 Characteristics of the Different Book Formats

In most markets, products are differentiated which means that the sold goods are similar but no perfect substitutes. Multimedia products are horizontally as well as vertically differentiated. Books are horizontally differentiated because consumers prefer distinct book genres and they are also vertically differentiated due to the differences in quality. For instance, book scientists measure the quality of a book based on tension curves or the share and distribution of direct speech.

This product differentiation of books is further increased due to the availability of different book formats. Thereby, it is not straightforward whether book formats are vertically or horizontally differentiated. However, the majority of studies assumes that print formats (paperback and hardcover) are vertically differentiated due to the various binding quality (see, e.g., Clerides (2002)). In contrast, the distinction between print and digital formats is mostly considered as horizontal product differentiation (Hao and Fan, 2014; Chen and King, 2017).

In the modern book market, we can differentiate between four different book formats. First, hardcover books are bound with rigid protective covers. This format is considered to be more durable than paperback books. However, they are also heavier and therefore less portable than their paperback equivalents. The lower-quality print format, called paperback, is characterised by a thick paper cover. This format is flexible, often held together by glue. Paperback books are generally lighter and smaller than their hardback equivalents. However, they are likely to be less durable.

Third, e-books (or electronic books) are books that have been made available in digital form, consisting of text, images or both. E-books are readable on computers, smartphones and dedicated e-reader devices such as the Amazon Kindle. Finally, audio books (or talking books) are a recording of a book's text being read. Nowadays, audio books are sold in the form of downloadable files which can be listened to on computers, smartphones and stereo equipment (for example, at home or in-car). In our empirical analysis, we will focus on the two print formats and e-books.

Basically, all contents can be published in both print and digital editions. However, this fact does not automatically imply that e-books and print books are perfect substitutes since e-books have some basic properties in terms of supply and demand that differentiate them from print books.

Generally, book publishing is characterized by high fixed costs and low marginal costs. Three basic steps are necessary to publish a book: manuscript preparation, book production as well as distribution and marketing. It is hard to quantify the exact costs of the first and third steps because, e.g., the costs for the manuscript preparation are mostly depending on the editor's time (opportunity costs). Nevertheless, costs

regarding the manuscript preparation as well as distribution and marketing hardly differ between print and digital books, or rather, are common costs for both formats.

In contrast, the costs of the production process itself can be calculated and significantly differ between print and digital formats. Since e-books are pure digital goods, the marginal production cost is almost zero. Furthermore, the retailers of e-books do not have to worry about the cost of storage and distribution. The production process of print books involves the three stages typesetting, printing, and binding. The cost of typesetting is fixed, while the cost of binding is almost entirely variable. Printing costs have a fixed component (the cost of plates) and a variable component (paper and press time). Marginal production costs are higher for hardcover books compared to the paperback format due to the cost of binding. Some authors argue that it might be appropriate to regard virtually all expenses incurred by the publisher (except for storage, transportation and handling) as fixed, sunk costs (see, e.g., Bittlingmayer (1992)).

Digital and print books also differ with respect to the demand side, in particular due to the different text comprehension. E-books have both advantages as well as disadvantages compared to print books. On the one hand, disadvantages are that consumers need to have an e-book reader device, install e-book software and set up an e-book store account. On the other hand, consumers can own and access thousands of e-books anytime with an e-book device, limited only by its memory capacity and do not need to pay shipping charges and wait for delivery when they buy e-books.

3.2 Data

Our empirical analysis is based on book data from Amazon in UK and Germany. Therefore, we have scraped the Amazon.co.uk and Amazon.de web pages for a list of publisher and imprint names to get retail prices for the three formats hardcover, paperback and e-book as well as further book characteristics available on the website. We describe the details on our data generating process in Appendix B.

Our final data set consists of 10,859 book titles being published on Amazon between March 2010 and March 2020. Thereby, our overall data set can be divided into a German and a British sample. The German sample consists 2,544 book titles being sold on Amazon.de by the publishers Springer Science & Business Media, Bonnier Media Deutschand GmbH, BookRix GmbH & Co. KG, C.H.Beck, Holtzbrinck Publishing Group, Klett Gruppe, Carl Hanser Verlag GmbH & Co. KG and the Suhrkamp Verlag AG. The sample for UK contains 8,315 titles being sold on Amazon.co.uk by the publishers Bloomsbury, Hachette, HarperCollins, Pan Macmillan, Penguin Random House and Simon & Schuster. Additionally, in both samples there is a category for smaller/independent publishers included.

Our variables of interest are the retail price, which is the price a consumer has to pay for a respective book, and the variable book sales rank, which can be interpreted as the Amazon bestseller rank. The latter is calculated based on the sold quantities of book titles. Beyond, we have data on several control variables for our empirical analysis. These variables comprise book characteristics as the book format, the book genre, the number of pages for the print version/ size of the e-book in KB, variables on book reviews as the star rating or the number of consumer and expert reviews of a book, variables containing information on the author or publisher of a book title and other variables as the publication date or the recommended retail price (RRP). Table 1 summarizes the descriptions of those variables.

Each book is a unique good written by an author and mostly published by one publisher. Thus, books are heterogeneous goods which make it impossible to actually compare the value of one specific book with one another. In order to provide an acceptable analysis, it is therefore also necessary to control for the genres of the several books. Hence, we use a Latent Dirichlet Allocation (LDA) to determine book genres

Variables	Information
Retail price	Retail price from the upper right *Buy-Box* on Amazon
Format	Hardcover, paperback, e-book
Star rating	Average star rating normalized to be between 0 and 1
No. customer reviews	Number of customer reviews
No. expert reviews	Number of expert reviews on Amazon
Series	Dummy variable whether book is part of a series
Book description	Detailed text-information for a book
Genre	Constructed by LDA from the descriptions and reviews
RRP	Recommended retail price of the print format. For e-books it is either related to the hardcover or paperback RRP
Author	Information on the author of a book
Title	Information on the title of a book
Kindle Size	Digital file size (in KB)
Pages	Number of pages for the print format
Publisher	Name of the publisher. We have different levels of aggregation (Imprint,Publisher,Publishing House)
Amazon rank	Uncategorized Amazon bestseller rank for either print books or e-books
Bestsellers	Number of bestsellers in the *Sunday Times* bestseller list conditional on the author's name
WeekInChart	Average number of weeks in the bestseller charts conditional on the author's name
Identifier	Aggregation of ASINs to verify the books
Date Retail	Period of time since the publication of a book title (in years)
Other	Other small control variables

Table 1: Relevant Variables in our data set and their information content.

from the descriptions and reviews of the individual books available on the Amazon web page.[7]

Such a text mining approach is necessary because on the Amazon web page the genre information is ambiguous and even not available for some book titles. Based on the coherence value across the estimated LDA models using smaller numbers of pages, we find that 15 genres provide the best statistical decomposition of our book description corpus for both countries.[8] The LDA decomposition gives a meaningful and easily interpretable genre classification of the book descriptions, which it does because our LDA approach identifies all important book genres and clearly delineates the topics.[9] Thus, our control variable *genre* should be able to capture specific effects within the individual genres.

In order to estimate cross-format elasticities, the observed Amazon sales rank for each book will be transformed into a measure of quantity by employing the approach used by Chevalier and Goolsbee (2003), while we will slightly modify their approach for our analysis.[10] Therefore, we use the *rank* for calculating the sales Q based on the following equation:

$$\log(Q) = \frac{\log(N) + \theta \log(k) - \log(Rank)}{\theta}. \tag{1}$$

where k and θ are the parameters of the assumed Pareto distribution.[11] The shape parameter θ indicates the relative frequency of large observations. In our analysis, we use values for θ ranging from 0.8 to 1.2 in steps of 0.1 units. We choose to set $k = 1$ as it refers to the minimum for a sold book. Moreover, N gives the total number of books in the sample and $Rank$ is the Amazon book sales rank of a book title in equation (1). For the exact determination of the demand values in the two countries, we use our full sample of books to incorporate N as large as possible in the calculations. We also normalize the demand values to be between 0 and 1 as the values do not necessarily reflect the true sales values.

---

[7]In the last years, there has been an explosion of empirical research in economics using text as data (see, e.g., Larsen and Thorsrud (2019) for an Latent Dirichlet Allocation (LDA) approach or Lenz and Winker (2020) for paragraph vector topic modelling).

[8]One caveat of the LDA estimation procedure is that it does not give the topics/genres any name or label. Thus, labels are subjectively given to each topic based on the most important words associated with each topic. In the most cases, it is conceptually simple to classify the topics. Besides, the exact labeling plays no material role in our empirical approach, it is just used as a convenient way of referring to the different topics instead of only using topic numbers.

[9]A detailed list of all 15 genre is presented in Table A.1 of Appendix A for the British sample. For Germany we find that 9 genres provide the best statistical decomposition. The Figures A.1 and A.3 in Appendix A show two examples of our 15 word clouds to visualize within the topic distribution of words by assigned probabilities through the LDA for books sold in the UK. We have labelled the genre in Figure A.1 *Crime novel/Thriller* and the genre in Figure A.3 *Politics*.

[10]In Chevalier and Goolsbee (2003), the equation to calculate the sales Q based on the sales rank is $\log(Q) = \frac{\log(N) + \theta \log(k) - \log(Rank-1)}{\theta}$. For the first sales rank, this formula is not defined as we would need to take the logarithm of zero. Therefore, we decided to add up each rank by one.

[11]We adhere to the methodology outlined in Chevalier and Goolsbee (2003, see Chapter 2 and Footnote 8) and their presented literature on the assumption that book ranks follow a Pareto distribution. Consequently, the probability of a book sale s exceeding a certain threshold level S follows an exponential distribution in the style of $Probability(s > S) = \left(\frac{k}{S}\right)^{\theta}$.

3.3 Descriptive Statistics

Our final data set consists of 10,859 book titles being published on Amazon between March 2010 and March 2020. Thereby, our overall data set can be divided into a German and a British sample. The German sample consists 2,544 book titles being sold on Amazon.de by the publishers Springer Science & Business Media, Bonnier Media Deutschand GmbH, BookRix GmbH & Co. KG, C.H.Beck, Holtzbrinck Publishing Group, Klett Gruppe, Carl Hanser Verlag GmbH & Co. KG and the Suhrkamp Verlag AG. The sample for UK contains 8,315 titles being sold on Amazon.co.uk by the publishers Bloomsbury, Hachette, HarperCollins, Pan Macmillan, Penguin Random House and Simon & Schuster. Additionally, in both samples there is a category for smaller/independent publishers included.

Type	Country	count	mean	std	min	25%	50%	75%	max
e-book Price	UK	6,539	5.95	4.31	0.25	3.94	4.99	7.49	47.49
Paperback Price	UK	6,539	9.32	4.56	0.96	6.55	8.19	10.99	67.11
e-Book Price	GER	2,239	10.08	7.23	0.49	4.99	9.99	12.99	79.99
Paperback Price	GER	2,239	13.92	8.15	3.00	9.99	11.99	15.05	99.99

(a) Subsamples e-book/paperback (in € for Germany and in £ for the UK).

Type	Country	count	mean	std	min	25%	50%	75%	max
e-Book Price	UK	5,506	8.19	5.46	0.49	4.49	7.29	9.99	83.79
Hardcover Price	UK	5,506	16.23	10.11	1.99	11.55	14.29	18.99	139.50
e-Book Price	GER	1,516	14.30	7.85	0.49	9.99	14.99	18.99	69.99
Hardcover Price	GER	1,516	20.74	8.65	3.50	16.60	20.00	24.00	98.00

(b) Subsamples e-book/hardcover (in € for Germany and in £ for the UK).

Table 2: Retail prices by book formats for the subsamples.

The main variables of our empirical approach are the prices and the sold quantities (calculated based on Amazon sales ranks, see Section 3.2) for the three book formats. We have divided the samples for both countries into an e-book/paperback and an e-book/hardcover subsample, respectively. Table 2a gives descriptive statistics for the retail prices of the e-book/paperback subsamples for Germany and the UK. It is obvious that the average retail price of e-books in Germany, which amounts to €10.08 (median: €9.99), is appreciably higher than the average e-book price in the UK (mean price: £5.95, the median is £4.99). The same pattern is true for the paperback average retail prices (without controlling for the different purchasing power and currencies) .

Table 2b presents the price statistics for the e-book/hardcover subsamples. It is worth mentioning that the e-book price statistics differ compared to the ones in Table 2a because we consider other subsamples now. Only e-books with a related paperback or hardcover version are contained in the subsamples, respectively. E-books, which only have a related paperback (e-book/paperback sample) or only a related hardcover (e-book/hardcover sample) version, are only included in one of the two subsamples. In the e-book/hardcover subsamples, the e-book retail prices are generally higher compared to the e-book/paperback subsamples and e-book prices are still higher in Germany than in the UK. In addition, Table 2b exhibits that on average also hardcover books are more expensive in Germany compared to the UK.

The distribution of retail prices for the three different book formats is shown in Figure 1. The retail prices for the subsample e-book/paperback are displayed in the first row of Figure 1 and the prices of the subsample e-book/hardcover in the second row (Germany on the left-, UK on the right-hand side). It is apparent that e-book prices in Germany (the UK) are mostly in the range between 0.49-€10 (0.49-£10),

most of the paperback prices are concentrated in the 10-€20 (10-£20) interval, and hardcover prices are generally higher in both countries. In addition, all three distributions exhibit significant mass points at candidate focal points in both countries at, e.g., €4.99 (£4.99) for e-books, €9.99 (£9.99) for paperbacks and €19.99 (£15.99) for the hardcover version. Focal point pricing is a frequently observed phenomena in the book market (see, e.g., Beck (2004)).
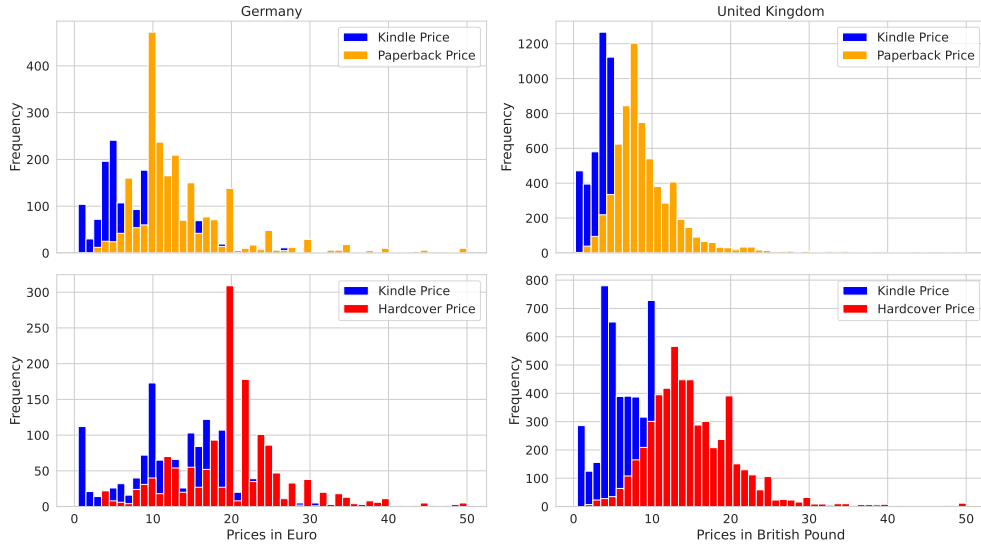

Figure 1: Distribution of retail prices for the three different book formats. First row refers to the respective e-book/paperback subsamples and the second row to the respective e-book/hardcover subsamples. Prices in the Figure have been limited to €50 (respectively £50).

Table 3 gives summary statistics for the remaining variables of the German subsample divided by publishers. In addition to the retail prices, we observe several characteristics for each book title such as the Amazon rank of a title on Amazon for every format,[12] the customer ratings, the number of customer and expert reviews, the number of pages for the print formats and the digital size (in KB) for e-books.[13] As shown in the table, titles from the Holtzbrinck Publishing Group have the lowest mean book rank (= most sold titles) for e-books and titles from C.H. Beck have the lowest mean book rank for the paperback version (upper chart in Table 3a) as well as for the hardcover version (lower chart in Table 3b). In addition, book titles from the Holtzbrinck Publishing Group have the highest average number of customer reviews on Amazon.de. Most of the other book characteristics are very similar across publishers. The low mean value for the variable *Date Retail* shows that the book titles in our subsample have been published in the recent past.

The summary statistics for the remaining variables of the British subsample divided by publishers are given in Table 4. Obviously, e-books published by Hachette are most popular on Amazon.co.uk since digital books from this publisher have the lowest mean book rank. With respect to the print formats, book titles published by Penguin Random House have the lowest mean book rank for the format hardcover, whereas Hachette also has the lowest mean book rank for paperback books. Books published by Pan Macmillan have the highest mean customer and expert reviews. Again, most of the other characteristics are very similar across publishers.

Figure 2 presents the price relation between the three book formats in Germany and the UK for the several publishers.[14] The two upper images display the relation of prices between e-books and paperbacks, the two lower images show the relation of

---

[12] Amazon reports book ranks for each format of a book independently. Therefore, e-books, paperbacks and hardcover have their own Amazon rank, respectively (see https://kdp.amazon.com/en_US/help/topic /G201648140).

[13] To keep the summary statistics clearly, only ranks and prices are depicted.

[14] All prices and quantities presented in the following figures are given in logarithms.

	Publisher	bc partners limited	bonnier media	bookrix	c.h.beck	dtv	ernst klett ag	hanser	holzbrinck	small/indie	suhrkamp verlag
	Count	182	812	104	94	68	72	53	288	488	78
Sales rank e-book	mean	501,416.15	156,501.35	139,136.38	222,434.05	115,856.35	182,240.24	166,875.75	110,194.85	319,848.76	220,119.54
	median	398,378.00	96,184.00	88,710.50	166,336.50	66,599.50	136,004.50	151,453.00	59,317.00	217,972.00	187,190.00
	std	404,097.24	177,058.44	169,060.83	194,562.08	138,348.49	173,862.43	130,413.19	159,167.57	329,557.12	177,942.32
	min	32,357.00	933.00	610.00	3,581.00	1,491.00	10,957.00	5,801.00	369.00	531.00	3,442.00
	max	2,440,364.00	1,305,034.00	1,029,737.00	1,180,047.00	756,333.00	1,119,141.00	494,458.00	1,394,378.00	2,118,026.00	687,869.00
Sales rank paperback	mean	446,288.35	358,806.05	1,377,435.10	286,894.21	413,844.53	159,016.25	272,010.64	302,704.31	732,156.49	273,733.79
	median	297,801.50	193,870.50	1,017,418.50	207,907.50	175,545.00	79,289.50	217,497.00	116,395.00	326,631.50	204,240.50
	std	837,889.09	560,349.15	1,306,395.96	292,146.55	663,429.05	229,141.21	251,399.59	535,872.19	1,228,516.77	301,808.18
	min	8,453.00	78.00	4,843.00	812.00	2,819.00	1,545.00	8,650.00	358.00	843.00	1,491.00
	max	10,427,065.00	10,415,483.00	5,995,113.00	1,723,927.00	3,606,255.00	1,193,925.00	1,110,655.00	4,587,144.00	10,405,808.00	1,544,436.00
Kindle Price	mean	20.53	8.20	3.40	11.18	9.53	15.20	16.67	11.73	9.99	8.52	16.28
	median	16.99	6.99	2.99	10.24	9.99	15.20	10.99	9.99	7.99	13.99
	std	12.48	4.40	1.53	4.46	2.17	6.15	2.63	3.19	7.12	8.17
	min	4.99	1.99	0.99	6.49	3.99	8.99	4.99	1.99	0.49	5.99
	max	79.99	44.99	7.99	26.99	18.99	19.99	19.99	49.99	74.99	57.99
Paperback Price	mean	27.84	12.00	10.51	13.98	10.73	20.39	11.90	11.25	13.54	16.28
	median	22.99	11.00	10.00	12.93	10.95	18.00	10.95	10.99	11.00	14.00
	std	14.33	4.24	2.88	5.10	1.53	8.29	2.45	1.92	8.61	8.17
	min	9.99	3.99	3.29	7.90	5.90	9.95	6.95	4.00	3.00	6.00
	max	99.99	39.99	16.99	34.00	16.95	45.00	22.00	20.00	96.30	58.00

(a) E-book/Paperback Subsample

	Publisher	bonnier media	c.h.beck	dtv	ernst klett ag	hanser	holzbrinck	small/indie	suhrkamp verlag
	Count	384	79	71	75	143	305	378	81
Sales rank e-book	mean	193,898.56	156,725.86	204,522.68	166,701.68	179,735.94	129,146.37	366,855.70	201,723.07
	median	119,827.50	137,427.00	131,284.00	119,605.00	134,970.00	71,888.00	231,448.00	169,960.00
	std	250,131.34	113,460.07	234,506.93	155,173.55	179,135.62	176,608.03	391,401.70	178,406.50
	min	1,272.00	3,581.00	1,491.00	3,808.00	3,325.00	369.00	531.00	3,442.00
	max	2,445,238.00	605,652.00	1,305,500.00	929,315.00	1,158,946.00	1,520,656.00	2,438,164.00	1,061,394.00
Sales rank hardcover	mean	179,927.31	162,893.43	285,253.93	195,868.28	253,555.77	211,090.44	818,955.25	224,070.90
	median	92,862.50	145,451.00	224,302.00	153,228.00	164,511.00	137,075.00	287,366.50	165,333.00
	std	223,172.34	138,744.00	341,480.04	233,487.47	280,974.90	274,434.33	1,205,102.98	201,255.25
	min	2.00	1,815.00	2,939.00	123.00	700.00	510.00	1,091.00	4,581.00
	max	1,431,818.00	690,799.00	2,367,890.00	1,566,746.00	1,869,210.00	2,457,002.00	8,023,168.00	1,016,172.00
Kindle Price	mean	14.66	19.91	13.47	17.74	17.92	15.40	8.84	19.55
	median	14.99	19.99	12.99	15.99	15.99	14.99	5.99	18.99
	std	5.68	5.77	3.94	5.81	8.48	6.57	8.59	7.44
	min	2.99	7.49	6.99	8.99	4.99	1.99	0.49	9.99
	max	44.99	49.99	22.99	35.99	47.99	69.99	44.99	57.99
Hardcover Price	mean	18.87	26.65	17.64	24.38	25.25	20.45	18.87	25.02
	median	19.99	26.00	18.00	22.00	23.00	20.00	18.00	24.00
	std	5.64	6.25	4.00	8.12	9.20	6.30	11.35	9.92
	min	3.99	16.00	9.95	12.00	10.00	4.95	3.50	10.00
	max	50.00	58.00	26.00	68.00	59.99	78.00	98.00	98.00

(b) E-book/Hardcover Subsample

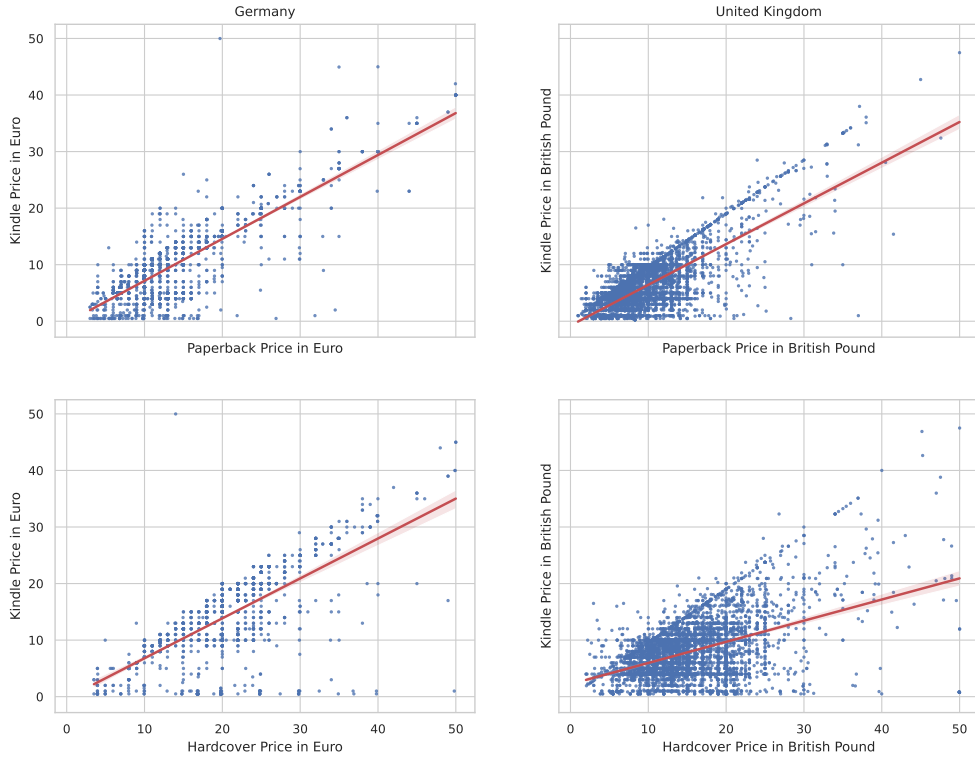**Table 3**: Summary statistics for the German subsamples.

Figure 2: Relation of prices for e-books to the price of paperback (above) and hardcover (below) versions.

prices between e-books and hardcover books (for both, Germany on the left- and the

Publisher	bloomsbury	faber	hachette	harpercollins	indie/small	oxford	pan macmillan	penguin random house	scholastic	simon & schuster
Count	213	107	1227	951	2270	92	288	1030	99	262
Sales rank e-book mean	554,060.50	366,049.65	222,208.46	409,376.46	626,943.49	908,883.95	403,418.42	333,933.95	490,242.00	521,048.55
median	321,764.00	260,317.00	105,198.00	222,719.00	383,996.00	773,413.50	228,407.50	151,974.00	366,349.00	271,614.50
std	563,689.60	447,400.24	322,216.84	512,022.56	659,726.55	672,738.63	497,151.20	466,947.63	487,489.79	608,012.04
min	2,945.00	7,137.00	70.00	9.00	2.00	43,091.00	151.00	17.00	12,443.00	143.00
max	3,301,125.00	3,310,000.00	2,129,743.00	3,314,026.00	3,328,748.00	2,698,746.00	3,313,802.00	3,312,467.00	2,564,675.00	3,314,538.00
Sales rank paperback mean	651,522.23	706,627.93	617,893.36	1,015,792.54	1,347,105.38	990,515.97	850,072.80	618,867.41	493,689.42	1,005,057.71
median	278,210.00	306,098.00	234,934.00	468,816.00	748,303.50	583,929.00	296,480.50	179,741.00	145,460.00	372,769.00
std	1,020,484.37	900,594.97	1,142,164.55	1,494,767.26	1,600,374.62	1,133,662.38	1,322,951.49	1,171,049.96	884,833.15	1,482,427.25
min	1,108.00	10,986.00	84.00	52.00	10.00	10,248.00	326.00	26.00	929.00	446.00
max	6,162,176.00	4,643,359.00	12,228,656.00	12,376,032.00	12,246,897.00	5,322,101.00	12,238,210.00	12,229,930.00	4,408,832.00	12,231,472.00
Kindle Price mean	9.51	5.59	4.55	4.77	6.36	16.21	6.06	6.04	4.72	6.76
median	7.17	5.03	3.99	4.40	4.68	14.37	5.71	4.99	4.68	6.49
std	5.53	2.47	1.98	2.54	5.49	9.12	2.36	2.54	1.16	3.03
min	1.42	1.89	0.99	0.75	0.25	3.32	0.99	0.99	0.98	0.75
max	27.54	15.20	27.99	14.99	47.49	46.80	13.12	19.99	8.04	16.49
Paperback Price mean	11.13	8.16	8.48	8.05	10.32	17.65	9.03	8.90	6.08	8.34
median	9.19	7.99	7.99	7.37	8.99	15.38	8.00	8.19	5.94	7.99
std	5.58	2.83	2.87	3.32	5.53	10.28	3.36	3.11	1.41	2.86
min	3.49	3.15	1.38	1.00	0.96	3.49	3.49	1.00	3.49	2.00
max	30.76	16.00	31.00	25.00	49.99	67.11	23.09	22.60	12.17	18.49

(a) E-book/Paperback Subsample

Publisher	bloomsbury	faber	hachette	harpercollins	indie/small	oxford	pan macmillan	penguin random house	scholastic	simon & schuster
Count	212	81	1039	701	1658	102	333	1054	45	281
Sales rank e-book mean	499,704.67	362,803.74	304,945.26	518,943.82	779,369.30	928,449.87	497,779.92	411,084.85	723,269.02	672,626.58
median	283,665.50	269,781.00	139,022.00	355,803.00	570,063.50	803,651.50	305,654.00	199,426.50	470,744.00	409,658.00
std	555,730.88	355,730.88	465,235.98	532,423.70	706,745.22	635,368.76	539,147.39	541,176.20	748,400.33	708,629.02
min	224.00	2,215.00	47.00	9.00	41.00	43,091.00	151.00	107.00	11,943.00	106.00
max	3,311,173.00	1,695,954.00	3,319,403.00	3,313,160.00	3,328,748.00	2,698,746.00	2,644,570.00	3,315,879.00	3,327,458.00	3,316,779.00
Sales rank hardcover mean	1,025,989.21	719,485.00	586,651.73	957,634.29	1,498,801.47	1,127,519.11	891,881.38	605,399.70	1,759,771.71	966,554.85
median	323,840.50	405,879.00	263,111.00	450,500.00	801,993.50	767,246.00	489,740.00	258,712.50	1,136,366.00	519,004.00
std	1,885,465.06	1,255,423.05	1,027,681.69	1,446,871.21	1,788,524.76	1,025,354.15	1,098,360.26	869,496.52	1,656,042.54	1,118,913.41
min	36.00	1,719.00	261.00	66.00	1,050.00	46,424.00	574.00	5.00	944.00	287.00
max	12,232,463.00	9,380,624.00	12,242,127.00	12,263,922.00	12,251,680.00	4,511,381.00	5,961,809.00	5,508,686.00	6,504,293.00	5,534,320.00
Kindle Price mean	10.50	6.74	6.17	7.32	9.11	16.92	8.31	7.98	5.65	8.84
median	8.72	6.02	4.99	6.99	7.83	15.20	7.70	7.99	4.76	9.49
std	7.35	4.25	3.45	3.50	7.25	8.52	4.11	3.29	1.96	3.38
min	2.05	0.98	0.99	0.75	0.49	5.03	0.99	0.99	1.84	0.80
max	78.11	35.99	27.99	29.99	83.79	50.65	32.30	40.00	10.44	20.99
Hardcover Price mean	20.68	14.33	14.58	14.30	18.31	35.65	14.43	14.47	11.19	14.50
median	15.32	12.05	13.87	13.35	16.02	22.66	13.64	13.73	12.96	13.29
std	16.32	11.52	4.46	4.99	12.41	30.42	4.97	4.86	3.16	6.93
min	5.16	5.14	2.03	5.23	2.05	8.47	1.99	2.48	2.58	2.96
max	83.99	100.00	39.81	55.00	137.00	139.50	37.34	40.00	15.99	49.95

(b) E-book/Hardcover Subsample

Table 4: Summary statistics for the British subsamples.

UK on the right-hand side). Every point is one observation in our data set. In general, Figure 2 outlines that digital and print format prices are positively correlated, so that the price for an e-book mirrors the price of the related print version (homothetic prices). This positive relation of print and digital books seems to be slightly stronger for the paperback version compared to the hardcover one. The positive correlation between digital and print books seems to be stronger in Germany than in the UK. This finding confirms previous empirical studies, whereupon the price homothety between digital and print books is higher in countries with FBP systems (as in Germany) than in countries without regulated prices (as in the UK) (see, e.g., Guillon and Thierry (2013)).[15]

Figure 3 constitutes the demand for e-books in our subsamples. The upper images display the relation between e-book quantities[16] and prices for our e-book/paperback subsample, and the lower ones for the e-book/hardcover subsample (Germany on the left- and UK on the right-hand side). We cannot observe any clear correlation in Figure 3 for both countries. There is a small negative relation for the British e-books, whereas in Germany there is no relationship at all between quantities and prices. However, focal point pricing is apparent for both countries again.

The demand (in logs) for paperback (upper images) and hardcover (lower images) books are presented for Germany and the UK in Figure 4. As for the demand of e-books, there is also no clear relationship apparent for prices and quantities of the two print formats. Only the focal point pricing becomes obvious again, which seems to be stronger for Germany than in the UK for print books.[17]

---

[15]We present the regression coefficients (including covariates) from Figure 2 in Table C.1 of Appendix C. First, prices of e-books are highly correlated to each other. Second, in Germany the coefficient of e-book prices on paperback or hardcover prices are 0.7332 and 0.5487, respectively. In comparison to the British coefficients, 0.618 and 0.2208, the German are large in magnitude. Empirically, these results back up the previous statement that price homothety is stronger in countries with FBP systems.

[16]As already mentioned above, we calculated book quantities based on Amazon sales ranks. See Section 3.2 for more details on that.

[17]Figure C.1 in Appendix C displays the relation between the sold quantities of e-books and the prices for the related paperback (above) and hardcover (below) versions. Assuming that digital and print books are substitutes, microeconomic theory would suggest a positive correlation between quantities of e-books and prices of the related print formats since the digital format would (*c.p.*) become more popular with higher print prices. However, there is no clear relationship apparent in Figure C.1
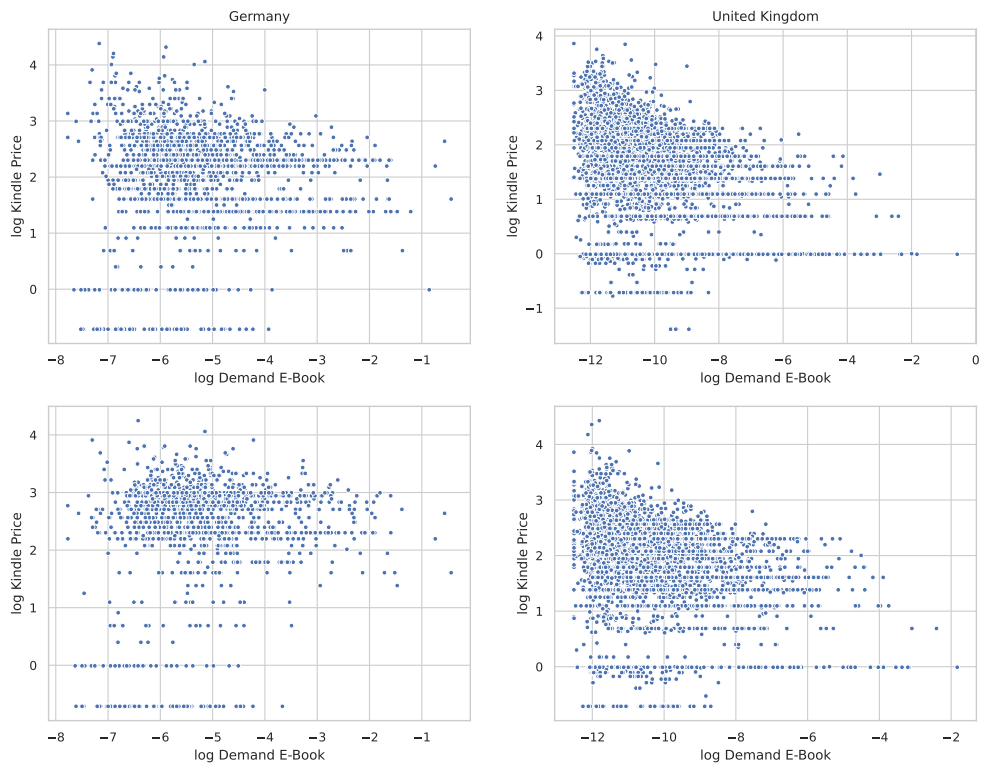
Figure 3: Demand of e-Books in Germany (left-hand side) and United Kingdom (right-hand side) for the e-book/paperback (above) and e-book/hardcover (below) subsamples.
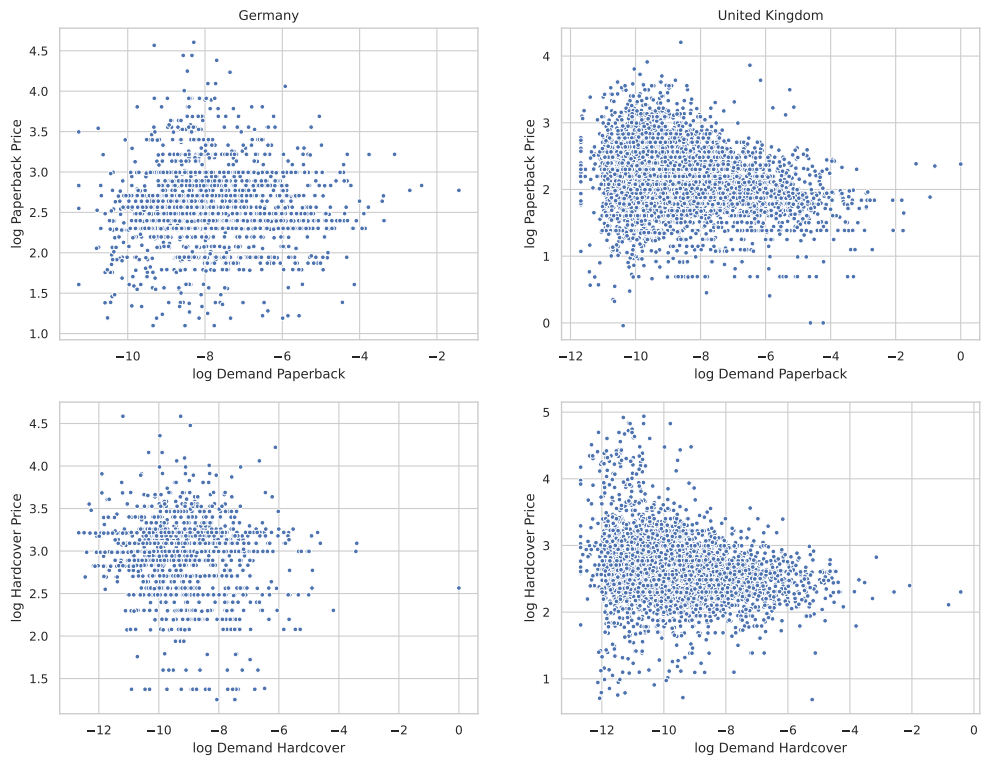

Figure 4: Demand of paperback and hardcover books for Germany (left-hand side) and United Kingdom (right-hand side) for the e-book/paperback (above) and e-book/hardcover (below) subsamples.

4 Empirical Analysis

In this chapter, we present our main empirical analysis. We estimate own- and cross-format elasticities for books by running an instrumental variable (IV) approach. In Section 4.1, we first explain our estimation strategy in more detail. In Section 4.2, we present present our estimation results.

4.1 Empirical Strategy

In our main empirical analysis, we estimate cross-format elasticities to analyze whether digital and print formats are substitutes or rather complements. In addition, we want to examine if there are differences among countries with regulated and unregulated book prices. Thus, the hypothesis to be tested in is:

Hypothesis 1 (H_0). *There is no relation between digital and print books from the consumers' perspective.*

If there will be a significant positive correlation between the retail price of print books (hardcover or paperback) and the demanded quantity for e-books, H_0 can be falsified and digital and print books are substitutes. Observing a significant negative correlation would also lead to a falsification of H_0, but digital and print books would be complements then.

To answer our research question, we estimate cross-format elasticities by running an instrumental variable (IV) estimation in the form of a two-stage least squares (2SLS) regression model. To deal with the problem that our regression models are not independent from each other, we also estimate a three-stage least squares (3SLS) model that accounts for several forms of endogeneity. This yields efficiency gains if there is substantial correlation of the error terms across equations and also poses the possibility to impose parameter restrictions across equations (Zellner, 1962; Zellner and Theil, 1962; Brundy and Jorgenson, 1971; Clerides, 2002). With respect to the book market, Clerides (2002) estimates a system of paperback and hardcover demand equations by employing a 3SLS estimation approach, while e-books did not play a role at that time. We rely to this approach by estimating own- and cross-price elasticities of e-books, paperback and hardcover in a similar way.

In our approach, we will even impose parameter restrictions across equations to equalize cross-price elasticities. This approach is unusual, but has its foundation stems from symmetry restrictions in structural demand systems of the Almost Ideal Demand System (AIDS). By imposing cross-price elasticity restrictions, we reflect the effects of utility maximization under budget constraints. In the AIDS model, $\gamma_{ij} = \gamma_{ji}$ represents the condition that reflects the Slutsky symmetry, ensuring that cross-price effects are symmetric. This condition implies that changes in relative prices translate into households' budget allocations where the relationships between goods should be consistent in magnitude (Deaton and Muellbauer, 1980).

Specifically, the Slutsky symmetry ensures that two goods are either substitutes, complements, or unrelated to each other to the same extent in terms of cross-price effects (see Deaton and Muellbauer (1980)). In our approach, we take this a step further by assuming that $\epsilon_{ij} = \epsilon_{ji}$, where ϵ_{ij} denotes the cross-price elasticity of good i with respect to the price of good j. This additional assumption ensures that in our estimation procedure the relationship between any two goods is mutual, meaning that the cross-price elasticities are the same in both directions. By imposing the restriction $\epsilon_{ij} = \epsilon_{ji}$, we also ensure that the magnitude of the relationship between the goods is symmetric.[18]

As we want to measure the price effect of print books on the demanded quantity of e-books, an OLS estimation approach would probably lead to biased estimators due to reverse causality between e-book quantities and the prices of related print formats. Thus, we apply an IV estimation approach in our empirical analysis. In order to get an unbiased and consistent estimator for the elasticities, it is essential to have exogenous variation that can distinguish demand- from supply-side effects. In this cross-sectional data set, identification relies on two instruments (see below): the number of pages of

---

[18]We will present the results of both an unrestricted and restricted version of 3SLS in Section 4.2.

a print book title and the digital size (in KB) of the matched e-book, both of which remain constant over time. Both instruments should be correlated with the prices of the respective book formats (relevance condition), but should not have a partial effect on the sold quantity of a given book title (orthogonality condition).

At first, we apply a 2SLS estimation approach to estimate own-format and cross-format elasticities. We run this estimation for the e-book/paperback as well as the e-book/hardcover subsamples of Germany and United Kingdom, which means that we have two regressions with two endogenous variables in the second stage, respectively. Thus, in the first stage we estimate two different regressions for every subsample, which can be formalized as follows:

$$P_{E,i} = \alpha_0 + \alpha_1 logKindle_{E,i} + \alpha_2 logPages_{P,i} + W\theta + \nu_i, \tag{2}$$

$$P_{P,i} = \beta_0 + \beta_1 logKindle_{E,i} + \beta_2 logPages_{P,i} + W\Psi + \epsilon_i. \tag{3}$$

Thereby, the dependent variables are the logged retail prices for e-books ($P_{E,i}$) and the respective print (hardcover or paperback) format ($P_{P,i}$) of a book title i. These prices are explained by the digital size of the e-book (variable $Kindle$) and the number of pages of the respective print format ($Pages$). All other book-specific covariates are included in the matrix W.

Making use of the first stages, we estimate two structural equations of our IV approach to estimate own-format and cross-format elasticities which take the following form, respectively:

$$Q_{E,i} = \gamma_0 + \gamma_1 \hat{P_{E,i}} + \gamma_2 \hat{P_{P,i}} + W\Omega + \eta_i, \tag{4}$$

$$Q_{P,i} = \gamma_3 + \gamma_4 \hat{P_{E,i}} + \gamma_5 \hat{P_{P,i}} + W\Omega + \mu_i. \tag{5}$$

In equations (4) and (5), the dependent variables $Q_{E,i}$ and $Q_{P,i}$ are the demanded quantity of e-book and print book title i. The fitted values from the first stage are captured by $\hat{P_{E,i}}$ and $\hat{P_{P,i}}$. Beside the coefficients γ_1 and γ_5 which give the own-format elasticities in both subsamples, the coefficient of interest in equations (4) and (5) are γ_2 and γ_4, which display the cross-format elasticity between e-books and the print format. Hence, a significant negative sign of, i.e. γ_2 would imply that e-books and print formats (paperback or hardcover) are complements, whereas a significant positive sign would suggest that digital and print formats are substitutes.

Estimating equations (4) and (5) solely might not be efficient and can also yield biased estimates. Both equations can be estimated as an equation system where we can also impose restrictions on the parameters. Therefore, we will also estimate the following regressions:

$$Q_{E,i} = \gamma_0 + \gamma_1 \hat{P_{E,i}} + \gamma_2 \hat{P_{P,i}} + W\Omega + \eta_i, \tag{6}$$

$$Q_{P,i} = \gamma_3 + \gamma_4 \hat{P_{E,i}} + \gamma_5 \hat{P_{P,i}} + W\Omega + \mu_i, \tag{7}$$

where the cross-format elasticity between e-books and the respective print format is restricted to be $\gamma_2 = \gamma_4$. Hence, we impose parameter restrictions across equations, which is possible within this SUR model. Moreover, the error terms η_i and μ_i are assumed to be correlated across the equations and endogeneity of sales and prices are still tackled by an IV approach. Therefore, we estimate equations (6) and (7) employing a 3SLS regression model in the style of Zellner and Theil (1962).[19]

---

[19]The results of 2SLS are nearly identical to 3SLS without any parameter restrictions. Thus, we will only present results for 3SLS with the imposition of parameter restrictions in Section 4.2.

4.2 Main Results

In the following, we will present the results of our 2SLS and 3SLS estimation approaches. As already explained in Section 4.1, the goal of our empirical analysis is to find out whether digital and print formats are substitutes or rather complements (or maybe are unrelated). Hence, we want to estimate cross-format elasticities using demanded quantities of book titles calculated based on the Amazon sales ranks (see Section 3.2 for details on the translation of sales ranks into quantities). Moreover, we run the IV estimation approach for the e-book/paperback and the e-book/hardcover subsamples using both 2SLS and 3SLS.[20]

Tables 5a and 5b present the results for our estimated cross-format elasticities in Germany and UK for the e-book/paperback (Table 5a) and e-book/hardcover (Tables 5b) subsamples.[21] The dependent variable is the logged quantity of e-books and the treatment variables are the logged retail prices for e-books (own-format elasticity) as well as the logged retail prices for the print formats (cross-format elasticity) (see equations (4)-(7)).[22] The first columns in both tables indicate the employed regression model (2SLS or 3SLS) and the respective country (Germany or UK). The following columns show the results for the own-format elasticity of e-books (OPF) (γ_1), the cross-format elasticities (CRFE) between the respective print format and the e-book (γ_2 and γ_4), and the OPE for the print format (γ_5). With respect to the shape parameters, we present the results for $\theta = 0.8$ and $\theta = 1.2$ (see Section 3.2 for more information on the shape parameter).

Remember that we expect an endogeneity issue in these regressions due to the reverse causality between book quantities and prices, so that the e-book prices are instrumented by the (logged) digital size of an e-book and the print prices are instrumented by the (logged) number of pages.

The estimation results for the paperback/e-book subsample in Table 5a exhibit a negative own-format elasticity for both book formats indicating that the product *book* generally is no Giffen good and follows the law of demand (see columns (I) and (IV)). Most of the estimated coefficients for the own-format elasticities are in absolute values larger than 1, which means that the demand for print and digital books in our sample is elastic.[23]

Shape Model/Country/DSet	OPE E-Book (I) $\kappa = 0.8$	$\kappa = 1.2$	CRFE E-Book to Paperback (II) $\kappa = 0.8$	$\kappa = 1.2$	CRFE Paperback to E-Book (III) $\kappa = 0.8$	$\kappa = 1.2$	OPE Paperback (IV) $\kappa = 0.8$	$\kappa = 1.2$
2-SLS/GER	-2.2247 (0.36)	-1.4831 (0.24)	3.0613 (0.42)	2.0409 (0.28)	3.9747 (0.42)	2.6498 (0.28)	-4.0695 (0.46)	-2.7130 (0.31)
2-SLS/UK	-3.3114 (0.46)	-2.2076 (0.31)	4.6439 (0.58)	3.0959 (0.39)	2.5923 (0.46)	1.7282 (0.31)	-3.4884 (0.56)	-2.3256 (0.37)
3-SLS/GER	-2.5906 (0.32)	-1.7271 (0.21)	3.5283 (0.31)	2.3522 (0.21)	3.5283 (0.31)	2.3522 (0.21)	-3.6711 (0.41)	-2.4474 (0.27)
3-SLS/UK	-2.5560 (0.33)	-1.7040 (0.22)	3.6739 (0.37)	2.4493 (0.25)	3.6739 (0.37)	2.4493 (0.25)	-4.6242 (0.53)	-3.0828 (0.35)

(a) E-Book/Paperback subsample

Shape Model/Country/DSet	OPE E-Book (I) $\kappa = 0.8$	$\kappa = 1.2$	CRFE E-Book to Hardcover (II) $\kappa = 0.8$	$\kappa = 1.2$	CRFE Hardcover to E-Book (III) $\kappa = 0.8$	$\kappa = 1.2$	OPE Hardcover (IV) $\kappa = 0.8$	$\kappa = 1.2$
2-SLS/GER/Hard	-0.9033 (0.35)	-0.6022 (0.23)	1.7375 (0.50)	1.1583 (0.34)	3.2472 (0.52)	2.1648 (0.35)	-3.6426 (0.75)	-2.4284 (0.50)
2-SLS/UK/Hard	-3.1122 (0.46)	-2.0748 (0.31)	5.0756 (0.67)	3.3837 (0.44)	2.4817 (0.36)	1.6545 (0.24)	-2.4848 (0.47)	-1.6565 (0.31)
3-SLS/GER/Hard	-1.5488 (0.29)	-1.0325 (0.19)	2.6894 (0.38)	1.7930 (0.25)	2.6894 (0.38)	1.7930 (0.25)	-3.0038 (0.59)	-2.0026 (0.39)
3-SLS/UK/Hard	-2.1707 (0.31)	-1.4471 (0.21)	3.5886 (0.36)	2.3924 (0.24)	3.5886 (0.36)	2.3924 (0.24)	-3.5801 (0.56)	-2.3867 (0.38)

(b) E-Book/Hardcover subsample

Table 5: Own-format and cross-format elasticities for the subsamples e-book/paperback and e-book/hardcover for Germany and the UK. Robust standard errors in parenthesis.

The estimated cross-format elasticities in Table 5a (see columns (II) and (III)) all have a positive sign pointing out that paperback books are rather substitutes for

---

[20]We present the results for naive OLS regressions to estimate the elasticities in Appendix F.

[21]The results for all first stages of the IV-approaches are presented in Appendix E.

[22]One exception in Tables 5a and 5b is column (IV), where we estimate the own-format elasticity for the print format so that the dependent variable is the (logged) quantity of the paperback book and the treatment variable is its (logged) (own) retail price.

[23]Own-format elasticities for books mostly exhibit high values due to the high number of substitution possibilities. Beyond, the estimated own-format elasticities for our sample are oftentimes larger than the ones calculated in previous studies because those studies only incorporate bestselling titles. In our sample, we also consider the "long tail", which increases the elasticities (in absolute terms) since the demand for "long tail" books is normally more elastic than the one for bestselling books.

digital books (and vice versa). For instance, the estimated coefficient for the German 2SLS-regression and the shape parameter $\theta = 1.2$ (2.0409) indicates that the demand for the e-book version of book title i increases by approx. 2% if the price of the related paperback version rises by 1%.

The results for the hardcover/e-book subsample in Table 5b are quite similar. The own-format elasticities for e-books (column (I)) as well as for the hardcover format (column (IV)) all have a negative sign and are mostly larger than 1 in absolute values, which means that also the demand for hardcover books in our sample is elastic. The estimated cross-format elasticities in Table 5b (see column (II) and (III)) all have a positive sign pointing out that also hardcover books are rather substitutes than complements for digital books (and vice versa) in both countries.

Two further findings are worth mentioning here. First, the demand for e-books seems to react more strongly when there is a price increase for the print formats in UK compared to Germany (see columns (II) in Tables 5a and 5b). This may be due to the unregulated book prices in the UK, which should (*c.p.*) lead to a more volatile demand for books in general. Second, whereas the cross-format elasticities in the German subsamples are larger for the paperback than for the hardcover version, we find the exact opposite for the British subsample. Most likely this is the case due to different consumer preferences in Germany and the UK and/or due to quality differences regarding the print formats in these two countries. The print and cover quality of the German paperback version is typically higher compared to the British one, so that more British readers probably consider the hardcover version as a better substitute for a digital book.[24]

In another approach, we analyze whether there the estimated elasticities in UK and Germany are different with respect to the two book genres fiction and non-fiction. Based on our LDA topic model, we split our data set into one fiction and one non-fiction sample and repeat the 3-SLS regression approach from Table 5. The corresponding results are presented in Table 6.

Genres	Shape Model/Country/DSet	OFE E-Book (I)		CRFE E-Book to Paperback (II)		CRFE Paperback to E-Book (III)		OFE Paperback (IV)	
		$\kappa = 0.8$	$\kappa = 1.2$	$\kappa = 0.8$	$\kappa = 1.2$	$\kappa = 0.8$	$\kappa = 1.2$	$\kappa = 0.8$	$\kappa = 1.2$
Fiction	3-SLS/GER/Paper	-4.4415 (3.30)	-2.9610 (2.20)	5.1574 (2.88)	3.4383 (1.92)	5.1574 (2.88)	3.4383 (1.92)	-5.1115 (2.70)	-3.4076 (1.80)
	3-SLS/UK/Paper	-3.1718 (0.71)	-2.1145 (0.47)	4.3202 (0.58)	2.8802 (0.39)	4.3202 (0.58)	2.8802 (0.39)	-5.1761 (0.71)	-3.4507 (0.48)
Non-Fiction	3-SLS/GER/Paper	-2.4661 (0.27)	-1.6441 (0.18)	3.3248 (0.27)	2.2166 (0.18)	3.3248 (0.27)	2.2166 (0.18)	-3.7171 (0.39)	-2.4780 (0.26)
	3-SLS/UK/Paper	-1.8357 (0.40)	-1.2238 (0.26)	2.5909 (0.56)	1.7273 (0.37)	2.5909 (0.56)	1.7273 (0.37)	-3.1491 (0.83)	-2.0994 (0.55)

(a) E-book/Paperback Subsample

Genres	Shape Model/Country/DSet	OFE E-Book (I)		CRFE E-Book to Hardcover (II)		CRFE Hardcover to E-Book (III)		OFE Hardcover (IV)	
		$\kappa = 0.8$	$\kappa = 1.2$	$\kappa = 0.8$	$\kappa = 1.2$	$\kappa = 0.8$	$\kappa = 1.2$	$\kappa = 0.8$	$\kappa = 1.2$
Fiction	3-SLS/GER/Hard	-2.2936 (0.74)	-1.5291 (0.50)	4.1532 (0.91)	2.7688 (0.61)	4.1532 (0.91)	2.7688 (0.61)	-4.6343 (1.41)	-3.0896 (0.94)
	3-SLS/UK/Hard	-1.9580 (0.64)	-1.3053 (0.43)	4.6178 (0.58)	3.0785 (0.39)	4.6178 (0.58)	3.0785 (0.39)	-4.2490 (1.02)	-2.8327 (0.68)
Non-Fiction	3-SLS/GER/Hard	-1.7246 (0.43)	-1.1497 (0.28)	2.4189 (0.53)	1.6126 (0.35)	2.4189 (0.53)	1.6126 (0.35)	-2.6043 (0.71)	-1.7362 (0.48)
	3-SLS/UK/Hard	-1.7673 (0.31)	-1.1782 (0.20)	2.5259 (0.39)	1.6840 (0.26)	2.5259 (0.39)	1.6840 (0.26)	-2.3880 (0.56)	-1.5920 (0.37)

(b) E-book/Hardcover Subsample

Table 6: Own-format and cross-format elasticities by fiction and Non-fiction for the subsamples e-book/paperback and e-book/hardcover for Germany and the UK. Robust standard errors in parenthesis.

The genre-specific estimation results in Table 6 first confirm two results from above, whereas the demand for all three book formats in our data set is elastic and digital and print books are (imperfect) substitutes (both independent of the specific genre). With respect two the two genres fiction and non-fiction, the estimated own-price elasticities (columns (I) and (IV)) imply that the demand for fiction books is more elastic than the one for non-fiction in both countries.

The estimated cross-format elasticities in Table 6 (columns (II) and (III)) suggest that for consumers the substitutability between digital and print book formats is stronger for fiction than for non-fiction books, independently of the specific print format (paperback or hardcover) and country. For instance, the estimated coefficient

---

[24]Both results are also visible in Table D.1 of Appendix D, where we present the ratios for the estimation results between the British and the German subsamples.

for the German e-book/paperback regression and the shape parameter $\theta = 0.8$ (see column (II) in Table 6a), which is 5.16, indicates that the demand for a fiction e-book increases by approx. 5.2% if the price of the related paperback version rises by 1% (on average). In contrast, the corresponding demand effect for a non-fiction e-book is only 3.3% on average.

Finally, the genre-specific results in Table 6 show that the finding for the UK from above, whereas hardcover books seem to be a better substitute for e-books than paperbacks, only holds for the genre fiction. With respect to non-fiction books, the estimated cross-format elasticities are slightly higher (on amount) for paperback than for hardcover books, which is the same like in Germany.

5 Discussion

Our results have important implications for policymakers as well as for the individual market players. Since we find that e-books are (imperfect) substitutes for print books from a consumers' perspective, an adequate pricing strategy for digital books seems to be very important for publishers. Setting too low prices for digital books might cannibalize the sales model of more profitable hardcover books.

The topic of setting e-book prices has also been investigated by the Department of Justice (DOJ) in the US antitrust case *United States v. Apple Inc. (2012)*, in which the Court ruled that Apple conspired to raise the price of e-books for five of the six major publishers (the "Big Six"). In the initial period following the introduction of the Kindle e-reader by Amazon in 2007, e-books have been sold under the classic wholesale model and Amazon pursued a low price strategy for e-books (e.g., \$9.99 for newly released e-books). The publishers were against this pricing strategy as they believed that the low price level might be a problem for their hardcover book sales and the consumer perceptions of the value of a book title (Li, 2019). They used several different strategies to fight against Amazon's pricing model, e.g. by "windowing" new releases (delaying the release of the e-book format).

In 2010, major publishers negotiated agency agreements (in combination with MFN clauses) with Apple to offer e-books in Apple's new iBookstore after the introduction of the iPad. The evidence on the effect of the agency agreements on the consumers prices for e-books is ambiguous so far. While De los Santos and Wildenbeest (2017) find a significant positive effect on prices for bestselling titles, the results of Gail and Klotz (2021) suggest that e-books sold under agency contracts are lower on average, which is especially driven by middle- and long tail book titles. Irrespective of the likely economic effect, the DOJ sued Apple and five of the Bix Six publishers in 2012 for fixing e-book prices.[25].

The results of this paper contribute to this discussion by implying that digital and print book formats are (imperfect) substitutes, which might cause a cannibalization of print book sales when setting e-book prices too low. Hence, publishers could lose a large part of their print book sales when the digital format of a title is sold significantly lower so that a certain share of readers decides to buy the digital instead of the print format. Publishers as well as authors should have an interest that digital and print book prices are relatively equal, not least to secure the consumer perceptions of the value of a book. Thus, strategies like "windowing" or keeping the pricing power should be used to hinder decreasing e-book prices.

From a policymaker perspective, our results suggest treating digital and print books uniformly because they are perceived as (imperfect) substitutes by the consumers. Thus, policymakers should promote the consumption of books independently of the individual format since otherwise biased consumption decisions could arise in this market. For instance, four countries of the OECD have a fixed book price system only for print books, but leave the price of e-books free (Poort and van Eijk, 2017). This can lead to a situation where the (fixed) print books prices are relatively high compared to the digital book prices because the retailers could rebate the e-books and, thus, there could be a cannibalization effect. Another policy question relates to the

---

[25]See United States v. Apple Inc., 12 Civ. 2826 (DLC)

taxation of different book formats. Several countries have implemented different VAT rates for print and digital books, whereby print books are mostly taxed with a lower rate and digital books with the normal VAT rate (Colbjørnsen, 2015). Even though our main finding, namely the substitutability of print and digital books, would suggest a uniform taxation of the individual formats, a lower VAT rate for print books could counteract the mentioned cannibalization effect at the same time.

6 Conclusion

The results of this paper imply that both print book formats (paperback and hardcover) are (imperfect) substitutes for digital books in Germany as well as in the UK. The estimated cross-format elasticities are higher for the British than for the German sample, which very likely reflects the higher demand volatility in a country with unregulated prices. We also find differences in the format preferences between British and German readers: while e-books seem to be better substitutes for paperback books in Germany (compared to hardcover ones), we find the exact opposite for the UK. Finally, our results suggest that the substitutability between digital and print formats is stronger for fiction than for non-fiction books in general. All these results are robust for a 2SLS IV estimation as well as for a 3SLS regression approach.

To the best of our knowledge, this paper is the first empirical analysis estimating cross-format elasticities between digital and print books. Since we find that consumers consider both print versions (paperback and hardcover) as a (imperfect) substitute for digital books in two countries with completely different market regulations, our results suggest taxing print and digital books uniformly. Different VAT rates for print and digital books can lead to biased consumption decisions and, therefore, might increase the deadweight loss in the affected book market. Moreover, the relative strong substitutability between digital and print books might cause a cannibalization of print book sales when setting e-book prices too low. Hence, strategies like "windowing" should be used by the publishers to prevent this cannibalization.

This research is limited because we cannot control for dynamic effects in our cross-sectional data set. On the one hand, cross-format elasticities might vary over time due to changing consumer preferences and, on the other hand, publishers release the various formats of a book title at different points in time to skim a higher share of the consumer's willingness to pay (second-degree price discrimination). Moreover, we have only incorporated the online retailer Amazon into our empirical analysis. Even though this provides a large market coverage of book sales, it would be interesting to examine a similar analysis for other book retailers. Thus, future research should concentrate on panel data and several book retailers to address dynamic as well as cross-retailer effects in the book market.

Acknowledgement(s)

Declarations

Authors' contributions:
Maximilian Maurice Gail: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Review & Editing.
Phil-Adrian Klotz: Conceptualization, Methodology, Validation, Formal analysis, Writing – Original Draft, Supervision.
Data Availability Statement: Data will be made available on request.

References

Barrot C, Becker JU, Clement M, et al (2015) Price elasticities for hardcover and paperback fiction books. Schmalenbach business review 67(1):73–91

Beck J (2004) Fixed, focal, fair? book prices under optional resale price maintenance. Book Prices Under Optional Resale Price Maintenance (December 2004) WZB, Markets and Political Economy Working Paper No SP II 15

Bittlingmayer G (1992) The elasticity of demand for books, resale price maintenance and the lerner index. Journal of institutional and theoretical economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft pp 588–606

Brundy JM, Jorgenson DW (1971) Efficient estimation of simultaneous equations by instrumental variables. The Review of Economics and Statistics pp 207–224

Chen H, Hu YJ, Smith MD (2019) The impact of e-book distribution on print sales: analysis of a natural experiment. Management Science 65(1):19–31

Chen L, King RC (2017) To lend is to own: A game theoretic analysis of the e-book lending market. International Journal of Electronic Commerce 21(3):334–362

Chevalier J, Goolsbee A (2003) Measuring prices and price competition online: Amazon. com and barnesandnoble. com. Quantitative marketing and Economics 1(2):203–222

Clerides SK (2002) Book value: intertemporal pricing and quality discrimination in the us market for books. International Journal of Industrial Organization 20(10):1385–1408

Colbjørnsen T (2015) What is the vat? the policies and practices of value added tax on ebooks in europe. International journal of cultural policy 21(3):326–343

Crosby P (2019) Don't judge a book by its cover: examining digital disruption in the book industry using a stated preference approach. Journal of Cultural Economics 43(4):607–637

Dang Nguyen G, Dejean S, Moreau F (2012) Are streaming and other music consumption modes substitutes or complements? Sylvain and Moreau, François, Are Streaming and Other Music Consumption Modes Substitutes or Complements

Deaton A, Muellbauer J (1980) An almost ideal demand system. The American economic review 70(3):312–326

Gail MM, Klotz PA (2021) The impact of the agency model on e-book prices: Evidence from the uk. Tech. rep., MAGKS Joint Discussion Paper Series in Economics

Gilbert RJ (2015) E-books: A tale of digital disruption. Journal of Economic Perspectives 29(3):165–84

Goetz G, Herold D, Klotz PA, et al (2020) The substitutability between brick-and-mortar stores and e-commerce-the case of books. Tech. rep., MAGKS Joint Discussion Paper Series in Economics

Gong J, Smith MD, Telang R (2015) Substitution or promotion? the impact of price discounts on cross-channel sales of digital movies. Journal of retailing 91(2):343–357

Grace RC, Kivell BM, Laugesen M (2014) Estimating cross-price elasticity of e-cigarettes using a simulated demand procedure. Nicotine & Tobacco Research 17(5):592–598

Guillon O, Thierry C (2013) Is ebook pricing structured to mirror paper book prices? the cases of france and the usa in 2011. Canadian journal of information and library science 37(3)

Hao L, Fan M (2014) An analysis of pricing models in the electronic book market. MIS quarterly 38(4):1017–1032

Hashim MJ, Ram S, Tang Z (2019) Uncovering the effects of digital movie format availability on physical movie sales. Decision Support Systems 117:75–86

Larsen VH, Thorsrud LA (2019) The value of news for economic developments. Journal of Econometrics 210(1):203–218

Lee E, Lee B (2013) Changing price elasticity of digital goods: Empirical study from the e-book industry. AIS Electronic Library (AISeL) URL https://core.ac.uk/outputs/301361227/?utm_source=pdf&utm_medium=banner&utm_campaign=pdf-decoration-v1

Lenz D, Winker P (2020) Measuring the diffusion of innovations with paragraph vector topic models. PloS one 15(1):e0226685

Li H (2013) The impact of ebooks on print book sales: Cannibalization and market expansion. Unpublished manuscript Retrieved from https://pdfs semanticscholar org/1b75/6e39e390de202d33756984028e6465970524 pdf

Li H (2019) Are e-books a different channel? multichannel management of digital products. Multichannel Management of Digital Products (August 20, 2019)

Meng Y, Brennan A, Purshouse R, et al (2014) Estimation of own and cross price elasticities of alcohol demand in the uk—a pseudo-panel approach using the living costs and food survey 2001–2009. Journal of health economics 34:96–103

Michel NJ (2006) The impact of digital file sharing on the music industry: An empirical analysis. The BE Journal of Economic Analysis & Policy 6(1)

OECD (2012) E-books: developments and policy considerations. OECD Publishing

Ogwang T, Cho DI (2009) Economic determinants of the consumption of alcoholic beverages in canada: a panel data analysis. Empirical Economics 37(3):599–613

Poort J, van Eijk N (2017) Digital fixation: the law and economics of a fixed e-book price. International Journal of Cultural Policy 23(4):464–481

Rainie L, Zickuhr K, Purcell K, et al (2012) The rise of e-reading. Pew Internet & American Life Project

Rao SS (2005) Electronic books: their integration into library and information centers. The electronic library

De los Santos B, Wildenbeest MR (2017) E-book pricing and vertical restraints. Quantitative Marketing and Economics 15(2):85–122

Singer LM, Alexander PA (2017) Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. The journal of experimental education 85(1):155–172

Stevans LK, Sessions DN (2005) An empirical investigation into the effect of music downloading on the consumer expenditure of recorded music: A time series approach. Journal of Consumer Policy 28(3):311–324

Tomlinson P, Branston JR (2014) The demand for uk beer: estimates of the long-run on-and off-trade beer price elasticities. Applied Economics Letters 21(3):209–214

Wang K, Goldfarb A (2017) Can offline stores drive online sales? Journal of Marketing Research 54(5):706–719

Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American statistical Association 57(298):348–368

Zellner A, Theil H (1962) Three-stage least squares: Simultaneous estimation of simultaneous equations. Econometrica 30(1):54–78. URL http://www.jstor.org/stable/1911287

Zhang Y, Kudva S (2013) Ebooks vs. print books: Readers' choices and preferences across contexts. Proceedings of the American Society for Information Science and Technology 50(1):1–4

Appendix

A LDA Approach

Topic	Genre
0	Philosophy
1	Textbooks
2	Childen and Youth
3	History
4	Guidebook
5	General Fiction
6	Fantasy Fiction
7	Personal development
8	Cultural
9	Self-Guidance
10	Thriller
11	Historic Novel
12	Social Novel
13	Crime Novel
14	Politics

Table A.1: 15 different genres identified by our LDA approach for the UK.

Figure A.1: Topic Crime Novel from the LDA (UK sample). Size is according to weight of word within the topic.

Figure A.2: Topic Thriller from the LDA (UK sample). Size is according to weight of word within the topic.
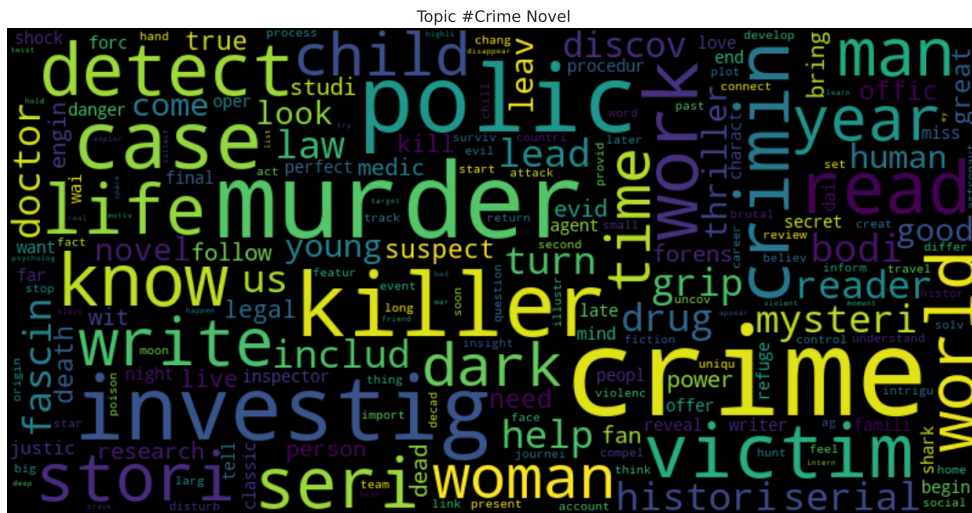

Figure A.3: Topic Politics from the LDA (UK sample). Size is according to weight of word within the topic.

B Data Set Construction

The data generating process is structured as follows. We have scraped the Amazon.co.uk and Amazon.de web pages for a list of publisher and imprint names to get retail prices for the three formats hardcover, paperback and e-book as well as further book characteristics available on the website. For creating this data set, we use *a priori* a list of publishing houses, publishers and imprints which is taken from historical bestseller lists from Germany and the UK. This procedure ensures that our data set only contains books from publishers with a relatively large market size.

With this list of publishers, we have first searched for e-books published between January 2019 and March 2020. After that, we have updated the original list, since the respective publisher names on Amazon are often written in different ways but the search itself must match every character exactly. Following this, the data set with publisher names has risen. With this newly obtained list, the actual data set is constructed on the above mentioned time frame by searching for all publisher names of the updated list. This procedure also incorporates books into our data set which have been published before 2019 since we have done the publisher search on Amazon independently of the format. Hence, it may have happened that for a book title we have found within our observation period another format of the same title has been published years before. However, we have ensured that no e-book is included in our working data set which has been published before March 2010.

All book titles on Amazon have different ASINs (Amazon Standard Identification Number), which leads to an identification problem for different format types of the same book. To solve this problem, we first extract as much data from the web page of a book as possible. Then, we open the list of all formats from the observed book title available on Amazon and extract the information about them, which only contain their prices. Unfortunately, in some cases different editions of a book title are sold by different publishers, which requires us to be even more precise for the identification of those books. Therefore, we take the newest, oldest and cheapest version of the format list and scrap detailed information for them. To prevent data failures, the ASINs of those three versions are combined to create a unique identifier. Finally, within the data conversion we only use the newest book of each format when there are various editions available.

Our raw data set consists of roughly one million observations, whereby each entry contains several information on different prices, formats, descriptions, ratings, reviews etc. being available on the Amazon website. Besides, for every book title there are three entries if all formats (hardcover, paperback, e-book) are available for the respective title. However, due to the usage of web-scraping methods the data set contains many entries that are duplicates or not of interest for our analysis. Hence, after the data cleansing process our final working data set consists of 8,315 unique e-book titles for the British sample and 2,544 e-book titles for the German one. We only use observations in our estimation approach for which all explanatory variables (book characteristics) are available.

C Descriptive Statistics

	Kindle Price in British Pound (I)	Kindle Price in British Pound (II)	Kindle Price in Euro (III)	Kindle Price in Euro (IV)
Paperback Price	0.6180		0.7332	
	(0.0165)		(0.0164)	
Hardcover Price		0.2208		0.5487
		(0.0162)		(0.0555)
Num.Obs.	6539	5506	2239	1516
R2	0.660	0.453	0.825	0.656
R2 Adj.	0.658	0.450	0.822	0.649
Covariates	Yes	Yes	Yes	Yes
Publisher Effects	Yes	Yes	Yes	Yes
Topic Effects	Yes	Yes	Yes	Yes
Data	UK	UK	GER	GER

Standard errors in parenthesis.

Table C.1: Regression of e-book prices on paperback or hardcover prices for Germany and the UK using the respective subsamples e-book/paperback and e-book/hardcover.
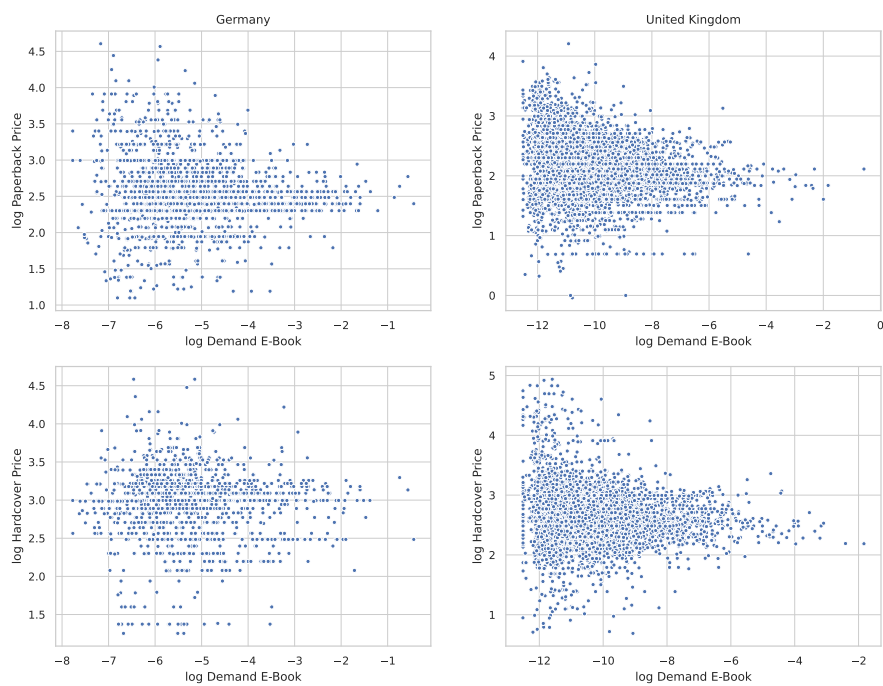

Figure C.1: Demand of e-Books in Germany (left-hand side) and United Kingdom (right-hand side) for the e-book/paperback (above) and e-book/hardcover (below) subsamples.

D Ratios of elasticities between Germany and the UK

Shape	OFE E-Book (I) $\kappa = 0.8$	$\kappa = 1.2$	CRFE E-Book to Paperback (II) $\kappa = 0.8$	$\kappa = 1.2$	CRFE Paperback to E-Book (III) $\kappa = 0.8$	$\kappa = 1.2$	OFE Paperback (IV) $\kappa = 0.8$	$\kappa = 1.2$
2-SLS/Paper	1.547311	1.547347	1.546884	1.546867	0.657360	0.657360	0.830117	0.830127
3-SLS/Paper	1.017313	1.017313	1.050797	1.050796	1.050797	1.050796	1.215935	1.215903

(a) E-book/Paperback Subsample

Shape	OFE E-Book (I) $\kappa = 0.8$	$\kappa = 1.2$	CRFE E-Book to Hardcover (II) $\kappa = 0.8$	$\kappa = 1.2$	CRFE Hardcover to E-Book (III) $\kappa = 0.8$	$\kappa = 1.2$	OFE Hardcover (IV) $\kappa = 0.8$	$\kappa = 1.2$
2-SLS/Hard	2.935207	2.935484	2.681955	2.682070	0.691301	0.691334	0.637521	0.637508
3-SLS/Hard	1.201738	1.201711	1.219161	1.219182	1.219161	1.219182	1.116235	1.116219

(b) E-book/Hardcover Subsample

Table D.1: Relative comparison of the estimated coefficients of UK and Germany for the e-book/paperback (top) and e-book/hardcover (bottom) subsamples.

E First Stage Regressions

	First Stage for E-Books log e-book price	log paperback price	First Stage for Paperback log paperback price	log e-book price
R-Squared	0.2832	0.3025	0.3025	0.2832
F Statistic	263.2520	448.8341	448.8341	263.2520
F P-Value	0.0000	0.0000	0.0000	0.0000
Wu-Hausman Statistic	103.2236	103.2236	19.3158	19.3158
Wu-Hausman P-Value	0.0000	0.0000	0.0000	0.0000
Wooldridge Test	173.3724	173.3724	34.5312	34.5312
Wooldridge P-Value	0.0000	0.0000	0.0000	0.0000

Table E.2: First stage statistics for equations (2) and (3) estimated on the British e-book/paperback sample.

	First Stage for E-Books log e-book price	log hardcover price	First Stage for Hardcover log e-book price	log hardcover price
R-Squared	0.3232	0.2770	0.3232	0.2770
F Statistic	256.3603	326.8567	256.3603	326.8567
F P-Value	0.0000	0.0000	0.0000	0.0000
Wu-Hausman Statistic	133.3216	133.3216	133.3216	133.3216
Wu-Hausman P-Value	0.0000	0.0000	0.0000	0.0000
Wooldridge Test	283.3577	283.3577	283.3577	283.3577
Wooldridge P-Value	0.0000	0.0000	0.0000	0.0000

Table E.3: First stage statistics for equations (2) and (3) estimated on the British e-book/hardcover sample.

| | First Stage for E-Books | | First Stage for Paperback | |
	log e-book price	log paperback price	log e-book price	log paperback price
R-Squared	0.4880	0.5157	0.4880	0.5157
F Statistic	183.5025	680.6766	183.5025	680.6766
F P-Value	0.0000	0.0000	0.0000	0.0000
Wu-Hausman Statistic	43.3043	43.3043	34.7960	34.7960
Wu-Hausman P-Value	0.0000	0.0000	0.0000	0.0000
Wooldridge Test	90.0448	90.0448	83.1638	83.1638
Wooldridge P-Value	0.0000	0.0000	0.0000	0.0000

Table E.4: First stage statistics for equations (2) and (3) estimated on the German e-book/paperback sample.

| | First Stage for E-Books | | First Stage for Hardcover | |
	log e-book price	log hardcover price	log hardcover price	log e-book price
R-Squared	0.5726	0.4686	0.4686	0.5726
F Statistic	212.1289	406.3235	406.3235	212.1289
F P-Value	0.0000	0.0000	0.0000	0.0000
Wu-Hausman Statistic	10.5542	10.5542	24.9485	24.9485
Wu-Hausman P-Value	0.0000	0.0000	0.0000	0.0000
Wooldridge Test	17.8763	17.8763	40.1768	40.1768
Wooldridge P-Value	0.0001	0.0001	0.0000	0.0000

Table E.5: First stage statistics equations (2) and (3) estimated on the German e-book/hardcover sample.

F Comparison of OLS to IV Results

Shape8	OLS E-Books	OLS Paperback	IV E-Books	IV Paperback
log e-book price	-0.4721 (0.0470)	0.4381 (0.0464)	-3.3114 (0.4599)	2.5923 (0.4613)
log paperback price	-0.0384 (0.0650)	-1.7024 (0.0809)	4.6439 (0.5777)	-3.4884 (0.5580)

Table F.1: OLS versus IV (2SLS) for log sales of E-Book and Paperback ($\theta = 0.8$) in UK using equations (4) and (5). Robust standard errors in parenthesis.

Shape12	OLS E-Books	OLS Paperback	IV E-Books	IV Paperback
log e-book price	-0.3147 (0.0313)	0.2921 (0.0309)	-2.2076 (0.3066)	1.7282 (0.3075)
log paperback price	-0.0256 (0.0433)	-1.1349 (0.0539)	3.0959 (0.3851)	-2.3256 (0.3720)

Table F.2: OLS versus IV (2SLS) for log sales of E-Book and Paperback ($\theta = 1.2$) in UK using equations (4) and (5). Robust standard errors in parenthesis.

Shape8	OLS E-Books	OLS Hardcover	IV E-Books	IV Hardcover
log e-book price	-0.2816 (0.0441)	0.7078 (0.0394)	-3.1122 (0.4584)	2.4817 (0.3573)
log hardcover price	-0.0175 (0.0567)	-1.3954 (0.0710)	5.0756 (0.6674)	-2.4848 (0.4662)

Table F.3: OLS versus IV (2SLS) for log sales of E-Book and Hardcover ($\theta = 0.8$) in UK using equations (4) and (5). Robust standard errors in parenthesis.

Shape12	OLS E-Books	OLS Hardcover	IV E-Books	IV Hardcover
log e-book price	-0.1878 (0.0294)	0.4719 (0.0263)	-2.0748 (0.3056)	1.6545 (0.2382)
log hardcover price	-0.0117 (0.0378)	-0.9303 (0.0473)	3.3837 (0.4449)	-1.6565 (0.3108)

Table F.4: OLS versus IV (2SLS) for log sales of E-Book and Hardcover ($\theta = 1.2$) in UK using equations (4) and (5). Robust standard errors in parenthesis.

Shape8	OLS E-Books	OLS Paperback	IV E-Books	IV Paperback
log e-book price	0.0581 (0.0602)	1.1428 (0.0675)	-2.2247 (0.3613)	3.9747 (0.4206)
log paperback price	0.3171 (0.0899)	-1.0626 (0.1134)	3.0613 (0.4237)	-4.0695 (0.4635)

Table F.5: OLS versus IV (2SLS) for log sales of E-Book and Paperback ($\theta = 0.8$) in Germany using equations (4) and (5). Robust standard errors in parenthesis.

Shape12	OLS E-Books	OLS Paperback	IV E-Books	IV Paperback
log e-book price	0.0387 (0.0401)	0.7619 (0.0450)	-1.4831 (0.2409)	2.6498 (0.2804)
log paperback price	0.2114 (0.0599)	-0.7084 (0.0756)	2.0409 (0.2825)	-2.7130 (0.3090)

Table F.6: OLS versus IV (2SLS) for log sales of E-Book and Paperback ($\theta = 1.2$) in Germany using equations (4) and (5). Robust standard errors in parenthesis.

Shape8	OLS E-Books	OLS Hardcover	IV E-Books	IV Hardcover
log e-book price	0.3642 (0.0607)	0.9168 (0.0621)	-0.9033 (0.3475)	3.2472 (0.5216)
log hardcover price	0.0068 (0.1090)	-1.1164 (0.1217)	1.7375 (0.5044)	-3.6426 (0.7509)

Table F.7: OLS versus IV (2SLS) for log sales of E-Book and Hardcover ($\theta = 0.8$) in Germany using equations (4) and (5). Robust standard errors in parenthesis.

Shape12	OLS E-Books	OLS Hardcover	IV E-Books	IV Hardcover
log e-book price	0.2428 (0.0405)	0.6112 (0.0414)	-0.6022 (0.2317)	2.1648 (0.3477)
log hardcover price	0.0045 (0.0727)	-0.7443 (0.0812)	1.1583 (0.3363)	-2.4284 (0.5006)

Table F.8: OLS versus IV (2SLS) for log sales of E-Book and Hardcover ($\theta = 1.2$) in Germany using equations (4) and (5). Robust standard errors in parenthesis.