



HÖGSKOLAN
DALARNA

Dalarna Doctoral Dissertations 15

Voluntary Information Sharing, Retail Pricing and Firm Performance

CHARLIE LINDGREN

Microdata Analysis
School of Technology and Business Studies
Dalarna University, Borlänge, Sweden
2021

Dissertation presented at Dalarna University to be publicly examined in a digital seminar, Friday, 4 June 2021 at 13:00 for the Degree of Doctor of Philosophy. The examination will be conducted in English. Opponent: Associate Professor Christian Sandström (Jönköping International Business School).

Abstract

Lindgren, C. 2021. Voluntary Information Sharing, Retail Pricing and Firm Performance. *Dalarna Doctoral Dissertations* 15. Borlänge: Dalarna University. ISBN 978-91-88679-13-0.

Online information sharing by firms has created an unprecedented amount of data to analyze by researchers. While conclusions from research should be drawn with basis in sound theory, there has also emerged a need to supplement these theoretical considerations with advanced data collection, storage and analysis as well as reporting to decisionmakers. As such, the emergence of the research field of microdata analysis has given aid in the process of gathering large quantities of data and managing databases, analyzing said data with knowledge in advanced areas, e.g., statistical inference, machine learning, artificial intelligence and the like, and presenting the results for decisionmakers and stakeholders in a clear, coherent way while also stating economic consequences to enable decision-making. This dissertation consists of five individual papers contributing to this field of research, and in the process answering a set of questions related to voluntary information sharing, retail pricing, and firm performance.

In the first paper, a large-scale data collection of corporate social responsibility reports was coupled with state-of-the-art topic modelling analysis to answer the question who the intended users of these reports are, and the results show that the shareholder perspective is more prominent rather than the stakeholder perspective. The second paper empirically shows the value of having lowest price or highest ratings on a price comparison website, with the former being found to have a profound impact on demand, while the effect of the latter is more unclear. The third paper relies on time series clustering analysis to test if intertemporal price discrimination is the cause of remaining price dispersion in low search cost markets. The empirical evidence rejects an established theory of explaining price dispersion in a wide range of markets characterized by low search costs. The fourth paper provides an investigation into how the increased use of a price comparison website affect pricing. It is found that an increased use of the platform and number of retailers entering lead to a reduction in average prices with substantial consumer savings as the general outcome. Following the results of the third paper, a more likely model to explain the persistent price dispersion in low search cost markets is also presented. The fifth and final paper combines two datasets with rigorous statistical analysis to answer why firms compete on price comparison websites, despite the threat of increased competition and reductions in prices. The results show that retailers competing on price comparison websites increase their productivity which creates increased profits that are shared between shareholders and employees. The different types of information sharing studied in this thesis is thus found to have profound impact on consumers, firms and society at large.

Keywords: Voluntary information sharing, online retailing, e-commerce, price comparison websites, CSR, consumer search

Charlie Lindgren, School of Technology and Business Studies, Microdata Analysis

© Charlie Lindgren 2021

ISBN 978-91-88679-13-0

urn:nbn:se:du-36389 (<http://urn.kb.se/resolve?urn=urn:nbn:se:du-36389>)

Acknowledgements

The long and winding road to completing this PhD has been filled with people and experiences that I greatly cherish. First and foremost, I would like to thank my supervision team consisting of Siril Yella, Niklas Rudholm and Johan Håkansson. You all have had the patience of saints, and time and again have shared your wisdom and advice to give me new perspectives on what it means to be a researcher. I would also like to extend my deepest gratitude to Sven-Olof Daunfeldt who has contributed vastly to my work and given me the resources, time, and energy to elevate the quality of the papers. I also want to thank the team at *PriceSpy*, and in particular Magnus Bengtsson, who made this dissertation possible and gave me such a warm welcome each time I visited.

To my Dalarna University colleagues, I cherish you all so much! As research leader during large parts of my PhD studies as well as co-author, Kenneth Carling, your work was invaluable and passing the torch to Yves Rybarczyk has been equally excellent. Perhaps we could increase the difficulty of the medicine ball sessions a notch or two, although the “easy” level was an illusion! My other co-authors, Asif M. Huq, with your sense of calm and reason, and Yujiao Li, always with a clear sense of direction and determination, have both contributed greatly. Catia Cialani, your passion for teaching and research has inspired me throughout the years, and I always appreciated your excellent tiramisu and occasional pizza! To all my PhD colleagues, I thank you for every happy moment you have given me, and for sacrificing your precious time to endure my endless talks.

To my family, although none of you admittedly really understood what I was doing, you could still read me like an open book and support me throughout this journey. My father, I have yet to find a person who would not argue that your energy and optimism are anything but inspiring. My mother, your love, care, and sacrifice for others has no bounds. To all five of my sisters living throughout this long country, your care for our family is what holds us together so strongly.

Last, and certainly not least, thank you, Yujiao Li, the love of my life. From the day I first saw you in class as my statistics teacher to the present day I have never stopped loving you. You gave this otherwise boring guy the most inspiring, vivid, and happy life he could ever have imagined, and now a second home and family in China.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I **Lindgren, C.**, Huq, A. M., & Carling, K. (2021). Who Are the Intended Users of CSR Reports? Insights from a Data-Driven Approach. *Sustainability*, 13(3):1-20.
- II **Lindgren, C.** (2021) Discontinuities: What is the value of having the lowest price or highest rating on a price comparison website? *HFI Working Paper No. 19*.
- III **Lindgren, C.**, Daunfeldt, S. O., Rudholm, N., & Yella, S. (2020). Is intertemporal price discrimination the cause of price dispersion in markets with low search costs? *Applied Economics Letters*, 1-4.
- IV **Lindgren, C.**, Daunfeldt, S. O., & Rudholm, N. (2021) Pricing in retail markets with low search costs: Evidence from a price comparison website. *HFI Working Paper No. 18*.
- V **Lindgren, C.**, Li, Y., & Rudholm, N. (2021) Why do firms compete on price comparison websites? The impact on productivity, profits, and wages. Submitted to *The International Review of Retail, Distribution and Consumer Research*.

Reprints were made with permission from the respective publishers.

Related report (not included in dissertation)

Rudholm, N., & **Lindgren, C.** (2019) Prisspridning på e-handelsmarknader med låga sökkostnader/Price dispersion in e-tailing markets with low search costs (in Swedish). The Swedish Competition Agency, No. 1.

Contents

Introduction.....	9
Theoretical background	12
Data collection and quality	14
CSR sustainability disclosures	15
PriceSpy data.....	16
The Wayback Machine.....	17
Firm annual reports	18
Methods	20
Latent Dirichlet Allocation topic modelling.....	20
Poisson pseudo-maximum likelihood regression	21
Time series clustering with Dynamic Time Warping.....	21
Fixed effects panel data regression with marginal effects.....	22
Coarsened Exact Matching and difference-in-difference model.....	23
Summary of the papers	24
Paper I	24
Paper II.....	25
Paper III.....	26
Paper IV	28
Paper V	29
Decision support and policy implications.....	31
Discussion.....	34
References.....	38

Introduction

This dissertation investigates what impact online information sharing has had in society on a variety of interest groups, such as consumers, firms, policy-makers, and investors. All the studies presented were produced in the field of microdata analysis, which is a multidisciplinary field that encompasses collecting, assessing, and storing, often large quantities, of data on the micro level, as well as analyzing the data and reporting the outcome to support decision-makers. This is what is referred to as the chain of microdata analysis. The studies included in this dissertation are produced in close collaboration with the price comparison website *PriceSpy*, contact with the Global Reporting Initiative (GRI), and with financing from the Swedish Retail and Wholesale Council.

Globally, it is expected that 463 exabytes of data will be created every day by 2025 (“How Much Data Is Generated Each Day? | World Economic Forum” 2019), and as data is becoming more openly shared, the question arises: who will be the winners and losers as the age of open data takes off? (“The Value of Data | World Economic Forum” 2017). While governments are becoming more aware of the importance of leveraging open data, it also means that firms, consumers, and investors could have an interest in having large quantities of data analyzed and explained. It is therefore important for researchers to find ways to understand how to analyze information to provide decision support to these, and other, entities.

In Sweden, as an example, legally binding obligations exist even for smaller firms to provide annual financial reports with relevant income statements and balance sheets, while in the United States the requirements are much more relaxed. However, despite the more relaxed requirements, firms still often choose to produce reports, accept scrutiny from accounting agencies, and freely disseminate these reports to the public at an increasing rate. This increase is not only restricted to financial reports, but Corporate Social Responsibility (CSR) reports are also produced and disseminated at an increasing rate. These reports are often used to assess how a firm position itself as a sustainable company and how it reports progress towards attaining sustainability goals. US-domiciled assets under management using sustainable investing strategies grew from \$12.0 trillion at the start of 2018 to \$17.1 trillion at the start of 2020, or 33 percent of the total US assets under management (“The Forum for Sustainable and Responsible Investment” 2020). With such trends in mind, content analysis of CSR reports has increased as of late

(Russo-Spena, Tregua, and De Chiara 2018), and one important question in the sustainability research area is: who are the intended users of CSR reports?

Information sharing also occurs where there are no opposing legal requirements from governments or authorities, but where it remains to be answered why such information sharing takes place to begin with. Consider a market where firms compete by selling goods to consumers and have the choice to have brick-and-mortar stores, an online presence, or both. Given that a firm chooses to voluntarily participate in the online market, what consequences would this then have on pricing, firm performance, and the society surrounding this market? These are important questions that are of interest to competition authorities, firms, consumers, and other stakeholders. Take the e-commerce marketplace for instance, where there has been substantial growth over the past two decades, with the recent entry of Amazon in Sweden in late 2020, and an expansion of the number of firms competing on price comparison websites. The data generated in these marketplaces have entered the realm of Big Data, with a need for large-scale data collection, storage and quality assessment, analysis based on theory, and a clear and concise presentation for decision-makers and others. In short, the chain of microdata analysis is suitable to apply in this domain.

There are four datasets used in this study to empirically answer the questions posed: CSR disclosures collected across the globe; product, pricing, firm, and consumer choice-related variables collected from a price comparison website in Sweden; data collected on past entry dates of firms on a price comparison website; and financial data collected from Bisnode, which is a company that offers decision support in the form of digital business, marketing, and credit information. To study who the intended users of CSR disclosures are, a large-scale data collection of such disclosures from the GRI database for 1998 – 2017 was conducted. Data from the price comparison website *PriceSpy*, Sweden, were used for answering the questions: what is the value of having the lowest price or highest consumer rating on a price comparison website?; is intertemporal price discrimination the cause of price dispersion in markets with low search costs?; and how does the increased use of price comparison websites affect price levels and price dispersion? Finally, two datasets were combined to examine why firms choose to compete on a price comparison website, one of which relates to financial data as recorded in annual reports by Swedish firms and collected by Bisnode, while the other pertains to entry data collected using the Internet Archive, and, in particular, web scraping of the Wayback Machine (<http://web.archive.org/>).

Paper I identify who the intended users of CSR disclosures are using Latent Dirichlet Allocation topic modelling. The results show that firms take a shareholder approach to reporting, e.g., they have a focus on investors rather than stakeholders. As opposed to previous literature in the field, this is determined in a data-driven fashion, while other studies have heavily relied on subjective topic labelling.

Paper II studies how holding the lowest price or the highest consumer rating on the price comparison website *PriceSpy*, Sweden, affects demand. The paper is structured as a replication study of Baye et al. (2009), with the addition of an analysis of the value of having the highest consumer ratings. By using Poisson pseudo-maximum likelihood regression, it is found that a discontinuity in demand at the lowest price remains important to account for when calculating demand elasticities. Decision-makers need to account for this discontinuity, and which category of products are being considered, as this has implications, for example, in retailer pricing decisions or setting tax policies.

Paper III tests whether intertemporal price discrimination is the cause of price dispersion in markets with low search costs. This is accomplished by performing time series clustering with Dynamic Time Warping (Rath and Manmatha 2003) on data from *PriceSpy*, Sweden. Remaining price dispersion in low search cost markets is coupled with weak evidence of firms randomizing prices, and, instead, what is found is that firm price offerings are clustered into persistently low, mid-range and high prices over time. These results therefore suggest that it is unlikely that intertemporal price discrimination is the cause of price dispersion in markets with low search costs.

Paper IV investigates how increased use of the Swedish price comparison website *PriceSpy* affects pricing for a sample of consumer electronics and durable goods. It is shown that the increase in consumers using the price comparison website during the period under study reduced average prices, resulting in consumer savings, but that an even stronger impact on consumer savings came from increased competition due to more retailers entering the site.

Paper V answers the question of why firms compete on price comparison websites, despite the expected outcome of lower prices due to fierce competition. The empirical evidence shows that firms participating in the *PriceSpy* marketplace have higher productivity, operating profits and gross wages compared to similar firms that are not on the website.

Theoretical background

Disclosure of CSR reports, whether integrated with financial reports or not, may be viewed as a vital tool for informing stakeholders, as a legitimizing effort and providing information to investors, as well as mitigating the information asymmetry between managers and various stakeholders (Bondy, Moon, and Matten 2012; Hahn and Kühnen 2013; Thorne, Mahoney, and Manetti 2014). Since CSR reports are voluntary, unless otherwise mandated, they are also subject to firm flexibility and latitude in reporting content. This flexibility may result in a variation of quality in reporting (Brammer and Pavelin 2008) but also inconsistencies that would undermine the value of reports to stakeholders, including investors using the reports to analyze sustainability risks in their investment choices. To mitigate such issues, reporting frameworks have been adopted to standardize CSR reporting, such as the GRI Standards and Sustainability Accounting Standards Board (SASB) materiality map. However, these guidelines are bound to be at odds as the former is more directed towards stakeholders, a wide spectrum of users, while the latter is directed at shareholders, a narrower set of users considering mainly material issues. Firms may resort to either of these perspectives. For instance, firms may want to participate in CSR reporting to legitimize their position in society (Hahn and Kühnen 2013; Salganik, Dodds, and Watts 2006; Cho et al. 2015; Deegan, Rankin, and Tobin 2002), indicating a tendency to favor the stakeholder perspective. Furthermore, firms in the financial sector may have to yield from pressure to address labor practices and employee health and safety (Sweeney and Coughlan 2008), even though these matters are not materially significant in this sector. Firms' business environment (Powell and DiMaggio 2012; Oliver 1991), environmental impact (Russo-Spena, Tregua, and De Chiara 2018; Dobbs and Van Staden 2016; Gill, Dickinson, and Scharl 2008; Lattemann et al. 2009), size (Lattemann et al. 2009; Cho et al. 2015; Thorne, Mahoney, and Manetti 2014; Andrikopoulos and Krikliani 2013; Duff 2016; Li et al. 2010), and stakeholder influence (Thijssens, Bollen, and Hassink 2015; Vitolla et al. 2019) have all been shown to influence disclosure behavior. The two types of perspectives are therefore likely to have competing interests, and it is therefore not clear who the intended users of the reports are. The specific question of whether the stakeholder or shareholder perspective is dominant in CSR reporting is hence investigated in Paper I.

With recent increases in the use of price comparison websites, decisions for firms to participate in such online marketplaces with fierce competition and

low search costs for information are expected to have important implications for pricing, firm performance, consumers, incumbent firms, and government policy. However, as economic theories may dictate a certain outcome in these regards, it is equally important to test these theories with rigorous empirical studies. For instance, with regards to pricing, in his seminal paper, Stigler (1961) showed that search costs would create market power and lead to prices set above competitive levels. In addition, it has been shown numerous times that price dispersion remains in markets with low search costs (Brynjolfsson and Smith 2001; Brown and Goolsbee 2002; Baye, Morgan, and Scholten 2004b; 2004a; Baye and Morgan 2009) and that it is important being the firm having the lowest price (Baye and Morgan 2002; Dana Jr 1994; Narasimhan 1988; Salop and Stiglitz 1977; Varian 1980). Varian (1980) furthermore showed that if retailers use intertemporal price discrimination to attract both informed and uninformed consumers, then an increase in the share of informed consumers will lead to lower prices. Price dispersion will remain in equilibrium, and it will be at its largest when the two groups, informed and uninformed consumers, are of equal size (Stahl 1989). However, there are limitations to the model proposed by Varian (1980), as Rosenthal (1980) and Stahl (1989) have shown that entry by more retailers into the market will, in these models, increase rather than decrease price levels, and ultimately the market will reach the monopoly price, provided there is a large enough number of retailers. It has also been shown that competition on price comparison websites lowers prices (Brown and Goolsbee 2002; Haynes and Thompson 2008; Tang, Smith, and Montgomery 2010), which begs the question why firms choose to enter such a marketplace. Papers II – V give evidence and challenge these conventional theories in the setting of a price comparison website, during a period of rapid increase in the use of the website by both consumers and firms.

Data collection and quality

There are four datasets used in this thesis: CSR sustainability disclosures collected as part of an ongoing large-scale data collection effort at Dalarna University; *PriceSpy* price comparison website data; *PriceSpy* firm entry data collected from the Wayback Machine, using web scraping; and financial data as recorded in annual reports by Swedish firms and managed by Bisnode.

This thesis employ the Quality Assurance Framework of the European Statistical System (Eurostat 2019), which is maintained to evaluate the quality of already produced statistical outputs based on the principles no. 11–15 of the European Statistics Code of Practice (ES CoP) (Eurostat 2017), in assessing data quality. Statistics that comply with the European quality standards and serve the needs of European institutions, governments, research institutions, business concerns and the public generally, are deemed to follow the ES CoP (Eurostat 2017). The principles, e.g., principles no. 11-15 of the ES CoP, consist of: Relevance, Accuracy and Reliability, Timeliness and Punctuality, Coherence and Comparability, and Accessibility and Clarity (Eurostat 2017). As these principles vary somewhat in interpretation, depending on context, definitions provided by Bergdahl et al. (2007) have been chosen as reference.

Relevance is the degree to which statistics meet current and potential user needs. It relates to whether necessary data is produced, and how suitable this data is to answer the research question. Accuracy, in the general statistical sense, denotes the closeness of computations or estimates to the (unknown) exact or true values. Statistics are never identical with the true values because of variability and bias. Timeliness of information reflects the length of time between its availability and the event or phenomenon it describes, while punctuality refers to the time lag between the release date of data and the target date when it should have been delivered. Coherence and Comparability states that data is consistent internally, over time and comparable between regions and countries; it is possible to combine and make joint use of related data from different data sources. Finally, accessibility refers to the physical conditions under which users can obtain data, while clarity refers to whether data are accompanied with appropriate documentation and metadata.

CSR sustainability disclosures

To investigate who the intended users of CSR reports are in a data-driven manner, there was a need to collect data on such reports, and a large-scale data collection effort was undertaken. After contacting GRI, a spreadsheet referencing 47093 aggregate data points from 1998 to the beginning of 2017, e.g., firms reporting sustainability disclosures, was retrieved. With a basis in these aggregate data points, web scraping procedures were used to collect both financially integrated and non-financially integrated reports. The reports were pre-processed by being converted into text files and subjected to language classification, in which 7881 English language reports could be retained. A stepwise procedure of making subsets of the English PDF reports was then committed to. Focus was directed towards the year 2016, which is the last full year of data. Integrated reports were excluded in the analysis, and sectors and legal origins with few reports were also excluded. The text reports were then transformed into so-called document-term matrices, which implies rows being documents and columns being terms, for final preprocessing and analysis.

Relevance: Out of the aggregate data points, a majority of the CSR reports were collected. While data collection is, as of publication of this thesis, still ongoing, the sample used in Paper I is deemed sufficient for answering the research question.

Accuracy and Reliability: The data on reports are provided through: a specially designed registration system in place to submit the GRI standards compliant reports; collaboration with GRI data partners (<https://database.globalreporting.org/datapartners/>), including such companies as Ernst & Young, Deloitte and the Governance and Accountability Institute, for each respective country; and email, whereby the GRI database team will then review the application, and post the report in the database.

Timeliness and Punctuality: The GRI data partners (who provide reports to the GRI database from their countries) account to some 40% of all reports. Reports not supplied through the standardized registration system or the GRI data partners can be submitted for free by email. Timing of entry into the GRI database therefore depends on each instance of submission. For Paper I, this was not of concern, since the reports are annual, and the reporting year is explicitly stated in the actual reports.

Coherence and Comparability: CSR reports are produced globally and are highly heterogeneous by design. However the spreadsheet maintained by GRI is structured in such a way that reports can be identified, and subsets can be retrieved according to different variables. Since only the year 2016 were studied, changes in rules and regulations were not of major concern.

Accessibility and Clarity: The database is owned and operated by GRI and is completely free for the general public. GRI provides an extensive data legend as documentation, including definitions and classifications.

PriceSpy data

Four of the articles in this dissertation relate to the setting of a price comparison website, and this work was only possible due to a close collaboration with the price comparison website *PriceSpy*, Sweden. Although web scraping such a resource is certainly technically feasible, this involves a large emphasis on sampling, errors, missing values, and so forth. Furthermore, the key variable of click-through to the firms' website used in Paper II and Paper IV is only retrievable by direct contact with the price comparison website. The geographical focus was set to the Swedish *PriceSpy* website. *PriceSpy* operates in 7 different countries, and since cross-country comparisons were not within the scope of the research questions posed in this thesis, the choice was made to focus on one country. Data were finally retrieved from January 2012 until February 2017.

Relevance: While collecting data in collaboration with *PriceSpy*, multiple visits, discussions, and correspondence with the company to assess data quality and assuring an efficient data collection process were made. The product categories were chosen to have a certain set of characteristics. Firstly, the products sold should be homogeneous between firms, which is important to isolate effects when studying how use of the price comparison website affects consumer demand for retailers as well as price levels and price dispersion. Secondly, and contrary to many previous studies, more heterogeneous product category choices were made, and the analysis could thus be done for durable goods, consumer electronics, experience goods and search goods separately. Thirdly, price levels should differ between product categories to represent both cheaper and more expensive alternatives for the consumer. Besides the price of the products, other variables, such as click-through to retailer websites and consumer ratings, were also collected. This ensured a representative sample suitable for answering the research questions related to this data.

Accuracy and Reliability: The data were retrieved directly in collaboration with *PriceSpy*, and to the extent that the company had collected data, the sample retrieved was representative of the products available on the price comparison website. As *PriceSpy* themselves collect their data from retailer sites, there is a need for data quality control also after having retrieved the data. Duplicate observations were removed, and outliers were screened for, with care taken not to remove data where high prices fell in line with, or close to, the manufacturer's suggested retail price, but rather removing abnormally high prices clearly identifiable as measurement errors.

Timeliness and Punctuality: Each data entry is, in practice, collected instantaneous, either through price files given by firms or by web scraping efforts by *PriceSpy*. However, the data provided is given in daily frequency, masking which time of the day the data were collected.

Coherence and Comparability: All data retrieved from *PriceSpy* is coherent as the site structure remained the same regardless of product category. CPI adjustments were made before any analysis was committed to.

Accessibility and Clarity: Contact with *PriceSpy* enabled detailed descriptions of, and insight into, the data.

The Wayback Machine

Sometimes the data that is necessary for studying a certain phenomenon is not readily available in a structured database, as secondary data, or even by conventional data collection due to the passage of time. This was certainly a concern for Paper V, where the intent was to study how the entry of firms into a price comparison website affected firm performance. As data on when such entry took place were, in a sense, “lost in time”, there was need to circumvent this issue. As a point of departure, the option of using the entry dates of the existing *PriceSpy* data was considered. But even though the data at hand was comprehensive, it still constituted a sample of all retailers having entered the website and was hence not optimal in measuring the entry effect. There was a need for a better solution. Fortunately, using the Internet Archive and, in particular, web scraping of the Wayback Machine, it was possible to retrieve firm entry data dating back to the inception of the price comparison website in 2002, when a mere 40 firms were listed, until 2017 and forward, where over 4000 firms, and counting, were listed on the website. This was challenging though, since there were several overhauls of the website over the years, and the web scraping code had to be flexible enough to handle HTML structure over two decades. By combining “carbon dating”, e.g., estimating the creation date of a website (SalahEldeen and Nelson 2013) for the firm information pages and collecting available past firm lists on the price comparison website, it was possible to identify close to perfect entry data for the years 2002 – 2010, and 2012 – 2016, while a gap in entry data was identified between 2010 and 2012. This gap would be addressed only if it proved to be problematic when merging with the annual firm financial report data, later on. Besides the entry data, and in order to make it possible to combine entry dates with annual report data, there was also a need to collect firm organization numbers. These were also collected by web scraping the Wayback Machine for all firm information pages used in the carbon dating of firm entry.

Relevance: The entry data gave, despite the aforementioned gap, an almost perfect panel of firms entering the price comparison website. Given the merging with, and data quality assessment of, the firm annual reports, the research question of Paper V could then be studied.

Accuracy and Reliability: As *PriceSpy* also posted official firm counts on their website over the years, it was possible to assess the accuracy of data

collection directly. As reported in Paper V, the accuracy was very high, except for the gap between 2010 and 2012.

Timeliness and Punctuality: The Wayback Machine crawls webpages from publicly available websites, but coverage varies substantially between pages and over time. This was the main cause for the gap in data. For the remaining years of available legacy *PriceSpy* website content, the coverage was more than sufficient.

Coherence and Comparability: Although the structure of the *PriceSpy* website differed greatly over the years since its inception, the variable of interest, namely entry of firms, is comparable over time. That is, the variable firm entry does not depend on the website content, such as images or HTML structure, as it is only a numeric variable. To reiterate, this variable could, however, not be retrieved where the gap in data existed.

Accessibility and Clarity: The Internet Archive, and in extension the Wayback Machine, provide free access to researchers, historians, scholars, and the general public. The Internet Archive is member of such organizations such as the American Library Association and the Digital Library Federation.

Firm annual reports

The data collected by use of the Wayback Machine were combined with annual report data of Swedish limited liability firms active during the years 2005 – 2015, using firm organization numbers. Sweden obliges firms to provide annual reports to the Swedish Companies Registration Office (CRO). These annual reports are then collected by Bisnode, which is a private consultancy agency specialized in producing high-quality business, credit, and market information. The firm annual report data is then retrieved from Bisnode. The financial reports contain information such as profits, wages, costs, revenues, and number of employees. The aforementioned gap in entry data, as retrieved by the Wayback Machine, did not, when merged with the Bisnode annual report data, have a significant impact on the overall linear trend of yearly firm entry. Therefore, it was possible to study the entire cohesive time period which had overlap between the two datasets, that is 2005 – 2015.

Relevance: In Paper V, details on Swedish firm-level financial data was necessary to study the impact on productivity, profits and wages.

Accuracy and Reliability: The data in the annual reports are subject to scrutiny by the Swedish Tax Authority when levying taxes on the limited liability firms, as well as submitted by the firms for external and independent audits, ensuring accuracy and reliability.

Coherence and Comparability: The methods of collecting and compiling data for annual reports are homogeneous across all limited liability firms in Sweden, and if there are changes in rules and regulations, these are imposed simultaneously. Annual report data is comparable to the Wayback Machine

data since firm organization numbers are standardized and maintained in Sweden.

Timeliness and Punctuality: The annual reports are collected during the year following the business year that it describes. Limited companies are obligated to submit their reports at latest 7 months after the end of the financial year, ensuring punctuality.

Accessibility and Clarity: The accessibility of this data is made possible due to the legal requirements in Sweden oblige firms to provide annual reports to the Swedish CRO, for which there are clear guidelines on the official website.

Methods

Latent Dirichlet Allocation topic modelling

The analysis in Paper I serves the purpose of determining who the intended users of CSR reports are, with the two contending views being shareholders or stakeholders. Since the data were a large corpus of CSR disclosures, a need for natural language processing was identified. Natural language processing is one of the most complex problems in artificial intelligence and computer science, which gives rise to several issues one needs to reckon with when analyzing text data (Hofmann 1999; Jelodar et al. 2019). Firm CSR disclosures are a collection of words occurring in some context, while there are common words that occur in multiple contexts. In natural language processing, a topic model is a statistical model used for identifying the abstract or latent topics within a collection of documents. One method is that of the unsupervised hierarchical Bayesian machine learning model, Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003), which is, in principle, similar to that of k-means clustering, where, a priori, a number of predefined clusters must be decided upon. From the context of Paper I, the documents to be analyzed are the CSR disclosures, while the topics, again, analogous to the clusters in k-means clustering, are given a priori by the dimensions and sub-dimensions of the SASB materiality map. The task at hand is then to compare the topic distributions retrieved from the Latent Dirichlet Allocation topic modelling and compare these distributions to those suggested by SASB materiality map to decide whether firms align with the shareholder approach, or the broader stakeholder approach. Previous studies on topic modelling of CSR disclosures have incorporated inductive approaches to choosing the number of topics, such as reasoning within the research group which topics are most prevalent (Russo-Spena, Tregua, and De Chiara 2018), while the analysis in Paper I instead incorporates a strategy to test the research question that is data-driven. Increased objectivity in the analysis has its virtue since the analysis is made reproducible and can therefore be extended to past, present, and future SASB materiality maps, as well as other settings, where text data can be related to distributions derived from some standard.

Poisson pseudo-maximum likelihood regression

The purpose of Paper II is to investigate the value in terms of increased consumer demand, measured as click-through from a price comparison website to retailers' websites, of having the lowest price and highest consumer rating on the Swedish price comparison website *PriceSpy*. Click-through, our proxy for consumer demand, is a count data variable with an excess of observations that are zero, and therefore a model specification suitable for this kind of data was required. While a specific distributional assumption of the stochastic process, for example Poisson or Negative binomial, with maximum likelihood estimation of the underlying parameters could be used, these would then rely on the assumption that the true stochastic process does not differ from the one used for obtaining the maximum likelihood estimates (Cameron and Trivedi 1986; Gourieroux, Monfort, and Trognon 1984b; 1984a). Gourieroux et al. (1984b) showed, roughly, for the pseudo-maximum likelihood approach that if the mean specification is correct, any estimator for the underlying parameters obtained by maximizing the likelihood function based on the linear exponential class will be consistent for these parameters, even if the underlying distribution is misspecified. The Poisson-based pseudo-maximum likelihood estimator is therefore used in Paper II since it is of the linear exponential class, unlike, for instance, negative binomial and other common specifications for count data, while also having the benefit of behaving well when the proportion of zeros in the sample is very large (Silva and Tenreiro 2011), which is generally the case for click-through data.

Time series clustering with Dynamic Time Warping

To answer the question whether intertemporal price discrimination using mixed strategies is the cause of price dispersion in markets with low search costs in Paper III, time series clustering was used. In particular, by using time series clustering it could be determined if there are clusters of retailers with similar time series pricing patterns found, rather than mixed strategies, e.g., that retailers change price randomly over time. After controlling for time-invariant retailer heterogeneity, time series clustering over a set of 14 products was employed. As opposed to Lach (2002), who used the quartiles of the price distributions to determine whether price movements were explained by intertemporal price discrimination, it is argued that it is arbitrary to use such grouping, and it is more reasonable to use clusters of retailers determined by the data instead. While searching for an approach to answering the research question, a seemingly unrelated study of tropical cyclone trajectories in the western North Pacific (Camargo et al. 2007) sparked the idea of combining the clustering techniques used in that paper combined with Markovian transition ma-

trices, to this end. More specifically, the final analysis relied upon a combination of Dynamic Time Warping (Rath and Manmatha 2003), to align the residual time series clustered; piecewise aggregation approximation (Keogh et al. 2001), to approximate the residual time series pattern over time; the ‘elbow method’ (Thorndike 1953), to determine the number of clusters; and, finally, Markovian transition matrices, to study retailer transition probabilities.

Dynamic Time Warping (Rath and Manmatha 2003) is suitable in this instance, since if a retailer change their price slightly before a competitor this will produce a larger distance between cluster members, potentially resulting in failure to show overall trends in clusters. Piecewise aggregate approximation (Keogh et al. 2001) was used to reduce dimensionality of the residual price series to 30-day windows, which is useful to avoid over-emphasizing frequent price movements, again, leading to missing the overall trends in clusters. The ‘elbow method’, used to validate the number of clusters, was favored over other cluster validation techniques such as the silhouette index (Rousseeuw 1987) or Dunn index (Dunn 1974) due its simplicity and objectivity in combination with a simple stopping rule. Having found the clusters, Markovian transition matrices were used to study the retailer transitions between clusters. The midpoint between the centroids, e.g., the boundary between the time series of the mean of each cluster, were used to register whether a retailer transitioned across clusters. The assumption of time-homogeneity is important to scrutinize when using Markovian transition matrices, and using χ^2 tests (Anderson and Goodman 1957), it was shown that for the majority of data in Paper III the assumption of time-homogeneity held true.

Fixed effects panel data regression with marginal effects

Estimating the impact of an increase in informed consumers on prices and price dispersion, in Paper IV, requires some methodological considerations. The paper is based on a theoretical model developed by Frank and Salkever (1993), which is suitable to analyze using fixed effects panel data regression. In the final empirical specification used in Paper IV, there are several interaction terms, of which the majority involves the number of informed consumers, which is a key variable in the study. In the absence of interaction terms, the coefficients can be easily interpreted, but once interaction terms are introduced the coefficients cannot be readily interpreted. This could have dire consequences for decisionmakers, if these make uninformed choices based on wrongful interpretations of empirical model coefficients. The use of marginal effects, through the calculation of partial derivatives when assessing parameter estimates, is therefore incorporated in Paper IV to ensure meaningful interpretation of parameter estimates.

Coarsened Exact Matching and difference-in-difference model

Paper V deals with estimating the impact of firms joining the *PriceSpy* marketplace on their productivity, operating profits, and gross wages. To do this, we use a two-step procedure. First, entering firms are unlikely to be similar to a random sample of non-entering firms. To avoid bias in the second step of our estimation procedure, it was necessary to first identify a control group of firms with similar pre-entry characteristics to those entering the *PriceSpy* website. The balance between the treatment and control groups is vital to ensure quality in the analysis, and Coarsened Exact Matching (CEM) (Blackwell et al. 2009; Iacus, King, and Porro 2011; 2012) is used to preprocess the data intended to be used in the second step of the estimation procedure. As opposed to propensity score matching, CEM has the advantage of improving balance in one covariate while not giving increased imbalance in other covariates. If successful, preprocessing the data using CEM reduces model dependence in the second step of the empirical analysis, which means that empirical findings will be more robust to the choice of estimation method and model specification in the second step (Ho et al. 2007; Iacus, King, and Porro 2011).

In the second step of the empirical analysis, a generalized difference-in-difference model was used to estimate the impact of *PriceSpy* participation on productivity, operating profits, and gross wages. The identification assumption in the final model is that firm entry is uncorrelated with the error term in the regression, conditional on the inclusion of firm- and time fixed effects, and, when the estimation is done, on matched data (Arcidiacono et al. 2020). Difference-in-difference models are among the most frequently used models when evaluating effects of different treatments on relevant outcome variables (Abadie 2005).

Summary of the papers

Paper I

In recent years voluntary non-financial disclosures, such as CSR reports, have received much attention. The need of firms for reporting (Landrum and Ohsowski 2018), disclosure level (Cho et al. 2015), frequency (Sethi, Martell, and Demir 2017), and verification of CSR disclosures (Kolk 2003) has increased significantly over the decades. However, what, and for whom, firms disclose information in their CSR reports are less examined (Clarkson et al. 2008; El-Bassiouny, Darrag, and Zahran 2018; Fifka 2013; Wanderley et al. 2008), ever-changing (Russo-Spena, Tregua, and De Chiara 2018), and requires more comprehensive content analysis (Wiese et al. 2012). The intended users can broadly adhere to two main frameworks for CSR reporting: SASB and GRI, where the former adopts a perspective aimed at “the reasonable investors”, while the latter focuses on a wider spectrum of users – “the stakeholders”.

The aim of Paper I was to identify the intended users of the CSR reports in a data-driven manner, where the materiality framework of SASB is used as a benchmark. SASB identifies financially material issues, which are issues reasonably likely to impact the financial condition or operating performance of a company and therefore are most important to investors. This benchmark specifically consists of the SASB materiality map (<https://www.sasb.org/standards-overview/materiality-map/>). The paper primarily contributes to the limited literature of materiality issues in non-financial corporate disclosures (e.g., CSR disclosures). Direct evidence on how effective the voluntary reporting guidelines are (i.e., do they codify what firms disclose in their reports) is contributed to the literature, where, in this case, the SASB materiality framework was evaluated for sustainability issues.

The data used were collected as part of a research project at the Microdata Analysis Department of Dalarna University, where a sample of 671 reports as of the year 2016 from 61 countries was analyzed. The reports were subject to language recognition to be able to isolate the English language reports, and they were also preprocessed to be represented as .txt files and subsequently document-term matrices for analysis.

To be able to quantify what firms disclose in CSR reports, the Latent Dirichlet Allocation topic modelling is used, coupled with the use of a novel stepwise approach of statistically testing the fit of the topic modelling to the

SASB benchmark. A good fit to the SASB benchmark would indicate more of a “reasonable investors” perspective, as such reporting is the stated goal of the SASB. The research question was therefore addressed in a systematic way to assess the concordance of the reports to determine who the intended users are. The concordance would then be measured in how well the topics found in the CSR disclosures align with the dimensions and sub-dimensions of the SASB materiality map. Most past studies that employ text analytics within the CSR disclosure domain are based on supervised learning methods, while here an unsupervised learning method is employed – which is extensively data-driven and thereby minimizes the subjectivity in the analysis. CSR disclosures, in their present form, are extremely difficult to process relying solely on manual investigation, where a particular stakeholder may easily get overburdened due to the sheer volume of disclosure, repetition, and use of boilerplate language.

It was found that firms, in general, adopt an investor perspective when disclosing CSR information – data showed strong evidence of this when evaluated at the high-level of the five sustainability dimensions of the SASB framework, and also showed indications even when evaluated at the micro level of the 26 sub-dimensions. Moreover, there was strong bias towards a shareholder orientation in the language of the reports even when firms use the more stakeholder attuned GRI framework or operate in a more stakeholder-oriented business environment, such as the civil-law legal environment. The results of this paper have implications for preparers, users, and standards-setters in that they may evaluate CSR disclosures more objectively as well as improve the effectiveness of reporting guidelines.

Paper II

Besides what is shown in Papers III and IV with regards to homogeneous products exhibiting price dispersion despite low search costs, there also exists evidence that the number of firms advertising prices for a given product changes with high frequency (Baye, Morgan, and Scholten 2004a; Baye et al. 2007; Rosenthal 1980; Varian 1980); and firms offering the lowest price change frequently (Baye, Morgan, and Scholten 2004a). The observation that price dispersion remains, coupled with firms’ supposedly changing lowest price frequently, begs the question: if the lowest price is not the key determinant of consumer demand measured by click-through, just how important is it?

The first empirical study to investigate if there is a discontinuity in demand from holding the lowest price was carried out by Baye et al. (2009), who found that, for a dataset of 18 personal digital assistants (PDAs), sold by 19 firms over 5 months in 2003 and early 2004 on the price comparison website Kelkoo.com, a firm enjoys a 60% increase in demand when it offers the lowest

price. It is unclear, though, whether this result extends to other categories of products, and if so, to which extent, and if the results also hold over time.

This study revisits the analysis of Baye et al. (2009), and aims to verify whether the results hold, as well as to which extent, for a larger, more recent and varied dataset from another country, while also including an analysis of how consumer ratings affect demand. The *PriceSpy* dataset is used, and both search and experience goods categories are studied, whereas Baye et al. (2009) investigated a single search goods category.

The results of the study show that the discontinuity in demand of being the firm holding the lowest price is statistically significant for all categories, and a firm has at least 58%, on average 92%, and as high as 154%, increase in demand in the categories considered when offering the lowest price. It is also found that the own price elasticity of demand estimates are distorted when not accounting for the discontinuity at the lowest price. The results pertaining to consumer ratings are found to be inconclusive. Also, using a measure created by Baye et al. (2009), the share of pure price shoppers in each category is calculated. This analysis shows that the share of price shoppers is low, despite recent advancements in popularity of price comparison websites as well as technology. In conclusion, there is strong evidence that supports considering the discontinuity of the lowest price when studying consumer demand on a price comparison website, evidence which also varies depending on category or product type.

Paper III

In his seminal article, Stigler (1961) pointed out the pervasive price dispersion for homogeneous products sold in well-developed markets. The most widely used models for explaining the existence of price dispersion, even in markets with low search costs, are so-called clearinghouse models. In essence, a clearinghouse model states that retailers must simultaneously appeal to two types of consumers: shoppers, who search and use available price information to buy from the retailer offering the lowest price, and non-shoppers, who do not engage in search but learn prices over time as they visit retailers' webpages or stores. The most widely cited clearinghouse model is that of Varian (1980). For price dispersion to remain in these models, there must be some consumers who are non-shoppers, and prices must be chosen using mixed strategies, with retailers changing their prices randomly over time. Otherwise, consumers will eventually learn which retailer has the lowest price, and all consumers will either patronize the lowest price retailer, or all retailers will charge the same price.

Paper III tests the predictions of the clearinghouse models using the fact that mixed strategies, as described in these models, has some empirically testable implications. First, there can be no grouping of retailers having similar,

and thus predictable, price strategies that remain over time. Second, the position of individual retailers within a cross-sectional price distribution will change randomly over time. Therefore, there will be no distinguishable patterns in a transition matrix of prices, and the probability of remaining in the same position in the transition matrix should be low. In Paper III, these predictions from clearinghouse models are tested using a dataset of identical products sold through the *PriceSpy* price comparison website. Compared to previous studies, a larger set of products (14) is observed with a higher frequency (daily) and for a longer period (up to 42 months). An important feature of the data is that the products are identical between retailers and that there is no ambiguity in product representation on the price comparison website. However, the average coefficient of variation for the price of these products still reveals substantial price dispersion.

This observed price dispersion can, in part, be due to retailer heterogeneity, and as such time-invariant retailer heterogeneity is first controlled for in the empirical analysis, deriving price series that are purged of differences due to heterogeneity among retailers. In previous studies, such retailer heterogeneity purged price series have then been divided into quartiles before investigating the frequency and likelihood of price movements between the quartiles to determine whether these movements followed the predictions of clearinghouse models. However, the grouping of prices into quartiles is arbitrary, and both the number and size of retailer price clusters are likely to differ between products. Therefore, clustering techniques are adopted to allow the data identify the number and size of retail price clusters instead of arbitrarily dividing the data into quartiles. If there are groups of retailers gathered in well-defined price clusters, these clusters remain over long periods, and the ranking of an individual retailer in a Markovian transition matrix of these clusters is likely to stay the same. This indicates that retailers do not follow the intertemporal price discrimination strategy suggested by clearinghouse models.

The results show that while substantial price dispersion remains even after controlling for heterogeneity in retailer offerings, there are clusters of retailers that maintain persistently high, mid-range, or low prices. In addition, the transition matrices show that diagonal rather than off-diagonal elements are more likely, on average, meaning that there is a low probability of transition between clusters. As such, the hypothesis that intertemporal price discrimination using mixed strategies is the cause of the observed price dispersion on the *PriceSpy* website is rejected. One possible explanation is that the share of consumers using price comparison websites in Sweden is still low enough to make it profitable for some retailers to focus on the group of uninformed consumers, while also listing their products on the price comparison website. This could certainly be the case, since only 4 to 33 percent of consumers depending on product category were found to be pure price shoppers on *PriceSpy* in Paper II.

Paper IV

Stigler (1961) emphasized that search costs are important because they create market power and lead to prices being set above the competitive level, while Varian (1980) showed that if retailers use intertemporal price discrimination to attract both uninformed and informed consumers, then an increase in the share of informed consumers will lead to lower prices. Price dispersion will remain in equilibrium, and it will be at its largest when the two groups, informed and uninformed consumers, are of equal size (Stahl 1989). However, the theoretical models that are based on Varian (1980) have some limitations. Besides the empirical evidence presented in Paper III, casting doubt on this theoretical model, Rosenthal (1980) and Stahl (1989) also showed that entry by more retailers into the market will in these models increase rather than decrease price levels, and ultimately the market will reach the monopoly price as the number of retailers gets large.

In Paper IV, a theoretical model developed by Frank and Salkever (1993) is instead used to investigate how the increased use of a price comparison website affects prices and price dispersion. The model predicts that reduced search costs for consumers will lead to lower prices, as will entry by additional retailers. If there are two types of retailers, a market leader and followers, the model also, under certain circumstances, predicts that price dispersion increases as the number of retailers increases, while the theoretical implications of reduced search costs for consumers on price dispersion are less clear.

The dataset used to study the effects on prices and price dispersion span, from January 2012 to February 2017, is measured on daily frequency, and includes durable goods as well as consumer electronics categories. Besides the prices, data also include clicks to retailer webpages, which are useful to study increases in consumer use of the price comparison website. A main advantage of the data is that it is retrieved directly from, and in collaboration with, *PriceSpy*, Sweden, and therefore the data quality is excellent with full coverage of prices and products for the selected categories.

The increase in consumers using the website created potential savings of 290 million SEK during the last year under study, which was 2016. However, the increase in the number of retailers competing on the price comparison website had an even larger effect, and the last year increase in competition on the website created potential consumer savings of 2.9 billion SEK. Total potential consumer savings during the last year under study, due to increased use of the price comparison website by both consumers and retailers, thus amounted to almost 3.2 billion SEK. As for theoretical considerations, the result favors the Frank and Salkever (1993) model over the clearinghouse models of Varian (1980) and Stahl (1989).

Paper V

The introduction of online retailing was expected to create almost perfectly competitive markets, with no excess profits for retailers competing in those marketplaces. There is a vast literature indicating that online competition in general (Brynjolfsson and Smith 2001; Baylis and Jeffery 2002; Brown and Goolsbee 2002; Clay, Krishnan, and Wolff 2001; Baye, Morgan, and Scholten 2004b), and competition on price comparison websites, in particular, (Brynjolfsson and Smith 2001; Haynes and Thompson 2008; Tang, Smith, and Montgomery 2010; Bodur, Klein, and Arora 2015; Thompson and Haynes 2017; Jolivet and Turon 2019), increase competition and lowers prices. With evidence from the price comparison websites *PriceSpy*, it has also been shown that there are, on average, large discontinuities in demand for the lowest price in Paper II, and lower prices and large consumer savings in Paper IV.

Despite the reports of increased competition and lower prices, coupled with the evidence found on the price comparison website *PriceSpy*, firms compete on price comparison websites more frequently than ever before. The increase in the use of the price comparison website *PriceSpy*, Sweden, during the years 2013 to 2016 is remarkable. Rudholm and Lindgren (2019), report detailed statistics over this development for an example product, games for the console PlayStation 4. The data show that in 2013 there were about 20 retailers marketing some 20 games on the *PriceSpy* website, while in 2016 this had increased to almost 60 retailers marketing approximately 600 products.

So why do firms compete in a marketplace with fierce competition that reduces prices? The purpose of the last paper in the thesis is to investigate how joining the *PriceSpy* marketplace affects productivity, operating profits, and gross wages to answer the main research question: Why do firms compete on price comparison websites?

Investigating the impact of *PriceSpy* market participation on productivity, profits, and wages empirically is not an easy task, since firms that compete on the *PriceSpy* marketplace are not likely to be a representative sample of the population of firms, and we need to address this selection problem. Using a two-step procedure, Paper V addresses this problem in the following way. In a first step, we control for differences in observables between entering firms and potential control group firms with a special focus on the development of output in the pre-entry period. This procedure reduces heterogeneity in pre-entry observables between the two groups and makes the pre-entry trends in the main outcome variable, CPI-adjusted sales, parallel for entering and selected control group firms. Then, in a second step, using a within firm difference-in-difference trans-log production function estimator, it is investigated how entry into the *PriceSpy* marketplace affects CPI-adjusted sales while holding inputs constant. One advantage of using a within firm estimator is that it also controls for time-invariant heterogeneity among entering and control

group firms, leaving time-variant firm heterogeneity as the main potential source of remaining bias in the estimations.

The results show that for all firms entering the *PriceSpy* marketplace during the years 2005 until 2015, there is, on average, an increase in CPI-adjusted sales while holding inputs constant of 11.63% when entering. For retail firms the increase was 17.35%, while for wholesale firms it was 12.75%, meaning that non-retail or wholesale firms that entered the *PriceSpy* website did not gain as much as retail- and wholesale firms did. The group of other firms is, however, very heterogeneous with firms from all different types of industries, making it difficult to say precisely why this is the case. One possible explanation could be that the retail- and wholesale firms that entered had more experience in online retailing, in general, and thus had a better understanding of how to use the *PriceSpy* market to increase sales.

The results also show that gross wages increase by between 15.03% and 17.35% when entering *PriceSpy*, depending on industry, while operating profits increase by 9.42% when analyzing the full sample of firms. However, for the retail- and wholesale firms in the sample studied, any statistically significant impact of *PriceSpy* market participation on operating profits is not found.

Decision support and policy implications

The papers in this dissertation cover topics related to how voluntarily shared information in terms of CSR disclosures and increased use of price comparison websites affect firms and consumers using a microdata perspective. However, the concepts of reports, pricing and firm performance are universal, and as such the empirical results add to the existing body of literature for consumers and firms, as well as policymakers and governments, for use around the world. It is therefore meaningful to distinguish between the micro- and macroeconomic decision support and policy implications.

With regards to the sharing of CSR disclosures, the main implication is that of reporting framework considerations, which has both micro- and macroeconomic consequences since there is a variety of stakeholders throughout society. While the result of Paper I strengthens the argument that a stakeholder focus is subordinate to a shareholder focus, which is also supported by previous literature (Harrison and van der Laan Smith 2015; Mitchell et al. 2015; Zeff 1978), this is partly expected since the SASB materiality map, by design, is based on what firms find useful to report to shareholders. However, since a large selection of the reports that are labelled as following the GRI guidelines directed at stakeholders, and for which there is evidence of language alignment with the SASB materiality map, standard setters need to take this into account when designing future frameworks or guidelines.

When considering the microeconomic perspective, in particular, Paper II, which provides evidence on the value of having the lowest price or highest ratings on a price comparison website, shows that not accounting for the discontinuity in demand will have implications for measuring price elasticities, which in turn affects firms' optimal pricing decisions. Although no conclusive evidence of such a discontinuity was found for consumer ratings, there was conclusive evidence of misleading estimates of price elasticities due to the discontinuity found when considering the lowest price, in all cases leading to higher elasticity estimates when excluding the discontinuity. The adjusted price elasticity estimates would then generally be useful for firms when making pricing decisions. While price ranking should be considered important and generally the lowest price is favored, the margins with which this price is held could be kept to a minimum, due to lower elasticities. For retailers competing on a platform, such as *PriceSpy* Sweden, the results from the clustering in Paper III have some key takeaways. For instance, since well-defined clusters are found, it is important for the individual retailer to find its "identity" within

these clusters. Is the retailer a low-, medium- or high-priced retailer? It could be that the number of shoppers in the marketplace is small enough to commit to such strategies as maintaining a higher price and capturing uninformed consumers, in which it makes little sense for entering retailers to commit to a randomization of prices, or voluntarily joining a low-price cluster of retailers. One of our main results from Paper IV is that the increased use of price comparison websites lowers prices, implying that these websites increase economic efficiency as retailers' prices are closer to marginal costs. The results of Papers III and IV are good for consumers since they may enjoy lower prices and also reduce their search costs as retailers tend to price in predictable clusters. However, the question is then raised why retailers participate in such markets, a question answered in Paper V, where it is found that *PriceSpy* participation clearly increases productivity and profits, the latter being divided among labor and capital owners, with labor having a somewhat larger share of the increase in profits. It is therefore clear that entry to *PriceSpy* Sweden has had overall positive effects for both consumers and firms.

When instead considering the macroeconomic perspective, primarily, Paper II provides excess burden of taxation estimates for policymakers considered when contemplating taxing online transactions. Many studies (A. Goolsbee 2000; Ballard and Lee 2007; Ellison and Ellison 2009; Alm, Melnik, and others 2010; A. Goolsbee, Lovenheim, and Slemrod 2010; Alm and Melnik 2012) analyze the sensitivity of e-commerce to differential effective sales tax rates. Although the estimated elasticities vary across the papers, the majority of this literature finds large responses of online transactions to sales taxation, suggesting that sales taxes and e-commerce have important economic interactions. Taxing such marketplaces is not a far-fetched idea, and a debate has been going on in a US setting (Einav et al. 2014). On June 21, 2018, the United States Supreme Court altered the rules for the collection of sales tax revenues from internet-based retailers. In its decision on *South Dakota v. Wayfair Inc.*, the Court concluded that individual states could require online sellers to collect sales taxes on their sales. This decision was a reversal of that made in its 1992 ruling on *Quill Corporation v. North Dakota*, which prohibited states from requiring businesses to collect sales taxes, unless those businesses had a physical presence in the state. As of 2020, over forty states have enacted new tax laws related to taxing internet sales, the presumption being that this will provide a boon to brick-and-mortar stores' sales and state tax revenues, but there is not universal agreement on this point (Jens, Patin, and Turpin 2020). From Paper II, one can deduce that making tax decisions based on price elasticity estimates that are larger when not accounting for the discontinuity in demand may lead to the introduction of taxes being over-rejected. However, Paper II furthermore shows that policymakers also must account for widely different outcomes in terms of welfare loss due to taxation, depending on product categories, suggesting that a suitable overall tax level may be difficult to set.

Also, in Paper IV substantial yearly price reductions due to increased use of price comparison websites were found, which implies an underlying downward pressure on inflation. In Sweden, a 2% inflation goal is stipulated by the Swedish Central Bank, and price comparison websites may therefore make it difficult to attain this goal. In fact, according to Statistics Sweden, inflation was below the price stability target of the Swedish Central Bank for every month of our study period. Even though increased use of price comparison websites is not the sole reason for not achieving the price stability goal, the empirical evidence in Paper IV supports the claim that increased use of such platforms might be a contributing factor.

In addition to creating problems reaching the set inflation target, increased e-commerce and use of price comparison websites could also create difficulties in the measurement of inflation. Goolsbee and Klenow (2018) report that inflation rates calculated, based on online transactions, are 1.3 percentage points lower than those calculated using traditional data collection methods. This is an indication that if the measurement methods used by central banks do not take increased use of online retailing and price comparison websites into account, inflation measures are likely to be upward biased. This is also recognized by Statistics Sweden (2020), which found that changes in consumer behavior due to ongoing digitalization that are not accounted for by traditional measurement methods might bias the measurement of the consumer price index, the most widely used measure of inflation in Sweden.

Discussion

This dissertation has studied how information sharing has had an effect in society on a variety of interest groups, such as consumers, firms, policymakers, and investors. In the context of CSR disclosures, as well as the price comparison websites marketplaces, there has been a massive growth in both firms disclosing CSR related topics and firms entering the price comparison website marketplace. Answers to a variety of research questions related to information sharing, pricing and firm performance have been examined and presented in the dissertation.

For reporting of CSR disclosures, the growing literature in the field has shown that the need of firms for reporting, disclosure level, frequency and verification of disclosures has increased significantly over the decades. Gaps have been filled when it comes to how disclosures are influenced by societal norms, and used for legitimizing efforts, for signaling asymmetric information and so forth. In Paper I, a data-driven approach to determine the intended users was developed and used for this specific purpose. The shareholder perspective was found to be the dominant perspective, and the SASB framework is based on historical value relevance, i.e., what firms themselves found useful to disclose to shareholders. One may then ask, since past frameworks (that is the historical materiality maps in themselves) carry weight when it comes to firms deciding on reporting practices, what are the long-term implications and trends for CSR disclosure reporting?. Paper I provided a snapshot of the year 2016, but this question can ideally be answered using the data-driven approach as a method of studying the longitudinal implications of competing frameworks, as more and more materiality maps are reported on a yearly basis. If such proposed studies would find that CSR disclosures conform to a single framework, the consequences for firms, stakeholders and shareholders need to be studied in-depth. For instance, stakeholders would have to be careful in how they interpret these disclosures when not being the intended users, while shareholders could then be more incentivized to use the reports for their investing practices.

With regards to solving the issue of providing a meaningful reporting framework, seemingly simple answers exist, such as issuing multiple reports classified as either generic or specialized (Azzone et al. 1997), or purposely reporting to stakeholders based on core business (Hörisch, Schaltegger, and Freeman 2020; Harrison and van der Laan Smith 2015; Mitchell et al. 2015).

Such advice, however, must be given with care since it may not be economically sound due to the associated costs. If society and policymakers deem it necessary to promote the comparably weaker stakeholder perspective, the research presented here shows that measures need to be taken to be able to move in such a direction, such as issuance of multiple reports, e.g., generic vs. specialized reports. This, however, would then, most likely, necessitate regulatory intervention, whereby Paper I can give guidance through empirical evidence retrieved by a data-driven method. The novel approach of Paper I provides the means of dealing with complexity in reporting volume, repetition, and use of boilerplate language, so that decision-makers can evaluate CSR disclosures as well as the effectiveness of reporting guidelines in a systematic fashion.

Online retailing, in general, has also grown considerably during the study period. During the period under examination in our paper, Swedish online retailing has grown from 31.6 billion SEK in 2012, to 67 billion SEK in 2017 (from 2.8, to 6.0 billion EUR) according to (PostNord 2019), with yearly growth rates in the interval of 14 to 17%. However, after 2017, the average yearly growth rates slowed somewhat and ranged between 12 and 14% (PostNord 2019), and online retailing was expected to grow from 87 billion SEK (7.8 billion EUR) in 2019, to a predicted 96.6 billion SEK (8.7 billion EUR) in 2020, equaling a yearly growth rate of 11%. Then COVID-19 struck. During the second quarter of 2020, online retailing was 49% higher than during the same quarter in 2019 (PostNord 2020a), and in the third quarter of 2020, the increase was 39%, compared to the same quarter in 2019 (PostNord 2020b). Online retailing in Sweden is now expected to increase from 87 billion SEK in 2019 to 115.7 billion SEK (10.4 billion EUR) in 2020, equaling a yearly growth rate of 33%.

With this in mind, the results in this dissertation demonstrate that increased entry of market participants, consumers as well as firms, on a price comparison website marketplace, led to lower prices, and, by extension, to savings for consumers. On the other hand, retailers, who entered the market and may have had initial fears of negative impacts from fierce competition, instead enjoy increases in productivity, operating profits, and gross wages. In Paper III it was also shown, besides the contribution of testing economics theory, that retailers consistently participate in this marketplace maintaining medium to high prices. Retailers could perhaps even optimize their pricing decisions in such an environment, given the leeway in pricing decisions that exists, using, for instance, the demand elasticities modelled in Paper II as a point of departure. Since retailers and consumers already enjoy these benefits, there could possibly even be room for some taxation by the government, or at least support for an argument in its favor.

This paints a somewhat rosy picture of the current circumstances, but will this last forever? As has been shown in Paper IV, and elsewhere, price dispersion is a persistent phenomenon, and Paper III empirically cast doubt on the

predictions of Varian (1980), who claim that intertemporal price discrimination is the cause of price dispersion. With the model suggested by Varian (1980), consumers would have to use the price comparison website frequently, while the empirical evidence of Paper III suggests that consumers could reduce their search costs by learning over time that retailers adhere to certain clusters of low, medium, and high prices. Lower prices were also shown to be the current trend in Paper IV, and as more retailers and consumers enter the price comparison website, this trend is reasonably expected to continue. However, at the time of publishing this dissertation, there has been a recent entry of the Amazon marketplace in Sweden in the fall of 2020. Not all products from this website are listed on *PriceSpy*, Amazon is not a profiled retailer on *PriceSpy* and does not have a link to their website, and what constitutes dual marketplaces seems to have emerged in this context. This calls into question how one would define a truly informed consumer, as albeit the consumer having a list of retailers to browse from, the lowest price is no longer certain to be uniquely lowest. Informed consumers would then have to browse both these marketplaces simultaneously, which would then paradoxically lead to higher search costs. As a platform, Amazon also gives smaller retailers the opportunity to compete as they gain relief from logistics and infrastructure they would otherwise have to bear (Handelsrådet 2018). These are trends to follow and be researched, but as it stands, this would most likely be a gradual transition over the years, rather than an instantaneous change.

Limitations of the work presented here mainly adhere to the different contexts and data availability from which the papers are written. The CSR disclosure analysis, although ambitious in its reach, only accounts for the year 2016 and using English language reports. Although out of scope for Paper I, multi-year analysis and accounting for cultural norms by incorporating reports in different languages are natural steps for future research. Furthermore, integrated reports were not included, and developing a means to isolate the CSR portion of these reports would greatly improve the data available for analysis.

With regards to the price comparison website *PriceSpy*, a global perspective would be interesting to study as opposed to the focus given to the Swedish context in this dissertation. *PriceSpy* operates in 7 different countries, and as the websites are largely homogeneous in structure, future studies could shed light on whether the questions asked in this dissertation could be shown to hold in these contexts as well. However, for such direct comparisons to take place, it would first be informative to study whether there is a high degree of correlation between the prices of products sold in Sweden and the other countries, and how have these correlations varied over time. Having knowledge of how correlated prices are across countries and the magnitude of price differences between countries has implications for assessing the impact of shocks to the economy, such as the, as of writing, ongoing COVID-19 pandemic. In the dissertation only prices on daily frequency were accounted for, but retailers are free to set prices during any time of the day, and *PriceSpy* also update

their prices at least 5 times a day. There is also the possibility that retailers use algorithmic pricing, trying to undercut other retailers with higher frequency in pricing. More alarmingly, if algorithmic pricing were to be employed in efforts to maintain cartel pricing by retailers, it would certainly be in the interest of competition authorities to have access to studies of such behavior. These are but some of the possible avenues for future research.

References

- Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies* 72 (1): 1–19.
- Alm, James, and Mikhail I Melnik. 2012. "Cross-Border Shopping and State Use Tax Liabilities: Evidence from Ebay Transactions." *Public Budgeting & Finance* 32 (1): 5–35.
- Alm, James, Mikhail I Melnik, and others. 2010. "Do EBAY Sellers Comply with State Sales Taxes?" *National Tax Journal* 63 (2): 215–36.
- Anderson, Theodore W, and Leo A Goodman. 1957. "Statistical Inference about Markov Chains." *The Annals of Mathematical Statistics*, 89–110.
- Andrikopoulos, Andreas, and Nikoleta Kriklani. 2013. "Environmental Disclosure and Financial Characteristics of the Firm: The Case of Denmark." *Corporate Social Responsibility and Environmental Management* 20 (1): 55–64.
- Arcidiacono, Peter, Paul B Ellickson, Carl F Mela, and John D Singleton. 2020. "The Competitive Effects of Entry: Evidence from Supercenter Expansion." *American Economic Journal: Applied Economics* 12 (3): 175–206.
- Azzone, Giovanni, Michael Brophy, Giuliano Noci, Richard Welford, and William Young. 1997. "A Stakeholders' View of Environmental Reporting." *Long Range Planning* 30 (5): 699–709.
- Ballard, Charles L, and Jaimin Lee. 2007. "Internet Purchases, Cross-Border Shopping, and Sales Taxes." *National Tax Journal*, 711–25.
- Baye, Michael R., John Morgan, and Patrick Scholten. 2004a. "Temporal Price Dispersion: Evidence from an Online Consumer Electronics Market." *Journal of Interactive Marketing* 18 (4): 101–15. <https://doi.org/10.1002/dir.20016>.
- Baye, Michael R, J Rupert J Gatti, Paul Kattuman, and John Morgan. 2007. "A Dashboard for Online Pricing." *California Management Review* 50 (1): 202–16.
- . 2009. "Clicks, Discontinuities, and Firm Demand Online." *Journal of Economics & Management Strategy* 18 (4): 935–75.
- Baye, Michael R, and John Morgan. 2002. "Information Gatekeepers and Price Discrimination on the Internet." *Economics Letters* 76 (1): 47–51.
- . 2009. "Brand and Price Advertising in Online Markets." *Management Science* 55 (7): 1139–51.
- Baye, Michael R, John Morgan, and Patrick Scholten. 2004b. "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site." *The Journal of Industrial Economics* 52 (4): 463–96.
- Baylis, Kathy, and Perloff M. Jeffery. 2002. "Price Dispersion on the Internet: Good Firms and Bad Firms." *Review of Industrial Organization* 21 (3): 305–24.
- Bergdahl, Mats, Manfred Ehling, Eva Elvers, Erika Földesi, Thomas Körner, Andrea Kron, Peter Lohauß, et al. 2007. "Handbook on Data Quality Assessment Methods and Tools." *Ehling, Manfred Körner, Thomas*.
- Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. "CEM: Coarsened Exact Matching in Stata." *The Stata Journal* 9 (4): 524–46.

- Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *The Journal of Machine Learning Research* 3: 993–1022.
- Bodur, H Onur, Noreen M Klein, and Neeraj Arora. 2015. "Online Price Search: Impact of Price Comparison Sites on Offline Price Evaluations." *Journal of Retailing* 91 (1): 125–39.
- Bondy, Krista, Jeremy Moon, and Dirk Matten. 2012. "An Institution of Corporate Social Responsibility (CSR) in Multi-National Corporations (MNCs): Form and Implications." *Journal of Business Ethics* 111 (2): 281–99.
- Brammer, Stephen, and Stephen Pavelin. 2008. "Factors Influencing the Quality of Corporate Environmental Disclosure." *Business Strategy and the Environment* 17 (2): 120–36.
- Brown, Jeffrey R, and Austan Goolsbee. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110 (3): 481–507.
- Brynjolfsson, Erik, and Michael D Smith. 2001. "The Great Equalizer? Consumer Choice Behavior at Internet Shopbots."
- Camargo, Suzana J, Andrew W Robertson, Scott J Gaffney, Padhraic Smyth, and Michael Ghil. 2007. "Cluster Analysis of Typhoon Tracks. Part I: General Properties." *Journal of Climate* 20 (14): 3635–53.
- Cameron, A Colin, and Pravin K Trivedi. 1986. "Econometric Models Based on Count Data. Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics* 1 (1): 29–53.
- Cho, Charles H, Giovanna Michelon, Dennis M Patten, and Robin W Roberts. 2015. "CSR Disclosure: The More Things Change...?" *Accounting, Auditing & Accountability Journal*.
- Clarkson, Peter M, Yue Li, Gordon D Richardson, and Florin P Vasvari. 2008. "Revisiting the Relation between Environmental Performance and Environmental Disclosure: An Empirical Analysis." *Accounting, Organizations and Society* 33 (4–5): 303–27.
- Clay, Karen, Ramayya Krishnan, and Eric Wolff. 2001. "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry." *The Journal of Industrial Economics* 49 (4): 521–39.
- Dana Jr, James D. 1994. "Learning in an Equilibrium Search Model." *International Economic Review*, 745–71.
- Deegan, Craig, Michaela Rankin, and John Tobin. 2002. "An Examination of the Corporate Social and Environmental Disclosures of BHP from 1983-1997: A Test of Legitimacy Theory." *Accounting, Auditing & Accountability Journal*.
- Dobbs, Stevie, and Chris Van Staden. 2016. "Motivations for Corporate Social and Environmental Reporting: New Zealand Evidence." *Sustainability Accounting, Management and Policy Journal*.
- Duff, Angus. 2016. "Corporate Social Responsibility Reporting in Professional Accounting Firms." *The British Accounting Review* 48 (1): 74–86.
- Dunn, Joseph C. 1974. "Well-Separated Clusters and Optimal Fuzzy Partitions." *Journal of Cybernetics* 4 (1): 95–104.
- Einav, Liran, Dan Knoepfle, Jonathan Levin, and Neel Sundaresan. 2014. "Sales Taxes and Internet Commerce." *American Economic Review* 104 (1): 1–26. <https://doi.org/10.1257/aer.104.1.1>.
- El-Bassiouny, Noha, Menatallah Darrag, and Nada Zahran. 2018. "Corporate Social Responsibility (CSR) Communication Patterns in an Emerging Market." *Journal of Organizational Change Management*.

- Ellison, Glenn, and Sara Fisher Ellison. 2009. "Tax Sensitivity and Home State Preferences in Internet Purchasing." *American Economic Journal: Economic Policy* 1 (2): 53–71.
- Eurostat. 2017. "European Statistics Code of Practice." 2017. <https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142>.
- . 2019. "Quality Assurance Framework of the European Statistical System. Version 2.0." 2019. <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V2.0-final.pdf>.
- Fifka, Matthias S. 2013. "Corporate Responsibility Reporting and Its Determinants in Comparative Perspective—a Review of the Empirical Literature and a Meta-Analysis." *Business Strategy and the Environment* 22 (1): 1–35.
- Frank, Richard G, and David S Salkever. 1993. "Pricing, Patent Loss and the Market for Pharmaceuticals." *Southern Economic Journal* 59 (2): 165–79.
- Gill, Donna Louise, Sonia Jane Dickinson, and Arno Scharl. 2008. "Communicating Sustainability." *Journal of Communication Management*.
- Goolsbee, Austan. 2000. "In a World without Borders: The Impact of Taxes on Internet Commerce." *The Quarterly Journal of Economics* 115 (2): 561–76.
- Goolsbee, Austan D, and Peter J Klenow. 2018. "Internet Rising, Prices Falling: Measuring Inflation in a World of e-Commerce." In *Aea Papers and Proceedings*, 108:488–92.
- Goolsbee, Austan, Michael F Lovenheim, and Joel Slemrod. 2010. "Playing with Fire: Cigarettes, Taxes, and Competition from the Internet." *American Economic Journal: Economic Policy* 2 (1): 131–54.
- Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984a. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica: Journal of the Econometric Society*, 701–20.
- . 1984b. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica: Journal of the Econometric Society*, 681–700.
- Hahn, Rüdiger, and Michael Kühnen. 2013. "Determinants of Sustainability Reporting: A Review of Results, Trends, Theory, and Opportunities in an Expanding Field of Research." *Journal of Cleaner Production* 59: 5–21.
- Handelsrådet. 2018. "Gränslös Handel: Sju Framgångsfaktorer För Global e-Handel." *Handelsrådet, Stockholm*.
- Harrison, Jeffrey S, and Joyce van der Laan Smith. 2015. "Responsible Accounting for Stakeholders." *Journal of Management Studies* 52 (7): 935–60.
- Haynes, Michelle, and Steve Thompson. 2008. "Price, Price Dispersion and Number of Sellers at a Low Entry Cost Shopbot." *International Journal of Industrial Organization* 26 (2): 459–72.
- Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.
- Hofmann, Thomas. 1999. "Probabilistic Latent Semantic Indexing." In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50–57.
- Hörisch, Jacob, Stefan Schaltegger, and R Edward Freeman. 2020. "Integrating Stakeholder Theory and Sustainability Accounting: A Conceptual Synthesis." *Journal of Cleaner Production* 275: 124097.
- "How Much Data Is Generated Each Day? | World Economic Forum." 2019. 2019. <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/>.

- Iacus, Stefano M, Gary King, and Giuseppe Porro. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association* 106 (493): 345–61.
- . 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Political Analysis*, 1–24.
- Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey." *Multimedia Tools and Applications* 78 (11): 15169–211.
- Jens, William, Jeanne-Claire Patin, and Lonnie Turpin. 2020. "Can Charging Sales Taxes on Internet Sales Save Brick-and-Mortar Stores?" *Journal of Accounting, Ethics & Public Policy* 21 (2): 263–74.
- Jolivet, Grégory, and H el ene Turon. 2019. "Consumer Search Costs and Preferences on the Internet." *The Review of Economic Studies* 86 (3): 1258–1300.
- Keogh, Eamonn, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. 2001. "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases." *Knowledge and Information Systems* 3 (3): 263–86.
- Kolk, Ans. 2003. "Trends in Sustainability Reporting by the Fortune Global 250." *Business Strategy and the Environment* 12 (5): 279–91.
- Lach, Saul. 2002. "Existence and Persistence of Price Dispersion : An Empirical Analysis Author (s): Saul Lach Source : The Review of Economics and Statistics , Vol . 84 , No . 3 (Aug . , 2002) , Pp . 433-444 Published by : The MIT Press Stable URL : <https://www.jstor.org>" 84 (3): 433–44.
- Landrum, Nancy E, and Brian Ohsowski. 2018. "Identifying Worldviews on Corporate Sustainability: A Content Analysis of Corporate Sustainability Reports." *Business Strategy and the Environment* 27 (1): 128–51.
- Lattemann, Christoph, Marc Fetscherin, Ilan Alon, Shaomin Li, and Anna-Maria Schneider. 2009. "CSR Communication Intensity in Chinese and Indian Multinational Companies." *Corporate Governance: An International Review* 17 (4): 426–42.
- Li, Shaomin, Marc Fetscherin, Ilan Alon, Christoph Lattemann, and Kuang Yeh. 2010. "Corporate Social Responsibility in Emerging Markets." *Management International Review* 50 (5): 635–54.
- Mitchell, Ronald K, Harry J Van Buren III, Michelle Greenwood, and R Edward Freeman. 2015. "Stakeholder Inclusion and Accounting for Stakeholders." Wiley Online Library.
- Narasimhan, Chakravarthi. 1988. "Competitive Promotional Strategies." *The Journal of Business* 61 (4): 427. <https://doi.org/10.1086/296442>.
- Oliver, Christine. 1991. "Strategic Responses to Institutional Processes." *Academy of Management Review* 16 (1): 145–79.
- PostNord. 2019. "E-Barometern  arsrapport 2019." Stockholm (in Swedish): PostNord.
- . 2020a. "E-Barometern Q2 2020." Stockholm (in Swedish): PostNord.
- . 2020b. "E-Barometern Q3 2020." Stockholm (in Swedish): PostNord.
- Powell, Walter W, and Paul J DiMaggio. 2012. *The New Institutionalism in Organizational Analysis*. University of Chicago press.
- Rath, Toni M., and R. Manmatha. 2003. "Word Image Matching Using Dynamic Time Warping." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2. <https://doi.org/10.1109/cvpr.2003.1211511>.

- Rosenthal, Robert W. 1980. "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price." *Econometrica* 48 (6): 1575. <https://doi.org/10.2307/1912828>.
- Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20: 53–65.
- Rudholm, Niklas, and Charlie Lindgren. 2019. "Prisspridning På E-Handelsmarknader Med Låga Sökkostnader/Price Dispersion in e-Tailing Markets with Low Search Costs (in Swedish)." *The Swedish Competition Agency*, no. 1.
- Russo-Spena, Tiziana, Marco Tregua, and Alessandra De Chiara. 2018. "Trends and Drivers in CSR Disclosure: A Focus on Reporting Practices in the Automotive Industry." *Journal of Business Ethics* 151 (2): 563–78.
- SalahEldeen, Hany M, and Michael L Nelson. 2013. "Carbon Dating the Web: Estimating the Age of Web Resources." In *Proceedings of the 22nd International Conference on World Wide Web*, 1075–82. <https://arxiv.org/pdf/1304.5213.pdf>.
- Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311 (5762): 854–56. <https://doi.org/10.1126/science.1121066>.
- Salop, Steven, and Joseph Stiglitz. 1977. "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion." *The Review of Economic Studies* 44 (3): 493–510.
- Sethi, S Prakash, Terrence F Martell, and Mert Demir. 2017. "Enhancing the Role and Effectiveness of Corporate Social Responsibility (CSR) Reports: The Missing Element of Content Verification and Integrity Assurance." *Journal of Business Ethics* 144 (1): 59–82.
- Silva, J M C Santos, and Silvana Tenreiro. 2011. "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator." *Economics Letters* 112 (2): 220–22.
- Stahl, Dale O. 1989. "Oligopolistic Pricing with Sequential Consumer Search." *The American Economic Review*, 700–712.
- Statistics Sweden. 2020. "Konsumentbeteende i Det Digitaliserade Samhället Och Risker För Bias i KPI." PM till nämnden för KPI, September 2020, Örebro, Sverige (in Swedish).
- Stigler, George J. 1961. "The Economics of Information." *The Journal of Political Economy* 69 (3): 213–25. [https://doi.org/10.1016/0377-2217\(87\)90117-2](https://doi.org/10.1016/0377-2217(87)90117-2).
- Sweeney, Lorraine, and Joseph Coughlan. 2008. "Do Different Industries Report Corporate Social Responsibility Differently? An Investigation through the Lens of Stakeholder Theory." *Journal of Marketing Communications* 14 (2): 113–24.
- Tang, Zhulei, Michael D Smith, and Alan Montgomery. 2010. "The Impact of Shopbot Use on Prices and Price Dispersion: Evidence from Online Book Retailing." *International Journal of Industrial Organization* 28 (6): 579–90.
- "The Forum for Sustainable and Responsible Investment." 2020. 2020. https://www.ussif.org/blog_home.asp?display=155.
- "The Value of Data | World Economic Forum." 2017. 2017. <https://www.weforum.org/agenda/2017/09/the-value-of-data/>.
- Thijssens, Thomas, Laury Bollen, and Harold Hassink. 2015. "Secondary Stakeholder Influence on CSR Disclosure: An Application of Stakeholder Salience Theory." *Journal of Business Ethics* 132 (4): 873–91.
- Thompson, Steve, and Michelle Haynes. 2017. "The Value of Online Seller Reputation: Evidence from a Price Comparison Site." *Managerial and Decision Economics* 38 (3): 302–13.

- Thorndike, Robert L. 1953. "Who Belongs in the Family." In *Psychometrika*.
- Thorne, Linda, Lois S Mahoney, and Giacomo Manetti. 2014. "Motivations for Issuing Standalone CSR Reports: A Survey of Canadian Firms." *Accounting, Auditing & Accountability Journal*.
- Varian, Hal R. 1980. "A Model of Sales." *The American Economic Review* 70 (4): 651–59.
- Vitolla, Filippo, Nicola Raimo, Michele Rubino, and Antonello Garzoni. 2019. "How Pressure from Stakeholders Affects Integrated Reporting Quality." *Corporate Social Responsibility and Environmental Management* 26 (6): 1591–1606.
- Wanderley, Lilian Soares Outtes, Rafael Lucian, Francisca Farache, and José Milton de Sousa Filho. 2008. "CSR Information Disclosure on the Web: A Context-Based Approach Analysing the Influence of Country of Origin and Industry Sector." *Journal of Business Ethics* 82 (2): 369–78.
- Wiese, Anne, Julian Kellner, Britta Lietke, Waldemar Toporowski, and Stephan Zielke. 2012. "Sustainability in Retailing--a Summative Content Analysis." *International Journal of Retail & Distribution Management*.
- Zeff, Stephen A. 1978. "The Rise of 'Economic Consequences.'"

