WEB SCRAPING FOR FOOD PRICE RESEARCH

Judith Hillen

judith.hillen@agroscope.admin.ch

Agroscope, Tänikon 1, CH-8356 Ettenhausen

2018

Vortrag anlässlich der 58. Jahrestagung der GEWISOLA
(Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.)
„Visionen für eine Agrar- und Ernährungspolitik nach 2020"
Kiel, 12. bis 14. September 2018

WEB SCRAPING FOR FOOD PRICE RESEARCH

Abstract

Web scraping is a method for extracting large amounts of data from online sources. In food price analysis, however, this data collection technique has not yet received a lot of attention. We discuss how this method can be used in food price research and identify areas of application. We find that web scraping is a promising method to collect customized, high-frequency data in real time, overcoming several limitations of currently used food price data. While today's applications mostly focus on (online) consumer prices, the scope of applications broadens as more and more price data are published online. To better deal with the technical and legal challenges of web scraping and to exploit its scalability, joint data collection projects in the field of agricultural and food economics should be considered.

Keywords

Food prices, data collection, digitalization, big data, e-commerce.

1 Introduction

Web scraping is a relatively new method for collecting online data. The term describes the automated process of accessing websites and downloading specific information, such as prices, from each (KIENLE et al., 2004). Allowing the creation of large, customized data sets at low costs, web scraping is already applied for scientific and commercial purposes in many areas, such as marketing, industrial organizations, or inflation measurement (for an overview, see CAVALLO and RIGOBON, 2016; EDELMAN, 2012).

In food price research, however, this data collection technique has not yet received a lot of attention. In agricultural economics and food system analysis, we mostly rely on more traditional data sources, such as official price indices or scanner data for consumer prices. Yet, there are some issues associated with these data sources.

Official prices and price indices for products, segments such as food, or even the whole economy, are mostly published on a monthly or quarterly basis, with some publication delay. The public provision by official agencies and the availability of long time series are attractive for research purposes. Yet, one must rely on correct data collection, weighting, and aggregation by official sources. Since normally no access to the raw data is given, it is not possible to detect errors or even manipulations (CAVALLO, 2013).

Scanner data obtained at the point-of-sale at retailers are available at a higher frequency (generally weekly) and provide more details at the product level. A main advantage is that they include transaction data, i.e., the quantities purchased of a good at a given price (CAMPBELL and EDEN, 2014; COTTERILL, 1994; SILVER and HERAVI, 2001). However, these data need to be purchased from market research institutes such as *Nielsen N.V.* and can be very costly, especially if longer time series or multiple retailers and locations are required.

As an increasing number of prices is published online and as online grocery retail is slowly gaining market shares in many parts of the world (NIELSEN N.V., 2015; RIGBY, 2018), web scraping may be a promising alternative to get data for food price research.

In the following, we will not give detailed instructions on how to build a web scraper, and we will omit technical details and coding issues. Rather, the aim is to discuss the method's potential for agricultural and food economics research. Chapter Two gives an overview on what exactly web scraping is and how it works, and weights the pros and cons regarding food price analysis. Chapter Three reviews existing applications and considers further applications for studying online and offline food prices. The article finishes with an outlook and suggestions on how this new data collection method could best be used in the discipline of agricultural and food economics.

2 A Treatise on Web Scraping

2.1 Definition

Throughout this article, we use the term *web scraping*. However, there are several related terms and concepts, which are not always distinctively defined (for definitions, see, e.g., KIENLE et al., 2004; MASSIMINO, 2016; NAKASH et al., 2015).

As a minimal definition, web scraping (or *screen scraping, information scraping*) describes the automated process of accessing web documents and downloading specific, pre-defined information, such as prices, from each, to then transform and save them into a structured format.

Web crawling on the other hand means accessing web content and indexing it via *hyperlinks*; i.e., only the URLs, but no specific information is extracted. Instead, the full content is made available through the hyperlink, but is generally not archived. Search engines, including Google, crawl the web, analyze the online content, and compile all the links they find to match the search request. Crawlers (or *spiders*) are also used for price comparison tools. *Shopbots* are programs that crawl websites to obtain price information from different sellers in order to find the lowest price (HEMENWAY and CALISHAIN, 2004).

For a conceptual distinction between scraping and crawling, see Table 1. The remainder of this article deals only with web scraping, because we are interested in collecting food price data for research from public websites.

Table 1: Distinction between Web Scraping and Web Crawling

	Web Scraping	Web Crawling
Process	Automatically requesting web documents and collecting information from it	Repetitively finding and fetching hyperlinks starting from a list of starting URLs
Target information	Pre-defined data on specific websites	URLs to access all kind of information, depending on search request
Output	Downloaded data in structured format	Indexed hyperlinks, stored in database
Use	Data collection (e.g., price series)	Ad-hoc requests (e.g., search engines, price comparison tools)

Source: own representation

2.2 Technical Procedure

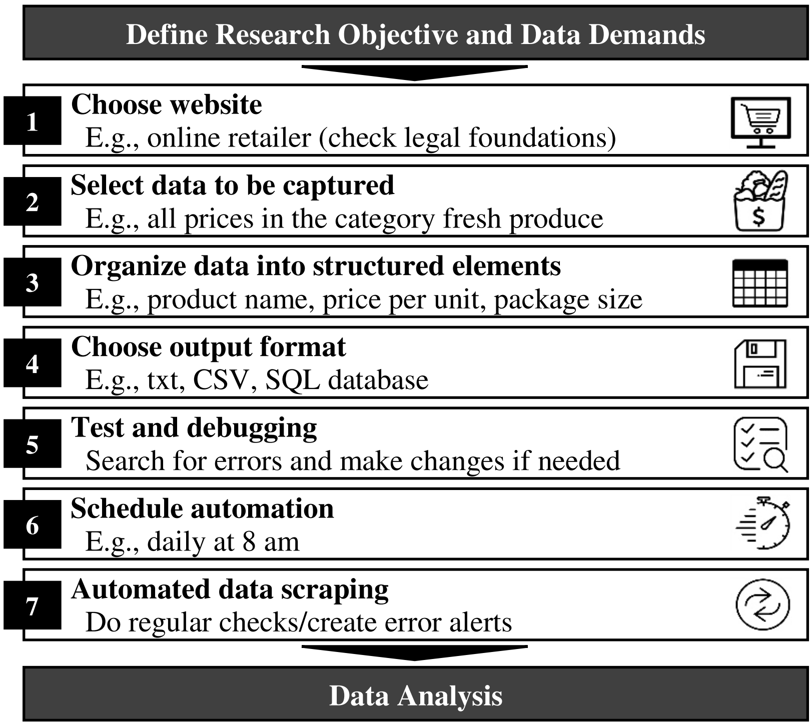
There are several ways to build a web scraper, and probably no one-size-fits-all approach exists. While this article does not aim to give detailed instructions on how to code a web scraper, we will briefly give an intuitive description of what a web scraper does technically and how this can be implemented for creating food price data sets.

What to do

Generally speaking, one needs to write a script that accesses the websites hosting the data, finds the relevant, previously defined elements, to then download and store them in structured data sets. In any case, the price, product name, and a timestamp recording when the content was accessed need to be stored to ensure a consistent output over time. If available and desired, additional information, such as package size, customer rating, category, country of origin, labels, etc., can be included. Recording a unique product ID and the URL can help to trace any irregularities.

In principle, the script imitates a web user, navigates through the sites, and extracts the pre-defined information. To keep the websites' traffic at a moderate rate, this download should happen with some delay between the requests (HEMENWAY and CALISHAIN, 2004).

Figure 1: Schematic Web-scraping Procedure



Source: Own representation, icons made by Freepik from www.flaticon.com

How to do that

Figure 1 gives a schematic overview of the several steps of building a web scraper. Technically, the script can be coded in most of the common programming languages (HEMENWAY and CALISHAIN, 2004; MASSIMINO, 2016). In some languages, such as Python or R, there are even pre-programmed packages available for this purpose. While such code elements are certainly helpful as building blocks, there is always a need to adapt the script to the targeted website, depending on how the website is set up, the page structure, authentication requirements, etc. As many websites use reactive elements, simple HTTP requests cannot be used. Instead, the script needs to navigate a real web browser and "click" through the website. Also, sufficient time should be allocated for debugging and testing. Once the script is written and tested, its execution can be fully automated: A scheduler starts the download at defined time intervals. However, the scraper can be sensitive to changes in a website's layout, product group structure, or other even minor changes. To be aware of such issues it may make sense to build in alerts, e.g., sending an e-mail if a script did not run through completely, or if the download size is unusually small. The accessed data can simply be saved in any format, e.g., as a text or CSV file, with defined elements (name, ID, price, timestamp, etc.). For each observation over time, the new data are simply added to this file, resulting in one consistent data set. If several websites (e.g., different retailers or locations) are scraped, separate files can be created, but formatting should be the same, simplifying later analyses.

As most applications require daily scraping, one may not want to occupy the personal computer with this task. Alternatives are a private server or dedicated single-board computers such as the *Raspberry Pi*.

If technological infrastructure or coding know-how are insufficient to build one's own solution, there are commercial providers of web-scraping services. However, such commercial options may lack the transparency and open code documentation required for most scientific research and peer-reviewed publications. For such purposes, it may be more appropriate to use and adapt open source libraries.

2.3 Advantages

For food price research, web scraping is a promising new method of data collection, as it helps to overcome some issues of traditional data sources, such as official statistics and scanner data, as Table 2 shows.

Table 2: Alternative data source comparison

	Scraped Data	Scanner Data	National Statistics Data*
Cost per observation	low	high	free**
Data frequency	daily	weekly	monthly
Real-time data	yes	no	no
Full product range	yes	no	no
Product details	yes	yes	limited
International comparability	yes	limited	limited
Transaction data	no	yes	yes (weighted)

* E.g., price indices ** if publicly available

Source: Own representation based on CAVALLO and RIGOBON 2016, p. 156

Low costs

Good data sets are expensive. Especially at the retail level, access to highly frequent and disaggregated data such as scanner data can be very costly. Collecting prices via web scraping is in principle free if done with open source software. The highest costs associated with web scraping consists of the time required to write and test the code. At the defined times when the script is executed, power and online access are necessary. With all this in place, scalability across countries and products is very high, decreasing the marginal cost per observation to almost zero (CAVALLO, 2015). Alternatively, if the budget allows, tasks can be outsourced to an increasing number of commercial providers offering *Data as a Service (DaaS)* (MASSIMINO, 2016).

Frequent, real-time sampling

Once the script is written, it is up to the user whether it should run and extract prices and other data monthly, weekly, daily, or even at a higher frequency (e.g., hourly, as done by ELLISON and ELLISON, 2009). For food price research, daily data is probably sufficient for most applications. Such a high sampling frequency allows for more detailed analysis of price dynamics and for the application of other statistical methods, compared to the analysis of time-aggregated price data (EDELMAN, 2012). Further, as data are collected in real time, availability is given without any publication delay. This is an advantage for analyzing very recent events or policy changes, as well as forecasting.

Product range and details

In official statistics, consumer prices are mostly reported for widely defined product or product category levels. Further details such as package size, brand, quality differentiation, etc., which may be interesting for food price analysis, are generally not available. In contrast, web scraping can be used to extract prices for precise products, including all desired product attributes. This may be relevant information for some research questions. If not, taking this as a starting point, the researchers themselves can aggregate to the level they need, using the methods they consider appropriate.

Store type

Most secondary data simply disclose that they were collected at "retail level," or in some lucky cases it is distinguished between supermarkets, discounters, and small single retailer stores. With web scraping, the researchers decide which stores, retailers, wholesalers, or online delivery service they are interested in and collect those prices, without any hidden aggregation. To access different websites in regions and countries, self-identification, e.g., through entering a ZIP code or choosing a country, may be necessary and can be included in the web scraper code.

Transparency and customizing

Finally, web scraping allows the creation of customized data sets, completely targeted to the respective needs and fully transparent in how the data were obtained, without omitted variables or black boxes. If open source programs are used and code and data are shared, this transparency and reproducibility is even given to the whole scientific community. Enabling researchers to integrate data collection into their empirical work instead of relying on "given" official or purchased data may improve the quality and precision of empirical research (CAVALLO and RIGOBON, 2016).

2.4 Limitations

Despite all the above-named advantages, the method has some limitations.

No historic data

Web scraping means collecting real-time data. Hence, to come up with a sufficiently long time series, one needs to start data collection from day zero of the respective time period. For ad-hoc analyses, this may not always be possible. Not being able to access historic data is certainly a drawback of data collection via web scraping.

Too Big Data

We saw that the marginal cost and effort of scraped data are minimal. Hence, it may be tempting to collect literally *big* data beyond what is needed to answer a given research question, or to start collecting data even before having a well-defined research question. Then, we may end up doing purely explorative data mining instead of theoretically motivated research (MASSIMINO, 2016). While computing power and storage possibilities are constantly improving, adequately analyzing the data may be challenging, especially for professionals and researchers in food price analysis who have little experience with big data handling. When considering web scraping, it is also worth considering what is sufficient for the planned research; regarding product scope, time frame, frequency, and level of detail. For example, if aiming to measure overall food price development in a country or region, the selection of representative retailers and product categories will be a core part of the research.

No transaction data

Generally, web scraping extracts only prices, but no data on how often products are clicked on, or eventually purchased, because these items of information are not publicly available. This lack of transaction data certainly is a drawback, as it may be that, especially for high prices, no purchase is done at all, making the price completely irrelevant to consumers (CHEVALIER and GOOLSBEE, 2003). However, GORODNICHENKO et al. (2014) compared such unweighted prices to price quotes weighted by clicks and found that they were quite similar. Categories like "bestsellers," or sorting by "most popular" can help to give some indication on frequently purchased products, if possible on a given website.

Online availability

Especially in some developing countries, in which food security, and hence food prices at consumer level are an issue and of interest for research, online availability of prices on standardized websites may still not be given for the majority of transactions, particularly if informal markets are taken into account. Yet, even in many developing countries, those prices available online may be a more reliable source than official statistics (CAVALLO, 2013). Moreover, the availability of prices published online is increasing globally. Regarding online grocery shopping, Asia-Pacific is leading in terms of market share. Developing markets in all parts of the world are quickly catching up, as internet and smartphone penetration rates increase, at least in urban areas (NIELSEN N.V., 2017).

Legal and ethical limitations

Web scraping itself is a technology and is not per se illegal or legal. Rather, one must assess the legal situation carefully for each individual application. Obviously, only public content should be accessed and copyright policies must be complied with. When downloading and using someone else's data, the respective *Terms of Use* or *Terms of Service* apply. Reading the Terms of Use and the *Robots Exclusion Protocol* (robot.txt file) are good starting points to see whether scraping data from a site is allowed or not (KIENLE et al., 2004). The robots.txt file is a clearly codified access policy in a standardized format and can be found at the URL http://[www.domain.com]/robots.txt (HEMENWAY and CALISHAIN, 2004). The Terms of Use are often available somewhere on a website, but one does not explicitly need to agree with these ("browse-wrap agreement"). Whether they are enforceable contracts and legally binding has been examined by courts with different outcomes on a case-by-case basis (TOTO and BUFFINGTON, 2016).

Also other aspects of how to deal with web scraping are still subject to discussion in the legal literature (see, for discussion, HIRSCHEY, 2014; ZHU and MADNICK, 2010) and the law is still evolving. Given this uncertain environment in most jurisdictions, one may want to seek professional legal advice before starting a larger research project.

3 Areas of research

So far, many web-scraping applications have been used to analyze online pricing of non-food consumer goods. As large-scale online retail started with product segments such as books and electronics, also most early research about pricing on the internet focused on these products (e.g., Bakos, 1997 and 1998; Chevalier and Goolsbee, 2003). Gorodnichenko and Talavera (2017) conducted a large study on online prices for more than 100,000 goods over five years, but even here fresh food products are not represented.

3.1 MIT Billion Prices Project

The largest web-scraping effort for scientific purposes is MIT's Billion Prices Project, launched in 2008. Initially looking for an alternative to manipulated official statistics about Argentina's inflation rates (Cavallo, 2013), the price collection through web scraping was soon scaled up to other countries and an even broader product range. Cavallo and Rigobon (2016) describe how the literally billions of online prices collected in this project do not only improve inflation measurement, but also help to answer other questions in international and macroeconomics. This project does pioneer work not just regarding the scope of the data collection, but also with its approach to make much of the data publicly accessible to other researchers.

3.2 Online retail pricing strategies

It is often assumed that online posted prices are more flexible compared to offline prices, which display rather rigid pricing patterns, except for temporary promotions (Herrmann et al., 2005). In online markets, menu costs to change prices are negligible, allowing to adjust prices in a high frequency, reacting to demand and supply changes, and ultimately leading to more efficient markets (Gorodnichenko and Talavera, 2017; Smith et al., 2001; Tang and Xing, 2001). Also on the consumer side, search costs for prices have decreased, and price comparison for a defined good is quick and easy, especially thanks to price comparison websites (Bakos, 1997; Gorodnichenko and Talavera, 2017).

This flexibility on the seller and buyer side is assumed to increase price conversion and price transmission. No longer having to deal with menu and search cost, which are commonly used to explain why the *law of one price* does not hold, there is hope that online prices allow for new insight regarding price transmission and conversion (Gorodnichenko et al., 2014). Some empirical studies have shown that online prices change more frequently, and in smaller magnitude, than offline prices (Brynjolfsson and Smith, 2000; Ellison and Ellison, 2009). However, these data were collected on marketplaces such as *eBay* or price comparison tools (*Google Shopping*), which may not be representative for overall online retail, as online retailers are heterogeneous in their characteristics and price setting (Einav et al., 2018; Pan et al., 2002).

An alternative strand of literature concludes that exactly the opposite is true, and prices do not converge more, but online markets give even more room to differentiate between customer groups with different price sensitivity and to apply targeted price discrimination (Ancarani, 2002; Baylis and Perloff, 2002). On top of that, the convenience of online shopping and home delivery may attract less price-sensitive consumer segments (Degeratu et al., 2000).

Online retail could even allow for *dynamic pricing* based on the analysis of current and past customer demand, competitor price setting and other factors such as holidays or weather conditions (Grewal et al., 2011; Shpanya, 2013). Detecting such practices, however, would require a very advanced web-scraping code, pretending to log on from different IPs, with different user profiles.

3.3 Online grocery retail

We saw that already in the quite well-studied non-food sector, there are contradictory findings. For online grocery retail, and fresh products in particular, even less is known about online pricing and empirical studies are scarce.

There is no clear evidence regarding *price level differences* between online and offline grocery sellers. Ad-hoc sample observations in the UK and U.S. suggest that online grocery retailers were more expensive for a long time, but have recently lowered prices, even below established offline supermarket prices (OLIVER WYMAN, 2014 and 2018).

CAVALLO (2017) included food items in his comparison of on- and offline prices, conducted in ten countries between 2015 and 2016. For the food subsample, there was a small markup of online prices (1%), while non-food items were on average even slightly cheaper online (drugstore -3%, household -2%). The results differ among countries and only measure within-retailer price dispersion for multi-channel retailers. Online-only retailers or brick and mortar stores, such as traditional small shops or discount stores, were not considered. Yet, it gives some indications that there are differences between food and non-food online pricing.

Also, the argument of reduced search costs through online price comparison websites does not seem to hold for online food retail. Currently, such tools are mostly available for homogeneous, durable, and easy-to-ship products, but not for (fresh) foods, with differences in appearance, freshness, and taste, and presumably limited arbitrage opportunities (GORODNICHENKO and TALAVERA, 2017). Further, groceries are generally not bought as individual items, but bundled (e.g., per week or at least per meal), making an individual price comparison futile.

FEDOSEEVA et al. (2017) analyze the price setting in the German online chocolate market and find no evidence for more homogeneous prices among different sellers (lower search cost hypothesis) or more frequent price adjustment (lower menu cost hypothesis) than in offline markets. Yet, as data were collected manually, sample size and time span are rather low. Applying web-scraping methods could help to understand pricing in the growing online food retail business.

3.4 Overall food price development

Certainly, web scraping can only access prices published online. Yet, it also gives insight into offline and overall price development. As discussed above, the online–offline level difference for food seems to be low in most countries (CAVALLO, 2015 and 2017; OLIVER WYMAN, 2014 and 2018). Also, in many countries offline retailers have started to publish their prices online (NIELSEN N.V., 2017). Hence, web scraping can provide data on overall price developments. Already today, this method is used to calculate general consumer price indices, both in research and by official statistical agencies and national banks (CAVALLO and RIGOBON, 2016).

3.5 Beyond consumer price research

So far we have focused on consumer price research, as these data are currently most widely available online. However, as online publishing becomes more common, applications could extend to wholesale and production level, as well as farm input prices.

Besides the use for scientific research, there are also more practical field applications. Scraping market prices and forwarding relevant price developments to farmers' mobile phones could help them to improve agricultural production and marketing decisions. There is plenty of evidence that access to information technology can improve farmers' and overall welfare (e.g., JENSEN, 2007; AKER, 2011). CAMACHO and CONOVER (2011) found that Colombian farmers who receive targeted digital price and weather information on their phones considered this useful and managed to reduce their crop loss.

Such projects could be scaled up at relatively low cost, exploiting the high scalability of web scraping. Also, the technology is not limited to prices but can be used for other reasonably well-structured information available online, such as weather data (YANG et al., 2010).

4 Conclusion

We have shown that web scraping is a promising, low-cost data collection method for food price research. Using web scraping to build customized data sets can help to overcome common problems such as incomplete data, omitted variables, and sample selection. So far, the main application consists of online retail prices because of their good availability. Data obtained through web scraping can help to analyze how prices are set in the growing online food retail, and how this may impact the value chain and potentially change global food systems. Also, online prices give insight into general (offline) price development and can help to fill in gaps where public data sources are not available or reliable. As more and more price data are published online, areas of application will also extend. Especially for global data collection and comparisons, this method will open up new research opportunities. Exploiting this new technology gives us the chance to find new answers to old questions, or even to ask new questions. Once aware of the technical possibilities, researchers and practitioners may come up with new innovative applications.

However, we saw that there are some limitations that may discourage or inhibit individual researchers to engage in web scraping, such as legal uncertainties, unavailability of historic data, and technical difficulties. Working together and centralizing data collection can help to overcome these barriers (MASSIMINO, 2016). Larger data collection projects could bundle technical and legal expertise and exploit the scalability of web-scraping technologies. A good example of such a project is the previously discussed MIT Billion Prices Project, making their scraped data publicly available for scientific use. A similar web-scraping project could be started for food price research, collecting and publishing consistent and cleaned data. After some time of ongoing real-time data collection, the data could benefit a wide range of researchers in agricultural economics, development economics, and related disciplines.

Yet, some universities, institutes, or associations would need to take the lead and make an initial investment. In a first step, a public repository could be organized, where researchers can upload and share their scraped data sets that may be relevant to colleagues who also work with food prices. While competition for the best scripts and biggest data may be, in part, fruitful, cooperating and sharing data sources and methods to advance in research is likely to lead to less biased and broader datasets as well as to better results.

References

AKER, J. C. (2011). Dial "A" for agriculture: a review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics*, *42*(6), 631–647.

ANCARANI, F. (2002). Pricing and the Internet: Frictionless Commerce or Pricer's Paradise? *European Management Journal*, *20*(6), 680–687.

BAKOS, J. Y. (1997). Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, *43*(12), 1676–1692.

BAKOS, Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM*, *41*(8), 35–42.

BAYLIS, K., PERLOFF, J. M. (2002). Price dispersion on the Internet: Good firms and bad firms. *Review of Industrial Organization*, *21*(3), 305–324.

BRYNJOLFSSON, E., SMITH, M. D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science, 46(4)*, 563–85.

CAMACHO, A., CONOVER, E. (2011). The impact of receiving price and climate information in the agricultural sector. IDB Working Paper Series, No. IDB-WP-220.

CAMPBELL, J. R., EDEN, B. (2014). Rigid prices: Evidence from U.S. scanner data. *International Economic Review*, *55*(2), 423–442.

CAVALLO, A. (2013). Online and Official Price Indexes: Measuring Argentina's Inflation. *Journal of Monetary Economics,* 60(2), 152–65.

CAVALLO, A. (2015). Scraped data and sticky prices. NBER Working Paper No. 21490.

CAVALLO, A. (2017). Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers. *The American Economic Review, 107*(1), 283–303.

CAVALLO, A., RIGOBON, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives, 30*(2), 151–178.

CHEVALIER, J., GOOLSBEE, A. (2003). Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative marketing and Economics*, *1*(2), 203–222.

COTTERILL, R. W. (1994). Scanner data: new opportunities for demand and competitive strategy analysis. *Agricultural and Resource Economics Review*, *23*(2), 125–139.

DEGERATU, A. M., RANGASWAMY, A., WU, J. (2000). Consumer Choice Behavior in Online and Traditional Supermarkets: The Effects of Brand Name, Price, and Other Search Attributes. *International Journal of Research in Marketing*, 17(1): 55–78.

EDELMAN, B. (2012). Using internet data for economic research. *The Journal of Economic Perspectives*, *26*(2), 189–206.

EINAV, L., FARRONATO, C., LEVIN, J., SUNDARESAN, N. (2018). Auctions versus posted prices in online markets. *Journal of Political Economy*, *126*(1), 178–215.

ELLISON, G., ELLISON, S. (2009). Search, Obfuscation, and Price Elasticities on the Internet. *Econometrica, 77(2)*, 427–52.

FEDOSEEVA, S., GREIN, T., HERRMANN, R. (2017). How German Online Retailers Price Foods: An Empirical Analysis for Chocolate Products. *International Journal on Food System Dynamics*, *8*(1), 32–44.

GORODNICHENKO, Y., SHEREMIROV, V., TALAVERA, O. (2014). Price Setting in Online Markets: Does It Click? (No. w20819). National Bureau of Economic Research.

GORODNICHENKO, Y., TALAVERA, O. (2017). Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration. *American Economic Review, 107*(1), 249–282.

GREWAL, D., AILAWADI, K. L., GAURI, D., HALL, K., KOPALLE, P., ROBERTSON, J. R. (2011). Innovations in retail pricing and promotions. *Journal of Retailing*, *87*, 43–52.

HEMENWAY, K., CALISHAIN, T. (2004). *Spidering Hacks*. Sebastopol, CA: O'Reilly Publishing. ISBN 0596005776.

HERRMANN, R., MOESER, A., WEBER, S. A. (2005). Price rigidity in the German grocery-retailing sector: scanner-data evidence on magnitude and causes. *Journal of Agricultural & Food Industrial Organization*, *3*(1), 1–37.

HIRSCHEY, J.K. (2014). Symbiotic Relationships: Pragmatic Acceptance of Data Scraping. *Berkeley Technology Law Journal, 29(4)*, 897–927.

JENSEN, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *The Quarterly Journal of Economics*, *122*(3), 879–924.

KIENLE, H. M., GERMAN, D., MULLER, H. (2004). Legal concerns of web site reverse engineering. *Web Site Evolution, Sixth IEEE International Workshop*, 41–50.

MASSIMINO, B. (2016). Accessing Online Data: Web-Crawling and Information-Scraping Techniques to Automate the Assembly of Research Data. *Journal of Business Logistics*, *37*(1), 34–42.

NAKASH, J., ANAS, S., AHMAD, S. M., AZAM, A. M., KHAN, T. (2015). Real Time Product Analysis using Data Mining. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *4*(3), 815–820.

NIELSEN N.V. (2015). The Future of Grocery. E-Commerce, Digital Technology and Changing Shopping Preferences around The World. Accessed 24 January 2018, available at https://www.nielsen.com/content/dam/nielsenglobal/vn/docs/Reports/2015/Nielsen%20Global%20E-Commerce%20and%20The%20New%20Retail%20Report%20APRIL%202015%20(Digital).pdf

NIELSEN N.V. (2017). What's In-Store for Online Grocery. Accessed 24 January 2018, available at http://www.nielsen.com/content/dam/nielsenglobal/de/docs/Nielsen%20Global%20Connected%20Commerce%20Report%20January%202017.pdf

OLIVER WYMAN (2014). AmazonFresh in the U.S. Accessed 24 January 2018, available at http://www.oliverwyman.de/content/dam/oliver-wyman/global/en/2014/aug/OW_AmazonFresh_ENG.pdf

OLIVER WYMAN (2018, January 4). Amazon Fresh Undercuts Supermarkets By 10–20 Percent. Accessed 24 January 2018, available at http://www.oliverwyman.com/our-expertise/insights/2018/jan/amazon-fresh-undercuts-supermarkets-by-1020-percent.html

PAN, X., RATCHFORD, B. T., SHANKAR, V. (2002). Can price dispersion in online markets be explained by differences in e-tailer service quality? *Journal of the Academy of Marketing science*, *30*(4), 433–445.

RIGBY, N. (2018, January 16). Amazon Grocery 2017 Review. *One Click Retail Study*. Accessed 24 January 2018, available at http://oneclickretail.com/amazon-grocery-2017-review/

SMITH, M. D., BAILEY, J., BRYNJOLFSSON, E. (2001). Understanding digital markets: Review and assessment. MIT Sloan School of Management Working Paper No. 4211-01.

SHPANYA, A. (2013, May 14). Five Trends to Anticipate in Dynamic Pricing [Blog post]. Retail Touchpoints. Accessed 24 January 2018, available at https://www.retailtouchpoints.com/features/executive-viewpoints/5-trends-to-anticipate-in-dynamic-pricing

SILVER, M., HERAVI, S. (2001). Scanner data and the measurement of inflation. *The Economic Journal*, *111*(472), 383–404.

TANG, F. F., XING, X. (2001). Will the growth of multi-channel retailing diminish the pricing efficiency of the web? *Journal of Retailing*, *77*(3), 319–333.

TOTO, C. S., BUFFINGTON, K. (2016, June 6). How Binding Is Your Browsewrap Agreement? [Blog post]. Pillsbury Winthrop Shaw Pittman LLP. Accessed 24 January 2018, available at https://www.lexology.com/library/detail.aspx?g=4b4b93da-c40a-4724-916c-3bdd9011698d

YANG, Y., WILSON, L. T., WANG, J. (2010). Development of an automated climatic data scraping, filtering and display system. *Computers and Electronics in Agriculture*, *71*(1), 77–87.

ZHU, H., MADNICK, S. E. (2010). Legal challenges and strategies for comparison shopping and data reuse. *Journal of Electronic Commerce Research*, *11*(3), 231–239.