

EXPLORING THE SUSTAINABILITY OF SWISS ONLINE SHOPS: PRELIMINARY EVIDENCE FROM A CLUSTERING APPROACH

THOMAS WOZNIAK, GUANG LU, DOMINIK GEORGI,
ANJA JANOSCHKA & ANTONIA STEIGERWALD

Lucerne School of Business, Institute of Communication and Marketing, Lucerne,
Switzerland; e-mail: thomas.wozniak@hslu.ch, guang.lu@hslu.ch,
dominik.georgi@hslu.ch, anja.janoschka@hslu.ch, antonia.steigerwald@hslu.ch

Abstract This paper proposes a framework for assessing the sustainability of online retailers. By adhering to the logic of the value chain, the framework captures all steps from procurement to shipping of products and handling product returns that typically occur in e-commerce. The framework is then applied to explore the sustainability of 227 Swiss online shops through a clustering approach. Core elements of the framework are dummified and condensed to 20 features using an autoencoder-based neural network. K-means clustering is used to identify three sustainability clusters that are visualised in the latent space. One-way ANOVA and chi-square analyses of the three-cluster solution indicate that the identified clusters are distinct across most elements of the proposed framework. These preliminary results hold a variety of future research avenues.

Keywords:
e-commerce,
online
shops,
sustainability,
clustering,
unsupervised
machine
learning

1 Introduction

Global e-commerce sales have continually grown over the past years and account for an increasing share of total retail sales. In 2019, emarketer predicted that global e-commerce sales would grow by 19 percent and account for 16 percent of total retail sales in 2020 (Lipsman, 2019). The recent COVID-19 pandemic has further increased the shift from offline to online retailing, especially due to social distancing requirements and lockdowns put in place to prevent spreading of the corona virus. For individual countries, e-commerce sales were thus predicted to grow more than 20 percent in 2020 (Cramer-Flood, 2020). It is expected that the accelerated trend towards online retailing will remain during the recovery from the COVID-19 pandemic (UNCTAD, 2021a).

E-commerce has environmental consequences (Fichter, 2003). Growth in e-commerce sales translates to, e.g., more intense use of ICT and logistics affecting the use of energy, material resources, and land, and the emission of greenhouse gas (Fichter, 2003). In light of climate change (Steffen et al., 2018) and even continents aiming to become climate-neutral (European Commission, 2020), analyzing the sustainability of e-commerce becomes more relevant than ever before.

This paper develops a framework for assessing the sustainability of online retailers (section 2.3), based on the concept of the value chain specifically in e-commerce (section 2.1) and considering extant research on sustainability in e-commerce (section 2.2). The framework serves as basis for developing a survey instrument for collecting empirical data from 227 online retailers in Switzerland (section 3.1). Using an autoencoder-based neural network, the data is condensed to 20 features for subsequent clustering (section 3.2.1). K-means clustering identifies three clusters of online shops which are visualized in the latent space (sections 3.2.2 and 3.2.3). The statistical analyses of the clusters indicate that they are distinct across most elements of the proposed framework (section 4). These preliminary results provide fertile ground for further and future research (section 5). Overall, this paper contributes to a better understanding of sustainability in e-commerce specifically by proposing an assessment framework, applying it to a unique dataset, and identifying three clusters of online shops according to their sustainability as assessed by our framework.

2 Conceptual background and framework development

2.1 The e-commerce value chain

The concept of the value chain, developed by (Porter, 1985), is widely used by academia and practice alike. Following a linear logic, the value chain entails five subsequent “generic categories of primary activities [...] involved in the physical creation and delivery of the product to the customers” (Ricciotti, 2020, p. 193), namely inbound logistics, operations, outbound logistics, marketing and sales, and service (Porter, 1985). These sequential primary activities are sustained by four support activities such as procurement, technology development, human resource management, and firm infrastructure (Porter, 1985; Ricciotti, 2020). Several scholars have adapted the generic value chain concept to more specific contexts (Ricciotti, 2020). Some have departed from its sequential and linear logic and embraced a network logic to better reflect that firms are part of stakeholder networks that compete with each other, resulting in the concept of the value network (Ricciotti, 2020). While such network logic may more closely represent today’s competitive environment, the simplicity of the value chain has not lost its appeal.

In the (online) retailing value chain, the most important players are manufacturers of products, institutional retailers, and consumers (Reinartz, Wiegand, & Imschloss, 2019). Such retailers may be traditional stationary retailers who also sell their products via online channels as part of multichannel strategies or so-called pure plays who sell online only (Reinartz et al., 2019). Manufacturers have also started running their own online operations aiming to engage directly with end customers, thereby circumventing retailers (Reinartz et al., 2019). In this paper, we focus on the value chain from the perspective of the online seller of products. This can include any type of online retailer as well as manufacturers maintaining direct-to-consumer (D2C) online operations. For selling products online, retailers and manufacturers need to run a website with an online shop.¹ Products sold online are typically stored in a warehouse. Once ordered, they are packaged, and delivered to the customer. The customer receives the product, keeps it, or may decide to return it.

¹ Some online sellers may choose to sell via downloadable apps or social media platforms such as Facebook or Instagram only. However, it can be assumed that most online sellers run a web-based online shop to reach the largest number of potential customers.

2.2 Sustainability in e-commerce

Research on the sustainability in e-commerce has addressed a wide array of aspects. This includes comparing the environmental impact of brick-and-mortar retail channels with online channels (Bertram & Chi, 2018; Pålsson, Pettersson, & Winslott Hiselius, 2017; P. Van Loon, McKinnon, Deketele, & Dewaele, 2014; Weber et al., 2009), sustainability in e-commerce packaging (Escursell, Llorach, & Roncero, 2020), sustainability of e-commerce logistics (Mangiaracina, Marchet, Perotti, & Tumino, 2015), lifecycle analysis of different fulfilment channel types (Patricia Van Loon, Deketele, Dewaele, McKinnon, & Rutherford, 2015), e-commerce customers' last-mile delivery preferences based on the sustainability impact (Ignat & Chankov, 2020), and product returns' effect on business, society, and the environment (Frei, Jack, & Brown, 2020). While these contributions are diverse in the topics they address and the methods applied, they have one important denominator in common: they address or can be related to specific segments or phases of the e-commerce value chain. This underlines the suitability of the value chain concept as a frame of reference for analyzing sustainability in e-commerce.

2.3 A framework for assessing the sustainability of online retailers

Adhering to the logic of the value chain and considering extant literature, we identified six core elements of the e-commerce value chain that need to be considered when assessing the sustainability of online shops: 1) *products*, 2) *online shop operations*, 3) *intralogistics and storage*, 4) *packaging*, 5) *delivery*, and 6) *product returns*. Some of these elements may involve the cooperation with other companies (e.g. suppliers or fulfilment/logistics service providers). We thus included 7) *cooperation with partners* as a further core element. In our framework (see Figure 1), this element is depicted in parallel to the six linear elements since it can relate to any of them.

To emphasize the goal of this framework, namely to guide assessing online retailers' sustainability, we nest the core framework within the society and the environment. The link between online retailers and the society is established through interactions with their stakeholders which are part of the society. Such stakeholders can include partners of the company (e.g. suppliers), employees, and customers, but also consumer and trade organizations, and entities such as states with their regulatory requirements. Communication plays a major role for such interactions (Lim &

Greenwood, 2017). We thus include *communication with stakeholders* as further element in our framework.

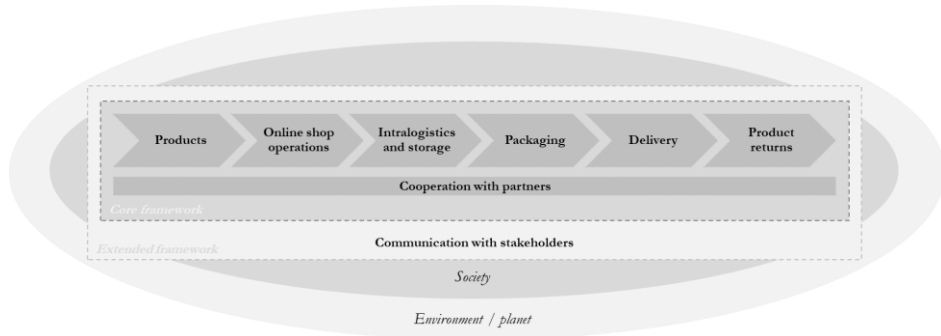


Figure 1: E-commerce sustainability assessment framework nested in society and environment

Nesting organizations and their processes and/or business models within external environments in the sense of context is not new. Prominent and well-known examples are the St. Gallen Management Model (Rüegg-Stürm & Grand, 2019) and the environment of the business model canvas (Osterwalder & Pigneur, 2010). Our approach is distinct from these examples in two ways: First, we purposely focus on society and environment, thereby simplifying the manifold and complex interactions between many types of stakeholders. Second, we nest our framework within society which is in turn nested within the environment. This is in line with that paradigm of sustainability which proposes the economy depends on the society and both depend on the environment, in that order (e.g. Griggs, 2013).

3 Methodology

3.1 Development of survey instrument, data collection, and sample

The framework developed in section 2.3 served as foundation for developing the survey instrument. For each of the framework's core elements, survey questions and items were designed to adequately address the specific nature of each core element. Further survey questions addressed communication and marketing, the assessment of customer needs, strategic priorities, and questions determining the size and nature

of the online shop (e.g. product categories and annual order volume). The final survey instrument was reviewed and pre-tested by experienced researchers, market research professionals, and e-commerce practitioners.

The online survey targeted online shops specifically operating in Switzerland and was administered from March to July 2020. Switzerland ranks first in the UNCTAD B2C E-commerce Index 2020, reflecting high scores in all four dimensions of the index, namely Internet penetration, banking coverage, secure server density, and postal reliability (UNCTAD, 2021b). An invitation to the online survey was sent to a large number of Swiss online shops. In addition, the link to the survey was distributed in social media and specific newsletters to reach out to the target group. The final sample consisted of n=227 completed online surveys.

In our sample, the most frequent product categories were fashion and food (27% each), houseware (21%), cosmetics and sporting goods (16% each), and toys (13%). In terms of business relationships, the majority stated to serve B2C customers (81%), which was the primary target group of the survey; 51% stated to serve B2B customers; 25% stated to follow a D2C model. Our sample included online shops with different annual order volumes: 32% reported less than 1,000 orders, 36% 1,000 to 9,999 orders, and 32% 10,000 and more orders in 2019.

3.2 Three-step clustering approach

3.2.1 Data preparation

From the overall survey instrument, we have selected a total of 10 questions to adequately cover all elements of our core framework, i.e. from products to returns, and cooperation with partners. The dummification of the relevant variables lead to a dataset containing 227 online shops and 150 features having only categorical values of “0” or “1”.

An autoencoder-based neural network has been used to perform the dimension reduction on the dataset obtained from the previous step. This is beneficial to reducing the noises underlying the original data, which is helpful for the clustering. Compared to the conventional linear dimension reduction techniques using e.g. Principal Component Analysis (PCA), autoencoder-based neural networks are

claimed to be more powerful as they can capture nonlinear correlations between features and perform nonlinear transformations to the data (Alkhayrat, Aljnidi, & Aljoumaa, 2020).

Figure 2 displays the structure of the stacked autoencoder. To train this autoencoder, 20,000 epochs have been run, with a batch size 32, a validation split 0.1, and a shuffle variable being set as “true”. Independent upon the size of the hidden layer, both the train and validation loss curves are getting flat after an epoch of approximately 10,000. At the end of the training process, the train loss and the validation loss are the smallest for a hidden layer size of 20, i.e. 0.1472 and 0.1450, respectively. Therefore, the original 150 features have been “condensed” into 20 new features in the latent space for further clustering analyses.

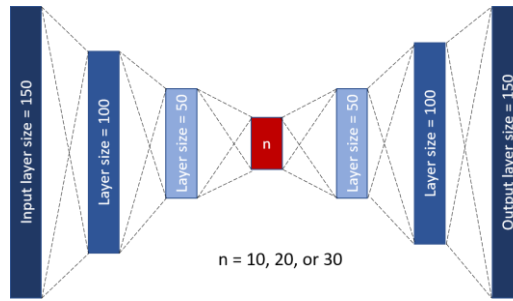
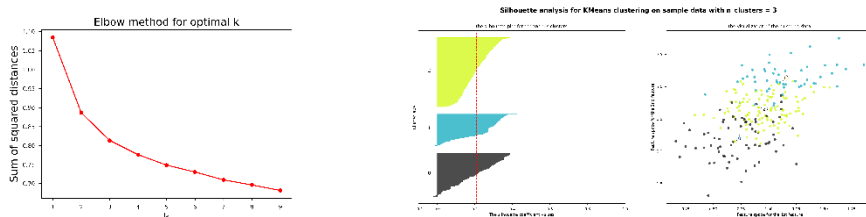


Figure 2: Structure of the stacked autoencoder. The input layers and the output layers are symmetric. n represents the number of neural nodes in the middle layer, which is varied using 10, 20, or 30

3.2.2 Clustering method

The conventional, unsupervised clustering method K-Means has been utilized to group the online shops that potentially have similar features. The elbow method has been used to determine the optimal number of clusters. Basically, the elbow method calculates the sum of squared distance between neighboring points (online shops) in the determined latent space. From Figure 3a, a clear “elbow” can be seen when the number of clusters equals 3. To verify the chosen number of clusters, a silhouette analysis has been performed for a cluster size of 2, 3, 4, 5, and 6. Correspondingly, the silhouette coefficients obtained are 0.289, 0.211, 0.159, 0.151, and 0.148. A cluster size of 4, 5, or 6 is relatively not good as it leads to a below-average silhouette

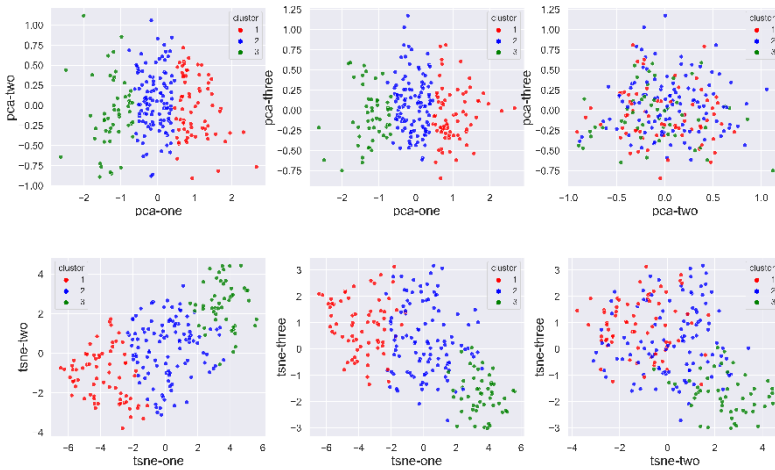
coefficient. In the end, the size of clusters is determined as 3. The online shops are grouped into three clusters (C1, C2, and C3) using the K-Means algorithm.



Figures 3a and b: Elbow method for determining the optimal number of clusters (a; left) and silhouette analysis for verifying the three-cluster solution (b; right)

3.2.3 Cluster visualization

To visualize the determined clusters of online shops on two-dimensional diagrams, both PCA and t-distributed stochastic neighbor embedding (t-SNE) methods are used. PCA is a deterministic, linear dimensionality reduction technique, which keeps the global structure of data by preserving their variance. Here, the first, second, and third principal components represent 78.4% of the variance in the original data, which is quite good. In comparison, t-SNE is a non-deterministic, non-linear dimensionality reduction technique, which is stronger at capturing local structures within the data and preserves distances between data points instead of their variances. In Figure 4, both the PCA and the t-SNE diagrams exhibit clear boundaries between clusters that can be distinguished mainly using the first principal component.



Figures 4a and b: PCA 2D plots (a; top) vs. t-SNE 2D plots (b; bottom)

4 Results

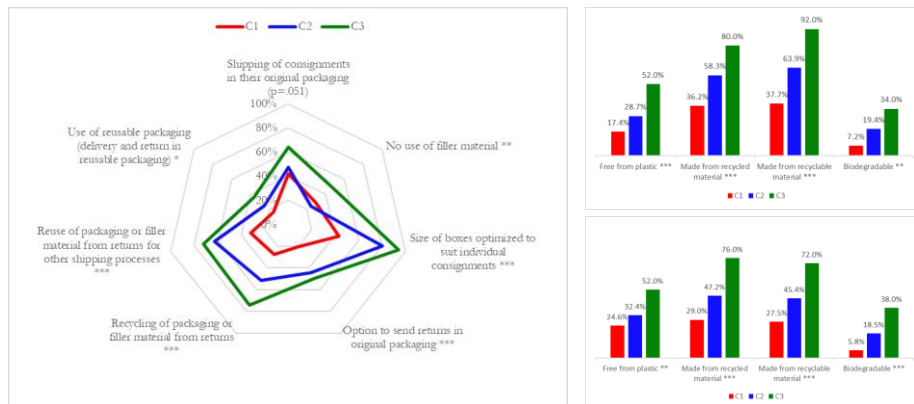
In the survey instrument, we used five items on a 5-point scale to measure sustainability related to the element *products*. Similarly, five 5-point-scale items were used to assess the elements *online shop operations*, three items for *intralogistics and storage*, six items for *product returns*, and seven items for *cooperation with partners*. One-way ANOVA was performed to detect statistically significant differences between the clusters. Table 1 details the results for two exemplary items per framework element, highlighting distinct characteristics of each cluster.

Table 1: ANOVA summary table for two exemplary items for each of the following elements: products PROD, online shop operations OSO, intralogistics and storage ILS, product returns PR, and cooperation with partners CP

Element of framework / Survey item		df	F <i>p</i>	C1	C2	C3
				M <i>SD</i>	M <i>SD</i>	M <i>SD</i>
PROD	Environmental aspects (e.g. resource conservation ... during manufacture) are just as important as product costs.	2, 223	6.1 .003	3.57 1.23	3.43 1.28	4.16 1.17
	Social aspects (e.g. fair working conditions during manufacture) are just as important as product costs.	2, 223	4.1 .018	4.03 1.16	4.06 1.11	4.54 0.91

OSO	We use IT solutions that leave the smallest possible carbon footprint (manufacture, operations, disposal).	2, 188	2.5 .081	2.95 1.33	2.81 1.32	3.37 1.54
	We use IT solutions that have minimal power requirements (incl. power for cooling).	2, 186	6.7 .002	2.85 1.35	2.78 1.40	3.67 1.34
I.L.S.	Our warehouse is operated in a resource-friendly manner (e.g. low power consumption).	2, 210	8.9 .000	3.62 1.07	3.76 1.19	4.45 0.91
	We optimize transport capacities ecologically (e.g. to branches or distribution center).	2, 209	12.4 .000	3.37 1.18	3.39 1.28	4.36 0.99
PR	Returns are passed on to specialized retailers.	2, 194	9.5 .000	2.12 1.27	1.49 1.06	1.22 0.80
	Returns are destroyed.	2, 203	3.2 .042	2.11 1.30	2.13 1.36	1.59 1.02
CP	When choosing possible partners, we consider environmental aspects (e.g. resource conservation, avoidance of carbon emissions).	2, 220	7.5 .001	3.18 1.21	2.87 1.36	3.74 1.35
	When choosing possible partners, we consider social aspects (e.g. fair working conditions).	2, 216	6.5 .002	3.62 1.28	3.43 1.45	4.27 1.24

For the element *packaging*, we developed lists of specific characteristics of the packaging and the packaging system as multiple-answer checkbox-type questions (e.g. if the packaging material is free of plastic or if the size of boxes is optimized to suit individual consignments). For the element *delivery*, we developed a similar list for a variety of respective options (e.g. carbon-neutral shipping or bundling of partial deliveries). Figures 5 and 6 detail how the respective characteristics and options vary across the clusters. Cross tabulation and chi-square analyses revealed that most of the characteristics of the packaging systems are unequally distributed across the clusters at $p < .001$ (***), $p < .01$ (**), or $p < .05$ level (*; see Figure 5a). Characteristics of the packing material (see Figure 5b) and the filler material (see Figure 5c) are unequally distributed across the clusters at $p < .001$ (***) or $p < .01$ (**) level. Taken together, these results suggest that the identified clusters are quite distinct with regards to the *packaging* core element of our framework. Out of the seven *delivery* options displayed in Figure 6, three are unequally distributed across the clusters at $p < .001$ (***), $p < .01$ (**), or $p < .05$ level (*), namely consolidated shipping, use of locally adapted logistics solutions, and delivery to third-party collection points.



Figures 5a, b, and c: Packaging system (a; left), packaging material (b; upper right), and filler material characteristics (c; lower right) – percentages by cluster

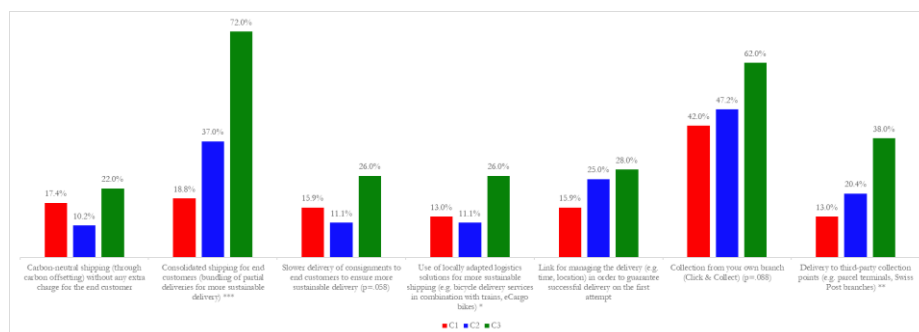


Figure 6: Delivery options offered to customers – percentages by cluster

5 Conclusion and outlook

Our clustering results indicate that online shops can be grouped into three clusters regarding their sustainability. With regard to many aspects of our proposed framework, the clusters are distinct from each other. Among them, cluster 3 has the most distinct characteristics and thus appears to comprise online shops that are most sustainable according to the criteria covered by our survey instrument. Further analysis of the present dataset needs to test to which extent the identified clusters also vary by structural properties of the online shops such as product categories, order volume, or for example D2C model. Future research should also investigate

the use of communication channels, possible differences between pure-play and omnichannel strategies, and potential latent drivers of sustainability engagement such as the values of the organizations behind the online shops.

Acknowledgements

The authors are grateful to Swiss Post for making possible this preliminary in-depth analysis of the original survey data. Detailed descriptive results of the survey are available at www.swisspost.ch/digital-commerce-studies.

References

- Alkhayrat, M., Aljnidi, M., & Aljoumaa, K. (2020). A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-0286-0>
- Bertram, R. F., & Chi, T. (2018). A study of companies' business responses to fashion e-commerce's environmental impact. *International Journal of Fashion Design, Technology and Education*, 11(2), 254–264. <https://doi.org/10.1080/17543266.2017.1406541>
- Cramer-Flood, E. (2020). Global Ecommerce 2020. Retrieved from <https://www.emarketer.com/content/global-ecommerce-2020>
- Escursell, S., Llorach, P., & Roncero, M. B. (2020). Sustainability in e-commerce packaging: A review. *Journal of Cleaner Production*, 280. <https://doi.org/10.1016/j.jclepro.2020.124314>
- European Commission. (2020). 2050 long-term strategy. Retrieved October 19, 2020, from https://ec.europa.eu/clima/policies/strategies/2050_en
- Fichter, K. (2003). E-Commerce: Sorting Out the Environmental Consequences. *Journal of Industrial Ecology*, 6(2), 25–41. <https://doi.org/10.1162/108819802763471762>
- Frei, R., Jack, L., & Brown, S. (2020). Product returns: a growing problem for business, society and environment. *International Journal of Operations & Production Management*. <https://doi.org/10.1108/IJOPM-02-2020-0083>
- Griggs, D. (2013). Sustainable development goals for people and planet. *Nature*, 495, 305–307. <https://doi.org/10.1038/495305a>
- Ignat, B., & Chankov, S. (2020). Do e-commerce customers change their preferred last-mile delivery based on its sustainability impact? *International Journal of Logistics Management*, 31(3), 521–548. <https://doi.org/10.1108/IJLM-11-2019-0305>
- Lim, J. S., & Greenwood, C. A. (2017). Communicating corporate social responsibility (CSR): Stakeholder responsiveness and engagement strategy to achieve CSR goals. *Public Relations Review*, 43(4), 768–776. <https://doi.org/10.1016/j.pubrev.2017.06.007>
- Lipsman, A. (2019). Global Ecommerce 2019. Retrieved October 13, 2020, from <https://www.emarketer.com/content/global-ecommerce-2019>
- Mangiaracina, R., Marchet, G., Perotti, S., & Tumino, A. (2015). A review of the environmental implications of B2C e-commerce: a logistics perspective. *International Journal of Physical Distribution and Logistics Management*, 45(6), 565–591. <https://doi.org/10.1108/IJPDLM-06-2014-0133>
- Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken, NJ: John Wiley & Sons.
- Pålsson, H., Pettersson, F., & Winslott Hiselius, L. (2017). Energy consumption in e-commerce versus conventional trade channels - Insights into packaging, the last mile, unsold products and product returns. *Journal of Cleaner Production*, 164, 765–778.

- <https://doi.org/10.1016/j.jclepro.2017.06.242>
- Porter, M. E. (1985). *The Competitive Advantage: Creating and Sustaining Superior Performance*. New York: Free Press.
- Reinartz, W., Wiegand, N., & Imschloss, M. (2019). The impact of digital transformation on the retailing value chain. *International Journal of Research in Marketing*, 36(3), 350–366. <https://doi.org/10.1016/j.ijresmar.2018.12.002>
- Ricciotti, F. (2020). From value chain to value network: a systematic literature review. *Management Review Quarterly*, 70(2), 191–212. <https://doi.org/10.1007/s11301-019-00164-7>
- Rüegg-Stürm, J., & Grand, S. (2019). *Managing in a Complex World*. Bern: Haupt.
- Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., ... Schellnhuber, H. J. (2018). Trajectories of the Earth System in the Anthropocene. *Proceedings of the National Academy of Sciences of the United States of America*, 115(33), 8252–8259. <https://doi.org/10.1073/pnas.1810141115>
- UNCTAD. (2021a). COVID-19 and e-commerce: a global review.
- UNCTAD. (2021b). *The UNCTAD B2C E-commerce Index 2020*.
- Van Loon, P., McKinnon, A. C., Deketele, L., & Dewaele, J. (2014). The growth of online retailing: A review of its carbon impacts. *Carbon Management*, 5(3), 285–292. <https://doi.org/10.1080/17583004.2014.982395>
- Van Loon, Patricia, Deketele, L., Dewaele, J., McKinnon, A., & Rutherford, C. (2015). A comparative analysis of carbon emissions from online retailing of fast moving consumer goods. *Journal of Cleaner Production*, 106(2015), 478–486. <https://doi.org/10.1016/j.jclepro.2014.06.060>
- Weber, C. L., Hendrickson, C. T., Matthews, H. S., Nagengast, A., Nealer, R., & Jaramillo, P. (2009). Life cycle comparison of traditional retail and e-commerce logistics for electronic products: A case study of buy.com. 2009 IEEE International Symposium on Sustainable Systems and Technology, ISSST '09 in Cooperation with 2009 IEEE International Symposium on Technology and Society, ISTAS. <https://doi.org/10.1109/ISSST.2009.5156681>

