Finance Research Letters xxx (xxxx) xxx

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

The present and future of sustainability disclosure in equity investment funds' pre-contractual documents: Mapping ESG discourse through STM

Caterina Cruciani [a,*], Carlo R.M.A. Santagiustina [b]

[a] Department of Management -Venice School of Management, Ca' Foscari University of Venice, Venice, Italy
[b] Venice Internation University, Venice, Italy

ARTICLE INFO

JEL codes:
G2
G4
C89
Q56

Keywords:
KIID
KID
ESG
Sustainability
SFDR, Structural topic modelling

ABSTRACT

To show how pre-contractual documents are currently employed to disclose sustainability and ESG-related information, we map and examine the contents of 945 Key Investor Information Documents (KIIDs) from ten major asset managers, including sustainable and non-sustainable funds. Through a Structural Topic Modelling approach, we infer sustainability-related topics and compare their contents and prevalence in different sections of these documents. Sustainability is a cross-cutting and multifaceted issue, discussed coherently with the SFDR fund classification across multiple sections of the KIID, making the option to convey sustainable information in a single section complex to implement.

1. Introduction

This paper focuses on the effectiveness of short-form financial information disclosure in communicating sustainability using the Key Investor Information Document (KIID) introduced in 2009 with the European Directive 2009/65/EC regarding Undertakings for Collective Investments in Transferable Securities (UCITS) (European Parliament and Council, 2009, page 7). KIIDs provide the blueprint for a financial documentation focused on increasing investor awareness using clear, fair and non-misleading information in a structured format to ensure easy comparability across financial products and product providers (European Parliament and Council, 2010).

Financial literature has shown that the format of financial-information disclosure affects individual decision making (see for instance, (Bateman et al., 2016)). Readability of financial documentation has been shown to increase willingness to invest in a specific financial product (Arora and Chakraborty, 2021) and to improve information processing and the level of reliance on the disclosed information (Rennekamp, 2012), while document length has an impact on investors' cognitive overload (Oehler and Wendt, 2017).

A recent trend in finance regards the introduction of sustainable investment products, often identified through sustainable labels. (Hartzmark and Sussman, 2019) look at the role of a popular industry-based sustainability label – the Morningstar sustainability rating – and show that investors look at it as a positive fund feature, increasing investment flows towards funds with the highest rating and decreasing flow into funds with the lowest ones. Introduced in 2019 in the European Union, the Sustainable Finance Disclosure

* Corresponding author.
E-mail address: cruciani@unive.it (C. Cruciani).

https://doi.org/10.1016/j.frl.2023.104033
Received 10 February 2023; Received in revised form 8 May 2023; Accepted 26 May 2023
Available online 27 May 2023

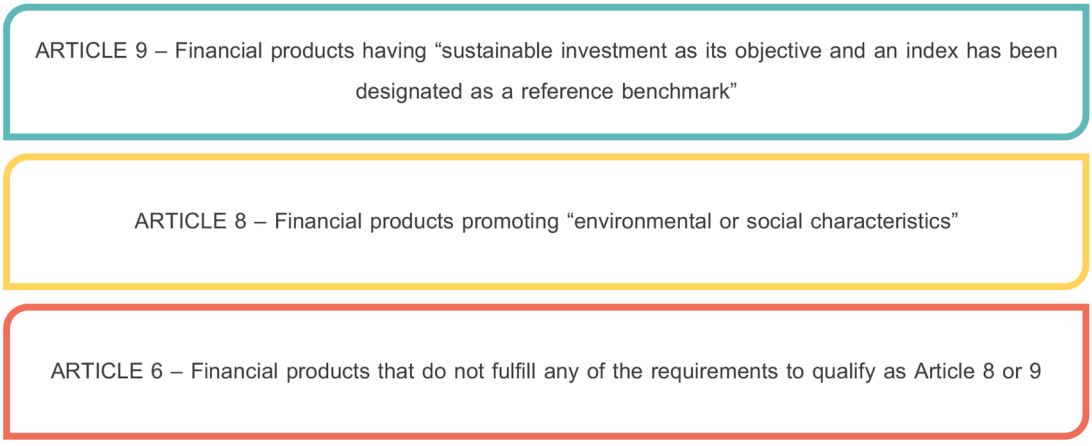

Fig. 1. SFDR classification by article and definitions.

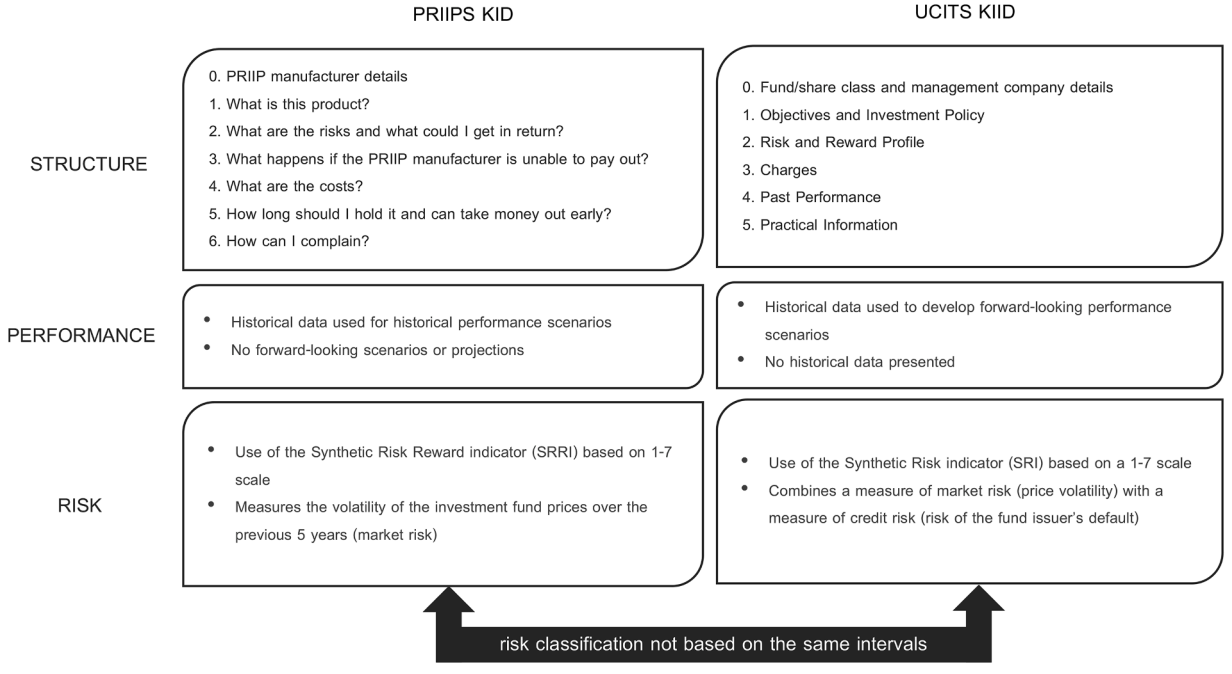

Fig. 2. UCITS KIID and the PRIIPS KID templates and section contents.

Directive (SFDR) has provided another regulatory investment-fund classification coherent with the European Taxonomy of sustainable activities (The European Parliament and the Council, 2019) summarized in Fig. 1.

Empirical research has shown that the introduction of the SFDR labels has a positive impact on the demand of Article 8 and 9 products (Becker et al., 2022), although increased investments flow more in response to industry ratings than in response to the regulatory SFDR classification according to (Ferriani, 2022).

Preliminary empirical research has shown that replacing prospectuses with shorter and more structured key information documents allows to effectively address some issues investors face when choosing financial investments (Oehler et al., 2014; Walther, 2015)

In 2014 EU Regulation 1286 extended the use of fixed-length pre-contractual documentation to all Packaged Retail and Insurance-based Investment Products (PRIIPs) (European Commission, 2014) introducing also for investment funds new Key Investor Document (KID), further defined in Annex I of Directive 653/2017 (European Commission, 2017).

Fig. 2 highlights the differences between the templates of the UCITS KIID and the PRIIPS KID.

The PRIIPS Directive exempted investment funds from providing the new KID even to retail clients, keeping the UCITS KIID in use. This exemption expired on January 1, 2023 (European Commission, 2021): now all PRIIPS (including investment funds) have to use KIDs to communicate mandatory pre-contractual information to potential investors.

Table 1
Collected KIIDs counts by asset manager and by SFDR article.

Asset manager	Article 6 funds	Article 8 funds	Article 9 funds
Amundi	102	111	26
Blackrock	88	57	48
BNP Paribas	2	31	11
Eurizon	5	6	0
Fidelity	18	71	0
JP Morgan	37	141	11
Nordea	32	56	9
Schroders	3	8	2
Swedbank	0	14	1
UBS	28	24	3
TOTAL BY ARTICLE	315	519	111

Table 2
Overview of KIIDs sample by risk category.

Risk category	Article 6	Article 8	Article 9
4	2,5%	1,2%	0,0%
5	6,3%	8,7%	3,6%
6	83,8%	88,8%	96,4%
7	7,3%	1,4%	0,0%
Tot.	100%	100%	100%

The debate on the effectiveness of sustainability disclosure within KID has already been raised by the European Supervisory Authorities that explicitly called for a revision of the current KID structure to accommodate sustainability concerns (European Supervisory Authorities, 2022). The ESAs proposed to introduce a separate section within the KID, titled "Does this product have a sustainable investment objective?" to highlight the investment contribution towards sustainable objectives.[1]

Given that KIDs are going to be the new documentation standard for all investment funds that used to fall under the UCITS Directive, an analysis of how sustainability is being currently addressed and discussed within KIIDs is relevant and timely for informing the debate and guiding future legislation in relation to possible integrations of SFDR and sustainability-related information into the KIDs.

We explore the distribution of sustainability and ESG-related contents in a corpus containing 945 documents from equity investment funds. We select the ten more relevant asset managers in terms of volumes of sustainable investment in 2021, as identified by Morningstar (Morningstar, 2022). Through a probabilistic topic model for textual data (Roberts et al., 2013), called Structural Topic Model (STM), we infer and map sustainability-related topics and their prevalence in documents, and show how and to which degree these issues are currently being addressed in the KIIDs. Differently from other topic modeling approaches, like LDA (Blei et al., 2003), STM allows to model topic contents and prevalence as a function of a set of covariates of interest. We rely on the SFDR classification, using it as a covariate in the STM, to test if KIIDs exhibit significant differences between Article-8 and Article-9 funds, regarding their approach to sustainability, in a way that is compatible with the SFDR. We hypothesize that Article-8 funds focus relatively more on investment strategies but relatively less on sustainability goals (with respect to Article-9 funds). Another hypothesis we consider, based on the recent ESAs' call for advice (European Supervisory Authorities, 2022), regards the feasibility of conveying all information on sustainability in a single section looking at where sustainability is currently discussed within KIIDs.

2. Methodology

A dataset of key investor documents has been created by identifying and downloading the KIIDs in English of all investment funds classified as Article-8 and Article-9 from the websites of the ten Asset Managers identified by Morningstar as the most relevant in terms of volumes (considering Art.8 and 9 fund assets combined) in 2021. Starting for the Asset Managers websites, we focus solely on equity funds[2] in the Euro currency, based in Europe (mostly in Ireland and Luxembourg), including both active and passive (ETF) funds.[3] Whenever an investment fund had more share classes, only one share class was included.[4] We only used retail-client sections of the Asset Manager websites available to the general public. A control group of non-sustainable funds (Article-6) was created choosing all

[1] The European regulation currently considers sustainability preferences as an add-on to the risk profile of investors. The guidelines on the introduction of sustainability concerns within the suitability questionnaire – the mandatory assessment all investment clients must fill out before purchasing financial products or receiving financial advice – suggest that questions about sustainability preferences should be introduced after the assessment on risk preferences and financial goals (Cruciani et al., 2022).

[2] Equity funds represent the majority of sustainable funds according to the same Morningstar source

[3] For a detailed description of the dataset and data-collection strategy please refer to Section S3 of the Supplementary Information Document

[4] Different share classes often imply different remunerations for the clients, but the description of the fund is always identical.

Topic proportion

$$\mu_{d,k} = X_d\gamma_k$$
$$\gamma_k \sim N(0, \sigma_k^2)$$
$$\sigma_k^2 \sim Gamma(s^\gamma, r^\gamma)$$

Language model

$$\theta_d \sim LogisticN(\mu_d, \Sigma)$$
$$z_{d,n} \sim Mult(\theta_d)$$
$$v_{d,n} \sim Mult\left(\beta_d^{k=z_{d,n}}\right)$$

Topic content

$$\beta_{d,v}^k \propto \exp\left(m_v + \kappa_v^k + \kappa_v^{y,} + \kappa_v^{y,k}\right)$$
$$\kappa_v^{y,k} \sim Laplace\left(0, \tau_v^{y,k}\right)$$
$$\tau_v^{y,k} \sim Gamma(s^\kappa, r^\kappa)$$

Fig. 3. STM model specification.

the available equity funds in Euro currency by the same Asset Managers for which a KIID in English was available, for a total of 945 KIIDs (see Table 1).

Table 2 summarizes the risk categories of the investment funds by SFDR article-type.

Since we are interested in understanding how the sustainability discourse is presented to potential investors within the KIIDs, the reason to focus on the ten more relevant Asset Managers in terms of Asset funds (covering almost 30% of the EU SFDR fund universe in terms of fund assets) ensures we include popular and well-distributed funds.

Since we are also interested in evaluating the degree to which sustainability and ESG related information is distributed within the different sections of the KIIDs, we define and use a set of regular expressions for automatically parsing the KIIDs for each of the selected asset managers, allowing to segment each document in the five UCITS KIIDs sections (excluding Section 0), described in Fig. 2.

We obtain a collection of 4725 KIID sections, each with its own metadata entries: (i) the section identifier of the document (called TYPE); (ii) the SFDR article number (called ARTICLE), (iii) the asset manager label (called ASSET MANAGER), (iv) the fund risk category (called RISK), and (v) the geographic area (called GEO AREA).

To infer the topics discussed in the corpus, we employ a Latent Graphical Modelling technique from the STM (Roberts et al., 2019) library for R. A STM with K topics is defined in Fig. 3 (refer to Roberts et al., 2013, for details).

Where topics are indexed by k, documents by d, and tokens by n. $X_d$is a $1-$by-5 vector, γ_k is a $5-$by$-K$ matrix of coefficients affecting topic proportions, and Σ is a $K-$by$-K$ topic proportion covariance matrix. The distribution over tokens is the combination of three effects: a topic effect $\kappa_v^{'k}$, a content covariate effect $\kappa_v^{y'}$, and a topic-covariate interaction effect $\kappa_v^{y,k}$. These effects are modelled as sparse deviations from a baseline frequency m_v.

To go beyond a bag-of-words approach and to appraise the syntagmatic structure of the inferred topics, following the method proposed by Santagiustina et al. (2022), we consider word associations (ordered pairs of adjacent words also called bigrams) as features in the token-by-document counts matrix. In total, after removing stopwords, numbers, tokens appearing less than twice in the whole corpus, and tokens that have less than three characters, our token-by-document counts matrix, contains 8245 columns, each representing the counts (by section for a specific KIID) of a token appearing in the corpus. 2685 unique tokens are words, and 5560 unique tokens are ordered pairs of words.

Since we want to understand if topic contents vary in relation to the SFDR classification, when estimating the STM, we allow the ARTICLE covariate to affect tokens-by-topic distributions through $\kappa_v^{y,}$ and $\kappa_v^{y,k}$. All metadata variables (ARTICLE, ASSET_MANAGER, TYPE, RISK, GEO_AREA) can affect topic proportions through $X_d\gamma_k$. This allows us to control for the effect of these covariates on the distribution of topics within KIIDs (see Table S4 in the Supplementary Information).

For choosing the number of topics, for each $K \in$ (5, 6, …, 19, 20, 25, …, 45, 50, 60, …, 90, 100) we estimate 10 times the model, holding out a random sample (25%) of the corpus, and then compute for each K the averages of four information criteria[5] suggested by Roberts et al. (2019). Based on these criteria we choose $K = 35$ and estimate again the STM with the whole dataset. The other parameters feeded to the STM algorithm are the following:

prevalence = ~ARTICLE+ASSET_MANAGER+TYPE+RISK+GEO_AREA; content = ~ARTICLE; max.em.its = 250; em.tol = 1e-06; init.type = "Spectral".

Further details about the methodology and the dataset are available in the online supplementary material at https://github.com/carlosantagiustina/mapping_ESG_discourse_through_STM/blob/main/Supplementary%20information.pdf

3. Results

For each topic, the STM gives as output a set of three token-by-topic distributions, one for each value of *ARTICLE*. These distributions represent the probabilities of tokens for a given topic, conditional on the value of the aforementioned covariate. To measure through a unique metric, called ESG score,[6] the degree to which a topic is related to sustainability issues, we choose a set of generic keywords related to sustainability (that is: "ESG", "sustainable", "sustainability", "social","environment", "environmental", "climate", "carbon") and, for each topic and *ARTICLE* covariate value, we compute the sum of their probabilities and of the probabilities of all pairs of words in the token vocabulary that contain them. Fig. 4 shows the ESG score by topic and by *ARTICLE* (on the left) and the most

[5] The four criteria are: exclusivity, semantic coherence, held-out likelihood, likelihood lower-bound

[6] The sustainability metric we develop is intended to be topic-specific (and not document-specific or fund-specific), and allows us to identify the location and proportion (within the latent space of topic discussed in the KIIDs) of sustainability-related discourse in different topics. We are by no means exploring whether the sustainability discourse within KIIDs matches or can proxy the sustainable rating of the funds by rating providers.

Fig. 4. On the left: ESG scores (in percentage) by topic and by article. On the right: sustainability-related keyword probabilities by topic and by article. Keywords and associations containing keywords are shown in the right panel only if their probability is higher than 0.01 for that topic and article class.

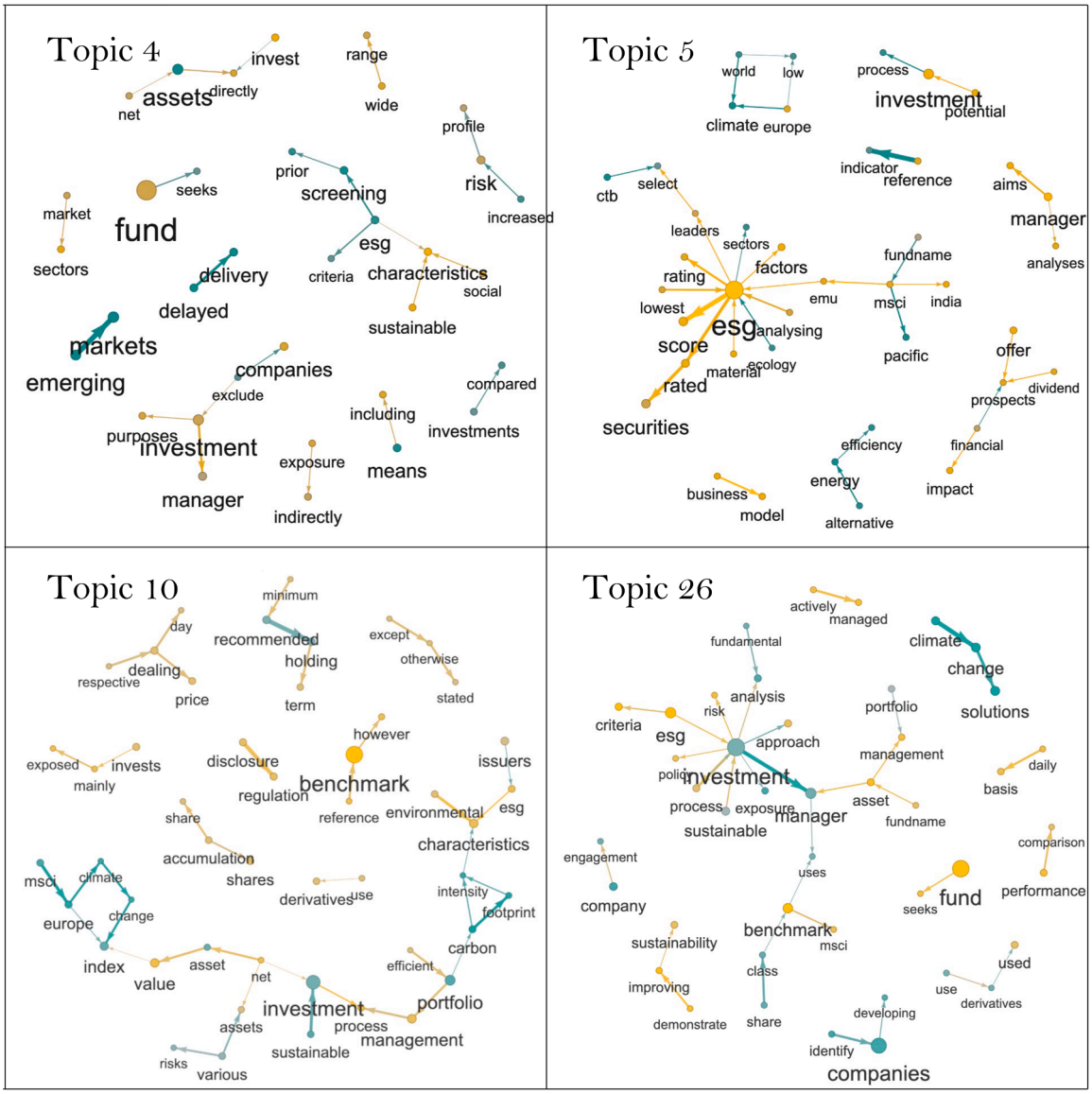

Fig. 5. Topic networks for four sustainability-related topics (top left: Topic 4; top right: Topic 5; bottom left: Topic 10; bottom right: Topic 26) for Article-8 Vs Article 9. Nodes and edges were filtered using a 97.5% percentile filter on topic-specific word probabilities (represented through the node size) as well as on associations probabilities (represented through the edge width) for Article 8 and Article 9 funds. Equal weights were used for Art.8 (0.5) and for Art.9 (0.5). Node and edge colors represent topic-specific differences in token probabilities between Art.9 and Art.8. The closer is the color to turquoise the more the token is relatively more frequent for Art.9 with respect to Art.8, the closer is the color to yellow the more the token is relatively more frequent for Art.8 with respect to Art.9. In gray tokens for which the differences are close to zero (0).

frequent tokens that characterize these topics from a sustainability perspective (on the right), keeping only those that have a probability higher than 0.01. The colored dots identify the KIID typology (Art. 6 in red, Art. 8 in yellow and Art. 9 in turquoise) and the size of the dot represents the probability with which that topic is present in the given investment-fund category. Keywords related to sustainability are in general more frequent in Article-8 and Article-9 KIIDs, but the sustainable themes often feature words and associations consistent with the way in which Article-6 describe their investment strategy and features.

Looking more closely at the themes with the highest ESG scores in Fig. 4, we can also remark that sustainability is cross cutting and enters different topics, like Topic 4, Topic 5, Topic 10 and Topic 26.

Fig. 5 compares, for four topics among those with the highest ESG scores (one in each box), the topic networks that can be reconstructed from the *ARTICLE*-specific tokens-by-topic distributions for Article-9 Vs Article-8 KIIDs (refer to Santagiustina et al., 2022, for details). In these networks nodes represent words and edges represent associations (i.e., pairs or words). The node size represents the probability of a word, and the edge width represents the probability of an association. In Fig. 5, a weighted average of Article-9 and Article-8 token probabilities is used as the reference distribution for each topic. To highlight the most important features,

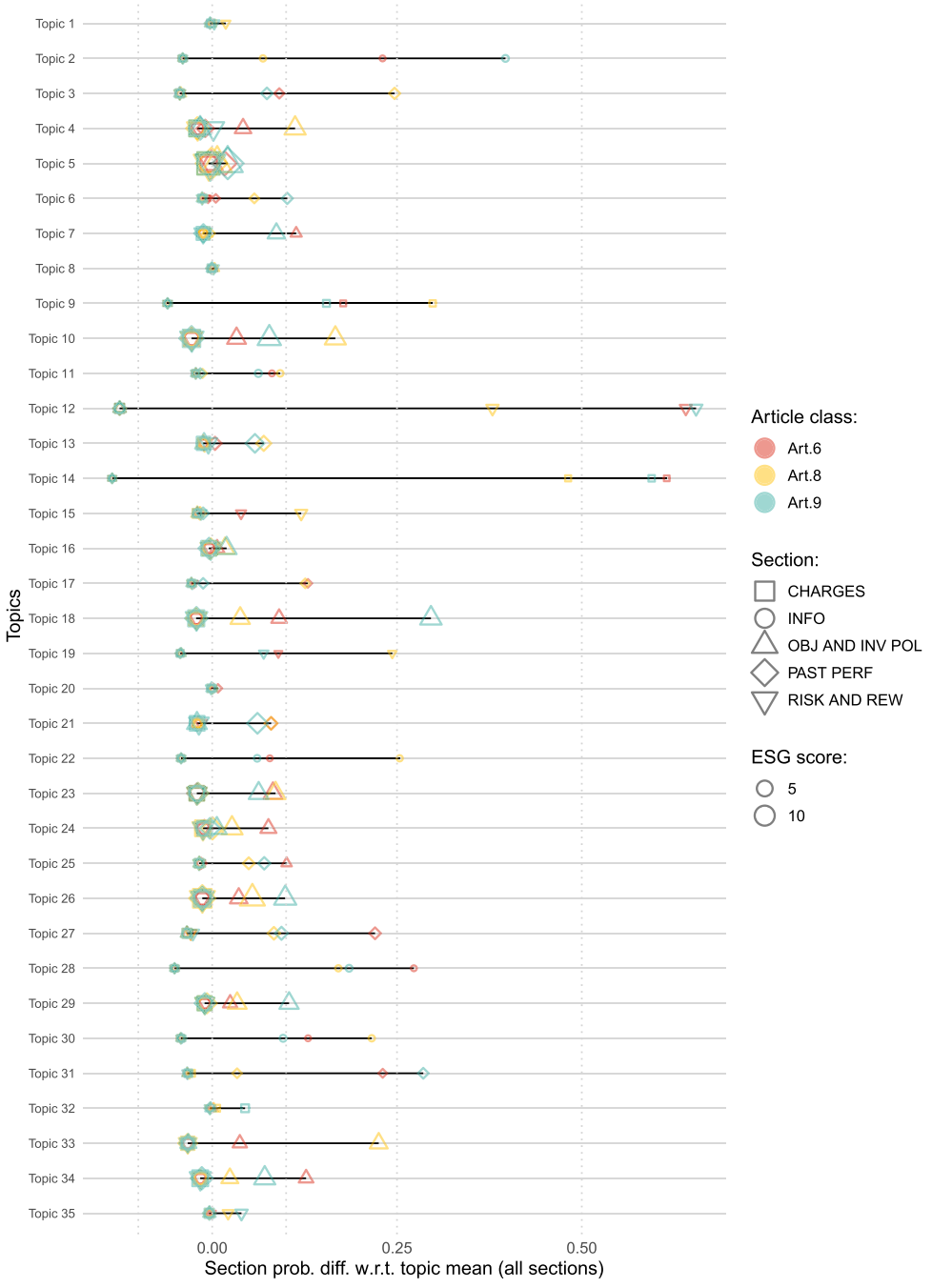

Fig. 6. Section-specific topic probabilities by article and by section. The markers' position on the X-axis represents the difference between the probability of a topic for a specific section and Article class, with respect to the average probability of that same topic in the whole dataset, considering all KIIDs sections and Article classes. The color of the markers is used to represent the Article class: in red Article 6 KIIDs, in yellow Article 8 KIIDs, and in turquoise Article 9 KIIDs; while the marker shape is used to represent the KIID section. Finally, the marker size is proportional to the ESG score of a topic for a given Article class.

each network is filtered using a 97.5% percentile filter on the probabilities of the nodes and on that of the edges. Using a color scale from turquoise (Article-9) to yellow (Article-8) passing through gray (at zero), we show the differences in token-by-topic probabilities between Article-9 and Article-8 KIIDs. Fig. 5 highlights the key implications of the SFDR within KIIDs. It can be noted that Article-9 KIIDs appear to discuss relatively more in the selected topics about "climate change", "pollution control", "carbon footprint" and other issues that are tied to the EU sustainable taxonomy and goals. Article-8 funds signal their commitment to sustainability in very generic terms, like using the term "ESG" relatively more often than in Article-9. They also mention more often "benchmarks" and "regulation"

alongside many terms that are related to the features of the investment such as “price” or “shares”. Lastly, while both mention “sustainable development”, the connection is stronger in Article-9 funds.

One last contribution of this paper is to look the sections of the current KIID more involved in the transmission of sustainability-related content to prospective investors. This assessment might be useful is shaping the transition from KIIDs to KIDs in compliance with PRIIPS regulation affecting investment funds. Fig. 6 indicates in which sections the different topics identified in the STM are discussed. Each section is represented through a different marker shape (e.g., a square marker for the "Charges” section), the colors used are the same as in previous figures, while relevance to sustainability is given by the marker size.

The ESG-related topics identified earlier, in particular Topic 5, exhibit similar topic proportions across the different sections of the KIIDs, even though the probability of finding these sustainability-related topics in all section is relatively small. For example, many markers overlap for Topic 5 indicating that this topic is discussed with similar proportions in the sections “Objectives and investment policy”, “Risk and reward” and “Past performance”. The most interesting comparisons are not within but between topics: if we focus on the larger markers, which indicate a higher ESG score for a topic, we can observe that KIIDs discuss sustainability in different ways across many sections, confirming that sustainability is a cross-cutting and multifaceted issue that is related to many of the topics discussed in funds’ pre-contractual documents.

4. Discussion

We believe that this analysis may contribute to the current and upcoming debate regarding pre-contractual information for prospective investors. Given the need to tailor the information previously included into KIIDs to the new document required by the PRIIPS documentation, it is important to better understand the strength and weaknesses of the current sustainability-related discourse within KIIDs. We have shown that while there is a significant difference between Article-8 and Article-9 funds in terms of how sustainable finance discourse is presented to investors, this is consistent with the SFDR regulation. Moreover, sustainability themes are overarchingly present and widely spread across different sections, making the ESA’s suggestion to include a new sustainability section not necessarily easy to implement. Sustainability is instrumental in discussing many of the features of the investment funds, including past performance, a domain that the KIDs do not feature at present. The topic-specific ESG metric produced in this article is not targeted at assessing the true sustainable nature of the fund as other available ESG scores do. We think that a future avenue of development regarding KIDs might explore whether the sustainability discourse within KIDs matches or can proxy the “true” sustainable nature of investment funds.

Author statement

All the authors contributed equally to the paper.

CRediT authorship contribution statement

Caterina Cruciani: Conceptualization, Methodology, Data curation, Writing – review & editing. **Carlo R.M.A. Santagiustina:** Conceptualization, Methodology, Data curation, Writing – review & editing.

Data availability

List of KIIDs of Art.8, 9 and 6 investment funds(on GitHub)

Acknowledgements

This paper was developed within the funded project NextGenerationEU – GRINS Growing Resilient INclusive and Sustainable (PE00000018) NRRP Mission 4, C2 Intervention 1.3. Carlo Santagiustina acknowledges financial support from PON R&I 2014–2020 (FSE REACT-EU). The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. We are grateful to the Editor, Samuel Vigne, and two anonymous reviewers for many insightful comments that helped improve the quality of this manuscript. We thank Pierangela Peruzzo and Francesco Michieli for excellent research assistance.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.frl.2023.104033.

References

Arora, J., Chakraborty, M., 2021. Does the ease of reading of financial disclosures influence investment decision? Econ. Lett. 204, 109883 https://doi.org/10.1016/j.econlet.2021.109883.

Bateman, H., Eckert, C., Geweke, J., Louviere, J., Satchell, S., Thorp, S., 2016. Risk presentation and portfolio choice. Rev. Financ. 20 (1), 201–229. https://doi.org/10.1093/rof/rfv001.

Becker, M.G., Martin, F., Walter, A., 2022. The power of ESG transparency: the effect of the new SFDR sustainability labels on mutual funds and individual investors. Finance Res. Lett. 47, 102708 https://doi.org/10.1016/j.frl.2022.102708. PB.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. J. Mach. Learn. Res. 3 (Jan), 993–1022.

Cruciani, C., Gardenal, G., Amitrano, G., 2022. Understanding financial risk tolerance. institutional, behavioral and normative dimensions. Palgrave Pivot Series. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-13131-8.

European Commission. (2014). Regulation (EU) No 1286/2014 of the European Parliament and of the council of 26 November 2014 on key information documents for packaged retail and insurance-based investment products (PRIIPs). 1–23.

European Commission. (2017). Commission delegated regulation (EU) 2017/653 of 8 March 2017 supplementing regulation (EU) No 1286/2014 of the European Parliament and of the council on key information documents for packaged retail and insurance-based investment products (PRIIPs).

European Commission. (2021). Directive (EU) 2021/2261 of the European Parliament and of the council of 15 December 2021 amending Directive 2009/65/EC as regards the use of key information documents by management companies of undertakings for collective investment in transferable securities. Official Journal of the European Union, 48–119.

European Parliament and Council. (2009). Directive 2009/65/EC of the European parliament and of the council. Official Journal of the European Union. https://doi.org/10.1002/9781119436775.ch25.

European Parliament and Council. (2010). Commission regulation (EU) No 583/2010. Official Journal of the European Communities.

European Parliament and Council. (2019). Regulation (EU) 2019/2088 of the European Parliament and of the council on sustainability-related disclosures in the financial services sector. Official Journal of the European Union, 2019(April), 1–16. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2088&from=EN.

European Supervisory Authorities. (2022). Call for advice on PRIIPs: ESA advice on the review of the PRIIPs regulation 2/150 contents (Issue April).

Ferriani, F., 2022. The importance of labels for sustainable investments: SFDR versus Morningstar globes. SSRN Electron. J. https://doi.org/10.2139/ssrn.4166932.

Hartzmark, S.M., Sussman, A.B., 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. J. Finance 74 (6), 2789–2837. https://doi.org/10.1111/jofi.12841.

Morningstar. (2022). SFDR article 8 and article 9 funds: 2021 in review. 1–35. http://global.morningstar.com/equitydisclosures.

Oehler, A., Höfer, A., Wendt, S., 2014. Do key investor information documents enhance retail investors' understanding of financial products? Empirical evidence. J. Financ. Regul. Compliance 22 (2), 115–127. https://doi.org/10.1108/JFRC-10-2013-0035.

Oehler, A., Wendt, S., 2017. Good consumer information: the information paradigm at its (Dead) end? J. Consum. Policy 40 (2), 179–191. https://doi.org/10.1007/s10603-016-9337-5.

Rennekamp, K., 2012. Processing fluency and investors' reactions to disclosure readability. J. Account. Res. 50 (5), 1319–1354. https://doi.org/10.1111/j.1475-679X.2012.00460.x.

Roberts, M., Stewart, B., Tingley, D., Airoldi, E., 2013. The Structural Topic Model and Applied Social Science. Neural Information Processing Society.

Roberts, M.E., Stewart, B.M., Tingley, D., 2019. STM: an R package for structural topic models. J. Stat. Softw. 91, 1–40.

Santagiustina, C.R.M.A., Warglien, M., 2022. The architecture of partisan debates: the online controversy on the no-deal Brexit. PLoS ONE 17 (6), e0270236. https://doi.org/10.1371/journal.pone.0270236.

Walther, T., 2015. Key investor documents and their consequences on investor behavior. J. Bus. Econ. 85 (2), 129–156. https://doi.org/10.1007/s11573-014-0724-6.