# Conservative and timely ESG–compliant investment screening using text mining

**Samuel Borms**[1], Kris Boudt[2], Frederiek Van Holle[3], Joeri Willems[4]

December 14, 2019

CFE 2019, London

[1]Université de Neuchâtel, Vrije Universiteit Brussel (samuel.borms@unine.ch)
[2]Universiteit Gent, Vrije Universiteit Brussel, Vrije Universiteit Amsterdam
[3]Degroof Petercam Asset Management
[4]Degroof Petercam Asset Management

# Overview

# Motivation

# Sustainable investing

ESG–compliant investors invest in companies and funds that comply with Environmental, Social and corporate Governance standards

**Selected ESG factors**

| Environment | Social | Governance |
|---|---|---|
| - Environment impact and risk management<br>- Environment performance<br>- Environment solution companies<br>- Climate-change impact and risk management<br>- Biodiversity impact and risk management<br>- Water scarcity and risk management<br>- Sector-specific issues, e.g. chemicals, timber, tar sands<br>- Allegations of environmental pollution or damage to biodiversity | - Human rights<br>- Supply-chain labour standards<br>- Relations with customers and suppliers<br>- Relations with employees<br>- Stakeholder engagement<br>- Community involvement<br>- Sector-specific issues, e.g. access to medicines<br>- Allegations of breaches of human rights norms and labour standards | - Board practice and structure<br>- Anti-bribery practices<br>- Codes of ethics<br>- ESG risk management<br>- Board-level responsibility for stakeholders<br>- Board-level gender diversity<br>- Allegations of bribery |

Source: EIRIS global sustainability ratings, as at 17 May 2017.

# Sustainable investing

Most investors obtain an investable universe of ESG–compliant assets using a best-of-class approach

They rely on third party agency reports and ESG ratings, **but** scores are...

- ... unstandardized and not transparent (Escrig–Olmedo & al. 2010)
- ... often reporting–driven, not signal–driven
- ... not updated frequently (up to monthly only)

# Text–based sustainability scoring

Why not add some other data points?

Text–based values will be (i) more frequently available and (ii) potentially carry complementary information value.

In many areas, textual analysis has proven valuable.

Recently in finance, Engle & al. (2019) use text–based indicators to form portfolios hedged against climate change news.

# Methodology

# Problem statement

We try to extract meaningful indicators from a large collection of texts.

Meaningful means:

- Represents one or more of the three key sustainability dimensions
- Represents a specific “entity” (here: company)
- Time–variation can be linked to real–time events and news stories
- Useful as an investment signal

This is a hard problem.

- No underlying objective function to optimize
- Data is not easy to explore
- Computationally intensive so cumbersome to iterate

# A step–wise approach

Step 1: Query relevant texts from a database

Step 2: Filter out the noise texts

Step 3: Compute text–based indicators

Step 4: Map the indicators into decision–making signals

Step 5: Validate the suite of indices

# Step 1: Query relevant texts from a database

Database = big corpus.

Querying is a function of entity name and relevant **keywords**.

Select and create subsets of corpora: one corpus for each company, and for each ESG dimension, where every subcorpus comprises at least one of the respective E, S or G keywords.

# Step 2: Filter out the noise texts

The texts are still very noisy...

*For example*: texts about a thief driving a BMW is not ESG–worthy news about the company BMW.

Run through some negative filters:

- Remove duplicates
- Remove near–duplicates
- Remove texts not in target language
- Remove irrelevant texts (e.g. about sports)
- Remove too short and too long texts

# Step 3: Compute text–based features...

A corpus for a given entity $e$ has a matrix $\boldsymbol{Z_e}$ associated to it:

$$\boldsymbol{Z_e} = \begin{bmatrix} n^e_{1,1} & s_{1,1} & n^E_{1,1} & n^S_{1,1} & n^G_{1,1} & \cdots \\ n^e_{2,1} & s_{2,1} & n^E_{2,1} & n^S_{2,1} & n^G_{2,1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n^e_{n,t} & s_{n,t} & n^E_{n,t} & n^S_{n,t} & n^G_{n,t} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n^e_{N-1,T} & s_{N-1,T} & n^E_{N-1,T} & n^S_{N-1,T} & n^G_{N-1,T} & \cdots \\ n^e_{N,T} & s_{N,T} & n^E_{N,T} & n^S_{N,T} & n^G_{N,T} & \cdots \end{bmatrix}.$$

The value $n^e_{n,t}$ is the count of entity mentions (in article $n$ at time $t$).

The value $s_{n,t}$ is a textual sentiment score.

The value $n^{\{E,S,G\}}_{n,t}$ is the count of dimension–specific keyword mentions.

# Step 3: ... and aggregate into text–based indicators

The indices are a function $g : \boldsymbol{Z_e} \mapsto \boldsymbol{I_e}$, where $\boldsymbol{I_e}$ is a time series matrix:

$$\boldsymbol{I_e} = \begin{bmatrix} n_1^e & s_1 & n_1^E & n_1^S & n_1^G & a_1^E & a_1^S & a_1^G & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n_t^e & s_t & n_t^E & n_t^S & n_t^G & a_t^E & a_t^S & a_t^G & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ n_T^e & s_T & n_T^E & n_T^S & n_T^G & a_T^E & a_T^S & a_T^G & \cdots \end{bmatrix}.$$

The aggregation can be (weighted) linear or nonlinear:

$$n_t^e = \sum_{i=1}^{N_t} \omega_i n_{i,t}^e, \quad a_t^{\{E,S,G\}} = \sum_{i=1}^{N_t} \mathbb{1}(n_{i,t}^{\{E,S,G\}} > 0), \quad \ldots$$

# Keywords

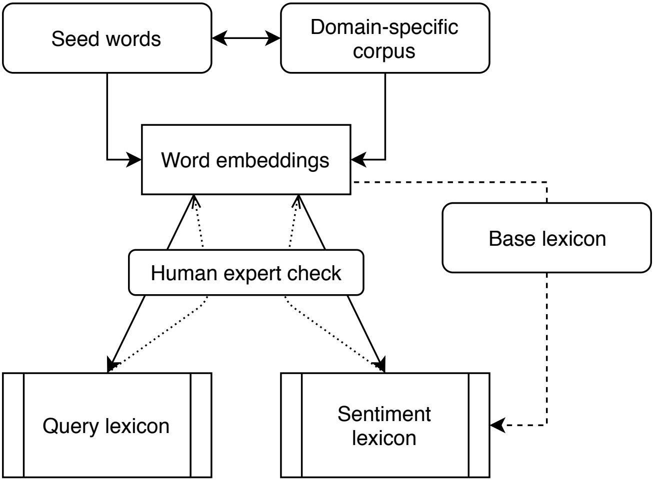
We need two sets of final keywords:

1. Query keywords (= query lexicon)
2. Sentiment keywords (= sentiment lexicon)

They have to be generated.

# How we generate keywords

Define a seed set of keywords, including a subset of sentiment words.



# Word embeddings

*"You shall know a word by the company it keeps."*

– J.R. Firth

Different approaches: LSA, Word2Vec, GloVe, fastText or BERT.

We use the **GloVe** method from Pennington & al. (2014). Essentially, it is a factorization method on the corpus word–word co–occurrence matrix.

# Word embeddings

For $i = 1, \ldots, V$ words in your corpus, the co–occurrence matrix $\boldsymbol{X}$ is:

$$\boldsymbol{X} = \begin{bmatrix} & \cdots & \\ \vdots & x_{ij} & \vdots \\ & \cdots & \end{bmatrix},$$

with $x_{ij}$ is the number of times word $j$ occurs in the context of word $i$, and $x_i = \sum_k x_{ik}$. The "context" means within $c$ words before/after word $i$.

Define $P_{ij} = P(j|i) = x_{ij}/x_i$ as the probability that word $j$ appears in the context of word $i$.

Take three words, $i$, $j$ and $k$. The ratio of probabilities $P_{ik}/P_{jk}$ hints at the association of word $k$ to word $i$ relative to word $j$.

If the ratio is high, word $k$ is more related to word $i$ than it is to word $j$. If the ratio is around 1, there is no discriminative power.

# Word embeddings

The probability ratios can be linked to the $d$–dimensional word vectors $w \in \mathbb{R}^d$ that want to be obtained:

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = P_{ik}/P_{jk},$$

which can be developed into:

$$w_i^T \tilde{w}_k = \log x_{ik} - \log x_i,$$

or:

$$w_i^T \tilde{w}_k + b_i + \tilde{b}_j - \log x_{ik} = 0.$$

Cast in a weighted least squares regression model with cost function:

$$\sum_{i=1}^{V} \sum_{j=1}^{V} \left(f(x_{ij})(w_i^T w_j + b_i + b_j - \log x_{ij})\right)^2$$

# From embeddings to lexicons

Every word is now represented as a numeric vector.

The similarity between a seed word and any word from the corpus can be measured with a distance metric (e.g. cosine similarity).

Take the words that are close enough to your seed words and add them.

For the sentiment lexicon: also add the words from a generic base lexicon which are not too distant from the seed words.

Upweight or downweight some words depending on the application.

# Step 4: Map the indicators into decision–making signals

Very application–specific – cf. infra.

# Step 5: Validate the suite of indices

Qualitative:

- News validation (degree of remaining noisiness)
- Graphical validation (appearance of events)

Quantitative:

- Timeliness (leading w.r.t. existing sustainability scores)
- Investment strategy (returns and interpretability of portfolios)

# Applications

# Two applications of interest

*Application #1*: Text–based ESG scoring as a **risk management** tool

- More timely screening of investment portfolios
- Positive and negative screening

*Application #2*: Text–based ESG scoring as an **investment** tool

- Additional value generation through stock picking

# Formulation of Application #1

Split a fixed universe of $N$ stocks at every time $t$ in two subuniverses (e.g. use cut-off based on an external data provider): a sustainable universe $S_t$ and a non–sustainable universe $\bar{S}_t = N \setminus S_t$.

Between a given time interval $[t_x, t_y]$, can the news–based indicators anticipate a change in the composition from $S_{t_x}$ to $S_{t_y}$?

Strategy:

1. Reset the scores to zero at every time point $t$
2. Cumulate daily news–based (sentiment) scores
3. If the cumulation exceeds a threshold, signal (a threshold level can be the average of the previous period news–based cumulation)

# Formulation of Application #1

Validation: false positives are acceptable but false negatives not.

| | | **Predicted state** | |
|---|---|---|---|
| | | $ESG$ | $\overline{ESG}$ |
| **Actual/future state** | $ESG$ | True Positive | False Negative |
| | $\overline{ESG}$ | False Positive | True Negative |

Note: $ESG$ means ESG compliant and $\overline{ESG}$ means not ESG compliant.

# Formulation of Application #2

Can we generate investment indicators from the news–based signals?

Follows the "hedging paradigm" from Engle et al. (2019).

Strategy:

1. Calculate single–factor beta's against a global news–based index
2. Sort the beta's, and apply investment strategy

A multi–factor approach is an obvious extension.

# Setup

Both applications are still in exploration, here focus on Application #2.

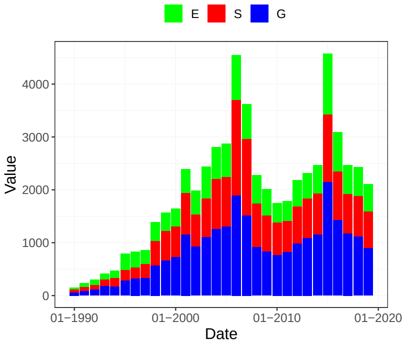
Data: corpus of millions of Flemish press articles. Includes both local and specialized financial/economic press. From 1990.

245 European equities.

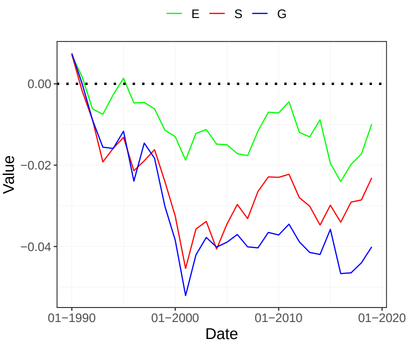
On average 1600–2000 starting articles per company, but very wide dispersion. Roughly 10%–30% filtered out.

We use the R package **sentometrics** (Ardia & al. 2019) to compute the textual indices in Step 3.

# Text indicators for **Volkswagen**



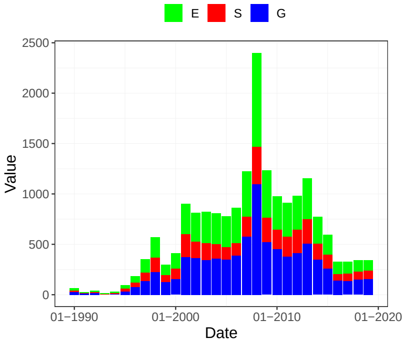

(a) Frequency of news articles



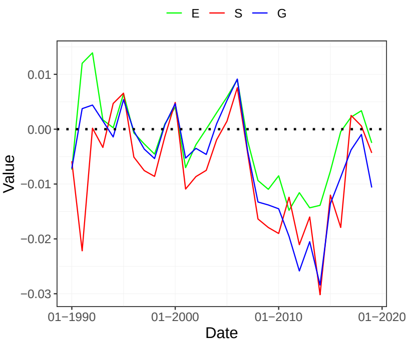

(b) Average sentiment

Events: big restructuring and emissions scandal.

# Text indicators for **UBS**



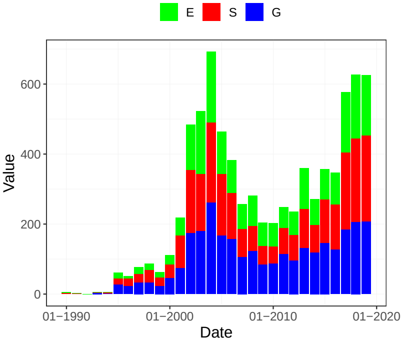

(a) Frequency of news articles



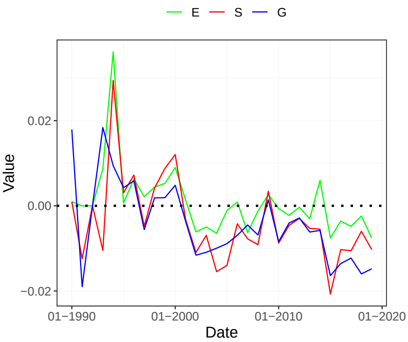

(b) Average sentiment

Events: financial crisis and forex scandal.

# Text indicators for **TUI**



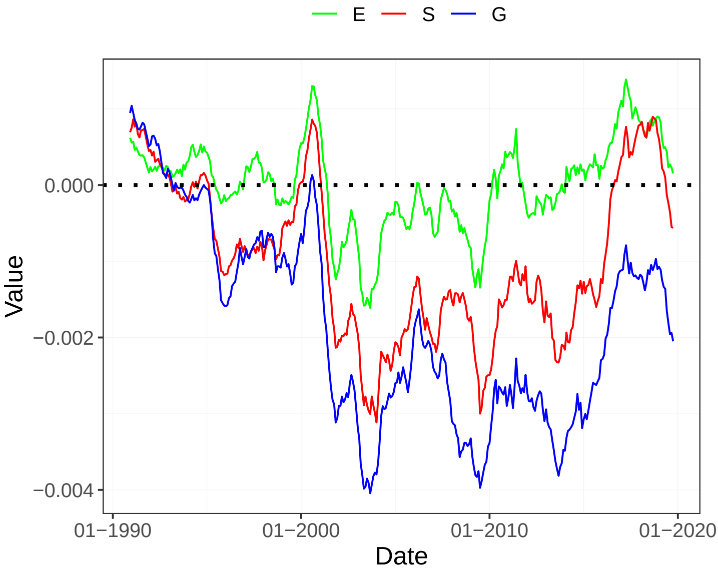

(a) Frequency of news articles




(b) Average sentiment

Events: sector–related Sobelair bankruptcy and multiple strikes.

# Average (monthly) company–wide sentiment



# Portfolio approach

80 different global indices (per E, S and G) to compute beta's against:

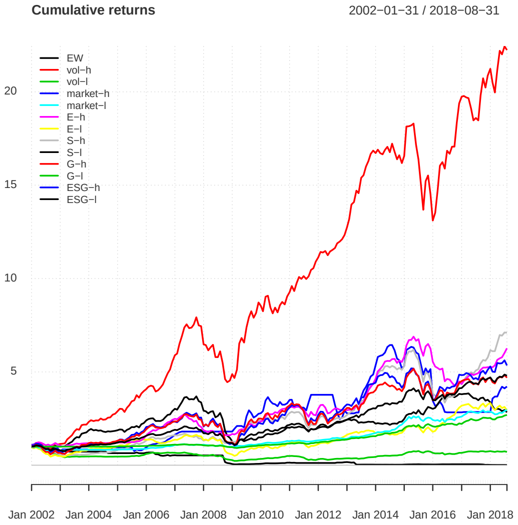
- Number of articles
- Number of keywords
- Sentiment weighted in eight ways (mean, sum; entity, keywords)
- The 10% minus 90% quantile version of the above
- The absolute value of the above
- The shocks in an $AR(1)$ model of the above (but short–term lag)

4 trading strategies:

- Invest in top–K/bottom–K
- Invest in companies in top–K/bottom–K for all E, S and G indices

Monthly rebalanced, equally weighted, 3–year rolling, $K = \{10, 30, 50\}$

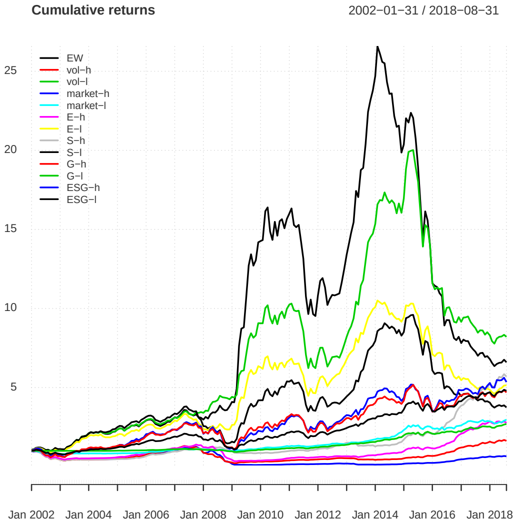
# Top–50, keywords frequency index (levels)



# Top–50, keywords frequency index (levels)

| | G-h | vol-l | market-l | E-h | EW | S-h | market-h | ESG-h | vol-h | S-l | E-l | G-l | ESG-l |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ann. return | 0.20 | 0.06 | 0.07 | 0.12 | 0.10 | 0.12 | 0.11 | 0.09 | 0.10 | 0.06 | 0.07 | -0.02 | -0.23 |
| Ann. std. dev. | 0.19 | 0.06 | 0.09 | 0.16 | 0.14 | 0.20 | 0.26 | 0.24 | 0.26 | 0.19 | 0.22 | 0.19 | 0.37 |
| Ann. Sharpe | 1.08 | 1.05 | 0.79 | 0.75 | 0.69 | 0.64 | 0.41 | 0.38 | 0.38 | 0.35 | 0.30 | -0.10 | -0.62 |

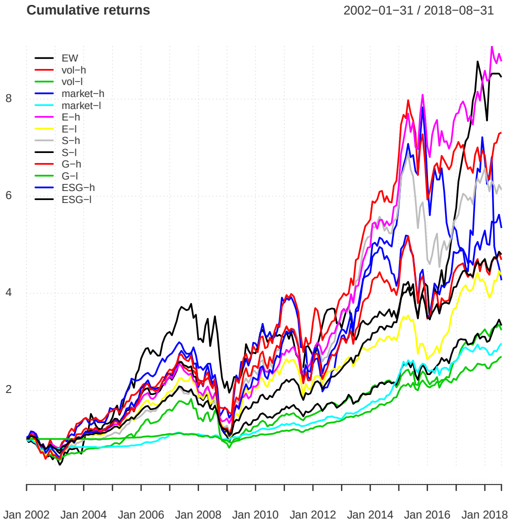
# Top–50, average sentiment index (levels)



# Top–50, average sentiment index (levels)

| | vol-l | market-l | S-h | EW | G-l | E-l | ESG-l | market-h | vol-h | S-l | E-h | G-h | ESG-h |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ann. return | 0.06 | 0.07 | 0.11 | 0.10 | 0.13 | 0.10 | 0.12 | 0.11 | 0.10 | 0.08 | 0.06 | 0.03 | -0.03 |
| Ann. std. dev. | 0.06 | 0.09 | 0.15 | 0.14 | 0.20 | 0.21 | 0.25 | 0.26 | 0.26 | 0.23 | 0.19 | 0.20 | 0.25 |
| Ann. Sharpe | 1.05 | 0.79 | 0.71 | 0.69 | 0.67 | 0.49 | 0.47 | 0.41 | 0.38 | 0.35 | 0.34 | 0.15 | -0.10 |

# Top–50, sentiment–keywords weighted index (shocks)



# Top–50, sentiment–keywords weighted index (shocks)

| | vol-l | E-h | market-l | EW | G-h | S-h | ESG-l | E-l | G-l | S-l | market-h | vol-h | ESG-h |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ann. return | 0.06 | 0.14 | 0.07 | 0.10 | 0.13 | 0.12 | 0.14 | 0.09 | 0.07 | 0.07 | 0.11 | 0.10 | 0.09 |
| Ann. std. dev. | 0.06 | 0.17 | 0.09 | 0.14 | 0.21 | 0.22 | 0.27 | 0.21 | 0.17 | 0.17 | 0.26 | 0.26 | 0.27 |
| Ann. Sharpe | 1.05 | 0.84 | 0.79 | 0.69 | 0.60 | 0.53 | 0.50 | 0.44 | 0.44 | 0.44 | 0.41 | 0.38 | 0.34 |

# Initial findings

Out–of–sample (2002–2018): higher returns, but lower Sharpe ratios.

Results depend heavily on the nature of the index (news frequency, sentiment, levels, shocks).

**The value to explore lies in meaningful index construction!**

# Thanks

Thanks for your attention! Any feedback?

# References


Ardia, Bluteau, Borms, & Boudt (2019). The R package **sentometrics** to compute, aggregate and predict with textual sentiment. The Journal of Statistical Software, revise and resubmit.

Engle, Giglio, Lee, Kelly, & Stroebel (2019). Hedging climate change news. Forthcoming in Review of Financial Studies.

Escrig–Olmedo, Muñoz–Torres & Fernandez-Izquierdo (2010). Socially responsible investing: sustainability indices, ESG rating and information provider agencies. International Journal of Sustainable Economy 2 (4), 442–461.

Pennington, Socher & Manning (2014). GloVe: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1532–1543.