



Online-Appendix

„The Impact of Prospectus Language on IPO Underpricing: A Textual Analysis of European IPOs“

Nicolas von Bodman

Technical University of Munich

Junior Management Science 9(4) (2024) 1934-1963

Appendix

Reduced Word List (fastText)	Sentence List (SBERT)
<p>uncertainty: ambiguity arbitrariness dependency deviate fluctuate randomize risk tend unconfirmed vagueness</p> <p>litigious: criminal demur depose lawsuit mediate nullify prehearing severable subleasehold subpoena</p> <p>negative: abdicates aberrant breach cancel concedes dishonest infringe misinform obstacle outage</p> <p>positive: able charitable conclusive enhance exceptional fantastic innovate outperform preeminent solving</p>	<p>uncertainty: Regulation might cause ambiguity Factors involve arbitrariness There may be a dependency The company might deviate from it Factors might fluctuate The process will be randomized The situation involves risk Results tend to change Results are unconfirmed The situation involves vagueness</p> <p>litigious: The company is criminally liable Demurring the allegation They are deposing witnesses Bring a lawsuit against the company Mediating the consequences Nullifying agreements or contracts A prehearing was held Severable parts create complexity A subleasehold exists The subpoena was issued</p> <p>negative: Abdication brings challenges Guidance is aberrant The breach has negative consequences It is required to cancel the activity A concession is required Dishonest actions have adverse effects Infringe on others' rights or property Misinforming stakeholders is problematic The company faces an obstacle The outage caused issues</p> <p>positive: The company was able to achieve targets For charitable purposes The outcomes are conclusive Enhanced performance improves outlook The performance is exceptional Fantastic results are achieved Innovated processes or products The company is outperforming expectations Preeminence is confirmed Solving the task successfully</p>

Table 1: Refined sentiment dictionaries

This table contains the sentiment dictionaries employed for the computation of fastText and SBERT scores. The lists are based on the LM word lists. The theoretical framework underlying the creation of these lists is detailed in Sections 3.2.1 and 3.2.2 for fastText and SBERT, respectively.

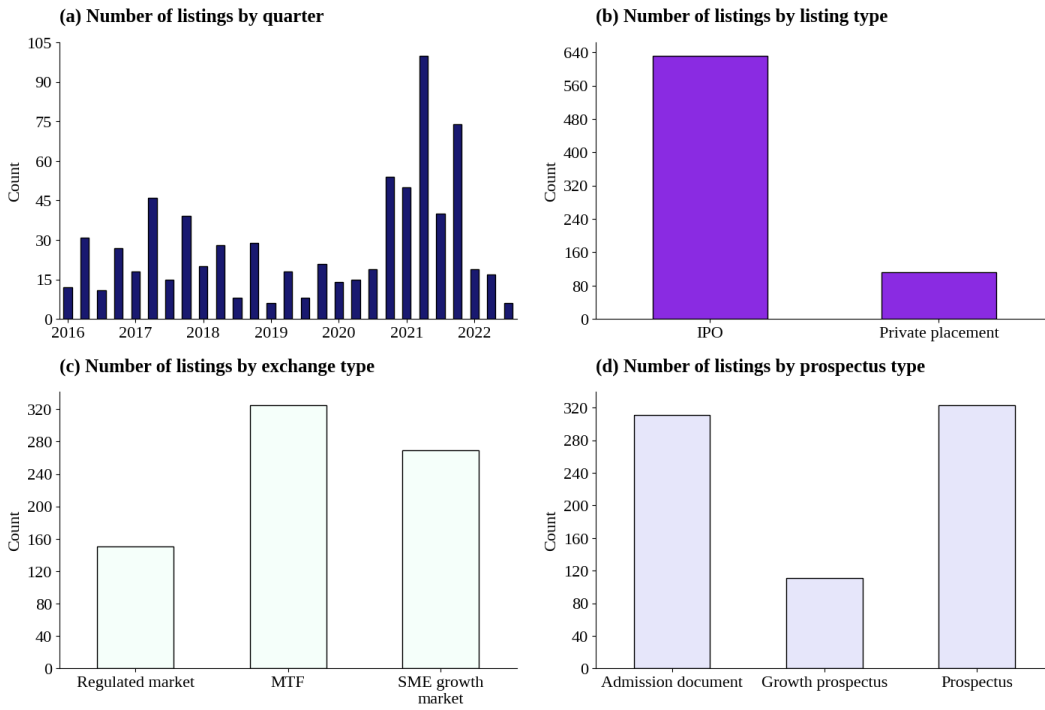


Figure 1: Number of listings across key categories

This figure includes four different bar charts representing the number of listings across different categorical variables. The subfigures include representations for county by year, listing type, exchange type and prospectus type.

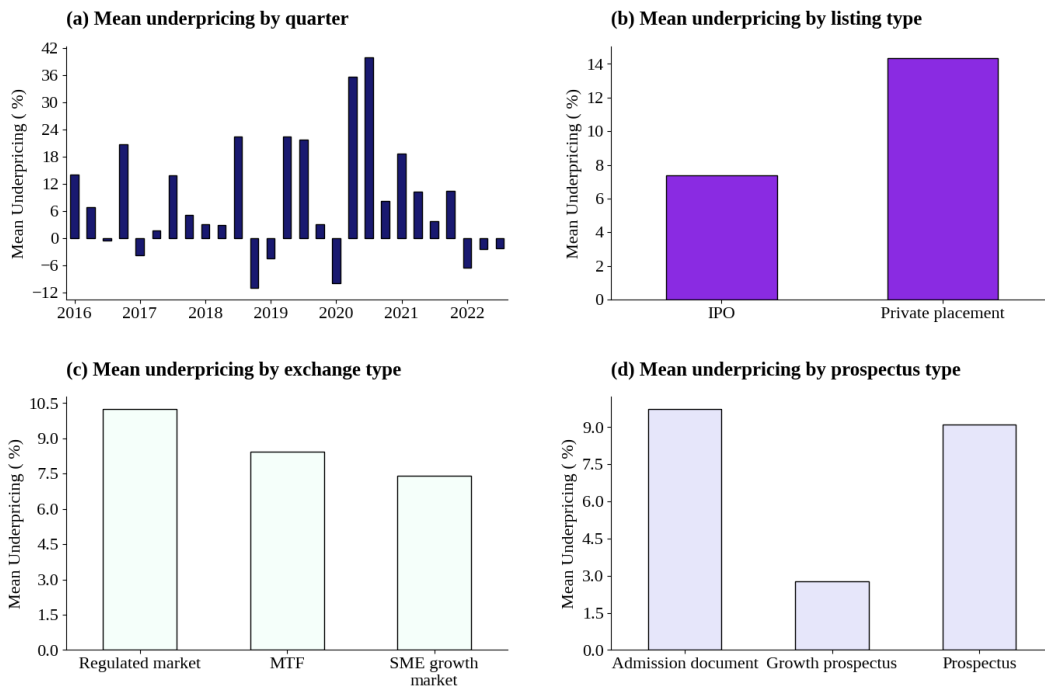


Figure 2: Mean underpricing across key categories

This figure includes four different bar charts representing the number of IPOs across different categorical variables. The subfigures include representations for IPO county by year, listing type, exchange type and prospectus type.

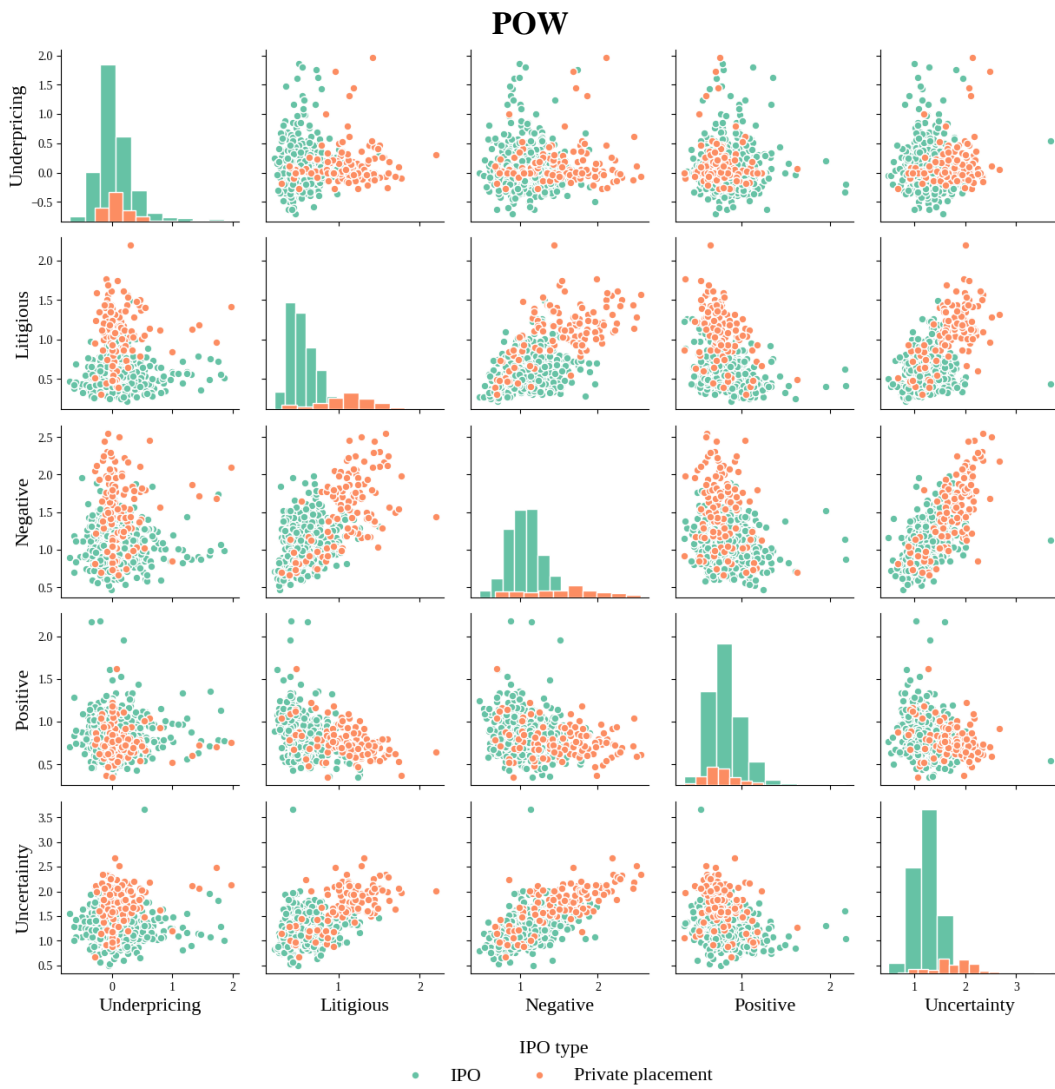


Figure 3: PairGrid analysis of underpricing and POW sentiment variables. The figure visualizes the relationships between underpricing and POW sentiment variables for sentiment categories: litigious, negative, positive, and uncertainty. Listing types: IPO and private placement are color-coded. The relationships between different variables are represented by scatter plots for each pairwise comparison of the variables. On the diagonal histograms are used to illustrate the distributions of each variable of interest (own representation).

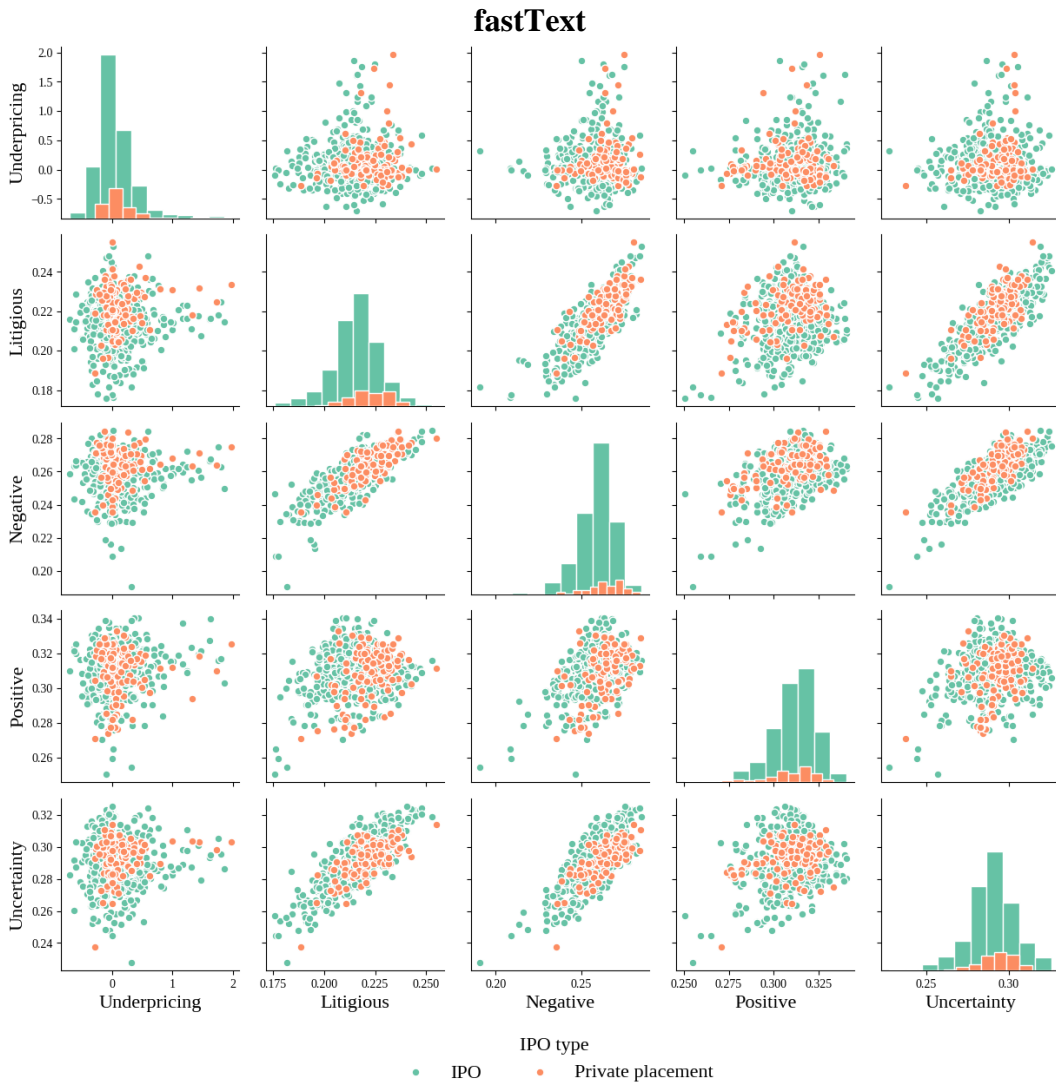


Figure 4: PairGrid analysis of underpricing and fastText sentiment variables. The figure visualizes the relationships between underpricing and POW sentiment variables for sentiment categories: litigious, negative, positive, and uncertainty. Listing types: IPO and private placement are color-coded. The relationships between different variables are represented by scatter plots for each pairwise comparison of the variables. On the diagonal histograms are used to illustrate the distributions of each variable of interest (own representation).

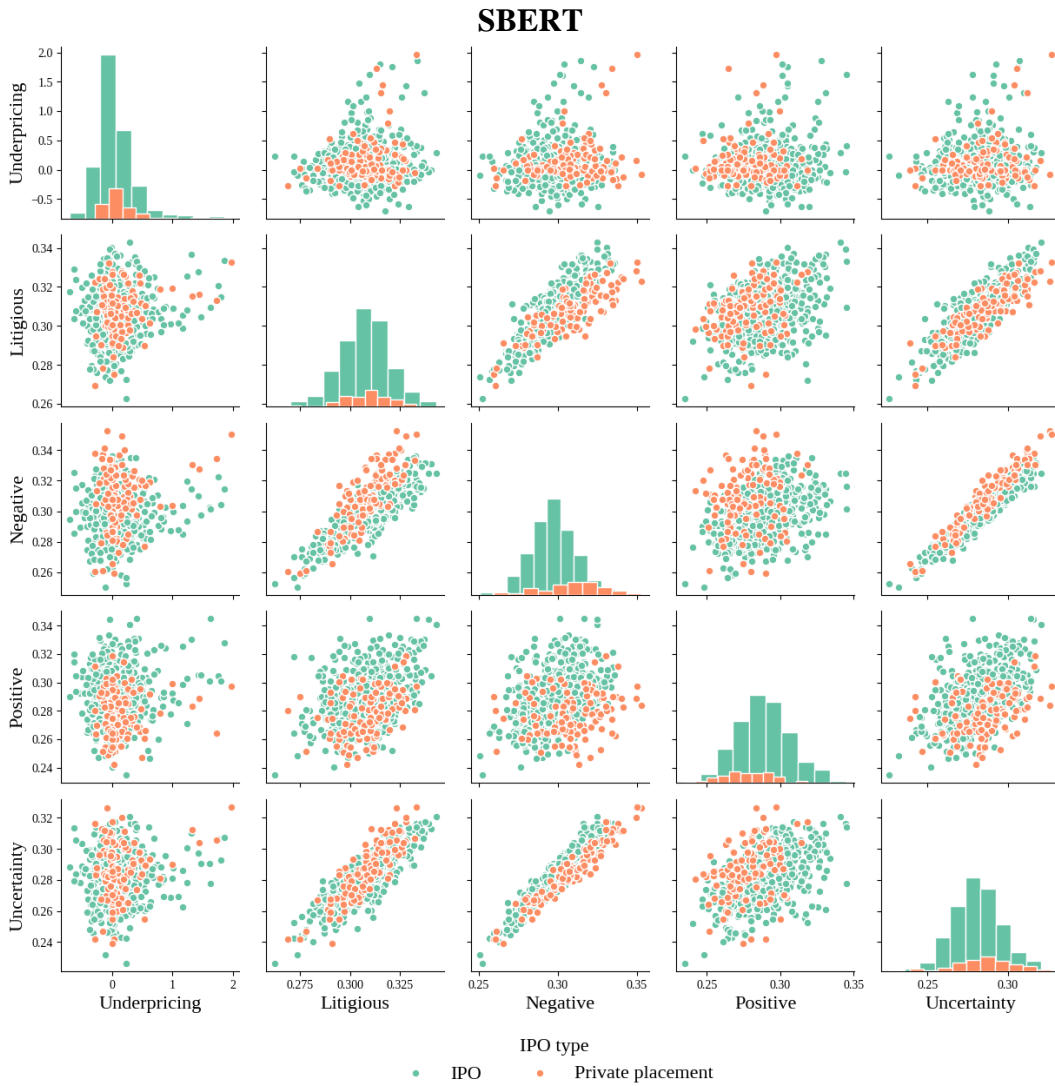


Figure 5: PairGrid analysis of underpricing and SBERT sentiment variables

The figure visualizes the relationships between underpricing and SBERT sentiment variables for sentiment categories: litigious, negative, positive, and uncertainty. Listing types: IPO and private placement are color-coded. The relationships between different variables are represented by scatter plots for each pairwise comparison of the variables. On the diagonal histograms are used to illustrate the distributions of each variable of interest.

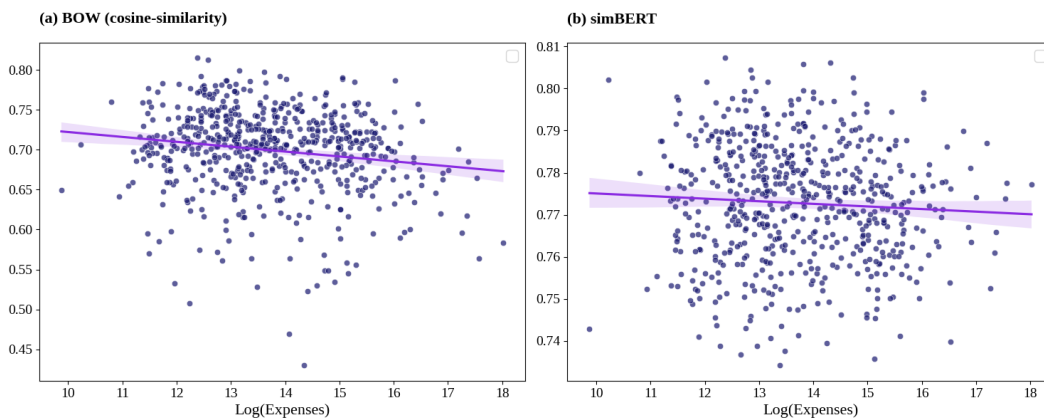


Figure 6: Scatterplots of similarity variables and listing expenses

This figure shows the relationships of log transformed listing expenses on the x-axis and textual similarity measures, BOW (cosine-similarity) in subfigure (a) and simBERT in subfigure (b) on the y-axis. The respective relationship is highlighted by a regression.

	N	Mean	SD	MIN	50%	MAX
Total						
Word count	745	38,610.489	27,046.784	7,156	28,393	197,549
Sentence count	745	1,493.125	914.249	336	1,212	6,106
Avg. sentence length	745	25.061	3.664	15.857	24.443	37.209
No. unique words	745	2,882.477	838.882	1,359	2,659	5,955
Full prospectus						
Word count	323	58,331.384	27,347.885	15,860	51,793	197,549
Sentence count	323	2,216.858	862.846	715	2,068	6,106
Avg. sentence length	323	25.669	3.514	16.774	24.981	37.209
No. unique words	323	3,446.585	753.827	1,972	3,355	5,955
Growth prospectus						
Word count	111	33,506.486	19,069.213	16,362	25,382	93,062
Sentence count	111	1,252.171	544.574	627	1,021	3,036
Avg. sentence length	111	25.697	3.767	20.141	24.410	36.824
No. unique words	111	2,849.135	769.273	1,871	2,563	4,835
Admission document						
Word count	311	19,950.344	9,268.420	7,156	17,769	86,699
Sentence count	311	827.466	341.193	336	774	2,723
Avg. sentence length	311	24.203	3.618	15.857	23.718	34.811
No. unique words	311	2,308.502	475.633	1,359	2,198	4,362

Table 2: Descriptive statistics of prospectus length variables

This table shows the number of observations (N), mean (Mean), standard deviation (SD), minimum (MIN), median (Median), and maximum (MAX) for different variables of document length.

	N	Mean	SD	MIN	50%	MAX
Total						
Underpricing (%)	745	0.085	0.330	-0.702	0.014	1.971
Sales (in M)	743	120.131	1,008.638	-0.463	3.009	25,012.691
Employees	733	319.905	1,786.501	0	24	40,131
NASDAQ return	744	0.016	0.050	-0.223	0.020	0.176
Tech company (D)	738	0.322	0.468	0	0	1
Regulated market (D)	745	0.203	0.402	0	0	1
IPO offering (D)	745	0.848	0.359	0	0	1
IPO						
Underpricing (%)	632	0.074	0.324	-0.702	0.011	1.854
Sales (in M)	631	128.915	1,079.714	-0.463	3.051	25,012.691
Employees	621	343.984	1,917.361	0	23	40,131
NASDAQ return	632	0.015	0.048	-0.223	0.018	0.144
Tech company (D)	625	0.342	0.475	0	0	1
Regulated market (D)	632	0.234	0.424	0	0	1
Private placement						
Underpricing (%)	113	0.145	0.360	-0.279	0.033	1.971
Sales (in M)	112	70.644	425.614	0.000	2.578	4,436.376
Employees	112	186.393	701.316	0	28	6,700
NASDAQ return	112	0.021	0.056	-0.112	0.027	0.176
Tech company (D)	113	0.212	0.411	0	0	1
Regulated market (D)	113	0.027	0.161	0	0	1

Table 3: Descriptive statistics of control variables

This table shows the number of observations (N), mean (Mean), standard deviation (SD), minimum (MIN), median (Median), and maximum (MAX) for selected variables of firm and offering characteristics. All currency values are expressed in Euros.

	N	Mean	SD	MIN	50%	MAX
Total						
Litigious-POW	745	0.648	0.290	0.211	0.563	2.198
Negative-POW	745	1.147	0.329	0.471	1.095	2.553
Positive-POW	745	0.816	0.202	0.346	0.788	2.179
Uncertainty-POW	745	1.325	0.341	0.499	1.282	3.669
Litigious-fastText	745	0.245	0.014	0.193	0.246	0.286
Negative-fastText	745	0.264	0.011	0.194	0.265	0.289
Positive-fastText	745	0.333	0.014	0.265	0.334	0.371
Uncertainty-fastText	745	0.300	0.013	0.231	0.301	0.334
Litigious-SBERT	745	0.308	0.012	0.263	0.308	0.343
Negative-SBERT	745	0.299	0.016	0.250	0.298	0.353
Postive-SBERT	745	0.287	0.018	0.235	0.286	0.345
Uncertainty-SBERT	745	0.282	0.015	0.226	0.282	0.327
IPO						
Litigious-POW	632	0.566	0.184	0.211	0.529	1.495
Negative-POW	632	1.074	0.235	0.471	1.063	1.984
Positive-POW	632	0.825	0.203	0.346	0.805	2.179
Uncertainty-POW	632	1.248	0.264	0.499	1.246	3.669
Litigious-fastText	632	0.244	0.014	0.193	0.245	0.284
Negative-fastText	632	0.263	0.011	0.194	0.264	0.289
Positive-fastText	632	0.333	0.014	0.265	0.334	0.371
Uncertainty-fastText	632	0.299	0.014	0.231	0.300	0.334
Litigious-SBERT	632	0.307	0.012	0.263	0.308	0.343
Negative-SBERT	632	0.297	0.014	0.250	0.296	0.336
Postive-SBERT	632	0.289	0.018	0.235	0.288	0.345
Uncertainty-SBERT	632	0.281	0.015	0.226	0.281	0.321
Private placement						
Litigious-POW	113	1.110	0.336	0.303	1.121	2.198
Negative-POW	113	1.551	0.464	0.672	1.569	2.553
Positive-POW	113	0.770	0.191	0.350	0.739	1.625
Uncertainty-POW	113	1.753	0.402	0.678	1.817	2.670
Litigious-fastText	113	0.249	0.012	0.210	0.251	0.286
Negative-fastText	113	0.268	0.011	0.239	0.269	0.289
Positive-fastText	113	0.330	0.016	0.291	0.333	0.355
Uncertainty-fastText	113	0.300	0.011	0.244	0.303	0.320
Litigious-SBERT	113	0.308	0.012	0.269	0.310	0.333
Negative-SBERT	113	0.309	0.020	0.259	0.309	0.353
Postive-SBERT	113	0.278	0.016	0.242	0.278	0.319
Uncertainty-SBERT	113	0.286	0.018	0.239	0.286	0.327

Table 4: Descriptive statistics sentiment analysis variables

This table reports the number of observations (N), mean value (Mean), standard deviation (SD), minimum value (MIN), median value (Median), and the maximum value (MAX) for the different types of sentiment variables. The POW variables represent the percentage of words of the respective word list in relation to total word count. The fastText and SBERT variables are expressed through cosine similarities.

	N	Mean	SD	MIN	50%	MAX
Total						
BOW (cosine-similarity)	695	0.697	0.055	0.392	0.706	0.815
simBERT	695	0.773	0.015	0.715	0.774	0.808
IPO						
BOW (cosine-similarity)	589	0.700	0.053	0.430	0.707	0.815
simBERT	589	0.773	0.014	0.734	0.773	0.807
Private placement						
BOW (cosine-similarity)	106	0.685	0.066	0.392	0.697	0.776
simBERT	106	0.777	0.016	0.715	0.779	0.808

Table 5: Descriptive statistics similarity analysis variables

This table reports the number of observations (N), mean value (Mean), standard deviation (SD), minimum value (MIN), median value (Median), and the maximum value (MAX) for the different types of similarity variables. BOW (cosine-similarity) and simBERT variables are expressed through cosine similarities.