JUSTUS LIEBIG UNIVERSITY
FACULTY OF ECONOMICS AND BUSINESS STUDIES
Department of Statistics and Econometrics

Three Essays on Applications of Text Analytics in Economics

A Thesis submitted in fulfillment of the requirements of
Doctor rerum politicarum (Dr.rer.pol.)

by
Dania Margarita Eugenidis
M.Sc., M.A.

Supervised By
Prof. Dr. Peter Winker
Prof. Dr. Alina Sorgner

January 24, 2023

Declaration of Authorship

I hereby declare that I completed the papers submitted and listed hereafter independently and with only those forms of support mentioned in the relevant paper or in the following supplementary list. When working with the authors listed, I contribute no less than a proportionate share of the work. In the analyses that I have conducted and to which I refer in the papers, I have followed the principles of good academic practice, as stated in the Statue of Justus Liebig University Giessen for Ensuring Good Scientific Practice.

Dania Eugenidis

.....

Submitted Articles

1. Eugenidis, D., Lenz, D., Leser, C., Schleer-van Gellecom, F. & Winker, P. (2020). Text-mining basierte Analyse der Kapitalmarktreaktionen auf Ad-hoc-Mitteilungen, *Corporate Finance*, 09-10.

2. Eugenidis, D., Kinne, J., & Lenz, D. (2022). Analysing gender equality at the firm level *MAGKS Joint Discussion Paper Series in Economics*, 14-2022.

3. Eugenidis, D., & Lenz, D. (2022). Measuring Gender Differences in Personalities through Natural Language in the Labor Force: Application of the 5-Factor Model, Philipps-Universität Marburg, Faculty of Business Administration and Economics, Department of Economics, (No. 202240).

Acknowledgments

I would like to express my deepest appreciation and thanks for Peter Winker. He always supported me with his invaluable guidance, support, expertise and helped me navigate through the complexities of research and academia. His constructive feedback, insightful comments, and suggestions have helped me refine my ideas and improve the quality of my work. In addition, I would like to thank my colleagues at work, who can probably best relate to the academic work. This goes to my research and office colleague David Lenz, Jenny Bethäuser, Elena Tönjes, Albina Latifi and Viktoriia Naboka. A special thanks goes to Carmen Hersener and her support with every possible concern. Even in stressful times, one could count on being warmly welcomed in the office.
Additionally, I would like to thank Alina Sorgner for her support and supervision of my dissertation. I would like to thank my co-authors Jan Kinne, Frauke Schleer-van Gellecom and Christoph Leser for their pleasant teamwork.
I would like to express my deepest gratitude to my family for the unwavering support throughout my academic journey, especially during the process of completing my thesis. This goes to Doris, Nikos and Alicia Eugenidis, as well as Georg and Annelise Fornoff.
Lastly, I would like to take this opportunity to express my sincere gratitude for the constant support of my friends Mara Czogalla, Kyra Hagge, Sinem Kandemir and Judith Rauch, with whom I share many interests, experiences and memories both professionally and personally.

Έν οίδα ότι ουδέν οιδα
Σωκράτης

Dania Eugenidis
February 25, 2023

Table of Contents

Chapter 1

Introduction

Text analytics plays an increasingly important role in various fields, including economics. Overall, text analyses can provide valuable insights into economic trends and events, helping to make better informed decisions in policy making, trend analysis and predictions. It provides a powerful tool for analyzing large amounts of mainly unstructured textual data and enlarges the spectrum of answering economic questions. It is possible to transform the field of Economics by enabling researchers to analyze unstructured, large and complex datasets and extract information in ways that have been impossible previously.
In this thesis, different applications of text analytics form the core of this work: While the first paper presents a capital market analysis, the following two articles (chapters 3 and 4) investigate a labor market related issue. Other fields of application include for example economic forecasting or policy analysis.

The first paper, "Text mining based analysis of capital market reactions to ad hoc announcements" (chapter 2), examines capital market reactions to the issuance of ad hoc announcements. Unforeseen business developments include, for example, changes in corporate structure or profit warnings. We show that these textual announcements can have a sometimes significant impact on firm value through additional stock price effects. We examine whether different contents of a predominantly negative nature lead to different price effects. Using current text mining techniques, we can identify topics that lead to significantly negative capital market reaction. This data-driven prediction method is a complement to classical approaches such as theoretical internal company models. This results in a gain in efficiency, more precise prediction quality in risky times and cost reduction through automation. In the long term, this can provide a hedge against short-term price fluctuations. Furthermore, terms such as "forecast" or "uncertainty" were identified within the ad-hoc messages, which, in combination with a

negative connotation, can cause a loss of value in the millions.

The second paper, "Analyzing gender equality at the firm level" (chapter 3), includes a model framework that, for the first time ever, can measure gender inequality at the firm level across an entire country and across all sectors. While a variety of quantitative metrics already exist to indicate gender imbalances in the labor market, such as the female participation rate in managerial positions or the gender wage gap, these are not typically applied to firm-level data. However, these are generally not collected at the company level. Thus, while structures can be revealed at an aggregate level, a deeper insight into structural grievances at the company level is not possible. The Gender Equality Firm Index (GEFI) aims to address these gaps and provide a first insight into gender equality. For this purpose, in addition to official company data, company website texts are used. The use of text data is a first approach to identify and quantify latent structures.

The third and final paper, "Measuring gender differences in personalities through natural language in the labor force: Application of the 5-Factor Model" (chapter 4), focuses on the perception and representation of individuals in the labor market depending on their gender. For this, gender stereotypes will be identified and quantified using website data and publicly available big data. Using current NLP methods and keyword searches, we first identify people and their genders and then investigate whether there are differences in the representation or description depending on gender through the use of certain terms. These are aligned with the five-factor model and a sublevel according to which there are different degrees of expression for men and women in the literature. We find statistically significant differences in how men and women are described. While men in our dataset are predominantly described using performance-related terms, women are mainly described using character-describing terms. If the individual terms are assigned to the Big 5 factors, the results are also in line with the literature. This approach combines novel text-based quantitative measurement techniques to map this latent level of potential inequality. It is also novel in that the analysis of gender stereotypes takes place from a combination of psychology, computer science, and linguistics in an enterprise context, which is low cost and at a very granular level.

Chapter 2

Text mining based analysis of capital market reactions to ad hoc announcements

Chapter 2: Text mining based analysis of capital market reactions to ad hoc announcements

This chapter is based on the article:

Title:	Text mining based analysis of capital market reactions to ad hoc announcements
Original Title:	Text-mining basierte Analyse der Kapitalmarktreaktionen auf Ad-hoc-Mitteilungen
Authors:	**Dania Eugenidis**: Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany; Contribution: 30%. **Dr. David Lenz**: Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany; Istari.Ai, Mannheim **Christoph Leser**: PricewaterhouseCoopers GmbH WPG **Prof. Dr. Frauke Schleer-van Gellecom:** PricewaterhouseCoopers GmbH WPG **Prof. Dr. Peter Winker:** Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany
Status:	Publication, *Corporate Finance* 2020, 09-10.

Note:

The original article is written and published in German. In order to ensure consistency of the dissertation, a translation into English has been made by the author of this dissertation. All associated errors and discrepancies are therefore solely their responsibility. In addition, adjustments were made to the numbering of figures and tables. The original German published version can be found in appendix IV of this chapter. The appendices listed in the original article have been translated into English for this thesis.

Text mining based analysis of capital market reactions to ad hoc announcements

Abstract

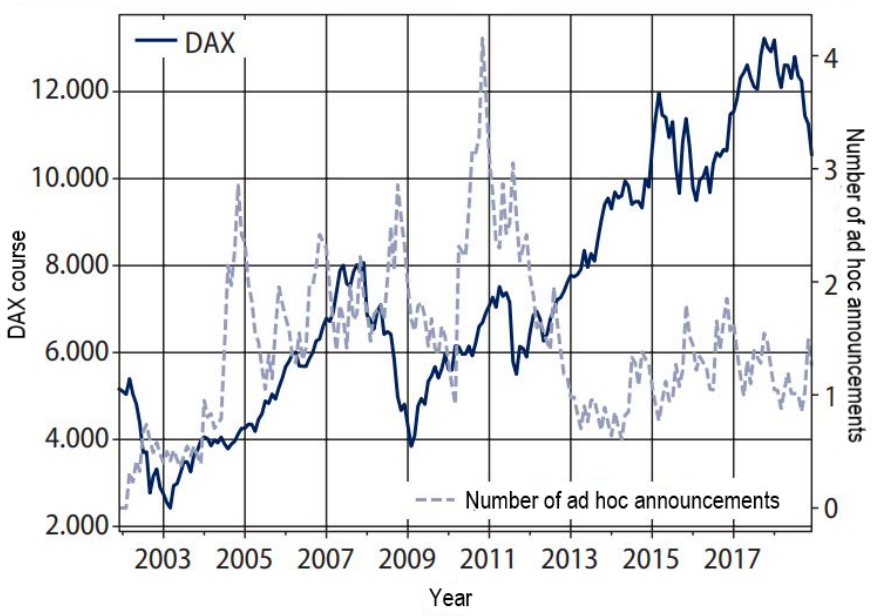
The reaction of the capital markets to unexpected business developments can be observed regularly when companies are forced to issue ad hoc announcements. This paper shows the significant impact these announcements can have on the value of a company. We analyze whether different contents of ad hoc announcements can be identified that lead to significant price effects. The study shows that for an average market capitalization in the DAX (40 billion euros) an expected loss in value of 158 million euros must be assumed as a result of the increased importance of the topic "forecast" in a negative connotation.

I. Introduction

Especially in times of high uncertainty or an (impending) economic downturn, ad hoc announcements and the associated profit warnings by companies listed on the capital market increase. Fig. 1 shows this correlation using the DAX as an example. The often contradictory development of share prices and the number of ad hoc announcements is apparent. This is particularly noticeable in the course of the financial market crisis in 2008/09 and in the years immediately following. This indicates that companies are often not sufficiently able to process information correctly and with foresight and to anticipate (sales/profit) developments with adequate accuracy. Therefore, incorrect assessments have to be corrected, which can lead to an ad hoc announcement or profit warning. For example, Baule and Tallau (2012) observe that ad hoc announcements with the same information as regular periodic reports lead to about twice as large market reactions. Moreover, positive news tend to be announced early via an ad hoc announcement, while negative news are published with a delay (Baule and Tallau, 2012).[1]
Similar effects are also evident in the following analysis. The average of absolute returns for observations belonging to ad hoc announcements with 4.5% is more than twice as high as for the other observations (2.15%), and this difference is highly significant. Fig. 2 shows histograms of the empirical distributions of the returns for the observations without (blue) and with (brown) ad hoc announcements.

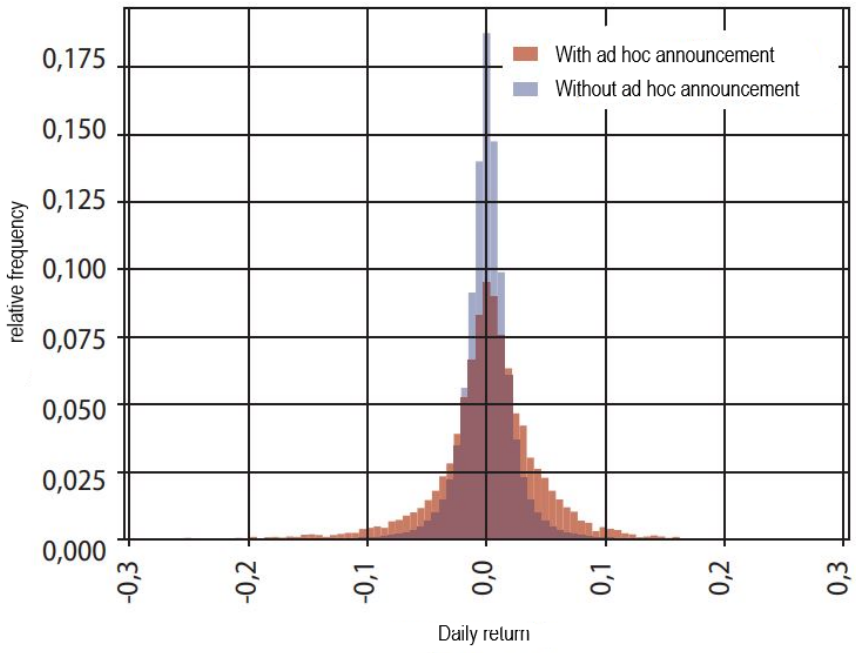
[1]Further analyses can be found at Muntermann and Guettler (2007), Feuerriegel, Neumann and Ratku (2016).

Figure 1: Development of the DAX (left scale) and number of ad hoc announcements per month on DAX stocks (smoothed exponentially over 12 months)



It is confirmed that returns are often larger for observations with ad hoc announcements, whereby both distributions exhibit a slight left skewness of similar magnitude (-0.970 without profit warning and -0.927 with profit warning). These results already indicate that ad hoc announcements induce uncertainty in the form of stronger price fluctuations in the market. The purpose of this study is to examine whether different content in ad hoc announcements can be identified that can be associated with stronger effects on share prices (especially in the case of negative news). We investigate whether the application of current text mining methods to texts of ad hoc announcements leads to topics with substantial content (topics). These topics are also put into relation with exogenously given conceptual fields such as "uncertain" and "forecast" in order to investigate whether topics in the context of these conceptual fields lead to a significantly negative capital market reaction. This analysis is particularly interesting in light of the current discussion on the application of Machine Learning (ML) based methods in the classical planning and forecasting process - especially for companies listed in the capital market. Traditional manual forecasting methods are time-consuming and often provide reliable results only in times of low uncertainty and in stable industries compared to ML-based methods (PwC, 2015). One reason for this is often that less data-driven approaches such as "gut feeling" or internal "political" company decisions influence forecasting results (PwC, 2011). In addition to the quality gain through objectification of the forecast creation as well as significant broadening of the considered data, ML methods provide a strong efficiency gain as well as a higher control focus, since the time saved in the forecast process can be invested in activities with more control relevance (e.g. scenario analyses). In addition,

Figure 2: Distributions of returns for observations without (blue) and with (brown) ad hoc announcements.



a high forecast quality strengthens investors' confidence in a company and thus makes it less vulnerable to short-term price fluctuations. In order to further support this hypothesis, this study also examines the capital market impact of profit warnings on topics in the "forecast" field.

II. Data base and preparation of the text data

The analysis contains price data for all shares included in the corresponding indices for DAX, MDAX, SDAX, FTSE 100, CAC40 and S&P 500 from January 3rd, 2000 until December 31st, 2018, for which there was at least one ad hoc announcement in the period under review.[2] Furthermore, ad hoc announcements for all these stocks are available for the period from Dec. 21, 2001 to Dec. 31, 2018. Continuous daily returns are calculated from the price data as the difference between the logarithmized prices for further analysis. A total of 7,743 ad hoc announcements are available as text data. The dataset contains a total of 2,707,629 observations of daily returns for stocks for which there was no ad hoc announcement at the respective time (observations without ad hoc announcement) and 7,455 daily returns for which there was an ad hoc announcement at the respective time (observation with ad hoc announcement).

[2]The data was obtained on the basis of the Capital IQ database.

Before using statistical methods for text analysis such as the Latent Dirichlet Allocation (LDA) topic model used here (see also Chapter III.2), it is necessary to prepare the text data. For this purpose, the usual techniques (see, e.g. Lenz and Winker, 2020; Hansen, McMahon and Prat, 2018; Larsen and Thorsrud 2019; Lüdering and Winker, 2016 and Hansen and McMahon, 2016) proposed in the literature are applied first: In the first step, punctuation marks and currency symbols, additional spaces and numbers are removed. In addition, so-called "stopwords" are eliminated. These are 532 words from a dictionary (English) that occur very frequently and usually have no relevance to the content, e.g. "which", "and" or "but". Finally, all words with a maximum of two letters are also removed. Tab. 2.1 shows the changes caused by this 1st stage of data preparation using two examples in the second column. This first stage of preprocessing

Table 1: Examples for preprocessing

Original ad hoc announcement	After preprocessing (1. stage)	After preprocessing (2. stage)
PSA Peugeot Citroen provided earnings guidance for full year 2006. The company cut its annual guidance, saying it no longer expects second-half operating profit margin to match the 2.4% recorded in the first half.	psa peugeot citroen provide earning guidance year company cut annual guidance say longer expect half operating profit margin match record half	provide guidance cut annual guidance say longer half operating profit margin match record half
Rio Tinto Group had its 2004 and 2005 earnings estimates raised by Morgan Stanley on expectations profit from its diamond unit will be higher than previously forecast. The 2004 estimate was raised 1.5 percent to USD 1.75 a share, and the 2005 forecast 1.2 percent to USD 1.93 a share.	rio tinto group earning estimate raise morgan stanley expectation profit diamond unit higher previously forecast estimate raise percent share forecast percent share	group estimate raise expectation profit unit higher previously forecast estimate raise percent forecast percent

reduces the number of all occurring words and thus the dimension of the word space, while at the same time only a minor loss of information is expected to occur. In a second stage of preprocessing, words are removed that either appear infrequently (here in less than 5 documents) or are contained in a very large proportion of all texts (here in more than 75% of the documents examined). The rationale for this is that extremely infrequently used words cannot be used to derive topics that characterize a particular type of announcement that occurs with at least a certain minimum frequency. Very frequently

occurring words, on the other hand, would be found in most of the identified topics and thus would not contribute to the differentiation of the topics. This second step is often seen as the removal of domain-specific stopwords, which are difficult to capture with general stopword lists as described in step 1. The texts resulting after both stages of preprocessing are shown for the examples in the third column of Tab. 1. After the two processing stages described above, a total of 1,975 words remain for the further analysis.

III. Topic modeling

1. Text data preparation

For the estimation of the topics, initially only 3,846 randomly selected documents are used (i.e. 50% of the available documents), in order to be able to check the derived statements on a test data set at a later time (external validation). Otherwise, there is a considerable risk of overfitting and bias generated by the method due to the numerous selection steps in the preparation and evaluation of the text data (Egami et al., 2018).

2. Implementation and results of the LDA model & topic interpretation

a) LDA model and Word Cloud

This section presents the procedure and results of the LDA model. LDA is a generative probabilistic model for collections of discrete data sets such as text corpora. Using LDA to extract topics has several advantages: First, it avoids subjective bias in manual topic extraction; second, it allows greater flexibility in topic selection to better fit into our corpus (text). In addition, the data mining approach used makes it possible to process thousands of documents and thus obtain a comprehensive picture of the ad hoc announcements. Such an exercise would be prohibitively difficult and costly with manual labeling.

We use the LDA model established in the literature to identify topics (Blei et al., 2003). The method underlying the LDA model is based on the idea that 1) topics can be described by the relative frequency of the words they contain, and 2) texts can be described by the relative relevance of the topics they contain. Specifically, each topic is represented as a probability distribution over all the "possible", i.e., words remaining after data preparation, while each document is represented as a probability distribution over all the "possible", i.e., topics identified during the estimation of the model. To put it simply, in the LDA model, texts are represented as a distribution over topics, and topics are represented as a distribution over words.

Figure 3: Generative process of texts from Topic according to Blei (2012)

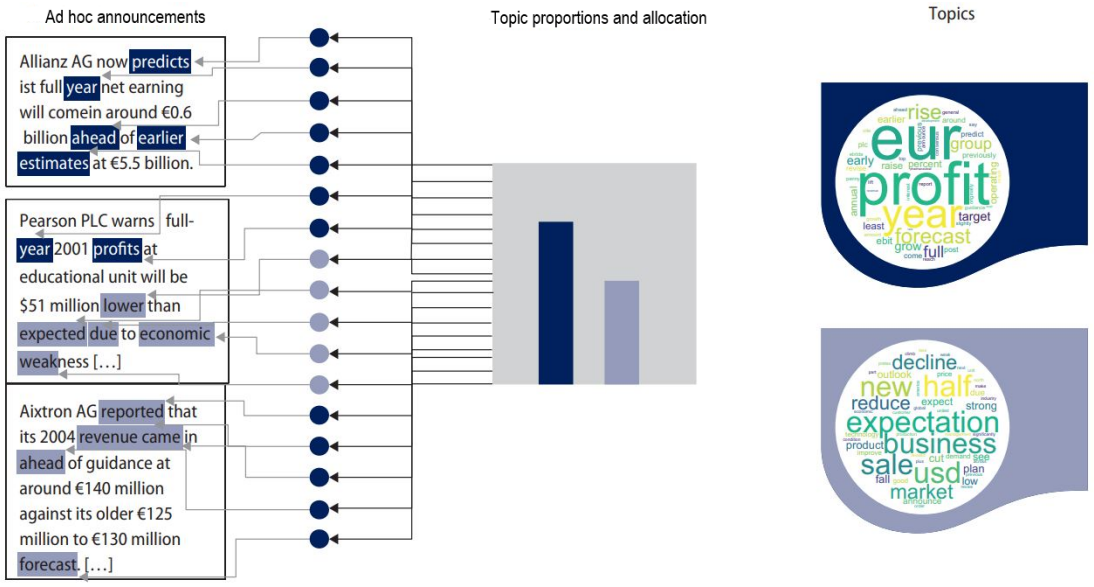


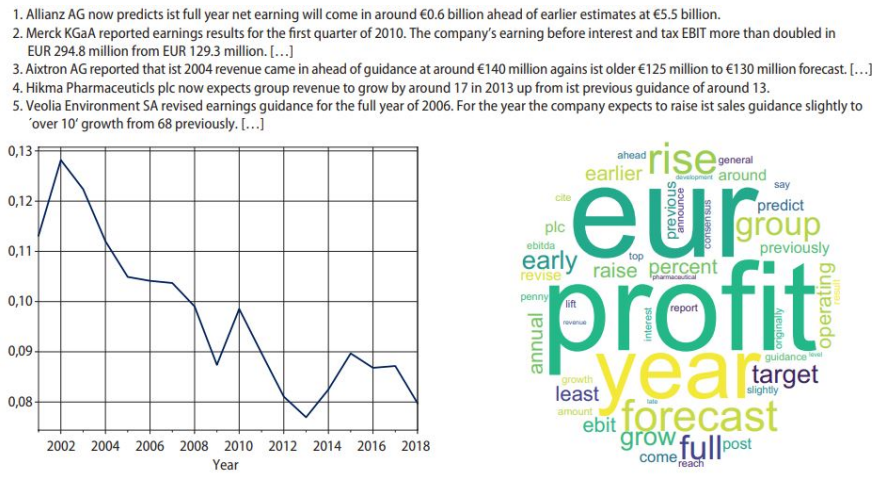
Figure 3 illustrates this concept with an example (according to Blei, 2012): The distribution of the topic fields across the texts, and of the words across the topic fields, is visualized for the topics "forecast" (yellow) and "uncertainty" (green). The topics are characterized by the words that are particularly relevant (= frequent occurrence relative to other topics). The challenge in estimating the model is to simultaneously determine the topics and their weighting in the texts on the basis of the available texts alone. This is done by a two-stage, approximate optimization procedure with the goal of achieving the best possible description of the texts based on a manageable, predefined number of topics.

In summary, 10 topics are identified[3]. Fig. 4 shows an example of a composite representation for topic 2[4]. To interpret a topic, a list of the most important 3-30 words of the topic is usually considered. On the basis of the words an idea of the underlying abstract topic is obtained. Another way of representing topics is via word clouds. In the upper left corner of Fig. 4, a word cloud is shown that contains the most relevant words for the topic ("Top Words"). The size of a word corresponds to the relevance of the word in the topic. Looking at the word cloud, topic 2, for example, could be identified as a "profit/forecast" topic. For the practical implementation of the LDA model, some parameter values need to be specified. We optimize the hyperparameters of the LDA model using a random search. This involves randomly drawing from a predetermined range of possible hyperparameter values. The set of hyperparameters that minimizes the selected information criterion is chosen. This meta-optimization resulted in values

[3]We name topics from 0-9, i.e. the first topic is indexed with "0".

[4]Corresponding summaries for the remaining topics can be found in the appendix in tab A3. In addition, a description of the relationship between topics and announcement types can be found in the appendix under II. The appendix is available online at: CF1343653.

Figure 4: Results of the LDA model for topic 2 - "profit/forecast"



for the doc-topic prior alpha of 0.1, for the topic-word prior beta of 0.5, and a learning decay of 0.7 for the data at hand, assuming that a number of ten topics should be sufficient for description given the very specific nature of the text.

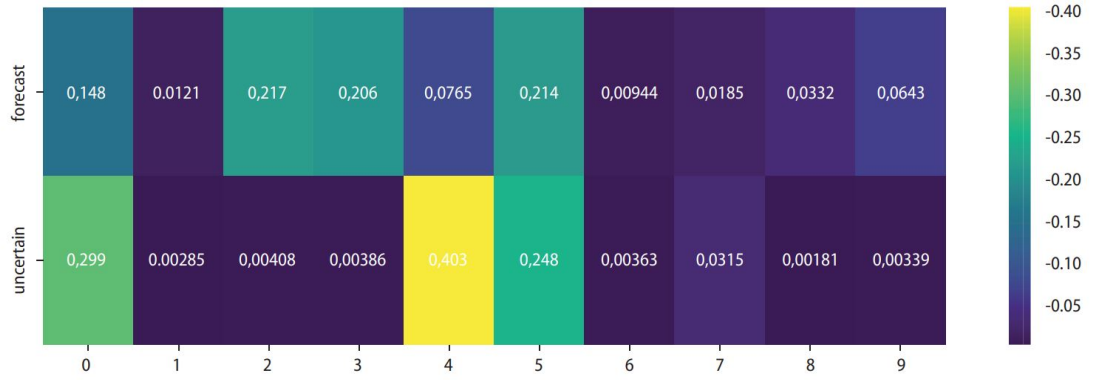
b) Keywords (term sets) and topics

In the course of an additional separate analysis, two term sets were formed following the approach of Baker et. al (2016). A term set is a list of words that have a high semantic similarity to a certain term, such as the word "prediction" to the word "forecast". We use termsets to simplify the interpretation of the topics found, since the topics generated by LDA models are not always easy to interpret (Baldwin et al., 2017). To define these term sets on the words "forecast" and "uncertain," we relied on an external data source that includes prebuilt models and corpora for computer-aided natural language processing[5]. The Word2Vec method (Mikolov et al., 2013) is used to identify words that have a high semantic similarity to each other. By relying on this method, the term sets are determined independently of the specific topics and the documents in the data set. The list of semantically most similar words for "forecast" and "uncertain" that also occur in the data set under analysis are reported as the "forecast" and "uncertain" term sets in the Appendix, Table A2.1.

Figure 5 shows the relevance of the two term sets for the respective topics. The numbers indicate how often words from the term sets occur in the documents that are assigned to the topics. The weights of the topics for the respective documents are also taken into account. The upper row in Fig. 5 shows that the words from the term set "forecast" appear relatively frequently in documents with a high weight of topics 2, 3 and 5, each

[5]RaRe-Technologies, Data repository for pretrained NLP models and NLP corpora – Gensim Data 2018, GitHub repository, http://hbfm.link/7870 (retrieved: 03.03.2020).

Figure 5: Relevance of term sets per topic



covering more than 20% of the total frequency. For the term set "uncertain" (lower row in Fig. 5), there is an even greater concentration for topics 0, 4 and 5.[6]

IV. Analysis of the relationship between topics and stock price development

1. Event regressions with symmetric effects

For this approach, it was assumed that the influence of the topic weights is symmetric for positive and negative price changes. Five estimates were made with a width of the event window of $j = 1, \ldots, 5$ days. The estimated model is:

$$\Delta P_{i,j} = \beta_1 + \beta_2 \Delta I_{i,j} + \beta_3 t_{0,t,i} + \beta_4 t_{1,t,i} + \cdots + \beta_{11} t_{9,t,i}, \tag{1}$$

where $\Delta P_{i,j}$ denotes the log return of the stock price associated with announcement i in the event window. $\Delta I_{i,j}$ is the log return of the associated market index. In addition, $t_{0,t,i}, \ldots, t_{9,t,i}$ represent the topic weights of the considered announcement[7]. The estimated coefficients with associated standard deviations (in parentheses) are presented in table 2.

As expected, there is a strong correlation between the returns of individual stocks and the associated index with estimated coefficients in the range of 0.95 - 1.16. In addition, there is a significant increase in the determination coefficient R^2 when an event window of length greater than 2 is selected. This may be due to the fact that for very short observation periods, depending on the time of the announcement, leading or lagging

[6]The appendix provides additional robustness analyses. The appendix is available online at: CF1343653 and in the appendix.

[7]Since the topic weights each add to one, we did not include one of the topic weights (for topic 6) to avoid the problem of perfect multicollinearity that would otherwise arise. This should be taken into account when interpreting the coefficients.

Table 2: Coefficients and standard deviations for model (1)

	$t_{-1} \to t_1$	$t_{-2} \to t_2$	$t_{-3} \to t_3$	$t_{-4} \to t_4$	$t_{-5} \to t_5$
Constant	-0.00567	0.01250	0.00243	0.00387	-0.00062
	(0.015)	(0.017)	(0.018)	(0.019)	(0.019)
Index	0.96560***	1.06247***	1.15809***	1.06921***	1.09845***
	(0.061)	(0.046)	(0.043)	(0.041)	(0.039)
Topic 0	-0.01986	-0.03513	-0.01777	-0.00882	-0.02503
	(0.027)	(0.030)	(0.032)	(0.033)	(0.034)
Topic 1	0.05927*	0.04937	0.08147**	0.07966**	0.08114**
	(0.030)	(0.034)	(0.036)	(0.037)	(0.038)
Topic 2	0.04828*	0.04269	0.06279**	0.06361**	0.07716**
	(0.025)	(0.027)	(0.029)	(0.030)	(0.031)
Topic 3	0.03219	-0.02093	-0.01266	-0.01793	-0.00994
	(0.031)	(0.034)	(0.036)	(0.038)	(0.039)
Topic 4	-0.08492***	-0.12320***	-0.13666***	-0.12539***	-0.12305***
	(0.032)	(0.035)	(0.037)	(0.039)	(0.040)
Topic 5	-0.03212	-0.06699**	-0.06228**	-0.07528***	-0.06195**
	(0.024)	(0.026)	(0.028)	(0.029)	(0.030)
Topic 7	-0.01969	-0.02496	-0.02671	-0.01338	-0.02549
	(0.029)	(0.032)	(0.034)	(0.036)	(0.037)
Topic 8	0.02456	0.01019	0.02630	0.01407	0.02614
	(0.022)	(0.024)	(0.025)	(0.026)	(0.027)
Topic 9	0.03272	0.01977	0.02986	0.02635	0.04460
	(0.024)	(0.027)	(0.029)	(0.030)	(0.030)
Stats					
No. Obs.	3605	3472	3447	3462	3491
R^2	0.07	0.14	0.19	0.18	0.20
DW	1.94	1.88	1.89	1.91	1.89
F-Stat	3.72***	4.53***	5.37***	4.43***	4.70***

$t_{-1} \to t_1$ represents the time window from t-1 to t+1, $t_{-2} \to t_2$ represents the time window from t-2 to t+2, etc. The columns represent the different event windows from ±1 to ±5 days around the time of the ad hoc announcement. In addition to the coefficients for index return (Index) and topic weights (topic 0, ...), the symbols *, **, and *** indicate whether the null hypothesis of a parameter value can be rejected at the 10, 5 and 1% level, respectively.

effects are not included in the event window. Therefore, these are not included into the analysis. We therefore recommend looking at somewhat longer periods when interpreting the results, e.g. ±3 days.

Comparing the coefficient of determination for the same regressions without the variables for the topic weights results in a value that is only about one percentage point lower. Nevertheless, the influence of the topic weights is relevant. Both in terms of the sign and in terms of the size and significance of the coefficients, there are only minor differences when considering different time windows from ±3 days. The weights for topic 1 and topic 2 show significantly positive effects, those for topic 4 and topic 5 significantly negative effects.

In addition, an F-test was used to examine whether the null hypothesis of no effect of

all topic weights on the modeled return should be rejected. The value of the relevant F-statistic is shown in the last line. In all cases, it is significantly above the critical value for the 1% level, so that the null hypothesis is rejected. Thus, we can assume a statistically highly significant additional influence of the topic weights. In addition, it should be recalled that the two topics (4 and 5) for which we consistently and significantly find negative effects on stock returns are exactly the two topics that also show the highest relevance with respect to the term set "uncertain". This can be seen as confirmation of the hypothesis that perceived uncertainty leads to disproportionate stock price losses.

2. Robustness of the results

In the results presented previously, the determination of the topics and the topic weights is based on the same data as the regression performed afterwards. At least theoretically, there is a risk of bias due to the use of calculated regressors (i.e. the topic weights). A simple test for the existence of such biases can be done in the present case by running the regression for the observations that were not used for the selection and estimation of the topic model. Repeating the regressions for this test data set does not show any qualitative differences to the results reported here. Therefore, no major distortions can be assumed, which also indicates an external validity of the results.
Two further variants of the estimation confirm the robustness of the results. First, a simplified relation for the abnormal returns $\Delta S_{i,j} = \Delta P_{i,j} - \Delta I_{i,1}$ was considered. This variant also confirms the results for the topic weights.[8] Table A4 in the appendix gives an overview of the results of the estimation with $\Delta S_{i,1}$ as dependent variable.
Furthermore, we consider a variant that additionally includes the sign of the associated index as an indicator variable. With the indicator variable, observations with positive returns or positive excess returns can be analyzed separately. The results of this analysis can be found in the appendix in tab A5. In particular, the negative influence of topic 4 "profit/ forecast" on returns is confirmed here.

3. Further contextual interpretation of the results

In order to simplify the interpretation of the effects found, the reference to the term sets "forecast" and "uncertain" is given again in conclusion.

[8]The analyses can be found in Appendix IV, available online at: CF1343653.

a) Forecast

With relative weights of at least 15%, the term set "forecast" with the associated terms is particularly relevant for topics 0, 2, 3 and 5. It can be seen that different signs are estimated for the influence of the weights of these topics. While the weight of topic 2 has a positive effect, a high proportion of topic 5 rather indicates a negative price reaction. The effects for topics 0 and 3, on the other hand, are not significantly different from zero. These ambiguous results reflect that there are no clear expectations regarding the sign, as the words contained in the term set can be interpreted both negatively ("decline," "slowdown," "drop") and positively ("optimistic," "increase").
If we look at the ten texts with the highest weights for the topics for which the "forecast" term set is particularly relevant, we find a clear indication in most cases that incorrect predictions lead to the announcements. Thereby, both positive and negative forecast corrections are components of these announcements. For example, the text with the highest weight for topic 5 states: The Parker-Hannifin Corp. lowered its earnings forecast for the current quarter and fiscal year.". The text with the highest weight for topic 2 contains: "Allianz AG now predicts its full year net earning will come in around 0.6 billion euro ahead of earlier estimates at 5.5 billion euro". In each of the ten texts with the highest weights for topic 2, an economically positive statement can be identified, e.g., an unanticipated increase in profits. This is consistent with the results from the estimation, where a significant positive coefficient for the weight of topic 2 is shown in each specification for each time window. Therefore, a general statement regarding the effects of the "forecast" topics is not possible, since both terms with positive and negative meanings are contained in the term set.
To estimate the magnitude of the effects, we assume, for example, a situation in which the weight for topic 5 in an ad-hoc announcement increases by one standard deviation (0.059). The effect for this weight at a time window of ± 3 days is -0.067 according to tab. 4. For a share with average market capitalization in the DAX (40 billion euro), this would result in an expected loss in value due to the increased importance of the topic "forecast" in a negative connotation as in topic 5 of -0.067 * 0.059 * 40 billion = 158 million euro. Thus, the magnitude of the measured effects is quite substantial.

b) Uncertain

Regarding the term set "uncertain", topics 0, 4 and 5 show the highest relevance. In line with the expectation of a rather negative influence of "uncertain", the weights for topics 4 and 5 also show negative signs in the estimates. Also for these topics with a high share of "uncertain", an estimation of the economic relevance of the effects found can be made on the basis of the average market capitalization per share in the indices

considered. As an example, we consider the effect for the weight in topic 4 from tab. 4. For this, we report a coefficient of -0.39 for an event window of length ±3 days. The standard deviation for the weight in the sample is 0.041. For a share with an average market capitalization of just under 40 billion euro in the DAX, an announcement with a weighting for Topic 4 that is one standard deviation (i.e. 0.041) higher would therefore result in an expected loss in value of -0.139*0.041*40 billion = 228 million euro. The magnitude of the measured effects is therefore quite substantial in this case as well.

V. Summary

In summary, the analysis of the texts of ad hoc announcements using current text mining methods leads to topics with substantial content, which can also be related to exogenously specified conceptual terms such as "uncertain" and "forecast". The weighting of these topic fields in the announcements is strongly correlated with the resulting price changes. This relationship is robust because it was reproduced qualitatively even with a partial data set that was not used for modeling and estimation of the model.
It can be seen that an increase in the weighting of sensitive topics in relation to "uncertain" and "forecast" leads to economically relevant changes in returns and thus - in the case of negative announcements - also to considerable losses in value. For a share with an average market capitalization in the DAX or S&P, these losses can amount to hundreds of millions
Conversely, it seems plausible that by improving the forecasting procedures, e.g. with ML-based methods, a part of these selectively occurring large value losses could be avoided. In addition to the gain in quality - and the associated possible reduction in company value losses - through the objectification of forecast generation, ML processes deliver a strong gain in efficiency as well as a higher control focus, since the time saved in the planning process can be invested in activities with control relevance. In addition, a high forecast quality strengthens investors' confidence in a company and thus makes it less susceptible to short-term price fluctuations.

References

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. The quarterly journal of economics, 131(4), 1593-1636.

Baule, R., & Tallau, C. (2012). Market response to mandatory pre-earnings-announce

ments-evidence from ad-hoc disclosures in Germany. Available at SSRN 1660679 (last received 03.03.2020).

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(1), 993-1022.

Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E., & Stewart, B. M. (2018). How to make causal inferences using texts. arXiv preprint arXiv:1802.02163. (last received: 03.03.2020).

Feuerriegel, S., Ratku, A., & Neumann, D. (2016). Analysis of how underlying topics in financial news affect stock prices using latent dirichlet allocation. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1072-1081). IEEE.

Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. Journal of International Economics, 99, S114-S133.

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: a computational linguistics approach. The Quarterly Journal of Economics, 133(2), 801-870.

Larsen, V. H., & Thorsrud, L. A. (2019). The value of news for economic developments. Journal of Econometrics, 210(1), 203-218.

Lenz, D., & Winker, P. (2020). Measuring the diffusion of innovations with paragraph vector topic models. PloS one, 15(1), e0226685.

Lüdering, J., & Winker, P. (2016). Forward or backward looking? The economic discourse and the observed reality. Jahrbücher für Nationalökonomie und Statistik, 236(4), 483-515.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.

Muntermann, J., & Guettler, A. (2007). Intraday stock price effects of ad hoc disclosures: the German case. Journal of International Financial Markets, Institutions and Money, 17(1), 1-24.

PricewaterhouseCoopers (2011): Financial planning: Realizing the value of budgeting and forecasting

PricewaterhouseCoopers (2015). Digital Controlling – Digitale Transformation im Controlling, http://hbfm.link/7799 (last received: 21.03.2020).

RaRe-Technologies (2018), Data repository for pretrained NLP models and NLP corpora – Gensim Data, GitHub repository, http://hbfm.link/7870 (last received: 03.03.2020).

Sorodoc, I., Lau, J. H., Aletras, N., & Baldwin, T. (2017, April). Multimodal topic labelling. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers (pp. 701-706). (last received: 10.09.2020).

Appendix

Appendix I: Word lists

Table A1: List of words of the respective term sets

forecast	uncertain
• projected, projection • estimate • forecasting, forecasted • estimate • outlook • growth • expected • prediction • profit • decline • rise • slowdown • drop • anticipated • revenue • inflation • output • optimistic • revised • increase • adjusted	• uncertainty, certainty • unlikely, likely • doubtful • optimistic • difficult • future • cautious • scenario • indication • due • outlook • confident • unpredictable

Appendix II: Topics

Figure A1.1: Results of the LDA model for Topic 0

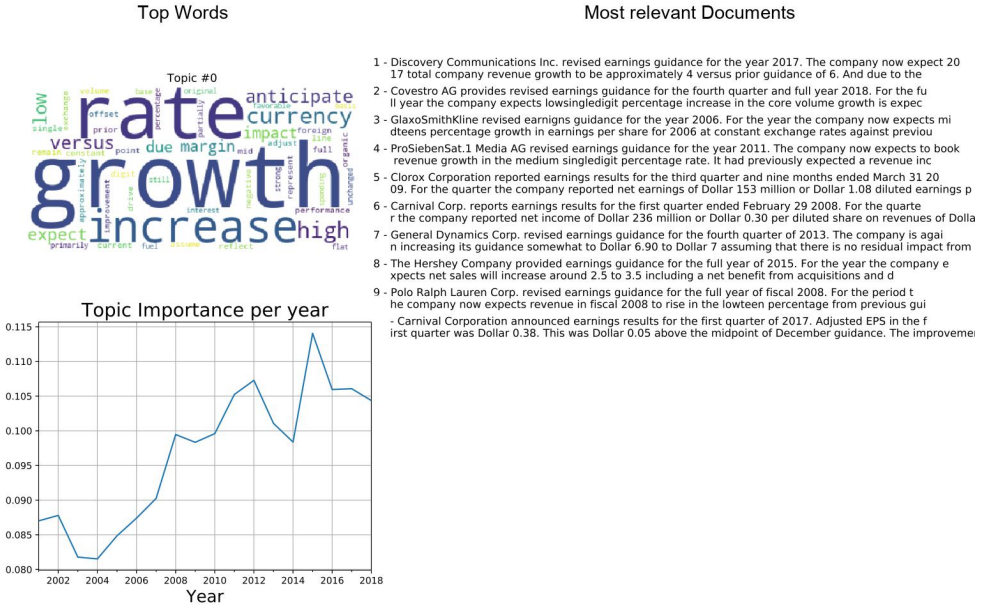


Figure A1.2: Results of the LDA model for Topic 1

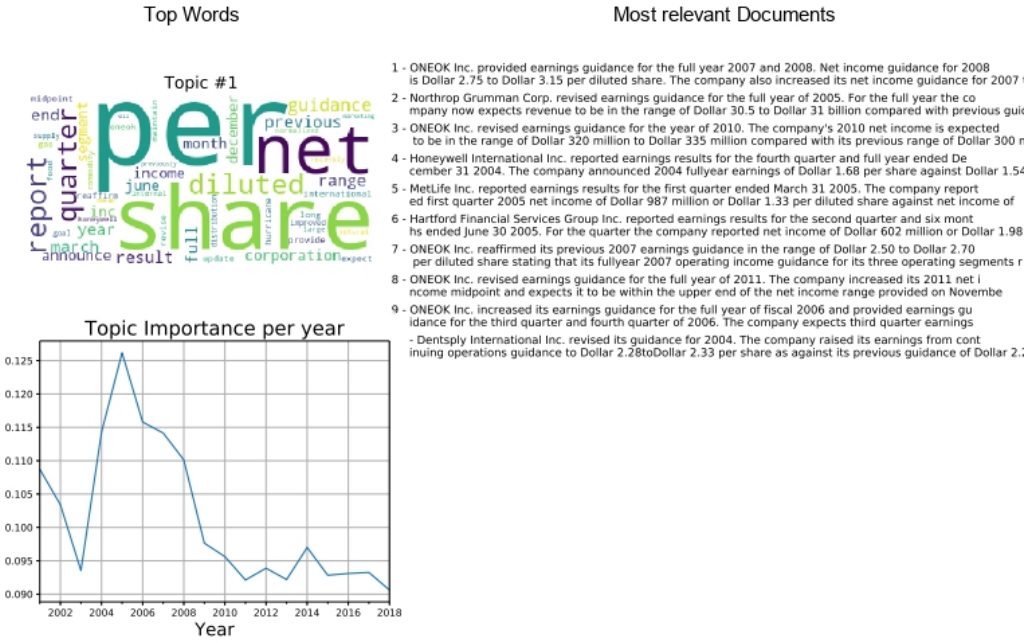


Figure A1.3: Results of the LDA model for Topic 2

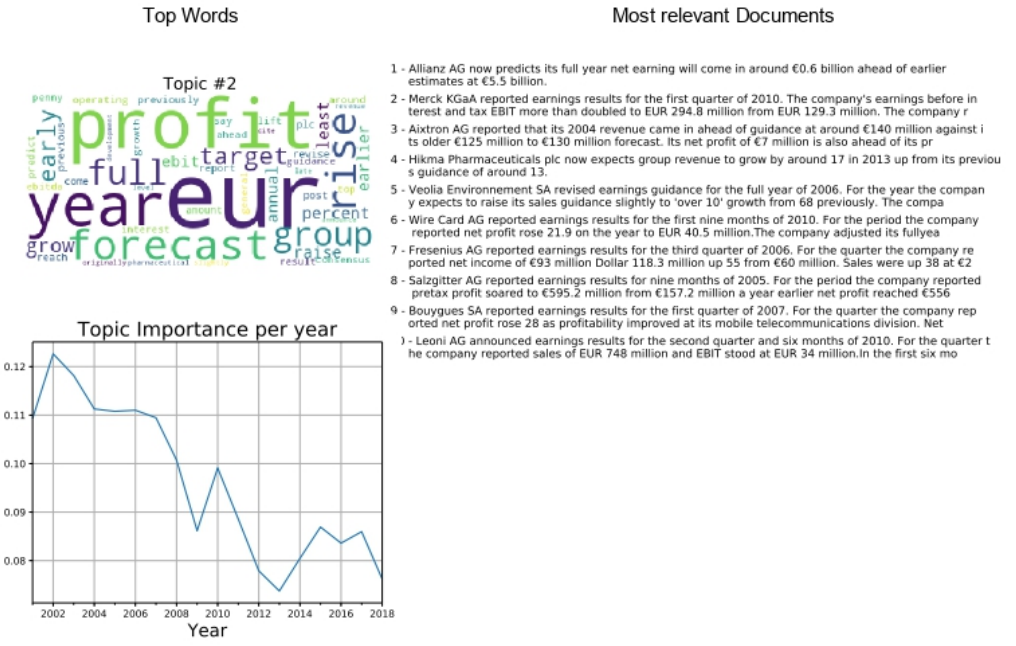


Figure A1.4: Results of the LDA model for Topic 3

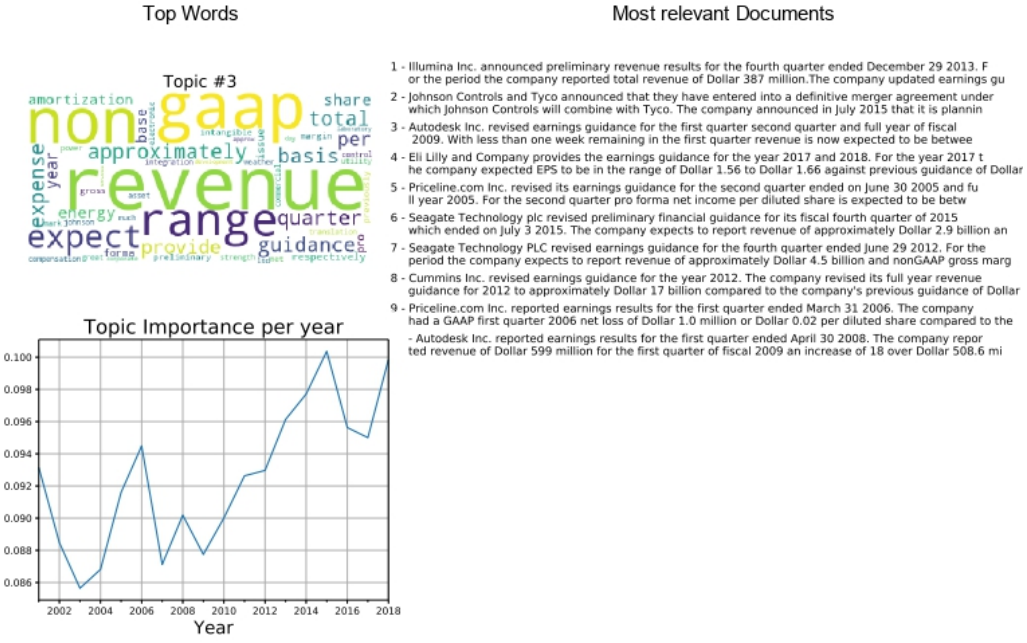


Figure A1.5: Results of the LDA model for Topic 4

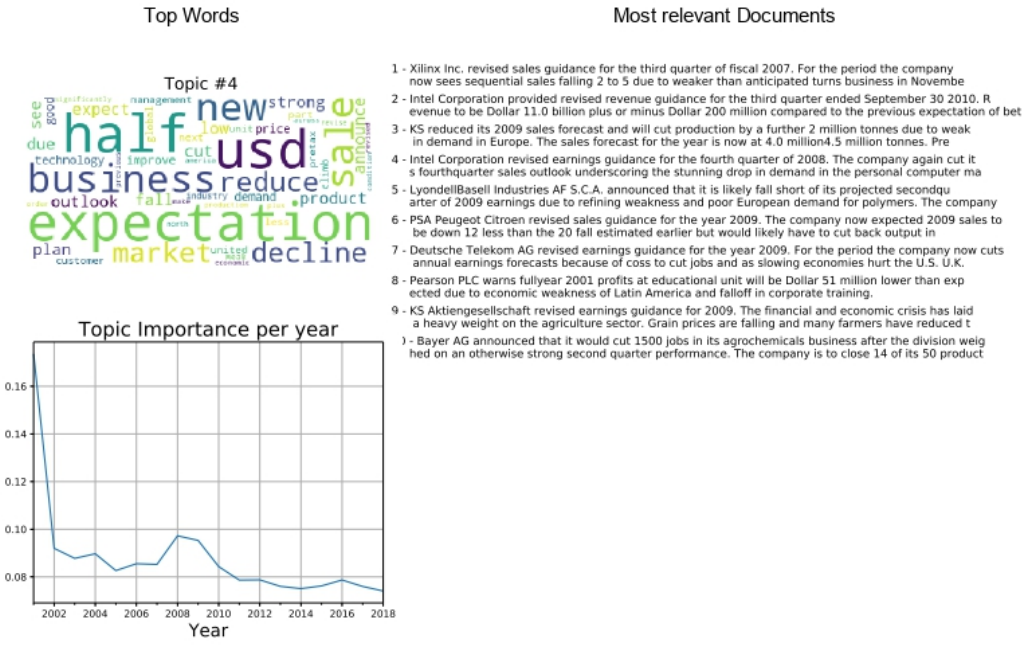


Figure A1.6: Results of the LDA model for Topic 5

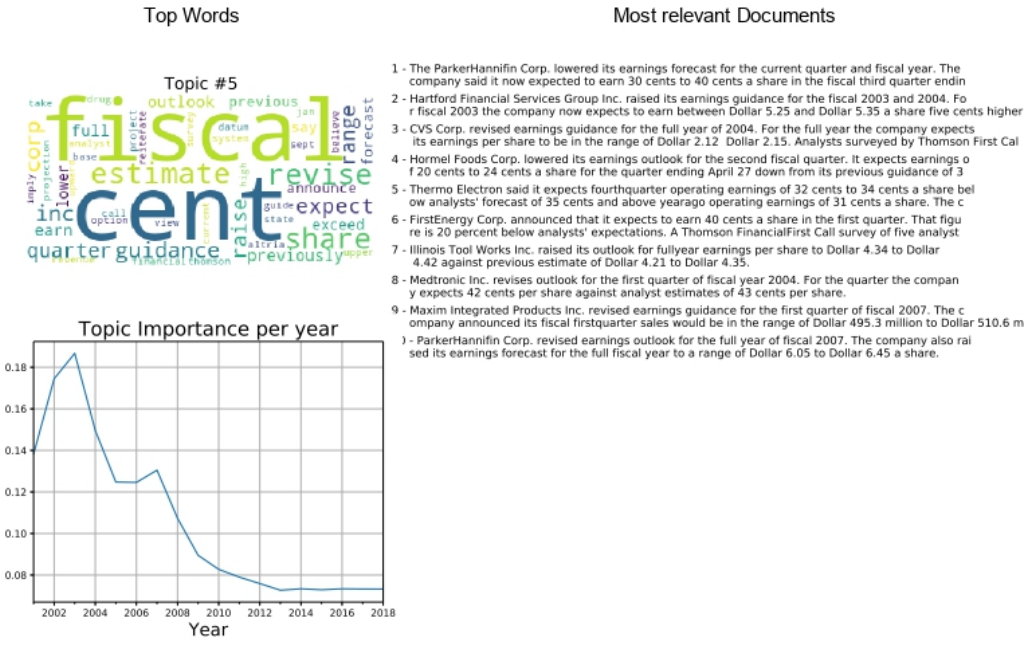


Figure A1.7: Results of the LDA model for Topic 6

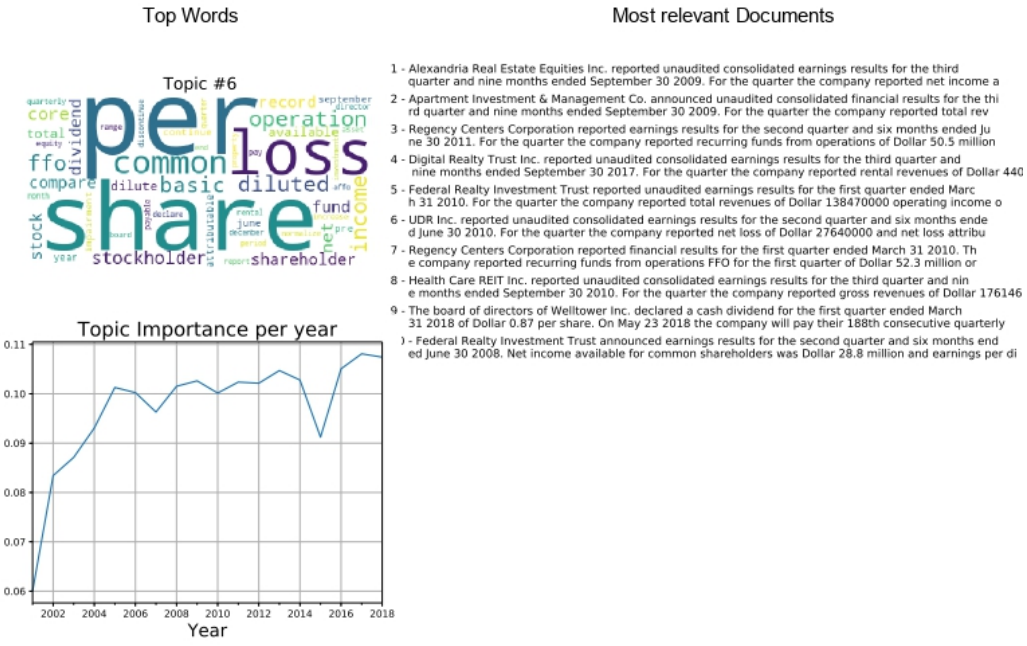


Figure A1.8: Results of the LDA model for Topic 7

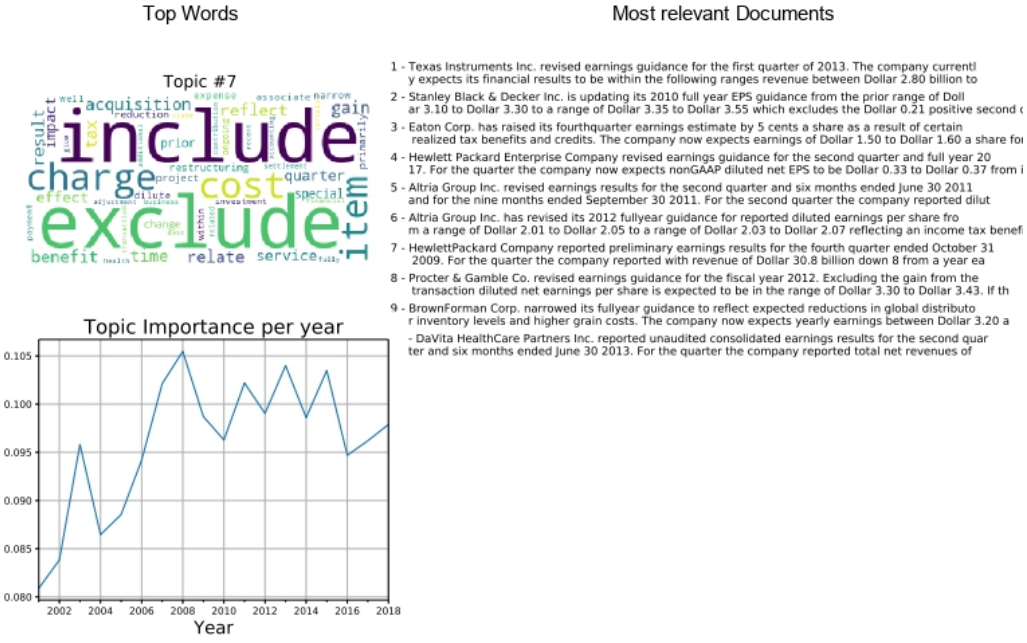


Figure A1.9: Results of the LDA model for Topic 8

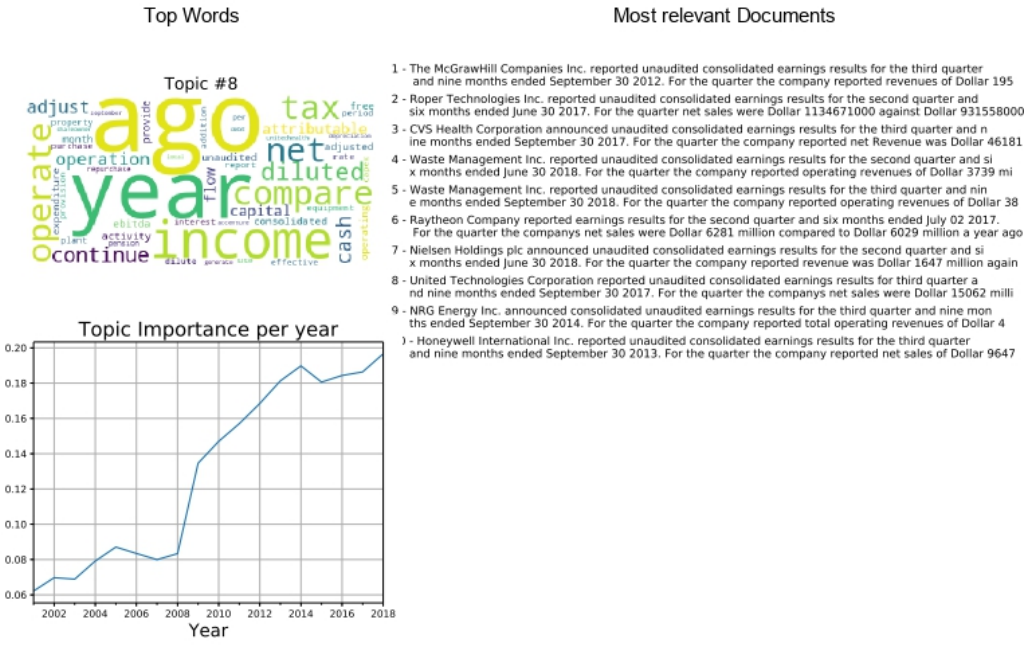
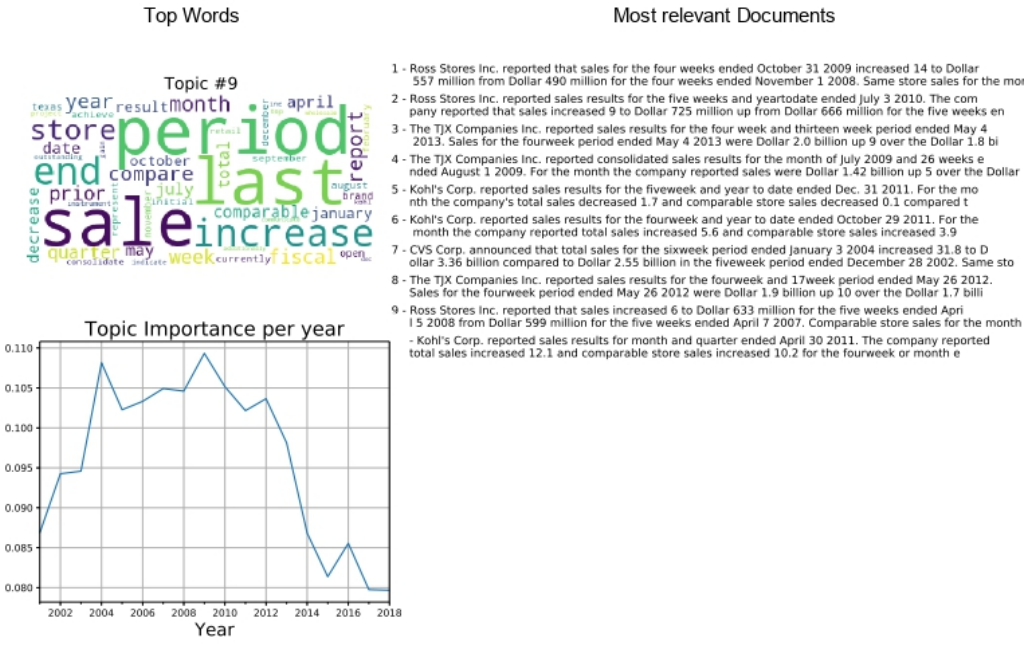


Figure A1.10: Results of the LDA model for Topic 9



Appendix III: Relationship between topics and message types

An alternative view of the topics is provided by the message types specified in the Capital IQ dataset for each individual document. Based on the (weighted) assignment of the documents to the topics, corresponding weights can also be determined at the topics level. Table A3.1 provides an overview of these weights for all message types. In table A1, the message type with the highest relative weight for each topic is shown as "Best Matching Type". For example, if we look at topic 2 - "profit/forecast", the associated

Table A3.1: Relationship between topics and message types (message type with the highest highest probability for the respective topic)

Topic	Best Matching Type
Topic 0	Debt Financing Related
Topic 1	Business Reorganizations
Topic 2	Address Changes
Topic 3	Buyback Transaction Announcements
Topic 4	Announcement of Interim Management Statement
Topic 5	Debt Financing Related
Topic 6	Executive Changes – CFO
Topic 7	Address Changes
Topic 8	Buyback - Change in Plan Terms
Topic 9	Dividend Decreases

documents are most frequently assigned to the message type "Address Changes", which initially appears somewhat counter intuitive when looking at the Word Cloud in Figure 4, since words relating to this do not have a high prevalence. However, the reverse assignment, shown in table 2, seems more relevant anyway. Here, for each of the 26 document types, the topics are shown which have the highest probability for the documents with the respective message type.

Table A3.2: Relationship between message types and topics (topic with highest probability for the respective message type)

Type	Best Matching topic
Lawsuits & Legal Issues	Topic 0
Business Reorganizations	Topic 1
Announcement of Operating Results, Address Changes, Buyback Tranche Update, Product-Related Announcements	Topic 2
Announcement of Interim Management Statement, Debt Financing Related	Topic 5
Executive Changes – CFO, Preferred Dividend Announcements of Earnings, Announcements of Sales/ Trading Statements, Business Expansions, Buyback - Change in Plan Term, Buyback Transaction Announcements, Corporate Guidance – Lowered, Corporate Guidance - New/Confirmed, Corporate Guidance – Raised, Dividend Increases, Impairments/Write Offs, Restatements of Operating Results, Special Dividend Announced, Corporate Guidance - Unusual Events, Discontinued Operations/Downsizings, Executive/Board Changes - Other	Topic 6
Dividend Affirmations, Dividend Decreases	Topic 8
	Topic 9

Appendix IV: Robustness and sensitivity analysis Regressions

1. Alternative modeling

Since the estimated coefficients for the index were close to one, as an alternative to model (1), a simplified relationship for the abnormal returns $\Delta S_{i,j} = \Delta P_{i,j} - \Delta I_{i,1}$ was considered, in which this coefficient is restricted to one:

$$\Delta S_{i,j} = \beta_1 + \beta_2 t_{0,t,i} + \beta_3 t_{1,t,i} + \cdots + \beta_{10} t_{9,t,i} \tag{2}$$

Table A4.1 summarizes the results of estimating model (2) with$\Delta S_{i,1}$ as the dependent variable: Compared with the results for model (1), there are no major changes in the coefficients, their sign and significance level. Thus, the overall results are largely robust with respect to the choice of event window size and the modeling of index returns. We still find significant positive effects on the (excess) return for the weights of topic 1 and topic 2 in the respective announcement text, while the weights of topic 4 and topic 5 show a negative impact. For the other topics there are no statistically significant effects, with the coefficients for the weights of topics 0 and 7 being estimated negatively and for those of topics 8 and 9 positively. This is also true for the influence of Topic 6, which is implicitly reflected in the constant. Again, on the basis of the F-test performed, the null hypothesis that the topic weights have no relevant explanatory power can be clearly rejected - despite the modest value of the coefficient of determination of about 0.01.

Table A4: Coefficients and standard deviations for model (2)

	$t_{-1} \to t_1$	$t_{-2} \to t_2$	$t_{-3} \to t_3$	$t_{-4} \to t_4$	$t_{-5} \to t_5$
Constant	-0.006	0.013	0.003	0.004	-0.001
	(0.015)	(0.017)	(0.018)	(0.019)	(0.019)
Topic 0	-0.020	-0.034	-0.015	-0.008	-0.023
	(0.027)	(0.030)	(0.032)	(0.033)	(0.034)
Topic 1	0.059*	0.050	0.084**	0.081**	0.083**
	(0.030)	(0.034)	(0.036)	(0.037)	(0.038)
Topic 2	0.048*	0.043	0.065**	0.064**	0.076**
	(0.025)	(0.027)	(0.029)	(0.030)	(0.031)
Topic 3	0.032	-0.022	-0.016	-0.019	-0.012
	(0.031)	(0.034)	(0.037)	(0.038)	(0.039)
Topic 4	-0.085***	-0.124***	-0.139***	-0.126***	-0.122***
	(0.032)	(0.035)	(0.038)	(0.039)	(0.040)
Topic 5	-0.032	-0.068***	-0.067**	-0.077***	-0.063**
	(0.024)	(0.026)	(0.028)	(0.029)	(0.030)
Topic 7	-0.020	-0.025	-0.027	-0.014	-0.023
	(0.029)	(0.032)	(0.034)	(0.036)	(0.037)
Topic 8	0.024	0.010	0.028	0.015	0.029
	(0.022)	(0.024)	(0.025)	(0.026)	(0.027)
Topic 9	0.033	0.019	0.028	0.026	0.045
	(0.024)	(0.027)	(0.029)	(0.030)	(0.030)
Stats					
No. Obs.	3605	3472	3447	3462	3491
R^2	0.01	0.01	0.01	0.01	0.01
DW	1.94	1.88	1.89	1.91	1.89
F-Stat	3.71***	4.62***	5.70***	4.56***	4.77***

To better assess the magnitude of the estimated parameters, we consider the coefficient of Topic 4 ("expectation/new") of -0.139 in model (2) for the time window ± 3 days. If the weight for this topic in the relevant announcement is 0.1 higher[9], ceteris paribus, this results in a $-0.139 * 0.1 = 0.0139$ or 1.39 percentage point lower return. If a message thus contains many statements assigned to this topic, we can assume an above-average negative reaction of the associated stock return, and the magnitude of the effect is also economically relevant.

Appendix V: Robustness and sensitivity analysis topic modeling

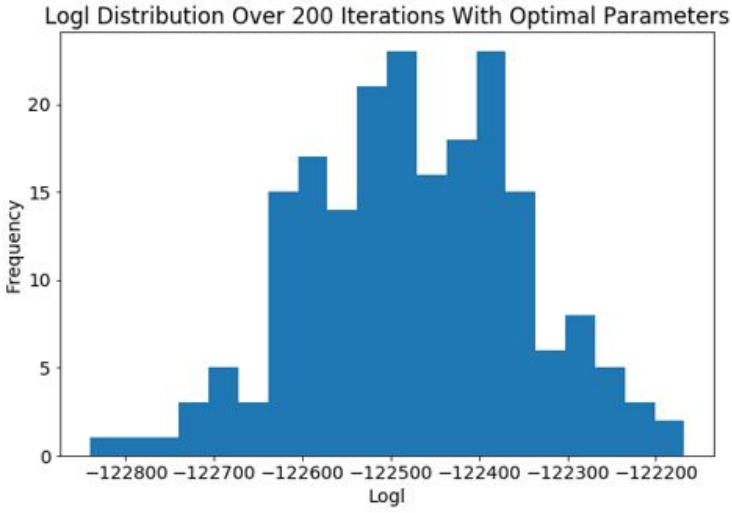
In this section, results on the robustness of the LDA model are provided, showing how sensitive the specific model is to changes in a parameter.

In total, the LDA model was estimated 200 times with the optimized parameters and

[9]Strictly speaking, the ceteris paribus statement is not valid here, since the weights of the topics add up to one i.e. the assumed increase in topic 4 leads inevitably to a reduction in other topics, which could reduce or increase the effect.

different "random seeds" (= starting point of a sequence of random numbers) with 100 iterations for each run to obtain an empirical distribution of likelihood values. Figure A5.1 presents the resulting distribution of the log-likelihood. It becomes clear that the

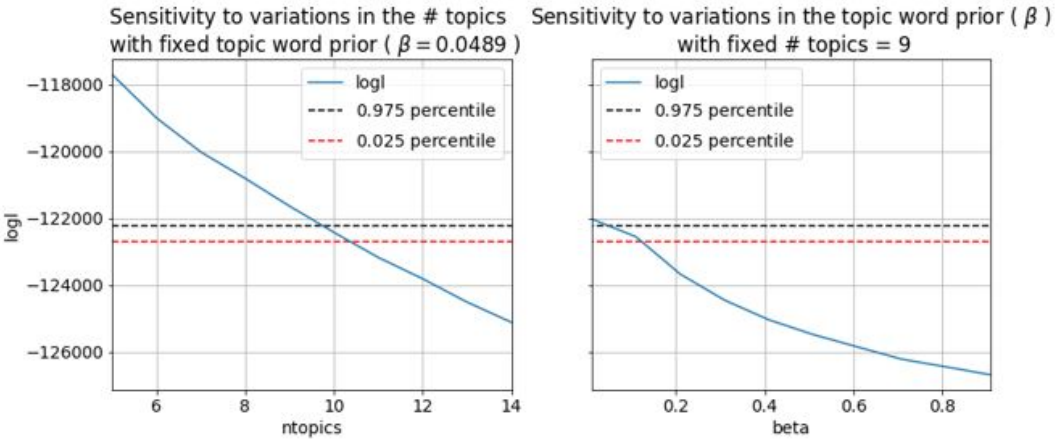
Figure A5.1: LDA distribution



results for the log-likelihood vary significantly between repeated runs of the optimization algorithm for the LDA model. The estimation of the model contains a not inconsiderable variance at the usual settings for the number of iterations (mostly even below 100), which results from the optimization algorithm. It is therefore possible that marginal changes in the model structure, such as alternative random seed values, lead to changes in the results. However, spot checks show that the basic structure of the topics and the assignment of the documents to these topics are only slightly affected.

Subsequently, the model was estimated with different parameter values (topic number or topic word prior β) than those of the optimized model and the results were compared with the estimated uncertainty range from the empirical distribution (95% interval). The sensitivity analysis of the log-likelihoods in Figure A5.2 shows how large the changes in the parameters must be in order to speak of a significant change in the log-likelihood. The upper and lower percentiles are represented by the red and black lines and show the range of variation of the model. Within this range is the log-likelihood at 200 model runs. On the left side, the number of topics varies with a fixed topic word prior, while on the right side the number of topics is kept constant and the prior takes on different values. The change in a parameter value is intended to test the change in the optimal model. If the value is not within the range, the alternative parameters differ significantly from the optimal parameters. The parameters quantified here as optimal and the resulting model are not robust for most expressions for beta or the topic number, since the value is not within the range in most cases. However, robust results are available for a small range around the optimized or selected values. For example, if we look at

Figure A5.2: Sensitivity analysis



beta, in addition to the value of 0.1 resulting from the optimization, smaller values such as 0.05 are still within the confidence interval.

Appendix VI: Original Publication (German)

Text-mining basierte Analyse der Kapitalmarktreaktionen auf Ad-hoc-Mitteilungen

Die Reaktion der Kapitalmärkte auf unvorhergesehene Geschäftsentwicklungen kann regelmäßig beobachtet werden, wenn Unternehmen gezwungen sind, Ad-hoc-Mitteilungen herauszugeben. Dieser Beitrag zeigt, welchen signifikanten Einfluss diese Mitteilungen auf den Unternehmenswert haben können. Dabei wurde analysiert, ob sich unterschiedliche Inhalte von Ad-hoc-Mitteilungen identifizieren lassen, die zu signifikanten Kurseffekten führen. Die Untersuchung zeigt dabei, dass bei einer durchschnittlichen Marktkapitalisierung im DAX (40 Mrd. Euro) von einem erwarteten Wertverlust durch die gestiegene Bedeutung des Themas "forecast" in einer negativen Konnotation von 158 Mio. Euro ausgegangen werden muss.

I. Einleitung

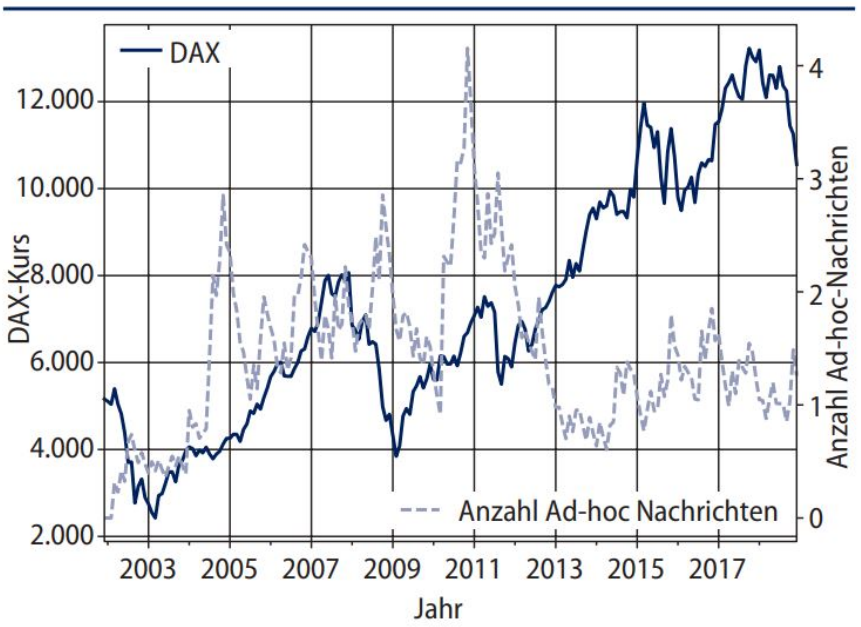
Gerade in Zeiten hoher Unsicherheit oder eines (drohenden) konjunkturellen Abschwungs nehmen Ad-hoc-Mitteilungen und damit verbundene Gewinnwarnungen durch am Kapitalmarkt gelistete Unternehmen zu. Diesen Zusammenhang zeigt Abb. 1 am Beispiel des DAX. Erkennbar ist die häufig gegenläufige Entwicklung von Aktienkursen und der Anzahl der Ad-hoc-Mitteilungen. Augenfällig ist dies insbesondere im Zuge der Finanzmarktkrise 2008/09 und in den unmittelbar folgenden Jahren. Das deutet darauf hin, dass Unternehmen häufig nicht ausreichend in der Lage sind, Informationen richtig und vorausschauend zu verarbeiten und die (Umsatz/Gewinn-)Entwicklung hinreichend genau zu antizipieren. Deshalb müssen falsche Einschätzungen korrigiert werden, was zu einer Ad-hoc-Mitteilung oder Gewinnwarnung führen kann. So stellen Baule und Tallau (2012) fest, dass Ad-hoc-Mitteilungen mit derselben Information wie reguläre periodische Berichte zu etwa doppelt so hohen Marktreaktionen führen. Zudem werden positive Nachrichten eher frühzeitig über eine Ad-hoc-Mitteilung bekannt gegeben, während negative Nachrichten mit einer Verzögerung publiziert werden.[1]

Ähnliche Effekte zeigen sich auch in der folgenden Analyse. Der Durchschnitt der absoluten Renditen für Beobachtungen, die zu Ad-hoc-Mitteilungen gehören, ist mit 4,5%

[1] Baule/Tallau, Market response to mandatory pre-earnings-announcements-evidence from ad-hoc disclosures in Germany, SSRN 1660679 2012, http://hbfm.link/7866 (Abruf: 03.03.2020). Weitere Analysen finden sich bei Guettler/Muntermann, Journal of International Financial Markets, Institutions and Money 2007 S. 1 ff. und Feuerriegel/Neumann/Ratku, Analysis of How Underlying Topics in Financial News Affect Stock Prices Using Latent Dirichlet Allocation, 49th Hawaii International Conference on System 2016, Sciences (HICSS), Koloa, S. 1072 ff.

mehr als doppelt so hoch als für die sonstigen Beobachtungen (2,15%), wobei dieser Unterschied hochgradig signifikant ist. Abb. 2 zeigt Histogramme der empirischen Verteilungen der Renditen für die Beobachtungen ohne (blau) und mit (braun) Ad-hoc-Mitteilung.

Abbildung 1: Verlauf des DAX (linke Skala) und Anzahl von Ad-hoc-Nachrichten pro Monat zu Aktien im DAX (exponentiell über 12 Monate geglättet)



Es bestätigt sich, dass die Renditen für Beobachtungen mit Ad-hoc-Mitteilung häufig größer ausfallen, wobei beide Verteilungen eine leichte Linksschiefe von ähnlicher Größenordnung aufweisen (-0,970 ohne Gewinnwarnung und -0,927 mit Gewinnwarnung). Diese Ergebnisse deuten bereits darauf hin, dass Ad-hoc-Mitteilungen Unsicherheit in Form von stärkeren Kursschwankungen am Markt auslösen.

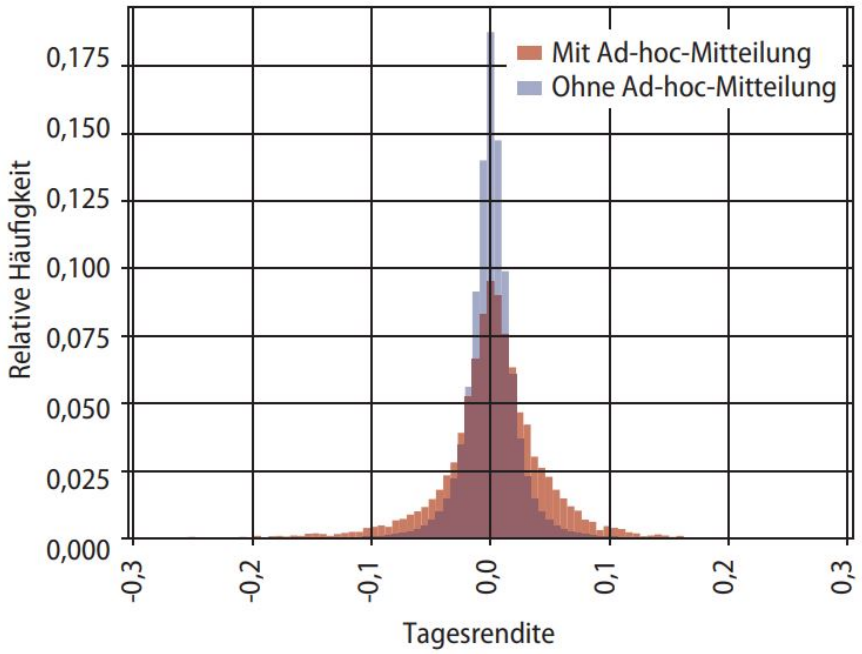
In dieser Studie soll nun überprüft werden, ob sich unterschiedliche Inhalte in Ad-hoc-Mitteilungen identifizieren lassen, die mit stärkeren Kurseffekten (vor allem im Falle negativer Nachrichten) in Verbindung gebracht werden können. Dabei wird untersucht, ob der Einsatz von aktuellen Verfahren des Textminings auf Texte von Ad-hoc-Mitteilungen zu inhaltlich gehaltvollen Topics (Themenfeldern) führt. Diese Topics werden auch mit exogen vorgegebenen Begriffsfeldern wie "uncertain" und "forecast"[2] in Beziehung gesetzt, um zu untersuchen, ob Topics im Kontext dieser Begriffsfelder zu einer signifikant negativen Kapitalmarktreaktion führen.

Diese Analyse ist vor allem vor dem Hintergrund der aktuellen Diskussion zum Einsatz von Machine Learning (ML) basierten Verfahren im klassischen Planungs- und Forecastingprozess – insbesondere bei kapitalmarktgelisteten Unternehmen – spannend.

[2]In Anlehnung an Baker/Bloom/Davis, The Quarterly Journal of Economics 2016 S. 1593 ff.

Manuelle Forecastingverfahren sind zeitintensiv und liefern häufig nur in Zeiten mit geringer Unsicherheit und in stabilen Branchen verlässliche Resultate im Vergleich zu ML basierten Verfahren.[3] Eine Ursache dafür liegt oftmals darin, dass weniger datengetriebene Ansätze wie das Bauchgefühl“ oder unternehmensinterne ”politische” Entscheidungen die Forecasting-Ergebnisse beeinflussen.[4] Neben dem Qualitätsgewinn durch Objektivierung der Forecast-Generierung sowie signifikanter Verbreiterung der in Betracht gezogenen Daten liefern ML-Verfahren einen starken Effizienzgewinn sowie einen höheren Steuerungsfokus, da die im Prognoseprozess eingesparte Zeit in Tätigkeiten mit mehr Steuerungsrelevanz (z.B. Szenarioanalysen) investiert werden kann. Darüber hinaus stärkt eine hohe Forecast-Güte das Vertrauen der Investoren in ein Unternehmen und macht dieses dadurch weniger anfällig für kurzfristige Kursschwankungen. Um diese Hypothese noch weiter zu untermauern, untersucht diese Studie vor allem auch den Kapitalmarkteinfluss von Gewinnwarnungen zu Topics im Begriffsfeld ”forecast”.

Abb. 2: Verteilungen der Renditen für Beobachtungen ohne (blau) und mit (braun) Ad-hoc-Mitteilungen



II. Datengrundlage und Aufbereitung der Textdaten

Der Analyse liegen Kursdaten für alle zum Zeitpunkt 27.05.2019 in den entsprechenden Indizes enthaltenen Aktien für DAX, MDAX, SDAX, FTSE 100, CAC40 und SP 500 von

[3] Vgl. PricewaterhouseCoopers, Digital Controlling, 2015, http://hbfm.link/7799 (Abruf: 21.03.2020).

[4] Vgl. PricewaterhouseCoopers, Financial planning: Realizing the value of budgeting and forecasting, 2011, S.1 ff.

03.01.2000 bis 31.12.2018 vor, zu denen es im betrachteten Zeitraum mindestens eine Ad-hoc-Mitteilung gab.[5] Ferner liegen die Ad-hoc-Mitteilungen zu all diesen Aktien für den Zeitraum vom 21.12.2001 bis zum 31.12.2018 vor. Aus den Kursdaten werden für die weitere Analyse stetige Tagesrenditen als Differenz der logarithmierten Preise berechnet. Als Textdaten liegen außerdem insgesamt 7.743 Adhoc-Mitteilungen vor. In dem Datensatz befinden sich insgesamt 2.707.629 Beobachtungen täglicher Renditen für Aktien, zu denen es zum jeweiligen Zeitpunkt keine Ad-hoc-Mitteilung gab (Beobachtungen ohne Ad-hoc-Mitteilung) und 7.455 tägliche Renditen, für die zum jeweiligen Zeitpunkt eine Ad-hoc-Mitteilung vorlag (Beobachtung mit Ad-hoc-Mitteilung).

Vor dem Einsatz statistischer Verfahren zur Textanalyse wie des hier eingesetzten Latent Dirichlet Allocation (LDA) Topic Modells (siehe dazu auch Kapitel III.2) ist eine Aufbereitung der Textdaten notwendig. Dafür werden zunächst die üblichen in der Literatur vorgeschlagenen[6] Techniken angewandt: Im ersten Schritt werden Satzzeichen und Währungssymbole, zusätzliche Leerzeichen und Zahlen entfernt. Außerdem werden sog. "Stopwords" ausgesondert. Es handelt sich dabei um 532 Wörter aus einem Wörterbuch (englisch), die sehr häufig auftreten und in der Regel keine inhaltliche Relevanz aufweisen, z.B. "which", "and" oder "but". Schließlich werden auch alle Wörter mit maximal zwei Buchstaben gelöscht. Tabelle 1 zeigt die Veränderungen durch diese 1. Stufe der Datenaufbereitung an zwei Beispielen in der zweiten Spalte. Diese erste Stufe des Preprocessing reduziert die Anzahl aller vorkommenden Wörter und damit die Dimension des Wortraums, während gleichzeitig erwartet wird, dass nur ein geringfügiger Informationsverlust eintritt.

In einer zweiten Stufe des Preprocessing werden Worte entfernt, die entweder nur selten auftauchen (hier in weniger als 5 Dokumenten) oder die in einem sehr großen Anteil aller Texte enthalten sind (hier in mehr als 75% der untersuchten Dokumente). Die Begründung hierfür ist, dass aus extrem selten benutzten Wörtern keine Topics abgeleitet werden können, die einen bestimmten Typ von Mitteilungen charakterisieren, der zumindest mit einer gewissen Mindesthäufigkeit auftaucht. Sehr häufig benutzte Worte wiederum würden in den meisten identifizierten Topics enthalten sein und damit keinen Beitrag zur Differenzierung der Topics leisten. Dieser zweite Schritt wird häufig als Entfernung von domain-spezifischen Stopwords angesehen, welche mit generellen Stopword-Listen, wie in Schritt 1 beschrieben, nur schwierig zu erfassen sind. Die nach

[5] Die Daten wurden auf Basis der Capital IQ-Datenbank bezogen

[6] Vgl. Bspw: Lenz/Winker, Measuring the diffusion of innovations with paragraph vector topic models, PLOS ONE 15(1): e0226685, http://hbfm.link/7867 (Abruf: 10.09.2020); Hansen/McMahon/Prat, The Quarterly Journal of Economics 2018 S. 801 ff.; Larsen/Thorsrud, Journal of Econometrics 2019 S. 203 ff.; Lüdering/Winker, Journal of Economics and Statistics 2016 S. 483 ff.; Hansen/McMahon, Journal of International Economics. 2016 S. 114 ff

Tabelle 1: Beispiele für Preprocessing

Ursprüngliche Ad-hoc-Mitteilung	Nach preprocessing (1. Stufe)	Nach preprocessing (2. Stufe)
PSA Peugeot Citroen provided earnings guidance for full year 2006. The company cut its annual guidance, saying it no longer expects second-half operating profit margin to match the 2.4% recorded in the first half.	psa peugeot citroen provide earning guidance year company cut annual guidance say longer expect half operating profit margin match record half	provide guidance cut annual guidance say longer half operating profit margin match record half
Rio Tinto Group had its 2004 and 2005 earnings estimates raised by Morgan Stanley on expectations profit from its diamond unit will be higher than previously forecast. The 2004 estimate was raised 1.5 percent to USD 1.75 a share, and the 2005 forecast 1.2 percent to USD 1.93 a share.	rio tinto group earning estimate raise morgan stanley expectation profit diamond unit higher previously forecast estimate raise percent share forecast percent share	group estimate raise expectation profit unit higher previously forecast estimate raise percent forecast percent

beiden Stufen des Preprocessing resultierenden Texte werden für die Beispiele in der dritten Spalte von Tab. 1 ausgewiesen. Nach den beiden beschriebenen Aufbereitungsstufen verbleiben insgesamt 1.975 Wörter für die weitere Analyse.

III. Topic Modelling

1. Textdatenaufbereitung

Für die Schätzung der Topics wurden zunächst nur 3.846 zufällig ausgewählte Dokumente benutzt (also 50% der verfügbaren Dokumente), um zu einem späteren Zeitpunkt die abgeleiteten Aussagen an einem Testdatensatz überprüfen zu können (externe Validierung). Anderenfalls besteht durch die vielen Auswahlschritte in der Aufbereitung und Auswertung der Textdaten ein erhebliches Risiko des Overfitting und einer durch die Methode erzeugten Verzerrung.[7]

[7] Vgl. Egami/Fong/Grimmer/Roberts/Stewart, How to make causal inferences using texts, arXiv:1802.02163, 2018, http://hbfm.link/7868 (Abruf: 03.03.2020).

2. Umsetzung und Ergebnisse des LDA-Modells & Topic Interpretation

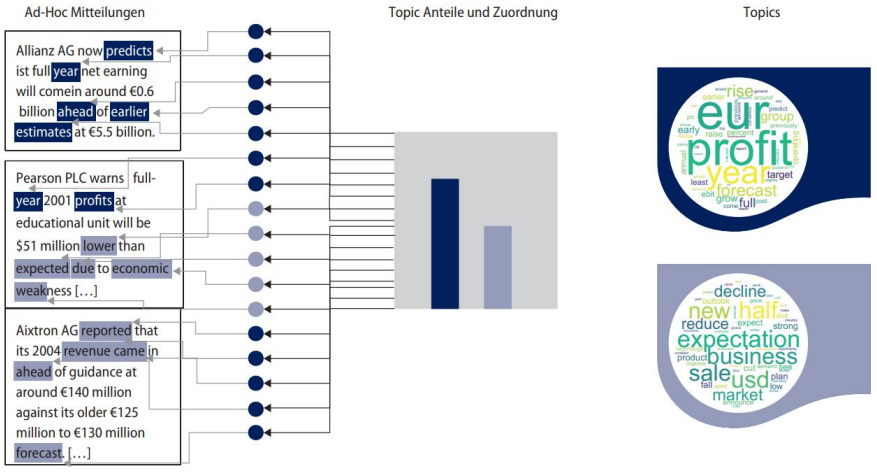
a) LDA-Modell und Word Cloud

In diesem Abschnitt werden das Verfahren und die Ergebnisse des LDA-Modells vorgestellt. LDA ist ein generatives probabilistisches Modell für Sammlungen diskreter Datensätze wie bspw. Textkorpora. Die Verwendung von LDA zur Extraktion von Themen hat verschiedene Vorteile: Erstens werden subjektive Verzerrung bei manueller Themenextraktion vermieden, zweitens ermöglicht es eine größere Flexibilität bei der Themenauswahl, um unserem Korpus (Text) besser gerecht zu werden. Darüber hinaus ermöglicht es der verwendete Data-Mining-Ansatz, Tausende von Dokumenten zu verarbeiten und so ein umfassendes Bild der Ad-hoc-Mitteilungen zu erhalten. Ein solches Unterfangen wäre mit einer manuellen Etikettierung prohibitiv schwierig und kostspielig.

Wir verwenden das in der Literatur etablierte LDA-Modell[8], um Topics zu identifizieren. Das dem LDA-Modell zugrundeliegende Verfahren basiert auf der Vorstellung, dass sich 1) Topics durch die relative Häufigkeit der darin vorkommenden Worte, und 2) Texte wiederum durch die relative Relevanz der darin auftauchen den Topics beschreiben lassen. Speziell wird jedes Topic als eine Wahrscheinlichkeitsverteilung über alle "möglichen", d.h. nach der Datenvorbereitung verbleibenden, Wörter dargestellt, während jedes Dokument als eine Wahrscheinlichkeitsverteilung über alle "möglichen", d.h. während der Schätzung des Modells identifizierten, Topics dargestellt wird. Vereinfacht ausgedrückt werden im LDA Modell Texte als eine Verteilung über Topics dargestellt und Topics als eine Verteilung über Wörter.

Abbildung 3[9] skizziert diesen Gedanken an einem Beispiel: Die Verteilung der Themen-

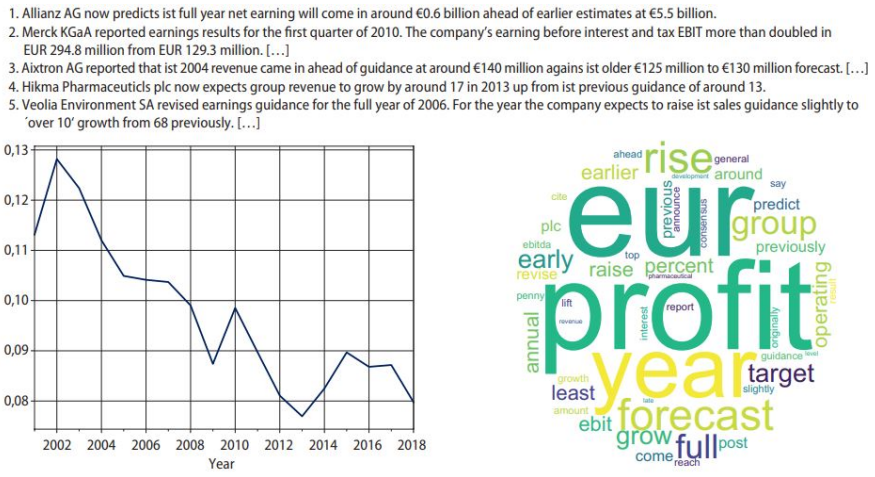
Abb. 3: Generativer Prozess der Texte aus Topic nach Blei (2012)



[8]Blei/Ng/Jordan, Journal of machine Learning research 03.01.2003 S. 993 ff

felder über die Texte, und der Wörter über die Themenfelder, ist für die Topics "forecast" (dunkelblau) und "uncertainty" (hellblau) visualisiert. Die Themenfelder werden durch die darin besonders relevanten (= häufiges Auftreten relativ zu anderen Themenfeldern) Wörter charakterisiert. Die Herausforderung bei der Schätzung des Modells besteht darin, allein auf Basis der vorliegenden Texte simultan die Topics und deren Gewichtung in den Texten zu bestimmen. Dies erfolgt durch ein zweistufiges, approximatives Optimierungsverfahren mit dem Ziel, auf Basis einer möglichst überschaubaren, vorher festgelegten Anzahl von Topics eine möglichst gute Beschreibung der Texte zu erreichen. Zusammenfassend wurden 10 Topics identifiziert.[10] Abb. 4 zeigt beispielhaft eine zusammenfassende Darstellung für Topic 2[11] Um ein Topic zu interpretieren, wird in der Regel

Abb. 4: Ergebnisse des LDA-Modells für Topic 2 – "profit/forecast"



eine Liste der wichtigsten 3-30 Wörter des Topics betrachtet und auf Basis der Wörter eine Vorstellung vom zugrundeliegenden abstrakten Topic gewonnen. Eine weitere Form der Darstellung von Topics findet via Wordclouds statt. Oben links in Abb. 4 befindet sich eine Wordcloud, welche die für das Topic relevantesten Wörter ("Top Words") enthält. Die Größe eines Wortes entspricht der Relevanz des Wortes im Topic. Bei Betrachtung der Wordcloud konnte Topic 2 bspw. als "profit/forecast"-Topic identifiziert werden. Für die praktische Implementierung des LDA-Modells sind einige Parameterwerte festzulegen. Wir optimieren die Hyperparameter des LDA Modells mithilfe einer Random Search. Dabei wird aus einer vorher festgelegten Spannweite von möglichen Hyperparameterwerten zufällig gezogen. Es wird das Set von Hyperparametern gewählt, welches das gewählte Informationskriterium minimiert. Diese Metaoptimierung ergab für die vorliegenden Daten Werte für den doc-topic prior alpha von 0.1, für den topic-word prior beta von 0.5, und einen learning decay von 0.7. Dabei wurde angesichts

[9] Vgl. Blei, Communications of the ACM 2012 S. 77 ff.

[10] Wir benennen Topics von 0-9, d.h. das erste Topic wird mit "0" indiziert.

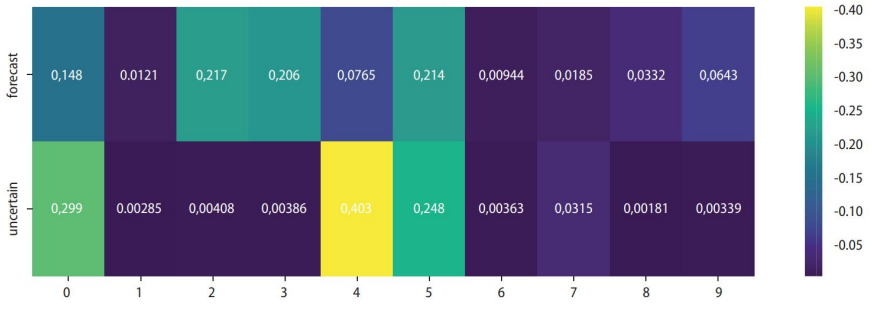
[11] Entsprechende Zusammenfassungen für die übrigen Topics können im Anhang in Tab. A3 gefunden werden. Zusätzlich findet sich im Anhang unter II. eine Darstellung des Zusammenhangs zwischen Topics und Nachrichtentypen. Der Anhang ist online abrufbar unter: CF1343653.

der sehr spezifischen Textart davon ausgegangen, dass eine Anzahl von zehn Topics zur Beschreibung ausreichen sollte.

b) Schlüsselbegriffe (Termsets) und Topics

Im Zuge einer weiteren, separaten Analyse wurden zwei Termsets in Anlehnung an den Ansatz von Baker et. al (2016)[12] gebildet. Bei einem Termsets handelt es sich um eine Liste von Wörtern, die eine hohe semantische Ähnlichkeit zu einem bestimmten Begriff aufweisen, wie bspw. das Wort "prediction" zum Wort "forecast". Wir verwenden Termsets, um die inhaltliche Interpretation der gefundenen Themenfelder zu vereinfachen, da die von LDA-Modellen generierten Topics nicht immer einfach zu interpretieren sind[13]. Zur Bestimmung dieser Termsets zu den Begriffen "forecast" und "uncertain" wurde auf eine externe Datenquelle zurückgegriffen, die vorgefertigte Modelle und Korpora für die computergestützte Verarbeitung natürlicher Sprache beinhaltet.[14] Mit der Word2Vec-Methode[15] werden Wörter identifiziert, die eine große semantische Ähnlichkeit zueinander aufweisen. Durch den Rückgriff auf diese Methode werden die Termsets unabhängig von den bestimmten Topics und den Dokumenten im Datensatz bestimmt. Die Liste der semantisch ähnlichsten Worte für "forecast" und "uncertain", die auch im zu analysierenden Datensatz vorkommen, werden als Termsets "forecast" und "uncertain" im Anhang, Tab. A2.1, ausgewiesen. Abb. 5 zeigt die Relevanz der

Abb. 5: Relevanz der Termsets je Topic



beiden Termsets für die jeweiligen Topics. Die Zahlen geben dabei an, wie häufig Worte aus den Termsets in den Dokumenten vorkommen, die den Topics zugeordnet werden. Dabei werden auch die Gewichte der Topics für die jeweiligen Dokumente berücksichtigt.

[12] Baker/Bloom/Davis, The Quarterly Journal of Economics 2016 S. 1593 ff.

[13] Vgl. Baldwin/Lau/Aletras/Sorodoc, Multimodal Topic Labelling. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 2: Short Papers, S. 701 ff., abrufbar unter: http://hbfm.link/7869 (Abruf: 10.09.2020).

[14] RaRe-Technologies, Data repository for pretrained NLP models and NLP corpora – Gensim Data 2018, GitHub repository, http://hbfm.link/7870 (Abruf: 03.03.2020).

[15] Mikolov/Sutskever/Chen/Corrado/Dean, Advances in neural information processing systems 2013 S. 3111 ff.

So weist die obere Zeile in Abb. 5 aus, dass die Worte aus dem Termset "forecast" relativ häufig in Dokumenten mit einem hohen Gewicht der Topics 2, 3 und 5 auftauchen, die jew. mehr als 20% der Gesamthäufigkeit abdecken. Für das Termset "uncertain" (untere Zeile in Abb. 5) findet sich eine noch deutlichere Konzentration für die Topics 0, 4 und 5.[16]

IV. Analyse des Zusammenhangs zwischen Topics und Aktienkursentwicklung

1. Eventregressionen mit symmetrischen Effekten

Für diesen ersten Ansatz wurde angenommen, dass der Einfluss der Topicgewichte für positive und negative Kursveränderungen symmetrisch ist. Es wurden fünf Schätzungen mit einer Breite des Eventfensters von $j = 1, \ldots, 5$ Tagen durchgeführt. Das geschätzte Modell lautet:

$$\Delta P_{i,j} = \beta_1 + \beta_2 \Delta I_{i,j} + \beta_3 t_{0,t,i} + \beta_4 t_{1,t,i} + \cdots + \beta_{11} t_{9,t,i}, \tag{3}$$

wobei $\Delta P_{i,j}$ die log-Rendite des zur Mitteilung i gehörenden Aktienkurses im Eventfenster bezeichnet. $\Delta I_{i,j}$ ist die log-Rendite des zugehörigen Marktindex. Zudem stehen $t_{0,t,i}, \ldots, t_{9,t,i}$ für die Topicgewichte der betrachteten Mitteilung.[17] Die geschätzten Koeffizienten mit zugehörigen Standardabweichungen (in Klammern) werden in Tab. 2 dargestellt. Wie erwartet zeigt sich ein starker Zusammenhang zwischen den Renditen einzelner Aktien und des zugehörigen Index mit geschätzten Koeffizienten in der Größenordnung von 0,95 – 1,16. Außerdem zeigt sich ein deutlicher Anstieg des Bestimmtheitsmaßes R^2, wenn ein Eventfenster der Länge größer 2 gewählt wird. Dies mag der Tatsache geschuldet sein, dass bei sehr kurzen Betrachtungszeiträumen je nach Uhrzeit der Ankündigung vor- oder nachlaufende Effekte nicht im Eventfenster beinhaltet sind und damit nicht in die Analyse einfließen. Wir empfehlen daher, bei der Interpretation eher etwas längere Zeiträume, z.B. ± 3 Tage, zu betrachten.
Bei einem Vergleich des Bestimmtheitsmaßes für dieselben Regressionen ohne die Variablen für die Topicgewichte ergibt sich ein nur um ungefähr einen Prozentpunkt geringerer Wert. Dennoch ist der Einfluss der Topicgewichte relevant. Sowohl im Hinblick auf das Vorzeichen als auch bezüglich der Größe und Signifikanz der Koeffizienten ergeben

[16] Im Anhang finden sich weitere Robustheitsanalysen. Der Anhang ist online abrufbar unter: CF1343653

[17] Da sich die Topicgewichte jew. zu eins addieren, wurde auf die Aufnahme eines der Topicgewichte (für Topic 6) verzichtet, um das sonst entstehende Problem der perfekten Multikollinearität zu vermeiden. Dies ist bei der Interpretation der Koeffizienten zu beachten.

Tabelle 2: Koeffizienten und Standardabweichungen für Modell (1)

	$t_{-1} \to t_1$	$t_{-2} \to t_2$	$t_{-3} \to t_3$	$t_{-4} \to t_4$	$t_{-5} \to t_5$
Konstante	-0.00567	0.01250	0.00243	0.00387	-0.00062
	(0.015)	(0.017)	(0.018)	(0.019)	(0.019)
Index	0.96560***	1.06247***	1.15809***	1.06921***	1.09845***
	(0.061)	(0.046)	(0.043)	(0.041)	(0.039)
Topic 0	-0.01986	-0.03513	-0.01777	-0.00882	-0.02503
	(0.027)	(0.030)	(0.032)	(0.033)	(0.034)
Topic 1	0.05927*	0.04937	0.08147**	0.07966**	0.08114**
	(0.030)	(0.034)	(0.036)	(0.037)	(0.038)
Topic 2	0.04828*	0.04269	0.06279**	0.06361**	0.07716**
	(0.025)	(0.027)	(0.029)	(0.030)	(0.031)
Topic 3	0.03219	-0.02093	-0.01266	-0.01793	-0.00994
	(0.031)	(0.034)	(0.036)	(0.038)	(0.039)
Topic 4	-0.08492***	-0.12320***	-0.13666***	-0.12539***	-0.12305***
	(0.032)	(0.035)	(0.037)	(0.039)	(0.040)
Topic 5	-0.03212	-0.06699**	-0.06228**	-0.07528***	-0.06195**
	(0.024)	(0.026)	(0.028)	(0.029)	(0.030)
Topic 7	-0.01969	-0.02496	-0.02671	-0.01338	-0.02549
	(0.029)	(0.032)	(0.034)	(0.036)	(0.037)
Topic 8	0.02456	0.01019	0.02630	0.01407	0.02614
	(0.022)	(0.024)	(0.025)	(0.026)	(0.027)
Topic 9	0.03272	0.01977	0.02986	0.02635	0.04460
	(0.024)	(0.027)	(0.029)	(0.030)	(0.030)
Stats					
No. Obs.	3605	3472	3447	3462	3491
R^2	0.07	0.14	0.19	0.18	0.20
DW	1.94	1.88	1.89	1.91	1.89
F-Stat	3.72***	4.53***	5.37***	4.43***	4.70***

$t_{-1} \to t_1$ steht für das Zeitfenster von t-1 bis t+1, $t_{-2} \to t_2$ für das Zeitfenster von t-2 bis t+2, etc. Die Spalten stehen für die unterschiedlichen Eventfenster von ±1 to ±5 Tagen um den Zeitpunkt der Ad-hoc-Mitteilung herum. Neben den Koeffizienten für Indexreturn (INDEX) und Topicgewichte (TOPIC 0 ...) wird jew. durch die Symbole *, ** und *** angezeigt, ob die Nullhypothese eines Parameterwerts von null zum 10, 5 bzw. 1%-Niveau verworfen werden kann.

sich nur geringe Unterschiede bei Betrachtung unterschiedlicher Zeitfenster ab ± 3 Tage. Die Gewichte für Topic 1 und Topic 2 weisen demnach signifikant positive, die für Topic 4 und Topic 5 signifikant negative Effekte aus.

Ergänzend wurde mittels des F-Tests untersucht, ob die Nullhypothese, dass alle Topicgewichte gemeinsam keinen Effekt auf die modellierte Rendite haben, verworfen werden muss. Der Wert der relevanten F-Statistik ist jew. in der letzten Zeile ausgewiesen. Er liegt in allen Fällen deutlich über dem kritischen Wert für das 1%-Niveau, sodass die Nullhypothese verworfen wird. Es kann also von einem statistisch hoch signifikanten zusätzlichen Einfluss der Topicgewichte ausgegangen werden. Zusätzlich sei daran erinnert, dass die beiden Topics (4 und 5), für die konsistent und deutlich signifikant negative Effekte auf die Aktienrenditen gefunden werden, genau die beiden Topics sind, die auch

in Bezug auf das Termset "uncertain" die höchste Relevanz aufweisen. Dies kann als Bestätigung für die Hypothese gesehen werden, dass wahrgenommene Unsicherheit zu überproportionalen Kursverlusten führt.

2. Robustheit der Ergebnisse

In den bisher vorgestellten Ergebnissen beruht die Bestimmung der Topics und der Topicgewichte auf denselben Daten wie die im Nachgang durchgeführte Regression. Hier besteht zumindest theoretisch die Gefahr von Verzerrungen aufgrund des Einsatzes von berechneten Regressoren (hier die Topicgewichte). Ein einfacher Test auf das Vorliegen solcher Verzerrungen kann im vorliegenden Fall dadurch erfolgen, dass die Regression für die Beobachtungen durchgeführt wird, die nicht für die Auswahl und Schätzung des Topicmodells verwendet worden sind. Eine Wiederholung der Regressionen für diesen Testdatensatz zeigt qualitativ keine Unterschiede zu den hier ausgewiesenen Resultaten. Von daher ist nicht von großen Verzerrungen auszugehen, womit auch Hinweise auf eine externe Validität der Ergebnisse vorliegen.
Zwei weitere Varianten der Schätzung bestätigen die Robustheit der Ergebnisse. Dabei wurde zuerst ein vereinfachter Zusammenhang für die abnormalen Renditen $\Delta S_{i,j} = \Delta P_{i,j} - \Delta I_{i,1}$ betrachtet. Auch in dieser Variante werden die Ergebnisse für die Topicgewichte bestätigt.[18] Tab. A4 im Anhang gibt einen Überblick über die Ergebnisse der Schätzung mit $\Delta S_{i,1}$ als abhängiger Variable.
Weiterhin betrachten wir eine Variante, die zusätzlich das Vorzeichen des zugehörigen Index als Indikatorvariable einbezieht. Mit der Indikatorvariablen können Beobachtungen mit positiven Renditen bzw. positiven Überrenditen gesondert analysiert werden. Die Ergebnisse dieser Analyse finden sich im Anhang in Tab. A5. Insbesondere der negative Einfluss von Topic 4 "profit/ forecast" auf die Renditen bestätigt sich hier.

3. Weitergehende inhaltliche Interpretation der Ergebnisse

Um die Interpretation der gefundenen Effekte zu erleichtern, erfolgt abschließend noch einmal der Bezug zu den Termsets "forecast" und "uncertain".

Forecast

Mit relativen Gewichten von mindestens 15% ist das Termset "forecast" mit den dazugehörigen Begriffen für Topic 0, 2, 3 und 5 besonders relevant. Es zeigt sich, dass für den Einfluss der Gewichte dieser Topics unterschiedliche Vorzeichen geschätzt werden. Während das

[18] Die Analysen finden sich im Anhang unter IV., online abrufbar unter: CF1343653

Gewicht von Topic 2 positiv wirkt, weist ein hoher Anteil von Topic 5 eher auf eine negative Kursreaktion hin. Die Effekte für die Topics 0 und 3 sind hingegen nicht signifikant von null verschieden. Diese nicht eindeutigen Ergebnisse reflektieren, dass keine eindeutigen Erwartungen hinsichtlich des Vorzeichens bestehen, da die im Termset enthaltenen Worte sowohl negativ ("decline", "slowdown", "drop") als auch positiv ("optimistic", "increase") interpretiert werden können. Betrachtet man jew. die zehn Texte mit den höchsten Gewichtungen für die Topics, bei denen das forecast"-Termset besonders relevant ist, so ergibt sich in den meisten Fällen ein klarer Hinweis darauf, dass inkorrekte Vorhersagen zu den Mitteilungen führen. Dabei sind sowohl positive als auch negative Vorhersagekorrekturen Bestandteil dieser Mitteilungen. So steht z.B. im Text mit dem höchsten Gewicht für Topic 5: "The Parker-Hannifin Corp. lowered its earnings forecast for the current quarter and fiscal year.". Im Text mit dem höchsten Gewicht für Topic 2 findet sich "Allianz AG now predicts its full year net earning will come in around 0.6 billion Euro ahead of earlier estimates at 5.5 billion Euro." In jedem der zehn Texte mit den höchsten Gewichten für Topic 2 kann eine ökonomisch positiv zu wertende Aussage identifiziert werden, also z.B. eine nicht antizipierte Erhöhung der Gewinne. Dies ist konsistent zu den Ergebnissen aus der Schätzung, bei der in jeder Spezifikation für jedes Zeitfenster ein signifikanter positiver Koeffizient für das Gewicht von Topic 2 ausgewiesen wird. Eine pauschale Aussage bezüglich der Effekte der "forecast"-Topics ist daher nicht möglich, da sowohl Begriffe mit positiver als auch negativer Bedeutung im Termset vorhanden sind. Um die Größenordnung der Effekte abzuschätzen, gehen wir z.B. von einer Situation aus, in der sich das Gewicht für Topic 5 in einer Mitteilung um eine Standardabweichung (0,059) erhöht. Der Effekt für dieses Gewicht bei einem Zeitfenster von $\pm$ 3 Tagen beträgt nach Tab. 4 -0,067. Für eine Aktie mit durchschnittlicher Marktkapitalisierung im DAX (40 Mrd. Euro) ergäbe sich damit ein erwarteter Wertverlust durch die gestiegene Bedeutung des Themas "forecast" in einer negativen Konnotation wie in Topic 5 von -0,067*0,059*40 Mrd. Euro = 158 Mio. Euro. Die Größenordnung der gemessenen Effekte ist also durchaus beträchtlich.

b) Uncertain

Bezüglich des Termsets "uncertain" weisen die Topics 0, 4 und 5 die höchste Relevanz aus. Der Erwartung eines eher negativen Einflusses von "uncertain" entsprechend weisen die Gewichte für die Topics 4 und 5 auch negative Vorzeichen in den Schätzungen auf. Auch für diese Topics mit hohem Anteil von "uncertain" kann eine Abschätzung der ökonomischen Relevanz der gefundenen Effekte auf Basis der durchschnittlichen Marktkapitalisierung pro Aktie in den betrachteten Indizes erfolgen. Als Beispiel betrachten wir den Effekt für das Gewicht in Topic 4 aus Tab. 4. Dafür wird ein Koeffizient von

-0,39 für ein Eventfenster der Länge ± 3 Tage ausgewiesen. Die Standardabweichung für das Gewicht in der Stichprobe beträgt 0,041. Für eine Aktie mit einer durchschnittlichen Marktkapitalisierung von knapp 40 Mrd. Euro im DAX ergäbe sich demnach bei einer Mitteilung mit einem Gewicht für Topic 4, das um eine Standardabweichung (also 0,041) höher ausfällt, ein erwarteter Wertverlust von -0,139*0,041*40 Mrd. Euro = 228 Mio. Euro. Die Größenordnung der gemessenen Effekte ist also auch in diesem Fall durchaus erheblich.

V. Zusammenfassung

Zusammenfassend lässt sich festhalten, dass die Auswertung der Texte von Ad-hoc-Mitteilungen mit aktuellen Verfahren des Textmining zu inhaltlich gehaltvollen Topics (Themenfeldern) führt, die auch mit exogen vorgegebenen Begriffsfeldern wie "uncertain" und "forecast" in Beziehung gesetzt werden können. Die Gewichtung dieser Themenfelder in den Mitteilungen ist stark mit den resultierenden Kursveränderungen korreliert. Dieser Zusammenhang ist robust, da er auch mit einem Teildatensatz, der nicht für die Modellierung und Schätzung des Modells benutzt wurde, qualitativ reproduziert werden konnte.

Es zeigt sich, dass eine Erhöhung der Gewichte sensibler Topics in Bezug auf "uncertain" und "forecast" zu ökonomisch relevanten Veränderungen in den Renditen und damit – im Falle negativer Mitteilungen – auch erheblichen Wertverlusten führt. Diese belaufen sich für eine Aktie mit durchschnittlicher Marktkapitalisierung im DAX oder S&P durchaus auf dreistellige Millionenbeträge.

Im Umkehrschluss erscheint es plausibel, dass durch eine Verbesserung der Prognoseverfahren bspw. mit auf ML-basierenden Methoden ein Teil dieser punktuell auftretenden großen Wertverluste vermieden werden könnte. Neben dem Qualitätsgewinn – und einer damit verbundenen möglichen Reduktion von Unternehmenswertverlusten – durch Objektivierung der ForecastGenerierung liefern ML-Verfahren einen starken Effizienzgewinn sowie einen höheren Steuerungsfokus, da die im Planungsprozess eingesparte Zeit in Tätigkeiten mit Steuerungsrelevanz investiert werden kann. Darüber hinaus stärkt eine hohe Forecast-Güte das Vertrauen der Investoren in ein Unternehmen und macht dieses dadurch weniger anfällig für kurzfristige Kursschwankungen.

Chapter 3

Analysing Gender Inequality at the Firm Level

Chapter 3: Analysing Gender Inequality at the Firm Level

This chapter is based on the article:

Title:	Analysing Gender Inequality at the Firm Level
Authors:	**Dania Eugenidis**: Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany; Contribution: 90%. **Dr. David Lenz**: Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany; Istari.Ai, Mannheim, Germany **Dr. Jan Kinne**: Istari.Ai, Mannheim, Germany
Status:	Working Paper
Available at:	http://hdl.handle.net/10419/262315 https://youtu.be/-adg3C3Z2Ns

An earlier version of this work has been presented at the following scientific conference:

63rd ISI World Statistics Congress 2021, Virtual, June 2021 (Presentation).

Analysing Gender Inequality at the Firm Level

Abstract

The role of gender in the labour force and potential inequalities between men and women have been widely discussed. Despite efforts to align gender roles in recent decades, high levels of inequality are not an exception but rather the standard. These inequalities can lead to the respective minorities' general dissatisfaction, which affects the working atmosphere and ultimately a firm's economic success (Hoogendoorn et al. 2013).
Recent quantitative studies confirm this dissatisfaction exists. However, analyses only take place at a country or regional level. Therefore, conclusions can be drawn on an aggregated level, whereas underlying structural differences between individual firms remain undetected. Alternative ways to measure inequalities include qualitative studies for individual companies. However, no generalized inference can be made.
Our proposed framework, the Gender Equality Firm Index (GEFI), allows for quantitative gender equality analysis at the company level. GEFI aims to explore the latent and the concrete implementation of gender equality in firms. Specifically, we derive firm-level measurements from large-scale data extraction of firm websites and combine them with official data. We consequently derive a gender equality score for each company, making it possible to draw conclusions at any given level of granularity. We demonstrate the applicability of our framework in a case study including nearly 1 million firms throughout Germany. Thereby, we find that mainly urban and western German firms in sectors such as health and social services comparably enforce gender equality the most, which is in line with the existing literature.

1 Introduction

Gender inequality plays an important role in many areas of society, such as working life. Equally qualified individuals may be offered different career opportunities and remuneration, depending on sociodemographic characteristics such as gender. According to official statistics, differences in income of an average of nearly 20% for the same work still existed between men and women in Germany in 2020 (Destatis 2021a). The distribution of the genders is also still strongly asymmetrical with regard to career opportunities, which is expressed, for example, by the lack of women on supervisory boards, leading to an executive glass ceiling for women (Kerevel 2019; Brader/Lewerenz 2006).

In addition to these scientifically quantified imbalances, differences also exist at latent

dimensions: Lacking efforts and measures to ensure equality among employees might lead to the (unintentional) discrimination of certain sociodemographic groups, such as women, irrespective of the qualification.
This not only affects the respective individuals, but can also lead to economic consequences for a company. Various theoretical and empirical studies confirm a negative relationship between gender inequality and economic growth (Klasen/Lamanna 2009).

The problems discussed above arise, take place, and, all things being equal, can be resolved at the firm level through concrete policies and measures. However, targeted policies to lower inequality are hard to implement without knowing the current state of the ecosystem. Typically, data about inequalities suffer from one of two problems: Quantitative data (e.g., official statistics) is not available at the necessary levels of granularity, while qualitative data (e.g., interview or questionnaire responses) cannot be conducted at large scale. While it is possible to carry out qualitative studies at the company level, quantitative analyses have so far been conducted at more aggregated levels. As the former covers only a small fraction of businesses or institutes, drawing conclusions for a society as a whole and identifying general structural patterns is not possible. In addition, qualitative data tools, such as surveys, often face the problem of being too dependent on the researcher's objective and are therefore not generalizable. They are also expensive and time-consuming (i.e., results can be obsolete before they are published).

We address these issues by proposing a composite index to quantify the degree of gender inequality at the firm level. The Gender Equality Firm Index (GEFI) fully measures equality automatically, objectively, at a large scale, and with high granularity and frequency using the information on company websites in combination with data provided by the companies. The GEFI framework makes it possible to gain a deeper insight into the structures of a labour market and thus identify certain dimensions (e.g., sector, firm size, or region) that may exhibit a higher degree of gender inequalities.

We combine company data and the website data to calculate a value for each company for our composite indicator. This value quantifies and measures gender inequality and is additionally evaluated with regard to the various dimensions.

To demonstrate the applicability of the GEFI framework, we conduct a case study for Germany. Our data set comprises all firms that are located and registered in Germany and have a web presence. We also retrieved supplementary firm meta-data from the Mannheimer Unternehmenspanel (MUP) database. Our findings confirm that firms located in urban and west German areas have the highest degree of equality on average. In addition, an increase in firm size accompanies higher gender equality. Furthermore,

we identify strong differences between certain industries and sectors: The highest values are found in the health sector, whereas construction is far below the average. This is consistent with the results of previous studies that indicate women with STEM degrees are more likely to switch to other sectors, such as healthcare (Beede et al. 2011).

An advantage of the GEFI framework is its applicability to other countries and languages: The company data we use in our analysis contain standardized information and are also available for other studies.
The remainder of this paper is structured as follows: In Section 2, we review previous literature on gender indices. In Section 3, we detail the components and technical steps to construct the GEFI. In Section 4, we provide a case study for Germany, starting with the description of the data sources, which is followed by the distinction of the different dimensions (e.g., sectors and regions) used in the analysis. The results are presented in Section 5. In Section 6, we validate the GEFI results with external data sources, before concluding in Section 7.

2 Literature overview

Gender inequality at the firm level

Both quantitative indices and qualitative survey analyses aim to identify structural inequalities that can be seen as an outcome of the unequal treatment of one gender by the other. To complete the picture, discussing the consequences for the corresponding minorities is necessary.

One can safely assume that women may not be accepted and promoted in the same way as men in certain professions, which is expressed in a disproportionately low number of women in higher management and executives level in comparison to low/middle-level positions.
Timmers et al. (2010) address the problem of vertical gender segregation depending on the management level. They identify individual, cultural and structural reasons for the underrepresentation of women in higher positions when examining the effectiveness of 19 gender equality measures at Dutch universities. For this purpose, internal and external reports concerning gender policy have been consulted in addition to information obtained through interviews with academic staff.
Relating the gender distribution of low/mid-level employees to the upper management levels (here: full and associate professors), one obtains the "Glass Ceiling Index" (GCI). Lower scores imply a more balanced gender distribution. The scores in their sample are comparatively low in universities with high numbers of policies in relation to universities

without gender equality policies, which supports the notion that such measures are effective. Unfortunately, the measures have not been evaluated by the universities in any case retrospectively. According to Jansen et al. (2001), this would have been possible by looking at the percentage change of the GCI before and after the measure. At the beginning of the observation period, which was conducted from 2000 to 2007, a decrease in GCI was observed by the authors in total and in each of the universities with values in a range of 2.5 (2.2) to 16 (9.5) in 2000 (2007). The presented evidence confirms a positive correlation between the change in GCI and the measures enforced.
Thus, the authors point out that while measures are often announced and discussed, at least one measure per institution was delayed or never actually enforced. Intrinsic motivation does not seem to be present in these cases. Instead, a predetermined necessity for inclusion of such measures.

Brader and Lewerenz (2006) investigate 16,000 mainly small- and medium-sized German firms. Their study was conducted at different management levels to ensure a representative and differentiated result. A decreasing proportion of women in the top management levels could be observed, while the overall proportion of women in lower management positions in the company mostly remained constant across firms, leading to an actual increase in the GCI. This is particularly the case with increasing firm size. For example, at the top management level, only one in four managers was a woman. The authors show that small- to medium-sized companies are more likely to be run by women, whereas a shortage of female managers still exists in large companies. In companies with more than 500 employees, only 4% of women work in the upper management level. When looking at the range of tasks at different management levels, significant differences can be identified: The tasks in middle management or at the second management level have significantly lower responsibility. Regardless of professional qualifications, women are not reasonably trusted (Brader/Lewerenz 2006).

Other studies argue similarly and mention that females are disadvantaged in male-dominated industries and excluded from certain technical tasks. Kanter (1977) finds that working groups with a majority of men tend to treat women as representatives of their category, i.e. as “symbols” or “tokens” rather than individuals (Kanter 1977). Men are still considered the main players in STEM fields and are more likely to be called upon for technological and industrial developments in particular. Employers assume that women do not possess an equivalent level of technical understanding, which implies that outdated gender roles remain preserved and are still part of the work structure (Berglund/ Thorslund 2012).

These circumstances arise for example from discursively created social stereotypes that lead to underrepresentation of females at the top management level (Carli/Eagly 2007).

Timmers et al. (2010) add individual and institutional factors to the cultural factors. Psychologically, this is due to subconscious constructions and beliefs that are socially and politically determined, as well as role-specific characteristics of men and women. Men are usually associated with characteristics such as leadership and power, whereas women tend to be associated with weak and adaptive characteristics. Therefore, they are perceived as unsuitable for senior management positions because female stereotypes are usually not consistent with the qualities and skills of a successful manager (Glick-/Rudman 2012).

Another reason that exacerbates the problem is the fact that women often choose paths that lead away from their academic majors and toward other professional groups, despite having the appropriate qualifications. Bede et al. (2011) point out that female STEM majors are twice as likely to switch their academic or business specialization to education or healthcare, thus supplying labour below their qualification (Bede et al. 2011). The above-mentioned literature gives sufficient evidence that a need for alignment exists in some areas of the labour force.

Gender indices

A variety of studies identifies gender inequality in the labour force on different levels of aggregation and the resulting personal and economic downsides. Kabeer and Natali (2013) point out a causal impact of gender equality on economic growth. Thus, if policies are implemented with regard to the latter, this will not reduce inequalities. Instead, gender policies should be implemented, which then indirectly affect growth as well (Kabeer/Natali 2013). Typically, different countries are compared by constructing an index with a score for each observation. Bericat (2012) gives an overview of the existing literature and works out differences in concepts and data sources (e.g. the "Gender Equality Index" by Straus and Sugarman (1988), the "Social Institutions and Gender Index" by Branisa et al. (2009) or the "Global Gender Gap Index" (Bekhouche et al. 2012)). Van Staveren (2013) finds average correlations of 50-80% between the concepts. These discrepancies arise due to data inconsistencies and make comparisons cumbersome. Thus, these indices only measure the target variable at the country level. Cross-country comparisons are possible, but supranational heterogeneities between industries or regions remain unidentified.
The "Gender-Equality Index" (Bloomberg 2022) represents the first specifically company-based quantitative measurement concept, designed through survey data from currently 418 selected international companies at regular intervals. However, one disadvantage concerns the sample: It only contains large international corporations, which might not be representative of society. Small- and medium-sized enterprises employ approximately

95% of the German labour force. In addition, due to the nature of a questionnaire, the data are neither freely accessible nor completely objective. With the GEFI, we address these issues by involving all German firms active in the labour market, without relying on survey data.

In addition to the well-established indices listed above, there have recently been attempts to reduce the aggregation level. Di Bella et al. (2020) address this problem with the Regional Gender Equality Index (R-GEI) based on the 20 NUTS-2 regions in Italy. They stress that nationally summarized indices cannot adequately reflect heterogeneities within the country. Most of these imbalances arise at a regional level. Supranational indicators must therefore be developed in order to map social processes correctly. If deficits are identified in certain domains, improvement measures can also be better enforced with greater precision. Italy for example exhibits strong regional heterogeneities between the south and the north. To maintain long-term nationwide convergence toward gender equality, policies may need to be tailored to specific environments (Di Bella et al. 2020).

The labour market in Germany shows strong heterogeneities with regard to various regions as well. Although the East–West reunification happened multiple decades ago, structural differences are still present. In addition, a high degree of sectoral diversity can be observed. Germany has 21 official sectors, such as construction and health and social services, and one would expect to observe imbalances between men and women within these sectors. We will analyse this in more detail in a case study starting from Section 4.

3 Gender equality firm index

We propose a highly granular quantitative gender equality index at the firm level that can be collected with high frequency and at a low cost through large-scale data extraction of firms' websites. Kinne and Lenz (2019) demonstrate the emergence of useful firm information from websites using deep-learning methods. The possibility of collecting independent text data through data extraction is advantageous and enriching. In particular, due to the continuous improvement in technical feasibility, transferring large amounts of unstructured text and documents into a form that can be used for economic analyses is possible (Lüdering/Winker 2016; Kinne/Lenz 2019; Lenz/Winker 2020).

Thus, to capture latent company characteristics, including web data in the analysis is necessary. Focusing only on the data provided by the companies at this level of aggregation could provide an incomplete concept. A web presence is among the most important

B2C channels, and the way a firm represents itself online could tell a lot about the underlying firm policy. We enable and simplify the quantification of these structures with the advancing digitalization of the collection of online mass data. Using this data (e.g., as we do with GEFI) is important. We are aware that we may only be able to measure marketing capabilities, because companies are aware of these characteristics when positioning themselves. We try to correct this with an extra indicator, as explained below.

Furthermore, we use the MUP panel database because it contains information of the total population of firms in Germany. In addition, we do not assume ex ante that there is an imbalance to the advantage of men, since it can also be the other way around. While the majority of empirical studies and official data point to a preference for men to the detriment of women, we want to remain as objective as possible. We do not equate gender imbalances to the disadvantage of women but aim to detect structural imbalances in the labour market in general.

The GEFI is a composite index, strongly oriented towards the SIGI, with one final score for each company. The GEFI consists of five components:

- Publicly announced pro-diversity programs ("pro-diversity-programs")
- Degree of gender diversity in leading positions ("Blau-executive")
- Discrepancy between representation and implementation ("purple-washer")
- Gender distribution on the website ("Blau-website")
- Number of children ("number-child")

That is, we assume that all sub-indicators should be weighted equally. The GEFI is built as an unweighted average, thus it has a logarithmic and exponential function. In addition, the GEFI does not have a hierarchical component because each indicator enters the index directly. The individual indicators are between zero and one and incorporated using the following formula, following the SIGI:

$$GEFI_i = \frac{1}{5}(e^{pro-div_i} + e^{Blau-exec_i} - e^{purple_i} + e^{Blau-website_i} + e^{child_i}) \tag{4}$$

This functional form allows for a higher degree of variability in the substitution rate between lower and higher levels of discrimination (OECD 2019). We detail each of these components in the following subsections.

3.1 Publicly announced pro-diversity programs (pro-diversity-programs)

The degree to which companies disclose information regarding their ethics or at a non-economic, non-business- or non-product-related level is captured with this component. It specifically accounts for concrete measures of achieving gender equality. The company is not only theoretically striving for equality, but also has policies to achieve this goal. As described in Section 2.1, Timmers et al. (2010) explicitly point to the positive relationship between gender measures and a reduction in the vertical gender segregation. Fehre et al. (2015) mention home office and flexible working hours as an important measure, especially for women. This also includes offers such as part-time jobs or job sharing. This is particularly suitable for parents who can or want to work only part time. The more that pleasant circumstances are offered, the greater the chance that qualified and motivated employees will be hired or remain loyal to the company (Fehre et al. 2015).

In contrast to the literature that is based mainly on questionnaires and interviews, this indicator is constructed using a keyword search on each firm website. It can be expected that companies with an appropriate gender policy implement equality measures and thus also achieve a comparatively high score. We included terms such as "childcare", "maternity leave" or "antidiscrimination law" and identified 15 exhaustive measures, which are listed in Appendix A. In a second step, we determine all relevant variants of the manually selected terms using word embeddings.[1] The variants can be used to find semantically similar words, such as "forecast" and "prediction". Employing this technique, we identify the 10 most similar words to the original words and include these in the keyword search. These words are listed in Appendix A as well. It is also possible that companies use different spellings or synonyms, which is taken into account that way.

For each company, the measures are counted and summarized. Words expressing the same measure (e.g., "parental leave" or "maternity leave") are not counted twice because they denote the same measure. Multiple mentions of a term does not guarantee added value, and in the worst case, distorts the results. Our main goal is to identify as many different measures as possible.

Because the share of the measures mentioned is marginal relative to the total number of words, we use a robust min-max normalization (i.e., we remove the median and scale the data according to a quantile range between the first and the third quartile). If, for example, a company mentions an equal opportunity representative on its website,

[1] This are a novel paradigm from computer science to represent words and phrases, as well as their compositionality (Mikolov et al. 2013).

then it can be assumed that this company promotes and implements equality between women and men and performs internal tasks in this respect, which ultimately leads to an increase in performance.

We normalize the sum by dividing it by the total number of words on the website. The normalization compensates for the firm-size effect because of the positive correlation between company size and Internet adoption (Pian/Teo 2004). A DAX-30 company is initially more likely to be able to offer in-house childcare than a small, family business with few employees and furthermore to mention this on its website. Additionally, Pfeffer (1977) argues that larger firms have more HR-related capital available to implement measures that address challenges associated with diversity (Pfeffer 1977).

3.2 Degree of gender diversity in leading positions (Blau-executive)

In this step, we aim to measure gender diversity at the management level. The use of a gender quota can be seen as a classic tool for measuring equality. Since information about the firm hierarchy is generally not available on the website, company executive data of the MUP is used here. This indicator is constructed using the gender of all executives as a basis. The degree of equality is measured for each firm, where the maximum value indicates a great orientation towards gender equality and a perfectly balanced composition of women and men, while values below the maximum show discrepancies. This information reveals whether a company actively implements gender mainstreaming at the executive level. A homogeneous composition denotes a bias towards the respective direction. It may happen that the same person has different top-level management positions within one company, thus each person is only counted once. For this purpose, the Blau Index (Blau 1977) is used as a basis. The score for each firm is calculated using the following formula:

$$Blau_i = \sum_{i=1}^{n} p_i^2, \tag{5}$$

where n is the number of groups (i.e., male or female) and p_i is the proportion of members, respectively. Yet, using this formula would yield only values from zero (minimum) to 0.5 (maximum). The score is rescaled when calculating the index to reach a maximum value of one. We deliberately do not use a women's quota, as this represents the proportion of women in all relevant persons. The maximum value would only be reached if all persons were female, which, however, does not represent gender diversity but rather perfect gender homogeneity.

3.3 Discrepancy between representation and implementation (purple-washer)

To achieve a better image among the public, companies have an incentive to highlight their efforts towards gender equality. However, these mainly latent measures cannot be clearly traced or quantified, resulting in a risk of non-objective things being mentioned. This is also known as purple washing, a term primarily influenced by so-called "green washing", which has not yet been applied in the academic literature. The focus here is on marketing methods that aim to present a company to the public as equitable without the necessary conditions being met.

Generally, in Germany, although gender diversity on executive and management boards is increasing, this growth is very sluggish. In 2018, the women's quota in Germany was just 26.1% for firms with up to 10 employees. For bigger firms, the quota has been even smaller (e.g., 17.1% for firms with 11 to 50 employees), resulting in an overall German average of 22.6%.

This indicator is constructed using the information collected in "pro-diversity-programs" and "Blau-executive": Companies identified as diversity-oriented at first (and ceteris paribus having a non-zero amount of pro-diversity measures) should employ both men and women in higher positions.[2] However, if the executive level only hosts members of a single gender, then the firm might only attempt to achieve equality, but not actually implement it. Therefore, we calculate the discrepancy between "Blau-executive" and a theoretically expected reference value. The latter is computed as the average of the expected score of the Blau Index with regard to size, sector, and region for each firm using data from the Federal Statistical Office (Destatis 2020a; 2020b; 2020c). Subsequently, we subtract this from the actual indicator score. If the result is negative (i.e., the actual equality is lower than theoretically expected), then it can be assumed that the firm is purple washing. It only pretends to be gender friendly, but the level of equality does not correspond to what one would expect, given measurements for similar companies. Formally, this can be presented as:

$$purple - washer_i = \begin{cases} Blau_{official,i} & pro - diversity - measures > 0 \\ 0 & pro - diversity - measures = 0 \end{cases} \tag{6}$$

For example, assume firm i reaches a Blau score of 0.15. We subtract this from its expected value (e.g., 0.22[3]), which would yield a value of 0.07. However, unlike other

[2]The value of purple-washer in the companies with no hits in pro-diversity-programs is zero, as there is no statement about equality.

[3]This serves as a synthetic example. Theoretical Blau scores are calculated for all companies of a certain size class, region, and sector. If an average score of 0.20 is calculated across all companies in a

indicators, "purple-washer" enters the index negatively, as a discrepancy between the representation and implementation that negatively effects gender equality.

Gender distribution on the website (Blau-website)

While the focus was formerly on the executive level, the following indicator concentrates on all persons in a firm. "Blau-website" measures the proportion of women and men mentioned on a firm's website. That is, an overall diversification quota based on website content is determined. Even without concrete firm enquiries, the gender distribution within the firm structure is approximated.
The implementation is done by matching first names; that is, filtering and counting female and male first names on the website. Specifically, a list of all male and female first names is taken and, in case of a match with a word on the website, the last names are checked.[4] Since there is no official list of last names, only words that begin with an uppercase letter followed by only lowercase letters are used as last names. A word recognized as a first name is followed by either a last name, a verb, or a punctuation mark. We calculate a Blau score based on the filtered website names. Equivalent to "Blau-executive", each person in a company is considered only once. A ratio between zero and 0.5 is subsequently rescaled to achieve a maximum value of one. In addition to the quota of official firm MUP data, this measure reveals additional information on the latent underlying company structure. If a company pursues a consistent gender equality policy, then both scores from "Blau-executive" and "Blau-website" should be relatively similar. However, as already highlighted, the proportion of women in top management is lower than in lower positions. Therefore, "Blau-executive" and "Blau-website" will likely differ.

3.5 Number of children (number-children)

For this indicator, we analyse the amount of children of executive level employees. To avoid withdrawing completely from working while parenting their children, employees often choose part-time employment. In 2018, almost 30% (11 million people) of the working population in Germany worked part-time.
Measures such as corporate childcare or maternity leave represent a major factor influencing people's decisions about whether to work for a company. If these benefits are

sector (for all sizes and regions), then another score of 0.28 is calculated for all companies in a size class (across all sectors and regions), and a third score of 0.18 is calculated across all companies in a region (for all companies in a size class and sector), then 0.22 would be obtained on average.

[4] We used a list hosting 40,000 first names in Germany. In split cases, we used the German connotation (e.g., Andrea is a female name in Germany, but male in Italy; https://pypi.org/project/gender-guesser/). Only names to which a gender could be clearly assigned are considered.

offered, then the company can be seen as gender or family friendly, which is reflected in a larger number of children. However, it should be noted that this factor only plays a role for an individual at the stage of life when their children are not yet adults. For this reason, only people who have not yet passed the age of 60 are considered.

We take the mean of the number of children per person at the executive level over each company from the MUP and normalize by the maximum value of the mean values over all firms to obtain a value between zero and one.

4 A Case Study for Germany

We demonstrate the applicability of the GEFI in a case study for Germany. We detail the data and the dimensions of the analysis in the following two sections.

4.1 Data

A company's public appearance, approximated through its website and enriched with firm data from the MUP, serves as a basis for the GEFI computation. MUP is updated semi-annually, and it maintains data for all companies based in Germany. It represents the most comprehensive panel company database outside official statistics. We checked whether the companies listed in the MUP have an Internet presence and, if they did, whether the texts and information contained there could be extracted. Of the original 1.3 million companies in Germany listed in the MUP, about 993,000 were used after filtering. This reduction is because companies either have no web address or do not provide sufficient information in the MUP.

To obtain the final data set, we use two pre-processing steps on the raw website data: We transform each word into its stem to filter out noise due to different word endings.[5] Further, to obtain comparable and consistent results, we deleted all non-German texts.

4.2 Company characteristics

We calculate the GEFI for all firms in the dataset described in the previous section. Given that firm-specific characteristics might exert influence on the degree of equality, clustering firms based on certain criteria is appropriate.

We first aggregate our results according to the official sector classification of the German

[5]Stemming is conducted following pro-diversity-programs because this step changes the words so that they would no longer be recognized as names.

Federal Statistical Office, which includes 21 branches (e.g., construction, healthcare, and social services). The list of all sectors is in Appendix B. We aggregate GEFI scores for every sector and consequently identify those with need for improvement.

Second, we define two regional classifications: Initially, the analysis is performed on the federal state level. We examine the differences in gender equality between the 16 federal states. Clear signs of economic imbalances exist between states. For example, the difference in GDP per capita between the strongest federal state (Hamburg) and the weakest (Saxony-Anhalt) is more than 130% (Destatis 2021b). Subsequently, an overarching, expanded, analysis of western and eastern Germany is conducted. Great efforts have been made to align Germany after the reunification. However, structural differences still exist (e.g., in average wages or productivity). Although there has been convergence in GDP growth between the regions since 1991 (1991: 10% in the east and 1.2% in the west vs. 2018: 1% in the east and 1.4% in the west; Destatis 2019c), differences become obvious when looking at the locations of the DAX companies. Not even one of the 30 companies comes from the former East German region, whereas more than half are located in Bavaria (27%) and North Rhine-Westphalia (30%) alone (Destatis 2021b).
Lastly, we aggregate with respect to firm size because we expect a positive relationship with gender equality. Segarra and Teruel (2017) argue that larger companies have to be more complex, offer more products, and solve tasks that are more specialized. A gender-friendly environment helps to cope with these tasks (Segarra and Teruel 2017). In line with this, other studies find that diversity has a positive influence on performance in large groups (Kanfer et al. 2008). Here, too, we follow the official size classifications when grouping (1-5, 6-25, 26-250, ¿250 employees).

5 Main Findings

In this section, we present the findings of our case study, which include selected results of the individual indicators obtained for each firm. Additionally, we discuss the GEFI scores of the sub-dimensions.
A need to improve gender-related working conditions becomes clear when looking at the results from "pro-diversity-programs". We identify topics and expressions that have been used the most among the firms in our sample: Only 7.6% mention alternative, family-friendly working time models with benefits such as "home office," "working time model," or "work life balance".

Legal regulations against gender discrimination, such as "prohibition of discrimination" and "anti-discrimination law" or "anti-discrimination policy," are found for 6.5% of

firms. Words such as “maternity leave” were use by 3.4% of companies that mentioned measures to support the period before and after childbirth. The possibility of supervising children after school within the company is mentioned by only 2.6% of the firms in the sample, using buzzwords such as “child care.” As stated above, this factor is extremely relevant for many families. This result highlights the alarmingly low attention paid to these measures.
Of additional noteworthiness is the discrepancy between ”Blau-website” and ”Blau-executive” over all companies: the average value of the former is 40% higher compared to the latter. That the gender distribution across management and employees is still asymmetrical becomes obviously clear.

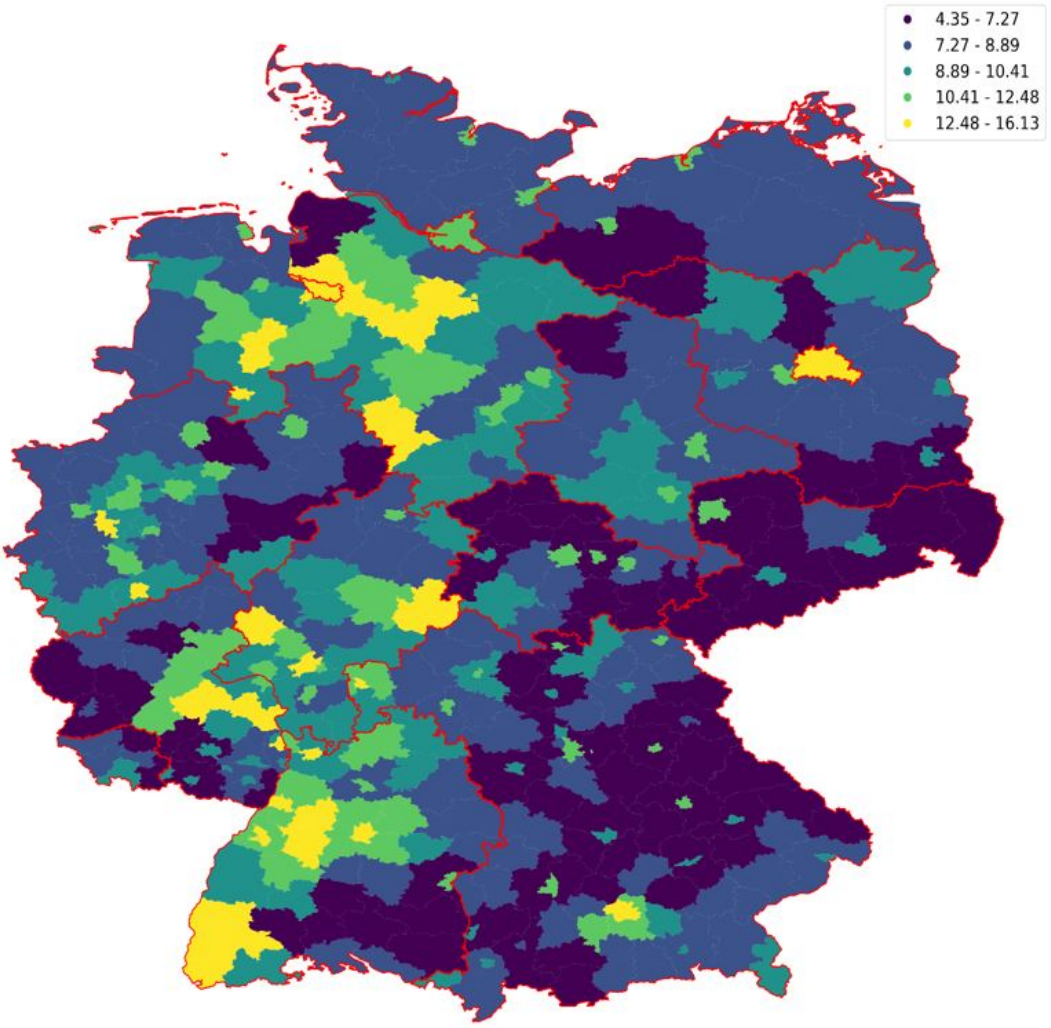
5.1 Regions

Within the following, we conduct a regional analysis of the GEFI scores. Figure 1 maps the overall results for all firms in Germany at the county level. Lighter areas imply greater equality and darker areas lower equality. The red lines detail the borders of the individual federal states.
What is immediately striking are the differences between the west and east. In the east (including Bavaria), most territories display a rather dark coloration, implying lower average GEFI values. The western states show an average score of 9.62, while the eastern states were slightly lower, averaging 9.09.[6] The lowest overall scores were found in Saxony, Thuringia and Brandenburg. This is consistent with regard to other economic indicators: The east still lags far behind western Germany in terms of economic performance, productivity, innovative capacity and unemployment (Borger and Müller 2014). What also stands out is that the city states (Bremen, Hamburg, and Berlin) report the highest average GEFI scores. Their average value (12.2) is more than 30% higher than the average value of the other federal states (9.3). Straus and Sugarman (1988) uncover and confirm a significant positive correlation between urbanization and gender equality. As the name suggests, the city states encompass only the territory of a city and its immediate surroundings. The population density here is noticeably higher than the national German average (233 people per km^2): Berlin has a population of 4,090 people, Hamburg 2,446, and Bremen 1,624 per km^2(Destatis 2021c).

When looking instead at the territorial states in particular, one can discern differences from the graph: Lower Saxony, Hessen, and North Rhine-Westphalia report values above average, whereas the opposite can be observed for Saxony, Brandenburg, and Thuringia.

[6] Excluding the city states, the value is 9.5 in the west and 7.9 in the east, so Berlin strongly alters the average in the east.

Figure 1: GEFI scores for Germany



In some cases, a clear border effect can be seen, such as in Lower Saxony and Saxony-Anhalt (i.e., geographically adjacent areas show differences in the GEFI scores). Here, too, the difference between east and west becomes clearly visible.

Finally, the figure also reveals differences between rural and urban territories. We find that the latter (11.89) has higher average values compared to the former (8.74). For example, when looking at Bavaria, most rural areas are coloured rather darkly, indicating less gender equality. However, scattered areas stand out relatively brightly. These are non-district cities, such as Munich, Nuremberg, or Würzburg. Similar to city states, the population density there is much higher, which confirms the results from the literature.[7] Again, Straus and Sugarman's (1988) hypothesis that more urban areas are associated

[7]The average population density of Bavaria is 186 inhabitants per km^2, thus population density varies widely within the individual districts. The population density in the Bavarian capital, Munich, is 4,777 inhabitants per km^2; in the district of Munich, 528 inhabitants per km2; and in the district of Lichtenfels, only 66 people per km^2.

with a higher degree of equality can be confirmed. When comparing the mean values, we confirm the statistical significance of this result for every level of significance. The degree of gender equality is therefore significantly higher in urban areas than in rural areas.

In summary, companies in urban or western areas show on average higher GEFI values, suggesting they implement more equality than rural and eastern areas.

When looking at the sub-indicators, "pro-diversity-programs" confirms that companies in the western federal states are found to have an on average a slightly lower number of measures on their websites compared to the eastern areas (14.9% vs. 15.2%). When looking at the individual states, the city states Hamburg (18.8%), Berlin (17.5%), and Bremen (17.3%) can be named as the areas with the most companies with at least one hit.

They also achieve above-average scores for "Blau-executive" (Berlin: 0.13, Hamburg: 0.12, Bremen: 0.11), compared to the average for all states (0.101). It is worth mentioning that the firms in the eastern states achieved higher values on average than those in the west did (0.104 vs. 0.097). This is consistent with official data of the Federal Statistical Office, which indicates higher official women's quotas in higher management levels. While the share of women on DAX boards in the east was 75% in 2019 (as opposed to 10% in the west), it should be noted that only four eastern German board members are represented in all DAX30 companies, which might affect the validity of this figure (Destatis 2019a; 2020d).The same results are achieved when looking at "Blau-website" scores. Again, the city states (Berlin and Hamburg: 0.16, Bremen: 0.15) perform best.

When looking at "purple-washer", an apparent inconsistency can be detected. As discussed above, "Blau-executive" scores are on average higher in the east, but those of the overall GEFI are higher in the west. The results in "purple-washer" may be able to explain this: Purple washing is most pronounced in the eastern states. In general, the average values in the city states and the western German territory states (0.25) are quite similar but lower than in the east (0.27). However, the share of firms doing purple washing relative to all firms are higher in the east.
With regard to "number-child", we find relatively similar results for all states. The highest values for this indicator are achieved in Saxony, Saxony-Anhalt, and Brandenburg (0.08 each) and the lowest in Bremen, Baden-Württemberg, and Lower Saxony (0.07 each). However, these differences are very marginal, with a fluctuation of only about 1 percentage point.[8] This is in line with other official statistics: On average, almost one

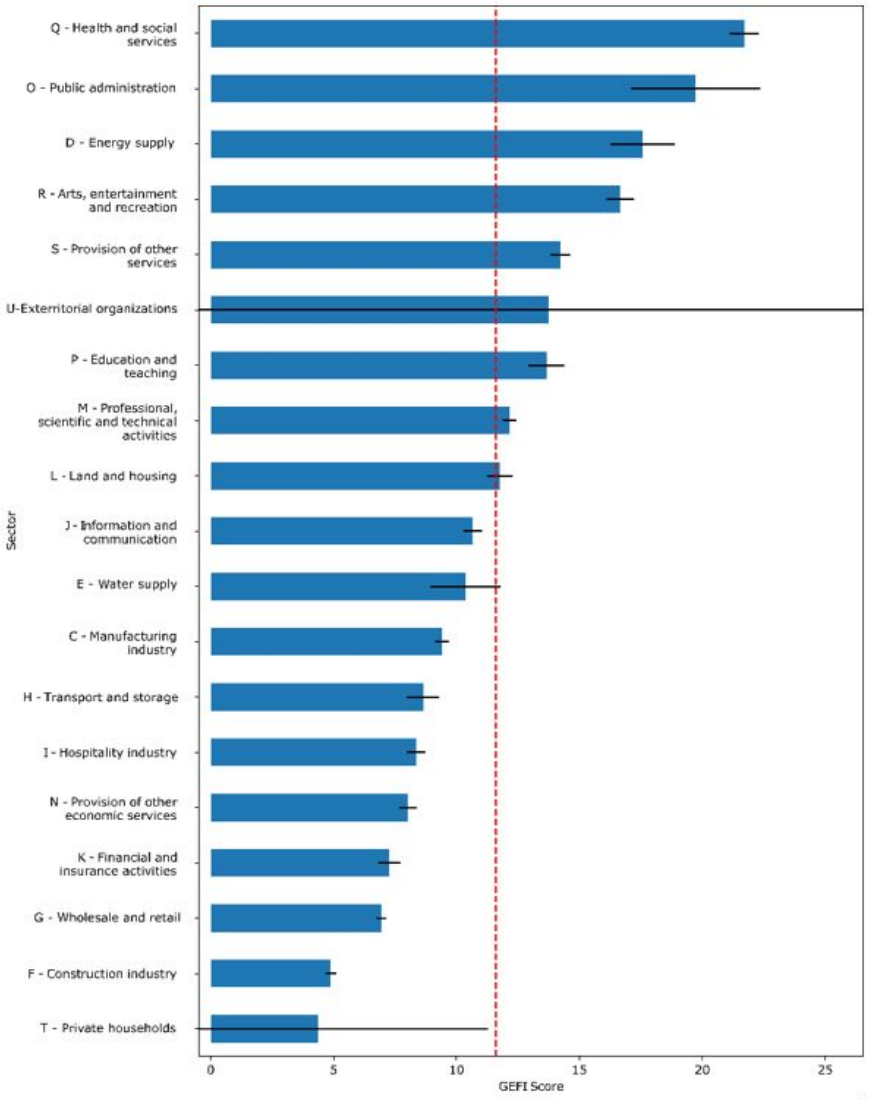
[8]We do not interpret the numbers directly because these are not the absolute numbers of children, but rather the normalized values.

in three western households (29%) had children in 2018. In the east, the proportion of children was slightly lower (23%), yet the female employment rate was marginally higher by about 3-5% (Destatis 2019b).

5.2 Sectors

Here we analyse GEFI scores achieved per sector across all companies in Germany. Figure 2 displays the values. The red vertical line represents the average across all sectors, and the black horizontal lines represent the 95% confidence intervals. Sectors T (private households with domestic staff) and U (exterritorial organisations) should be treated with caution, since there are very few observations (only about 0.01% and 0.017% of all firms), which cause very wide interval boundaries.

Figure 2: GEFI scores per sector in Germany



As one would expect, the highest scores are found in the health and social services sector. Compared to the average (11.8), the firms in this sector score more than 80% higher (21.7). The sector for public administration and defence follows, still being nearly 67% above average (19.7). The education sector is also in the upper range (13.7), which is consistent with previous studies. For example, Eriksson (2014) examines a secondary school in Sweden in her qualitative study on the effect of gender perspective on innovations. She emphasizes that the school emerged as the most successful of seven institutions in her investigation with regard to the generation and implementation of innovation within a gender-mainstreaming project. In line with our results, she mentions the comparatively high degree of gender equality in public sector organizations, which is advantageous to achieving desirable social outcomes (Eriksson 2014).

In contrast, the construction industry scored the lowest (4.9), which may not be surprising because the majority of this sector is dominated by small firms, often still run by men. For example, Davey et al. (2010) investigate the position of women in the construction sector in the UK and name the educational segregation in this context that causes the occupational segregation. They point out that women were only 7% of architecture students and 8% of engineering and technology students, especially at higher educational levels, at the time of their study. The majority of women employed in the UK construction sector work in office and secretarial jobs, almost half of them part time. When looking at management and administration, women are mainly employed in areas such as human resources or public relations and are left out of concrete and strategic operations. In terms of the professional/technical category, women are severely underrepresented. Reasons cited include the image of the sector, selection criteria and male-dominated courses, recruitment practices and procedures, sexist attitudes, male-dominated culture, and the work environment (Davey et al., 2010). This is also consistent with Beede et al. (2011), who report that women are not only a minority in STEM professions, but are also less likely to pursue a profession with a STEM degree (Beede et al. 2011).

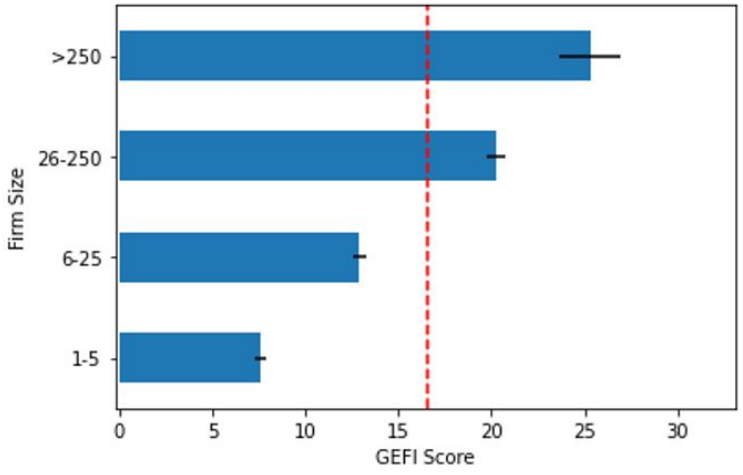
In conclusion, the GEFI results with regard to the sector dimension are in line with findings previously reported in the literature. We are also able to derive direct implications for the individual indicators. Regarding "pro-diversity-programs", we find that gender equality does not seem to play a central or at least noteworthy role in company policy in most cases. However, substantial differences exist between the individual sectors. While 61% of the firms in the public administration sector name at least one measure, only 9% could be found in construction.
For "Blau-executive", our results are consistently similar to the literature. Health and social services (0.21) and public administration (0.16) have the highest values, while the construction industry (0.07) has been identified as the least diverse. Furthermore,

"Blau-website" results reveal that the educational sector, which is more in the middle of the range (0.12) for the MUP data, is now one of the sectors with the highest scores (0.2). One would expect this result in view of the above-mentioned literature. Most sectors show higher average values relative to the official data. However, the trend is different in the construction sector: While it is the least gender friendly compared to other sectors, its average value also decreased (0.57).
The sectors that achieve the highest Blau scores (health and social services: 0.2; public administration: 0.21) also achieved the best results in "purple-washer" relative to the sector average (0.24).
Lastly, when looking at "number-child", it is noticeable that the sectors that previously always performed best only achieved below-average values for this indicator (health and social services: 0.03; public administration: 0.02), while trade, private households with domestic staff, and other economic services (0.09 each) performed best.

5.3 Firm size

Finally, we relate firm size to GEFI scores. Figure 3 shows the average GEFI score in relation to the number of employees.

Figure 3: GEFI score per firm size group in Germany



It becomes obvious that an increase in company size is accompanied by an increase in gender equality. A positive relationship between firm size and gender friendliness can be observed here, which supports the above-mentioned theory of Segarra and Teruel (2017). Every next higher group achieves a higher GEFI value.[9] While firms in the largest size category are almost 68% above the average (25.3 vs. 15.1), the smallest firms with only

[9]In some companies, individual details are missing (e.g., with regard to the sector or the size). This missing information cannot then be taken into account when calculating the mean value, which is why it differs between the respective dimensions.

1-5 employees are just under half that (7.6). This may also be due to larger companies having more detailed and better-planned websites (e.g., providing more space to mention equality measures). We tried to account for this by performing the count normalization in "pro-diversity-programs"; however, future research needs to determine the effect of company size on web-based company measurements.
In the examination of the individual indicators, the following results can be mentioned for "pro-diversity-programs": While in the groups of 1-5 and 26-250 employees, measures were found in only slightly more than 10% of the companies, in the group of 6-25 employees or groups with more than 250 employees, the figures are 46% and 57%, respectively.

For "Blau-executive", the largest companies (with more than 250 employees) also achieved the highest scores. Similar results were observed with regard to "Blau-websites": The larger the company is, the higher the score. Again, in many cases the values are far above those from "Blau-executive".
The "purple-washer" scores are in line with the overall results: An inverse relationship exists between the firm size and the value. However, for the group of 1-5 employees, one should be careful with the interpretation: If a company consists only of a single person, values close to zero for the Blau index result automatically. This might affect the value of this indicator, and future research needs to examine the effects of this further.
The results for "number-child" are not in line with the overall results or those of the above-mentioned indicators. While the scores increased with firm size, above-average values were achieved, especially for the group of small (1-5 and 6-25 employees) companies.

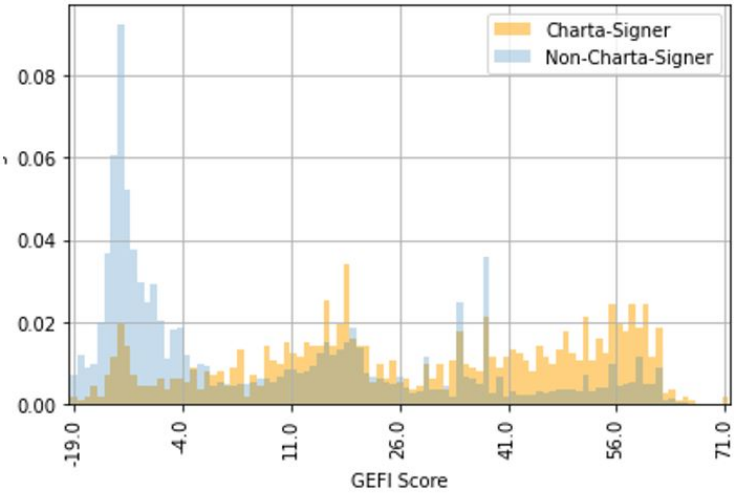
6 Validation

During the time of writing, no comparable gender equality index exists at the firm level of granularity. Nevertheless, evaluating whether the calculated GEFI scores correspond reliably to the value originally intended to be measured is necessary. Since no direct comparison is possible, we decided to conduct an indirect evaluation through external proxies. In particular, we compare GEFI scores from companies that explicitly expressed the commitment to a working environment free of discrimination and prejudice with a random sample of those that have not done so. The sample of commited firms consists of companies that have signed the Charta der Vielfalt (Diversity Charter)[10]. The Charta der Vielfalt is a self-commitment published in 2006 and an association under the patronage of the German Chancellor, which is committed to an unprejudiced working

[10]https://www.charta-der-vielfalt.de/

environment. By signing the charter, employers declare that they will create or promote equal opportunities for their employees. Their commitment to this charta should be reflected in a higher average GEFI score. The charter currently has almost 4,000 signatories, from which we were able to match 1,120 companies. Thus, the overall sample size is 2240 companies. We evaluated the resulting GEFI distributions. Specifically, we compared the mean values of the GEFI scores of both samples. Figure 4 shows the histograms of the empirical distribution of scores for non-signers (blue) and signers (yellow). The visual analysis of the figure suggests that the scores for signers are higher than those of non-signers are.

Figure 4: Distribution of GEFI scores for non-signers (blue) and signers (yellow)



The distribution of non-signers also shows a clear leftward slope. The average scores for signers is 30.29, which is much higher than the average is for the random sample of non-signers (9.85). This difference is highly significant ($p<0.01$). These findings suggest that the GEFI is capable of providing grounded approximations to gender equality at the firm level and and also hint towards the applicability of the approach, as it appears that the GEFI indeed measures the intended dimension.

7 Conclusion

The GEFI is a composite index measuring the degree of equality between men and women in firms. In our study, we incorporate not only official company data but also latent variable observations, quantified through large-scale data extraction and analysis of company websites. We proposed the GEFI, a composite index consisting of five sub-indicators, to quantify these visible and latent factors. Thereby, we were able to compute a score for each company serving as an indicator of equality achievement.

This concept is completely new and innovative relative to previous quantitative studies because it allows us to measure equality at the firm level objectively and can be fully automated for an entire economy.

To gather first insights into the validity of the GEFI, we conducted a case study for Germany. Our results are consistent with expectations and the results from previous gender equality studies. Although previous studies were conducted at a different level of aggregation, we were still able to confirm their results qualitatively. Our findings confirm the hypotheses that equality is positively correlated with urbanization and that it is enforced more in western Germany. In addition, a disproportionately high imbalance between men and women still exists at the executive level and between employees.

When focusing on the branches, the results are also in line with our expectations. Sectors such as construction are still mainly male-dominated, if only for physical reasons, and this is reflected in GEFI scores far below the average. In sectors such as healthcare, our results are in line with previous findings. When classifying by company size, a clear trend can be observed. As the number of employees increases, higher GEFI values can also be realized. This is not surprising and can be explained by the fact that larger companies spend more financial resources on employee well-being. Among other things, this can be communicated to the public through a company's web presence, which results in a higher GEFI value.

Our findings suggest that the GEFI can be a valuable tool for policymakers to guide targeted equity and improvement efforts. First, our descriptive results point to structural deficits in the labour market that can be specifically targeted (e.g., through corresponding requirements from politics and lawmakers). Furthermore, the GEFI could be used in the future for quantitative analyses; for example, to measure the influence of gender equality on the economic success of a company.

This concept is aiming to quantify vertical gender segregation for all domains, therefore it is important to find adequate ways of measuring. Expanding the data base and improving the methodology is a beneficial way to improve results. Additional indicators could be added or existing ones improved. Future research could even go one level deeper and analyze individuals within firms across all firms. After all, the employees listed on the websites are already known, so it would be beneficial to also analyze structural dynamics within firms. This could be done, for example, using a keyword search to look for differences in how men and women are portrayed by the employer on the website.

References

Beede, D. N., Doms, M. E., Julian, T. A., G., Khan, B., Langdon, D., McKittrick (2011), Women in STEM: A gender gap to innovation. Economics and Statistics Administration Issue Brief, 04(11).

Bekhouche, Y., Hausmann, R., Tyson, L. D., S. Zahidi (2012), The global gender gap index 2012. The Global Gender Gap Report, 2012: 3-27.

Bloomberg (2022), Gender Equality Index, can be downloaded at: https://www.bloomberg.com/gei. (webpage dated 2022, last retrieved 04.02.2022).

di Bella, E., Gandullia, L., Leporatti, L., F. Maggino (2020), Proposing a regional gender equality index (R-GEI) with an application to Italy. Regional Studies: 1-12.

Berglund, K., J. Granat Thorslund (2012), Innovative policies? Entrepreneurship and innovation policy from a gender perspective. In S. Andersson, K. Berglund, E. Gunnarsson, & E. Sundin (Eds.), Promoting Innovation: Policies, practices and procedure. Vinnova: Stockholm.

Bericat, E. (2012), The European gender equality index: Conceptual and analytical issues. Social Indicators Research 108(1): 1-28.

Blau, P. M. (1977), Inequality and heterogeneity: A primitive theory of social structure (Vol. 7). New York: Free Press.

Borger, K., M. Müller (2014), In der Normalität angekommen–Deutschland 25 Jahre nach dem Mauerfall. Fokus Volkswirtschaft. KfW Economic Research 73.

Brader, D., J. Lewerenz (2006), Frauen in Führungspositionen: An der Spitze ist die Luft dünn (No. 2/2006). IAB-Kurzbericht, Can be downloaded at: http://hdl.handle.net/10419/158209.

Branisa, B., Klasen, S., M. Ziegler (2009), The construction of the social institutions and gender index (SIGI) (No. 184). IAI Discussion Papers, Can be downloaded at: http://hdl.handle.net/10419/57322.

Campbell, L. G., Dozier, M. E., Mehtani, S., J. Rinehart (2013), Gender-heterogeneous working groups produce higher quality science. PloS one, 8(10).

Carli, L. L., Eagly, A. H. (2007). Through the labyrinth: The truth about how women become leaders. Boston, MA: Harvard Business School Press, Can be downloaded at: https://doi.org/10.2189/asqu.53.2.363.

Davey, C. L., Davidson, M. J., Fielden, S. L., A. W. Gale (2000), Women in construction: the untapped resource. Construction Management Economics, 18(1): 113-121.

Destatis (2019a), Frauenanteile unter West- und Ostdeutschen in führenden wirtschaftlichen Positionen in Deutschland im Jahr 2019, Statistisches Bundesamt (Destatis).

Destatis (2019b), 30 Jahre Mauerfall: Familien in Ost- und Westdeutschland werden sich immer ähnlicher, Statistisches Bundesamt (Destatis).

Destatis (2019c), Wie sich die Wirtschaft in Ost und West verändert, Statistisches Bundesamt (Destatis).

Destatis (2020a), Frauenanteil in Führungspositionen in Deutschland nach Anzahl der Mitarbeiter im Unternehmen im Jahr 2018, Statistisches Bundesamt (Destatis).

Destatis (2020b), Frauenanteil in Führungspositionen in Deutschland nach Branchen im Jahr 2018, Statistisches Bundesamt (Destatis).

Destatis (2020c), Frauenanteil in Führungspositionen in Deutschland nach Bundesländern im Jahr 2018, Statistisches Bundesamt (Destatis).

Destatis (2020d), Anzahl der DAX, MDAX und SDAX-Unternehmen nach Bundesländern im Jahr 2020, Statistisches Bundesamt (Destatis).

Destatis (2021a), Gender Pay Gap: Verdienstabstand zwischen Männern und Frauen in Deutschland von 1995 bis 2019, Statistisches Bundesamt (Destatis).

Destatis (2021b), Bruttoinlandsprodukt (BIP) je Einwohner nach Bundesländern im Jahr 2020, Statistisches Bundesamt (Destatis).

Destatis (2021c), Bevölkerungsdichte (Einwohner je km^2) in Deutschland nach Bundesländern zum 31. Dezember 2020, Statistisches Bundesamt (Destatis).

Etzkowitz, H., M. Ranga (2010), Athena in the world of techne: The gender dimension of technology, innovation and entrepreneurship. Journal of technology management & innovation, 5(1): 1-12.

Eriksson, A. F. (2014), A gender perspective as trigger and facilitator of innovation. International Journal of Gender and Entrepreneurship: 163-180.

Fehre, K., Lindstädt, H., A. Merklein (2015), Diversity Management in Deutschland: Empirische Untersuchung der Maßnahmenebene–Fokus: Gender (Frauen). Abschlusspräsentation und Bericht zur Studie. Karlsruhe: KIT, Institut für Unternehmensführung.

Glick, P., L.A. Rudman (2012), The social psychology of gender: How power and intimacy shape gender relations. New York, NY: Guilford.

Hoogendoorn, S., Oosterbeek, H., M. Van Praag (2013), The impact of gender diversity on the performance of business teams: Evidence from a field experiment. Management Science 59(7): 1514-1528.

Jansen, P. G., van der Velde, M. E., I. A. Telting (2001), The effectiveness of human resource practices on advancing men's and women's ranks. Journal of Management Development.

Kabeer, N., L. Natali (2013), Gender equality and economic growth: Is there a win-win?. IDS Working Papers, 2013(417): 1-58.

Kanfer, R., Neubach, B., Roth, C., Schmidt, K. H., J. Wegge (2008), Age and gender diversity as determinants of performance and health in a public organization: the role of task complexity and group size. Journal of Applied Psychology, 93(6): 1301.

Kanter, R. M. (1977), Men and women of the corporation: New edition. Basic books.

Kinne, J., D. Lenz (2019), Predicting innovative firms using web mining and deep learning. ZEW-Centre for European Economic Research Discussion Paper, 19-01.

Kerevel, Y. (2019), Empowering Women? Gender Quotas and Women's Political Careers. The Journal of Politics, 81(4): 1167-1180.

Klasen, S., F. Lamanna (2009), The impact of gender inequality in education and employment on economic growth: new evidence for a panel of countries. Feminist economics, 15(3): 91-132.

Lenz, D., P. Winker (2020), Measuring the diffusion of innovations with paragraph vector topic models. PloS one, 15(1).

Lüdering, J., P. Winker (2016), Forward or backward looking? The economic discourse and the observed reality (No. 07-2016). MAGKS Joint Discussion Paper Series in Economics, Can be downloaded at: https://doi.org/10.1515/jbnst-2015-1026.

Mikolov, T., Chen, K., Corrado, G., J. Dean (2013), Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

OECD (2019), SIGI 2019 Global Report: Transforming Challenges into Opportunities, Social Institutions and Gender Index, OECD Publishing, Paris. Can be downloaded at: https://doi.org/10.1787/bc56d212-en . Accessed 23 February 2020.

Pfeffer, J. (1977), The ambiguity of leadership. Academy of management review, 2(1): 104-112.

Pian, Y., T.S. Teo (2004), A model for web adoption. Information Management, 41(4)

: 457-468.

Segarra, A., M. Teruel (2017), The link between gender diversity and innovation: What is the role of firm size?. International Review of Entrepreneurship, 15(3): 319-340.

Straus, M. A., D.B. Sugarman (1988), Indicators of gender equality for American states and regions. Social Indicators Research, 20(3), 229-270.

Unit, E. I. (2010), Women's Economic Opportunity–A New Pilot Index and Global Ranking. Findings and Methodology. Can be downloaded at: http://graphics.eiu.com/upload/WEO_report_June_2010.pdf. Accessed 04 January 2021.

Van Staveren, I. (2013), To measure is to know? A comparative analysis of gender indices. Review of Social Economy, 71(3): 339-372.

Appendix

Appendix A

List of terms from "pro-diversity-programs" sub indicator of the GEFI

Gender Equality Measure	Buzzwords (German)
1. Possibility to be legally absent from work and paid before and after having a baby (e.g. Maternity Leave, parental leave period)	Elternzeit, Mutterschaftsurlaub, Erziehungsurlaub, Mutterschutz(gesetz), Vaterschaftsurlaub, Mutterschutzzeit, Schwangerschaftsurlaub, Erziehungsurlaub, Elternurlaub, Babypause, Mutterschaftsurlaub, Mutterschutz,Babyjahr, Babyzeit, (Kinder-) erziehungszeit
2. Possibility within the company to supervise children after school (e.g. child care, infant care)	Schulkinderbetreuung, Kinderbetreuungsmöglichkeiten, Kinderbetreuung(sangebote), Kindergartenbetreuung, Babybetreuung, Kleinkindbetreuung, Betreuungsplan
3. Communication and feedback between employer and employee (employee survey, feedback conversation, ...)	Mitarbeitergespräch, Jahresgespräch, Fördergespräch, Zielvereinbarungsgespräch, Beurteilungsgespräch, Feedbackgespräch, Kritikgespräch, Mitarbeiterbefragung
4. Family-friendly firm measures for employees with children to continue with job (e.g. family friendly, child-friendly)	Familienfreundlich, kinderfreundlich
5. Measures and decisions that represent a concretization of the firm philosophy regarding gender equality (e.g. gender equality policy, women's promotion plan)	Frauenförderplan, Frauenpolitik, Gleichstellungspolitik, Mädchenförderung, Frauenförderpreis
6. A person/office that is responsible to make sure that employees and applicants are treated fairly regardless of their sex (e.g. Equal Opportunities Officer, Women's Representative)	Gleichstellungsstelle, Gleichstellungsbeauftragte(r), Gleichstellungsbüro, Gleichstellungskommission, Frauenbeauftragte(r), Gleichstellungsplan
7. Promotion of equitable access to equal positions/conditions (e.g. Equal opportunities, Freedom from Discrimination)	Gleichstellung, Geschlechtergerechtigkeit, Anti Diskriminierung, genderorientiert, Geschlechterperspektive, Geschlechter(un)gleichheit, Gleichberechtigung, Geschlechtergleichstellung, Chancengerecht(igkeit), Diskriminierungsfrei(heit), Chancengleich(heit), Geschlechterdiskriminierung
8. Additional payment a woman/couple receives to help pay for the costs of taking care of children (e.g. parental benefit, childcare support)	Familiengeld, Elterngeld, Erziehungsgeld, (Kinder-)betreuungsgeld, Landeserziehungsgeld, Kindergeld, Familienleistung, Zusatzleistung

9. A person/institution to give advice to other people about gender-related conflicts and pregnancies in an organized way (e.g. conflict advisor, pregnancy conflict advisory)	Schwangerschaftskonfliktberatung, Konfliktberater, Schwangerenkonfliktberatung, Konfliktbearbeitung, Krisenberatung, Konfliktcoaching, Konfliktklärung
10. Gender-related goals (e.g. Equality objectives, equality standards)	Gleichstellungskonzept, Gleichstellungsziele, Gleichstellungsstandards, Gleichstellungsrecht, Gleichstellungsmaßnahmen, Gleichstellungsreferat, Gleichstellungsausschuss, Gleichstellungskommission
11. Gender-based quota system for fillings to committees or posts (e.g. women's quota)	Frauenquote, Geschlechterquote
12. Legal regulations against gender discrimination (e.g. Anti-discrimination laws/regulations)	Gleichstellungsgesetz, Gleichbehandlungsgebot, Diskriminierungsverbot, Gleichberechtigungsgesetz, Gleichbehandlungsgrundsatz, Gleichstellungsarbeit, Anti Diskriminierungsgesetz, Antidiskriminierungspolitik, Chancengleichheitsgesetz, Diskriminierungsschutz, Gleichbehandlungsgesetz
13. Strategy to promote gender equality (e.g. gender mainstreaming)	Gender Mainstreaming, Genderpolitik
14. Alternative, family-friendly working time models (e.g. Home office, work life balance)	Teilzeitmodell, (Gleit-)zeitmodell, Arbeitszeitmanagement, Arbeitszeitregelung, Arbeitszeitverteilung, homeoffice, Arbeitszeitmodell, Work Life Blending, Work Life Management, Work Life Balance, Familienarbeitszeit, Beruf und Familie, Arbeitszufriedenheit
15. Compensation for an economic (dis-) advantage caused by sex (e.g. (dis-) advantage compensation)	Nachteilsausgleich, Vorteilsausgleich

Appendix B

List of sectors or economic activities, respectively. Classified by the Federal Statistical Office and used in German official statistics

A Agriculture, forestry, fisheries

B Mining and quarrying

C Manufacturing industry

D Energy supply

E Water supply; sewerage, waste management and remediation activities

F Construction industry

G Wholesale and retail trade; maintenance and repair of motor vehicles

H Transport and storage

I Hospitality industry

J Information and communication

K Financial and insurance activities

L Land and housing

M Professional, scientific and technical activities

N Provision of other economic services

O Public administration and defence; compulsory social security

P Education and teaching

Q Health and social services

R Arts, entertainment and recreation

S Provision of other services

T Private households with domestic staff, undifferentiated goods- and services-producing activities of private households for own use

U Exterritorial organisations and corporation bodies

Chapter 4

Gender Perception in the Labor Force

Chapter 4: Measuring gender differences in personalities through natural language in the labor force: Application of the 5-Factor Model

This chapter is based on the article:

Title:	Measuring gender differences in personalities through natural language in the labor force: Application of the 5-Factor Model
Authors:	**Dania Eugenidis**: Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany; Contribution: 90%. **Dr. David Lenz**: Department of Econometrics and Statistics, Justus-Liebig-University, Gießen, Germany; Istari.Ai, Mannheim, Germany
Status:	Working Paper, submitted to *Journal of Computational Social Science*
Available at:	https://www.uni-marburg.de/en/fb02/research-groups/economics/macroeconomics/research/magks-joint-discussion-papers-in-economics

Earlier versions of this work have been presented at the following scientific conferences with review process:

- 16th International Conference on Computational and Financial Econometrics (CFE), December 2022, London, UK (Presentation).
- 24th International Conference on Computational Statistics (Compstat), August 2022, Bologna, Italy (Presentation).

Measuring gender differences in personalities through natural language in the labor force: Application of the 5-Factor Model

Abstract

Gender stereotypes still play a major role in the perception and representation of people in the workplace. Measuring the effects of those stereotypes quantitatively is very hard though. Traditional methods, such as questionnaires, struggle to provide the full picture, for example through misunderstanding, omission or incorrect answering of questions. However, evidence-based policy making requires accurate indicators of gender inequalities to promote equality. We present a framework measuring gender stereotypes on company level using publicly available big data. Specifically, we analyse the one million websites of all German companies using natural language processing with regard to differences in their portrayal of genders through the use of certain terms. We then contextualize the gender stereotype measures following the personality traits of the Five Factor Model and their sub levels. Statistical analysis of the results indicates significant stereotypes within personality traits for large portions of the sample. The qualitative differences in gender presentation are mostly consistent with those found in the literature, which serves as a validation for the presented framework. The presented approach complements traditional quantitative measurement techniques by capturing a mainly latent level of inequality. The fully automated and comprehensive analysis of the linguistic portrayal of gender stereotypes in a corporate context is at low cost, with little delay and at a granular basis.

Introduction

Gender inequalities in the labor market have already been identified in numerous studies. Current research finds, for example, that women are not accepted and promoted in certain professions in the same way as men. Official diversity management studies show that around 80% of the representative companies in Germany surveyed aim for a balanced gender ratio (Destatis, 2021a). However, when looking at the distribution among men and women on the management boards of the 200 largest German companies, the proportion of women was just under 15% in 2021 (Destatis, 2021b). Thus, among executive positions, the proportion was only 22.6% (Destatis, 2021c). A disproportionate number of women work in low-level positions, which is partly due to the fact

that with equivalent qualifications, the work of male employees is comparably higher valued (Etzkowitz & Ragna, 2010).

In the comprehensive current corporate context, quantitative studies investigating vertical gender segregation are mostly limited to specific dimensions, such as women's quotas. Structural differences in the representation of men and women have not been captured yet. We assume that fundamental differences in the perception of individuals and the associated assessment of their abilities and character traits are strongly dependent on gender and that this ultimately leads to the imbalances measured quantitatively. It is therefore of essential relevance to analyse this latent level of gender stereotypes in addition to quantitative measures in order to derive conclusions. For example, it can be assumed that companies that strive for gender equality may be more innovative or economically successful. In addition to pure performance-related figures, corporate social sustainability plays an important role in the current work society. Companies attaching importance to corporate social responsibility are found to be more economically successful (Orlitzky et al., 2003). Especially with regard to employee satisfaction, priorities are set in various areas in order to achieve an equal working environment for all social subgroups (Yoon & Chung, 2018).

Existing approaches measuring company level data in a psychological context include qualitative studies for individual companies.[1] However, typical data tools (such as questionnaires) face the problem of being very expensive, not generalizable and time-consuming. It is therefore not possible to obtain generally valid results in the short term or to track changes over time by repeatedly conducting the survey at high frequency. With our approach, it is possible to conduct an almost comprehensive and fully automated analysis of the linguistic portrayal of gender stereotypes in a corporate context. Our framework is not limited to individual sample companies, but is applicable nationwide for companies across all regions and sectors.

We focus specifically on stereotypes in the external presentation within the companies and consult the respective websites for this purpose. The latter represents an important communication channel between the company and external stakeholders such as customers, cooperation partners and potential employees. More than 80% of the total population in Germany and even 100% of 14-49 year-olds are considered regular internet users (Destatis, 2022a). 61% consult the internet for searching information about goods and services (Destatis, 2022b) and 95% made purchases online in 2021 (Destatis, 2022c). These figures make it clear that an appealing internet presence is enormously important

[1] Caliendo et al. (2014) measure market entry and exit conditions at the individual level in Germany, using survey data from the Socioeconomic Panel (GSOEP), a representative household survey. This is described in more detail in section two.

with regard to the general success of the company. In the external presentation, the websites are used to build up sympathy, to strengthen the trust of visitors and customers, to stand out from competitors and to communicate information efficiently and promptly. Employees are also introduced to the visitors, including their personal competencies and characteristics. Subconsciously, gender stereotypes might be transmitted to the public, which constitute the main subject of our analysis.
In this context, the following question is to be answered: Are there systematic differences in the (external) perception and presentation of men and women on corporate websites? Differences regarding predefined personal characteristics are to be quantified using an approach from organizational psychology, namely the Five Factor Model (FFM) of personalities (Costa & McCrae, 1991).
Since more recent studies (DeYoung et al., 2007; Weisberg et al., 2011) found gender differences within each factor that cannot be revealed when focusing only on this level, a subdivision into so-called aspects is conducted for each factor.
The remainder of this paper is structured as follows: In Section 2, a review of the relevant literature regarding gender inequality and the organizational psychological models is presented. In Section 3, we detail the data and methods used in our framework. Section 4 provides the main results for the German labor market which is subsequently compared to the literature in Section 5. Section 6 gives an additional insight into different dimensions of the labor market and examines the validity of our framework. Section 7 concludes and outlines future research.

2 Literature

2.1 Gender inequality in the labor market

As already described in the introduction, there is still a strong imbalance at management level in Germany to the detriment of women. Berglund and Thorslund (2012) document that women are disadvantaged in male-dominated industries and excluded from certain technical tasks. Employers assume that women do not have an equivalent level of technical understanding and assertiveness, which means that specific gender roles remain and are still part of the work structure (Berglund & Thorslund, 2012). Kanter (1977) has similar findings and adds that in gender-heterogeneous work groups women are often excluded in decision-making and development processes and treated as “tokens”, i.e. as representatives of their category rather than individuals. Only after a certain limit with members of a minority is reached, a so-called ”critical mass”, influence can be exerted by this subgroup (Kanter, 1977; Brahma et al., 2021).

These findings are consistent with those of Brader and Lewerenz (2006), who studied the gender distribution at different management levels of 16,000 mainly small and medium-sized German companies and relate it to data of the Glass Ceiling Index (GCI). This indicator quantifies the distribution of male and female low- and mid-level employees to the upper level. They report high scores for the GCI, caused by an imbalance in the share of women in high management positions compared to low/middle positions. This phenomenon is amplified as the size of the company increases. The tasks of the respective levels also differ greatly: the upper management level has a disproportionately large range of tasks and responsibilities compared to lower levels (Brader & Lewerenz, 2006). Discursively created social stereotypes explain the above-mentioned imbalances in the labor force that might further lead to the outcome of the GCI (Carli & Eagly, 2007).

Timmers et al. (2010) pursue a similar hypothesis, analysing the impact of 19 gender equality measures[2] at Dutch universities with the help of the GCI, where the upper level corresponds to full and associate professors. They report comparably lower scores and ceteris paribus less imbalances for universities with a high number of adapted measures and point to the effectiveness of the latter that way (Timmers et al., 2010).

Asymmetries in the employment of men and women with otherwise equal qualifications can be explained by discursively created social stereotypes (Carli & Eagly, 2007): Men tend to be assigned performance-related characteristics, according to which they are then described, while women are associated with adaptive and character-related characteristics. Inter alia, men are generally perceived as more assertive compared to women. This trait often tends to have a negative connotation for women, while being seen as positive for men (Weisberg et al., 2011). Therefore, the characteristics and skills of a successful manager sometimes do not match those of the classical female stereotype (Glick & Rudman, 2012), indicating that the work performance of two equally qualified individuals might be evaluated differently depending on their gender.
This is a well-analysed phenomenon in the field of psychology. Abele and Wojciszke (2007) divide human behavior into two categories, Agency and Communion. The former represents self-efficacy and individual aspiration with traits like friendly, empathetic, helpful, while the latter characterizes the desire to form and maintain interpersonal relationships with traits such as dynamic, self-confident and efficient. Pietraszkiewicz et al. (2018) base their study on this framework and develop dictionaries of approximately 400 terms that approximate a division into communal and agentic terms[3]. Both dictionaries are subsequently applied to language in job applications. They find that male-dominated jobs, i.e. jobs with a low share of women, are mainly advertised with agentic terms, while

[2]These include, for example, women in committees, women's networks or bonuses for hiring a woman.
[3]The authors build and evaluate internal consistency of their dictionaries from a mix of different data sets, such as blog posts, natural speech transcriptions and *New York Times* articles.

female-dominated ads are described with communal words.
Consistent to our hypothesis, gender-related differences in external perceptions and their impact on career trajectories can be explained by social and political unconscious constructs and beliefs, in addition to role-specific gender characteristics, finally explaining the results of the GCI in many disciplines. Men are typically associated with traits such as leadership and power, women with adaptive and non-performance traits. The latter might be perceived as unsuitable for leadership positions, as female stereotypes generally do not match the characteristics and skills of a successful leader (Glick & Rudman, 2012). This phenomenon will be discussed in the following.

2.2 Five Factor Model and subdivisions

Our framework aims to analyse the labor market with respect to gender equality perception and stereotypes by combining paradigms from classical organizational psychology with current text data analysis methods.
The theoretical basis for our framework is the FFM of personality. This model is the most widely used classification system for personality traits, identifying five major areas of personality: Openness/Intellect, Conscientiousness, Extraversion, Agreeableness and Neuroticism. These together form the acronym "OCEAN" (Costa & McCrae, 1992a; Digman, 1990; John & Srivastava, 1999). It is assumed that the majority of the variance of an individuals' personality can be explained on these scales, i.e. observable differences in patterns of thought, feeling and behavior. A wide range of traits is assigned to each factor with a description of each given below. The paradigm is universally applicable, regardless of age, gender or other socio-demographic attributes and the individual expressions of the traits usually do not change over time or environment (Costa & McCrae, 1992).

It should be noted that previous studies (e.g. Borghuis et al., 2017; Pietraszkiewicz et al. 2018) are based on questionnaires completed by individuals for themselves. Thus, only self-perception is analysed and measured. Our framework makes it possible to address the same question with completely different measurement instruments. The focus here is on the external representation of individuals caused by gender stereotypes. Yet, Costa and McCrae (1989) find in their questionnaire-based study that personality as perceived by others shows correlations up to 80% with self-assessment. Although there is still a 20% gap, the implications gained in this framework can create comparability and extensions to the literature.

Following the prevailing classical paradigm, suggestions for extensions on the five factors are established by DeYoung et al. (2007): They introduce an intermediate level,

where the factors are subsumed into so-called aspects, representing related but separate dimensions. With the help of a factor analysis, two aspects are assigned to each factor (DeYoung et al., 2007). These are represented in the following:

Openness/Intellect

DeYoung (2015) highlights the relevance of the Openness/Intellect factor, describing it as an empirically derived personality dimension that quantifies gender-related individual differences in the ability to recognize and evaluate complex patterns of information, i.e. cognitive exploration. What is to be measured as a matter of priority is the expression of innovative ability and imagination.[4] Social interaction, the making of tools and human goals require complex mechanisms for the cognitive exploration (DeYoung, 2015).
The differentiation takes place into the aspects of Openness and Intellect. While the former comprises facets with traits such as being artistic, aesthetic, perceptive, poetic or fantasy-prone, the latter describes characteristics like being intellectual, intelligent, philosophical and clever.
Weisberg et al. (2011) use data from the International Personality Item Pool (IPIP; Goldberg, 1999) for their study and find significantly greater average scores for Openness for women and significantly higher scores for Intellect for men. This is explained partly by the fact that men more often overestimate their own intelligence and tend to be comparatively less reserved, which makes them appear more intellectual. Women, on the other hand, are often more open regarding aesthetically oriented characteristics related to preoccupation with sensations and perceptions (for example art), which is reflected in a higher score in this aspect (Costa et al., 2001; DeYoung, 2020).

Conscientiousness

The ability for self-control and rule-following is expressed with the help of the factor Conscientiousness, which is subdivided into the aspects Orderliness and Industriousness (Feingold, 1994; Costa et al., 2001). While the former describes an individual tendency, the latter is seen as the result of an ordering process or behavior (DeYoung et al., 2007). Sax (2008) investigates the gender gap in academia by comparing the learning behavior of 17.000 students from 200 institutions and finds that women are more likely, inter alia, to take notes, seek feedback and finish homework and assignments. They achieve higher scores for Orderliness, such as with regard to the facets self-discipline or dutifulness.
For many aspects, significant differences and clear tendencies towards one gender can be identified, however, the results from previous studies are rather mixed for the aspect Industriousness. Costa et al. (2001) list cultural reasons that cause gender differences. In their cross-cultural study, Kusnierz et al. (2020) examine the academic success of male and female students in Poland and Ukraine. They find a higher level of internal and

[4] The author notes that the term "imagination" would probably be more appropriate than "openness". But since the acronym of the Big Five ("OCEAN") would have to be renamed, the latter is still in use.

external motivation for women (Kusnierz et al., 2020), which might subsequently translate to a higher level of industriousness. Feingolds (1994) results point into the same direction by reporting a higher level of self-discipline for women. This is contrasted by the results of Leon et al. (2017), who find higher scores for males when replicating Weisberg et al.'s (2011) study with a Peruvian sample.

Extraversion

According to DeYoung et al. (2007), Extraversion is subdivided into Assertiveness and Enthusiasm. Assertiveness is derived from Depue and Collins' (1999) definition of agency and describes the tendency of a social dominance and leadership role and a distinct ambition to achieve goals.

Contrary to this, Enthusiasm is linked to the trait of associating positive feelings with the expectation of rewards. Social interaction is perceived as rewarding, which is conducive to social behavior and ceteris paribus to extraversion (Depue & Collins, 1999).

Higher average scores are found for women for Enthusiasm, whereas men score higher for Assertiveness (Weisberg et al., 2011; Feingold, 1994; Costa et al., 2001). Women tend to score higher overall at the factor level, however, certain character traits, such as dominance are measured as a part of Assertiveness in which men score higher (Helgeson & Fritz, 1999).

This is in line with the cross-cultural analysis of Williams et al. (1999) who requested college students of 25 countries to assign certain character traits to the respective gender. The authors report a comparably higher level of Extraversion for women in their study.[5]

Agreeableness

Agreeableness describes character traits associated with empathy, altruism and niceness. Agreeable people are willing to compromise and are in need of harmony. In addition, the degree of consideration and respect for others' desires and needs is quantified here. Based on this, the subdivision takes place into the aspects Compassion and Politeness (DeYoung et al., 2007). Women are found to achieve significantly higher average scores in both aspects (Weisberg et al., 2011).

Costa et al. (2001) cite biological origins, explaining that evolutionarily, stronger bonds prevailed between women and children, which is why they are more agreeable and nurturing (Costa et al., 2001).

Neuroticism Neuroticism explains the tendency to react to negative emotions, associated processes and punishment in response to a perceived threat. This manifests itself, for example, in depression or emotional instability. The two aspects are labelled Volatility and Withdrawal. Volatility describes, among other things, emotional instability or

[5]The method they used is described in section 3 in more detail.

difficulties in impulse control, addressing the expressions of emotions. Withdrawal is reflected by susceptibility to a class of negative affect, capturing intrinsic feelings (Weisberg et al., 2011). Women show significantly higher scores for most facets whereas men score higher only in the case of anger in the questionnaire-based literature (Costa et al., 2001). Regardless of gender, it is assumed that companies do not describe their employees by traits indicative of a neurotic personality. Since no reasonable results are expected, we need to adjust this negatively connotated factor. That is, Neuroticism is replaced by Emotional Stability (Caliendo et al., 2014). Emotionally stable people are characterized as being self-confident, relaxed and robust towards stressful situations. Completing and managing daily tasks, as well as managing stress and uncertainty does not pose a problem to them (Caliendo et al., 2014).
To the best of our knowledge, no literature exists on the aspect level yet, which is why this study suggests a replacement of the aspect Volatility by Consistency. In contrast to volatile individuals, consistent people have control over their emotions and are less likely to experience mood swings. The second aspect Withdrawal is replaced by "Robustness" here. Individuals with a high expression of this aspect rarely feel sad, anxious or overwhelmed.[6] According to Weisberg et al. (2011), women score higher for Volatility as well as Withdrawal on average. Using the negative specification of the two aspects, this would mean that men should score higher.[7]

Table 1 summarizes the results from the literature with the respective dominant gender for each aspect. While the paradigm of the FFM is used standardly on an individual

Table 1: Literature overview of the dominant gender per aspect

Aspect	Dominating mean (Literature)
Openness	W (Weisberg et al., 2011; Costa et al., 2001; DeYoung, 2020)
Intellect	M (Weisberg et al., 2011; Costa et al., 2001; DeYoung, 2020)
Orderliness	W (Weisberg et al., 2011; Sax, 2008)
Industriousness	M (Weisberg et al., 2011; Leon et al., 2017) W (Kusnierz et al., 2020; Feingold, 1994)
Assertiveness	M (Weisberg et al., 2011; Feingold, 1994; Costa et al., 2001; Helgeson & Fritz, 1999)
Enthusiasm	W (Weisberg et al., 2011; Williams et al., 1999; Helgeson & Fritz, 1999)
Compassion	W (Weisberg et al., 2011; Costa et al., 2001)
Politeness	W (Weisberg et al., 2011)
Volatility	W (Weisberg et al., 2011; Costa et al., 2001)
Withdrawal	W (Weisberg et al., 2011; Costa et al., 2001)

[6] The terms of the opposite specification introduced here were determined according to the closest orientation to the individual items from IPIP.

[7] "W" means that women score higher on average, whereas "M" depicts the cases with higher average male scores.

level in psychological research as well as in the field of personnel economics, its application in an economic-quantitative context on a firm level in very large samples (both in terms of the number of firms and in terms of the number of employees considered) has not been done so far. The technical implementation is described hereafter.

3 Data and Method

There are approximately 1.3 million companies economically active in Germany that are listed in the all-encompassing panel company database of the "Mannheimer Unternehmenspanel" (Mannheim Enterprise Panel; MUP). A detailed description of this all-encompassing database can be found in Bersch et al. (2014). In addition to official government firm statistics, this represents the largest company database in Germany. Updated every six months, it provides official company data, including the respective web addresses, i.e. URLs. The URLs are used to call up the web pages and to extract the information and texts that serve as the basis for our analysis. Since not every company provides a web presence, the number of relevant firms is reduced to about 910,000. Our analysis is limited to texts using the German language. All web pages are available in German, yet the English versions are either congruent or more sparsely available in terms of content, making the additional inclusion obsolete.

The assignment of gender to a person follows the procedure of Eugenidis et al. (2022): To identify employees and persons named on the website, each word listed on the website is checked to see if it is a first name and, if so, which gender that first name is assigned to. In addition to the first name search, a keyword search is also performed with terms like "Mr." and "Ms."/"Mrs.", covering the cases that persons are not listed by their first name. A subsequent analysis of the surnames assures that the word found as a person is indeed a physical person rather than a street or product name, for example. Unfortunately, unlike the first names, there is no complete list of surnames. For this reason, only words beginning with an uppercase letter followed by only lowercase letters are recognized as surnames. Usually, first names are followed by verbs, surnames or punctuation marks - thus, our technique filters out all irrelevant cases. Acronyms, which usually have additional capital letters within the word, are also identified in this way. This makes it possible to identify the firm level distribution without manually analyzing each firm individually.

In order to be able to structurally and specifically analyze the external representation with the associated gender stereotyping, dictionaries based on the ten aspects for the FFM are constructed. In the textual context of the person, descriptions are searched for according to the aspects that can be assigned to this person explicitly. The assignment

to each person on the website is conducted by determining a maximum distance of four words from each word recognized as a name. The aspect term in this environment (i.e. a word from one of the word lists) is assumed to be referencing to the person. For example, "Mr. Müller is a very assertive employee": "Mr. Müller" would be recognized as a name, and the adjective "assertive" as a term of the aspect Assertiveness.

We include all synonyms and the superlative forms for each aspect as the tone may vary from company to company. "Industrious", for example, can also be referred to adjectives such as "proficient" or to nouns like "proficiency". For this purpose, we use the item list of the aspects scales as a basis. For each aspect, ten items are questioned that describe the character facets. For example, "I understand things quickly" within the Intellect aspect describes astuteness or alertness. For the two aspects in the Neuroticism factor, the opposite traits were used. The facet "I get angry easily" is not measured by irascibility, but by calmness or meekness. In addition to a manual determination of all relevant terms, the Word2Vec method (Mikolov et al., 2013) is additionally used to identify words with semantically high similarity . This results in 758 terms (including the superlatives) in ten dictionaries in total, of which we were able to match 482 terms across all companies and websites. The full list of words can be found in the appendix, table A1. That is, in each of these cases, a word recognized as a name is in close proximity.

Our method might be associated with measurement inaccuracy, since character traits that are not in immediate proximity to the name cannot be detected. For example, "Mr. Müller" could still be described as "likeable" in a further sentence, but this would not be recorded. Algorithm-based analysis methods from the field of co-reference resolution could therefore replace the buzzword search as a central method in the future. This approach is used to identify statements in the text that refer to specific entities, such as people (Clark & Menning, 2016). In this way, from the original 910,000 companies, 279,000 companies with 193,000 men and 105,000 women are suitable for our dataset, resulting in a women's quota of 35.23%. The gender distribution is only marginally affected by this procedure. Looking at the distribution over all persons without including the aspect terms, it is 32.39%.

This reduction is firstly due to the fact that physical persons including gender cannot be clearly identified, or secondly that persons are identified but no character trait can be assigned to them. In this framework, only those cases for which all this is true are relevant, which explains the large reduction of firms. Exceptional cases where the gender could not be assigned without any doubt are excluded. This refers, for example, to "Kim" and "Michelle", both of which can be male or female first names.

It should be noted that not all companies perform public employee descriptions with the same resources. Large companies may bring in professionals (i.e. marketing or HR) since they have the resources and knowledge to ensure the best possible representation, which might not be the case for smaller firms. The corresponding analysis of this dimension takes place in section 6.1.3.
The concept of classifying adjectives according to gender is already introduced in a study by Williams and Best (1982). In their cross-cultural international study, participants were asked to assign adjectives to one of the two genders, or to indicate gender neutrality. From this, they then calculated the proportion of terms that were more likely to be associated with each gender. They found some individual cultural country-specific differences in the characteristics of how men and women are described, but across countries a high level of similarity was found (Williams & Best, 1982).
Once all characteristics are assigned to the individuals, the mentions of each term per gender can be recorded. Classic two-sample z-tests for testing equal proportions are used to examine whether there are statistically significant differences between men and women within the aspects of the FFM[8] (Wilson, 1927). We conduct this test for each term and aspect for all firms and across sectors, firm sizes and regions, assuming that differences exist at different dimensions.

Since the two populations are not equal in size, with about twice as many findings identified as male names, a normalization of the data is performed in the course of the analysis. For this purpose, the number of hits per word is divided by the total number of individuals of the respective gender.
A generalization of the methodological approach in the direction of an intersectional view is also possible. For instance, differences may not only exist between men and women but a migration background might also play a role when analysing equality in the labor market. Using 1000 job openings with 5 applicants per job from the Finnish national employment service, Ahmad (2020) finds that employers significantly prefer Finnish applicants over foreign candidates, and within them, they prefer candidates with a European name over a non-European name. This effect is strengthened by gender: Minority males are consistently subject to a much larger degree of discrimination than minority females (Dahl & Krog, 2018). Based on name recognition of foreign and domestic names, one could perform a similar analysis.[9]

[8]Men and women are considered as independent populations following a binomial distribution. The null hypothesis states that there are no gender differences between the proportion values for each term between the populations. The alternative states the opposite, namely significant differences between men and women when using specific terms to describe a personality.

[9]Since in this study we would like to introduce the new concept first and apply it to the dimension "gender," it appears more appropriate to apply other social demographic dimensions in more advanced studies in the future.

4 Main findings

This section provides a description of the main results of our analysis, namely the findings of whether there exists a significant difference in the way men and women are described using the terms of our aspect dictionaries within the FFM. The interpretation of the results and the inclusion to the findings of the literature is provided in the subsequent section 5.

On the level of individual terms, significant differences are found in 252 of 482 cases (in 30 cases at the 10% level, 54 cases at the 5% level, and 168 cases at the 1% level). Hence, our results support the hypothesis that certain terms and, ceteris paribus, certain character traits are assigned to individuals on the basis of gender.

When comparing the hits of all 482 terms for men and women, the largest differences in proportion scores in favour of men are found for "discrete" with 21 times more occurrences than for women,[10] "progressive" with 20 times more hits for men and "self-confident" with 17 times. The traits used most often for women are "friendly" with 40 times more hits, "easy-going " 20 times and "artistic" with 7 times more hits for women than for men. All differences are significant. These findings provide a first impression about possible systematic differences in the description and perception of people driven by their gender. The findings are in line with those of Abele and Wojciszke (2007), who identify words such as "self-confident" as an agentic (i.e. masculine) trait and "friendly" as a communal (i.e. feminine) trait.

Table 2 details the results for the associated terms of each aspect with significance levels given in parentheses. The first two columns depict the number of terms per aspect for which more hits are found for men and women respectively. For example, within the aspect Assertiveness, men are able to achieve a higher score in 30 cases, 16 of which are significant (in eleven cases at the 1% level, once at 5% and four times at the 10% level). Women are able to score higher in 16 cases, yet significance has to be rejected in ten cases (in contrast to significant results twice for the 10%, 5% and 1% level, respectively).

The third column indicates which of the two genders[11] has a higher mean value across all terms per aspect. The first letter indicates the gender that shows a higher mean across all terms of an aspect in our sample, while the letter in parentheses displays the

[10] 39 men and one woman are described by this term in absolute occurrences. To ensure comparability between both populations, the ratios are formed using the normalized values. Absolute values of the remaining terms can be provided by the authors upon request.

[11] "W" means that women score higher on average, whereas "M" depicts the cases with higher average male scores.

results from the literature described in table 1.

Table 2: Results of the two-sample z-test for equality of proportions

	Male dominated ($p < 0.01$; $p < 0.05$; $p < 0.1$; $p > 0.1$)	Female dominated ($p < 0.01$; $p < 0.05$; $p < 0.1$; $p > 0.1$)	Dominating mean (estimation)
Assertiveness	30 (11; 1; 4; 14)	16 (2; 2; 2; 10)	M (M) ($p < 0.01$)
Enthusiasm	36 (15; 4; 1; 16)	36 (10; 9; 2; 15)	W (W) ($p < 0.01$)
Industriousness	18 (7; 4; 0; 7)	16 (1; 1; 1; 13)	M (M/W) ($p < 0.01$)
Orderliness	28 (13; 3; 2; 10)	9 (4; 0; 0; 5)	M (W) ($p < 0.01$)
Openness	20 (7; 1; 1; 11)	27 (7; 1; 3; 16)	W (W) ($p < 0.01$)
Intellect	45 (23; 9; 1; 12)	16 (5; 2; 0; 9)	M (M) ($p < 0.01$)
Compassion	17 (6; 2; 2; 7)	17 (9; 1; 1; 6)	W (W) ($p < 0.01$)
Politeness	42 (12; 6; 1; 23)	39 (12; 3; 2; 22)	W (W) ($p < 0.01$)
Consistency	8 (4; 1; 1; 2)	12 (5; 2; 0; 5)	W ($p < 0.01$)
Robustness	27 (9; 1; 3; 14)	23 (6; 1; 3; 13)	M ($p < 0.01$)

4.1 Term Level

Table 2 points to the necessity for the subdivision of the factors since the dominant gender differs between the aspects in some cases. The findings of our sample thus support the initial hypothesis of within-factor gender differences.
Regarding Openness/Intellect, for example, there are comparably more terms for Openness found for women. Yet the opposite is true for Intellect. In this case, if only the factor level were examined, it would not be possible to identify these gender divergences. Our results are qualitatively consistent with those of Weisberg et al. (2011), who describe women as more open, whereas men are more intellectual.

The result for Industriousness within Conscientiousness is showing a similar tendency for both genders, pointing slightly towards a dominance of men. The majority with 18 of the 34 terms are used more frequently for men, in addition, the findings for women are insignificant in almost all cases. The second aspect shows similar results: Orderliness is a trait that is clearly assigned to men, with 28 out of 37 terms in our sample.
Whereas within the factor Extraversion Assertiveness is a trait clearly attributable to men, no tendency can be identified for Enthusiasm when examining the individual terms. Each gender dominates for half of the terms, and the number of significant differences is high and almost the same for both genders.
The same can be observed with regard to Compassion in the factor Agreeableness. With 17 terms each, too, the distribution in terms of gender dominance is approximately the same. However, what contradicts the results of the literature (for example Weisberg et al., 2011) are the findings for Politeness. Overall, there is a high number of (significant)

terms for both genders with a slight tendency towards men's dominance.
Lastly, the results for Emotional Stability with Consistency and Robustness show a female dominance for the former and male dominance for the latter, indicating the existence of a within-factor difference which is revealed by the breakdown into aspects. Emotional Stability specified as the opposite of Neuroticism and should therefore also show correspondingly opposite results. Consistency describes the expression via emotion and indicates higher average female dominance. This is in contrast to the relevant literature, in which women achieve a higher score for Volatility as well. Robustness describing intrinsic emotions unveils higher scores for men for a higher share of terms. This is consistent to our expectations.

4.2 Aspect Level

As stated in section 2, several studies find higher mean values for men for the aspects Assertiveness and Intellect and for women for Enthusiasm, Orderliness, Openness, Compassion, Politeness, Volatility and Withdrawal. Overall, seven of the nine results are consistent with those from the literature, namely for the aspects Openness, Intellect, Assertiveness, Enthusiasm, Compassion, Politeness and Robustness. All test statistics are significant. Our results regarding the Orderliness and Consistency aspects differ from those in the literature and favour the opposite gender, possible reasons for this are discussed in section five. Industriousness represents a special case here, as there are no clear results regarding gender dominance in preceding studies. As described earlier, this trait is found to be stronger represented for males in one part of the studies and for females in another part.
Although more male-dominant hits are found for the aspect Politeness, women are able to achieve a higher average score. The significance of the individual terms depends on the size of an absolute difference in the population and on the number of observations. The chance of obtaining a significant result is higher for these cases. The difference in the aspect might therefore be significant because one adjective has particularly many observations with a significant difference.
Looking at the distribution in the usage across all terms of each aspect, men are most often described with terms of the aspects Orderliness, Assertiveness, and Intellect, while women are mainly characterized with Compassion, Politeness, and Openness, which is consistent with expectations from the literature to a large extent.

5. Discussion of the results

As already described in section 4, consistent results to the self-assessing literature are found in seven out of nine aspect cases. It should be noted, however, that a direct comparison is of limited value since we are examining the external presentation. Nevertheless, it can be seen as much more of a validation of our results.

Industriousness

We find insignificant differences for a majority of the cases, which is in line with the mixed results from the literature. Consistent and valid with past studies, therefore, no clear tendency can be identified with respect to one gender. Our data indicates a higher average score for men, which contradicts to the results of Kusnierz et al. (2020). They document a greater willingness to learn due to higher intrinsic motivation, which results in better average academic performance. Keiser et al. (2016) find similar results and attribute different university performances with equal prerequisites between the genders to the fact that women overall invest more effort in their academic education (Keiser et al., 2016). This contrasts with the results of Weisberg et al. (2011), who report a higher but insignificant value for men. Possible reasons are not discussed further, presumably because the effect is relatively small and insignificant.

Orderliness

Our result for Orderliness and a male dominance differs to those from the literature. Weisberg et al. (2011) report higher (insignificant) scores for women and base their results on data from North America.

Casimir and Waldmann (2007) explain these differences by describing Orderliness as being one of the traits that is culture-dependent, that is differences between studies may arise due to this impact. Schneider and Littrell (2003) documents the character trait of orderliness (along with planning reliability, punctuality and accuracy) to be comparatively highly valued among German managers. Characteristics such as creativity, on the other hand, are hardly required. One of the main tasks of a German manager is to establish and maintain order and reduce uncertainty. German corporate governance is comparatively long-term oriented, requiring a greater degree of planning, orderliness and control. This is opposed, for example, to the English leadership style, which focuses primarily on control after the actual occurrence of uncertainties (Schneider & Littrell, 2003).

In the German language, grammatically masculine role nouns exist for job titles such as ”manager” or ”managing director” that can be used for both women and men, but often only consider men when interpreted. This creates biases in favor of men, which can reinforce gender inequalities in the labor market (Misersky et al., 2019). Since there

are more male managers at the executive level, the character trait Orderliness as a core competence of a leader tends to be assigned to men.

Consistency

Unfortunately, there is no literature describing the opposite specification of the aspect Volatility, a consistent categorization to the literature is only possible and meaningful to a limited extent. We assume consistency to be a trait that incudes the control of emotions and moods and opposing volatility, therefore the opposite gender should also dominate. Weisberg et al. (2011) find a higher score for women for Volatility, which is contrary to our results, which suggest female dominance for Consistency. Women are generally considered to be more emotional than men, which is expressed in more volatile behavior (see section two). This is in line with Borghuis et al. (2017) who investigate changes in the five factors for Dutch adolescents and young adults and find temporal dips and a lower degree of emotional stability for girls.

In the following, possible reasons causing diverging results in comparison to the relevant literature are elaborated: Firstly, differences may be due to the different methods used. We exclusively utilize quantitative text analysis-based website data compared to questionnaire-based data to measure gender stereotypes. Also, it should be pointed out again that our model refers to external perceptions, whereas the results of previous studies were obtained mainly on the basis of self-perceptions. Secondly, structural social differences may also impact the outcome. Weisberg et al's (2011) analysis is based on a data set with participants from Canada and the United States. The social structures and dynamics in a foreign labor market may not be comparable to those from Germany, which is considered to be significantly less flexible and dynamic than the U.S. labor market (Jakob & Weiss, 2008).

The third factor that might play a role is language difference and the associated influence in cognition which possibly affects the perception of others or oneself. Different national languages arise from a process of social development (Neis, 2003). The outcome might therefore be highly dependable of the language. Not only the variety of cultures might have an influence, but also the structure of the language and the connection of the relations of all traits at a low level (Saucier et al., 2005). Language thinking also varies between different societies and possibly influences outcomes like the way people are described. Psychological and neuroscientific research shows that language influences not only communication but also brain activity related to behavior, emotions and emotional regulation (Lindquist, 2021). Lastly, as already mentioned in section two, Costa et al. (2001) list biological and cultural reasons to cause gender differences. It might be misleading to argue for biological reasons, since then the relevant differences would be the same across different cultures (Costa et al., 2001). However, the authors note

that results are not consistent across cultures, thus there may be differences due to the different languages in addition, in line with Saucier (2005). Kusnierz et al. (2020) support this hypothesis and point out that the FFM is cross-culturally applicable, but the results are dependent on the sample.

6 Additional results and validation

In addition to the overall labor market, our data set provides information for individual analyses of regions, firm sizes and sectors. It is to be assumed that the quality and choice of words differs across firms. The resources available to companies to design their website and thus external presentation vary. In addition, regional language differences and dialects may also play a role.[12] We examine the validity of our framework and our findings with results from previous studies and provide a deeper insight into these dimensions. Labor market related studies, such as Haltiwanger et al. (2014), are typically subdivided into these dimensions. It should be mentioned that the normalization now takes place at the corresponding dimension, rather than the aggregate level over all firms.

6.1.1 Sectors

The distribution among men and women is in part very heterogeneous between the various sectors in Germany. While sectors such as healthcare are comparatively balanced, construction and mechanical engineering are still male-dominated. The first column in table 3 presents the distributions of official data on women's quotas[13] ("Official quota"; Destatis, 2022d). Only those sectors are listed here that are available with both official data and our data set and are thus directly comparable. The second and third column depict the quotes based on our sample. One quota is formed based on all names mentioned online ("website quota"), not filtering out those cases for which no aspect term could be found. Another quota is calculated on the basis of the names with traits assigned ("Website quota including traits").

Two things are noticeable when comparing the distributions: first, our sample is qualitatively similarly distributed, with a high women's quota in education and health care and relatively low values in construction and mechanical engineering. Second, the quotas we calculate are substantially higher in absolute terms in most cases. This can be

[12] For an overview of regional linguistic variations see for example Ammon et al. (2014).

[13] These are the proportions of women in management positions in Germany in 2022. Unfortunately, no official data exists at the company level, as can be found in our sample. Primarily, a qualitative comparison is to be conducted, in such a way that similar and opposing tendencies can be identified.

Table 3: Women's quotas by sector (in percent)

Sector	Official quota	Website quota*	Website quota including traits
Education	73.4**	40.89	50.00
Health	36.90	47.41	56.33
Retail	26.20	32.1	30.62
Publishing/Media	24.40	30.14	31.80
Total	**24.10**	**25.57**	**31.28**
Real Estate	22.10	29.57	34.55
Logistics/Transportation	19.60	22.27	29.51
Manufacturing	18.30	23.95	35.95
Pharma	17.60	29.46	25.35
Financial services	17.50	21.53	26.25
Agriculture and forestry	16.70	30.3	28.51
(mech.) engineering	9.80	14.45	26.07
Construction	9.70	14.73	18.61

*: Estimates of the remaining MUP classification sectors according are not listed here, however, can be provided from the authors upon request.
**: Official data on the education sector in Germany only include the proportion of female teachers in general education schools (Destatis, 2022e). This does not include, for example, student advisors, research assistants, etc., and is therefore only representative to a limited extent.

attributed to the fact that our data set covers the entire gender distribution of a firm, while official data are only available for the share of women in managerial positions.[14] This supports the findings of the literature, for example on the GCI (see section 2), which shows that women are underrepresented in senior management overall.
When looking at the most frequently occurring word per sector, "competent" (in Intellect) is most frequently used for men in eight out of 43 sectors, "friendly" (Politeness) in seven sectors and "strong" (Assertiveness) in three sectors. For women, "cordial" (Enthusiasm) is found most often in 14 sectors, followed by "friendly" and "social" (both Enthusiasm) in four and three cases, respectively. Again, this actuates Abele and Wojciszke's (2007) theory that men tend to be described in agentic terms and women in communal terms.

Gender differences in different sectors are also examined in Bergmann et al. (2018). According to this study, institutional conditions influence income opportunities. Public sectors with a high women's quota, such as healthcare, have income disparities, while private sectors with a balanced gender distribution have smaller gaps. In addition, pay mechanisms such as bonuses and pay rates are mostly managed on an individual basis and the authors find that this is to the disadvantage of women and further increases pay disparities. Information on wage differentials is unfortunately not given from the website data, but our results are consistent and complementary to those of Bergmann

[14]Some companies do not provide information on sector or location. These companies are therefore not included in the calculation and hence the women's quota differs to the one in section 3.

et. al (2018): Healthcare has the highest women's quota at 56% (6212 firms), while finance has a much lower rate at 26% (3596 firms).
Looking at the individual aspects, men in the healthcare sector are comparatively most often characterized with the aspects Orderliness, Robustness and Assertiveness, while women are described as enthusiastic, polite and consistent. The trend is similar in the financial sector, where women are most often perceived as enthusiastic, industrious and polite. Men, on the other hand, are comparatively considered assertive, orderly, but also consistent. The latter differs from the results in the healthcare sector. This can be explained by the fact that in this male-dominated industry, a lot of monetary responsibility lies in the hands of employees and controlled, non-volatile behavior and appearance is crucial for trust in the employee by external parties.
The divergent results for Consistency in the two sectors show that an analysis that is not aggregated across all firms is important for understanding the role of gender in the German labor market: Overall, in the health care sector, a dominance of women over men is found in only three out of ten aspects, whereas in the financial sector it is found for five of the ten aspects. This (private) sector shows greater efforts to reduce gender inequality and to achieve equalization over time, for example also visible through the reduction the gender pay gap: While the gap in the financial sector was still just over 30% in 2010, a reduction of six percentage points to 24.2% was recorded in 2018, while in the health sector it fell by only 0.62 percentage points from 24.6% to 23.98% (Eurostat LFS, 2010 & 2018). Accordingly, Bergmann et al.'s (2018) hypothesis regarding structural differences in different sectors fostering imbalances can be confirmed.

6.1.2 Regions

The German labor market is characterized with regional heterogeneities, partly caused by the former separation into East and West. This is also noted in Dudel et al. (2021): Eastern[15] German women are more likely to work full time, are less likely to work part time or to be a housewife. Motherhood therefore has comparatively a greater negative effect on employment in the former Western states. Our data confirms these findings for the most part: We find a marginally higher women's quota in the Eastern states (29.4% vs 28.27% in the West). The state with the highest womens's quota, Mecklenburg-Western Pomerania (1493 firms), is located in the East, while the two states Bremen (910) and Saarland (1038), each with the lowest rate, are in the West. When comparing the choice of words between Eastern and Western firms, no significant differences could be found.
In Mecklenburg-Western Pomerania, the state with the highest women's quota, terms

[15]The federal states from the former GDR are East Berlin, Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, Thuringia.

relating to Intellect, Assertiveness and Openness are mainly used for men, while women are described as having enthusiastic, industrious and robust characteristics. In Saarland, men are named as industrious, assertive, and open-minded, while women are robust, consistent, and intellectual. The last result is unusual, since this characteristic could previously always be assigned to men. However, when comparing the choice of words between these two states, no significant differences could be found.

6.1.3 Firm Size

Finally, we examine the dimension of company size. Overall, companies are classified into four groups depending on the number of employees. The gender distribution is very similar across all company groups. The comparatively highest proportion of women is realized in the group of the largest companies (¿250 employees) with 38%, followed by the second largest group (36.6%, 50-250 employees), the smallest companies (35.7%; 1-10 employees) and finally the group with 11-49 employees, 34.3%. Accordingly, the differences in the women's quotas are so small that it can be assumed that company size does not play a role.
As already stated in section 3, it is to be assumed that the quality and choice of words differs depending on firm size. We checked for possible differences in the probability of (normalized) occurrences of terms for small firms (1-10 employees) and large firms (>250 employees) using a two-sample z-test. and conclude that there are no significant differences in the probability of occurrence of a term between small and large firms. Firm size therefore does not play a role regarding quality and choice of words.

6.2 Validation

Section two presented Abele and Wojciszke's (2007) model for describing persons, according to which individuals can be classified as being agentic or communal. Uchronski (2010) collects a total of 69 characteristic words that are suitable for both fundamental dimensions in terms of their representativity. 33 words form the respective dictionaries, yet three terms are not suitable for validating our framework due to their negative connotation.

To check the validity of our estimation, we examined which of the two genders dominates in our sample for each term and compared this to Uchronski's (2010) lists.
Agentic traits are mainly assigned to men in the literature; in our sample we could find a total of ten matches to the 33 trait terms defined by Uchronskis (2010). In six cases we find a male dominance, three of them significant; in the remaining four cases

women score higher, two times significant. On aspect level, two words each come from the aspects Robustness, Orderliness, Industriousness and Assertiveness and one word each from Intellect and Enthusiasm. If we exclude the rather ambiguous findings for Industriousness, all results besides Enthusiasm are consistent with our previous results and a male dominance (see Table 2).
Communal traits are found primarily among women, with an overlap in twelve of 33 cases. Nine times women scored higher (six of them significantly), the remaining three times men, with one significant case. Four words are found in each of the aspects Compassion and Enthusiasm, two in Politeness, and one each in Consistency and Orderliness. With the exception of the last two aspects, these all represent cases where the literature and our results are able to identify female dominance.

Conclusion

We presented a framework to conduct fully automated and highly granular measurements of gender stereotypes based on external perception via publicly available big data. We then applied this framework to the landscape of German companies in a pilot study for the domestic labor market.

The high innovative potential lies in the application of the website data in combination with official firm data and the application of our algorithm. With the presented framework it is possible to analyse and quantify external gender perception in an established psychological framework (FFM) and the corresponding sub dimensions.

The dimension of gender and the associated inequality in the labor market has already been studied on different quantitative and qualitative levels (see, section 2), but not yet in a framework as ours. The whole spectrum of companies can be included, which makes general and all-encompassing statements possible and also enables a comparison between all sectors and regions. We focus exclusively on external perceptions and representations by others, which allows us to uncover possible gender stereotyping.

Our study offers concrete evidence that firms still display gender stereotypical patterns today. With the help of our results, targeted measures can be implemented to reduce vertical gender segregation through language, i.e. through advice and tips on gender-neutral terms. As already described in section 5, distortions due to grammatically masculine role nouns might discriminate against women in terms of pay or promotion (Misersky et al., 2019).

In conclusion, a more in-depth sociocultural and linguistic analysis might necessary in

order to identify all relevant facets of a working society. Our analysis here is intended to serve as a cornerstone for further research in this area.

It must also be considered that there may be a self-selection bias, that women are more likely to select occupations in sectors such as education and men are more likely to select occupations such as engineering that require certain character traits, and that the differences are caused by specific jobs and not by gender.

References

Abele, Andrea E., and Bogdan Wojciszke. 2007. "Agency and communion from the perspective of self versus others." Journal of personality and social psychology 93(5): 751. https://doi.org/10.1037/0022-3514.93.5.751.

Ahmad, Akhlaq. 2020. "When the name matters: An experimental investigation of ethnic discrimination in the Finnish labor market." Sociological Inquiry 90(3): 468-496. https://doi.org/10.1111/soin.12276.

Ammon, Ullrich, Hans Bickel, Jakob Ebner, Ruth Esterhammer, Markus Gasser, Lorenz Hofer, Birte Kellermeier-Rehbein, Heinrich Löffler, Doris Mangott, Hans Moser, Robert Schläpfer, Michael Schloßmacher, Regula Schmidlin and Günter Vallaster. 2011. Variantenwörterbuch des Deutschen: Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol.“ Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110905816

Berglund, Karin, and Jennie Granat Thorslund. 2012. "Innovative policies? Entrepreneurship and innovation policy from a gender perspective." In Promoting Innovation: Policies, practices and procedure, edited by Andersson, Susanne, Karin Berglund, Ewa Gunnarsson and Elisabeth Sundin, 25-46. Vinnova: Stockholm.

Bergmann, Nadja, Alexandra Scheele, and Claudia Sorger. 2019. "Variations of the same? A sectoral analysis of the gender pay gap in Germany and Austria." Gender, Work & Organization 26(5): 668-687. https://doi.org/10.1111/gwao.12299.

Bersch, Johannes, Sandra Gottschalk, Bettina Mueller, and Michaela Nifert. 2014."The Mannheim Enterprise Panel (MUP) and Firm Statistics for Germany." ZEW-Centre for European Economic Research Discussion Paper 14-104.
http://dx.doi.org/10.2139/ssrn.2548385.

Borghuis, Jeroen, Jaap J.A. Denissen, Daniel Oberski, Klaas Sijtsma, Wim H.J. Meeus,

Susan Branje, Hans M. Koot, and Wiebke Bleidorn. 2017. "Big Five personality stability, change, and codevelopment across adolescence and early adulthood." Journal of Personality and Social Psychology 113(4): 641. https://doi.org/10.1037/pspp0000138.

Brader, D., and J. Lewerenz. 2006. Frauen in Führungspositionen: An der Spitze ist die Luft dünn. IAB-Kurzbericht, http://hdl.handle.net/10419/158209

Brahma, Sanjukta, Chioma Nwafor, and Agyenim Boateng. 2021."Board gender diversity and firm performance: The UK evidence." International Journal of Finance Economics 26(4): 5704-5719. https://doi.org/10.1002/ijfe.2089

Caliendo, Marco, Frank Fossen, and Alexander S. Kritikos. 2014. "Personality characteristics and the decisions to become and stay self-employed." Small Business Economics 42(4): 787-814. https://doi.org/10.1007/s11187-013-9514-8

Carli, Linda L., and Alice H. Eagly. 2008. "Through the Labyrinth: The Truth about How Women Become Leaders". Administrative Science Quarterly. 53(2): 363-366. doi:10.2189/asqu.53.2.363

Clark, Kevin, and Christopher D. Manning. 2016. "Improving coreference resolution by learning entity-level distributed representations." arXiv preprint arXiv:1606.01323. https://doi.org/10.48550/arXiv.1606.01323

Costa, Paul T., and Robert R. McCrae. 1989. "Rotation to maximize the construct validity of factors in the NEO Personality Inventory." Multivariate Behavioral Research 24(1): 107-124. https://doi.org/10.1207/s15327906mbr2401_7

Costa, Paul T., and Robert R. McCrae. 1991. "The NEO Personality Inventory: Using the five-factor modeI in counseling." Journal of Counseling & Development 69(4): 367-372. https://doi.org/10.1002/j.1556-6676.1991.tb01524.x

Costa, Paul T., and Robert R. McCrae. 1992. "The five-factor model of personality and its relevance to personality disorders." Journal of personality disorders 6: 343-343. https://doi.org/10.1521/pedi.1992.6.4.343

Costa Jr, Paul T., Antonio Terracciano, and Robert R. McCrae. 2001. "Gender differences in personality traits across cultures: robust and surprising findings." Journal of personality and social psychology 81(2): 322. https://doi.org/10.1037/0022-3514.81.2.322.

Dahl, Malte, and Niels Krog. 2018."Experimental evidence of discrimination in the

labour market: Intersections between ethnicity, gender, and socio-economic status." European Sociological Review 34(4): 402-417. https://doi.org/10.1093/esr/jcy020.

Depue, Richard A., and Paul F. Collins. 1999. "Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion." Behavioral and brain sciences 22(3): 491-517. https://doi.org/10.1017/S0140525X99002046.

Destatis

- PageGroup, cited by Destatis. 2021a. "Schwerpunktthemen im Diversity Management von Unternehmen in Deutschland 2020." Chart. 1. Accessed December 14, 2021, https://de-statista-com/statistik/daten/studie/1226989/umfrage/schwerpunktthemen-im-diversity-management-von-unternehmen/
- DIW Berlin, cited by Destatis. 2021b. "Frauenanteil in den Vorständen der 100 bzw. 200 größten deutschen Unternehmen von 2006 bis 2020." Accessed September 09, 2021, https://de-statista-com/statistik/daten/studie/180102/umfrage/frauenanteil-in-den-vorstaenden-der-200-groessten-deutschen-unternehmen/
- CRIF GmbH, cited by Destatis. 2021c. "Frauenanteil in Führungspositionen in Deutschland nach Branchen im Jahr 2021." Accessed December 15, 2021, https://de-statista-com /statistik/daten/studie/575509/umfrage/frauenanteil-in-fuehrungspositionen-in-deutschland-nach-branchen/
- ARD, and ZDF, cited by Destatis. 2022a. "Anteil der Internetnutzer nach Altersgruppen in Deutschland in den Jahren 1997 bis 2021." Accessed August 04, 2022, https://de-statista-com/statistik/daten/studie/36149/umfrage/anteil-der-internetnutzer-in-deutschland-nach-altersgruppen-seit-1997/
- Statistisches Bundesamt (Destatis). 2022b. Internetaktivität zu privaten Zwecken". Accessed June 30, 2022, https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/IT-Nutzung/_inhalt.htmlsprg229130
- Bitkom, cited by Destatis. 2022c. "Anteil der Online-Käufer an den Internetnutzern in Deutschland in ausgewählten Jahren zwischen 2002 und 2021." Accessed March 14 2022, https://de-statista-com/statistik/daten/studie/76228/umfrage/anteil-der-online-shopper-an-den-internetnutzern-in-deutschland-seit-2002/
- CRIF GmbH, cited by Destatis. 2022d. "Frauenanteil in Führungspositionen in Deutschland nach Branchen im Jahr 2022."

Accessed September 1, 2022, https://de-statista-com.ezproxy.uni-giessen.de/statistik/daten/studie/575509/umfrage/frauenanteil-in-fuehrungspositionen-in-deutschland-nach-branchen/

- Statistisches Bundesamt (Destatis). 2022e. "Anteil der weiblichen Lehrkräfte an allgemeinbildenden Schulen in Deutschland im Schuljahr 2020/2021 nach Schulart ." Accessed September 04, 2022, https://de-statista-com/statistik/daten/studie/1129852/umfrage/frauenanteil-unter-den-lehrkraeften-in-deutschland-nach-schulart/

DeYoung, Colin G. 2015. "Openness/intellect: A dimension of personality reflecting cognitive exploration." In APA handbook of personality and social psychology, Vol. 4. Personality processes and individual differences, edited by Mikulincer, Mario, Phillip R. Shaver, Kendra M. L. Cooper, and Richard J. Larsen, 369-399. American Psychological Association. https://doi.org/10.1037/14343-017

DeYoung, Colin G., Lena C. Quilty, and Jordan B. Peterson. 2007."Between facets and domains: 10 aspects of the Big Five." Journal of personality and social psychology 93(5): 880-896. https://doi.org/10.1037/0022-3514.93.5.880.

DeYoung, Colin G. 2020. "Intelligence and personality". In The Cambridge handbook of intelligence, edited by Robert J. Sternberg, 1011–1047. Cambridge University Press. https://doi.org/10.1017/9781108770422.043.

Digman, John M. 1990. "Personality structure: Emergence of the five-factor model." Annual review of psychology 41(1): 417-440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Dudel, Christian, Elke Loichinger, Sebastian Klüsener, Harun Sulak und Mikko Myrskylä. 2021. "The extension of late working life in Germany: trends, inequalities, and the East-West divide.", BiB Working Paper, 38. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-77024-2.

Eugenidis, Dania, Jan Kinne, and David Lenz. 2022. "Analysing gender equality at the firm level". MAGKS Joint Discussion Paper Series in Economics. No. 14-2022. http://hdl.handle.net/10419/262315.

Eurostat. 2010. "Mean hourly earnings by sex, age and economic activity", Accessed August 30, 2022, https://ec.europa.eu/eurostat/databrowser/view/earn_ses10_13/default/table?lang=en

Eurostat. 2018. "Mean hourly earnings by sex, age and economic activity", Accessed

August 30, 2022, https://ec.europa.eu/eurostat/databrowser/view/earn_ses18_13/default/table?lang=en

Feingold, Alan. 1994. "Gender differences in personality: a meta-analysis." Psychological bulletin 116(3): 429-456. https://doi.org/10.1037/0033-2909.116.3.429

Goldberg, Lewis R. 1999. "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models." Personality psychology in Europe 7(1): 7-28.

Haltiwanger, John, Stefano Scarpetta, and Helena Schweiger. 2014. "Cross country differences in job reallocation: The role of industry, firm size and regulations." Labour Economics 26: 11-25. 10.1016/j.labeco.2013.10.001

Helgeson, Vicki S., and Heidi L. Fritz. 1999. "Unmitigated agency and unmitigated communion: Distinctions from agency and communion." Journal of Research in Personality 33(2): 131-158. https://doi.org/10.1006/jrpe.1999.2241.

Jacob, Marita, and Felix Weiss. 2010. "From higher education to work patterns of labor market entry in Germany and the US." Higher Education 60(5): 529-542. https://doi.org/10.1007/s10734-010-9313-y.

Kanter, Rosabeth M. 1977. Men and Women of the Corporation. Basic Books, New York.

Keiser, Heidi N., Paul R. Sackett, Nathan R. Kuncel, and Thomas Brothen. 2016. "Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns." Journal of Applied Psychology 101(4): 569-581. https://doi.org/10.1037/apl0000069

Kusnierz, Cezary, Aleksandra M. Rogowska, and Iuliia Pavlova. 2020. "Examining gender differences, personality traits, academic performance, and motivation in Ukrainian and Polish students of physical education: A cross-cultural study." International journal of environmental research and public health 17(16): 5729. https://doi.org/10.3390/ijerph17165729

Leon, Federico R., Oswaldo Morales, Hugo Vértiz, and Andrés Burga-León. 2017. "Universality of gender differences in 10 aspects of personality: A study of younger and older adult Peruvians." Personality and Individual Differences 112: 124-127. https://doi.org/10.1016/j.paid.2017.02.060

Lindquist, Kristen A. 2021. "Language and emotion: introduction to the special issue."

Affective Science 2(2): 91-98. https://doi.org/10.1007/s42761-021-00049-7

Schneider, Judith, and Romie F. Littrell. 2003. "Leadership preferences of German and English managers", Journal of Management Development, 22(2): 130-148. https://doi.org/10.1108/02621710310459694

Mikolov, Thomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient estimation of word representations in vector space"
arXiv preprint arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781

Misersky, Julia, Asifa Majid, and Tineke M. Snijders. 2019. "Grammatical gender in German influences how role-nouns are interpreted: Evidence from ERPs." Discourse Processes 56(8): 643-654. https://doi.org/10.1080/0163853X.2018.1541382

Neis, Cordula. 2003. Anthropologie im Sprachdenken des 18. Jahrhunderts: die Berliner Preisfrage nach dem Ursprung der Sprache" (1771). Walter de Gruyter.

Orlitzky, Marc, Frank L. Schmidt, and Sara L. Rynes. 2003. "Corporate social and financial performance: A meta-analysis." Organization studies 24(3): 403-441. https://doi.org/10.1177/0170840603024003910

Pietraszkiewicz, Agnieszka, Magdalena Formanowicz, Marie Gustafsson Sendén, Ryan L. Boyd, Sverker Sikström, and Sabine Sczesny. 2019. "The big two dictionaries: Capturing agency and communion in natural language". European journal of social psychology, 49(5), 871-887. https://doi.org/10.1002/ejsp.2561

Ranga, Marina, and Henry Etzkowitz. 2010. "Athena in the world of techne: The gender dimension of technology, innovation and entrepreneurship." Journal of technology management innovation 5(1): 1-12. http://dx.doi.org/10.4067/S0718-27242010000100001

Rudman, Laurie A., and Peter Glick. 2021. The social psychology of gender: How power and intimacy shape gender relations. Guilford Publications.

Saucier, Gerard. 1992."Openness versus intellect: Much ado about nothing?." European journal of Personality 6(5): 381-386. https://doi.org/10.1002/per.2410060506

Sax, Linda J. 2008. "The gender gap in college: Maximizing the developmental potential of women and men". San Francisco, CA, Jossey-Bass/Wiley.

Timmers, Tanya M., Tineke M. Willemsen, and Kea G. Tijdens. 2010. "Gender diversity policies in universities: a multi-perspective framework of policy measures." Higher Education 59(6): 719-735. https://doi.org/10.1007/s10734-009-9276-z

Uchronski, Mirjam. 2010. Das agentische und kommunale Selbstkonzept: Die situative Variabilität von Selbstbeschreibungen.“ Dissertation, Friedrich-Alexander-Universitaet Erlangen-Nuernberg (Germany).

Weisberg, Yanna J., Colin G. DeYoung, and Jacob B. Hirsh. 2011. "Gender differences in personality across the ten aspects of the Big Five." Frontiers in psychology 2(178). https://doi.org/10.3389/fpsyg.2011.00178.

Williams, John E., and Deborah L. Best. 1977. "Sex stereotypes and trait favorability on the Adjective Check List." Educational and Psychological Measurement 37(1): 101-110. https://doi.org/10.1177/001316447703700111

Williams, John E., Robert C. Satterwhite, and Deborah L. Best. 1999. "Pancultural gender stereotypes revisited: The five factor model." Sex roles 40(7): 513-525. https://doi.org/10.1023/A:1018831928829.

Wilson, Edwin B. 1927. "Probable inference, the law of succession, and statistical inference." Journal of the American Statistical Association 22(158): 209-212. 10.1080/01621459.1927.10502953

Yoon, Borham, and Yeasun Chung. 2018. "The effects of corporate social responsibility on firm performance: A stakeholder approach." Journal of hospitality and tourism management 37: 89-96. https://doi.org/10.1016/j.jhtm.2018.10.005.

Appendix

In the following, all aspects are listed with the respective terms in German and English. We made the translations to the best of our knowledge. Though, it may happen that the English translation of the individual terms does not always completely correspond to the German meaning.

Openness

Original terms: inspirierend, inspiriert, stürmisch, glühend, brennend, Inspiration, Stürmigkeit, künstler- isch, kreativ, kunstvoll, schön, idealistisch, Kreativität, Schönheit, Idealismus, dichterisch, lyrisch, anregend, Lyrik, Anregung, musikalisch, musisch, melodisch, Musikalität, schöpferisch, fruchtbar, harmonisch, gefühlsbetont, gefühlvoll, eingebungsvoll, sentimental, Fruchtbarkeit, Harmonie, Gefühlsbetonung, Sentimentalität, geistsprühend, gestalterisch, poetisch, belletristisch, Belletristik, gedankenverloren, versunken, zerstreut, gedankenreich, tiefgründig, Gedankenreichtum, Tiefgründigkeit, abstrakt, vertieft, träumerisch, fantasievoll, Vertiefung, Fantasie, empfindsam, ästhetisch, intuitiv, originell, unkonventionell, rührig, Empfindsamkeit, Ästhetik, Intuition, Originalität, Unkonventionalität.

English translation:

inspiring, inspired, impetuous, ardent, burning, inspiration, impetuosity, artistic, artful, beautiful, idealistic, creativity, beauty, idealism, poetic, lyrical, lyricism, stimulation, musical, melodic, musicality, fertile, harmonious, emotive, sentimental, fertility, harmony, sentimentality, spirited, creative, belletristic, fiction, thoughtful, absorbed, scattered, thoughtfulness, profound, profoundness, abstract, deep, dreamy, fanciful, deepening, imagination, aesthetic, intuitive, original, unconventional, touching, sensibility, intuition, originality, unconventionality

Intellect

Original terms: scharfsinnig, aufgeweckt, begnadet, smart, weise, Scharfsinnigkeit, Aufgewecktheit, Weisheit, lernbereit, kompetent, wissend, geistig, gebildet, intellektuell, reflektierend, renommiert, reflektiert, begabt, hochbegabt, akademisch, kundig, Lernbereitschaft, Kompetenz, Wissen, Geistigkeit, Intellekt, Renommee, Begabung, klug, weitblickend, erfinderisch, analytisch, intelligent, sachlich, kognitiv, fundiert, pfiffig, erfindungsreich, Klugheit, Weitblick, Intelligenz, Sachlichkeit, Kognition, Fundiertheit, Pfiffigkeit, Erfindungsreichtum, kultiviert, gewandt, fundiert, gelehrt, Kultiviertheit, Fundierung, fortschrittlich, überdacht, Fortschrittlichkeit, Überdachtheit, verstandesgemäß, durchdacht, aufgeklärt, nüchtern, Aufgeklärtheit, Nüchternheit, geistig, umsichtig, ideell, idealistisch, Geistigkeit, Umsichtigkeit, Idealismus, spirituell, moralistisch, politisch, ideologisch, Spiritualität, Moral, Ideologie, kenntnisreich, Kenntnisreichtum

English translation:

astute, alert, wise, astuteness, alertness, wisdom, willing to learn, educated, intellectual, reflective, renowned, reflected, highly gifted, academic, Learning, Competent, Intellect, Reputation, Gifted, Clever, Far-Sighted, Analytical, Intelligent, Cognitive, Smart, Inventive, Cleverness, Far-Sightedness, Intelligence, Factual, cognition, soundness, shrewdness, inventiveness, cultured, adroit, sound, erudite, sophistication, founding, progressive, progressiveness, thoughtfulness, rational, thoughtful, enlightened, sober, enlightenedness, sobriety, prudent, idealistic, prudence, idealism, spiritual, moralistic, political, ideological, spirituality, morality, ideology, knowledgeable, knowledgeable-richness

Orderliness

Original terms: routiniert, geübt, Routine, ordnungsliebend, ordnungsgemäß, ordentlich, geordnet, Ordnungsliebe, Ordnungsmäßigkeit, Ordentlichkeit, organisiert, sauber, achtsam, Organisationstalent, Sauberkeit, Achtsamkeit, geregelt, strukturiert, systematisch, Struktur, sorgfältig, präzise, penibel, solide, Sorgfältigkeit, Präzision, akribisch, akkurat, geradlinig, konsequent, Akribie, Geradlinigkeit, Konsequenz, gewissenhaft, gründlich, genau, Gewissenhaftigkeit, Gründlichkeit, Genauigkeit

English translation:

practiced, routine, orderly, orderliness, organized, clean, mindful, organizational, neatness, attentive, regulated, structured, systematic, structure, careful, precise, solid, diligence, precision, meticulous, accurate, straightforward, consistent, meticulousness, straightforwardness, consistency, conscientious, thorough, exact, conscientiousness, thoroughness, accuracy

Industriousness

Original terms: ambitioniert, Ambition, Diszipliniert, fleißig, eifrig, tüchtig, emsig, Disziplin, Fleiß, Eifer, Tüchtigkeit, Emsigkeit, aktiv, agierend, leistungsbereit, professionell, fähig, sachkundig, qualifiziert, Leistungsbereitschaft, Professionalität, Fähigkeit, Sachkundigkeit, Qualifikation, flott, zügig, geschwind, flink, Schnelligkeit, pflichtbewusst, verlässlich, Pflichtbewusstsein, Verlässlichkeit, ordnungsgemäß, Ordnung, konzentriert, zielstrebig, zielsicher, Konzentriertheit, Konzentration, Zielstrebigkeit, Zielsicherheit, fokussiert, Fokussiertheit, Fokus

English translation:

ambitious, ambition, disciplined, diligent, eager, discipline, diligence, zeal, industriousness, active, acting, performing, professional, capable, proficient, qualified, performance, professionalism, ability, proficiency, qualification, brisk, briskness, speed, dutiful, reliable, dutifulness, reliability, orderly, purposeful, focusedness, concentration, purposefulness, focused, focus

Assertiveness

Original terms: verantwortungsbewusst, verantwortungsvoll, verfassungsmäßig,

pflichtbewusst, pflichtgetreu, verantwortungsfreudig, Verantwortungsbewusstsein, Pflichtbewusstsein, Pflichttreue, stark, belastbar, fit, kernig, tapfer, heroisch, mutig, autoritär, markig, imposant, imponierend, tonangebend, Stärke, Belastbarkeit, Kernigkeit, Tapferkeit, Mut, Autorität, energiereich, schlagfertig, Schlagfertigkeit, Energiereichtum, entschlossen, draufgängerisch, bestimmend, Draufgänger, Bestimmtheit, einflussreich, durchsetzungsfähig, unnachgiebig, durchsetzungsstark, dickköpfig, hartnäckig, standhaft, kämpferisch, beharrlich, willensstark, Einflussreichtum, Durchsetzungsfähigkeit, Unnachgiebigkeit, Durchsetzungsstärke, Dickköpfigkeit, Hartnäckigkeit, Standhaftigkeit, Beharrlichkeit, Willensstärke, proaktiv, aktiv, tatkräftig, handelnd, tollkühn, Tatkräftigkeit, Tollkühnheit, offensiv, unmittelbar, Unmittelbarkeit, mächtig, dominant, dominierend, Dominanz, forsch, ungestüm, unbescheiden, penetrant, Direktheit, Unbescheidenheit, Penetranz

English translation:

responsible, constitutional, dutiful, sense of responsibility, sense of duty, dutifulness, strong, resilient, fit, brave, heroic, courageous, authoritative, pithy, imposing, setting the tone, strength, resilience, grit, bravery, courage, authority, quick-witted, quick-wittedness, wealth of energy, determined, go-getter, decisive, influential, intransigent, assertive, bull-headed, steadfast, pugnacious, persistent, strong-willed, wealth of influence, intransigence, assertiveness, bull-headedness, stubbornness, steadfastness, perseverance, willfulness, proactive, active, energetic, acting, foolhardy, vigor, foolhardiness, offensive, direct, immediacy, powerful, dominant, domination, brash, impetuous, immodest, penetrating, directness, immodesty, penetrance

Enthusiasm

Original terms: einladend, integrierend, sympathisch, nett, Sympathie, Nettigkeit, kontaktfreudig, sozial, umgänglich, ungezwungen, offen, kommunikationsfähig, empfänglich, kommunikationsstark, Kontaktfreudigkeit, Umgänglichkeit, Ungezwungenheit, Offenheit, Kommunikationsfähigkeit, Empfänglichkeit, Kommunikationsstärke, gefühlvoll, gefühlsbetont, rührselig, beflügelt, ausdrucksstark, Rührseligkeit, Ausdrucksstärke, aufgedreht, erregt, glühend, lebhaft, lebendig, dynamisch, strahlend, energiegeladen, energievoll, beliebt, Aufgedrehtheit, Erregtheit, Lebhaftigkeit, Lebendigkeit, Dynamik, Energiegeladenheit, Beliebtheit, lustig, munter, Lustigkeit, Munterkeit, umgänglich, kontaktfreudig, aufgeschlossen, Umgänglichkeit, Kontaktfreudigkeit, Aufgeschlossenheit, zudringlich, nahbar, zugänglich, freundschaftlich, herzlich, gesellig, teamfähig, Zudringlichkeit, Nahbarkeit, Zugänglichkeit, Freundschaftlichkeit, Herzlichkeit, Geselligkeit, Teamfähigkeit, extrovertiert, aufgeschlossen,

gesprächig, mitteilsam, aufrichtig, offenherzig, selbstsicher, kommunikativ, unverschlossen, redsam, Extrovertiertheit, Aufgeschlossenheit, Gesprächigkeit, Mitteilsamkeit, Aufrichtigkeit, Offenherzigkeit, Selbstsicherheit, Unverschlossenheit, Redsamkeit, enthusiastisch, leidenschaftlich, begeistert, engagiert, motiviert, passioniert, begeisterungsfähig, euphorisch, ekstatisch, beschwingt, mitreißend, mitgerissen, überschwänglich, hingebungsvoll, anregend, angeregt, hinreißend, sinnenfreudig, unternehmungslustig, Enthusiasmus, Leidenschaftlichkeit, Begeisterung, Engagement, Motivation, Passion, Begeisterungsfähigkeit, Euphorie, Ekstase, Beschwingtheit, Überschwänglichkeit, Hingabe, Anregung, Sinnenfreudigkeit, Unternehmungslust, temperamentvoll, begeisternd, einnehmend, fesselnd, charismatisch, ansprechend, Temperament, Begeisterung, Einnahme, Charisma

English translation:

inviting, integrating, sympathetic, nice, sympathy, niceness, informal, open, receptive, informality, openness, receptivity, communicative strength, soulful, emotional, sentimental, impulsive, expressive, emotionality, expressiveness, agitated, ardent, vivacious, lively, dynamic, bright, energy, energetic, popular, upbeat, agitation, vivacity, liveliness, dynamism, energy-charged, popularity, merry, cheerful, levity, pushy, approachable, accessible, amicable, cordial, sociable, team player, pushiness, approachability, accessibility, amicability, cordiality, sociability, team-spirited, extroverted, outgoing, communicative, sincere, open-hearted, self-confident, uncloseted, talkative, extroversion, open-mindedness, communicativeness, sincerity, open-heartedness, self-confidence, unclosetedness, talkativeness, committed, motivated, euphoric, ecstatic, elated, rousing, carried away, exuberant, devoted, stimulating, excited, ravishing, sensual, enterprising, passionate, commitment, motivation, passion, euphoria, ecstasy, exhilaration, exuberance, devotion, excitement, sensuality, enterprise, spirited, enthusiastic, engaging, captivating, charismatic, appealing, temperament, enthusiasm, ingestion, charisma,

Compassion

Original terms: empathisch, emotional, gefühlvoll, mitfühlend, zart, empfindsam, einfühlsam, sensibel, vermittelnd, Empathie, Emotionalität, Mitgefühl, Zärtlichkeit, Empfindsamkeit, Einfühlsamkeit, Sensibilität, besorgt, sorgsam, warm, behutsam, feinfühlig, sensitiv, vertrauensvoll, einfühlend, liebevoll, partnerschaftlich, zartfühlend, Behutsamkeit, Feinfühligkeit, Sensitivität, Partnerschaftlichkeit, empfindend, feinfühlig, feinfühlend, teilnehmend, menschlich, Teilnahme, Menschlichkeit, hilfsbereit, solidarisch, altruistisch, aufbauend, barmherzig, nobel, gefällig, Hilfsbereitschaft, Solidarität, Barmherzigkeit, Noblesse, Gefälligkeit, rücksichtsvoll, verständnisvoll, Rücksichtnahme, Verständnis, großherzig, Großherzigkeit, altruistisch, fürsorglich, Altruismus, Fürsorge, gefühlvoll

English translation:

empathic, emotional, compassionate, tender, mediating, emotionality, compassion, tenderness, sensibility, empathy, concerned, warm, cautious, trusting, empathetic, loving, tender-hearted, cautiousness, delicacy, sensitivity, partnership, feeling, delicate, sensitive, participating, humane, participation, humanity, helpful, constructive, merciful, noble, pleasing, helpfulness, solidarity, mercy, noblesse, complaisance, considerate, thoughtful, understanding, generous, big-hearted, altruistic, altruism, caring, soulful

Politeness

Original terms: zuvorkommend, unterwürfig, gehorsam, folgsam, artig, akzeptierend, ergeben, sittsam, gesittet, respektabel, Unterwürfigkeit, Gehorsamkeit, Folgsamkeit, Artigkeit, Akzeptanz, Ergebung, Sittsamkeit, Respekt, diplomatisch, verschwiegen, defensiv, geheimnisvoll, abwartend, unnahbar, bedächtig, scheu, schüchtern, still, introvertiert, demütig, gehemmt, unauffällig, maßvoll, ehrerbietig, manierlich, Diplomatie, Verschwiegenheit, Unnahbarkeit, Bedächtigkeit, Schüchternheit, Introvertiertheit, Demut, Unauffälligkeit, nachgiebig, devot, fügsam, gesittet, gehorsam, entgegenkommend, ergeben, bereitwillig, unaufdringlich, dezent, Nachgiebigkeit, Fügsamkeit, Gehorsamkeit, Ergebenheit, Bereitwilligkeit, ermunternd, motivierend, anfeuernd, anregend, anspornend, bestärkend, animierend, ermutigend, Motivation, Anregung, Ermunterung, Ansporn, Bestärkung, Animierung, Ermutigung, achtungsvoll, freundlich, hochachtungsvoll, höflich, rücksichtsvoll, wohlerzogen, charmant, gutartig, nett, umgänglich, warmherzig, wohlwollend, taktvoll, anständig, Freundlichkeit, Höflichkeit, Wohlerzogenheit, Gutartigkeit, Nettigkeit, Umgänglichkeit, Warmherzigkeit, Wohlwollen, Anständigkeit, bodenständig, selbstkritisch, uneingebildet, bescheiden, willfähig, galant, wohlerzogen, Bodenständigkeit, Selbstkritik, Bescheidenheit, begünstigen, kompensieren, helfen, zuvorkommend, konfliktscheu, respektvoll, friedvoll, friedfertig, ruhig, gewaltfrei, einträchtig, pazifistisch, unaggressiv, ausgeglichen, harmoniebedürftig, friedlich, friedliebend, Friedfertigkeit, Ruhe, Gewaltfreiheit, Einträchtigkeit, Pazifismus, Ausgeglichenheit, Harmoniebedürftigkeit, Friedliebigkeit, uneigennützig, altruistisch, solidarisch, fürsorglich, uneigennützig, aufopferungsvoll, wohltätig, uneigensinnig, karitativ, gemeinnützig, aufopfernd, freigiebig, edel, ehrbar, Uneigennützigkeit, Altruismus, Solidarität, Fürsorglichkeit, Uneigennutz, Wohltat, Freigiebigkeit, Ehrbarkeit

English translation:

accepting, respectable, submissiveness, kindness, acceptance, modesty, respect, diplomatic, defensive, secretive, wait-and-see, aloof, deliberate, shy, quiet, introverted, inhibited, moderate, reverent, mannerly, diplomacy, secrecy, aloofness, deliberateness, shyness, introversion, humility, inconspicuousness, submissive, docile, demure, obedient, accommodating, devoted, unobtrusive, discreet, yielding, docility, obedience, surrender, willingness, motivating, cheering, stimulating, inciting, encouraging, animating,

motivation, stimulation, incitement, animation, encouragement, friendly, considerate, charming, kind, affable, warm, benevolent, tactful, decent, friendliness, courtesy, benign, niceness, affability, warm-heartedness, benevolence, decency, unconceited, modest, willing, gallant, well-mannered, down-to-earth, self-critical, humble, favoring, compensating, helping, courteous, conflict averse, respectful, peaceable, calm, non-violent, amicable, pacifist, unaggressive, peaceful, peace-loving, peacefulness, tranquility, nonviolence, harmony, pacifism, balanced, harmony-seeking, unselfish, self-sacrificing, charitable, altruistic, bountiful, noble, honorable, altruism, solidarity, caring, unselfishness, beneficence, bounty, honorability,

Volatility

Original terms: gelassen, sanftmutig, Gelassenheit, Sanftmut, bedächtig, gefasst, Bedacht, Gefasstheit, beständig, unproblematisch, Beständigkeit, ausgeglichen, Ausgeglichenheit, beherrscht, beruhigend, ruhig, Beherrschtheit, Beherrschung, Ruhe, entspannt, gechillt, friedvoll, konfliktfähig, korrekt, Konfliktfähigkeit, Korrektheit, kontrolliert, bedacht, besonnen, Besonnenheit

English translation:

gentle, serenity, gentleness, thoughtful, composed, steady, unproblematic, steadfast, balanced, equilibrium, sedate, calm, restraint, composure, relaxed, chilled, peaceful, conflict, capable, correct, conflict ability, correctness, controlled, considered, prudent, levelheadedness

Withdrawal

Original terms: zuversichtlich, robust, unbeschwert, Zuversichtlichkeit, Unbeschwertheit, Robustheit, zugänglich, angstfrei, kühn, unerschrocken, zutraulich, Zugänglichkeit, Kühnheit, Kuehnheit, Zutraulichkeit, lebensfroh, sorglos, unbesorgt, unbekümmert, unbekuemmert, zwanglos, sorgenfrei, vorurteilsfrei, Sorglosigkeit, Unbekümmertheit, Unbekümmertheit, ermutigend, couragiert, Courage, Ausgeglichen, unverkrampft, belastbar, Belastbarkeit, unängstlich, furchtlos, selbstsicher, wagemutig, heroisch, unbesorgt, gefestigt, unbeschwert, selbstbewusst, Furchtlosigkeit, Selbstsicherheit, Selbstbewusstsein, Wagemut, Wagemutigkeit, freudig, positiv, heiter, lebensbejahend, lebenslustig, vital, Heiterkeit, Lebenslust, Vitalität, glücklich, erfüllt, erfuellt, wohlbehaglich, Erfüllung, Wohlbehaglichkeit, Glück, optimistisch, fröhlich, Optimismus, Fröhlichkeit, enthemmt, freimütig, Freiheit, Enthemmtheit, Freimütigkeit

English translation:

robust, confidentness, light-heartedness, robustness, accessible, bold, undaunted, trusting, accessibility, boldness, trustfulness, unworried, unprejudiced, carefree, unconcernedness, encouraging, courageous, courageousness, balanced, unconstrained, resilient, resilience, unafraid, fearless, confident, daring, heroic, unconcerned, steady, light-hearted, self-assured, fearlessness, self-assurance, self-confidence, daringness,

joyful, positive, cheerful, life-affirming, fun-loving, vital, lust for life, vitality, happy, fulfilled, fulfilledness, well-being, happiness, optimistic, cheerfulness, optimism, jollity, disinhibited, liberty, freedom, disinhibition, libertinage