

The EFA Annual Meeting 2020 in Helsinki, Finland: Conference Digitalization and Textual Analyses of EFA Papers 2009 and 2010

Wolfgang Breuer*

I. Overview

The 47th Annual Conference of the European Finance Association (EFA) took place virtually, but was hosted by the Aalto University School of Business in Helsinki, Finland, from 19 to 22 August 2020. Professor Ulrike Malmendier from the University of California, Berkeley, could be won as the keynote speaker.

Similarly to the previous year, 243 of the approximately 1,884 papers submitted were accepted for presentation. Consequently, the acceptance rate remains at the usual low level of around 12.90%. Table 1 presents conference key figures since 2012 in detail.

A total of 631 scholars contributed to the 243 papers presented at the conference, with 40 authors – who participated in 23 papers – being from 14 universities and institutions based in Germany. The ratio of papers with German contribution amounts to $23/243 = 9.47\%$ and is thus somewhat below the long-term average of 10.08% realized between 2012 and 2019.

Measured by the number of downloads from the Social Science Research Network (SSRN; deadline: 2020/10/13), the three most successful contributions with German participation were

1. *Heimer, R.* (Boston College)/*Iliewa, Z.* (Max Planck Institute for Research on Collective Goods)/*Imas, A.* (Carnegie Mellon University)/*Weber, M.* (University of Mannheim): Dynamic Inconsistency in Risky Choice: Evidence from

* Univ.-Prof. Dr. Wolfgang Breuer, Rheinisch-Westfälische Technische Hochschule Aachen, Department of Finance, School of Business and Economics, RWTH Aachen University, Templergraben 64, D-52056 Aachen, E-Mail: wolfgang.breuer@bfw.rwth-aachen.de

I would like to thank Nadine Schramm and Anthony Haake very much for their invaluable support when retrieving and analyzing all the data on conference papers 2009, 2010, and 2020. Of course, the usual disclaimer applies.

Table 1
Selected Key Figures in Annual Comparison 2012 to 2020

| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|-------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Submissions | 1,600 | 1,713 | 1,700 | 1,700 | 1,853 | 1,800 | 1,900 | 1,900 | 1,884 |
| Acceptance | 216 | 240 | 267 | 240 | 243 | 222 | 243 | 243 | 243 |
| Acceptance rate | 13.50 % | 14.01 % | 15.71 % | 14.12 % | 13.11 % | 12.33 % | 12.79 % | 12.79 % | 12.90 % |
| “German” papers | 21 | 19 | 30 | 18 | 29 | 27 | 22 | 27 | 23 |
| Share Germany | 9.72 % | 7.92 % | 11.24 % | 7.50 % | 11.93 % | 12.16 % | 9.05 % | 11.11 % | 9.47 % |
| Downloads total | 32,936 | 30,109 | 31,923 | 32,613 | 35,454 | 34,523 | 44,646 | 36,372 | 51,057 |
| Available via SSRN | 170 | 163 | 181 | 145 | 144 | 148 | 176 | 174 | 185 |
| Rel. availability | 78.70 % | 67.92 % | 67.79 % | 60.42 % | 59.26 % | 66.67 % | 72.43 % | 71.60 % | 76.13 % |
| Downloads per paper | 194 | 185 | 176 | 225 | 246 | 233 | 254 | 209 | 276 |
| Downloads German Top 3 | 1,074 | 1,424 | 1,607 | 685 | 1,236 | 1,729 | 2,213 | 2,421 | 915 |
| Ranking German Top 3 | 26.67 | 13.67 | 11 | 47.67 | 15 | 15.67 | 13.33 | 12.33 | 36.67 |
| Downloads Top 7 | 6,880 | 7,095 | 6,445 | 7,328 | 9,313 | 8,370 | 10,173 | 8,498 | 17,187 |
| Percentage of downloads Top 7 | 20.89 % | 23.56 % | 20.19 % | 22.47 % | 26.27 % | 24.24 % | 22.79 % | 23.36 % | 33.66 % |

the Lab and Financial Markets, 323 Downloads, ranking position 34 based on total downloads.

2. *Maug, E.* (University of Mannheim)/*Levit, D.* (University of Washington)/*Malenko, N.* (University of Michigan): Trading and Shareholder Democracy, 310 Downloads, ranking position 35 based on total downloads.

3. *Klos, A.* (Kiel University)/*Koehl, A.* (Kiel University)/*Rottke, S.* (University of Amsterdam): Streaks in Daily Returns, 282 Downloads, ranking position 41 based on total downloads.

This leads to a total number of downloads of 915 with an average ranking position of 36.67, which has been one of worst results since 2012 (see Table 1 again). The download figures reported for the years 2012 to 2020 are based on the results from the first half of October of the respective year. In general, 185 of the 243 papers accepted at the Helsinki meeting are available for download via SSRN. The resulting rate of 76.13 % is noticeably higher than the average of 68.10 % achieved so far since 2012. The same applies to the total number of downloads of 51,057 (until 2020/10/13) and the number of downloads per paper of 276, both exceeding considerably the respective average values of 34,822 and 215 of the years 2012 to 2019.

Table 2 shows the shares of authors from selected countries over the years 2012 to 2020 as a percentage of the respective total number of presentations. The authors were assigned to countries according to the location of the university or institution where they work. If there are several locations mentioned for one author, only the first one from the list was taken into account. Furthermore,

each author was weighted according to his or her proportionate participation in a conference contribution (e.g. with 0.5 participation points for two authors or 0.33 for three authors). Germany achieves a rather comparatively low value of 5.4% falling from ranking position #3 in 2019 to position #6 in 2020. As expected, China now has outperformed Germany rather clearly achieving the second position directly behind the all-time winner USA making the EFA conference a meeting which is in fact more “international” than “European”. Remarkably, even Canada and Switzerland were able to place themselves before Germany. However, the downfall of Great Britain is even more pronounced, as the UK share of all submissions went down to about only 50% of its value in 2019.

Consistent with our findings in previous years, the host country exhibits an extraordinarily high share of all contributions: While the average for Finland for 2012 to 2019 only amounts to 0.65% without any clear trend, we observe a share of 1.06% in 2020 which is higher than any other fraction for 2012 to 2019. In previous reports, we mused whether this quite robust anomaly is caused by lower travel expenses or some kind of national nepotism. However, in 2020 the EFA conference took place in a virtual way rendering the travel expenses argument more or less invalid. Nevertheless, Finland is too small a country to elicit sufficiently high (absolute) effects.

Table 3 shows the relevance of different topics of the EFA 2020 meeting according to the respective number of accepted manuscripts and the corresponding downloads. Similarly to previous conferences, there is the triumvirate of empirical papers in the fields Corporate Finance and Governance, Asset Pricing, and Financial Intermediation and Institutions which account for the first three positions representing almost 60% of all papers and more than 72% of all downloads. In contrast, theoretical papers still are far less prevalent at the EFA conferences. Moreover, it is remarkable that the number of special sessions sponsored by external institutions like the Academy of Finland or the European Central Bank was increased to a total of five. However, with 10 out of 15 papers available on the SSRN platform and an average download number of as little as 116 per paper, one may pose the question whether it would not be preferable – from a scientific point of view – to increase the acceptance rate for “regular” sessions instead of wasting capacity presumably due to monetary reasons.

II. The Most Important Contributions

The below-average performance of German contributions with respect to the 2020 EFA conference is also reflected by the fact that there has not been any paper with German contribution able to place itself among the top 7 according to total downloads. 33.66% of all downloads are accounted for by these top 7,

which is far above the average share observed since 2012. In what follows, the contents of these seven papers are presented in somewhat more detail.

1. Berg, F. (Massachusetts Institute of Technology)/Koelbel, J. (Massachusetts Institute of Technology)/Rigobon, R. (Massachusetts Institute of Technology): *Aggregate Confusion: The Divergence of ESG Ratings*. 8,757 downloads, ranking position 1 based on downloads per day.

Environmental, social, and governance (ESG) ratings become increasingly relevant, as more and more investors take ESG issues into account when allocating their wealth across different stocks. However, it is a well-known fact that ESG ratings vary across different providers. In this paper, six different raters – KLD (MSCI Stats), Sustainalytics, Vigeo Eiris (Moody's), RobecoSAM (S&P Global), Asset4 (Refinitiv), and MSCI, are examined. The correlations between the ratings of these providers are on average as little as only 54%. Three causes for these discrepancies are considered: (1) scope divergence which means that ratings are derived from different sets of attributes, (2) measurement divergence, i. e., different indicators are employed for the same attribute, and (3) weight divergence which is a consequence of rating agencies putting different weights on the relevance of attributes. Measurement and scope divergence turn out to be the main drivers of differences in ESG ratings with measurement divergence being at least partly driven by a rater effect which may be caused by rating agencies typically allocating tasks among analysts by firm rather than by category. These results are of theoretical and practical importance. For example, the degree of dispersion in ESG ratings may have a direct effect on asset pricing with higher negative CAPM alphas for good ESG performance in the case of greater unanimity among ESG raters.

2. Chen, H. (University of Notre Dame)/Cohen, L. (Harvard Business School)/Gurun, U. (University of Texas at Dallas): *Don't Take Their Word For It: The Misclassification of Bond Mutual Funds*. 1,622 downloads, ranking position 8 based on downloads per day.

It is particularly difficult to assess the quality of bond funds, as the average bond fund holds about six times the number of different issues as the average equity fund. Among others, Morningstar therefore serves as an information intermediary for the rating of fixed income funds to which are allocated – depending on their quality – so-called Morningstar stars. Unfortunately, Morningstar bases its decision solely on data about bond quality self-reported by bond fund managers. It is shown that this self-reported information is highly unreliable. More than 31% of all funds are misclassified as more secure than they really are leading on average to higher expected returns than typical for bonds of the respective (falsely documented) rating class and implying about 12.3% additional, unjustified Morningstar stars. This higher number of stars enables bond

Table 2
Authors' Share of Total Contributions by Country of Origin

| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| USA | 49.9% | 53.0% | 42.0% | 44.0% | 40.2% | 43.0% | 49.3% | 51.7% | 46.2% |
| China | 3.2% | 2.5% | 3.5% | 4.2% | 3.6% | 5.7% | 3.0% | 5.7% | 6.4% |
| Canada | 4.4% | 3.5% | 4.0% | 6.8% | 4.3% | 6.9% | 3.6% | 3.8% | 6.2% |
| Switzerland | 5.1% | 6.3% | 7.2% | 2.8% | 4.8% | 3.5% | 3.6% | 3.5% | 6.0% |
| United Kingdom | 7.6% | 11.5% | 12.3% | 11.2% | 9.7% | 9.5% | 10.2% | 11.0% | 5.6% |
| Germany | 7.0% | 5.2% | 7.2% | 5.3% | 7.9% | 8.1% | 6.0% | 5.8% | 5.4% |
| Netherlands | 4.3% | 4.0% | 2.5% | 4.7% | 4.8% | 2.5% | 2.9% | 3.1% | 3.8% |
| France | 4.6% | 3.0% | 4.8% | 4.8% | 4.7% | 2.1% | 5.5% | 3.7% | 3.2% |
| Australia | 0.7% | 0.7% | 2.0% | 3.2% | 2.7% | 2.6% | 2.0% | 1.3% | 2.9% |
| Denmark | 2.3% | 0.7% | 0.7% | 1.3% | 2.3% | 0.7% | 2.4% | 0.6% | 2.0% |
| Italy | 2.5% | 0.0% | 2.8% | 0.9% | 2.3% | 2.0% | 2.6% | 1.1% | 1.8% |
| Sweden | 1.8% | 0.8% | 2.0% | 1.9% | 1.8% | 3.3% | 1.4% | 1.6% | 1.4% |

Table 3
SSRN Downloads and Conference Contributions per Topic Area

| | Number of Downloads | Share of Total Downloads | Number of Conference Contributions | Share of total Conference Contributions |
|--|---------------------|--------------------------|------------------------------------|---|
| Corporate Fin. and Gov., Empirical | 17,341 | 33.96 % | 54 | 22.23 % |
| Asset Pricing, Empirical | 12,438 | 24.36 % | 45 | 18.52 % |
| Financial Interm. and Instit., Empirical | 7,434 | 14.56 % | 45 | 18.52 % |
| Market Microstructure | 2,999 | 5.87 % | 15 | 6.17 % |
| Asset Pricing, Theory | 2,821 | 5.53 % | 15 | 6.17 % |
| Behavioral Finance | 2,359 | 4.62 % | 18 | 7.42 % |
| International Finance | 1,827 | 3.58 % | 6 | 2.47 % |
| Corporate Fin. and Gov., Theory | 1,262 | 2.47 % | 12 | 4.94 % |
| Household Finance | 868 | 1.70 % | 12 | 4.94 % |
| Financial Interm. and Instit., Theory | 544 | 1.07 % | 6 | 2.47 % |
| Academy of Finland | 465 | 0.91 % | 3 | 1.23 % |
| Bank for International Settlement | 404 | 0.79 % | 3 | 1.23 % |
| Norges Bank Investment Management | 132 | 0.26 % | 3 | 1.23 % |
| European Central Bank | 96 | 0.19 % | 3 | 1.23 % |
| Bank of Finland | 67 | 0.13 % | 3 | 1.23 % |
| Total | 51,057 | 100 % | 243 | 100 % |

Notes: Fin.: Finance, Gov.: Governance, Interm. & Instit.: Intermediation & Institutions.

funds to charge expense ratios that are *ceteris paribus* 11.4 basis points higher and to acquire larger funds inflows. However, after correcting for these misqualifications the respective funds turn out to be only mediocre. Young fund managers on an early stage of their careers are particularly inclined to falsely report bond qualities.

3. Bolton, P. (Columbia University)/Kacperczyk, M. (Imperial College): Do Investors Care about Carbon Risk? 1,591 downloads, ranking position 12 based on downloads per day.

Ongoing climate change and political activities handling this issue give rise to special earnings risks for companies with high carbon emissions. In this empirical study, three hypotheses related to this problem are tested. According to the carbon risk premium hypothesis, a positive relationship between a firm's CO₂ emission and its stock returns is expected. The market inefficiency or carbon alpha hypothesis states that the risk from carbon emissions is underpriced making it possible to earn positive abnormal returns by selling short companies with high carbon emissions and investing in firms with low emissions. The last hypothesis is called the divestment hypothesis and addresses the possibility that stocks of firms with high carbon emission are shunned by investors as sin stocks leading to higher return demands for these kinds of stocks. Based on a comprehensive sample that comprises about 1,000 listed companies since 2005 and more than 2,900 listed companies since 2016 in the US, the authors only find evidence for the first hypothesis. There is a special carbon risk premium which is mainly related to the total emission and the change in total emission of a firm (about 15 to 33 basis point reaction of stock returns to an increase of total emission or a change of total emission of one standard deviation, respectively). Moreover, some firms from industries with a high intensity of carbon emission are indeed avoided by institutional investors, but the emission intensity does not turn out to be a driving force for stock returns. Moreover, in additional analyses, it is shown that a substantial carbon risk premium is a recent phenomenon highlighting the change in environmental awareness in society and economics.

4. Bybee, L. (Yale University)/Kelly, B. (Yale University)/Manela, A. (Washington University in St. Louis)/Xiu, D. (University of Chicago): The Structure of Economic News. 1,546 downloads, ranking position 11 based on downloads per day.

In this paper, with the help of machine-learning tools a full text analysis of all about 800,000 articles of the Wall Street Journal between 1984 and 2017 is performed. A limited number of 180 topics is identified in such a way that the variation in term usage across articles can be explained best. For each article and thus each different point in time, the relative importance of each of these 180 topics can be determined. The authors find that there are three different general

kinds of topics: (1) recurrent ones which attract media attention throughout the sample period (like “Federal Reserve”), (2) seasonal ones with periodically rising relevance (like “presidential election”), and (3) episodic ones that only gain immediate relevance at irregular time intervals due to sudden events (like “terrorism”). By means of additional machine-learning algorithms, the 180 base topics are organized in a hierarchical structure with certain high-level metatopics like “markets” and “government” on top whose relevance stays almost constant over time, while there is more pronounced variation in attention on lower hierarchy levels. The main value-added from these text-analyses is that business news are suited to describe a large scope of economic circumstances and may be helpful for, e.g., forecasting purposes, as this data covers information which is not part of standard quantitative (macro-) economic regression approaches. Moreover, by identifying ex post certain news related to an economic shock, it is possible to analyze in detail the events which may have caused the shock in the first place.

5. Lopez-Lira, A. (University of Pennsylvania): Risk Factors that Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns. 1,417 downloads, ranking position 17 based on downloads per day.

The author utilizes machine-learning methods to analyze a part of firms’ 10-K annual reports from 2006 to 2019 that address particularly firm management’s own risk assessment: Item 1A Risk Factors. In total, 79,304 documents are examined. While unrestricted algorithms would end up with 25 different risk factors, the author restricts himself to only four factors for reasons of comparability with traditional factor models. Resulting systematic risk factors (i.e. risk factors increasing the covariance between stock prices of different companies) are eventually International, China, Oil, and Credit Risk. An asset pricing model based on these four factors performs approximately as well as the leading models in the literature like the Fama-French four-factor model. This is rather remarkable, as in the model with firm identified risk factors only, no information about historical stock returns are used.

6. Ayyagari, M. (School of Business, George Washington University)/Demircug-Kunt, A. (The World Bank)/Maksimovic, V. (School of Business at the University of Maryland): The Rise of Star Firms: Intangible Capital and Competition. 1,202 downloads, ranking position 26 based on downloads per day.

This empirical paper examines a dataset of publicly listed firms from the Compustat database with respect to the specific features of so-called star firms, i.e. firms that seem to offer extraordinarily high returns on capital. The aim of this study is to explain this gap between star firms and other firms. In particular, the issue is to be resolved whether the abuse of market power or efficiency advantages lie at the root of these return discrepancies. The authors conclude

that the large dispersion across firms regarding return on invested capital is mainly a consequence of a measurement error in accounting for intangible capital (like a highly skilled labor force) in a consistent way. In a similar manner, seemingly excessive pricing mark-ups over time disappear when intangible capital is considered adequately. Though mark-ups are positively connected to the probability of being a star firm, these positive correlations may not be an indicator for market power, as higher mark-ups over time may be consistent with decreasing prices due to efficiency gains. All these findings are relevant from a political point of view as well. While there is apparently little evidence of activities to increase profits by making use of monopolistic power, star firms may utilize their productivity advantages for setting rather low prices thus giving up some part of short-term gains in order to grow quickly and secure their market position in the long run. Such a strategy may be relevant from a political point of view.

7. Boyarchenko, N. (Federal Reserve Bank of New York)/Larcen, L.C. (Copenhagen Business School)/Whelan, P. (Copenhagen Business School): The Overnight Drift. 1,052 downloads, ranking position 7 based on downloads per day.

The authors show that returns between 2:00 a.m. and 3:00 a.m. (Eastern Time) on US equity futures amount to 3.6% p.a. on average. As the return simply seems to be shifted to the right during this hour, the authors coin this phenomenon the overnight drift. Their paper mainly aims at exploring the possible explanations for this observation. While they find no evidence that the overnight drift can be explained by incoming news (like earnings announcements) after US cash markets have closed, the overnight drift may be a consequence of order imbalances inducing market makers' inventory management activities in a global market for equity risk and of end-of-day volatility shocks implying higher expected overnight returns. Regarding inventory management, the authors are able to show that overnight returns are negatively connected to the closing order imbalance of the preceding day. Moreover, only negative order imbalances (market sell-offs) are followed by positive overnight drift. Also consistent with the order imbalance hypothesis, the authors find out that the order book is deeper on the ask side for closing order imbalance being negative and that price reversals are more pronounced in high volatility states. In addition, due to daylight savings time in the US, but not in Japan, the Japanese markets open one hour earlier in winter than in summer from a US point of view. This can be interpreted as an exogenous variation in the arrival-time of Japan-based clients. As a consequence, it can be shown that US market makers are able to off-load earlier a part of their inventory during the winter time at the Tokyo Stock Exchange instead of waiting for the London Stock Exchange to open. While all this evidence is consistent with inventory management as the main determinant of overnight drift, risk sentiment may offer an additional explanation. Adverse

end-of-day volatility shocks in the form of unexpected bad news about, e. g., the economic outlook may lead to the sale of risky positions and subsequent market rebounds overnight.

Apparently, environmental issues (papers #1 and #3) and the application of new technologies for textual analyses (papers #4 and #5) are currently on the rise and attract much of the attention of the scientific community. It will be interesting to investigate whether this development will gain momentum in the years to come.

III. Differences Between Published and Unpublished EFA Papers

According to the preceding section, textual analyses currently seem to be quite topical. In fact, using the right software makes it easy to examine large volumes of texts. In what follows, we want to take a closer look at the contents of EFA conference papers 2009 (Bergen, Norway) and 2010 (Frankfurt, Germany) and in particular potential differences between those papers which were eventually published and those which could not be identified as published, because they turned out to be unpublishable or at least went through such a complete overhaul that they were no longer retrievable.

Table 4 presents a contingency table of the whole sample of 430 papers. The acceptance rate of papers in journals presented at an EFA conference is about 69% for EFA 2009 and about 71% for EFA 2010 and the average time until publication after an EFA presentation is 3.17 years. While our sample contains the same number of papers for both conferences, the average lead time until publication is considerably shorter for the 2010 conference.

For both years, the titles and abstracts from the conference paper manuscripts were collected together with the information whether, when and in which journal a paper was published. To focus on keywords, stop words like “and” or “or” were ignored, and all plural words swapped to the singular.

Using this sample, first, a co-occurrences analysis of paper titles is performed, highlighting which words commonly occur together. Figure 1 documents the results from two different angles.

Panel A compares the co-occurrences of keywords from papers published in scientific journals with those that are not. Apparently, there is a more pronounced “core” of a large number of keywords for the subset of published papers than for unpublished ones and, moreover, the most relevant keywords differ for published and unpublished papers. For the published ones, terms like risk, bank, market, liquidity, stock, price, and evidence are closely interrelated, while there is a lack of a similar topical center for unpublished papers and in addition, new terms like governance, corporate and effect gain relevance. It

Table 4
Contingency Table of Publication Dates

| Year of publication | EFA 2009 (NOR) | 2010 (GER) | Total |
|-------------------------|----------------|------------|-------|
| 2009 | 0 | – | 0 |
| 2010 | 16 | 4 | 20 |
| 2011 | 31 | 25 | 56 |
| 2012 | 42 | 37 | 79 |
| 2013 | 27 | 37 | 64 |
| 2014 | 11 | 34 | 45 |
| 2015 | 12 | 7 | 19 |
| 2016 | 3 | 4 | 7 |
| 2017 | 4 | 2 | 6 |
| 2018 | 0 | 1 | 1 |
| 2019 | 2 | 1 | 3 |
| 2020 | 0 | 0 | 0 |
| Published | 148 | 152 | 300 |
| Not published | 67 | 63 | 130 |
| Total in sample | 215 | 215 | 430 |
| Missing information | 1 | 5 | 6 |
| Papers accepted | 216 | 220 | 436 |
| Years until publication | 3.46 | 2.88 | 3.17 |

Notes: This table shows the absolute number of papers in our sample presented at the EFA conference in the years 2009 and 2010 by year of publication and in aggregate.

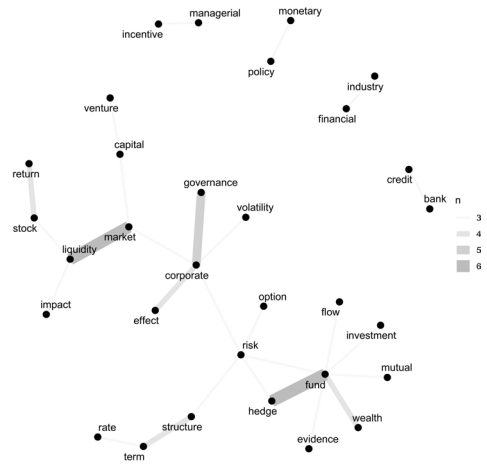
seems that published papers are more concerned with issues of asset pricing while unsuccessful ones are addressing typically “other” topics like governance issues.

In Panel B, publications are distinguished between those in low-impact journals on the one side and high-impact journals on the other. Low-impact journals are defined as such with only up to 0.5 points according to the ranking for business administration of the German newspaper Handelsblatt as of 2012.¹ High-impact journals are those with more than 0.5 Handelsblatt points (HB) with a possible maximum value of 1 point. Apparently, the results for published papers in Panel A are mainly caused by those published in high-impact journals, thus confirming our previous findings.

¹ <https://www.handelsblatt.com/politik/konjunktur/bwl-ranking/-bwl-ranking-2012-bwl-ranking-2012-methodik-und-zeitschriftenliste/6758368.html>.

Panel A: Networks by publication success

Unpublished papers



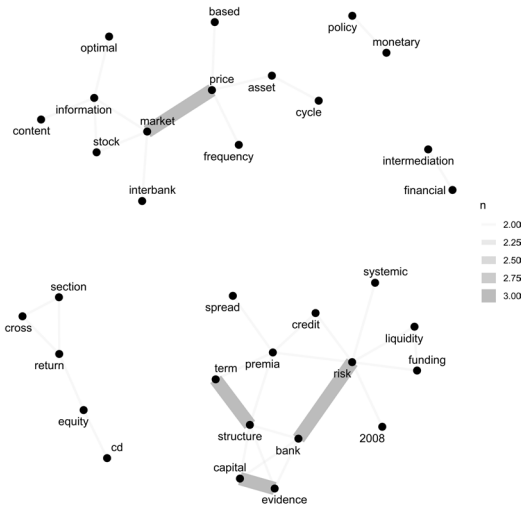
Published papers



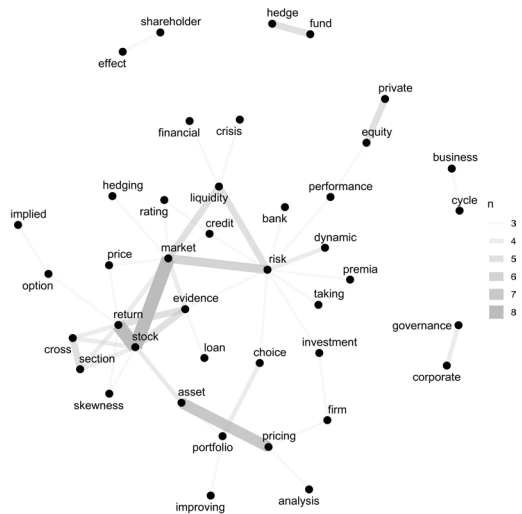
Figure 1: Bi-Grams of EFA Conference Papers (Panel A)

Panel B: Networks by Publication Quality

Low impact ($HB \leq 0.5$)



High impact ($HB > 0.5$)



Notes: This figure presents the most common word pairs in working paper titles presented at EFA conferences in the years 2009 and 2010. Panel A shows the most common word pairs for unpublished and published papers and Panel B for low and high impact publications, respectively. The number of co-occurrences (n) is indicated by the thickness of the connection line; the most common words are located in the center.

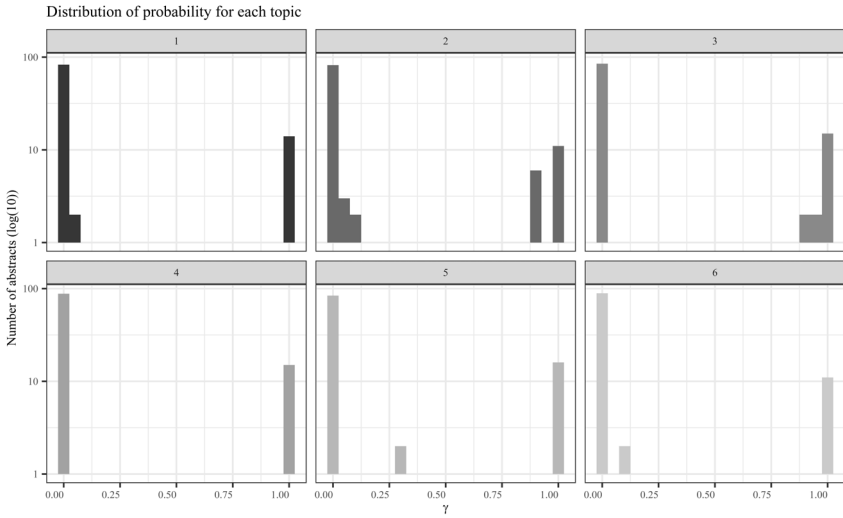
Figure 1: Bi-Grams of EFA Conference Papers (Panel B)

An additional analysis is carried out applying the state-of-the-art machine learning method “Latent Dirichlet Allocation” (LDA) (Blei et al., 2003) to the abstracts of the EFA 2009 and 2010 conference papers. LDA is an advanced textual analysis technique that regards each abstract as a union of topics and each topic as a union of words. This approach mimics natural language processing, as it classifies documents into natural groups without any pre-specified topics. LDA applies mathematics that reduces the dimensionality of datasets and hence compares to a factor analysis (Dyer et al. 2017). This reduction is achieved by defining a topic as a collection of words where each word is assigned a probability of belonging to a topic. Thus, LDA connects documents with probability distributions belonging to topics, so that one document can contain several topics. For further reference, *Silge and Robinson (2016)* provide an overview of how to apply text mining with the programming language R.

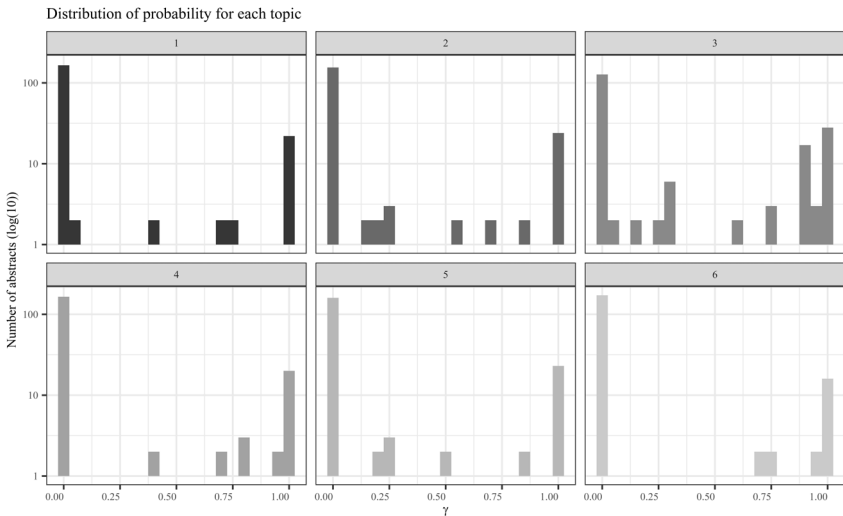
For both published and unpublished papers, the most relevant six topics each described by eight words are determined. Due to space constraints and in the light of the findings according to Figure 1, these results are only available from the author upon request. Instead, Figure 2 presents for each topic of unpublished and published papers, the probability (γ) that a given abstract belongs to a given topic. Without going into details, it is apparent that for unpublished papers, probabilities are typically either near zero or near one. For example, in Panel A (unpublished), it can be seen that the majority of the unpublished papers does not cover Topic 4 (characterized by the terms bank, credit, loan, risk, capital, stock, performance, sample) at all (0%), whereas most of the remaining abstracts exhibit a coverage of 100% for this topic. This makes it an “either-or topic”. The other unpublished papers have a similar U-shaped distribution among their related topics with clear classifications.

In contrast, for Panel B (published) the probability that a given abstract belongs to a given topic is more dispersed. For example, for Topics 2 (price, return, model, risk, asset, stock, investor, portfolio) and 3 (risk, market, factor, return, stock, bond, model, option), some abstracts are assigned probabilities of around 25% to 75% of belonging to the respective topic. Thus, published papers seem to touch on rather many identified topics which indicates that the set of published papers differs only by nuances or uses a similar language to describe their contents in the abstract while the subjects of the unpublished papers seem to be more distinctive. Rather interestingly, this finding contradicts to some degree that of *Breuer and Steininger (2020)* with respect to journal and conference papers in the context of real estate finance. As a possible aim of future research, it might pay to dig into this issue somewhat deeper.

Panel A: Unpublished



Panel B: Published



Notes: This figure shows the probability (γ) that an abstract from EFA conference papers belongs to a specific topic. The y-axis counts the number of abstracts and is log(10) scaled. Panel A depicts the papers that were not published and Panel B the papers that were published in the years after the conference.

Figure 2: Probability of an Abstract Belonging to a Topic

IV. Conclusion

The EFA annual meeting 2020 was special as it took place in a completely virtual way. Digitalization as a buzzword of our time also showed some impact with respect to the most topical presentations of this conference. Machine learning algorithms are gaining relevance. As an example of their application, differences between published and unpublished papers from the EFA annual meetings 2009 in Bergen and 2010 in Frankfurt were presented. Published papers center mainly around issues of asset pricing while unsuccessful ones are more concerned with other issues like corporate governance and exhibit larger topical differences between each other. The question of what distinguishes a successful conference paper from a one with more problems getting published in a (high-quality) scientific journal seems to be of general interest and might be addressed in more detail by future research.

References

- Blei, D. M./Ng, A. Y./Jordan, M. I. (2003):* Latent dirichlet allocation. *Journal of Machine Learning Research* Vol. 3(1): 993–1022.
- Breuer, W./Steininger, B. I. (2020):* Recent trends in real estate research: a comparison of recent working papers and publications using machine learning algorithms. *Journal of Business Economics* Vol. 90 (1): 963–974.
- Dyer, T./Lang, M./Stice-Lawrence, L. (2017):* The evolution of 10-K textual disclosure: evidence from latent dirichlet allocation, *Journal of Accounting and Economics* Vol. 64 (2): 221–245.
- Silge, J./Robinson, D. (2016):* tidytext: text mining and analysis using tidy data principles in R, *Journal of Open Source Software* Vol. 3(1): 37.