KING'S
College
LONDON

The visibility of women in the UK business and finance press

Laura Jones and Dr Ylva Baeckstrom

September 2020

Laura Jones is a Research Associate at the Global Institute for Women's Leadership at King's College London

Dr Ylva Baeckstrom is a Lecturer in Banking and Finance at King's Business School

Introduction

Relative to men, women continue to be severely under-represented and underpaid at senior levels in business, especially in the finance industry. This issue has gained increasing prominence in the UK since the Gadhia review, the launch of the HM Treasury Women in Finance Charter in March 2016 and other initiatives. Yet, analysis of progress suggests that while improvements are happening, they are at best slow, incremental and vulnerable to regression.[1]

There are many reasons why improving gender balance in finance is important, including the multiple business benefits such as the imperative to retain key talent and to attract the growing market of female consumers. Ensuring that women's voices are visible in public reporting of business and finance news is critical. Not only to ensure that diverse perspectives are heard, but also to role model and normalise female leadership for the next generation. Role modelling senior women in business and finance is imperative to challenge the stereotype that business, leadership and finance are male domains. It creates associations between the female gender and success, making careers in finance accessible for young women and being consumers of financial services attractive to women in our society.

Therefore, with this research we complement the analysis of the gender gap in female representation and pay in the sector by examining how visible women are in the business and finance press, and how this has changed across time. We make use of big data techniques to analyse over 240,000 newspaper articles from the Guardian, the Financial Times, the Economist and City A.M.

Background and our approach

Traditionally media researchers have hand coded small numbers of newspaper articles to understand how often women are mentioned in certain types of news stories or featured as experts. More recent studies have started to exploit the potential of big data to explore the representation of women in the media at scale. Researchers from Bristol and Cardiff Universities who analysed over 2 million articles from 950 English language news outlets from 2015 found that women made up just 20% of people featured in articles classified as 'business', with business ranking as the second most gender unequal article category behind sport.[2]

Sleeman (2019) made use of the Guardian's open access archive to look at the changing representation of women in the creative sections of the Guardian (e.g. fashion, music and film) over the last twenty years. Sleeman finds that the representation of women and women's voices was stagnant between 2000 and 2014 at just under 30%, with a sharp increase in recent years to just over 40%.[3]

Inspired by these developments and to fill the gap in research that considers news reporting of women in business we investigate the gender balance in press representation in the business and finance press over time.

Our analysis is split into two sections.

Part 1. Guardian analysis 2000–2020 (Figs 1–3)

We analyse 283,909 articles published in the Guardian: 128,828 from the Business section, 100,085 from the Politics section and 54,996 from the Money section.

- The Guardian was selected because of its wide online readership and because it offers open access to its current and historical content to enable analysis using Natural Language Processing ('NLP') techniques.

- We track mentions of male and female **pronouns** across time in the Business, Politics and Money sections (fig 1), and then look at particular finance related **key-word tags** in Business articles (fig 2 and fig 3).

Part 2. The financial press 2014–2020 (Figs 4–8)

We then focus on three business and finance focused papers: The Financial Times, The Economist and City A.M. We sourced 112,377 articles published between January 2014 and June 2020 from the API provided by website eventregistry.org. This allowed access to the articles in a format that facilitates computational analysis.

- Looking at articles classified into particular **categories** within the financial press (fig 4) we repeat the **pronoun analysis**. This is complemented with **entity extraction**, where we identify individual mentions of named persons (fig 5),

and **analysis of quotations** (fig 6) to work out the proportion of quotations that can be attributed to men or women. The Wikidata Query Service is then used to extract information about the occupations and genders of notable individuals mentions in the financial press (figs 7 and 8).

1. Guardian analysis 2000–2020

Slow progress and backsliding

Fig 1: The percentage of gendered pronouns that are female in the Business, Money and Politics sections of the Guardian 2000-2020
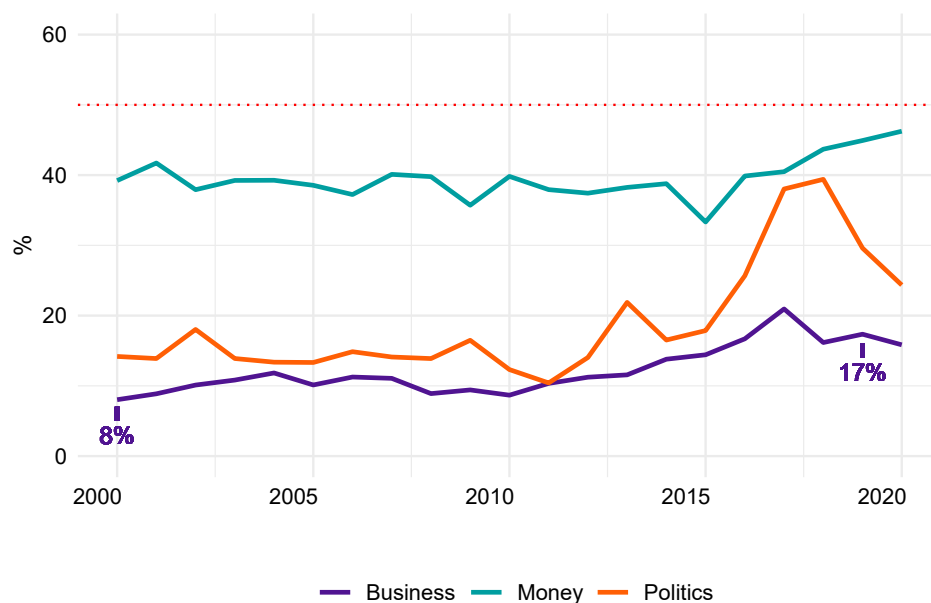
Based on pronoun analysis of 128,828 articles from the Business section, 54,996 articles from the Money section and 100,085 articles from the Politics section 2000-2020

There has been steady but very slow progress from a low base over the last 20 years. In 2000, 8% of gendered pronouns mentioned in the Business section of the Guardian were female, increasing to 11% in 2007 before stagnating, slipping then slowly rising after 2013. We note a Theresa May related spike in 2017. Despite this, we only found a 9 percentage point rise between 2000 and 2019: a rate of less than half a percentage point a year. If this trend continues it will take at least a further 60 years to reach equal gender representation.

Reporting in the Business section covers both international and domestic topics, and reports on figures from the world of business, politics and beyond. However the 17% female representation in the Guardian during 2019 trails far behind the 32% female representation in senior management across the Women in Finance Charter Signatory firms identified by New Financial in 2019[4] – or the 23% representation of women on boards across the financial sector in 2015 included in the Gadhia review.[5]

By comparison female pronouns are proportionally more common in articles in the Money (focused on personal finance) and Politics sections. In the latter section the female share of pronouns rose from 12% in 1999 to 24% during the first half of 2020, with a spike at around 38% during the Theresa May era.

Steady change in some finance related topics but hitting a ceiling with others

As the Guardian tags articles with keywords we could zoom in on the representation of women in articles that cover particular finance related topics. We selected keywords that appeared often enough to have substantial samples in each year.

Most finance related topics have seen a slow increase in female representation over the years. Articles tagged as related to 'banking', for example, had just 6% female pronouns in 2007 – rising to 20% in 2019. Other topics, such as retail, have seen much more fluctuation over the years but peaked above the all-business average of 24% in 2013.

Fig 2: The percentage of gendered pronouns that are female in the Business section of The Guardian, by select keywords

A dashed line indicates that the tag had less than 100 articles in the year the line begins
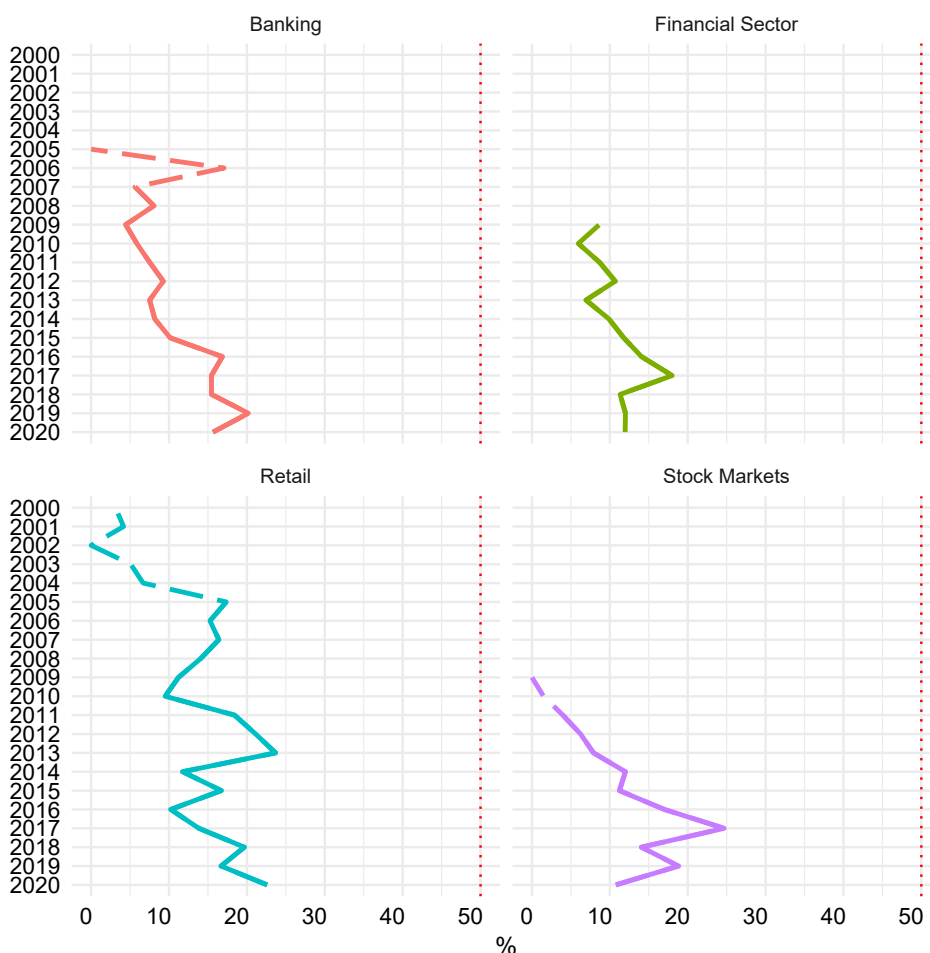
Pronoun analysis of 128,828 articles from the Guardian Business section 2000-2020

Pooling data across the years provides a clearer picture of long-term trends and allows analysis of less frequent tags. We compare the years 2011-2015 and 2016-2020.

Fig 3: The percentage of gendered pronouns that are female in the Business section of The Guardian, by select keywords, 2011-2015 and 2016-2020
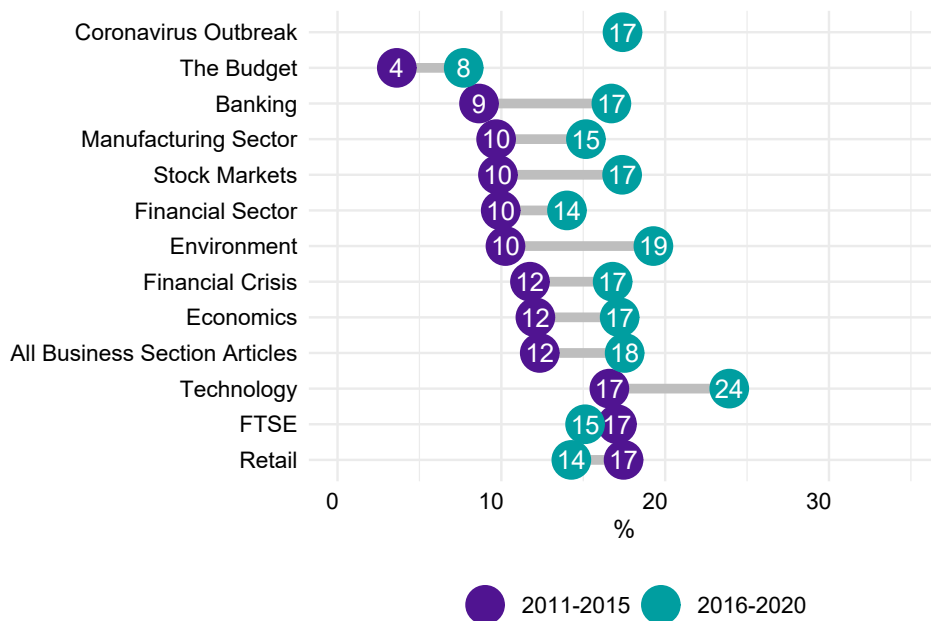
Pronoun analysis of 128,828 articles from the Guardian Business section 2000-2020

Again, the overall picture is one of slow, incremental progress in topics like banking, the manufacturing sector and stock markets, which started from very low bases. More worryingly in the finance related topics that started out with greater female representation – such as FTSE and retail – there has been a *decrease* in the last four years, with an apparent ceiling of female representation at about 17%.

Only the articles tagged 'technology' seem to have been able to buck this trend, outperforming the others to reach 24% female representation between 2016-2020. Unlike finance, this is above the estimated 21% female employment in the technology sector in 2018.[6]

In the context of ensuring that female voices are present in key decision-making moments two other findings are worth highlighting. The first is the low representation of female pronouns in articles related to the annual Budget: 4% in 2011-2015 and 8% in 2016-2020. While this represents a doubling of female pronouns between the time periods, it remains a depressingly small proportion.

Finally, as businesses face unprecedented challenges in the wake of the COVID-19 pandemic it is more important than ever that a diversity of voices are heard. Yet only 17% of pronouns in business articles related to the pandemic are female.

2. The financial press 2014–2020

For the second part of this analysis we focus on three business and finance focused sources – The Financial Times, The Economist and City A.M, analysing articles from 2014-2020.

Fig 4: Percentage of gendered pronouns that are female in the financial press, and the Guardian Business section, 2014-2020
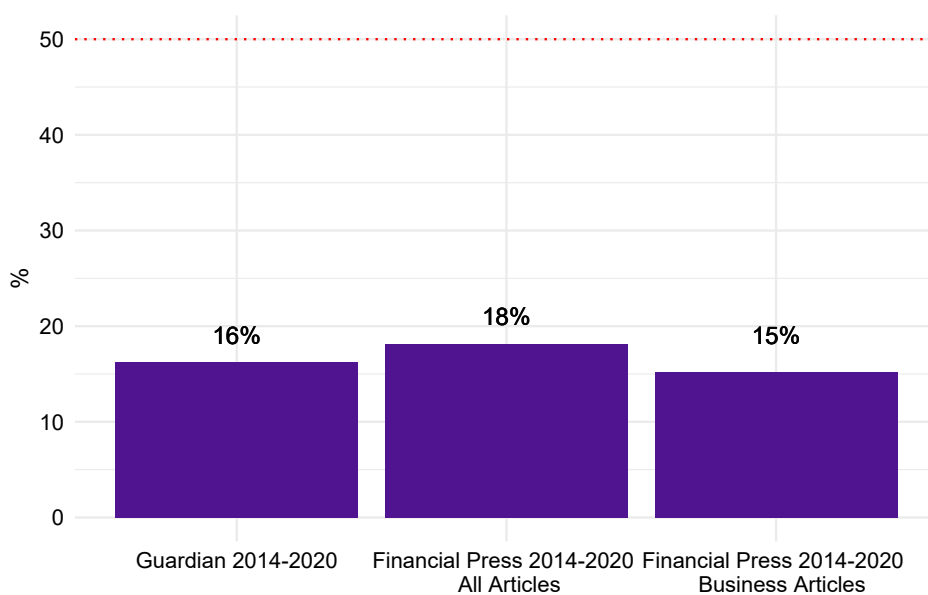
Pronoun analysis of 27,508 articles from the Guardian Business section 2014-2020 and 112,377 from the Financial Times, The Economist and City A.M. 2014-2020

Overall, 18% of pronouns in the financial press are female - slightly more than in the Guardian Business section. However, this doesn't represent a straightforward comparison as not all articles in these outlets will be business focused. To create a fairer comparison we use standardised topic classification tags to select only those articles that are focused on business and finance topics. These tags assign articles to one of 8 pre-defined categories based on their linguistic content.[7] We classify 72% of the articles from the financial press as belonging to the business category. Female pronouns *fall* from 18% to 15%, i.e. lower than the female representation in the Guardian.

Beyond pronouns to individuals

Next, we apply an algorithm to extract every mention of a named person from the text and infer their gender. The algorithm then resolves co-references within an article i.e., identifies that "Alison Rose" and "Rose" within the same article refer to the same person. In doing so we identify a list of the references

to unique individuals, along with their gender, to analyse female representation with a higher degree of sophistication than the pronoun approach.

Fig 5: Percentage of gendered pronouns and individual mentions of people that are female in business articles in the Financial Press 2014-2020
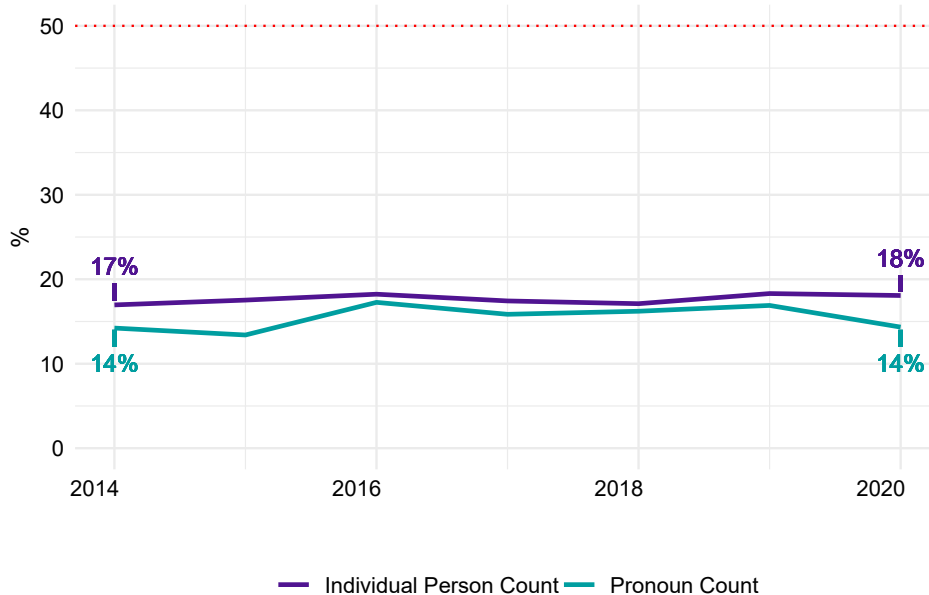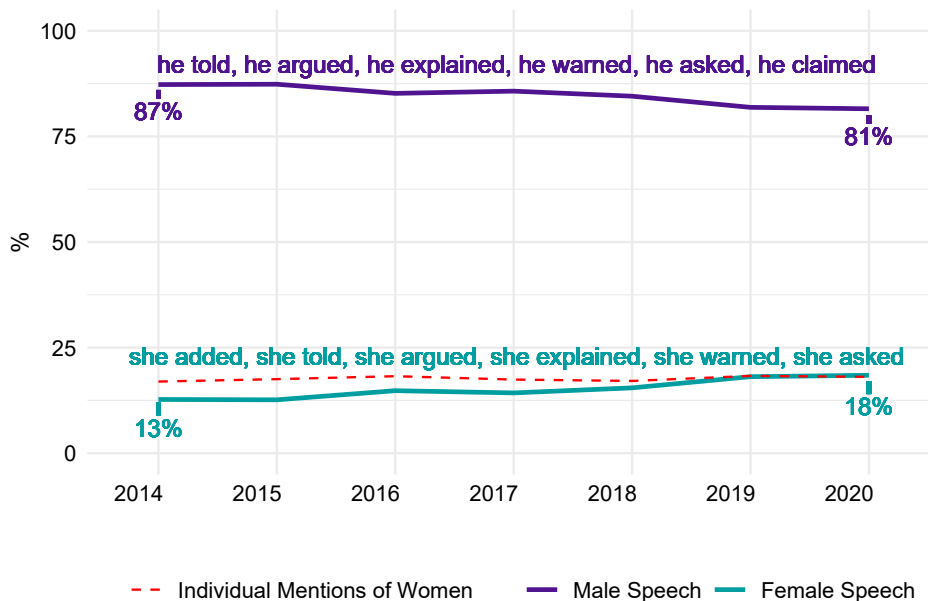

Named entity recognition and pronoun analysis of 81,187 business articles from the Financial Times, The Economist and City A.M. 2014-2020

Comparing named individual and gender pronoun mentions yield very similar results. Neither sees a significant change across the four-year period, although the 'named person' mentions tell a slightly more optimistic story. One reason why female representation might be slightly higher among *individuals* mentioned than among gendered *pronouns* could be that women, when mentioned, receive less article space and therefore generate fewer associated gendered pronouns.

A seat at the table?

To differentiate between a woman being merely *mentioned* and *receiving space* to express her opinions in the article, we extracted every instance where a gendered pronoun was followed by a verb. Every verb that appeared at least 100 times was tagged as indicative of speech (e.g., "he said", "she argued") or not (e.g., "he went").

Fig 6: Ratio of male to female speech in business articles in the financial press


Named entity recognition and quotation analysis of 81,187 business articles from the Financial Times, The Economist and City A.M. 2014-2020

Business related articles are heavily dominated by male voices. In 2014 only 13% of quotes were attributed to women, and women were even less likely to be quoted than to be mentioned. But, unlike with individual mentions of women, there are signs of a slow but steady increase in female voices between 2014 and 2020. By 2020, 18% of quotes were attributable to women. Until 2019 women were slightly more likely to be mentioned than have their voices reported, but in the last two years this imbalance had been redressed.

Role models for future generations?

In our final analysis we wanted to understand the roles that the women represented in the press were playing. If an article contained a reference to a well-known person then the article was tagged with that person's Wikipedia page. We then used this information to automatically pull from Wikipedia information on that person's gender and occupation. The results are displayed in figure 7 below. The size of the blocks corresponds to the relative frequency with with people from that occupational grouping were mentioned.
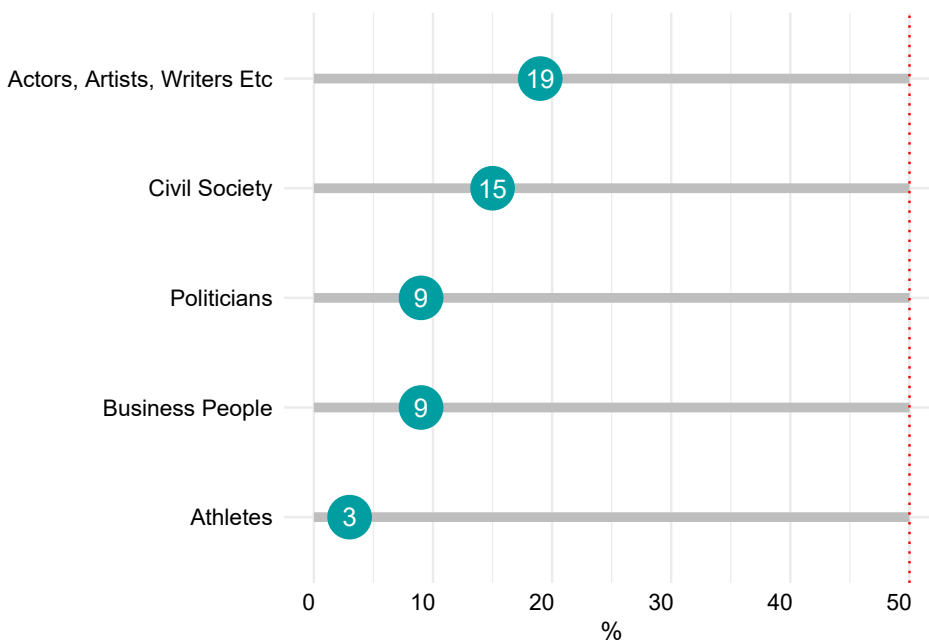
Fig 7: Occupations of well-known people mentioned in the financial press
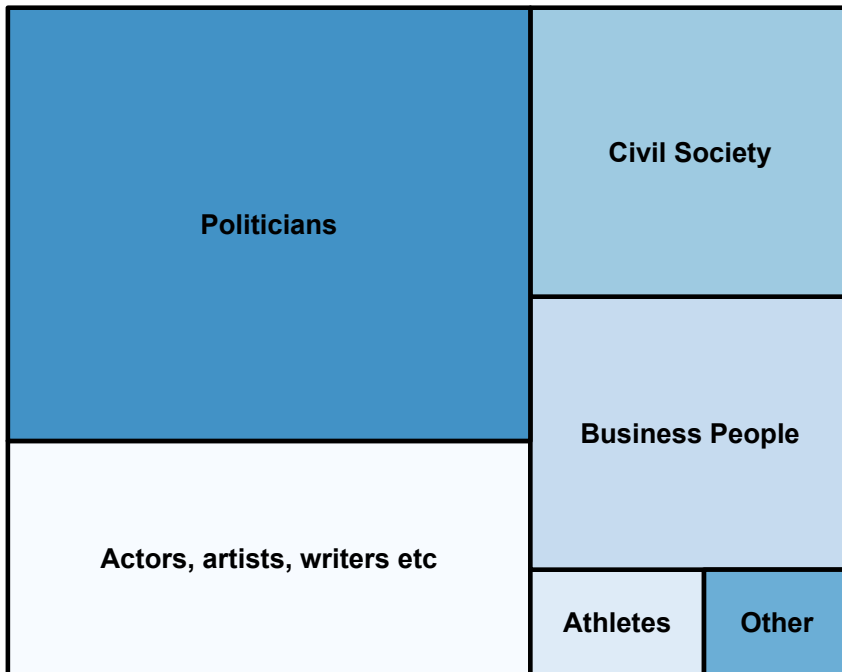
Based on information extracted from Wikipedia on 18,838 unique well-known people mentioned 156,070 times in 112,377 articles from the Financial Times, the Economist and City A.M. 2014-2020

Using this information we were able to analyse women's share of representation among each category (figure 8).

Fig 8: Women's representation among notable people mentioned in the financial press 2014-2020

Based on information extracted from Wikipedia on 18,838 unique well-known people mentioned 156,070 times in 112,377 articles from the Financial Times, the Economist and City A.M. 2014-2020

Just 9% of the mentions of politicians and notable business people in the financial press are of women. This figure is lower than the female share of quoted speech (between 13% and 18%), and much lower than women's representation within these professions in the wider population. In the UK women make up 34% of MPs[8] and 33% of those sitting on FTSE boards.[9]

A fifth of the space

This report demonstrates how gender inequality in female representation in leadership positions and pay in business and finance extends to the press. Women are in fact more underrepresented in the financial press than they are in leadership positions.

Whether documenting women's relative share of gendered pronouns, individual mentions, quoted speech or notable people, a consistent story emerges with women receiving *less than a fifth* of the space. Although there are signs of slow but steady progress, this research demonstrates just how badly business and finance fares compared to other areas.

These stark gender imbalances in the business and finance press identify an urgent need for more female role models, for women to be seen and their voices to be heard. The lack of female role models exacerbates stereotypes and gender bias instead of challenging them. Today it seems that written media tell us that women's successes and voices are worth less than men's and there may well be a strong link between gender inequality in female representation in leadership positions and pay in finance and female presence in business press.

This emphasises the importance of initiatives like the Women in Finance Charter to keep pushing for more change.

Methodology endnote

Data

The analysis in this paper is based on two key sources of information:

283,909 articles from the Guardian Jan 1999 – June 2020 (Figs 1-3)

The focus of the first part of the analysis was on 128,828 articles published in the Business section of the Guardian between January 1999 and June 2020. The Guardian was selected for analysis because, unlike other major British newspapers, it offers open access to its current and historical content via an application programming interface (API). This makes its straightforward to analyse its content using Natural Language Processing (NLP) techniques. The Business section of the Guardian was selected as its content – "the latest financial, market & economic news and analysis" – was the most relevant for analysing the representation of women in banking and finance.

Alongside these articles a further 100,085 articles from the Politics section and 54,996 articles from the Money section were selected to act as comparators. Articles in the Money section focus on "the latest personal finance and money news".

112,377 articles from three top financial papers January 2014 – June 2020 (Figs 4-8)

Although the Guardian's open access API makes it the most convenient source for longitudinal analysis of the representation of women in the finance pages it may not necessarily be the most relevant or visible source of financial and business news. Therefore, for the second part of this analysis we focus on three business and finance focused sources – The Financial Times, The Economist and City A.M. We sourced 112,377 articles printed on these newspapers' websites between January 2014 and June 2020 from the API provided by website eventregistry.org. This allows us to access the articles in a format that facilitates easy computational analysis, but we are unable to access articles printed before 2014. It should be noted that 72% of these articles were drawn from ft.com, and so analysis from this section is heavily weighted towards articles published by the Financial Times.

Methodology

To analyse the representation of women within these sources we applied a variety of techniques drawn from the fields of machine learning and natural language processing.

Pronoun analysis (Figs 1-5) – Each article was analysed to collect mentions of male and female pronouns. We then looked at how the distribution of these pronouns varied across time and across various keyword tags (the Guardian) or by article topic classification (Finance sources).

Entity extraction (Fig 5) – We analysed the full text of every article from the finance sources using a suite of NLP tools developed by the University of Sheffield.[10] These extract every mention of a named person and then infer their gender. We are additionally able to use these tools to resolve co-references within an article i.e. to identify that "Boris Johnson" and "Johnson" within the same article refer to the same person. This means we can identify, for each article, a list of the references to unique individuals, along with an estimation of their gender. Although an adapted version of this tool has been used previously on a very similar set of documents with a high degree of accuracy,[11] it is not 100% accurate, and so we therefore complement the entity extraction with pronoun analysis detailed above, and the analysis of quotations described below.

Quotations (Fig 6) – Following the analysis by Sleeman 2019[12] we analysed every article in the finance sources by searching for a gendered pronoun, followed by a verb. We manually tagged every verb that appeared at least 100 times as indicative of speech (e.g. "he said", "she argued") or not (e.g. "he went").

Entity linking (Figs 7 and 8) – As a final piece of analysis we make use of some of the additional meta-information provided by eventregistry. org alongside each article. Eventregistry uses AI to scan the content of each article, identify whether a notable person is mentioned, and provide a link to their Wikipedia page as meta-data. For the articles published by the financial press this resulted in 156,070 mentions of 18,838 unique people. Using the Wiki Data Query Service we were able to pull information on gender and occupation for 81% of these people. Occupations that appeared at least 25 times were manually coded into one of 6 categories – Politicians; Business People; Athletes; Civil Society; Actors, artists, writers etc; and Other. Using this method we were able to categorise 96% of those for whom we had information. To prevent people that were listed as having more than one occupation on Wikipedia from being double counted in the results[13] we then established a hierarchy of professions and assigned those with two categories into the top category in the hierarchy. Politicians were ranked first, as they often have previous careers but are generally appearing in the press in their role as a politician. Business people were ranked second as this was an analysis of the financial press, next civil society and athletes, and finally actors, artists, writers etc, as this also included the category 'writer' which was a second categorisation for a lot of people. As a robustness check we tested running the analysis again and allowing people to appear in multiple categories. Although this shifted the relative size of the blocks in the tree map (fig 7), it did not materially affect the proportion of women in each category (fig 8).

References

1. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894197/HM_Treasury_Women_in_Finance_Charter_-_Annual_Review_2019.pdf

2. Jia, S., Lansdall-Welfare, T., Sudhahar, S., Carter, C., Cristianini, N., 2016. Women Are Seen More than Heard in Online Newspapers. PLOS ONE 11, e0148434. https://doi.org/10.1371/journal.pone.0148434

3. https://data-viz.nesta.org.uk/measuring-gender-imbalances-reporting-creative-industries/index.html

4. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894197/HM_Treasury_Women_in_Finance_Charter_-_Annual_Review_2019.pdf

5. https://uk.virginmoney.com/virgin/assets/pdf/Virgin-Money-Empowering-Productivity-Report.pdf

6. https://www.gov.uk/government/statistics/dcms-sectors-economic-estimates-2018-employment

7. Categorisation based on 8 general categories used to categorize news content – Business, Politics, Technology, Arts and Entertainment, Sports, Health, Science and Environment.

8. https://commonslibrary.parliament.uk/research-briefings/sn01250/

9. https://www.businessleader.co.uk/one-third-of-all-board-members-at-ftse-100-companies-are-now-female-according-to-new-data-but-the-government-says-more-work-is-still-needed-to-create-a-true-gender-balance/79187/

10. The ANNIE plugin of GATE. Annie stands for 'A Nearly-New Information Extraction System'. Gate stands for 'General Architecture for Text Engineering'.

11. Jia, S., et al., 2016. Op cit.

12. https://data-viz.nesta.org.uk/measuring-gender-imbalances-reporting-creative-industries/index.html

13. For example, Boris Johnson is listed as both 'politician' and 'journalist'.

giwl.kcl.ac.uk
@GIWLkings
#womensleadership
giwl@kcl.ac.uk