Coverage bias on Wikipedia? Evidence from biographies of German Members of Parliament[*†]

Anna Kerkhof[‡] Johannes Münster[§]

January 10, 2017

Abstract

We investigate whether Wikipedia has a coverage bias in its biographies of members of the current German Parliament (Bundestag). Biographies of MPs belonging to the Social Democrats (SPD) are, on average, shorter than biographies of MPs from other parties. This correlation remains after controlling for individual characteristics and demographic properties of the electoral district.

We discuss several potential explanations for these correlations. First, they could be driven by unobserved heterogeneity between the politicians of the different parties, which might even lead neutral authors to allocate less space to MPs from the SPD. Second, the differences could be due to partisan writing. To weigh up the plausibility of these two competing explanations, we compare MPs' coverage on the German Wikipedia with the English Wikipedia, arguing that partisan writers have no incentive to contribute to biographies of German MPs on the English language version of Wikipedia. Difference-in-differences estimation tends to support the role of lower partisan writing on behalf of the SPD.

[*]We thank Felix Arnold, Adrian Chadi, Maria Cubel, Matthew Ellman, Marc Fischer, Oliver Gürtler, Matthias Heinz, David Jaeger, André Kaiser, Sebastian Mertesacker, Florian Morath, Susanne Prantl, Santiago Sanchez-Pages, Wendelin Schnedler, Detlef Schoder, Marius Vogel, Dainis Zegners, participants of the SMYE 2016, of the Ammersee Workshop 2016, the 14th Annual Media Economics Workshop, and seminar participants of the DICE Brown Bag Seminar and the Paderborn Research Seminar for helpful comments and suggestions. We also thank Annika Bohnhorst and Julian Freitag for excellent research assistance, and Marc Egger for supporting us with his programming skills.

[†]Preliminary draft. Please check with the authors for updates before citing!

[‡]University of Cologne. E-mail: anna.kerkhof@wiso.uni-koeln.de

[§]University of Cologne. E-mail: johannes.muenster@uni-koeln.de

1 Introduction

Media, often referred to as "forth estate", are essential for modern democracies and societies. Recent research in economics has convincingly demonstrated the media's causal impact on a wide range of political and social outcomes, ranging from election outcomes to consumption and saving decisions to personal choices such as marriage, fertility, and divorce. Against this background, the existence of media biases is raising concerns.[1] The lion's share of the existing literature on media bias studies traditional media such as newspapers, television and radio. Today's media landscape, however, is changing at high speed, and the importance of the so called "fifth estate", i.e., social media and user generated content, is growing swiftly. Therefore, it does not suffice to investigate the biases of traditional media outlets. Rather, it is crucial to also study biases of online media.

In this paper, we study Wikipedia biographies of members of the current German Parliament, the 18th Bundestag. Our focus is on a clean identification whether Wikipedia has a *coverage bias* against MPs from specific political parties, defining coverage bias as an unequal coverage of equivalent cases.[2] In other words, we investigate whether MPs from a particular party receive less coverage relative to comparable MPs from another party. Investigating coverage, as opposed to content, has the advantage that it is not only a very comprehensible and traceable, but also a completely objective measure.

A prime challenge in investigating coverage bias so defined is the need to find equivalent cases. The current German Bundestag is particularly interesting in this respect, since there is a coalition government by Germany's two main parties, the center-right CDU/CSU (Christian Democrats) on the one hand, and the center-left SPD (Social Democrats) on the other. Therefore, any imbalanced coverage of members of these parties cannot simply be due to differences between the coverage of opposition versus government,[3] or of major versus small parties.

The German case is also interesting as almost all research on media bias has focussed on English language media, while research on biases of media in other languages is rare. Greenstein and Zhu (2012) and Brown (2011) have found a slight and decreasing bias towards the Democratic Party in the English Wikipedia's treatment of US politics. Given the profound differences between both the political systems and the media systems of the USA and Germany (Persson and Tabellini 2003, Hallin and Mancini 2004), it is a priori unclear whether these results have any external validity to the German

[1] E.g., Druckmann and Parkin (2005), DellaVigna and Kaplan (2007), and Chiang and Knight (2011) assess significant effects of media bias on voting decisions and public opinion.

[2] See D'Alessio and Allen (2000) and Groeling (2013) for discussions of different definitions of media bias.

[3] As pointed out by Green-Pedersen et al. (2015, p.1), a central finding of the literature on media coverage of politicians is the "incumbency bonus: the incumbent party or government leader gets more media coverage than the opposition or opposition leader". According to van Dalen (2012), actors from the government are more than twice as often the main actors in news items in German newspapers and television news broadcasts than actors from the parliament. See Vos (2014) for an overview over this literature.

Wikipedia. Several recent papers have argued that the content, reception and effects of media in multi-party democracies as in Germany differ in important respects from those in two-party democracies such as the USA (Benson 2009, Baum 2013, Wessler and Rinke 2014, Knobloch-Westerwick et al. 2015). Indeed, a naive extrapolation of the result that the English Wikipidia has a pro-liberal bias (Greenstein and Zhu 2012, Brown 2011) might suggest that, when comparing the two biggest German parties, the German Wikipedia is biased in favor of the SPD relative to the CDU/CSU. Our results, however, show exactly the opposite.

We find that biographies of MPs from the SPD are, on average, shorter than biographies of MPs from other parties; in particular, they are about 1/2 of a page shorter than biographies of MPs from the more conservative CDU/CSU.[4] The correlation between biography length and party membership remains after (i) controlling for observable MP characteristics, such as gender, political offices and experience, (ii) controlling for demographic properties of the electoral districts such as education, age, and population density, and (iii) eliminating obvious outliers such as current or former members of the cabinet (ministers) from better comparable MPs.

Our starting point for explaining this correlation is that there are two types of authors active on Wikipedia. On the one hand, there are neutral contributors who stick to the Wikipedia guidelines of maintaining a Neutral Point of View (NPOV, see also Section 2) and who allocate time and space to subjects according to how much can be said about them. On the other hand, there are partisan writers who want to extend and brighten the biographies of their respective MPs, because in low-information elections - such as the elections for the Bundestag - a high amount of coverage is usually beneficial for politicians (Burden 2002). There is no shortage of anecdotes on how politicians have manipulated their Wikipedia biographies in the past; journalists even claim that Wikipedia has become "part of the political strategy" (Cohen 2008).[5]

Thus, there are two potential explanations for the correlation between party membership and biography length. First, there could be unobserved heterogeneity that leads even neutral authors to allocate less space to MPs from the SPD. Second, the differences could be due to the activities of partisan writers. To weigh up the plausibility of these competing explanations, we exploit variation between the German and the English language versions of Wikipedia, and perform a difference-in-differences analysis. Our key argument is that partisan writers have no incentive to contribute to biographies of German MPs on the English Wikipedia, because German voters are unlikely to access English language information on German politicians. Under the additional

[4] To put the effect size into perspective, the median length corresponds to roughly 2 pages. This measure corresponds to the document obtained by choosing "Printable version" on Wikipedia. The effect of party membership is, however, much smaller than the effect of political standing such being a party head or a member of the cabinet.

[5] E.g., Sarah Palin's Wikipedia biography was extended only hours before she announced her candidateship for vice-president in 2008 (Cohen 2008). Noguchi (2008) reports that "[p]oliticians, including Vice President Joe Biden, [...] have their paid staff edit their Wikipedia pages to make them more favorable." In 2012, the biography of Christian Lindner, leader of the German Liberal Party, was presumably brightened by his staff (Beucker, 2013).

assumption that unobserved heterogeneity affects the length of German and English biographies equivalently, the difference-in-differences estimation allows us to obtain an unbiased estimate for the relative effect of party affiliation on German Wikipedia biography length. An estimate unequal to zero would indicate that Wikipedia coverage of comparable MPs from different parties is unbalanced, and hence that coverage is biased.

Our identification strategy differs slightly from conventional difference-in-differences designs. There is no time dimension and we do not consider a policy change, either. Instead, German versus English corresponds to after and before treatment, and being a member of a particular party corresponds to a particular type of treatment.[6] A further challenge for our identification strategy is that English Wikipedia biographies exist for only about 1/3 of the MPs. We conduct a Heckman Two Step and a Maximum Likelihood estimation to take potential selection bias into account.

The difference-in-differences regression results reveal that there is less partisan writing on behalf of the SPD than on behalf of any other party, hence, there is a coverage bias against MPs from the SPD. We point out that this result does not just reflect generic differences in party coverage in the German media landscape. Dallmann et al. (2015) investigate party coverage on four major German online news outlets, and find no evidence that the SPD is covered any less than the other parties.[7]

Additional results on patterns of partisan activity are in line with our main results: biographies of MPs from the SPD exhibit fewer images, a lower ratio of adjectives to words, and a smaller number of links to external websites under the control of the MP or her party. Images and adjectives brighten texts and contribute to a more positive coverage, and a high number of weblinks under party control indicates that Wikipedia is used more for election campaigns. Moreover, we find some evidence for fewer partisan writers being active on behalf of the SPD.

The paper proceeds as follows. Section 2 discusses the institutional background and related literature. Section 3 describes our dataset. The empirical strategy is discussed in Section 4; Section 5 presents the corresponding regression results. Section 6 illustrates further patterns of partisan activity on Wikipedia. Section 7 concludes.

2 Background

Wikipedia is a free online encyclopedia, its content is written by volunteers, and its operation is financed by donations. It attracts millions of readers every month (Wikipedia:Statistics). The German Wikipedia is the second oldest version of Wikipedia

[6] This approach is related to Mayzlin et al. (2014) who compare hotel reviews on Expedia with reviews on TripAdvisor. Unlike TripAdvisor, Expedia requires its users to have actually spent at least one night at a particular hotel to be able to publish a review on it, which makes it much more costly to fake a review on Expedia.

[7] For example, Dallmann et al. (2015) show that on any of the four outlets they investigate (faz.net, spiegel.de, taz.de, and zeit.de), the party acronym 'SPD' is contained in more article titles than the acronym of any other party, even when adding the counts for CDU and CSU and counting them as one party. Moreover, the average sentiment scales for the parties do not differ significantly.

(it started in March 2001) and the second biggest version of Wikipedia in terms of number of authors, edits, and admins. Moreover, with about 1.9 million articles, it is the fourth biggest by number of articles (after English, Swedish and Cebuano; the latter two are mainly written by bots). Wikipedia is ranked No. 7 on a list of Top Sites in Germany by alexa traffic, and it is often the first address in Internet research (Bitkom, 2011). Moreover, it plays an important role in the media information food chain: many media outlets use information from Wikipedia.[8]

Neutrality is one of Wikipedia's most fundamental principles: "All encyclopedic content on Wikipedia must be written from a neutral point of view (NPOV), which means representing fairly, proportionally, and, as far as possible, without editorial bias, all of the significant views that have been published by reliable sources on a topic" (Wikipedia:Neutral point of view). In a tutorial for Wikipedia authors, it is further explained that "[d]ifferent views don't all deserve equal space. Articles need to be interesting to attract and keep the attention of readers. For an entry in an encyclopedia, ideas also need to be important. The amount of space they deserve depends on their importance and how many interesting things can be said about them." (Wikipedia:NPOV tutorial). To study whether Wikipedia lives up to these aspirations, we compare the length of MPs' biographies, controlling both for observable characteristics, and for characteristics of the MPs' electoral districts.

Wikipedia's quality control mechanisms are very efficient.[9] The primary control is the community itself. Larry Sanger, co-founder of Wikipedia, refers to famous E. S. Raymond (1999): "Given enough eyeballs, all errors are shallow". According to Wikipedia, there are hundreds of times more well-meaning authors than bad ones, which means that although poor information can be added easily, fraudulent content, downright lies, and vandalism are rapidly spotted and removed by the community. On average, only a few minutes lie between a blatantly bad or harmful edit, and some editor acting on it (Wikipedia:Editorial oversight and control). In our analysis, we will therefore ignore the possibility that partisan writers might also obviously sabotage the Wikipedia biographies of MPs affiliated to other parties.

Various studies establish that Wikipedia exhibits only few factual errors (Giles 2005, for a survey see Mesgari et al. 2014), but there are frequent errors of omission (e.g., Bragues 2007, Mühlhauser and Oser 2008, Brown 2011), which motivates our focus on coverage. Moreover, it is often stated that coverage in Wikipedia suffers from a systemic bias induced by its contributors' demographics who are mostly English-

[8] This can be very nicely illustrated: The German politician *zu Guttenberg* is famous for his many first names. When he became minister in 2009, a couple of German newspapers published a full list of his first names, including "Wilhelm", which does *not* belong to them. Shortly before that incidence, an anonymous user smuggled "Wilhelm" into zu Guttenberg's Wikipedia biography. Originally, he wanted to find out how long it would take until someone detected the mistake (BILDblog 2009). Apparently, the journalists did not do any research themselves, but simply copied from Wikipedia. *SPIEGEL Online* even reported that zu Guttenberg had introduced himself as "Wilhelm". This allowed the anonymous Wikipedian to include a valid reference for the fake first name into zu Guttenberg's Wikipedia biography. As a consequence, even high quality newspapers such as *Sueddeutsche Zeitung* and *Rheinische Post* started to report "Wilhelm" as a first name.

[9] See Stvilia et al. (2007) for a survey on Wikipedia quality control mechanisms.

speaking, white, male, and Internet affine (Halavais and Lackaff 2008), leading, e.g., to the underrepresentation of women (Reagle and Rhue 2011). In our analysis, we therefore also investigate whether the gender of an MP makes a difference for Wikipedia coverage. While biographies of women are shorter on average, the effect loses statistical significance when we control for the MPs' political experiences, and even changes sign (while still being statistically insignificant) when adding controls for the demography of the MPs' electoral districts. We thus do not find evidence for a gender bias in our data.

Literature. Our paper contributes to several strands of literature. First of all, it is linked to the growing economics literature on political media bias. The empirical research in this area can be roughly classified into three fields (see, e.g., Puglisi and Snyder 2016). One group of papers analyzes the *explicit* political behavior of the media. E.g., Ansolabehere et al. (2006) measure endorsement patterns of a series of newspapers between 1940 and 2002. Puglisi and Snyder (2015) use data on ballot propositions to determine the relative ideological positions of newspapers, voters, interest groups, and political parties. Our investigation of coverage bias is, however, more closely related to research on *implicit* political behavior.

There are several papers that measure implicit political behavior by comparing the speech of media outlets with the speech of politicians. In their seminal paper, Groseclose and Milyo (2005) construct an ideological index by counting the times media outlets cite particular think tanks, and compare this number to the times members of Congress cite the same groups. Gentzkow and Shapiro (2010) measure the similarity of a media outlet's language to that of a congressional Republican or Democrat to construct a slant index.

Given our dataset, the present paper is related to Greenstein and Zhu (2012), who apply the method by Gentzkow and Shapiro (2010) to measure the slant in Wikipedia articles on political issues. They find evidence for a bias in favor of the Democrats in Wikipedia's early years that has, however, diminished due to the entry of articles with opposite slants. Similarly, Greenstein and Zhu (2015) examine Wikipedia's slant relative to the Encyclopedia Britannica. In absolute terms, they find Wikipedia's slant to be stronger; regarding slant per word, however, the encyclopedias are very similar. Our work differs from these approaches in that we do not measure media bias in terms of ideological *content*, but in terms of *coverage*.

Our paper is most tightly linked to a third group of empirical studies on media bias that measures the implicit political behavior of the media by the amount of coverage devoted to various issues. The majority of research examines the agenda-setting behavior of the media. Studies agree that the abuse of agenda-setting power belongs to the most harmful behaviors of the media, because the amount of coverage of a topic is likely to influence consumers' perception on its importance (e.g., Larcinese et al. 2011). Puglisi (2011) examines a dataset of the New York Times from 1946 to 1997. He observes that during US presidential campaigns, the New York Times gives more emphasis to topics perceived as favorable for the Democrats when the incumbent president is Republican. Puglisi and Snyder (2011) find that partisan biases on the editorial pages of newspapers

are strongly correlated with partisan biases in the coverage of scandals. Larcinese et al. (2011) investigate the correlation between the endorsement policy of newspapers and the coverage of economic news. They show that, compared to pro-Republican newspapers, pro-Democratic newspapers systematically devote more coverage to high unemployment when the incumbent president is Republican.

Though closely related, our analysis of coverage bias exhibits subtle differences to former research in this area. First of all, most studies mentioned above focus on bias in the traditional media. Meanwhile, little is known about political bias in the online media. By studying coverage bias on a UGC platform, we contribute to fill this gap. Second, while traditional media outlets are clearly actively involved in the emergence of media bias, this is not so clear for UGC platforms. Here, media bias might not be induced by (decisions of) the outlet itself, but by external contributors. E.g., Wikipedia does not face the typical media outlet's decision whether to report on a specific topic or not. According to Wikipedia's notability guidelines (Wikipedia:Notability), MPs are generally relevant enough to be written about. Indeed, every MP in our dataset is "covered".[10] Consequently, any media bias can exclusively originate from differences in the *extent* of coverage, which is affected by numerous external factors. Third, while media bias has been studied extensively for US media outlets, the German media landscape remains largely unexplored. Using biographies of German MPs on Wikipedia, our paper contributes to form a clearer image about potential biases in the German media landscape.

Partisan bias has been studied in the context of political science, too (see D'Alessio and Allen 2000, and Groeling 2013 for surveys). Closely related to our work is Brown (2011), who analyzes Wikipedia coverage of US gubernatorial candidates, although not explicitly against the background of media bias. He finds a positive correlation between the probability of having a Wikipedia biography and being legislative leader, experienced, technologically savvy, and having a high district population. Being Republican lowers the probability of having a biography; the length of an article is affected similarly. In contrast to Brown (2011), by comparing coverage on the German and English version of Wikipedia, we present evidence for a *causal* relationship of party affiliation on coverage.

Since we pursue the clean measurement of coverage bias on a UGC media platform, our investigation ties together research on media bias and on user generated content. The paper is closely related to the empirical methods that former studies on UGC have employed (see Luca 2016 for a survey). Studies on promotional content are particularly close to our work. E.g., Mayzlin et al. (2014) compare hotel reviews from Expedia and TripAdvisor exploiting the fact that, unlike TripAdvisor, Expedia identifies people who have booked a hotel through its platform. Hence, they use the fact that it is more costly to fake a review on Expedia than on TripAdvisor for identification.

[10] There is a Wikipedia biography on every member of the Bundestag since the founding of the FRG.

We use a similar identification strategy to disclose partisan writing on Wikipedia. We argue that German voters are unlikely to read the English language biographies of German MPs. Therefore, it is much less attractive for partisan writers to extend those articles. Consequently, unlike the German sites, we expect the English language biographies to correspond to neutral writing, and use this variation for a difference-in-differences estimation. The novelty of our approach is thus to use variation *within* a UGC media platform instead of variation across platforms. A further difference to the literature on promotional content is the focus of our study. Previous research has primarily examined promotional reviews that might affect consumer demand. In contrast to that, we examine partisan writing on a UGC platform that might have an effect on voting decisions.

Finally, our paper contributes to Wikipedia specific research in various fields. A first branch of literature studies Wikipedia authors and contributions. Zhang and Zhu (2011) exploit an exogenous decrease of group size in the Chinese Wikipedia to provide evidence for a causal relationship between group size and incentives to contribute to Wikipedia; they find that the number of potential readers has a positive impact on contribution levels. Aaltonen and Seiler (2014) show that cumulative past contributions to Wikipedia articles lead to significantly more editing activity.[11] Greenstein and Zhu (2016b) investigate the ideological segregation among contributors to articles on US politics in the English Wikipedia. They find that contributors often offer content to those articles which have a different point of view and that they moderate their contributions over time.

Wikipedia's impact on financial markets has recently attracted some attention, too. Scholars argue that Wikipedia's information provision mitigates the information asymmetry (i) among investors and managers (Xu and Zhang 2013) and (ii) among individual investors, institutional investors and corporate insiders (Wu et al. 2014).

3 Data

We combine several data sources for our analysis. Kürschner (2015) lists all members of the 18th German Bundestag along with comprehensive information such as party affiliation, current offices, education, electoral districts, and political experience.[12] Dropping

[11] Note that Aaltonen and Seiler (2014) base their study on articles on the Roman Empire to exclude the possibility that current events have an impact on article length. In contrast to that, our analysis deals with biographies of *living* people.

[12] The 18th Bundestag consists of 631 MPs (regularly, there are only 598 seats). 299 MPs are directly elected in electoral districts following plurality vote (first vote); the second votes determine the composition of the Bundestag, i.e., the overall proportion of seats devoted to the respective parties. Seats that are not filled with directly elected candidates are given to candidates from ballot lists on state level. It is possible that a party wins more electoral districts via first votes than it wins seats according to the proportional representation determined by second votes. In that case, the parliament is amplified with so-called "overhang seats". There are currently four of them. In addition to that, there are currently 29 "compensatory mandates" in order to re-establish the proportional representation determined by the second votes.

chancellor Angela Merkel as an outlier, and dropping 12 MPs who have already left the 18th Bundestag[13] and were replaced, we obtain a sample of 630 observations, where 310 MPs are affiliated to the CDU/CSU, 193 to the SPD, 63 to the Left, and 64 to the Greens. Germany's two main parties CDU/CSU and SPD currently govern in a grand coalition. Considering the four parties that comprise the 18th Bundestag, they are located furthest in the center of the political spectrum.

To determine MPs' *international* political experience, we combine data from the Bundestag's homepage, bundestag.de, with data from Kürschner (2015). The Bundestag's homepage lists which MPs currently hold offices that feature an international background, i.e., which MPs participate in the Commission for Foreign Affairs, in the Commission for European Affairs, and which are head of an international parliamentary group (e.g., the "German-British parliamentary group"). Kürschner (2015) contains information on how often MPs have been member of the European Parliament.[14]

German MPs are obliged to disclose their ancillary income on the basis of different categories. "Abgeordnetenwatch" ("MP-watch") uses this information to estimate upper and lower bounds for MPs' outside earnings.[15] Following the common procedure in the literature, we use the corresponding mean values in our analysis.[16]

Electoral district specific data (e.g., demographics) stem from the "Bundeswahlleiter" ("electoral management body").[17] Germany is divided into 299 electoral districts of approximately equal population. Without Angela Merkel's district, there are 298 districts in our dataset. Several large cities such as Cologne and Berlin are divided into several electoral districts. The data on those cities are aggregated, such that we do not have 298 independent observations. We account for that by clustering the standard errors on the city level, and obtain 271 clusters.

Finally, we use Wikipedia itself. There are German language biographies for all MPs, English biographies for 170 MPs, and occasionally other foreign language biographies (French, Spanish, etc.). Each article links to a background site ("site information") containing meta-information, including its length (in bytes), the number of edits conducted so far, and - for the German Wikipedia - a full list of unique authors.[18] Figures 1 and 2 illustrate the average biography length per party. The numbers of words, adjectives and images are obtained via an automatic code that accesses the API of Wikipedia. The code failed for 13 MPs such that we obtain only 617 observations for the according variables. Information on the number of external links under party control, translation indicators, MPs' English homepages, criticism, and authors' activities is hand coded.

[13] See Appendix A for a robustness check where we include these observations.

[14] Note that according to 2002/772/EG (Article 6), it is not possible to be member of the Bundestag and the European Parliament at the same time.

[15] See www.abgeordnetenwatch.de. Viewed: November 2015.

[16] See, e.g., Becker et al. (2009).

[17] See http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/strukturdaten/ Accessed: June 2015.

[18] Data on length was extracted within a single day (October 12th, 2015) to exclude the possibility that a major political event could affect one part of our sample.

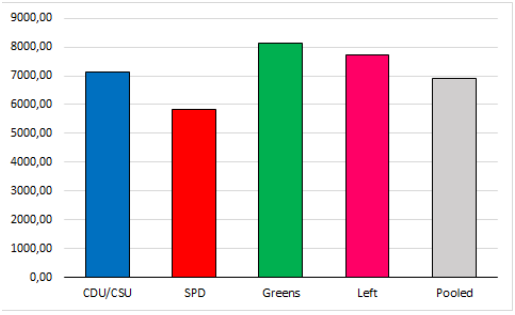

Figure 1: Average length of German Wikipedia biographies in bytes

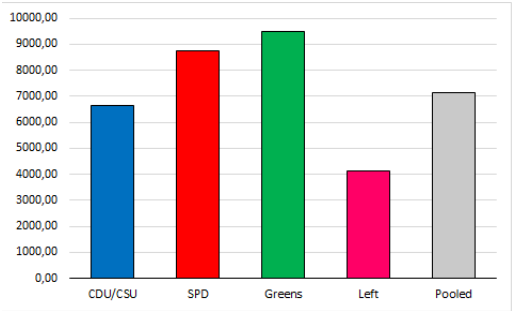

Figure 2: Average length of English Wikipedia biographies in bytes

Table 1 summarizes all variables employed in this paper.

4 Empirical strategy

The aim of this paper is to examine whether Wikipedia has a coverage bias in its biographies on German members of Parliament. We argue that there are two types of authors active on Wikipedia. On the one hand, there are neutral contributors who stick to the Wikipedia guidelines of maintaining a Neutral Point of View (NPOV, see also Section 2) and who allocate time and space to subjects according to how much can be said about them. On the other hand, there are partisan writers who want to extend and brighten the biographies of their respective MPs. We assume that "any coverage is good" for the politicians. To probe the validity of this assumption, in Appendix B we investigate negative coverage contained in the Wikipedia biographies.

Following the political science literature, we say that there is a coverage bias if otherwise equivalent members of two different parties are covered unequally. We presume that Wikipedians devote an *equal* amount of time and space to equivalent MPs. Thus, coverage bias on Wikipedia can arise solely through contributions of partisan writers.[19] Partisan writing, however, does not constitute a coverage bias per se. If the amount of partisan writing is equal for two parties, Wikipedia coverage is still balanced. Only if the amount of partisan writing differs, equivalent cases are covered to an unequal extent and coverage bias arises.

4.1 Basic regression

The starting point of our empirical investigation is to explain Wikipedia coverage in terms of biography length by

$$length_i^G = \beta_0^G + P_i\beta_1^G + X_i\beta_2^G + a_i^G + \varepsilon_i^G, \tag{1}$$

[19] The easiest way to visualize our notions of neutral and partisan writing is to assume that neutral authors report all objectively important facts about an MP, while partisan writers make contributions on top of that. Still, it is possible that some articles are largely generated by politically motivated authors. In this case, we classify everything that a neutral contributor would also have reported as "neutral coverage", and everything that goes beyond as "partisan writing".

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
German biography length (bytes)	6892.121	7301.929	1118	68623	630
English biography length (bytes)	7154.882	8201.174	838	55688	170
CDU/CSU	0.492	0.5	0	1	630
SPD	0.306	0.461	0	1	630
Greens	0.1	0.3	0	1	630
Left	0.102	0.302	0	1	630
Female	0.365	0.482	0	1	630
Sum of former periods in Bundestag	1.705	1.877	0	11	630
Sum of former periods in government	0.048	0.295	0	3	630
Current minister	0.021	0.142	0	1	630
Party head	0.013	0.112	0	1	630
Average outside earnings (Euros)	26342.357	104955.943	0	1411000	628
Doctoral degree	0.184	0.388	0	1	630
Directly elected	0.463	0.499	0	1	630
Population density	886.257	1431.155	38	12842.9	292
Fraction population 18-25 years	20.077	2.67	15.6	28.9	292
Fraction population with Abitur	36.617	7.505	21.7	59.7	292
Sum of international offices	0.206	0.493	0	3	630
Sum of former periods in EP	0.013	0.138	0	2	630
Number of adjectivesj	78.198	90.193	13	758	617
Number of images	2.783	1.924	0	19	617
Weblinks under party control	1.487	0.662	0	4	630
Unique authors	66.477	96.541	5	802	630
Number of words (German)	674.245	702.592	102	6421	617
Sum of former periods in Landtag	0.352	1.035	0	7	630
Academic studies	0.798	0.402	0	1	630
English homepage	0.025	0.157	0	1	630
Translation template	0.021	0.142	0	1	630
Number of critical sentences	0.498	1.857	0	20	630

where P_i is a vector of party dummies with SPD as the omitted category, and X_i is a vector of control variables including observable characteristics of MP i (e.g., political experience, current offices) and electoral district specific variables (e.g., demographics).[20] The superscript G clarifies that we consider biographies from the German Wikipedia version. We denote by a_i^G all unobserved characteristics of MP i, including, e.g., their ability, their good looks, or their wittiness. Conditional on a_i^G, we assume that ε_i^G is white noise and hence orthogonal to all covariates.

The parameter of interest is β_1^G. Conditional on X_i and a_i^G, β_1^G measures the pure impact of party affiliation on biography length relative to the omitted category SPD. Thus, β_1^G corresponds to the coverage bias against MPs from the SPD relative to each party in P_i. In contrast to that, β_1^G does not measure the *absolute* amount of partisan writing. If no party was active on Wikipedia, party affiliation would not affect biography length once we controlled for X_i and a_i^G. If all parties were equally active on Wikipedia, total coverage would still be balanced, and there would be no coverage bias, either. Thus, β_1^G can only indicate *differences* in partisan writing.

As noted by Groeling (2013), any approach to measure media bias by controlling for observable differences between politicians is only as good as the controls involved. If party affiliation was random conditional on observable differences captured in X_i, we could estimate

$$length_i^G = \beta_0^G + P_i\beta_1^G + X_i\beta_2^G + u_i^G, \tag{2}$$

with $u_i^G = a_i^G + \varepsilon_i^G$, by OLS to obtain an unbiased estimate of β_1^G. It is, however, possible that P_i is correlated with unobserved factors in a_i^G (e.g., if more able politicians self-select into a particular party). Given that we cannot condition on a_i^G, an OLS estimation of equation (2) might suffer from an omitted variable bias. The direction of the bias is unclear, since it depends on whether, e.g., high ability politicians tend to self-select into the SPD or into other parties. In other words, we cannot rule out the possibility that differences in coverage may be driven by unobserved heterogeneity between MPs, leading even neutral authors to allocate more or less space to MPs affiliated to a particular party. Therefore, we cannot infer the existence of a coverage bias from an OLS regression of equation (2).[21] Given these concerns, we proceed with a difference-in-differences analysis to recover an unbiased estimate of β_1^G.

4.2 Difference-in-differences estimation

The English Wikipedia version provides biographies on 170 German MPs. We exploit variation in biography length between the German and the English Wikipedia for a difference-in-differences estimation. The key assumption of our analysis is that partisan writers have no incentive to spend effort on English language biographies, because they are unlikely to be read by German voters.

[20] Using the SPD as omitted category is convenient, because we will focus on the comparison of Germany's two main parties SPD and CDU/CSU in the remainder of the paper.

[21] In addition to that, some of our regressors are "bad controls" in the spirit of Angrist and Pischke (2009). For instance, political experience could be an outcome itself.

4.2.1 Identifying assumptions

Assuming that the length of an English Wikipedia biography is determined analogously to the length of a German biography, i.e.,

$$length_i^E = \beta_0^E + P_i\beta_1^E + X_i\beta_2^E + a_i^E + \varepsilon_i^E, \tag{3}$$

our key premise would correspond to

$$\beta_1^E = 0, \qquad \textbf{(Identifying Assumption I)}$$

given X_i and a_i^E. As argued in the previous section, P_i might be correlated with a_i^E, in which case an OLS regression of

$$length_i^E = \beta_0^E + P_i\beta_1^E + X_i\beta_2^E + u_i^E, \tag{4}$$

with $u_i^E = a_i^E + \varepsilon_i^E$, would yield biased estimates. Thus, we cannot directly test our identifying assumption I. We discuss its plausibility below.

Taking the difference between the German and the English biography length, given in equations (1) and (3), yields

$$length_i^G - length_i^E = \beta_0 + P_i\beta_1 + a_i^G - a_i^E + X_i\beta_2 + \varepsilon_i, \tag{5}$$

with $\beta_k = \beta_k^G - \beta_k^E$ and $\varepsilon_i = \varepsilon_i^G - \varepsilon_i^E$. Our second identifying assumption is that, conditional on the covariates in X_i, the unobserved heterogeneity captured in a_i^G and a_i^E affects the English and the German length equally, i.e.,

$$a_i^G = a_i^E = a_i. \qquad \textbf{(Identifying Assumption II)}$$

Then, the unobserved heterogeneity differences out in (5), and we can estimate

$$length_i^G - length_i^E = \beta_0 + P_i\beta_1 + X_i\beta_2 + \varepsilon_i, \tag{6}$$

by OLS without being concerned about a potential omitted variable bias.[22]

The empirical strategy allows us to directly recover β_1^G from equation (6). To see that, note that $\beta_1^G = \beta_1 + \beta_1^E$. Assuming $\beta_1^E = 0$ implies that

$$\beta_1 = \beta_1^G.$$

The intuition for this result is as follows. Conditional on X_i, β_1 measures the impact of party affiliation on the *difference* in biography length between the German and the English Wikipedia, relative to the omitted category SPD. By assumption, there is no partisan writing on English biographies and the *difference* in biography length can exclusively be affected via the German Wikipedia biographies. Thus, although the dependent variable in equation (6) is $\left(length_i^G - length_i^E\right)$, β_1 measures differences in partisan writing on the *German* biographies, and is therefore equivalent to β_1^G, the parameter of interest.

[22] Identifying Assumption II is actually a bit stronger than what we need, which is that $a_i^G - a_i^E$ is orthogonal to $(X_i, P_i, \varepsilon_i)$.

4.2.2 Discussion

To discuss the identifying assumptions behind our empirical strategy, we now compare it with standard difference-in-differences procedures. Our approach is unconventional in so far as there is no time dimension, and we do not examine the effects of a policy change, either. The difference between the English and the German language version of Wikipedia, however, is similar to the conventional 'before versus after treatment'. Moreover, the party membership can be thought of as constituting one particular treatment (there are thus no untreated observations).

Identifying Assumption I There are two underlying assumptions in the identification of our key parameter of interest β_1^G. First, there is no partisan writing on the English biographies ($\beta_1^E = 0$). This assumption allows us to interpret the length of the English biographies as unaffected by the treatments, similar to the pre-treatment observations in conventional difference-in-differences designs. As pointed out above, we cannot directly test this assumption. The absence of partisan writing on the English Wikipedia is plausible, however, given that only few of the MPs in our data even have an English homepage. In our sample, only 16 out of 630 MPs provide significant parts of their personal homepage in English, and only 8 out of these 16 have an English Wikipedia biography.[23] Before bothering to manipulate an English Wikipedia entry, surely the MPs would start setting up their own web presence in English first. The robustness checks in Appendix A show that our results are not driven by these observations.

Our identification strategy would also be undermined if the English language biographies were not edited independently, but mere translations of their German counterparts. In this case, partisan writing conducted on the German biographies would be transferred to the English ones, and we could no longer credibly assume that $\beta_1^E = 0$. Wikipedia, however, advises *against* one-to-one translations. Instead, contributors are encouraged to critically examine the original articles and to deviate from their content if it seems appropriate. The Wikipedia guidelines state that having no translation is preferred over machine translations (Wikipedia:Translation). Moreover, Wikipedia's translation guidelines require that "the new translated article must credit the source article". In practice, articles that are to some extent translated are marked by the template "Translated page" on the article's talk page and by an interlanguage link to the source article. In our sample, only 13 out of 170 English biographies are indicated as partly translated. Thus, lack of independent editing does not seem to be a major concern. The robustness checks in Appendix A confirm that our results are not driven by these observations.

Relatedly, if it was a common procedure to translate the Wikipedia biographies of German MPs from English *into* German, partisan authors could have an incentive to extend English biographies to provide the basis for a lengthy translation. In our sample, however, no German biography is translated from a foreign language into German.

[23] By "significant parts" we mean everything beyond a short CV in English.

Irrespective of direct translations, it is easier to write a long English biography if the German counterpart is long, because long articles tend to contain more references. A higher number of references facilitates the access to information on MPs for Wikipedia contributors. In our data, Wikipedia biographies of MPs from the SPD are shortest on average. Thus, it might be harder to write English biographies of SPD MPs compared to, e.g., MPs from the CDU/CSU. We do not take this effect into account in our analysis, since taking it into consideration would only strengthen our result of a coverage bias against the SPD. Our findings might therefore underestimate the coverage bias.

A related argument holds for MPs who provide their personal homepage in English. Similar to a high number of references, detailed English homepages might facilitate the writing of long English Wikipedia biographies. As mentioned above, only few MPs have an English webpage. We thus believe that the provision of English homepages is at most a minor concern.

We also take into account that English texts are generally shorter than German texts. According to anecdotal evidence by translators, English texts are about a fourth to a fifth shorter, but there is no consensus on a general factor.[24] As a robustness check, we scale the length of the German biographies in equation (6) with a broad spectrum of factors, ranging from 0.6 to 0.9, and repeat the difference-in-differences estimation. In addition to that, we use the difference in logs, i.e., $\ln\left(length_i^G\right) - \ln\left(length_i^E\right) = \ln\left(length_i^G / length_i^E\right)$ as dependent variable. This corresponds to studying the ratio of German to English biography length, such that scaling issues do not matter. We find that our results are qualitatively unaffected. The results of these robustness checks are discussed in Appendix A.

Finally, it is not likely that the assumption of no partisan writing (identifying assumption I) is valid for every foreign language. For instance, we cannot use languages that are spoken in countries adjacent to Germany. Voters who live, e.g., in electoral districts at the French border are more likely to stem from a French-speaking household than voters who live elsewhere in Germany. Hence, politicians who run in such districts may have an incentive to manipulate their French Wikipedia biographies. The same argument applies to languages spoken by large minorities. E.g., politicians who run in electoral districts with a large fraction of Turkish-speaking households may have an incentive to manipulate their Turkish Wikipedia biographies. English Wikipedia biographies, on the other hand, do not raise such concerns, since Germany has not direct border with any English speaking country, and the number of immigrants whose native language is English is fairly low (Statistisches Bundesamt).

Identifying Assumption II Our second identifying assumption that, conditional on observables, unobserved heterogeneity affects the English and German biographies similarly, can be understood as a parallel trends assumption. If unobserved heterogeneity would affect the different language versions differently, and would also be correlated with party membership, then we might find a significant estimate of β_1^G even in the absence of partisan writing on the German Wikipedia. In conventional difference-in-

[24] See, e.g., orbis-uebersetzungen.de. Accessed: February 2016.

differences estimations, state of the art practice recommends to check the validity of the parallel trends assumption by investigating pre-treatment patters. While we do not have an explicit time dimension here, in principle it would be possible to conduct a similar exercise if we had more than one foreign language, with a sufficiently high number of Wikipedia biographies of the MPs, and plausibly unaffected by partisan writing. As discussed above, however, English seems to be the only language that meets all these requirements. For Spanish, Italian, Portuguese, Swedish, Finnish, and Norwegian, there are far too little observations per language - usually around four or five - to draw any reliable conclusions.

We can show, however, that a_i^G and a_i^E are positively correlated. The variances of the residuals u_i^G and u_i^E in equations (2) and (4) are given by

$$V\left(u_i^l\right) = V\left(a_i^l + \varepsilon_i^l\right) = V\left(a_i^l\right) + V\left(\varepsilon_i^l\right), \quad l = G, E$$

since by assumption ε_i^l is white noise. Let V^{sum} denote the variance of the sum of these residuals:

$$V^{sum} := V\left(a_i^G\right) + V\left(\varepsilon_i^G\right) + V\left(a_i^E\right) + V\left(\varepsilon_i^E\right).$$

The variance of the residual in equation (5) is given by

$$V^{diff} := V\left(a_i^G + \varepsilon_i^G - \left(a_i^E + \varepsilon_i^G\right)\right) = V\left(a_i^G\right) + V\left(a_i^E\right) + V\left(\varepsilon_i^G\right) + V\left(\varepsilon_i^E\right) - 2cov\left(a_i^G, a_i^E\right).$$

If $a_i^G = a_i^E$, then the ratio

$$\frac{V^{diff}}{V^{sum}} = \frac{V\left(\varepsilon_i^G\right) + V\left(\varepsilon_i^E\right)}{V\left(a_i^G\right) + V\left(\varepsilon_i^G\right) + V\left(a_i^E\right) + V\left(\varepsilon_i^E\right)}$$

is smaller than one.[25] This empiricially observable implication of our identifying assumption II is indeed supported by our data (see Appendix C). Although this exercise does, of course, not prove of identifying assumption II, it shows that the unobserved heterogeneity impacts the length of the English versus German Wikipedia in the same direction.

4.3 Sample selection

Unfortunately, there are only 170 observations on English language Wikipedia biographies in our data.[26] Naively regressing equation (6) on the subsample of MPs where we observe English biography length could induce sample selection bias. Thus, we consider a selection model consisting of equation (6) and the selection equation

$$d_i = 1[Z_i\delta + u_i > 0], \tag{7}$$

[25] In case of no unobserved heterogeneity at all, $a_i^G = a_i^E = 0$ and $V^{sum} = V^{diff}$.

[26] Note that while German MPs are automatically relevant for the German Wikipedia, they have to meet slightly different criteria for being of relevance in the English Wikpedia version. Although the English notability criteria do not make an explicit statement here, German members of parliament are likely to meet them. (Wikipedia:Notability)

where $Z_i = (P_i, X_i, I_i)$, where the vector I_i contains variables on MPs' international political relevance, and where d_i is a binary dependent variable equal to 1 if there exists an English biography for MP i, and 0 otherwise. The binary outcome d_i is always observed; the difference in biography length between German and English Wikipedia, $\left(length_i^G - length_i^E\right)$, is observed if $d_i = 1$, and missing otherwise.

The identification of the selection model requires the existence of exclusion restrictions. We argue that the covariates in I_i are excluded from equation (6), because their impact on biography length is negligible. Past and current "international" offices - having been a member of the European Parliament or the current number of offices exhibiting an international background - are not very salient; beyond the statement that a particular MP holds such an office, only little can be reported in Wikipedia biographies. Consequently, there is usually less than one sentence per biography devoted to such information. On the other hand, it is likely that the covariates in I_i strongly affect the *likeliness* that Wikipedians spend the effort to set up a foreign language biography, because they determine how relevant MP i is from an international perspective. Hence, I_i is included in selection equation (7) and fulfills the requirements on exclusion restrictions.[27]

4.3.1 Two Step estimation

There are several ways to account for nonrandom selection. The Heckman Two Step ("Heckit") procedure (Heckman 1976, 1979) is popular for its easy implementation, and the results are straightforward to interpret. The Heckit model corrects for sample selection bias using

$$E\left(length_i^G - length_i^E | Z_i, d_i = 1\right) = \beta_0 + P_i\beta_1 + X_i\beta_2 + \gamma\lambda_i\left(Z_i\delta\right), \tag{8}$$

where $\lambda\left(Z_i\delta\right) \equiv \phi\left(Z_i\delta\right)/\Phi\left(Z_i\delta\right)$ denotes the inverse Mills ratio (see, e.g., Wooldridge 2010, Ch. 17, for technical details). The Two Step estimation is, however, susceptible to collinearity problems. In the absence of appropriate exclusion restrictions, the coefficients are only identified through the non-linearity of the inverse Mills ratio $\lambda\left(\cdot\right)$, which can lead to severe multicollinearity of $\lambda\left(\cdot\right)$ and the covariates. Although this does not affect the point estimates, standard errors would inflate.

4.3.2 Maximum likelihood estimation

An alternative way to estimate the selection model is maximum likelihood. The procedure is more efficient than the Two Step estimation, but less common since the computation is burdensome and has required much time until recently. Today, the estimation procedure is implemented in most statistical software. The full log-likelihood function can be found in various econometrics textbooks (e.g., Wooldridge 2010).

[27] We do not believe that the probability of having an English Wikipedia biography is affected by individual MP preferences. As noted above, only few MPs provide English homepages. Moreover, only half of the MPs who have English homepages also have English Wikipedia biographies.

5 Results

5.1 Basic regressions

The results of an OLS regression of equation (2) are presented in Table 2.[28] Model 1 explains the length of a Wikipedia biography in terms of party affiliation only and thus replicates the party averages in biography length relative to the omitted category SPD. To provide a vivid image, the average difference in length between MPs of CDU/CSU and SPD roughly corresponds to 1/2 DIN-A4 page if the biographies were printed out.[29] The average difference between Greens and SPD amounts to slightly less than one page, and the average difference between Left and SPD lies in between. To put these numbers into perspective, note that the average length across all parties is around 2 1/2, and the median length around 2 pages.

A plausible explanation for this finding is MP heterogeneity. E.g., MPs from the SPD might be less politically experienced and have fewer doctoral degrees or side jobs than MPs from other parties, which would reduce the amount of noteworthy information on them. To take this into account, we control for gender (Model 2), for political experience (Model 3), and for further individual characteristics (Model 4). Subsequently adding these sets of control variables hardly affects the estimates for the party coefficients, though, so differences in observable MP characteristics do not seem to be responsible for unequal Wikipedia coverage.[30]

Note that the coefficients on the party dummies are, roughly speaking, an order of magnitude smaller than the coefficients on political standing such as being a party head or a current of former member of the cabinet (minister). Moreover, party membership alone explains only a minor part of the variation in our data, whereas addding variables for political standing and experience increases the R^2 notably. These results are in line with the literature on media coverage of politicians, where political standing has often been found to have the largest effects (see Vos 2014 for a survey).

Differences in the biographies' noteworthiness could also drive our findings.[31] In other words, MPs affiliated to a particular party might attract a larger audience, e.g., because their voters are more Internet affine. According to Forschungsgruppe Wahlen (2014), 75% of CDU/CSU and SPD voters, 80% of Left voters and even 93% of the Greens' voters use the Internet as a source for political information, which might partly explain why the Wikipedia biographies of MPs from the latter parties are longest on average. To take reader characteristics into account, Model 5 controls for the de-

[28] All regression results in this section are based on a sample excluding the 12 observations on early resigns. Regressions *including* these observations are conducted as robustness checks; the results are presented in the Appendix.

[29] This measure corresponds to the document generated if you click on "Printable Version".

[30] Careers of MPs from the SPD *prior* to becoming politicians might be less noteworthy than the prior careers of MPs affiliated to other parties. We do not control for this, but can exclude the possibility that there are past "superstars" in the Bundestag: there are some former actors, one singer, and one wrestler, but none of them is particularly famous, and they are not affiliated to the same party.

[31] Zhang and Zhu (2011) find a causal positive impact of group size on Wikipedia contributions.

mographics of MPs' electoral districts.[32] Plausibly, population density has a highly significant effect on biography length. Broadband connections in urban areas are usually better than in rural areas, which facilitates the use of Wikipedia and increases the biographies' likelihood to be read. In contrast to that, age and education do not seem to make any difference. Relative to the previous specifications, the CDU/CSU estimate increases (it corresponds to slightly more than 1/2 DIN-A4 page); it is statistically significant at the 5%-level.

Interestingly, the estimate for *Female* is negative and significant at the 5%-Level in Model 2; the effect size corresponds to 1/2 DIN-A4 page and might be related to a gender bias as discussed in the literature. We point out that the effect is no longer significant once we control for further characteristics, and even changes sign in Model 5. We thus do not find evidence for a gender bias in our data.[33]

As has been argued, detection of coverage bias requires the comparison of equivalent cases. The four parties that currently constitute the German Bundestag differ, however, in several respects. First, the parties differ in their political positions: CDU/CSU is center-right, SPD is center-left, Greens and Left are located even further left. Second, CDU/CSU and SPD are Germany's main parties and larger than Greens and Left in terms of members and voters. Third, CDU/CSU and SPD currently govern in a grand coalition while Greens and Left constitute the opposition. Thus, CDU/CSU and SPD on the one hand, and Greens and Left on the other, are quite similar. Note that this is a particular strength of our data: If we compare Wikipedia coverage of MPs affiliated to Germany's main parties, any differences cannot be the result of Wikipedians' different treatment of government and opposition parties. Given that CDU/CSU and SPD are currently in power, they are of particular interest, and we will focus on the coverage gap between CDU/CSU and SPD in the remainder of the analysis.

Similarly, MPs in our dataset are not very homogenous with respect to their individual characteristics. We separate obvious outliers - party heads, current and former ministers - from the remaining "ordinary" MPs to obtain two subsamples that are respectively more homogenous. For the sake of clarity, we call the former subsample the "Stars Subsample" (32 observations), and the latter the "Ordinary Subsample" (598 observations). A further advantage is the following: The previous government (CDU/CSU and FDP, 2009 - 2013) did *not* include the SPD, raising the concern that differences in the *history* of government involvement might drive our results. Regressions that exclude former ministers rule out this possibility.[34] An additional benefit

[32] As only 299 German MPs are directly elected, we lose about one half of our observations. On top of that, we drop the directly elected MPs of Left and Greens, because there are too few observations (4 Left, 1 Green).

[33] This results is similar to a common finding in the literature on media coverage of politicians is that females are covered less, but the effect disappears when controls for personal characteristics are added (see Voss 2014).

[34] The literature on news coverage of politicians has found a clear incumbency bonus. Members of the cabinet get more coverage. The bonus does not seem to extend, however, to all members of of the majority party. Vos (2014, p. 2449) summarizes the literature: "Politicians of majority parties do not receive additional coverage."

Table 2: Main regression on Full Sample

	Model 1	Model 2	Model 3	Model 4	Model 5
CDU/CSU	1297.3**	1088.2*	977.9*	801.4	1701.7**
	(626.1)	(653.3)	(511.0)	(527.0)	(813.7)
Greens	2297.8*	2447.0**	2064.9**	2066.0**	-
	(1240.7)	(1235.9)	(986.2)	(992.8)	
Left	1913.6*	2110.7**	2591.8***	2564.4***	-
	(1005.5)	(1007.0)	(805.6)	(751.1)	
Female		-1244.0**	-980.4**	-651.2	154.3
		(601.2)	(483.9)	(466.5)	(789.5)
Sum former periods in the Bundestag			700.0***	692.9***	472.1**
			(169.2)	(170.7)	(195.1)
Sum former periods in the government			9103.0***	8869.0***	8834.0***
			(2136.9)	(2100.0)	(1651.6)
Current minister			11410.3***	10992.5***	14113.2***
			(3773.9)	(3645.3)	(4447.3)
Party head			13653.0***	13614.7***	7190.5
			(4460.6)	(4342.7)	(4736.6)
Average outside earnings				0.00387	0.00176
				(0.00280)	(0.00166)
Doctoral degree				2213.8***	1887.6*
				(690.0)	(995.8)
Population density					0.845***
					(0.295)
Fraction population 18 - 25 years					205.4
					(173.6)
Fraction population with Abitur					-15.67
					(52.24)
Constant	5829.6***	6351.7***	4263.4***	3712.0***	-798.4
	(466.3)	(595.9)	(564.1)	(586.6)	(3449.5)
N	630	630	630	628	286
R^2	0.011	0.018	0.406	0.430	0.528

Robust standard errors in parentheses. Standard errors are clustered at the city-level in Model 5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

of focussing on the Ordinary Subsample is that, for these backbenchers, our assumption that any coverage is good is particularly convincing. As we show in Appendix B, their biographies contain almost no negative coverage. Futhermore, for less well known politicians, even negative coverage may increase name recognition and thereby increase election chances (Burden 2002).

Table 3: Main regression Stars Subsample

	Model 1
CDU/CSU	1887.4
	(7739.7)
Greens	2620.7
	(7944.0)
Left	2902.2
	(15283.6)
Constant	23868.3***
	(6671.2)
N	32
R^2	0.004

Robust standard errors in parantheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Due to its small size, we only perform a regression according to Model 1 on the Stars Subsample; the results are displayed in Table 3. The results for the Ordinary Subsample are presented in Table 4. In terms of magnitude, the estimates for the Ordinary Subsample are similar to the previous regression results, but statistically more significant. Interestingly, we do not find evidence for a gender bias anymore. Though still negative in Models 2 to 4, the estimate for *Female* is statistically insignificant throughout all specifications, which implies that the gender effect is driven by outliers. In other words, when we consider only comparable members of the Bundestag, women are not systematically covered less.

In sum, the OLS regression demonstrates that there is a persistent coverage gap between MPs from the SPD relative to MPs from the remaining parties, in particular to MPs from the CDU/CSU. It remains to establish whether this effect stems from differences in partisan writing.

Blinder-Oaxaca decomposition We have shown that observable MP and voter characteristics cannot fully explain differences in Wikipedia coverage. Another way to illustrate that is to perform a Blinder-Oaxaca decomposition (Blinder 1973, Oaxaca 1973). The Blinder-Oaxaca decomposition divides an outcome differential between two groups into (i) a part that is explained by group differences in the regressors and (ii) a residual part that cannot be accounted for by such differences.[35] The outcome differential considered in our application is the average coverage gap between two parties. Thus, the decomposition determines the proportion that can be explained by the regressors used in this section. The residual part may include unobserved heterogeneity in MP and voter characteristics, but also differences in partisan activities. A brief illustration of the technique is provided in Appendix D.

We conduct a Blinder-Oaxaca decomposition for MPs from the parties of main interest, i.e., CDU/CSU and SPD. The decomposition results are given in Table 5.

[35] See Fortin et al. (2011) for a recent discussion and for technical details.

Table 4: Main regression on Ordinary Subsample

	Model 1	Model 2	Model 3	Model 4	Model 5
CDU/CSU	1165.8***	1057.0***	971.3***	870.4**	1614.8**
	(374.1)	(385.9)	(366.6)	(353.1)	(738.7)
Greens	1597.2	1692.7*	1736.4*	1733.5*	-
	(972.0)	(985.2)	(906.8)	(917.7)	
Left	2182.2***	2297.7***	2388.3***	2396.4***	-
	(728.7)	(730.6)	(694.6)	(656.1)	
Female		-689.1	-474.8	-216.0	299.0
		(435.6)	(403.9)	(402.0)	(749.8)
Sum former periods in the Bundestag			869.6***	875.5***	723.1***
			(161.5)	(160.2)	(197.6)
Average outside earnings				0.00200	0.00168
				(0.00170)	(0.00162)
Doctoral degree				1915.3***	886.2
				(608.9)	(868.8)
Population density					0.896***
					(0.278)
Fraction population 18 - 25 years					255.2
					(155.5)
Fraction population with Abitur					-1.997
					(48.21)
Constant	4947.3***	5231.9***	3810.6***	3327.6***	-2578.2
	(232.7)	(305.5)	(347.0)	(373.3)	(3436.9)
N	598	598	598	596	266
R^2	0.021	0.025	0.121	0.150	0.185

Robust standard errors in parentheses. Standard errors are clustered at the city-level in Model 5.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Depending on the weight chosen for the "neutral" coefficient vector $\hat{\beta}^*$, around a third of the coverage gap between MPs from the SPD relative to MPs from the CDU/CSU can be explained by observable factors. The residual part of the decomposition, i.e., unobserved heterogeneity and potential partisan activities, constitutes the lion's share of the coverage gap.

Table 5: Blinder-Oaxace decomposition CDU/CSU v. SPD

	(1) SPD benchmark	(2) CDU/CSU benchmark	(3) Pooled
Differential			
Prediction CDU	6037.8***	6037.8***	6037.8***
	(285.2)	(285.2)	(283.4)
Prediction SPD	4908.6***	4908.6***	4908.6***
	(232.8)	(232.8)	(230.2)
Difference	1129.2***	1129.2***	1129.2***
	(368.2)	(368.2)	(365.1)
Decomposition			
Explained Part	371.7	162.0	211.1
	(229.5)	(183.9)	(159.9)
Unexplained Part	757.4*	967.2***	918.1***
	(399.4)	(365.3)	(350.7)
N	476	476	476

Standard errors in parentheses. See Appendix for technichal details.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.2 Difference-in-differences estimation

As laid out in Section 4, we perform a difference-in-differences estimation to ensure that our findings are not driven by unobserved heterogeneity. Some regressors in the Stars Subsample (e.g., being a current minister or party head) are perfect predictors of selection, so we restrict the analysis to the Ordinary Subsample. We do not restrict the analysis on directly elected MPs, as was done above in Model 5, to avoid having even less observations on English language biographies.

Table 6 presents the results of an OLS regression of equation (4). We find that the CDU/CSU estimate is very small and statistically insignificant throughout all specifications, which matches Identifying Assumption I. This result must be taken with caution, though, since we cannot rule out that the regression suffers from an omitted variable bias.

The estimates of the difference-in-differences regression given by the selection model (6) and (7) are presented in Table 7; the corresponding results for the selection equation

Table 6: Main regression of English Length on Ordinary Subsample

	Model 1	Model 2	Model 3	Model 4	Model 5
CDU/CSU	-119.2	-141.8	-128.0	-381.2	-432.2
	(952.9)	(947.4)	(955.0)	(1000.6)	(1368.8)
Greens	2869.0**	2554.0**	2524.9**	2453.8*	-
	(1323.4)	(1284.3)	(1236.6)	(1277.2)	
Left	-1542.0	-2061.2*	-1951.1*	-2012.5*	-
	(1019.3)	(1072.6)	(1116.5)	(1156.7)	
Female		1268.2	1335.0	1424.3	2213.6
		(926.5)	(890.9)	(911.9)	(1576.7)
Sum former periods in the Bundestag			479.6**	491.3**	431.5
			(226.8)	(230.9)	(422.1)
Average outside earnings				0.00317	-0.00266
				(0.00240)	(0.00586)
Doctoral degree				499.9	-751.5
				(587.2)	(1265.6)
Population density					1.138
					(0.686)
Fraction population 18 - 25 years					-10.25
					(238.0)
Fraction population with Abitur					4.356
					(69.83)
Constant	4833.2***	4551.4***	3364.3***	3235.2**	3065.4
	(826.1)	(863.4)	(1132.5)	(1250.0)	(6858.1)
N	138	138	138	136	71
R^2	0.080	0.098	0.145	0.156	0.166

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

are given in Table 8. Column 1 contains the results of a potentially biased OLS estimation. To put the magnitudes into perspective, the estimate roughly corresponds to one DIN-A4 page for the CDU/CSU, to slightly more for the Greens, and to around 2 1/3 DIN-A4 pages for the Left. The estimate for the CDU/CSU is not statistically significant, so we do not find any evidence for a coverage bias favoring them relative to the SPD here. Large standard errors could, however, also be an artifact of the strongly decreased sample size.

The models in Columns 2 and 3 account for potential sample selection. Column 2 presents the results of the Two Step estimation. To enable a direct comparison of ML and OLS estimates, we display all estimates' partial average effects (PAEs) in Table 9. We find that the PAEs of the Two Step estimation are very similar to the OLS estimates. In contrast to the previous specification, however, the CDU/CSU estimate is significant at the 10%-Level and thereby provides weak evidence for a coverage bias against the SPD. Interestingly, the inverse Mills ratio - though very large - is not statistically significant. Thus, we cannot reject the null hypothesis of no sample selection here.[36] A regression of $\lambda(\cdot)$ on P_i and X_i, however, indicates that our exclusion restrictions are not very powerful, since it reveals multicollinearity between the inverse Mills ratio and the covariates ($R^2 = 0.85$). As argued in Section 4.3.1, multicollinearity entails inflated standard errors. Hence, we cannot take a statistically insignificant estimate for the inverse Mills ratio as a reliable indicator for the absence of sample selection.[37]

Column 3 presents the results of a maximum likelihood estimation of the selection model. This is our preferred specification for several reasons. First, the ML estimation is generally more efficient than the Two Step procedure. Second, it is less susceptible to multicollinearity problems regarding the inverse Mills ratio. Third, we find that the estimate for the inverse Mills ratio is highly statistically significant here. Given that we take the ML specification to be more reliable than the Two Step procedure, we conclude that sample selection is nonrandom, which furthermore implies that the OLS estimates in Column 1 are biased.

The PAEs of the ML estimates are slightly larger than the PAEs of the previous specifications, but altogether very similar. The CDU/CSU estimate corresponds to about one and a third DIN-A4 page, the Greens estimate to slightly more than one and a half pages, and the Left estimate to about two and a half pages. Importantly, the CDU/CSU estimates is significant at the 1%-level. Thus, our preferred specification of the selection model given by equations (6) and (7) provides reliable evidence for a difference in partisan writing between CDU/CSU and SPD that amounts to 1 1/3 DIN-A4 pages on average. In other words, having ruled out that unobserved heterogeneity might affect our results, the difference-in-differences analysis provides evidence for a coverage bias on the German Wikipedia biographies against MPs from the SPD relative to MPs from the CDU/CSU.

[36] The OLS estimates would be unbiased, then.

[37] If the range of predicted probabilities from the selection equation was very narrow, the inverse Mills ratio would be very similar to a constant term, which could also induce multicollinearity. This is, however, not the case; predicted probabilities range from 0.054 to 0.965, with a mean of 0.230.

Table 7: Difference-in-differences estimation

	(1) OLS	(2) Two Step	(3) ML
CDU/CSU	2230.8	3165.4*	6010.6***
	(1378.3)	(1749.1)	(1720.4)
Greens	2901.8	4365.4*	7941.9***
	(2805.8)	(2510.7)	(2432.1)
Left	6447.0***	7892.4***	11842.6***
	(1627.8)	(2489.3)	(2365.9)
Female	-1237.8	-1409.9	-1443.3
	(1265.4)	(1284.4)	(1419.8)
Sum former periods in Bundestag	670.0**	1030.4**	1902.7***
	(332.0)	(472.2)	(370.7)
Doctoral degree	2509.2*	3180.6**	3845.3**
	(1477.9)	(1518.9)	(1599.4)
Average outside earnings	0.0102	0.00984	0.00140
	(0.00985)	(0.00804)	(0.00747)
Constant	-245.1	-5430.1	-18997.1***
	(1607.3)	(5347.4)	(2375.3)
Mills Lambda		2837.6	10558.6 ***
		(2792.9)	(906.1)
N	136	596	596
R^2	0.124		

Standard errors in parentheses. In Models 2 and 3, there are 136 uncensored and 460 censored observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Selection equation

	Two Step	ML
CDU/CSU	0.403***	0.370**
	(0.148)	(0.146)
Greens	0.559**	0.578***
	(0.221)	(0.212)
Left	0.670***	0.637***
	(0.212)	(0.206)
Female	-0.102	-0.131
	(0.133)	(0.128)
Sum former periods in Bundestag	0.205***	0.208***
	(0.0336)	(0.0319)
Doctoral degree degree	0.297*	0.371**
	(0.154)	(0.145)
Average outside earnings	0.000	0.000
	(0.000)	(0.000)
Sum former periods in European Parliament	1.355**	0.345
	(0.585)	(0.414)
Sum international offices	0.178	0.214***
	(0.112)	(0.0533)
Constant	-1.518***	-1.524***
	(0.156)	(0.152)
N	596	596

Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Partial Average Effects (PAE)

Variable	Two Step	ML
CDU/CSU	2275.546	2962.996
Greens	3183.494	3384.95
Left	6488.188	6843.826
Female	-1184.665	-356.773

PAEs for a discrete change of the dummy variables from 0 to 1.

6 Patterns of partisan activity

The results in Section 5.2 tend to support the existence of differences in partisan writing leading to a coverage bias against MPs from the SPD relative to the CDU/CSU. In this section, we present additional evidence on partisan activity on the German Wikipedia.

6.1 Adjectives

The literature on sentiment analyses states that a high proportion of adjectives often indicates that a text is particularly brightened (see Pang and Lee 2008 for a survey). Figure 3 shows that Wikipedia biographies of MPs from the SPD contain on average less adjectives than biographies of MPs from other parties; in particular, they contain about 20 adjectives less than biographies of MPs from the CDU/CSU. A regression using the number of adjectives shows that this difference is significant at the 5%-level; the results are displayed in Column 1 of Table 10. Biographies of MPs from the SPD are, however, shortest on average, so they might trivially contain less adjectives.[38] To take this effect into account, we normalize the number of adjectives per biography with its length. Figure 4 displays the average adjective/word ratio per party. We find that the SPD biographies also contain *relatively* less adjectives than CDU/CSU biographies, although the differences across parties become smaller. This relative difference is also significant at the 5%-level, as shown in Column 2 of Table 10.

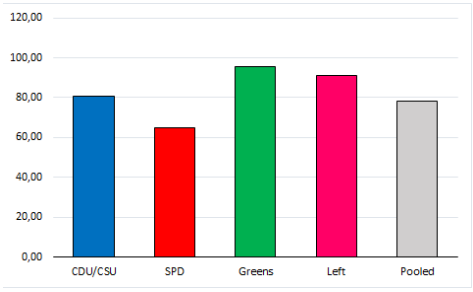

Figure 3: Average number of adjectives per German Wikipedia biography

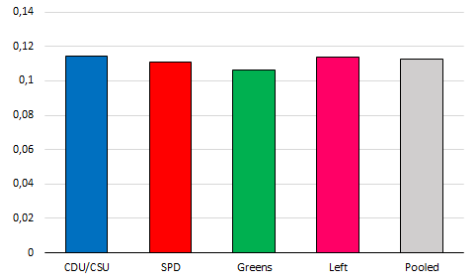

Figure 4: Average adjective/word ratio per German Wikipedia biography

6.2 Images

Similar to the argument supporting the role of adjectives as indicator for brightened texts, it can be argued that a high amount of images may point to partisan activity, since a Wikipedia biography that contains more images might be perceived as more lively and more attractive by its readers. Figure 5 displays the average number of images per Wikipedia biography.

[38] You could, however, also argue that it is necessary to artificially extend a biography to "hide" a lot of beneficial adjectives in a text, in which case the *absolute* amount of adjectives would be the more interesting measure.

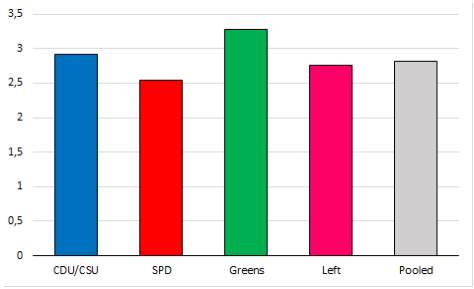

Figure 5: Average number of images per German Wikipedia biography

We find that biographies on MPs affiliated to the SPD contain the smallest number of images, and, in particular, about 1/3 image less than biographies on MPs belonging to the CDU/CSU; as illustrated in Table 10, the difference is significant at the 1%-level. These results match the pattern identified in Section 6.1 and provide further evidence for the SPD being less active on Wikipedia.

6.3 Weblinks under party control

The majority of Wikipedia biographies provides a list of links ("weblinks") to external websites.[39] We say that such a weblink is "under party control" if it links to a website that is under obvious and substantial influence of the party the respective MP is affiliated to. E.g., we count links to MPs' personal or party homepages as "weblinks under party control", but ignore links to the Bundestag's homepage.[40] We argue that a high number of weblinks under party control indicates partisan activity, because it facilitates the use of Wikipedia as a platform for political advertising.

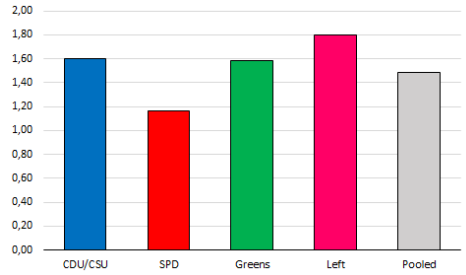

Figure 6: Average number of weblinks under party control per German Wikipedia biography

Figure 6 displays the party averages for weblinks under party control. We find that Wikipedia biographies on MPs from the SPD feature the least weblinks under party control on average, and, in particular, about half a weblink less than biographies of MPs from the CDU/CSU (significant at the 1%-level), which - again - matches the pattern of partisan activity identified in this Section and which confirms our findings from Section 5.

The finding is further supported by one particular incident. According to the Wikipedia discussion sites, it is not so easy to incorporate weblinks into articles. There exists, however, a user called "Cducsu" that has written a program to facilitate the procedure and used it to install weblinks underneath biographies of MPs affiliated to the CDU/CSU that redirect to the homepage of the CDU/CSU parliamentary group.

[39] Note that a weblink is *not* equal to a reference.

[40] The weblinks to bundestag.de redirect to the biographies of MPs found there. These biographies are highly standardized.

6.4 Authors

Wikipedia is a collaborative project that depends heavily on the active participation of its users.[41] Stvilia et al. (2007) find that contributions are often motivated by Wikipedians' interests in specific areas, but also by altruistic motives such as the general desire to fix quality problems or to help the community. Lih (2004) argues that the number of unique authors can serve as objective metric for the quality of a Wikipedia article, because it leads to deeper scrutiny of content.[42] This argument might, however, not be appropriate under the presence of partisan writing.

Figure 7 displays average number of different ("unique") authors per party. We find that biographies of MPs affiliated to the small parties, Greens and Left, exhibit far more unique authors than the biographies of SPD and CDU/CSU. As shown in Table 10, the difference between the SPD and the CDU/CSU is significant at the 5%-level. Against the background of potential partisan activities, this result is not surprising. Users who potentially engage in writing and editing Green and Left biographies are more Internet affine, which might explain the high activity on those sites.

Although it is not feasible to find hard evidence for the existence of partisan writers, we can present some evidence that matches the general pattern of the SPD being less active on Wikipedia. Along with their total number, the German Wikipedia provides a complete list of the unique authors per article, including their logon name or their IP address in case of unregistered users. In a first step, we check for the number of "repetitive" authors per party, i.e., we check how many authors contribute to at least 10% of the biographies per party. This corresponds to contributing to at least 32 biographies of MPs from the CDU/CSU, to at least 20 biographies of MPs of the SPD, and to at least 7 biographies of MPs from the Greens and the Left, respectively. In a second step, we approach to separate neutral Wikipedians from partisan writers. While we expect that partisan writers only spend effort on editing biographies of MPs affiliated to one particular party, Wikipedians that are interested in politics might have an incentive to edit biographies across parties. Thus, we identify which authors appear as "repetitive" authors for just one particular party, and which authors appear as "repetitive" authors for more than one particular party. The former "party specific repetitive" authors might be a rough indicator for the amount of partisan writing per party, while the latter are taken for Wikipedians.

The results are illustrated in Figure 8. We find that there exist both much less "repetitive" and "party specific repetitive" authors for the SPD than for the CDU/CSU. This is particularly remarkable given that we are considering a relative measure here:

[41] The formal organization of the Wikipedia community is as follows (Stvilia et al. 2007): First, there are anonymous users who can only be tracked via their IP address. Second, there are registered users that are identified and tracked by their logon name. Third, there are administrators who are registered users with special system permissions or priviledges such as deleting or undeleting Wikipedia articles or blocking users. Registered users can be promoted to administrators based on their performance, their knowledge of Wikipedia, and the need for additional quality insurance. Users are often assisted by bots (automatic processes) to automate simple, repetitive tasks such as updating templates, spell checking, mining edits for vandalism, and checking for copyright violations.

[42] A similar argument holds for the number of edits.

Table 10: Alternative dependent variables

	Adjectives	Ratio	Images	Weblinks	Authors
CDU/CSU	9.752**	0.00359**	0.340***	0.408***	7.282**
	(4.380)	(0.00165)	(0.100)	(0.0567)	(3.606)
Greens	21.14	-0.00473*	0.756***	0.399***	31.44**
	(13.06)	(0.00253)	(0.254)	(0.0946)	(12.40)
Left	30.42***	0.00286	0.451**	0.643***	31.92***
	(8.189)	(0.00211)	(0.176)	(0.103)	(9.803)
Female	-3.766	0.00000	-0.181*	-0.057	0.782
	(5.138)	(0.00151)	(0.101)	(0.0555)	(4.721)
Sum former periods in Bundestag	11.03***	0.00087**	0.134***	0.062***	14.91***
	(2.072)	(0.00039)	(0.0425)	(0.0139)	(1.820)
Average outside earnings	0.00	0.00000	0.000	0.000	0.000
	(0.000)	(0.00000)	(0.000)	(0.0000)	(0.000)
Doctoral degree	19.31**	-0.00025	0.263**	0.00588	12.87*
	(7.618)	(0.00180)	(0.125)	(0.0714)	(6.977)
Constant	35.67***	0.109***	1.989***	1.111***	14.96***
	(4.754)	(0.00162)	(0.109)	(0.0559)	(4.585)
N	584	584	584	596	596

Robust standard errors in parentheses. All analyses are restricted to Model 4 from Section 5.1.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to be counted as a (party specific) repetitive author on behalf of the SPD, authors "only" had to contribute to at least 20 biographies as opposed to contributing to at least 32, as is the case for the CDU/CSU.

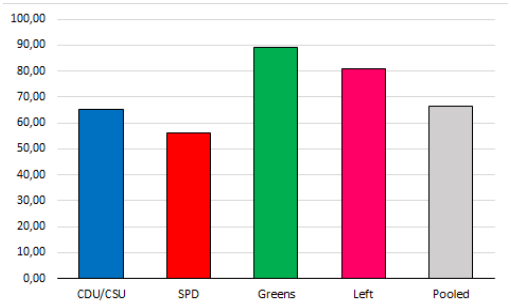

Figure 7: Average number of authors per Wikipedia biography

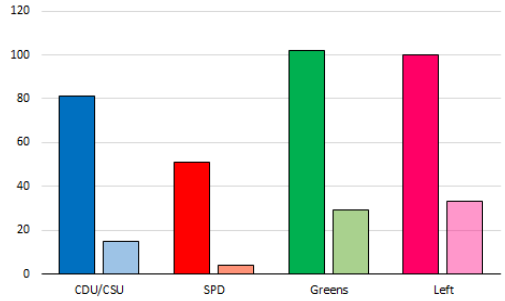

Figure 8: Left bar: Repetitive authors. Right bar: Party specific repetitive authors.

We point out that we do not claim this to be a comprehensive or very reliable measure. It is, of course, possible that partisan authors use many different logon names, that MPs let their staff edit only their personal Wikipedia biography, or that there are Wikipedians who are solely interested in editing the biographies of one particular party. Still, the results nicely fit into the overall impression obtained in this section.

7 Conclusion

We find that Wikipedia exhibits a coverage bias against the SPD in its biographies of members of the 18th German Bundestag, where we define coverage bias as unequal treatment of equivalent cases. We argue that there are two types of writers active on Wikipedia: neutral contributors ("Wikipedians") that devote time and space according to a neutral point of view, and partisan writers, who want to extent the coverage of MPs affiliated to a particular party. Coverage bias on Wikipedia can only arise through *differences* in partisan writing that result in unbalanced coverage.

Biographies of MPs from the SPD are, on average, shorter than biographies of MPs from other parties. This correlation remains after controlling for individual characteristics and demographic properties of the electoral district. We discuss several potential explanations for these correlations. First, it could be that they are driven by unobserved heterogeneity between the politicians of the different parties, which might even lead neutral authors to allocate less space to MPs from the SPD. Second, the differences could be the due to partisan activities.

We use a difference-in-differences approach using MPs' English biographies to measure the effect of party affiliation on biography length. Our main argument is that partisan writers have no incentive to contribute to biographies of German MPs on the English language version of Wikipedia, thus, the English biographies can serve as a benchmark for neutral coverage. Under the additional assumption that unobserved heterogeneity affects English and German biography length equally, and therefore cancels

out, the difference-in-differences estimation supports the role of fewer partisan activity on behalf of the SPD. Further results with respect to the pattern of partisan activities on Wikipedia and a series of robustness checks confirm our results. Importantly, these results do not just reflect a general coverage bias in the German media landscape; as shown by Dallmann et al. (2015), four big German online newspapers do not cover the SPD less than the CDU/CSU.

Our investigation has several limitations, though. First, the external validity of our results beyond the 18th Bundestag is unclear.[43] Given that our main objective is the clean measurement of coverage bias, however, this is a price we are willing to pay. Still, it would be interesting to extend the analysis to, e.g., state parliaments and to check whether the patterns of partisan writing are similar.

Second, our unconventional application of difference-in-differences, though innovative, entails certain disadvantages. Our identification strategy relies on two assumptions, that we cannot test. We assume that (i) the English Wikipedia is free of partisan writing and that (ii) unobserved factors impact German and English length equally. In our analysis, there is no time dimension, and we do not consider the effects of a policy change. Thus, we cannot conduct standard tests such as the investigation of pre-treatment patterns.

Third, we can only measure relative differences in coverage and must speculate about their origin.[44] It is equally plausible to surmise their cause on the demand side than to suspect that they stem from the supply side. There are less potential voters of the SPD than of the CDU/CSU. In particular, four times as many potential SPD voters have recently announced to abstain from voting, e.g., because they are disappointed by the SPD's current policy or because they do not think that the party has a chance to win the elections anyway (Forschungsgruppe Wahlen 2013, 2016). Thus, there could be relatively less demand for Wikipedia biographies of MPs from the SPD such that their extension yields a relatively smaller payoff than the extension of Wikipedia biographies of MPs from the CDU/CSU. Potential SPD voters might also perceive Wikipedia as a less important source of information. Regarding the supply side, it is possible that the relative differences in partisan writing simply reflect idiosyncratic party differences. As argued above, parties could differ in their perception of the importance of an extensive Wikipedia presence or Internet presence in general. Thus, the SPD might not consider Wikipedia as an important tool in election campaigns while other parties do. E.g., the CDU/CSU Secretary-General Peter Tauber (2013) provides a social media compendium that also points to the importance of Wikipedia (p. 12), while nothing comparable exists for the SPD. It is in particular possible that MPs affiliated to the left-wing parties

[43] One potential concern is that the probability of getting elected into the Bundestag may depend on Wikipedia coverage. For out of sample predictions about politicians who are not elected into the Bundestag, one would need to take the sample selection into account. Note that this is a different sample selection issue than the selection into having an English Wikipedia biography discussed above (see Section 4.3).

[44] We contacted the parties' press offices to inquire whether there are coordinated party activities on Wikipedia. According to all parties' press offices replies, there exist no official guidelines for the handling of Wikipedia; every MP is self-responsible for his Wikipedia biography.

Greens and Left perceive a thorough Internet presence as relatively more important than members of Germany's main parties SPD and CDU/CSU.

In sum, there are plausible reasons for our findings, which are not mutually exclusive. Our empirical analysis, however, does not identify the reasons for the differences in partisan writing that we have detected. A comprehensive analysis of this must be left for future research.

A Robustness checks

In this section we perform a series of robustness checks to endorse our results from Section 5 and to show that the finding of a coverage bias against the SPD is not just a coincidence, but robust to variations of the model. To keep the analysis tractable, we focus this section's analysis on the Ordinary Subsample. In addition to that, we restrict the robustness checks to (i) Model (4) of the main analysis, because it is the best controlled specification keeping the majority of observations, and to (ii) Model (3) of the difference-in-difference analysis, since the ML specification is most efficient and we consider it to be most reliable.

A.1 Alternative independent variables

Translations and English homepages It might be easier to write long English biographies if the authors translate from the German counterparts or if significant parts of the MPs' English homepages are provided in English. To take this effect into account, we control for direct translations (biographies marked by a translation template) and English homepages using dummy variables. On top of that, we drop the respective observations from our sample.

The results are given in Columns 1 to 4 in Table 11 and Table 12, respectively. In Table 11, the estimates for the dummy variables are positive, but insignificant (Columns 1 and 3). Relative to the regression results in Section 5.1, magnitude and statistical significance hardly change. The same is true for the specifications in Columns 2 and 4, where we drop the respective observations. Regarding the amplified difference-in-differences estimation in Table 12, we find that the estimates for the dummy variables are statistically significant, but that they do not have the expected sign. If directly translating from German Wikipedia biographies or copying from English homepages would facilitate writing longer English biographies, the impact on the *difference* in length between German and English biographies would be negative. We constitute, however, exactly the opposite. Dropping the respective observations hardly affects the estimates as compared to Section 5.2. Thus, our finding of coverage bias is not affected by the provision of English homepages or by direct translations.

Academic studies and German Landtag In our main analysis, we neither control for whether an MP has studied or not, nor for whether and how many times an MP has been member of a German Landtag (state parliament). Here, we show that these variables do not explain much of the variation in biography length and that they may be legitimately omitted from the main analysis.

The results are given in Column 5 in Table 11 and Table 12, respectively. In both cases, the respective estimates are very small and statistically insignificant. Moreover, relative to Section 5, they do not affect the remaining estimates, either.

Early resigns Throughout the paper, we have excluded twelve observations on MPs who had left the Bundestag until the date of data collection, arguing that they are

not well comparable to the remaining MPs (e.g., involved in scandals). Here, we confirm that including these observations does not affect our results. To avoid multicollinearity, we must exclude the control variable for outside earnings from equation (2) and the selection model consisting of equations (6) and (7), because these observations are missing for all MPs who have resigned early.

The results displayed in Column 6 of Table 11 and Table 12, respectively, confirm that the exclusion of those MPs does not drive our results. The estimates for early resigns are very large and statistically significant at the 5%-level, legitimizing the presumption that they are not well comparable to other MPs in the sample. Still, including them in the analyses hardly affects the remaining estimates.

A.2 Alternative dependent variables

Word count as dependent variable Wikipedia biographies of MPs from the CDU/CSU might be longer on average, because they contain more technical terms and loanwords. To check whether the length of words drives the results in Section 5, we replace the dependent variable $length_i^G$ with the number of words per biography.[45] The results are displayed in Table 12. The estimates are qualitatively unchanged; the magnitudes correspond to what we find in Section 5. Thus, our finding of coverage bias does not stem from the CDU/CSU using more technical terms.

Log-length as dependent variable We also check whether our results from Section 5 are robust to the choice of functional form. Hence, we replace the dependent variables $length_i^G$ with $\ln\left(length_i^G\right)$ and $length_i^E$ with $\ln\left(length_i^E\right)$ such that the dependent variable in the selection model given by equations (6) and (7) becomes $\ln\left(length^G/length^E\right)$. This has the advantage of considering the ratio of German biography length to English biography length instead of their difference, which makes the regression independent of any scaling issues. The results are displayed in Tables 13 and 14. The estimates' signs and magnitudes are qualitatively unaffected, but they are less statistically significant.

Scale German length An alternative way to deal with intrinsic language differences between English and German is to scale down the length of German Wikipedia biographies. There is anecdotal evidence by translators that English texts are about a quarter to a fifth shorter than German texts. To cover a broad spectrum, we scale the German length $length_i^G$ with the factors 0.6, 0.75, 0.8, and 0.9 and re-run the difference-in-differences estimation using $(length_i^G * Factor - length_i^E)$ as dependent variable. The results are displayed in Table 14. Trivially, the smaller the scaling factor, the smaller are the estimates from the difference-in-differences regression. Still,

[45] Since we do not have data on the number of words of the English Wikipedia biographies, we cannot test whether te results from our difference-in-differences estimation are robust to using the number of words instead of biography length in bytes, too.

all estimates are as statistically significant as in Section 5.2, hence, our results are qualitatively robust to scaling down the German biography length.

Table 11: Robustness checks I

	(1)	(2)	(3)	(4))	(5)	(6)
CDU/CSU	864.8**	869.2**	885.0**	981.6***	848.1**	1047.913***
	(354.8)	(358.0)	(352.1)	(353.4)	(353.0)	(381.4)
Greens	1735.5*	1746.7*	1429.9*	1048.8	1599.9*	1910.3*
	(918.5)	(918.6)	(770.1)	(700.0)	(972.5)	(911.8)
Left	2386.4***	2397.8***	2447.1***	2414.3***	2171.8***	2347.9***
	(660.0)	(674.7)	(659.4)	(654.8)	(677.4)	(662.4)
Female	-215.8	-247.0	-182.4	-10.44	-200.6	-421.7
	(402.2)	(408.9)	(393.0)	(379.4)	(406.4)	(408.1)
Sum former periods	875.3***	876.1***	863.7***	824.8***	876.7***	905.5***
in the Bundestag	(160.3)	(161.0)	(156.7)	(150.2)	(160.2)	(159.4)
Average outside	0.00201	0.00204	0.00214	0.00214	0.00205	-
earnings	(0.00170)	(0.00171)	(0.00170)	(0.00169)	(0.00175)	
Doctoral degree	1913.3***	1863.0***	1803.8***	1823.7***	1837.6***	1917.1***
	(609.9)	(618.8)	(623.2)	(618.3)	(609.9)	(613.3)
Translation	374.1					
template	(760.1)					
English			3454.3			
homepage			(2691.9)			
Sum former periods					197.6	
German Landtag					(172.8)	
Academic studies					265.2	
					(536.9)	
Early resign						7629.5**
						(3397.5)
_cons	3325.9***	3339.5***	3291.2***	3274.8***	3087.2***	3402.3***
	(373.1)	(375.5)	(375.9)	(374.6)	(469.1)	(372.1)
N	596	587	596	583	596	607
R^2	0.150	0.149	0.160	0.150	0.152	0.173

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Robustness checks II

	(1)	(2)	(3)	(4)	(5)
CDU/CSU	5752.8***	6200.6***	5993.8***	6054.1***	5879.3***
	(1687.9)	(1808.4)	(1695.4)	(1607.4)	(1718.0)
Greens	7907.4***	8516.5***	6803.9***	6034.9**	7733.7***
	(2387.5)	(2537.8)	(2490.3)	(2393.4)	(2435.4)
Left	11275.2***	11996.4***	11872.5***	11452.2***	11554.3***
	(2328.1)	(2524.7)	(2335.1)	(2194.4)	(2364.0)
Female	-1520.8	-1758.0	-1463.3	-938.9	-1368.0
	(1405.1)	(1518.5)	(1410.5)	(1325.4)	(1425.5)
Sum former periods in the Bundestag	1918.0***	2034.8***	1791.5***	1744.5***	1904.4***
	(367.5)	(390.0)	(370.1)	(354.3)	(371.9)
Doctoral degree	3652.4**	3640.3**	3532.7**	3061.2**	3610.0**
	(1579.3)	(1693.4)	(1593.8)	(1518.4)	(1637.0)
Average outside earnings	0.00152	0.000839	0.00191	0.00193	0.00205
	(0.00741)	(0.00785)	(0.00739)	(0.00688)	(0.00753)
Translation template	10882.6***				
	(3702.1)				
English homepage			7144.3*		
			(3959.6)		
German Landtag					626.0
					(591.0)
Academic studies					951.6
					(1745.7)
Constant	-18692.4***	-19941.7***	-18613.3***	-17638.2***	-19841.3***
	(2348.3)	(2521.9)	(2367.6)	(2261.2)	(2808.9)
Mills	10281.6***	10912.2***	10428.6***	9621.8***	10542.1***
Lambda	885.0	962.4	906.5	881.7	902.5
N	596	587	596	583	596

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Robustness checks III

	(1)	(2)
CDU/CSU	67.73**	0.0889*
	(34.14)	(0.0468)
Greens	177.5*	0.174**
	(94.31)	(0.0783)
Left	242.4***	0.333***
	(66.70)	(0.0747)
Female	-31.03	-0.0513
	(39.30)	(0.0461)
Sum former periods	86.17***	0.116***
in the Bundestag	(15.75)	(0.0148)
Average outside	0.00	0.000
earnings	(0.00)	(0.0000)
Doctoral degree	168.3***	0.229***
	(58.71)	(0.0644)
Constant	341.2***	8.174***
	(36.14)	(0.0443)
N	584	596
R^2	0.151	0.180

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Robustness checks IV

	Loglength	Factor 0.6	Factor 0.75	Factor 0.8	Factor 0.9
CDU/CSU	0.314	3836.8***	4767.6***	5032.0***	5533.5***
	(0.233)	(1212.8)	(1405.9)	(1466.2)	(1590.2)
Greens	0.00663	4359.3**	5818.2***	6255.7***	7111.0***
	(0.347)	(1745.9)	(1993.2)	(2076.7)	(2249.0)
Left	1.035***	8622.4***	9979.4***	10373.3***	11128.3***
	(0.345)	(1714.2)	(1944.0)	(2023.7)	(2189.9)
Female	-0.232	-1570.5	-1506.5	-1499.7	-1483.2
	(0.142)	(999.5)	(1148.4)	(1200.0)	(1306.8)
Sum former periods	0.0645	987.7***	1348.2***	1458.8***	1678.5***
in the Bundestag	(0.0737)	(265.4)	(303.4)	(316.2)	(342.9)
Doctoral degree	0.138	2565.7**	3007.0**	3167.8**	3501.7**
	(0.193)	(1123.5)	(1293.2)	(1351.8)	(1473.3)
Average outside	0.000	0.001	0.001	0.001	0.0012
earnings	(0.000)	(0.005)	(0.006)	(0.006)	(0.007)
Constant	-0.0713	-13683.6***	-15861.5***	-16490.7***	-17738.0***
	(0.952)	(1909.9)	(2043.0)	(2096.4)	(2220.4)
Mills	0.284	6793.5***	8244.2***	8699.7***	9621.6***
Lambda	(0.510)	(812.0)	(807.0)	(820.1)	(854.2)
N	596	596	596	596	596

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Negative coverage

Our analysis is based on the presumption that extensive Wikipedia biographies are beneficial for MPs. Of course, this raises concerns on how to deal with criticism that is included in the biographies. As has been argued in Section 2, downright lies, vandalism, and subjective statements are usually quickly detected by Wikipedia's control mechanisms and are thereupon erased. Still, there might exist objective criticism or discussions. If such content was perceived as harmful rather than beneficial for MPs, our results would be invalid. To probe our empirical strategy, this section examines criticism in Wikipedia biographies more closely.

B.1 Beneficial and harmful criticism

The first important point to note is that there exist several forms of "criticism". On the one hand, there might be criticism that is unambiguously harmful to the MPs, including, e.g., scandals such as plagiarism or consumption of illegal drugs. Major scandals, however, usually result in MPs' resignation, and these observations are excluded from our analysis in Section 5. Wikipedia biographies often discuss MPs' controversial political positions and opinions, which could be termed "criticism", too. It is, however, reasonable to assume that readers' political attitudes are heterogenous. Some readers - reasonably the MPs' very voters - might share the MPs' opinions, while others do not. Thus, a fraction of readers might approve a particular MP *because of* potential controversies. As a consequence, one cannot perceive their coverage as criticism. Similarly, the discussion of controversies cannot be perceived as thoroughly harmful for MPs. If they convince some readers of the MPs' political views, coverage can even be beneficial. Against that background, treating controversies in Wikipedia biographies as neutral content appears to be legitimate.

Second, it could plausibly be the case that MPs benefit from *any* coverage, regardless of whether it is positive or negative. Burden (2002) argues that name recognition constitutes an important factor, in particular if information levels are low. He finds that in such elections (i) the volume rather than the tone of coverage affected voters' assessments of the candidates, and therefore (ii) even negative coverage can benefit a campaign, because it raises voter familiarity with the candidate. It can be argued that this applies to elections for the German Bundestag, too, and in particular to the less prominent MPs in the Ordinary Subsample. Many German citizens often do not know the candidates that run in the respective electoral districts very well. Hence, any Wikipedia coverage can be viewed as beneficial for them.

B.2 Quantity of criticism

To obtain an objective (but only rough) measure of the amount of objective criticism contained in our sample, we count for each biography the number of sentences that criticize the respective MP. We searched each biography for the word stems of "Kritik" / "kritisieren" ("criticism" / "to criticize"), "Diskussion" / "diskutieren" ("discussion"

/ "to discuss"), "Rück- / Austritt" ("resignation"), "Skandal" ("scandal") and "Affaire" ("affair"). We then distinguish between MPs *expressing* criticism and MPs *being* criticized, where only the latter is taken as criticism. English language biographies are searched likewise using the corresponding key phrases.

We find that Wikipedia biographies on MPs affiliated to the CDU/CSU and to the Left contain on average the highest amount of criticizing sentences. Considering the parties that currently constitute the Bundestag, these parties are located at the outer political spectrum, and their respective positions might therefore provoke more controversies. The same argument is valid for the very salient MPs in the Bundestag: if we only consider the Ordinary Subsample, on which the lion's share of our analysis is based, we find much less criticizing sentences. Figure 9 illustrates the party averages of criticizing sentences using both the full sample and the Ordinary Subsample.

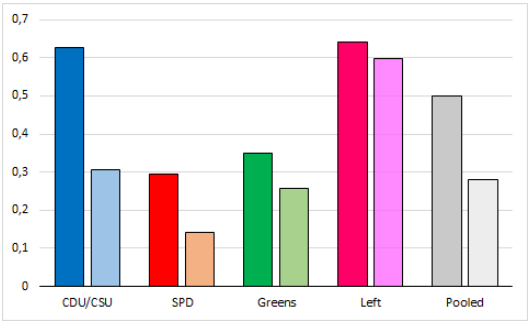

Figure 9: Average number of critical sentences per German Wikipedia biography. Left bar: full sample. Right bar: Ordinary Subsample.

Table 15: Criticizing Sentences

Number	Frequency	Percent	Cumulative
0	533	89.13	89.13
1	24	4.01	93.14
2	19	3.18	96.32
3	7	1.17	97.49
4	5	0.84	98.33
5	4	0.67	99.00
6	2	0.33	99.33
7	2	0.33	99.67
9	1	0.17	99.83
10	1	0.17	100.00

Table 15 illustrates that less than 10% of the biographies in the Ordinary Subsample exhibit more than one single sentence of criticism; about 90% do not exhibit any criticism at all. The patterns of criticism are similar for the English language biographies. Thus, criticism as measured here does not seem to of major concern in our analysis.

To confirm that our results from Section 5 are not driven by differing amounts of criticism in the biographies, we re-run Model (4) of the main analysis and Model (3) of the difference-in-differences analysis on the subsample of the Ordinary Subsample consisting of the MPs whose biographies do not contain any criticism at all. The results are presented in Table 16. The estimates are qualitatively similar to the estimates obtained in Section 5, but smaller and less statistically significant. This is plausible, given that biographies of MPs from the CDU/CSU containing more controversies. Still, the results in Table 15 demonstrate that the coverage bias against MPs from the SPD is not entirely driven by this effect.

B.3 Forks

One further concern relates to "Wikipedia forks"; i.e., independent articles that might cover scandals involving certain MPs. If that happened frequently, our measure of coverage - the length of MPs' *biographies* - would be biased. The phrase "→ Hauptartikel" ("→ main article"), indicates whether particular life events of MPs are covered in separate articles. We searched all biographies for the phrase "→ Hauptartikel" and found a total of 11 biographies that link to forked articles; regarding the Ordinary Subsample, there are only 5 (note that the Ordinary Subsample excludes early resigns, and thereby MPs involved in scandals). MPs Axel Fischer (CDU/CSU) and Michael Hartmann (SPD) are mentioned only briefly in forked articles.[46] The biographies of Annette Groth, Heike Hänsel and Inge Höger, all members of the Left, redirect to an article called "Toiletten-Affaire", which reports on argument between them and their party leader Gregor Gysi that gained certain prominence. The article is, however, rather short. In sum, we do not consider Wikipedia forks as a major problem.

[46] Axel Fischer is mentioned in the article "Internet-Phänomen" ("Internet phenomenon"), Michael Hartmann is mentioned in the article on the "Edathy affair".

Table 16: Regressions on subsample without any criticism

	(1)	(2)	(3)	(4)	(5)	(DD)
CDU/CSU	510.0**	451.1*	439.5*	454.7*	704.2	2637.8**
	(257.5)	(255.0)	(247.3)	(238.3)	(493.0)	(1058.4)
Greens	131.9	185.5	285.4	323.8	-	1605.4
	(312.0)	(309.2)	(300.6)	(295.7)		(1632.6)
Left	575.3	634.3*	728.9**	814.6**	-	3533.0**
	(356.9)	(362.6)	(349.0)	(338.5)		(1714.6)
Female		-396.5*	-312.1	-179.7	-132.9	-831.7
		(226.5)	(219.1)	(214.4)	(336.7)	(948.9)
Sum former periods			370.2***	387.1***	354.1***	714.5***
in the Bundestag			(79.43)	(78.72)	(118.7)	(254.3)
Average outside				0.00166	0.00166	0.00677
earnings				(0.00112)	(0.00120)	(0.00637)
Doctoral degrees				1006.7***	446.8	2264.4**
				(344.0)	(564.5)	(1056.1)
Population density					0.719***	
					(0.200)	
Fraction population					39.03	
18 - 25 years					(117.9)	
Fraction population					-35.23	
with Abitur					(30.28)	
Constant	4539.5***	4703.6***	4121.9***	3810.0***	3781.0*	-8869.0***
	(175.4)	(196.1)	(198.6)	(188.9)	(1951.8)	(1933.6)
Mills						5485.4***
Lambda						(865.9)
N	533	533	533	532	236	539
R^2	0.009	0.014	0.069	0.095	0.132	

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C Variance decomposition

Here we describe the procedure of the variance decomposition from Section 4.2.2 more closely.

First, given that we want to check whether a_i^G and a_i^E are positively correlated, we drop all observations without English Wikipedia biographies, i.e., all observations where we cannot compute $V\left(u_i^E\right)$ and V^{sum}. On top of that, we restrict the analysis to the Ordinary Subsample (See Section 5), on which most of our results are based on, and obtain a total of 138 observations.

Next, we estimate equations (2) and (4) by OLS and save the residuals u_i^G and u_i^E (i) for all observations and (ii) for all parties separately. This allows us to check not only whether a_i^G and a_i^E are on average positively correlated, but also whether this is true for each single party. Using u_i^G and u_i^E, we generate $V\left(u_i^G\right)$, $V\left(u_i^E\right)$, and V^{sum}. Moreover, we compute the difference $u_i^G - u_i^E$ and generate V^{diff}.

Finally, we generate the ratio

$$\frac{V^{diff}}{V^{sum}} = \frac{V\left(\varepsilon_i^G\right) + V\left(\varepsilon_i^E\right)}{V\left(a_i^G\right) + V\left(\varepsilon_i^G\right) + V\left(a_i^E\right) + V\left(\varepsilon_i^E\right)}$$

and obtain the following results

Table 17:

	V^{diff}/V^{sum}
CDU/CSU	0.615
SPD	0.852
Greens	0.722
Left	0.583
Pooled	0.663

D Blinder-Oaxaca decomposition

Let Δ^{μ} denote the difference in average biography length between two parties $j = A, B$, such that

$$\Delta^{\mu} = E\left(lenght_A|X_A\right) - E\left(length_B|X_B\right) = E\left(X_A\right)'\beta_A - E\left(X_B\right)'\beta_B, \tag{9}$$

with $length_j = X_j'\beta_j + \varepsilon_j$ and $E\left(\varepsilon_j\right) = 0$. Equation (9) can be rearranged to

$$\Delta^{\mu} = \underbrace{\left[E\left(X_A\right) - E\left(X_B\right)\right]'\beta^*}_{\text{Explained part}} + \underbrace{\left[E\left(X_A\right)'\left(\beta_A - \beta^*\right) + E\left(X_B\right)'\left(\beta^* - \beta_B\right)\right]}_{\text{Unexplained part}}, \tag{10}$$

where β^* denotes an unknown unbiased coefficient vector. By means of equation (10), we can identify the proportion of the coverage gap explained by differences in the regressors; this is the first component of (10). The second component of (10), the unexplained part of the coverage gap, indicates coverage bias. Replacing the expected values $E\left(X_j\right)$ with their sample averages $\bar{X}_j$, equation (10) is estimated as

$$\hat{\Delta}^{\mu} = \left(\bar{X}_A - \bar{X}_B\right)'\hat{\beta}^* + \bar{X}_A'\left(\hat{\beta}_A - \hat{\beta}^*\right) + \bar{X}_B'\left(\hat{\beta}^* - \hat{\beta}_B\right).$$

The literature suggests several estimators for β^*. For instance, there might be reason to assume that only one group suffers from coverage bias, such that $\hat{\beta}^* = \hat{\beta}_A$ or $\hat{\beta}^* = \hat{\beta}_B$. Neumark (1988) suggests to use the coefficient vector of a pooled regression over both groups as estimate for β^*. We do not have a preference for either estimator and present results for all of them.

References

[1] Aaltonen, Aleksi Ville, and Stephan Seiler. Quantifying spillovers in open source content production: evidence from Wikipedia. London School of Economics and Political Science, LSE Library, 2014.

[2] Angrist, Joshua David, Jörn-Steffen Pischke, and Jörn-Steffen Pischke. Mostly harmless econometrics: an empiricist's companion. Vol. 1. Princeton: Princeton university press, 2009.

[3] Ansolabehere, Stephen, Rebecca Lessem, and James M. Snyder Jr. "The Orientation of Newspaper Endorsements in US Elections, 1940–2002." Quarterly Journal of political science 1.4 (2006): 393.

[4] Baum, Matthew A. "The Iraq Coalition of the Willing and (Politically) Able: Party Systems, the Press, and Public Influence on Foreign Policy." American Journal of Political Science 57.2 (2013): 442-458.

[5] Becker, Johannes, Andreas Peichl, and Johannes Rincke. "Politicians' outside earnings and electoral competition." Public Choice 140.3-4 (2009): 379-394.

[6] Beucker, Pascal. "Wer bin ich?" In: taz. Published: January 10, 2010. URL: http://www.taz.de/!5075666/

[7] BILDblog. "Wie ich Freiherr von Guttenberg zu Wilhelm machte." Published: February 2nd, 2009. URL: http://www.bildblog.de/5704/wie-ich-freiherr-von-guttenberg-zu-wilhelm-machte/. Viewed: November 2016.

[8] Bitkom. (n.d.). Kennen bzw. nutzen Sie die Online-Enzyklopädie Wikipedia?. In Statista - Das Statistik-Portal. Published: January 2011. URL: https://de.statista.com/statistik/daten/studie/170241/umfrage/bekanntheit-und-nutzung-von-wikipedia/. Viewed: November 2016.

[9] Blinder, Alan S. "Wage discrimination: reduced form and structural estimates." Journal of Human resources (1973): 436-455.

[10] Bragues, George. "Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's Entries on Seven Great Minds." MediaTropes 2.1 (2011): 117-158.

[11] Brown, Adam R. "Wikipedia as a data source for political scientists: Accuracy and completeness of coverage." PS: Political Science & Politics 44.02 (2011): 339-343.

[12] Burden, Barry C. "When Bad Press Is Good News The Surprising Benefits of Negative Campaign Coverage." The Harvard International Journal of Press/Politics 7.3 (2002): 76-89.

[13] Chiang, Chun-Fand, and Brian Knight. "Media Bias and Influence: Evidence from Newspaper Endorsements." The Review of Economic Studies (2011): 795-820.

[14] Cohen. "Don't Like Palin's Wikipedia Story? Change It". In: New York Times. Published: August 31, 2008. URL: http://www.nytimes.com/ 2008/09/01/technology/01link.html?_r=0

[15] D'Alessio, Dave, and Mike Allen. "Media bias in presidential elections: A meta-analysis." Journal of communication 50.4 (2000): 133-156.

[16] Dallmann, Alexander, et al. "Media bias in german online newspapers." Proceedings of the 26th ACM Conference on Hypertext & Social Media. ACM, 2015.

[17] DellaVigna, Stefano, and Ethan Kaplan. "The Fox News Effect: Media Bias and Voting." The Quarterly Journal of Economics (2007): 1187-1234.

[18] Druckman, James N., and Michael Parkin. "The impact of media bias: How editorial slant affects voters." Journal of Politics 67.4 (2005): 1030-1049.

[19] Forschungsgruppe Wahlen. "Entpolarisierung statt Mobilisierung. Die Bundestagswahl vom 27. September 2009." In: Weßels, Schoen, Gabriel: "Wahlen und Wähler". Springer Fachmedien Wiesbaden (2013): 30-62.

[20] Forschungsgruppe Wahlen. "Politbarometer Juli II". URL: http:// www.forschungsgruppe.de/Aktuelles/Politbarometer/. Viewed: July 2016.

[21] Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. "Decomposition methods in economics". In: Ashenfelter and Card (eds.): Handbook of Labor Economics, Vol. 4a. North-Holland, 2011.

[22] Gentzkow, Matthew, and Jesse M. Shapiro. "What drives media slant? Evidence from US daily newspapers." Econometrica 78.1 (2010): 35-71.

[23] Giles, Jim. "Internet encyclopaedias go head to head." Nature 438.7070 (2005): 900-901.

[24] Green-Pedersen, Christoffer, Peter B. Mortensen, and Gunnar Thesen. "The incumbency bonus revisited: Causes and consequences of media dominance." British Journal of Political Science (2015): 1-18.

[25] Greenstein, Shane, and Feng Zhu. "Is Wikipedia Biased?." The American Economic Review (2012): 343-348.

[26] Greenstein, Shane, and Feng Zhu. Do Experts or Collective Intelligence Write with More Bias? Evidence from Encyclopædia Britannica and Wikipedia. 2015.

[27] Groeling, Tim. "Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news." Political Science 16.1 (2013): 129.

[28] Groseclose, Tim, and Jeffrey Milyo. "A measure of media bias." The Quarterly Journal of Economics (2005): 1191-1237.

[29] Halavais, Alexander, and Derek Lackaff. "An analysis of topical coverage of Wikipedia." Journal of Computer-Mediated Communication 13.2 (2008): 429-440.

[30] Hallin, Daniel C., and Paolo Mancini. Comparing media systems: Three models of media and politics. Cambridge University Press, 2004.

[31] Heckman, James J. "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." Annals of Economic and Social Measurement, Volume 5, number 4. NBER, 1976. 475-492.

[32] Heckman, James J. "Sample Selection Bias as a Specification Error." Econometrica: Journal of the Econometric Society (1979): 153-161.

[33] Knobloch-Westerwick, S., Mothes, C., Johnson, B. K., Westerwick, A., & Donsbach, W. (2015). "Political online information searching in Germany and the United States: confirmation bias, source credibility, and attitude impacts." Journal of Communication, 65(3), 489-511.

[34] Kürschner Volkshandbuch Deutscher Bundestag. Klaus-J. Holzapfel (ed.). Neue Darmstädter Verlagsanstalt, Rheinbreitbach, 2015.

[35] Larcinese, Valentino, Riccardo Puglisi, and James M. Snyder. "Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers." Journal of Public Economics 95.9 (2011): 1178-1189.

[36] Lih, Andrew. "Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource." Nature (2004).

[37] Luca, Michael. "User-generated content and social media." In: Anderson, Stromberg, Waldfogel (eds.): Handbook of Media Economics, Vol 1b. North-Holland, 2016.

[38] Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. "Promotional reviews: An empirical investigation of online review manipulation." The American Economic Review 104.8 (2014): 2421-2455.

[39] Mesgari, Mostafa, et al. "“The sum of all human knowledge”: A systematic review of scholarly research on the content of Wikipedia." Journal of the Association for Information Science and Technology 66.2 (2015): 219-245.

[40] Mühlhauser, Ingrid, and Friederike Oser. "Does Wikipedia provide evidence based health care information? A content analysis." Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 102.7 (2008): e1-e7.

[41] Neumark, David. "Employers' discriminatory behavior and the estimation of wage discrimination." Journal of Human resources (1988): 279-295.

[42] Noguchi, Yuki. “Wikipedia's Help from the Hill.” In: The Washington Post. Published: February 9, 2006. URL: http://www.washingtonpost.com/wp-dyn/content/article/2006/02/08/AR2006020802212.html

[43] Oaxaca, Ronald. "Male-female wage differentials in urban labor markets." International economic review (1973): 693-709.

[44] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and trends in information retrieval 2.1-2 (2008): 1-135.

[45] Persson, Torsten, and Guido Tabellini. The economic effects of constitutions. MIT Press, 2013.

[46] Puglisi, Riccardo. "Being the New York Times: the political behaviour of a newspaper." The BE Journal of Economic Analysis & Policy 11.1 (2011).

[47] Puglisi, Riccardo, and James M. Snyder. "Newspaper coverage of political scandals." The Journal of Politics 73.03 (2011): 931-950.

[48] Puglisi, Riccardo, and James M. Snyder. "The balanced US press." Journal of the European Economic Association 13.2 (2015): 240-264.

[49] Puglisi, Riccardo, and James M. Snyder. "Empirical Studies of Media Bias." In: Anderson, Stromberg, Waldfogel (eds.): Handbook of Media Economics, Vol 1b. North-Holland, 2016.

[50] Raymond, Eric. "The cathedral and the bazaar." Knowledge, Technology & Policy 12.3 (1999): 23-49.

[51] Reagle, Joseph, and Lauren Rhue. "Gender bias in Wikipedia and Britannica." International Journal of Communication 5 (2011): 21.

[52] Stvilia, Besiki, et al. "Information quality work organization in Wikipedia." Journal of the American society for information science and technology 59.6 (2008): 983-1001.

[53] Tauber, Peter. "Social Media Leitfaden 2013." University of Leipzig. URL: http://www.petertauber.de/wp-content/uploads/2013/04/SocialMedia-Leitfaden-Peter-Tauber-2013.pdf. Viewed: December 2015.

[54] Van Dalen, Arjen. "Structural bias in cross-national perspective how political systems and journalism cultures influence government dominance in the news." The International Journal of Press/Politics 17.1 (2012): 32-55.

[55] Vos, Debby. "Which politicians pass the news gates and why? Explaining inconsistencies in research on news coverage of individual politicians." International Journal of Communication 8 (2014): 2438-2461.

[56] Wessler, Hartmut, and Eike Mark Rinke. "Deliberative performance of television news in three types of democracy: Insights from the United States, Germany, and Russia." Journal of Communication 64.5 (2014): 827-851.

[57] Wooldridge, Jeffrey M. Econometric analysis of cross section and panel data. MIT press, 2010.

[58] Wu, Weifang, Xiaoquan Michael Zhang, and Rong Zheng. "Crowd Governance: The Monitoring Role of Wikipedia in the Financial Market." (2014).

[59] Xu, Sean Xin, and Xiaoquan (Michael) Zhang. "Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction." MIS Quarterly 37.4 (2013): 1043-1068.

[60] Zhang, Xiaoquan Michael, and Feng Zhu. "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia." The American Economic Review 101 (2011): 1601-1615.