

Studying Women's Substantive Representation in India through Parliamentary Questions

Saloni Bhogale¹

Draft Paper: Please do not cite

Abstract

While arguing for a higher representation of female members of parliament (MPs), journalists and researchers have often argued that women MPs are more effective in representing concerns held by Indian women.

In order to test this, texts from the parliamentary proceedings of the Question Hour over a 20-year period are used to discover who is raising concerns about women in the Lok Sabha. This involves the creation of a unique, publicly available dataset of over 300,000 questions, which is processed using unstructured machine learning methods.

Finally, using OLS and NBR regression models, a link is observed between the gender of the MP raising the question and the type of issue raised. While both female and male MPs represent a variety of women's issues across topics, on the whole, there is discernable link between asking questions about women's issues and the gender of the MP raising these concerns.

Keywords: *parliament, representation, India, gender, question hour*

¹ Research Fellow, Trivedi Centre for Political Data, Ashoka University

I. Introduction

Over the last few decades, India has witnessed a closing of the gender bias in political participation (Kapoor and Ravi 2014). This means that more female voters are turning out to vote, and in comparison, the share of male voters has been relatively stagnant. While women in India are able to equally exercise their political preference through the franchisee, the low number of women contesting and winning in parliamentary and state elections begs the question whether there has been a meaningful representation of almost half the population at the various levels of governance. In order to limit this trend, quotas for women were introduced at the local levels in 1993, and they have played a role in increasing the number of female candidates who contest and do well at the national and state legislature (O'Connell 2018). However, only a fraction of such candidates actually win, and even those who do, exercise limited control in influencing policy through their political party organizations. This also means that they had limited power if their goal was the improvement of the status of women in Indian society (Yadav and Verma 1996).

Under this backdrop, a bill was introduced in the Lok Sabha in 2008 for introducing quotas for women at the state level and national level legislatures. The arguments in favor of the bill provide a range of reasons: the need for the presence of feminist women who can strongly represent women (Dasson 2019), enabling women to have a bigger voice in decision making (Khanna and Sharma 2019), raising economic performance (as women are less likely to be criminal or corrupt) (Guha 2018), combating the trend of “women’s issues” being relegated to the private sphere (Roy 2019) and accommodation of women’s perspectives and interests (Lenneberg 1994). Lastly, a common reason for advocating an increase in women in the legislature is the evidence of the differential allocation of public goods by female politicians (Chattopadhyay and Duflo 2004) or less likely to be corrupt (Beaman, et al. 2011), which makes them more effective politicians. These arguments are based on the premise that women perform their duties as legislators in different ways. Lovenduski and Norris elaborate on this:

“There is a question that precedes the idea that an increase in women parliamentarians makes a significant difference in policy – that is, whether female politicians, on account

of their lived experiences, are concerned about issues that are different than that of male politicians.” (Lovenduski and Norris 2003)

These lived experiences may be significantly different from the experiences of males in society and may manifest in different forms. Phillips posits that this distinctive group identity is based upon shared common interests on issues such as abortion, childcare or equal opportunities in education and the labor force (Phillips 1995). At the same time, men and women could share some core beliefs and some concerns may apply to both. However, the argument is that there are significant departures in the experiences and lives of women that makes their concerns unique and form a separate political group.

Such divergences in experience may be even more pronounced in the case of Indian women, who have a lower participation in the workforce, have a lesser say in household finances, have lower access to higher education and also experience bias in workplaces (Batra and Reio Jr 2016). This is in addition to several other forms of discrimination that exist in Indian society against women. However, few political parties have appealed to women as a unique political group by being a champion of their cause. Social attributes derived from religion, caste and class have become the salient traits in the formation of specific group interests and issues of women are typically relegated to the private sphere.

The question then, is that in what form are women’s issues being represented by their elected representatives? Dovetailed on this is the question of whether female politicians more likely to raise such issues than male politicians. This becomes a necessary question to address while thinking about the debates on institutionalizing gender-based quotas.

To answer this, this paper seeks to determine the parliamentary questions which concern women’s issues so that it is possible to study what kind of women’s issues are being represented. Moreover, it becomes possible to answer whether women MPs elected to the Lok Sabha (the lower house of the Indian Parliament) are more likely to raise concerns about women, thereby bringing more attention to women’s issues in the highest legislative chambers.

This is done by looking at the content of questions raised by MPs during the Question Hour of the Indian Parliament. Using a unique dataset of questions and tools from computer science, it is possible to computationally determine the quantum of women's issues that find salience in the house and also show the variety of issues raised. The accuracy of these data is determined by comparing the results of human coders with that of the algorithms. Lastly, the link between legislative behavior that is in the substantive interest of women and the gender of the MP is explored using a simple OLS regression test. This new evidence helps in providing empirical grounding to some of the extant opinions and theories and opens an avenue for understanding representation in the Indian context through the evidence of revealed preferences and behavior.

The next section surveys the comparative literature on this issue - what are the ways in which substantive representation of women has been tested, and what have been the outcome of such studies. Section three introduces the dataset and the methods used for the automated discovery along with reporting the results of the reliability tests and topics discovered. Section four presents the results of a statistical test, answering who is representing women's issues in the Lok Sabha. This is followed by a few observations and a conclusion in the final section.

II. Studying Women's Substantive Representation

This section looks at how scholars have attempted to operationalize substantive representation, that is, in what ways is women's substantive representation manifested and how it can be measured, specifically pertaining to the legislative behavior of elected representatives. While the size and scale of identities in India's context demands particular study, it is possible to draw from the existing comparative work based on other countries, as well as some of the work based in India.

Such studies focus on various aspects of a representative's roles - their voting behavior in the legislatures, debates/motions brought forth by them, performance in committee hearing debates or even through estimating their priorities towards certain kinds of policies using survey data. I briefly summarize some of these studies here before focusing on evidence from India on women's representation.

In a study of the Hong Kong *LegCo*, Waikeung Tam asks whether female legislators have had different policy priorities by looking at the debates and motions introduced. This involved hand coding of debates and motions into several categories (like women's rights, children & families, healthcare, education etc.) (Tam 2017). A similar method is applied to the work on Pakistani legislature (Chowdhury 2018), and comparing Argentina and US legislatures (Jones 1997). Studies on the US Senate involve coding debates/bulletins along with bills introduced and floor speeches, finding that female legislators give a higher priority to women's issues while carrying out their duties as elected representatives (Osborn and Mendez 2010). A different approach is conducting survey studies, for instance, Wangnerud uses data from the parliamentary survey studies in Sweden and compares it with surveys of voters, observing differences between female and male parliamentarians in their views about the duty to promote the interests of women (Wangnerud 2000).

These studies illustrate how the substantive representation of groups can be manifested in a variety of activities carried out by legislators. In the particular case of India, some of these ideas have been tested through the methods and data available on Indian politicians.

While female politicians in other countries express a tendency to focus on women's issues when interviewed (Jones 1997), female legislators in India share a contrasting idea. During interviews that Singh held with MPs from Uttar Pradesh (a populous state in Northern India) in the early 2000s, women legislators shared that they, along with male legislators, raise women specific and strategic issues in the Assembly (Singh 2003). He writes:

“In fact, the legislators were insistent that the tendency to limit women legislators concerns to only women specific issues might harm the cause rather than further it since women's issues were not merely women's concerns but men as well.” (Singh 2003)

Additionally, when MPs were asked what kind of parliamentary questions they raised, the answers were wide ranging – more than 80% said that they raise issues pertaining to their constituencies, about 11% said they raise labour-related issues, crime in western Uttar Pradesh & water/sanitation/education in eastern Uttar Pradesh . Issues of interest to politicians seemed to vary

by the type of constituency (geography, rural/urban) rather than the gender of the politician (Singh 2003).

Another aspect where the substantive representation can be manifested is in the allocation of public goods, however these have often led to conflicting results. While Chattopadhyay and Duflo found that female politicians at the local level in India allocate resources differently than male politicians (Chattopadhyay and Duflo 2004) Bardhan et. al find a worsening of resources targeted by female pradhans at villagers belonging to Scheduled Castes or Scheduled Tribes (Bardhan, Mookherjee and Torrado 2010). Similarly, Iyer and Mani's survey shows that the presence of women village leaders had no impact on any measures of electoral participation for women (Iyer 2019). One explanation for this could be that the elected women are not rooted in social movements that are focused on women's development. One female MP explains it as follows:

“Because of the women that are coming into the Parliament now, they are not necessarily women that have been linked to women's organizations and movements. I won't say they are not interested in women's issues, but they've not been involved in women's organizations and movements. That makes a big difference.” (Rai and Spary 2018)²

The conflicting findings call for more work on this topic. However, studies on Indian legislators cannot exploit conventional strategies like the use of roll-call behavior as an avenue of representation due to the peculiar rules of the Indian Parliament that do not permit a high degree of autonomy to individual legislators. This is due to the rules of the “Anti-Defection Law”, under which an MP has to necessarily vote on legislation as directed by his or her party through the whip (FE Online 2018). This form of control pervades legislative debates as well - there have been cases where MPs have been removed from parties (and thereby forced to vacate their seats in the House) if they fail to toe the party line.

Fortunately, there are other avenues of representation carried out by Indian politicians that can be studied: like introduction of legislation by private members, allocation of public funds, politician surveys and so on. A new way of studying this is by looking at the content of the questions raised by parliamentarians during the Question Hour in the Lok Sabha. One advantage

² Interview with MP 26

of using parliamentary questions to study representation is to overcome problems pertaining to desirability bias that is inherent in interviews with legislators. That is, some politicians may be more likely to share socially desirable reasons for their behavior on account of the public facing nature of their jobs. As there are no reported barriers towards raising questions in the parliament, MPs can exhibit a great amount of autonomy on the kinds of questions they want to ask of the ruling government.

Thus, instead of studying how they ‘say’ they behave, it is possible to study their revealed behavior itself. In the following section, I introduce the dataset of parliamentary questions and the methods used for identifying instances of women’s substantive representation.

III. Data and Methods

The Secretariat of the Lok Sabha and Rajya Sabha maintain a record of questions asked by parliamentarians during each day that the House is in session. The complete data for these questions is available on the Lok Sabha website³ under the Questions Search tab for the last four Lok Sabha Assemblies. For these analyses, only the periods for which the entire set of questions is available at the time of the study, i.e. Winter Session of the 13th Lok Sabha to Winter Session of the 16th Lok Sabha has been used. These were scraped using web scraping scripts written on Python making use of Selenium to crawl through the web pages of the Lok Sabha website. However, even basic information like constituency and gender of an MP are not available in the metadata of the questions available on the website and need to be merged using other data sources.

The Election Commission of India (ECI) is the authoritative source for such data. The ECI releases constituency-level datasets of the General Elections from 1961 onwards. These have been digitized in analyzable formats from PDF files and made available on the website of the Trivedi Centre for Political Data (Jensenius and Verniers 2017). For the purpose of this study, the winners from the elections held between 1999 and 2014 (towards the 13th, 14th, 15th and 16th Lok Sabha) have been considered and we thus obtain the electoral and gender attributes of the MPs raising

³ <https://loksabha.nic.in>

questions. Information on number of terms for candidates has been obtained from the dataset on Individual Trajectories of Candidates in Indian General Elections (Hangal and Verniers 2019).

While the length of an MP's term remains constant across these 20 years as each Lok Sabha lasted its full term (5 years) - the exact numbers for these are indicated by the number of days of parliament in between the date of oath taking for the MP and the date of resignation⁴. For the MPs who held ministerial positions, the net term as an opposition MP/backbencher MP is computed by using the date of appointment and date of relinquishing the cabinet position from the dataset on the Indian Cabinet put together by the Trivedi Centre for Political Data. Using these dates, the total number of days for which an MP was a backbencher, or a part of the opposition was calculated. Thus, the average number of questions raised by an MP per day can be calculated for all of the 2300 MPs. MPs who were ministers for the entire term ($n = 121$) have been excluded from the analysis as it is not their prerogative to raise questions in the house.

Method to extract questions on women

Before elaborating on the method used to extract questions of interest that could be indicative of women's representation, it is first necessary to define the scope of what a concern for women means in this context. Only when a parliamentarian is making explicit reference to some particular concerns that women may have in Indian society across a range of issues, it can be considered as a performance or carrying out of some form of representation. That is, such questions are better defined as what politicians *perceive* to be women's issues.

These concerns may be very different from actual issues in Indian society. Surveying the literature on women's issues in India, crimes against women, violence by spouse/family, women in the labor force, inequality in politics and the lack of financial inclusion and healthcare are among the major issues faced by women in India (Ravi and Jayaram 2017). However, these may (or may not) be included, or cover the spectrum of concerns raised by politicians in the parliament.

⁴ These were obtained from the Lok Sabha's website using the debates published by the secretariat. The data will be shortly made available on the website of the Trivedi Centre for Political Data.

This idea was first posited by Suraj Jacob, who looks at gender and legislative behavior in the Indian Parliament by looking at parliamentary questions (Jacob 2014). In his analysis of questions raised between 1980 and 2009, he finds that female parliamentarians are slightly less likely to raise questions than male parliamentarians in the Lok Sabha. Additionally, for the years between 1999 and 2009, he employs a hand-coding method to categorize parliamentary questions. Given the availability of novel computational tools, hand coding is a painstaking process and falls short of the efficiency (and arguably, the objectivity) provided by using computational tools. Moreover, while the exercise undertaken by Jacob includes detection of words in one-line summaries suggested by the Secretariat resulting in a small set of observations⁵.

This study improves on the idea by using the entire content of questions to determine the topical categories. One improvement is that the use of computational methods that significantly increase the size of observations ($n = 8000$) identified as “women’s issues”. I will elaborate on the method employed in this work in the following section and also compare these with manual coding methods.

The task of finding the questions about women raised in the Parliament was performed using an automated approach using natural language processing tools and a neural network algorithm., finding words which indicate an explicit concern about women. This can be done by finding words that are *similar* to the word woman that appear in the dataset, both in meaning and context. The presence of these terms would indicate that the question concerns some aspect about women.

This subsequently creates the data about women raised in the Lok Sabha as detected by the algorithm. This subset can now be analyzed to determine what kind of topics concerning women are raised in the parliamentary questions. Further, I will elaborate on the major steps in this process: determining words that are similar to ‘woman’ to create a subset of questions on women’s issues, discovering topics that occur within this subset and determining the accuracy of the datasets created.

⁵ Jacob looks at a sample of 3000 questions to determine what words would be indicative of a topic. For instance, he finds that questions concerning womens issues consisted of at least one of the following words or its grammatical forms: woman, mother, girl, balika, gender, mahila, nari, maternal, widow, NCW, maid, dowry, Swadhar, Swayamsidha, surrogacy, devadasi, rape, sati, eve teasing, prostitution. These are extremely similar to the words suggested by the algorithm on using word vectors.

(i) *Determining the words that are similar to woman*

In order to maintain contextual and logical similarity of words indicating a woman related theme, a word-embedding algorithm called word2vec (Mikolov, et al. 2013) was trained to produce a local model, using the Lok Sabha's Question Hour Dataset of over 300,000 questions and answers raised in the last 20 years. All unique words were used to create vector representations of each word so that word similarity can be determined.

The top 500 words similar to woman were retrieved from which 92 words were shortlisted that indicate some reference to women or women's issues in the parliamentary questions. The shortlisted terms from the dataset are provided in Table 1.

The advantages of using a local corpus for training become quickly evident, as some of these words are not conventional English terms but are nouns for Hindi words typically used in the Indian Parliament in place of the word woman (for example, words like *sabla*, *janani*, *mahila* etc.). The algorithm is able to suggest these terms as it is trying to learn both the context and form of the word: that is, it will infer what the word is about and also take into consideration whether it is a noun, verb, etc. to make more meaningful suggestions about semantic similarity.

Abbreviations of schemes/policies/agendas aimed specifically at women are also inferred to be 'close' to the term *woman* (for example, abbreviations like JSY and IGNWPS, which stand for Janani Suraksha Yojana (a scheme for pregnant mothers) and Indira Gandhi National Widows Pension Scheme are examples of policies, while CAW stands for Crimes Against Women). These abbreviations are also a good indicator of the fact that the questions shortlisted pertain to some form or aspect of women's development. Lastly, care has been taken to remove any words that may have multiple meanings. For example, WIPS (which may mean Women in Public Sector or Wireless Intrusion Prevention System) is suggested by the algorithm but not included in the list of shortlisted words to control for disambiguation related problems. The model performs well as it has a reasonable window size ($n = 5$) and looks for topical associations in addition to syntactic relations.

Table 1: Words similar to ‘woman’ discovered by word embeddings

- *Insert Table 1* -

(ii) *Discovering Topics*

A topic model was used to discover meaningful topics from the shortlisted dataset on women’s issues using *number of topics = 50*. The terms indicating these topics were further used as keywords onto the word-embedding algorithm to obtain terms that have a high vector similarity to that of the keyword. Thus, a set of words indicating the topic was retrieved.

Each question can have multiple MPs who have raised the question. At the same time, each question may have n number of topics (where n can be any number between 0 and n-max (19) which is the total number of topics). In order to calculate the number of questions categorized under each topic, we need both the number of MPs raising the question (m) and the number of topics associated with the question (n) for each question.

Now for each topic t, the weight of questions associated with the topic can be determined for each question i:

$$t_i = \Sigma \frac{m_i}{n_i}$$

Table 2 provides the distribution of topics as computed by the above formula. Note that this includes 92% of the questions that could be classified by the algorithm. The topics for the remaining 8% could not be determined and hence have not been reported below.

Table 2: Percentage of Questions on Each Topic by Male & Female MPs ⁶

- *Insert Table 2 Here* -

⁶ Calculated using the dataset of questions on women’s issues detected by the unsupervised algorithm from *Bhogale(2018), Compilation of Parliamentary Questions in the Lok Sabha 1999-2018* which can be currently accessed through GitHub: https://github.com/salonibhogale1/LS_QuestionHour

As we can see, female MPs raise a slightly larger proportion of questions on women's issues on any of these topics when compared to the total strength of female MPs in the house⁷. Their interests seem to be distributed across all topics, which is not very different from the behavior observed for their male colleagues. While this is an analysis obtained by looking at the questions actually asked, section VI will provide a statistical analysis of the questions that includes all MPs, including those who have raised no questions on women's issues.

(iii) *Determining Accuracy*

In order to arrive at meaningful inferences from the dataset created, it is important to check the accuracy of the dataset. This was done by comparing these results with independent human coders, first through an open selection of topics, and later by using a technique similar to the one devised by Grimmer, Westwood and Messing (Grimmer, et al. 2014).

The first round of testing involved two human coders determining whether questions were about women or not, and then classifying it into topics from a set of 20 topics provided in a code book. This method was not effective because of two problems. Firstly, the task of choosing from over 20 topics was overwhelming for the coders. Since questions are not bound within a fixed number of topics, the classification involved multiple iterations of checking whether all the topics had been detected (even when the coders did not know how many topics were involved and when they should stop). The second problem was the codebook itself - since the words used in the codebook could be priming, very few words could be used for describing each topic. Thus the resulting code book sufficiently captured the meaning of each topic while being vague enough so as to avoid to priming of the coders⁸.

In order to get around the first problem, the design of the reliability test now became that of classifying whether or not the coders would agree with a certain classification. That is, an observation was provided, and the coders were asked whether they would agree or disagree with the classification. Each observation consisted of a question text shortlisted as a 'woman's issue',

⁷ During this period, women roughly made up about 10.5% of the total strength of the house across the four terms

⁸ Codebook provided in Appendix B

along with a topic assigned by the algorithm. The first task was to identify whether the question text is indicative of an interest in women's issues or not (i.e. the theme). This resulted in an accuracy of 90%. The second task was determining if the topic was appropriate for the question provided – they agreed with the algorithm 75% of the time. Consequently, both coders were asked to review their differences, and this resulted in an accuracy of 95% for the theme. The following section conducts a statistical analysis of the MPs raising themes pertaining to women's issues.

IV. Analysis of Women's Issues in the Lok Sabha

On retrieving the set of questions raised by MPs about women, the dataset of such MPs raising questions resulted in 2299 observations over 4 terms. The final dataset includes 2170 observations, after excluding the MPs who have been ministers the entire time (121 observations) or resigned immediately after winning an election as they had contested from two seats (8 observations). Some of these may be instances of the same person if an MP has contested and won in multiple terms.

Data on questions about women forms the dependent variable and is obtained by running the algorithms mentioned in section V. This indicates, for each member of parliament, how many total questions were raised during the question hour which indicated some perceived concern for the welfare of women. While these questions may be starred or unstarred⁹, for the purpose of this study, these are aggregated. Thus, the dependent variable is the sum of starred and unstarred questions raised by MPs in the Lok Sabha which concern women, divided by the total number of days that the MP held a seat in the Lok Sabha¹⁰.

In order to analyze the data, an OLS model is used, with the average number of questions concerning women raised in a single day as the dependent variable (see Figure 1 for a comparison). A significant number of the MPs have raised no questions on issues pertaining to women, and thus the value of the dependent variable for these observations is 0. The duration for which the MP was

⁹ From the Lok Sabha's website: A starred question is one to which a member desires an oral answer from the Minister in the House and is required to be distinguished by him/her with an asterisk. Answer to such a question may be followed by supplementary questions by members. An unstarred question is one to which written answer is desired by the member and is deemed to be laid on the Table of the House by Minister. Thus, it is not called for oral answer in the House and no supplementary question can be asked thereon.

¹⁰ Summary Statistics provided in Appendix B

in the opposition or a backbencher (i.e. not a minister) was computed using the data on the Indian Cabinet.

Figure 1: Average number of questions raised per day (by gender)

- *Insert Figure 1 Here* -

The analysis explores the relationship between gender and the questions raised on women's issues in the parliament. Column 2 of Table 3 presents the OLS results for the volume of questions on women's issues raised per day. The results indicate that the seniority of an MP (indicated by the number of terms for which they have been elected to the Parliament) is negatively associated with the questions such MPs raise on women's issues. Similarly, MPs holding ministerial positions are less likely to raise questions on women's issues.

Additionally, the NBR model is used to confirm these results. Here, the dependent variable is the number of parliamentary questions on women's issues raised by the MP. The NBR model is applicable to model count data wherein the dispersion exceeds the mean, and is the standard model used in the literature (Ayyangar and Jacob 2014). An offset indicating the number of functioning days for which the MP holds a seat in the parliament is added to the model. Using the model, it becomes easier to infer some of the trends observed in the data. The results on applying the NBR regression is shown in Table 3.

Table 3: Indian Legislators & Parliamentary Questions on Women's issues¹¹

- *Insert Table 3 Here* -

We similarly observe a statistically significant effect of gender on the content of parliamentary questions. That is, overall, male MPs are less likely to raise women's issues in the house. These observations are confirmed by both the OLS and the NBR model. The IRR is calculated as an exponentiated coefficient as NBR results can be interpreted as incidence rate ratios (IRRs). Table 4 shows that the Gender of an MP is significant at the 99 percent confidence level. The IRR is

¹¹ Calculated using the dataset of questions on women's issues detected by the unsupervised algorithm from Bhogale(2018), *Compilation of Parliamentary Questions in the Lok Sabha 1999-2018* which can be currently accessed through GitHub: https://github.com/salonibhogale1/LS_QuestionHour

0.75, which means that men raise 25% fewer questions than women about women's issues in the parliament. Similarly, MPs who hold ministerial positions during the term ask 26% fewer questions, and senior MPs ask 11% fewer questions. On the other hand, the types of constituencies (which are related to the caste of the MP) does not affect the propensity of an MP to ask questions about women's issues in the Lok Sabha. However, there may be significant differences in the characteristics of MPs apart from gender and comparing MPs who have roughly similar attributes may decrease this effect entirely. This phenomenon has been tested in earlier work (Jacob 2014) which employs a Regression Discontinuity design to establish the same. Thus, while these conclusions hold for the overall trend observed in the Lok Sabha, we are not able to conclude whether gender is the only reason behind the observed differences in the MP's questioning behavior in the Parliament.

V. Conclusion

This paper has so far laid down an alternate approach to create and analyze a dataset that studies the substantive representation of women in the Indian Parliament: through the analysis of parliamentary questions. While the results show that overall, female MPs do not behave in a particularly different manner than their male colleagues (as female MPs overall ask about 13% of the questions, which is slightly larger than the number of seats held by them in the Lok Sabha), we see a slight variation in the average number of questions raised by female and male MPs. Moreover, extraneous factors like number of elected terms, holding ministerial positions, education, and constituency type are also significantly affecting the quantum of questions raised on women's issues by MPs. Hence, an important question is whether the male and female MPs are actually comparable given the other attributes on which they differ. Thus, in addition to the NBR regression employed in this paper, sophisticated methods such as a regression discontinuity design (Jacob 2014) might be better suited to carry out the analysis.

Secondly, while the methods described may be novel when applied to Indian political data, they have their promises and pitfalls. The promise is in identifying themes, which the algorithm is able to do with a very good accuracy. However, accurately classifying the questions into topics is a harder task resulting in a lower accuracy. This raises questions about the statistical treatment of

such results, which may be misleading due to the variability in the accuracy of the dataset, which is itself a difficult problem to capture and measure.

Finally, the results in this paper are in line with what we observed earlier about women in politics: while their low numbers in legislatures remain an issue to deliberate upon, an argument about their effectiveness about their substantive representation requires more operationalizing by further studying the various roles played by MPs once they are elected, and the different capacities in which they are able to champion the causes that are important to them on account of their interests or identities.

Bibliography

- Kapoor, Mudit, and Shamika Ravi. 2014. "Women Voters in Indian Democracy: A Silent Revolution." *Economic and Political Weekly* 49 (12): 63-67.
- Kapoor, Mudit, and Shamika Ravi. 2014. "Women Voters in Indian Democracy: A Silent Revolution." *Economic and Political Weekly* 49 (12): 63-67.
- O'Connell, Stephen. 2018. "Can Quotas Increase the Supply of Candidates for High Level Positions? Evidence from Local Government in India." *IZA Institute of Labor Economics Discussion Paper Series*.
- Yadav, Gyanendra, and Ravindra Kumar Verma. 1996. "Women in Bihar Politics." *Economic and Political Weekly* 31 (15): 935-937.
- Khanna, Pretika, and Neetu Chandra Sharma. 2019. "Elections 2019: Women Reservation in Politics Back on the Agenda." *Live Mint*, March 13.
- Guha, Ramachandra. 2018. "Why the Women's Reservation Bill must be revived." *The Telegraph*, November 24.
- Roy, Pragya. 2019. "Why the Women's Reservation Bill is the Need of the Hour." *Feminism in India*, June 18.
- Chattopadhyay, Raghavendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409-1443.
- Beaman, Lori A, Esther Duflo, Rohini Pande, and Petia Topalova. 2011. "Political Reservation and Substantive Representation: Evidence from Indian Village Councils." *India Policy Forum* 159-201.

- Lenneberg, Cornelia. 1994. "Women and Political Leadership in India: Able Politicians or Token Presences?" *Asian Studies Review* 17 (3): 6-14.
- Dasson, Bharvi. 2019. "We not only need women candidates, we need feminist women in Parliament. In conversation with Kamla Bhasin." *Times Now News*, April 25.
- Lovenduski, Joni, and Pippa Norris. 2003. "Westminster Women: The Politics of Presence." *Political Studies* 51 (1): 84-102.
- Phillips, Anne. 1995. *The Politics of Presence*. Oxford: Oxford University Press.
- Batra, Renu, and Thomas G Reio Jr. 2016. "Gender Inequality Issues in India." *Advances in Developing Human Resources* 18 (1): 88-101.
- Tam, Waikung. 2017. "Do Female Legislators have different Policy Priorities than their Male Colleagues in an Undemocratic/Semi-Democratic Legislature? The Case of Hong Kong." *The Journal of Legislative Studies* 1-27.
- Chowdhury, Nusrat Jahan. 2018. "Who Speaks for Women in Parliament? Patriarchy and Women MNAs in Pakistan." In *Women in Governing Institutions in South Asia*, by N. Ahmed. Cham: Palgrave Macmillan.
- Jones, Mark P. 1997. "Legislator Gender and Legislator Policy Priorities in the Argentine Chamber of Deputies and the United States House of Representatives." *Policy Studies Journal* 25 (4): 613-629.
- Osborn, Tracy, and Jeanette Morehouse Mendez. 2010. "Speaking as Women: Women and Floor Speeches in the Senate." *Journal of Women, Politics & Policy* 31: 1-21.
- Wangnerud, Lena. 2000. "Testing the Politics of Presence: Women's Representation in the Swedish Riksdag." *Scandinavian Political Studies* 23 (1): 67-91.
- Singh, Pitam. 2003. *Women Legislators in Indian Politics*. Concept Publishing Company.
- Bardhan, Pranab K., Dilip Mookherjee, and Monica Parra Torrado. 2010. "Impact of Political Reservations in West Bengal Local Governments on Anti-Poverty Targeting." *De Gruyter* 1 (1): 1-38.
- Iyer, Lakshmi. 2019. "Getting More Women into Politics: Evidence from Elections in India." *VOX CEPR Policy Portal*. May 24. Accessed August 2019. <https://voxeu.org/>.
- Rai, Shirin M., and Carole Spary. 2018. *Performing Representation*. Oxford: Oxford University Press.
- FE Online. 2018. "What is a Whip? What role with the Chief Whip play when Modi Government faces No-Confidence Motion in the Lok Sabha? ." *Financial Express*, July 19.
- Jensenius, Francesca, and Gilles Verniers. 2017. "Studying Indian Politics with Large-Scale Data: Indian Election Data 1961 - Today." *Studies in Indian Politics* 5 (2): 269-275.

- Hangal, Sudheendra, and Gilles Verniers. 2019. *Individual Trajectories of Candidates in Indian General Elections 1962 - current*. Trivedi Centre for Political Data.
- Ravi, Shamika, and Nirupama Jayaram. 2017. "Gender Issues in India: An Amalgamation of Research." *Brookings*, March 10.
- Jacob, Suraj. 2014. "Gender and Legislative Performance in India." *Politics and Gender* 10: 236-264.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv*.
- Grimmer, Justin, Sean Westwood, Messing, and Solomon. 2014. *The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability*. Princeton University Press.
- Ayyangar, Srikrishna, and Suraj Jacob. 2014. "Studying the Indian Legislature: What does the Question Hour Reveal? ." *Studies in Indian Politics* 2 (1): 1-19.

Tables

Table 1: Words similar to 'woman' discovered by word embeddings

abortion	abortions	antenatal	balika	BBBP ¹²	beti	brides
bride	CAW ¹³	CEDAW ¹⁴	creche	creches	daughter	daughters
dowry	EWR ¹⁵	EWRs	female	females	foeticide	foeticides
girl	girls	housemaids	IGMSY ¹⁶	IGNWPS ¹⁷	janani	JSSK ¹⁸
JSY ¹⁹	kishori	KSY ²⁰	lady	ladies	mahila	maid
maids	maternal	maternity	matritva	menstrual	midwife	midwives
midwifery	mother	mothers	motherhood	MTP ²¹	nari	neonatal

¹² Beti Bachao Beti Padhao

¹³ Crimes Against Women

¹⁴ Convention on the Elimination of all Forms of Discrimination Against Women

¹⁵ Elected Women Representatives

¹⁶ Indira Gandhi Matritva Sahyog Yojana

¹⁷ Indira Gandhi National Widow Pension Scheme

¹⁸ Janani Shishu Suraksha Karyakaram

¹⁹ Janani Suraksha Yojana

²⁰ Kishori Shakti Yojana

²¹ Medical Termination of Pregnancy

PMSMA ²²	pregnancy	pregnancies	prostitutes	prostitution	rape	rapes
RGNCS ²³	sabla	samakhya ²⁴	samridhi	sati	sportswomen	stree
surrogacy	surrogate	swadhar ²⁵	Swayam-siddha	swayam-sidha	widow	widowed
widows	wife	wives	woman	women	womens	

Table 2: Percentage of Questions on Each Topic by Male & Female MPs

Topic	% of questions	Percentage by Male MPs	Percentage by Female MPs
Employment & Skill Development	17.4%	86.4%	13.6%
Crimes against Women & General Safety - General, Sexual Crimes, Prostitution, Trafficking, Foeticide, Dowry & Womens Safety	18.3%	86.7%	13.3%
Womens Education	9.4%	87.2%	12.8%
Health Issues	8.1%	86.2%	13.8%
Pregnancy, Surrogacy & Abortion	6.8%	87.0%	13.0%
Development	8.6%	85.1%	14.9%
Reservation/Representation & Legislation	2.0%	86.3%	13.7%
Intersectionality: with Caste, Religion, Tribes	5.6%	89.1%	10.9%
Infrastructure: Hostels Orphanages Prisons	3.2%	87.0%	13.0%
Welfare of Widows	8.1%	86.2%	13.8%
Marriage, Family & Child Rearing	2.4%	86.3%	13.7%
Poverty	2.2%	88.7%	11.3%
Women and Army / Defence	2.0%	86.3%	13.7%

²² Pradhan Mantri Surakshit Matritva Abhiyan

²³ Rajiv Gandhi National Creche Scheme

²⁴ Mahila Samakhya (Education for Womens Equality)

²⁵ Swadhar Greh Scheme

Sports	1.2%	84.7%	15.3%
Women in Science / Technology	1.1%	85.3%	14.7%
Water and Sanitation	1.1%	88.6%	11.4%
Agriculture & Farmer Issues	0.9%	78.5%	21.5%
Censorship: Indecent Representation of Women	0.5%	91.1%	8.9%
Children	2.7%	90.3%	9.7%
All Questions	100%	86.7%	13.3%

Table 3: Indian Legislators & Parliamentary Questions on Women's issues

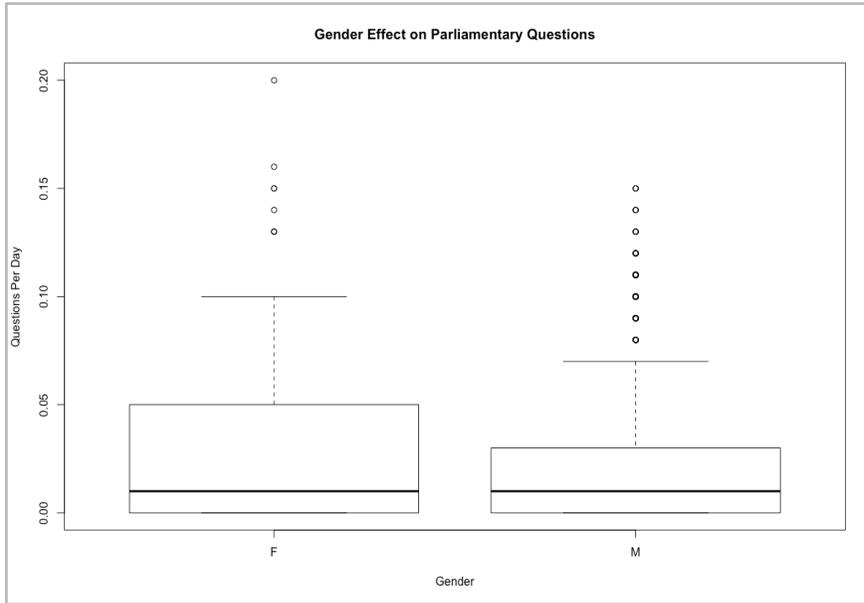
<i>Variables</i>	<i>(1) OLS</i>	<i>(2) NBR</i>
Gender (Male)	- 0.007 ** (0.002)	- 0.294 ** (0.107)
Seniority of MP	- 0.002 *** (0.000)	- 0.112 *** (0.021)
SC Constituency	- 0.000 (0.002)	- 0.002 (0.092)
ST Constituency	- 0.001 (0.002)	- 0.115 (0.121)
Held a Ministerial Post during Term	- 0.006 ** (0.002)	- 0.307 * (0.127)
Observations	2170	2170

Note:Parameter estimates calculated using lm() and glm() in R version 3.5.3

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Figures

Figure 1: Average number of questions raised per day (by gender)



Appendix A

Topic No.	Topic	Terms indicating the topic
1	Women's Education	education,literacy,teacher,teachers,teach,learning,schooling,school,schools ,enrolment,rmsa,rashtriya_madhyamik_shiksha_abhiyan,sarva_shiksha_abhiyan,girl_students,ssa,secondary_education,elementary_education,rashtriya_madyamik_shiksha_abhiyan,educational,gross_enrolment_ratio,higher_secondary_schools,udise,kasturba_gandhi_balika_vidyalaya_scheme,text books,dropout,samakhya,university,ugc,students,ssc,vidyalayas,vidyalaya, scholarships,scholarship,college,illiteracy
2	Crimes against Women & General Safety - General, Sexual Crimes,	foeticide,infanticide,foeticides,female_foeticide,sex_selective,sex_determination,crime,crimes,atrocities,cyber_crime,investigation,cyber_crime_prevention,ccpwc,kidnapped,firs,offences,iucaw,fast_track_courts,criminal,kidnapping,kidnappers,caw,kidnappers,criminal_justices_ystem,assault,attack

	Prostitution, Trafficking, Foeticide, Dowry & Women's Safety	,attacks,tease,teasing,acid,slavery,slaves,sexual,harassment,harassments,molestation,rape,rapes,sexual_offences,sexual_harassment,nirbhaya,dowry,violence,mutilation,devdasi,sati,prostitution,prostitute,prostitutes,brothels,pimp,pimps,porn,pornography,safety,trafficking,immoral_trafficking,human_trafficking
3	Health Issues	healthcare,health_care,health,doctor,cancer,hospital,nutrition,malnutrition,aids,hiv,breast,hygiene,disease,diseases,anaemia,nutritious
4	Welfare of Widows	widow,widows,ignwps,martyrs,widowed
5	Marriage, Family & Child Rearing	family,marriage,marriageable,divorce,marital,brides,bride,familial,child_marriage,family_courts,marriages,bigamy,polygamy,hindu_succession_act
6	Pregnancy, Surrogacy & Abortion	pregnancy,maternal,maternity,abortion,pregnant,pregnancies,childbirth,natal, prenatal,antenatal,fertility,abortions,caesarean,janani_surakshya_yojana,jsy,foetus,contraceptive,contraceptives,baby,babies,igmsy,janani_suraksha_yojana,indira_matritva_sahyag_yojana,jssk,surrogacy,surrogate,assisted_reproductive_technology
7	Intersectionality: with Caste, Religion, Tribes	muslim,minority,minorities, madarsa, madarsas, madrasa, madrasas, muslims, minority_women, christians, jains, buddhists, parsis, jains, tribe, tribal, tribes, scheduled_tribe, scheduled_tribes, sts, st, indigenous, adivasi, caste, scheduled_caste, scheduled_castes, castes, dalit, scs, sc, dalits
8	Women in Science / Technology	science_and_technology, research, scientists, biotechnology, technologists, women_scientist, women_scientists, women_scientists_scheme, women_scientists_scholarship_scheme, women_scientists, women_technology_parks
9	Infrastructure: Hostels Orphanages Prisons	hostel, hostels, shelter, shelters, nikan, nikan, orphanages, disabled, disability, orphan, orphans, jail, jails, tihar, prisoners, prisoner, undertrial, tihar_jail
10	Sports	sport, sports, sports_women, football, games, martial, kho, national_sports_talent_contest_scheme, sports_promotional_schemes, hockey, nstc, national_sports_talent_contest, olympics, judo, sporting, sports_authority, wrestling, physical_education, tennis, rajiv_gandhi_khel_abhiyan, athletics, khokho, national_sports_development_fund, national_services_scheme, special_area_games,

		north_east_sports_festival,wrestler,wrestlers,athletes,rural_sports_program,umpires,football,weightlifting,olympics,olympic_games,gymnastics,shooting,kabaddi,nis,olympic,asian_games,rgka,national_sports_federations,national_championship,sag
11	Employment & Skill Development	unemployment,employment,employed,remuneration,working_women,working_mother,working_mothers,women_working,women_worker,women_workers,working_condition,working_conditions,workplace,entrepreneur,entrepreneurs,entrepreneurship,bank,banks,loan,loans,credit,labourer,labourers,mgnrega,worker,enterprises,employee,employees,job,jobs,savings,financing,cooperative,workers,industry,income,director,directors,wage,wages,unskilled,skilled,swarozgar,rajeshwari,lic,vacancy,vacancies,insurance,sgsy,recruitment,promotion,promotions,recruit,recruitments,recruited,ceos,rashtriya_mahila_kosh,rmk,training,vocational,skill,industrial_training_institutes,itis,itcs,modular_employable_skills,ncvt,jss,seekho_aur_kamao,pradhan_mantri_kaushal_vikas_yojana,pmkvy,vtc,vtcs
12	Women in Army / Defence	armed, defence, army, armed, forces, navy, bsf, crpf, airforce, navy, cisf security_personnel, pmf, battalion, jawan, military, battalions, battalion
13	Reservation/Representation & Legislation	reservation,rights,reservations_for_women,reservation_for_women,representation_of_women,women_representation,law,laws,legislation,legislations,bill,bills
14	Censorship: Indecent Representation of Women	sensor,film,films,television,censor_board,press_council,obscene,tv,irwa,censorship,indecent_representation,nude,nudity,indecent
15	General Development	shgs,shg,self_help_group,self_help_groups,cedaw,pmuy,balika_samridhi_yojana,nwc,ncw,national_commission,ksy,swadhar_yojana_scheme,bbbp,kishori_shakti_yojana,beti_bachao_beti_padhao,mahila_samridhi_yojana,mahila_samaridhi_yojana,indira_mahila_yojana,swadhar_yojana,lpg,jgsy,implads,pris,pri,panchayati_raj_institutions,panchayati_raj,pmeysa,panchayat,sex_ratio,rajiv_gandhi_kishori_sashaktikaran_yojana,adolescent,swayam_siddha,sukanya,samriddhi,swayamsiddha

16	Water & Sanitation	toilet,toilets,water,sanitation,drinking_water,potable,community_sanitary_complex,bathing,latrines,swachh_vidyalaya,washing,ihhl,ihhls,individual_household_latrines,drains,swachh_vidyalaya_initiative
17	Agriculture & Farmer Issues	farmers,farms,agriculture,krishi,agro
18	Poverty	poverty,poverity_line,poorer,backward_classes,backward_class,poor
19	Children	child,children,cwsn,creche,creches,child_development,icds,integrated_child_development_service,infant,child_protection,ncper

Appendix B

Descriptive Statistics

<i>Variable</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Number of Questions per day	0.02	0.03	0	0.2
Total Number of Questions raised in term	6.3	8.63	0	66
Gender	0.89	0.30	0	1
Seniority of MP	2.15	1.61	1	10
Constituency Type: GEN	0.77	0.42	0	1
Constituency Type: SC	0.14	0.35	0	1
Constituency Type: ST	0.08	0.27	0	1
Minister During Term	0.08	0.28	0	1