Appendices/Supplementary Information for Online Publication: Electoral Systems and the Substantive Representation of Marginalized Groups

Contents

Appendix A Parochial Bills and Electoral Systems **A3**

Appendix B Detailed Explanation of Machine Learning Process **A4**

Appendix C Descriptive Statistics **A8**

Appendix D Predicting Sponsorship and Passage using the Alternative Categorization of Women's Bills **A9**

Appendix E Description of Randomly Sampled Bills **A17**

Appendix F Main Analyses **A21**

Appendix A Parochial Bills and Electoral Systems

In Table A below, we calculate the proportion of parochial bills among women-related legislation (defined in the methods section) introduced by legislators in PR and SMD seats. Parochial bills are defined as those that target narrowly defined beneficiaries, such as specific groups of individuals, institutions, or geographical regions. These groups were identified by a keyword search of bill titles that include words such as single parent, Korean expats abroad, war veterans, retired soldiers, teachers, mixed blood individuals, North Korean refugees, victims of land mines/natural disasters/man-made disasters, Korean Japanese, artists, farmers, and public officials. Specific institutions include research institutes, religious organizations, military, government institutions, or educational institutions. Targeted regions include areas affected by natural disasters, areas near by military bases, areas near by large infrastructures (e.g. dam, nuclear power plants...etc), or rural areas.

While these data only provide descriptive evidence, the table confirms that politicians in SMD seats have a greater propensity to target specific groups of individuals or regions.

Table A1: Share of Parochial Bills Sponsored by Legislators in Different Electoral Systems, 2004-2016

	Parochial bills (%)			
Congress #	17th	18th	19th	Avg.
PR legislators	10.07	9.56	9.03	9.55
SMD legislators	20.30	17.83	15.62	17.92

Appendix B Detailed Explanation of Machine Learning Process

Introduction: We followed a systematic process for categorizing the approximately 63,000 Korean legislative bills into 24 categories, of which 21 are adopted from categorizes used in Comparative Policy Agenda as listed below. We added two additional categories that are specific to South Korean contexts: North Korea and History (e.g. Gwangju democratic movements, Truth reconciliation committees, Korean war...etc). Finally, we added an 'others' category for bills that cannot be categorized under these 23 labels.

21 Bill Categories adopted from Comparative Policy Agenda projects

Macroeconomics; Civil rights & Liberty and Minority issues; Health; Agriculture & Fisheries; Labor and Employment; Education; Environment; Energy; Immigration; Transportation; Law & Crime; Social policy; Regional and urban policy & planning; Banking, finance and internal trade; Defense; Space, Science & Communications; Foreign trade; International affairs&Foreign aid; Governance & Government operations; Public lands & Water management; Culture & Media

In order to categorize bills, we split our process into iterations, which consisted of first hand coding a small subset of bills, and second using supervised machine learning techniques to categorize a large subset of bills (all based on bill titles). We followed this sequential process five times. Had we stopped after the first iteration we could have categorized all remaining bills with 81 percent accuracy; however, we deemed this level of accuracy insufficient. By adhering to our iteration process, we were able to successfully categorize all approximately 63,000 bills with less than two percent error.

The classifier used is "softmax regression" from the LogisticRegression Python function. Python can apply either a logistic regression or softmax regression model depending on how many classes there are. Note that we have 24 classes rather than two.

Steps:

1. Clean Korean bill titles

   (a) Split every character

   (b) Split and keep all numbers

2. Divide data into train, validate, and test sets

   (a) Approximately 60-20-20 percentage split

3. Make a pipeline that applies the TF-IDF vectorizer and a logistic regression algorithm to the training dataset in sync to avoid unwarranted information leakage

4. Analyze the error on the validation set

   (a) Find total number incorrect

   (b) Calculate percentage incorrect

5. Iterate through the prior steps if unsatisfied with the error on the validation set

6. Analyze the error on the test set

7. Apply algorithm to unseen data

8. Filter probability at 90 percent

9. Persist with those bills with higher than .9 probability

   (a) Push those bills with lower than or equal to .9 probability to next iteration

10. Repeat steps on next iteration

Train-Validate-Test Procedure: The training set is used to train the algorithm. The validation set is used to validate the training of the algorithm. The validation set is a semi-unseen dataset: we allow the algorithm to see it after it is trained, but the validation set is not used to train the algorithm directly. If the algorithm was allowed to see the validation set, information would be "leaked" into the training dataset. This process is valid because we use the test set at the end, before the algorithm classifies truly unseen data. So, applying our algorithm on the test set tells us how our algorithm is likely to perform on truly unseen data. This means we can test different algorithms on the training dataset as long as what we do helps to better predict

on the validation set. Generally, in machine learning, the algorithm tends to do slightly worse on the test set than on the validation set. Usually, this difference is negligible, as it was in our case.

Each iteration proceeded as follows. First, a few thousand bills were hand-coded by the authors into one of the 24 categories using the text of the bill titles. Unlike in many legislatures, bill titles in Korea are highly descriptive and individual bills encompass only one topic area.[12] Because there are over 30,000 bills in the sample, it is impractical to read bill text and classify accordingly. To ensure that bill titles accurately reflect bill content, we randomly sampled 20 bills, read the full text, and compared their content to the title description. The results are shown in Table E1 in the appendix, and confirm the usefulness of using titles to categories bills.

The hand-coded bills were split into three categories: train, validate, and test. The first subset of bills was used to train the algorithm, which uses the text in the already classified data to determine how to classify additional bills. The validate subset of the hand-coded data is used to determine how well the algorithm classifies bills as compared to the human coder. That is, it compares its own classification to the hand-coded classification and generates an error rate (the percentage of bills classified as falling into different topic areas by the algorithm and the human coder). Finally, the last subset of data is classified using the algorithm. The algorithm does not have access to the test set of bills until the error rate for the validation set is sufficiently low, to ensure that the error rate will be generalizable to the test set. Thus, the error rate is a measure of how well the algorithm will do in classifying the next set of uncategorized bills *a priori*.

For each iteration, the error rate was less than 2%, meaning for the subset of data on which the algorithm classification was compared to the human classification, there was a topic area discrepancy for less than 2% of the bills. We repeated the above process through five iterations, with a sixth iteration in which the remaining bills were human coded. Iterations are necessary because the algorithm classifies bills into one of the 24 topic areas and assigns a probability that each bill falls into that category. We required that the algorithm be more than 90% confident that a bill fit into that category, a highly conservative (i.e., bills are highly unlikely to be misclassified) threshold. If the algorithm was not 90% sure that a bill fits into a category, it was

[12] In the United States, for example, many bills are composed of a combination of smaller bills, making them harder to categorize.

not classified. Thus, each iteration finishes with a successful classification of a certain number of bills, and remaining bills which cannot be classified with 90% certainty. A subset of these additional bills were hand-coded, and the process repeated. Bills which could not previously be classified by the algorithm can be classified in a future iteration because as more bills are hand-coded, the algorithm "learns" how to classify additional bills. We set the 90% certainty threshold because that maintained an error rate under 2%; as the threshold declines, the algorithm assigns codes to a greater set of bills for which it is less confident, but more discrepancies between human coding and machine coding occur.

Table 3: Details of Supervised Machine Learning Outcomes for Each Iteration

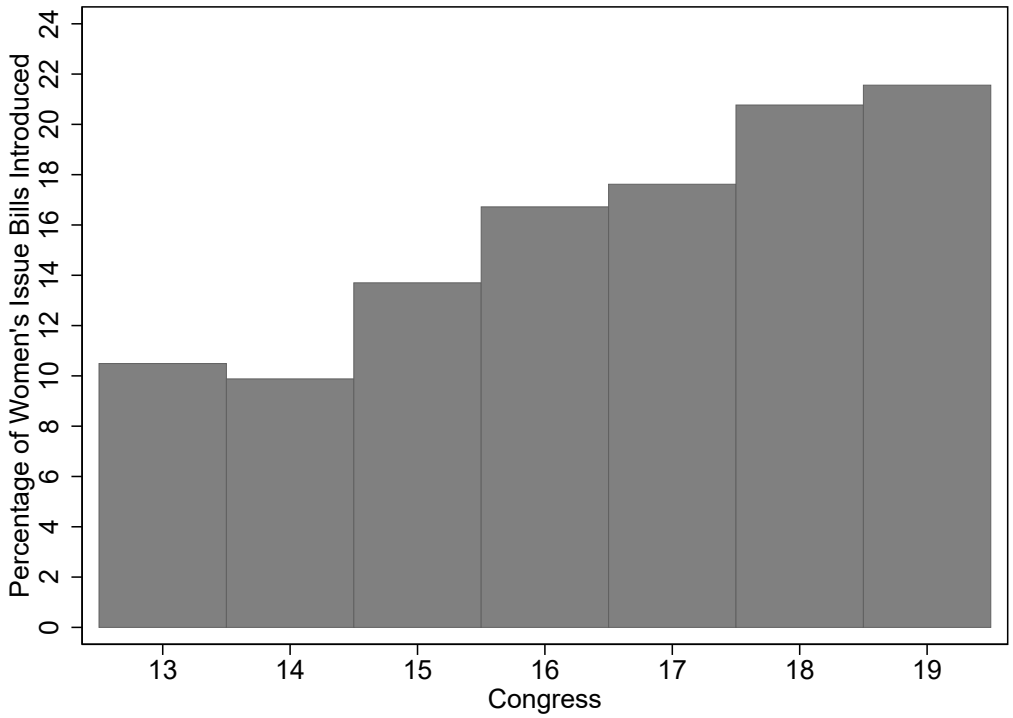
Iteration	No. Of Unclassified Bills at Start	No. Hand Coded	No. Machine Classified	Percentage Error	No. of Unclassified Bills at End
1	62923	6003	24849	1.11%	32071
2	32071	1998	8775	1.80%	21298
3	21298	2031	4678	1.57%	14589
4	14589	2011	3420	1.59%	9158
5	9158	2031	1754	1.97%	5373
6	5373	5373	*NA*	NA	*NA*
Total	62923	19447	43476	1.61%	*NA*

Appendix C Descriptive Statistics

Table C1: Percentage of Korean Legislators by Electoral System and Gender, 2004-2016

Congress	17th		18th		19th		Total
Gender							
	PR	SMD	PR	SMD	PR	SMD	
Men	8.5	72.91	7.37	77.17	8.23	73.18	85.26
Women	14.74	3.86	10.15	5.13	10.55	8.05	14.74
Across all Congresses							
	PR	SMD					
Male	6.56	78.71					
Female	9.24	5.5					
Total	15.80	84.2					

Figure C1: Percentage of Women's Issue Bills Introduced in Korean National Assembly by Congress



Note: Bars show the percentage of sponsored bills classified as women's issue bills. PR system introduced in the Korean Assembly in the 17th Congress.

Appendix D Predicting Sponsorship and Passage using the Alternative Categorization of Women's Bills

As noted in the text, we construct our own bill categorization as a robustness check using a keyword search of bill titles. Korean bill titles are long and descriptive and provide information on the content of the legislation. Further, Korean bills only cover one subject. This categorization is different from the one created in the text in that rather than using a supervised machine learning process, we simply categorize a bill as directly addressing a women's issue if it contains at least one of the following keywords. The key words commonly appear in bill titles and explicitly target women's issues, narrowly defined, consistent with an approach which typically categorize women's interests as those that directly affect women (e.g., reproductive rights, gender-based violence), or those related to women's traditional roles as caregivers (e.g., childcare) O'Brien and Piscopo (2019). The number of women's bills categorized using this method is far fewer than the number found by the supervised machine learning process (10,729 bills or 17% of the sample are identified using the machine learning process, whereas only 1,844 bills or 4.34% are identified using the key word search process.)

Words/phrases used to identify women's issue bills

- "daycare"
- "childcare" *or* "infant care"
- "child education support"
- "gender equality"
- "mother-child welfare"
- "single parents"
- "sexual harassment" *or* "sexual violence" *or* "sexual assault" *or* "domestic violence"
- "prostitution"
- "female scientists"
- "Committee of women"
- "gender discrimination"

- " women's jobs" *or* "women's career" *or* female employment
- "pregnant women" *or* "pregnancy" *or* "child birth"
- "family-friendly business"
- "women in agriculture"

The results shown below are robustness checks of the main results presented in the text. The results are consistent with those presented in Table 1. Figure D1 shows the predicted probabilities for the interaction term in model 2 in Table D1, and the differences between legislators in SMD and PR districts are consistent with the results shown in Figure 1 though the estimated effect is much smaller, due to the smaller number of bills classified as addressing women's issues. Note that we only show results for women's issue bills because the keyword search identifies those. The results for all bills are shown in the main text.

The results using the dependent variable confirm that both women and men in PR sponsor more women's issue bills than representatives in SMD seats, though women in either system sponsor more than men in either system, supporting Hypothesis 1 and 2. Differently from the results in the main text however, women in PR do not sponsor more women's issue bills than women in SMD at a statistically significant level, though the point estimate is higher.

Similarly, the results in Figure D2 showing the conditional effect of gender and electoral system on legislative passage are also very similar. Notably, and consistent with the results in the main text, women in PR systems are more successful at passing their legislation as compared to men in SMD systems. The differences between men in PR and women in PR or SMD are not statistically significant, as they are in the main text. Again, we attribute the large confidence interval to the paucity of observations. Broadly these results are consistent with the main results, while also demonstrating the utility of using the supervised machine learning process to identify a much larger number of women's issue bills than would be identified by a simple keyword search.

While this narrow definition of women's issue bills addresses issues that are directly relevant to women, it is important to note that other issues such as minimum wage laws, and regulations related to part-time jobs that may not only target women but disproportionally affect females are excluded. Given their socio-economic status in the society, a majority of part-time, irregular jobs with low wage and little employment protection are taken by women. In addition, although legislations related to welfare for other individuals such as the elderly are not included in this alternative definition of women's issue bills, one can argue that when elderly care is not subsidized by government, the burden of care work are most likely to fall on

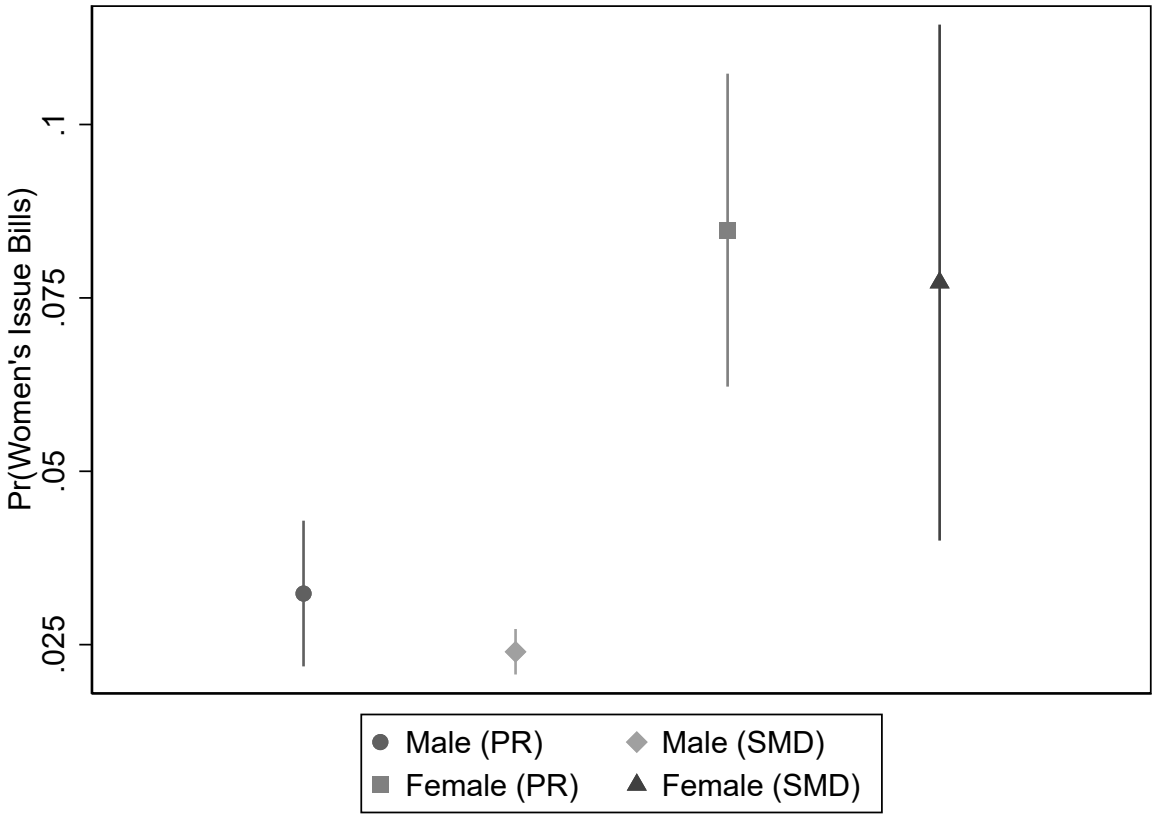
women's shoulders, particularly given the socio-cultural context of South Korea. Therefore, we believe that though our alternative analysis shows that there exists a large gendered difference in sponsorship of bills related to childcare and sexual violence, regardless of electoral systems, representation styles are still very important in promoting (not undermining) the substantive representation of women, as politicians in PR seats are more likely to legislate bills, such as those related to labor market, social welfare, and civil rights, that hugely affect women's rights and welfare in multifaceted ways.

Table D1: The Effects of Electoral Systems and Gender on Women's Bill Sponsorship—Alternative DV

	(1)	(2)
Gender of Sponsor (Female=1)	1.14*	1.23*
	(0.19)	(0.28)
Electoral System of Sponsor (PR=1)	0.20	0.31
	(0.20)	(0.19)
Terms Served of Sponsor	-0.15*	-0.14*
	(0.07)	(0.07)
Sponsor University Educated (Yes=1)	0.01	0.03
	(0.24)	(0.23)
Sponsor Age	0.17#	0.18#
	(0.10)	(0.10)
Sponsor Age^2	-0.00*	-0.003*
	(0.00)	(0.001)
Number of Bills Introduced by Sponsor	-0.004*	-0.004*
	(0.002)	(0.002)
Sponsor Moved to PR	-1.40*	-1.45*
	(0.65)	(0.63)
Sponsor Moved to SMD	-0.09	-0.13
	(0.28)	(0.29)
Number of Bill Cosponsors	-0.00	-0.00
	(0.00)	(0.00)
Yearly GDP Per Capita	1.45*	1.48*
	(0.71)	(0.71)
Gender x Electoral System		-0.21
		(0.36)
Constant	-2.91	-3.20
	(2.90)	(2.99)
Party Fixed Effects	Yes	Yes
Congress Fixed Effects	Yes	Yes
Pseudo R-squared	0.05	0.05
N	30,230	30,230

Note: The dependent variable is whether the bill is classified as addressing women's issues based on containing one of the keywords listed in Appendix B. GDP Per Capita measured in tens of thousands of won. All models are logistic regression with standard errors clustered by legislator (596 clusters). Standard errors in parentheses. *p<.05

Figure D1: Predicted Probability of Electoral System and Gender on the Sponsorship of Women's Issue Bills—Alternative DV



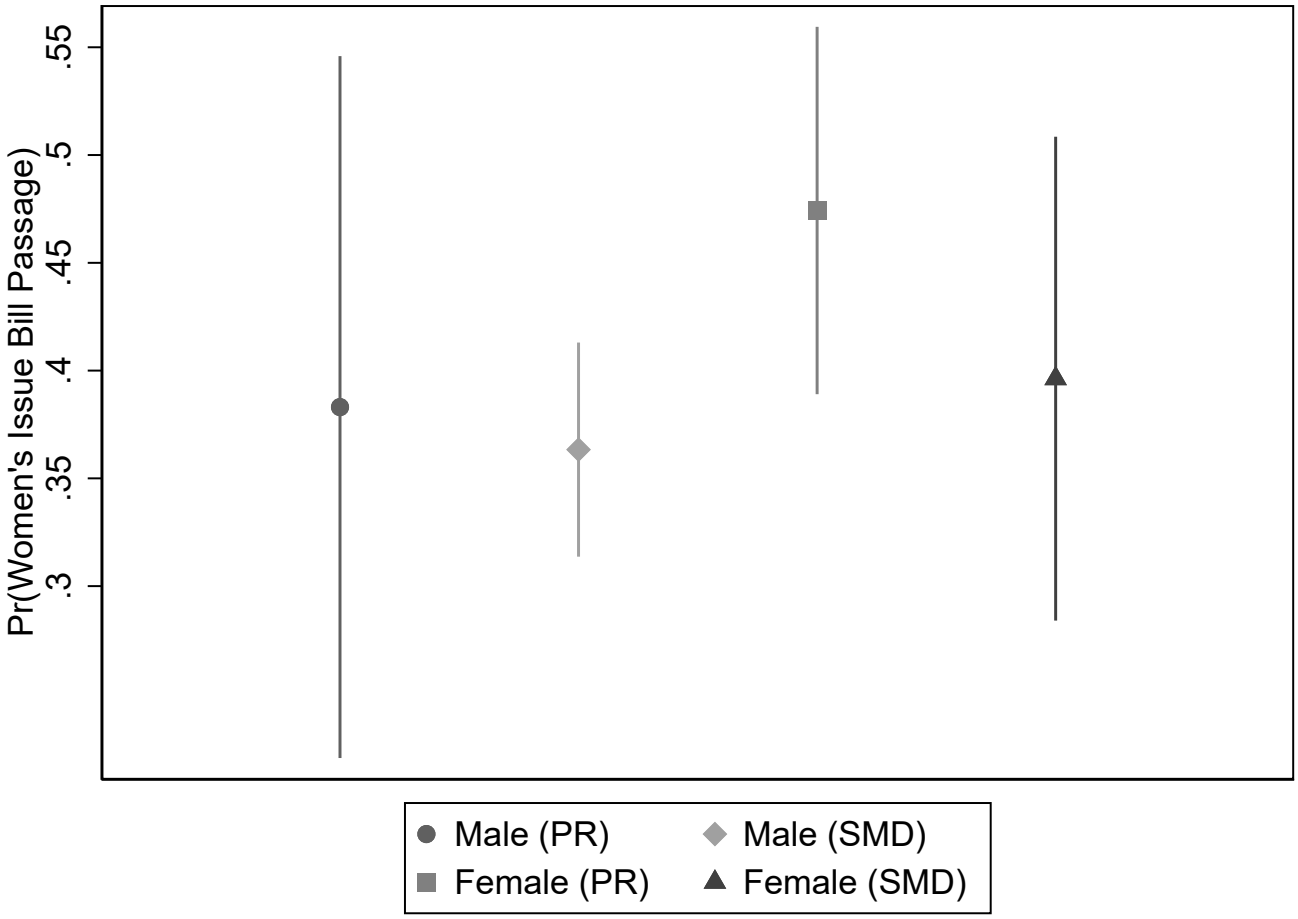
Note: Predicted probabilities from results presented in model 2 in Table D1. Lines above and below point estimates display 95% confidence intervals.

Table D2: The Effects of Electoral Systems and Gender on Women's Issue Bill Passage—Alternative DV

	(1)	(2)
Gender of Sponsor (Female=1)	0.25	0.15
	(0.24)	(0.29)
Electoral System of Sponsor (PR=1)	0.23	0.09
	(0.24)	(0.40)
Terms Served of Sponsor	-0.08	-0.09
	(0.10)	(0.11)
Sponsor University Educated (Yes=1)	-0.03	-0.08
	(0.33)	(0.34)
Sponsor Age	-0.05	-0.06
	(0.14)	(0.14)
Sponsor Age^2	0.001	0.001
	(0.002)	(0.002)
Number of Bills Introduced by Sponsor	0.0002	0.0002
	(0.003)	(0.0003)
Sponsor Moved to SMD	-0.33	-0.28
	(0.24)	(0.26)
Number of Bill Cosponsors	-0.02*	-0.02*
	(0.01)	(0.01)
Yearly GDP Per Capita	-4.84*	-4.87*
	(1.35)	(1.36)
Gender x Electoral System		0.25
		(0.50)
Constant	9.95*	10.32*
	(4.42)	(4.56)
Party Fixed Effects	Yes	Yes
Congress Fixed Effects	Yes	Yes
Pseudo R-squared	0.04	0.04
N	1,047	1,047

Note: The dependent variable is whether the bill passed the National Assembly for women's issue bills only using the alternative dependent variable described in Appendix D. GDP Per Capita measured in tens of thousands of won. Sponsor Moved to PR variable is not included because all values are zero. All models are logistic regression with clustered standard errors by legislator (284 clusters). Standard errors in parentheses. *p$<$.05

Figure D2: Predicted Probability of Electoral System and Gender on Women's Issue Bill Passage—Alternative DV



Note: Predicted probabilities from results presented in model 2 in Table D2. Lines above and below point estimates display 95% confidence intervals.

Appendix E Description of Randomly Sampled Bills

Table E1

Congress Number	Bill Number	Gender of Sponsor	Title/Summary
18	1803028	M	**Act for the Enhancement of Convenience for the Disabled, Elderly, and Pregnant Women** Establish accountability systems in each relevant bureau and merge multiple and redundant committees in order to enhance the effectiveness of the decision-making process.
18	1811008	M	**Partial Amendment to the Elementary and Secondary Education Act** Grant full-time teacher status to health instructors to enhance the quality of education in the context of increased school violence and student health issues related to drugs, smoking, drinking, and sexual misconduct.
17	172494	M	**Partial Amendment to the Elementary and Secondary Education Act** In order to enhance education development in rural areas and low-income neighborhoods in cities, governments are to relax regulations on public schools and allow hybrid forms of public schools (e.g., charter schools).
19	1909688	M	**Partial Amendment to the Infant Care Act** Current legislation requires proof of employment from both parents for a student to be eligible for priority in daycare admission. However, often the children of parents in the agricultural or fishing sectors have been excluded from such support due to their inability to provide official documentation. The proposed bill aims to address this problem by including these children and consider their parents as full-time workers
18	1814214	F	**Partial Amendment to the Child Welfare Act** Provide ”monthly children cash subsidies” to minors under the age of 12.
19	1901614	F	**Partial Amendment to the Child Welfare Act** Mandate education for adults who are convicted of child abuse, including those who are suspended from prosecution. Also, the definition of child abuse should include ”domestic violence”.
19	1903335	M	**Partial Amendment to the Child Welfare Act** Mandates a childs guardian to take the necessary measures for the regular health screening for a child.
18	1807987	M	**Partial Amendment to the Act on Sexual Protection of Children and Adolescents** The term, ”Yeoja” used in the current Act on Sexual Protection of Children and Adolescents tends to imply a pejorative concept of women and stereotyped role of females. Therefore, this term is to be replaced with ”Yeoseong” to promote a more gender equal concept.
17	173480	M	**Elderly Pension (”Hyodo” Pension) Act** Provide a national pension to senior citizens aged 65 or older who meet certain economic criteria in order to reduce rates of elderly poverty.

Table E1 (continued)

Congress Number	Bill Number	Gender of Sponsor	Title/Summary
17	176273	F	**Partial Amendment to the Act on Sexual Protection of Youth** In order to prevent the recurrence of offense, the government should expand and enhance the registration system, access system, and employment restriction system. Also, in the case of rape of a child under the age of 13, probation and suspension of the prosecution should be restricted.
19	1917907	F	**Partial Amendment to the Act on Enforcement of Child Support** Those who have not provided child support can be banned from traveling abroad. Additionally, relevant government bureaus can provide information about the person's income and assets without his/her consent.
18	1808451	M	**Partial Amendment to the Act on Equal Employment and Work-Family Compatibility Support** Expand parental leave to parents who adopt a child
17	174684	F	**Partial Amendment to the Basic Women's Development Act** In order to enhance the outcomes of women-related facilities, national and local governments should evaluate the performance of its centers and facilities related to women (e.g. women's career development center) and publicly report the evaluation results. These results will be used in assessing the future budget and government subsidies.
19	1900281	M	**Partial Amendment to the National Health Promotion Act** Establish electronic systems which allow local governments to assess household economic circumstances in order to enhance the effectiveness of child care subsidies.
19	1915988	M	**Legislation on the use of hospice and palliative care** Establish the system for individuals to make end-of-life decisions that can enhance the quality of life for patients and their guardians.
17	172561	F	**Partial Amendment to the Maternal and Child Health Act** National and local governments should fund all preventative care and vaccinations for pregnant women, infants, and children.
19	1902867	M	**Partial Amendment to the Long-term Care Insurance Act** Remove any existing language in the present versions of the law that can implicitly exclude senior citizens who live alone from accessing long-term insurance coverage for at-home care.

Table E1 (continued)

Congress Number	Bill Number	Gender of Sponsor	Title/Summary
19	1913090	F	**Partial Amendment to the Basic Act on National Demographics** National and local governments should prepare a separate budget to address challenges related to reduced birthrates and population aging.
18	1813881	M	**Partial Amendment to the Act on the Prevention of Sexual Violence and Victim Protection** Institutional capacity development to support disabled victims of sexual violence.
19	1915961	M	**Partial Amendment to the Welfare Act for the Disabled** Provide tax subsidies to aid the transportation needs of disabled people living in rural areas or those with low incomes.

Appendix F Main Analyses

Table 4: The Effects of Electoral Systems and Gender on Women's Bill Sponsorship

	(1)	(2)
Gender of Sponsor (Female=1)	0.44*	0.56*
	(0.15)	(0.19)
Electoral System of Sponsor (PR=1)	0.45*	0.54*
	(0.16)	(0.19)
Terms Served of Sponsor	-0.01	-0.01
	(0.03)	(0.03)
Sponsor University Educated (Yes=1)	-0.27	-0.24
	(0.23)	(0.23)
Sponsor Age	0.06	0.06
	(0.06)	(0.06)
Sponsor Age^2	-0.001	-0.001
	(0.001)	(0.001)
Number of Bills Introduced by Sponsor	0.04	0.05
	(0.2)	(0.1)
Sponsor Moved from SMD to PR	-1.79*	-1.83*
	(0.35)	(0.38)
Sponsor Moved from PR to SMD	-0.24	-0.29
	(0.19)	(0.20)
Number of Bill Cosponsors	-0.01*	-0.01*
	(0.00)	(0.00)
Yearly GDP Per Capita	-0.09	-0.08
	(0.53)	(0.52)
Gender x Electoral System		-0.24
		(0.28)
Constant	-3.80	-4.10
	(2.17)	(2.17)
Party Fixed Effects	Yes	Yes
Congress Fixed Effects	Yes	Yes
Pseudo R-squared	0.03	0.03
N	30,252	30,252

Note: The dependent variable is whether the bill is classified as addressing women's issues.
GDP per capita measured in tens of thousands of won.
Coefficients for Number of Bills Introduced multiplied by 100.
All models are logistic regression with standard errors clustered by legislator (598 clusters).
Standard errors in parentheses. *p<.05

Table 5: The Effects of Electoral Systems and Gender on Bill Passage

	All Bills		**Women's Issue Bills**	
	(1)	(2)	(3)	(4)
Gender of Sponsor (Female=1)	-0.14*	0.26	0.05	-0.07
	(0.07)	(0.13)	(0.09)	(0.16)
Electoral System of Sponsor (PR=1)	0.11	0.03	0.38*	0.41*
	(0.07)	(0.09)	(0.12)	(0.15)
Terms Served of Sponsor	-0.04	-0.04	0.02	0.02
	(0.02)	(0.02)	(0.04)	(0.04)
Sponsor University Educated (Yes=1)	0.12	0.10	0.21	0.21
	(0.09)	(0.09)	(0.17)	(0.17)
Sponsor Age	0.003	-0.002	0.03	0.03
	(0.04)	(0.04)	(0.06)	(0.06)
Sponsor Age^2	0.2	0.3	-0.004	-0.002
	(0.04)	(0.04)	(0.1)	(0.1)
Number of Bills Introduced by Sponsor	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Sponsor Moved to PR	0.80*	0.84*	0.68	0.67
	(0.34)	(0.37)	(0.97)	(0.97)
Sponsor Moved to SMD	-0.09	-0.05	-0.11	-0.12
	(0.12)	(0.11)	(0.15)	(0.16)
Number of Bill Cosponsors	-0.0002	-0.0002	-0.01*	-0.01*
	(0.002)	(0.0002)	(0.00)	(0.00)
Yearly GDP Per Capita	-3.49*	-3.49*	-3.94*	-3.93*
	(0.35)	(0.36)	(0.63)	(0.63)
Gender x Electoral System		0.23		-0.07
		(0.13)		(0.23)
Constant	3.89*	4.15*	2.43	2.33
	(1.18)	(1.26)	(2.22)	(2.21)
Party Fixed Effects	Yes	Yes	Yes	Yes
Congress Fixed Effects	Yes	Yes	Yes	Yes
Pseudo R-squared	0.02	0.02	0.03	0.03
N	30,279	30,279	6,758	6,758

Note: The dependent variable is whether the bill passed the National Assembly.
Models 1 and 2 show the results for all bills, models 3 and 4 restrict the sample to bills classified as addressing women's issues. GDP Per Capita measured in tens of thousands of won.
Coefficients for Sponsor Age^2 are multiplied by 100.
All models are logistic regression with clustered standard errors by legislator.

(599 clusters in models 1 and 2, 513 clusters in models 3 and 4).
Standard errors in parentheses. *p<.05