Replication and Extension of "Gendered Language on the Economics Job Market Rumors Forum"

Martin Huang

1.Introduction and Paper Overview

Alice H. Wu's (2018) paper, "Gendered Language on the Economics Job Market Rumors Forum," provided a way to show the academic atmosphere for women working in this field. This paper employed text analytic methods to examine gendered discourse within the anonymous Economics Job Market Rumors (EJMR) forum. Using a LASSO logistic model, Wu demonstrated systematic differences in how forum participants discussed women and men. Her analysis showed that posts referring to women disproportionately emphasized physical appearance and personal characteristics, without focusing too much on their academic behavior. However, posts referring to men largely focused on professional and academic achievements. Wu thought that this linguistic asymmetry reflected an exclusionary professional environment to some extent.

This paper reproduces and extends Wu's research. We first replicated her analysis using the EJMR post samples from Wu's work and retained the original LASSO logistic regression model with only minor code adjustments to ensure compatibility. The replication process encountered practical constraints and employed tracing methods to verify the accurate execution of the original R code. Despite these adjustments, our replication results confirm Wu's core findings and align with her original tables and figures.

Building on the replication, we introduce a random forest classifier as a methodological extension. This approach uses the same training data and textual features as the LASSO model,

enabling a comparison of predicted word rankings across different machine learning frameworks.

2.Discussion of Replication Attempt and Challenges Encountered

This section discusses our attempt to reproduce the findings of Wu (2018) and the primary limitations encountered during the replication process. Overall, the replication successfully reproduced the main empirical results, confirming the robustness of the original findings.

Several practical challenges arose during the replication. First, minor code modifications were required to ensure compatibility with current software environments and package versions. These adjustments did not alter the model specification or estimation procedures, but they illustrated the sensitivity of computational replications to software versions. Tracing methods were employed to verify that all estimation steps were executed as intended.

Second, computational resource constraints limited our ability to fully replicate all subsample analyses. While the LASSO logit model was rerun on the full sample, the pronoun subsample relied on intermediate outputs provided by the authors under the same model specification.

In addition, the extension using a random forest classifier yielded differences in model structure relative to the original LASSO approach. Unlike the LASSO model, which produces signed coefficients, the random forest relies on feature importance measures. As a result, comparisons across models require auxiliary assumptions and should be interpreted in conjunction with the LASSO coefficients.

Despite these limitations, our replication and extension produced results that are highly consistent with those reported in Wu (2018), suggesting that the main empirical findings are

robust.

3.Replication Results

Our replication successfully reproduced the empirical findings reported by Wu (2018) in Table 1, Table 2, and Figure 1. The analysis focused on gendered word identification and key robustness checks. These results closely match those in the original study.

3.1 Gendered Words Findings

The replicated LASSO logistic regression identified predictive word sets for female- and male-associated posts consistent with Wu (2018). Table 1 reports the top ten words and their marginal effects from the full sample. These closely match the original estimates.

Table 1: Replication and Extension Table 1: Word Importance Analysis

ID	Word	ME	Word	ME
8915	hotter	0.422	homo	-0.303
4136	pregnant	0.323	testosterone	-0.195
9138	plow	0.277	chapters	-0.189
3299	marry	0.275	satisfaction	-0.187
1107	hot	0.271	fieckers	-0.181
7147	marrying	0.260	macroeconomics	-0.180
5068	pregnancy	0.254	cuny	-0.180
1697	attractive	0.245	thrust	-0.169
1531	beautiful	0.240	nk	-0.165
7890	breast	0.227	macro	-0.163

Female-associated words primarily emphasized physical appearance and personal or family roles, with relatively few academic or professional terms. Male-associated words largely reflected academic, professional, and institutional contexts. Both the rankings and effect magnitudes closely mirror those in the original study. This confirms the accuracy of the replication.

3.2 Robustness Check

Robustness checks using pronoun-gendered posts yielded results consistent with the original

findings. Table 2 presents estimates from the pronoun subsample. It shows that core gendered language patterns persist even after excluding potentially non-professional relationships.

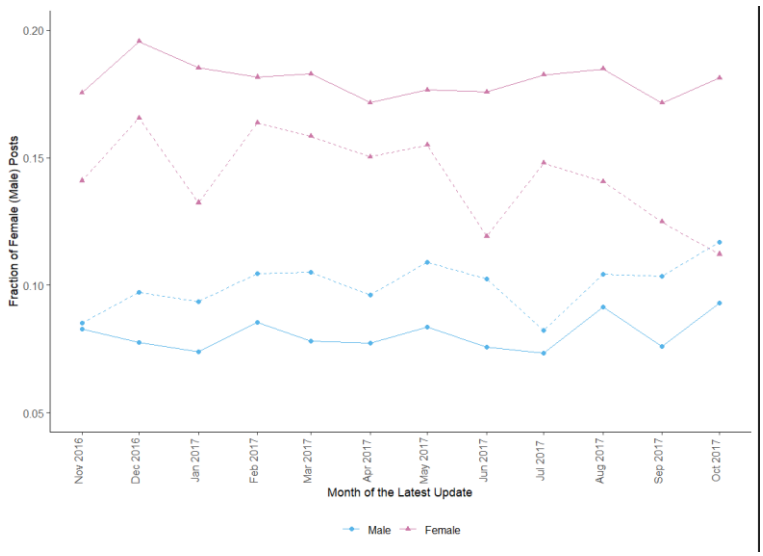
Table 2: Replication and Extension Table 2: Word Importance with Pronoun

ID	Word	ME_pronoun	Word	ME_pronoun
5068	pregnancy	0.292	knocking	-0.329
8915	hotter	0.289	testosterone	-0.204
4136	pregnant	0.258	blog	-0.183
8660	hp	0.238	hateukbro	-0.176
9018	vagina	0.228	adviser	-0.175
7890	breast	0.220	hero	-0.174
9138	plow	0.219	cuny	-0.173
4308	shopping	0.207	handsome	-0.166
3299	marry	0.207	mod	-0.166
7859	gorgeous	0.201	homo	-0.160

Within this subsample, female-associated words continued to emphasize physical and personal attributes. Male-associated words remained centered on academic and professional themes, although some informal or negative terminology did exist.

3.3 Prevalence Trends

Our replication also reproduced the temporal patterns of gendered language documented by Wu (2018). Figure 1 shows the monthly share of posts containing at least one of the top 50 predictive words for women or men from November 2016 to October 2017.



Across all months, female-associated words appeared nearly twice as frequently as male-associated words. Seasonal peaks around job market periods, particularly in late 2016 and early 2017, indicate heightened use of gendered language during periods of increased competition.

3.4 Summary of Replication Findings

Despite minor technical adjustments, the replication consistently matches Wu's (2018) original results. This supports the robustness and replicability of the main empirical findings.

4. Methodological Extension: Random Forest Analysis

This section introduces a Random Forest classifier (Breiman, 2001) as an extension to assess whether the gendered language patterns identified by Wu (2018) persist under an alternative modeling approach. Random Forest is a learning method that aggregates predictions from multiple decision trees and measures feature importance based on each variable's contribution to classification accuracy. Unlike LASSO regression, which assumes linear relationships, Random forests capture nonlinearities and interactions between features without assuming a specific functional form.

4.1 Model Implementation

We trained the Random Forest model using the identical text features, training subset, and preprocessing pipeline employed in the LASSO model. This ensures that any observed differences in results stem from the model rather than variations in data preparation. The classifier was configured with 300 trees. To maintain interpretability while controlling overfitting, we limited the maximum depth of each tree to 20.

4.2 Comparative Framework

To facilitate meaningful comparison with Wu's original findings, we implemented a two-

stage analytical procedure. First, each word was assigned to a gender category based on the sign of its coefficient from the replicated LASSO model. Second, within each gender group, words were ranked according to their importance scores derived from the Random Forest model. This hybrid approach preserves the directional interpretation established in LASSO while incorporating Random Forest's non-parametric assessment of feature relevance.

4.3 Results and Comparison

Table 3 presents the top ten words for each gender category based on Random Forest importance scores, following the classification methodology described above.

Table 3: Replication and Extension Table 3: Word Importance by Gender (Random Forest)

Female Words	RF Imp.	Male Words	RF Imp.
hot	0.0608	economics	0.0091
sex	0.0300	fat	0.0083
beautiful	0.0216	papers	0.0074
marry	0.0195	fiecking	0.0065
dating	0.0173	nobel	0.0058
kids	0.0171	data	0.0057
married	0.0165	people	0.0054
paper	0.0132	children	0.0053
macro	0.0115	economic	0.0052
ugly	0.0113	read	0.0049

The findings largely align with the LASSO method, though some differences exist. In the Random Forest model, words associated with female authors included academic words such as “paper” which did not appear in the LASSO method. In contrast, terms associated with male authors exhibited greater lexical diversity and contained both professional and informal words. Furthermore, the distribution of feature importance scores was more uniform than the marginal effects estimated by the LASSO method. This indicates that the Random Forest model exhibited lower dependence on a small number of predictor variables.

4.4 Discussion

The outputs of the Random Forest and LASSO models are similar. This corroborates the findings of Wu (2018). Both models indicate that people use more personalizing words when discussing women on the forum, while employing more professionalizing words when discussing men on the forum.

Random Forest proves more effective than LASSO in capturing nonlinear word relationships. This potentially explains why academic terms such as “paper” stand out more prominently among female-related words in the Random Forest results. However, LASSO offers clearer interpretability through signed coefficients, whereas Random Forest importance scores lack directional meaning. The results from different methods are generally consistent, yet they also reflect the trade-off between interpretability (LASSO) and flexibility (Random Forest) in text analysis.

5. Conclusion

This study replicates and extends Wu (2018). This confirms the gender asymmetry observed in EJMR forum discussions: discussions about women focus more on personal and physical attributes, while discussions about men emphasize more on professional and academic achievements. Despite certain technical challenges, our replication results show high consistency with the original findings.

The extension using a random forest model further validated these findings. It indicates that results remain consistent across different modeling frameworks. This suggests the observed pattern is a robust feature of the forum's culture rather than a methodological artifact.

References

Wu, A. H. (2018). Gendered language on the Economics Job Market Rumors forum. AEA Papers and Proceedings, 108, 175–179. https://doi.org/10.1257/pandp.20181101

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324