

Replication of Wu (2018): Gendered Language on the Economics Job Market Rumors Forum

Alejandro De La Torre*

February 13, 2026

1 Overview

Wu [2018] studies gendered language in anonymous posts on the Economics Job Market Rumors (EJMR) forum. The paper frames the exercise as predictive: given a post’s word usage, which terms are most informative about whether the post refers to a woman or a man? The headline takeaway is qualitative rather than purely statistical—female-referenced posts are disproportionately associated with words about appearance and personal traits, while male-referenced posts are more associated with professional roles and status.

This report does three things to meet the assignment requirements. First, I locate and run the author-provided replication package and reproduce Table 1, Table 2, and Figure 1 from the published article. Second, I document the (minor) engineering work needed to run the legacy code on a modern Python/R stack, keeping the original analysis logic unchanged. Third, as a modest extension, I re-estimate the “Table 1” idea with two alternative prediction models (a linear probability model and a random forest) using the same underlying feature matrix and train/test split, and I compare how the resulting “most predictive words” lists change across models.

2 Data and replication package

I use the OpenICPSR replication package linked from the paper and its online appendix [Wu, 2019, Wu]. The key inputs are (i) `gendered_posts.csv` with post-level labels and split indicators, (ii) `X_word_count.npz`, a sparse word-count matrix for the top 10,000 words, (iii) `vocab10K.csv` with the vocabulary and excluded terms, and (iv) `trend_stats.csv` for Figure 1. These files define the same design matrix, labels, and train/test split used by the original scripts.

3 Replication procedure

I built a single-command pipeline that runs the upstream Python Lasso scripts and the upstream R script for tables and figures:

```
python src/run_all.py
```

The pipeline executes the two Lasso-logit programs in the raw package directory so relative paths resolve, then runs `tables-figures.R` to produce Table 1, Table 2, and Figure 1. All runs were

*Replication repository: <https://github.com/adelatorre2/wu2018-gendered-language-replication>.

executed on macOS arm64 in a conda environment (wu2018) with Python 3.11 and R + ggplot2. Logs are saved in `output/logs/`.

4 Obstacles and fixes

Two compatibility issues required minor, non-substantive patches to the upstream code. First, deprecated pandas `as_matrix()` calls were replaced with `to_numpy()`. Second, NumPy now blocks loading pickled objects in `np.load` by default, so I added `allow_pickle=True` for the sparse word-count matrix. These changes restore compatibility but do not alter analysis logic.

5 Replication results

Table 1, Table 2, and Figure 1 reproduce the original results in Wu [2018]. The word lists and trends match the published paper up to rounding, which is expected because I use the same data, preprocessing, splits, and upstream hyperparameters. The qualitative pattern is the same: female-associated words emphasize appearance or personal traits, while male-associated words emphasize professional roles.

Most female		Most male	
Word	ME	Word	ME
Hotter	0.422	Homo	-0.303
Pregnant	0.323	Testosterone	-0.195
Plow	0.277	Chapters	-0.189
Marry	0.275	Satisfaction	-0.187
Hot	0.271	Fieckers	-0.181
Marrying	0.260	Macroeconomics	-0.180
Pregnancy	0.254	Cuny	-0.180
Attractive	0.245	Thrust	-0.169
Beautiful	0.240	Nk	-0.165
Breast	0.227	Macro	-0.163

Table 1: Top 10 words most predictive of female- and male-referenced posts (full sample).

Most female		Most male	
Word	ME	Word	ME
Pregnancy	0.292	Knocking	-0.329
Hotter	0.289	Testosterone	-0.204
Pregnant	0.258	Blog	-0.183
Hp	0.238	Hateukbro	-0.176
Vagina	0.228	Adviser	-0.175
Breast	0.220	Hero	-0.174
Plow	0.219	Cuny	-0.173
Shopping	0.207	Handsome	-0.166
Marry	0.207	Mod	-0.166
Gorgeous	0.201	Homo	-0.160

Table 2: Top 10 words most predictive of female- and male-referenced posts (pronoun sample).

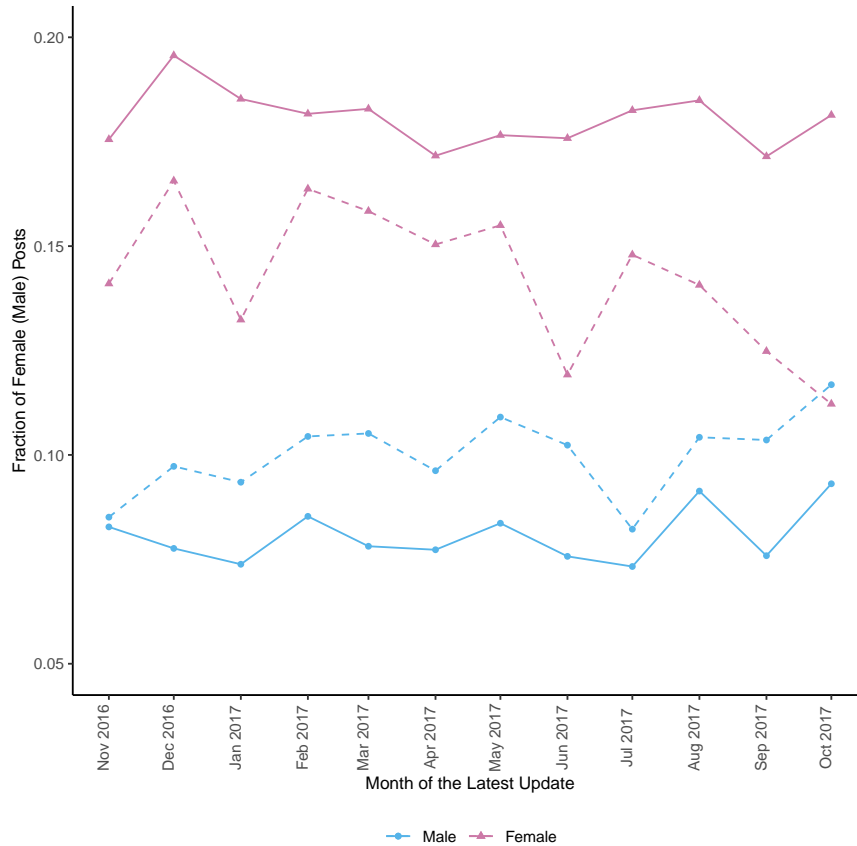


Figure 1: Fraction of female and male posts containing any of the top 50 gendered words.

6 Modest extension: alternative prediction models

To satisfy the assignment extension, I re-estimate the “Table 1 idea” using two alternative models: OLS/LPM and a Random Forest classifier. I intentionally reuse the same upstream dataset,

preprocessing, and split logic as the replication package: the same design matrix X , the same female indicator, and the same training/test split defined by the `training` column. This keeps the comparison apples-to-apples with Wu [2018]. Posts with both female and male classifiers (where `training` is NA) are excluded, matching the upstream scripts. These are predictive associations in text, not causal effects; labels are classifier-driven and therefore imperfect.

6.1 OLS / Linear Probability Model (LPM)

The LPM is a linear baseline with easily interpretable coefficients but a misspecified probability model: predictions can fall outside $[0, 1]$. I clip predictions to $[0, 1]$ for threshold-based metrics and report the raw range (min = -16.04 , max = 7.41) to show why clipping is needed. On the held-out test set, the LPM achieves accuracy = 0.792, precision = 0.669, recall = 0.157, F1 = 0.255, and ROC-AUC = 0.734 (AUC computed on raw scores). Table 3 reports the most predictive words under this model.

Most female (OLS)	Score	Most male (OLS)	Score
hotter	0.4885	thrust	-0.3924
plow	0.3907	pros	-0.2263
pregnant	0.3830	testosterone	-0.2155
marrying	0.3626	homo	-0.2065
marry	0.3389	belonging	-0.2041
hot	0.3253	satisfaction	-0.1971
misogynistic	0.3166	ante	-0.1969
breast	0.3068	drown	-0.1908
q3	0.3017	theatre	-0.1893
whore	0.2954	muscles	-0.1872

Table 3: Top 10 words most predictive of female- and male-referenced posts under OLS/LPM (extension).

6.2 Random Forest classifier

The Random Forest (RF) allows nonlinear interactions that the LPM cannot capture but is less interpretable and more computationally expensive on sparse text. To keep the extension feasible, I restrict the feature set to the top 1,000 terms by document frequency and subsample to 20,000 training and 10,000 test observations (random seed 123). With these settings, RF achieves accuracy = 0.783, precision = 0.571, recall = 0.117, F1 = 0.194, and ROC-AUC = 0.661. Feature importances are Gini-based and biased toward frequent terms, so I use a simple directional sign (mean difference by gender) to label “female” versus “male” words for a Table 1-style list.

Most female (RF)	Score	Most male (RF)	Score
you	0.0101	.	0.0232
hot	0.0100	the	0.0158
your	0.0088	,	0.0113
i	0.0086	is	0.0111
my	0.0060	a	0.0106
with	0.0057	to	0.0103
!	0.0055	?	0.0099
sex	0.0053	and	0.0097
date	0.0043	of	0.0096
if	0.0042	in	0.0089

Table 4: Top 10 words most predictive of female- and male-referenced posts under Random Forest (extension).

Comparison. Relative to Lasso-logit, the alternative models trade sparse regularization for either a simple linear baseline (LPM) or flexible nonlinear interactions (RF). The qualitative lists remain broadly similar in theme, but the RF rankings are more sensitive to feature frequency and sampling choices, which is expected for this model class.

7 Conclusion

The replication reproduces the main tables and figure from Wu [2018] using the OpenICPSR package and a single-command pipeline. I also implement two alternative models (OLS/LPM and Random Forest) on the same design matrix and split, illustrating the tradeoff between interpretability and flexibility. Overall, the extension confirms that the core descriptive pattern is robust across modeling choices, while reinforcing that these are predictive associations rather than causal claims.

References

- Alice H Wu. Online Appendix to “Gendered Language on the Economics Job Market Rumors Forum”.
- Alice H. Wu. Gendered Language on the Economics Job Market Rumors Forum. *AEA Papers and Proceedings*, 108:175–179, May 2018. ISSN 2574-0768, 2574-0776. doi: 10.1257/pandp.20181101.
- Alice H. Wu. Replication data for: Gendered Language on the Economics Job Market Rumors Forum, October 2019.