

Supplemental Materials

Gendered behavior as a disadvantage in open source software development

Balazs Vedres¹, Orsolya Vasarhelyi¹

¹ Department of Network and Data Science, Central European University,
Nador u. 9., 1051. Budapest, Hungary

	(1)		(2)		Dependent Variable:				(3)		(4)		(5)	
	Controls		Controls + Languages		Success (number of stars)				Controls + DT classes		5% gender swaps		10% gender swaps	
	Zero-inflation model	Count model	Zero-inflation model	Count model	Zero-inflation model	Count model	Sign.	Coef.	Sign.	Coef.	Count model	Sign.	Coef.	Count model
Female	Coef. 0.666 (0.357)	Sign. 0.618 * (0.314)	Coef. 0.766 * (0.350)	Sign. 0.704 * (0.309)	Coef. 0.517 (0.375)	Sign. 0.424 (0.308)			Coef. 0.645 (0.375)	Sign. 0.424 (0.308)	Coef. 0.645 (0.375)	Sign. 0.424 (0.308)	Coef. 0.526 (0.375)	Sign. 0.424 (0.308)
Female:ness	1.492 *** (0.404)	-4.123 *** (0.273)	1.326 ** (0.406)	-3.108 *** (0.297)	0.996 * (0.496)	-4.132 *** (0.381)			-3.416 *** (0.496)	-4.132 *** (0.381)	-3.416 *** (0.496)	-4.132 *** (0.381)	-3.741 *** (0.496)	-5.027, -1.663
Female:Female:ness	1.031 * (0.484)	1.503 *** (0.403)	0.599 (0.480)	0.823 * (0.418)	1.274 * (0.513)	1.258 ** (0.409)			0.976 (0.513)	1.258 ** (0.409)	0.976 (0.513)	1.258 ** (0.409)	1.149 (0.513)	-2.433, 3.472
Name frequency	-0.048 (0.279)	0.034 (0.191)	-0.186 (0.273)	-0.293 (0.192)	0.172 (0.288)	0.290 (0.203)			0.111 (0.288)	0.290 (0.203)	0.111 (0.288)	0.290 (0.203)	0.134 (0.288)	-0.665, 1.054
Female:Name frequency	-0.966 * (0.394)	-1.841 *** (0.341)	-0.918 * (0.390)	-0.061 *** (0.014)	-1.013 * (0.410)	-1.579 *** (0.340)			-1.651 *** (0.410)	-1.579 *** (0.340)	-1.651 *** (0.410)	-1.579 *** (0.340)	-1.573 (0.410)	-3.241, 0.826
Followers (log)	-0.678 *** (0.032)	0.669 (0.016)	-0.651 *** (0.032)	0.718 *** (0.018)	-0.610 *** (0.032)	0.735 *** (0.017)			0.624 (0.032)	0.735 *** (0.017)	0.624 (0.032)	0.735 *** (0.017)	0.630 (0.032)	0.598, 0.670
Tenure	-0.903 *** (0.050)	-0.502 *** (0.055)	-0.850 *** (0.051)	-0.418 *** (0.053)	-0.911 *** (0.051)	-0.397 *** (0.056)			-0.318 *** (0.051)	-0.397 *** (0.056)	-0.318 *** (0.051)	-0.397 *** (0.056)	-0.311 (0.056)	-0.407, -0.184
No of own repositories (log)	-0.119 (0.068)	0.327 *** (0.076)	-0.290 ** (0.093)	0.097 (0.078)	-0.407 *** (0.097)	0.173 * (0.076)			0.347 (0.097)	0.173 * (0.076)	0.347 (0.097)	0.173 * (0.076)	0.333 (0.097)	0.218, 0.429
No of touched repositories (log)	-0.590 *** (0.134)	-0.287 ** (0.110)	-0.628 *** (0.143)	-0.044 (0.117)	-0.227 (0.138)	-0.421 *** (0.110)			-0.227 *** (0.138)	-0.421 *** (0.110)	-0.227 *** (0.138)	-0.421 *** (0.110)	-0.205 (0.138)	-0.360, -0.064
No of collaborators (log)	-0.349 *** (0.068)	0.393 *** (0.060)	-0.377 *** (0.069)	0.244 *** (0.062)	-0.156 * (0.073)	0.353 *** (0.063)			0.354 (0.073)	0.353 *** (0.063)	0.354 (0.073)	0.353 *** (0.063)	0.334 (0.073)	0.244, 0.428
Potential bookmarkers	0.140 ** (0.046)	0.021 (0.030)	0.059 (0.046)	-0.031 (0.034)	0.127 ** (0.047)	0.004 (0.032)			0.054 (0.047)	0.004 (0.032)	0.054 (0.047)	0.004 (0.032)	0.053 (0.047)	0.022, 0.079
Intercept	4.706 *** (0.453)	2.658 *** (0.355)	5.818 *** (0.499)	2.449 *** (0.401)	4.434 *** (0.559)	2.620 *** (0.471)			1.702 (0.559)	2.620 *** (0.471)	1.702 (0.559)	2.620 *** (0.471)	1.825 (0.559)	0.673, 2.881
Observations	20000		20000		20000				20000		20000		20000	
Languages included	N_o	Y_{es}	N_o	Y_{es}	N_o	Y_{es}			N_o		N_o		N_o	
DT classes included	N_o		N_o		N_o				N_o		N_o		N_o	

Note: *p<0.05; **p<0.01; ***p<0.001

Zero-inflated negative binomial models

SI. 1: Zero-inflated negative binomial models of success for men and women

	Dependent Variable:									
	(1)		(2)		(3)		(4)		(5)	
	Controls		Controls + Languages		Controls + DT classes		5% gender swaps		10% gender swaps	
	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	95% CI	Coef.	95% CI
Female	-0.041 (0.027)		-0.051 (0.027)		-0.059 (0.027)	*	0.645 (0.646, 1.774)		0.526 (-1.379, 2.051)	
Femaleness	-0.367 (0.029)	***	-0.295 (0.030)	***	-0.343 (0.032)	***	-3.416 (-2.771, 2.705)	-4.432, -1.724	-3.741 (-2.433, 3.472)	-5.027, -1.663
Female:Femaleness	0.060 (0.033)		0.075 (0.033)	*	0.089 (0.033)	**	0.976 (0.111, 0.568)	-2.771, 2.705	1.149 (0.134, -0.665, 1.054)	-2.433, 3.472
Name frequency	0.014 (0.022)		0.012 (0.022)		0.010 (0.022)		0.111 (-1.651, -2.722, 0.053)	-0.481, 0.568	0.134 (-1.573, -3.241, 0.826)	-0.665, 1.054
Female:Name frequency	-0.037 (0.028)		-0.004 (0.028)	***	0.089 (0.028)		-1.651 (0.028)	-2.722, 0.053	-1.573 (0.630, 0.598, 0.670)	-3.241, 0.826
Followers (log)	0.131 (0.004)	***	0.125 (0.004)	***	0.130 (0.004)	***	0.624 (-0.318, -0.372, -0.266)	0.596, 0.663	0.630 (-0.311, -0.407, -0.184)	0.598, 0.670
Tenure	0.003 (0.003)		0.004 (0.003)		0.005 (0.003)		-0.318 (0.003)	-0.372, -0.266	-0.311 (0.333, 0.218, 0.429)	-0.407, -0.184
No of own repositories (log)	0.019 (0.004)	***	0.013 (0.004)	**	0.020 (0.004)	***	0.347 (0.241, 0.412)	0.241, 0.412	0.333 (-0.205, -0.360, -0.064)	0.218, 0.429
No of touched repositories (log)	0.085 (0.007)	***	0.085 (0.008)	***	0.055 (0.007)	***	-0.227 (0.032, 0.038)	-0.330, -0.113	-0.205 (0.334, 0.244, 0.428)	-0.360, -0.064
No of collaborators (log)	0.034 (0.004)	***	0.028 (0.004)	***	0.032 (0.004)	***	0.354 (0.054, 0.053)	0.280, 0.438	0.334 (0.053, 0.022, 0.079)	0.244, 0.428
Potential bookmarks	-0.011 (0.003)	***	-0.017 (0.003)	***	-0.009 (0.003)	**	0.054 (0.834, 2.417)	0.026, 0.077	0.053 (1.825, 0.673, 2.881)	0.022, 0.079
Intercept	-0.174 (0.031)	***	-0.157 (0.034)	***	-0.091 (0.033)	**	1.702 (20000, 20000)	0.834, 2.417	1.825 (20000, 20000)	0.673, 2.881
Observations	20000		20000		20000		20000		20000	
Adjusted R2	0.298		0.312		0.311					
Languages included	No		Yes		No		No		No	
DT classes included	No		No		Yes		No		No	

Note: *p<0.05; **p<0.01; ***p<0.001
OLS models

SI. 2: OLS models of log(success+1)

	Dependent Variable:									
	(1)		(2)		(3)		(4)		(5)	
	Controls		Controls + Languages		Controls + DT classes		5% gender swaps		10% gender swaps	
	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	95% CI	Coef.	95% CI
Female	-0.286 (0.256)		-0.173 (0.259)		0.009 (0.257)		-0.380 (0.257)	-0.633, -0.089	-0.346 (0.257)	-0.650, -0.047
Femaleness	-2.934 (0.290)	***	-2.769 (0.297)	***	-2.592 (0.331)	***	-2.748 (0.331)	-3.150, -2.445	-3.021 (0.331)	-3.43, -2.623
Female:Femaleness	0.010 (0.326)		-0.156 (0.331)		-0.409 (0.269)		0.234 (0.269)	-0.112, 0.627	0.236 (0.269)	-0.167, 0.789
Name frequency	0.072 (0.213)		0.076 (0.214)		0.107 (0.208)		0.109 (0.208)	-0.013, 0.279	0.124 (0.208)	-0.025, 0.273
Female:Name frequency	-0.323 (0.274)		0.128 (0.016)	***	-0.409 (0.269)		-0.329 (0.269)	-0.610, -0.115	-0.295 (0.269)	-0.544, -0.050
Followers (log)	0.219 (0.024)	***	0.210 (0.024)	***	0.207 (0.024)	***	0.219 (0.024)	0.215, 0.224	0.221 (0.024)	0.214, 0.228
Tenure	-0.635 (0.033)	***	-0.571 (0.034)	***	-0.615 (0.034)	***	-0.465 (0.034)	-0.472, -0.456	-0.454 (0.034)	-0.467, -0.445
No of own repositories (log)	-0.061 (0.046)		-0.091 (0.046)	*	0.048 (0.046)		-0.082 (0.046)	-0.094, -0.068	-0.088 (0.046)	-0.104, -0.069
No of touched repositories (log)	0.205 (0.074)	**	0.177 (0.082)	*	0.140 (0.074)		0.245 (0.074)	0.219, 0.273	0.257 (0.074)	0.226, 0.290
No of collaborators (log)	0.420 (0.041)	***	0.432 (0.042)	***	0.218 (0.045)	***	0.341 (0.045)	0.322, 0.356	0.327 (0.045)	0.305, 0.351
Potential bookmarks	0.472 (0.019)	***	0.421 (0.021)	***	-0.391 (0.337)		0.463 (0.337)	0.453, 0.473	0.460 (0.337)	0.447, 0.474
Intercept	1.177 (0.297)	***	1.118 (0.318)	***	3.842 (0.329)	***	0.778 (0.329)	0.512, 1.021	0.867 (0.329)	0.601, 1.146
Observations	20000		20000		20000		20000		20000	
Languages included	No		Yes		No		No		No	
DT classes included	No		No		Yes		No		No	

Note: *p<0.05; **p<0.01; ***p<0.001
Logit models

SI. 3: Logit models of survival

	(1) Controls				(2) Controls + Languages				(3) Controls + DT classes				(4) Controls, Users without name			
	Zero-inflation model		Count model		Zero-inflation model		Count model		Zero-inflation model		Count model		Zero-inflation model		Count model	
	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.
Femaleness	6.113 (1.076)	***	-4.329 (0.897)	***	5.314 (1.157)	***	-4.701 (0.905)	***	2.907 (1.355)	*	-5.805 (1.165)	***	7.930 (1.360)	***	-0.516 (1.363)	
Followers (log)	-0.763 (0.054)	***	0.575 (0.027)	***	-0.747 (0.058)	***	0.635 (0.032)	***	-0.624 (0.054)	***	0.651 (0.031)	***	-0.858 (0.074)	***	0.544 (0.037)	***
Tenure	-1.285 (0.070)	***	-0.846 (0.083)	***	-1.229 (0.075)	***	-0.752 (0.081)	***	-1.272 (0.076)	***	-0.617 (0.086)	***	-1.344 (0.089)	***	-0.966 (0.119)	***
No of own repositories (log)	0.023 (0.092)		0.181 (0.132)		-0.053 (0.095)		0.196 (0.143)		-0.427 (0.109)	**	0.029 (0.134)		0.270 (0.114)		0.520 (0.217)	*
No of touched repositories (log)	-0.420 (0.184)	*	0.835 (0.172)	***	-0.258 (0.207)		0.830 (0.180)	***	0.182 (0.191)		0.603 (0.174)	***	-0.888 (0.271)	**	0.256 (0.270)	
No of collaborators (log)	-0.455 (0.092)	***	-0.035 (0.091)		-0.425 (0.095)	***	0.029 (0.093)		-0.448 (0.109)	***	-0.229 (0.088)	**	-0.234 (0.114)	*	0.430 (0.132)	**
Potential bookmarks	0.199 (0.027)	***	-0.221 (0.023)	***	0.214 (0.034)	***	-0.245 (0.028)	***	0.148 (0.032)	***	-0.243 (0.028)	***	0.272 (0.033)	***	-0.173 (0.028)	***
Intercept	1.773 (0.753)	*	3.305 (0.656)	***	1.783 (0.845)	*	3.143 (0.667)	***	3.275 (1.020)	**	4.196 (0.867)	***	1.090 (0.945)	***	1.369 (0.988)	
Observations	10000				10000				10000				7231			
Languages included	No				Yes				No				No			
DT classes included	No				No				Yes				No			

Note: *p<0.05; **p<0.01; ***p<0.001

Zero-inflated negative binomial models

SI. 4: Zero-inflated negative binomial models of success for users with unknown gender

	Dependent Variable:							
	(1)		(2)		(3)		(4)	
	Controls		Controls + Languages		Controls + DT classes		Controls, Users without name	
	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.
Femaleness	-0.584 (0.067)	***	-0.535 (0.072)	***	-0.599 (0.075)	***	-0.479 (0.070)	***
Followers (log)	0.124 (0.006)	***	0.119 (0.006)	***	0.122 (0.006)	***	0.114 (0.008)	***
Tenure	0.028 (0.003)	***	0.024 (0.003)	***	0.027 (0.003)	***	0.026 (0.003)	***
No of own repositories (log)	0.012 (0.005)	*	0.014 (0.005)	**	0.019 (0.005)	***	0.013 (0.005)	**
No of touched repositories (log)	0.104 (0.009)	***	0.091 (0.010)	***	0.069 (0.010)	***	0.085 (0.010)	***
No of collaborators (log)	0.029 (0.005)	***	0.028 (0.005)	***	0.022 (0.006)	***	0.028 (0.005)	***
Potential bookmarks	-0.023 (0.002)	***	-0.023 (0.002)	***	-0.022 (0.002)	***	-0.021 (0.002)	***
Intercept	-0.002 (0.054)		0.012 (0.056)		0.103 (0.061)		-0.004 (0.058)	
Observations	10000		10000		10000		7231	
Languages included	No		Yes		No		No	
DT classes included	No		No		Yes		No	

Note: Heteroscedasticity-robust SEs are in parentheses.

p<0.05; **p<0.01; ***p<0.001

OLS models

SI. 5: OLS models of log(success+1) for users with unknown gender

	Dependent Variable:							
	(1)		(2)		(3)		(4)	
	Controls		Controls + Languages		Controls + DT classes		Controls, Users without name	
	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.	Coef.	Sign.
Femaleness	-2.987 (0.639)	***	-2.110 (0.697)	**	-1.684 (0.749)	*	-3.341 (0.707)	***
Followers (log)	0.446 (0.038)	***	0.410 (0.039)	***	0.363 (0.039)	***	0.416 (0.045)	***
Tenure	-0.486 (0.035)	***	-0.474 (0.036)	***	-0.515 (0.036)	***	-0.487 (0.039)	***
No of own repositories (log)	-0.015 (0.056)		-0.020 (0.057)		0.045 (0.060)		-0.069 (0.062)	
No of touched repositories (log)	0.221 (0.092)	*	0.115 (0.098)		0.066 (0.098)		0.261 (0.103)	*
No of collaborators (log)	0.209 (0.049)	***	0.206 (0.050)	***	0.197 (0.059)	***	0.180 (0.054)	***
Potential bookmarks	0.649 (0.017)	***	0.583 (0.020)	***	0.651 (0.020)	***	0.637 (0.018)	***
Intercept	-0.323 (0.477)		-0.451 (0.513)		-0.678 (0.572)		-0.079 (0.529)	
Observations	10000		10000		10000		7231	
Adjusted R2	0.265		0.278		0.282		0.227	
Languages included	No		Yes		No		No	
DT classes included	No		No		Yes		No	

Note: *p<0.05; **p<0.01; ***p<0.001

Logit models

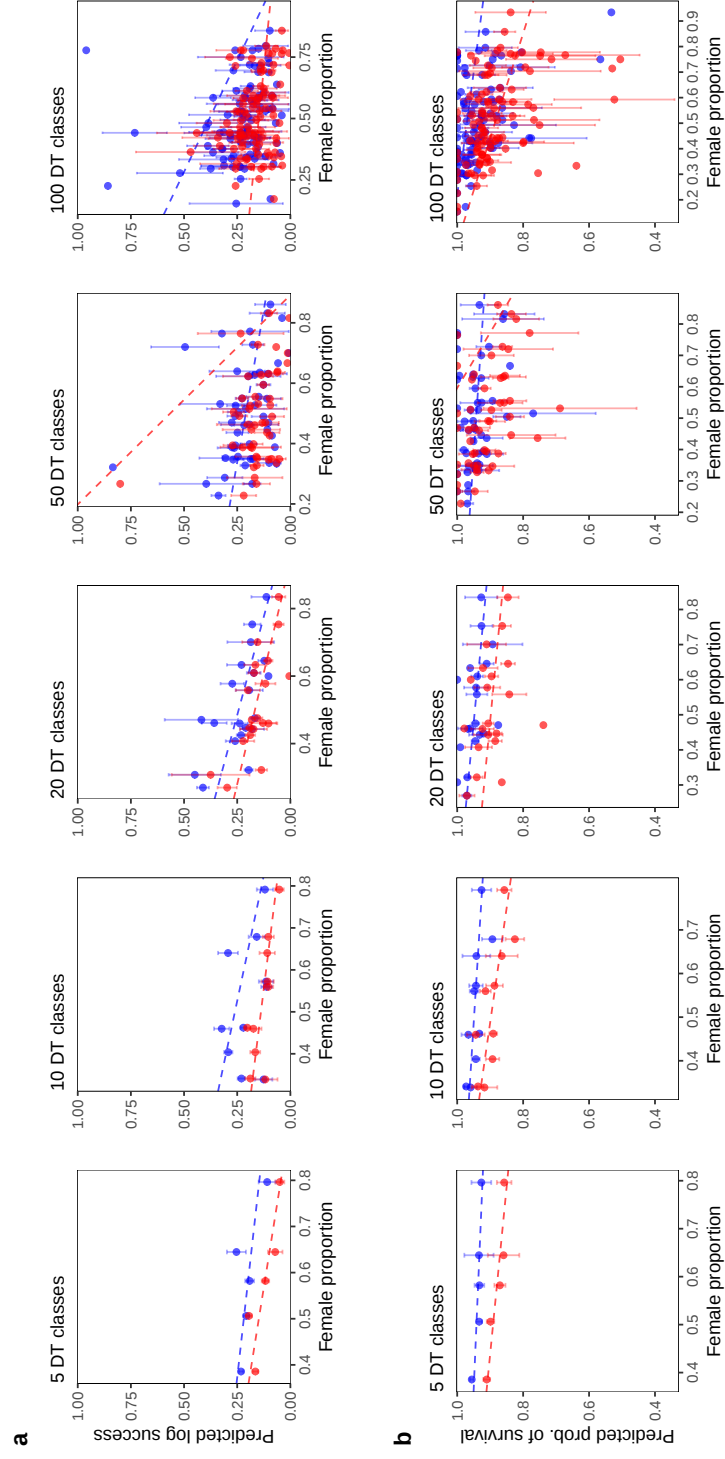
SI. 6: Logit models of survival for users with unknown gender

Dependent Variable:
Success (count)

	Zero-inflation model		Count model	
	Coef.	Sign.	Coef.	Sign.
Female	0.956 (0.382)	*	1.030 (0.327)	**
Femaleness	2.578 (0.457)	***	-3.313 (0.271)	***
Female:Femaleness	1.096 (0.565)		1.624 (0.433)	***
2015-16	2.886 (0.616)	***	4.364 (0.742)	***
Female:2015-16	-0.862 (0.773)		-1.668 (0.914)	
Femaleness:2015-16	-2.561 (1.212)	*	-6.854 (1.342)	***
Female:Femaleness:2015-16	0.742 (1.438)		2.082 (1.597)	
Name frequency	0.072 (0.294)		0.247 (0.204)	
Female:Name frequency	-1.271 (0.403)	**	-2.434 (0.353)	***
Followers (log)	2.886 (0.616)	***	4.364 (0.742)	
Tenure	-0.716 (0.032)	***	0.649 (0.016)	***
No of own repositories (log)	-0.077 (0.106)		0.364 (0.079)	***
No of touched repositories (log)	-0.603 (0.146)	***	-0.283 (0.115)	*
No of collaborators (log)	-0.394 (0.072)	***	0.360 (0.063)	***
Potential bookmarks	0.149 (0.049)	**	0.046 (0.030)	
Intercept	1.681 (0.468)	***	0.627 (0.331)	
Observations	20000		20000	
Languages included	No		No	
DT classes included	No		No	

Note: *p<0.05; **p<0.01; ***p<0.001

Zero-inflated negative binomial models



SI. 8: OLS model of $\log(\text{success}+1)$ and Logit model of survival with classes of gendered behavior

Dependent Variables:
Success (log) & Survival (yes=1, no=0)

	OLS model		Logit model	
	Coef.	Sign.	Coef.	Sign.
Female	-0.021 (0.028)		-0.343 (0.209)	
DT Female Propotion	-0.279 (0.031)	***	-1.976 (0.213)	***
Name frequency (log)	0.017 (0.020)		0.103 (0.142)	
Female:Name frequency (log)	-0.042 (0.028)		0.089 (0.221)	
Female:DT Female Propotion	0.000 (0.040)		-0.143 (0.230)	
Followers (log)	0.130 (0.002)	***	0.620 (0.014)	***
Tenure	0.027 (0.004)	***	0.698 (0.030)	***
No of own repositories (log)	0.017 (0.008)	***	0.292 (0.039)	***
No of touched repositories (log)	0.081 (0.008)	***	0.052 (0.068)	
No of collaborators (log)	0.029 (0.013)	***	0.137 (0.040)	***
Potential Bookmarkers	0.0285 (0.040)	*	0.682 (0.114)	***
Intercept	-0.448 (0.063)	***	-8.278 (0.542)	***
Observations	20000		20000	
Adjusted R2 / McFadden's pseudo R2	0.293		0.269	

Note: *p<0.05; **p<0.01; ***p<0.001

SI. 9: Robustness of classes of gendered behavior