

Project Part 3: Exploratory Analysis of Collaborative Study in Genomics Revolution - Panel Model for Success with Faculty and Network Records

Md. Rafiul Amin and MD Tanim Hasan
University of Houston
Texas, USA

January 22, 2021

Contributions from Each Member

Md. Rafiul Amin

- 1) Pooled cross-sectional model - robustness check.
- 2) Regression table generation.
- 3) Critical thinking over all the generated plots and models in the report.
- 4) About 50% of the report writing.

MD Tanim Hasan

- 1) Linear modeling of career data set pooled cross-sectional model.
- 2) The plot of the cross-sectional regression model.
- 3) Critical thinking over all the generated plots and models in the report.
- 4) About 50% of the report writing.

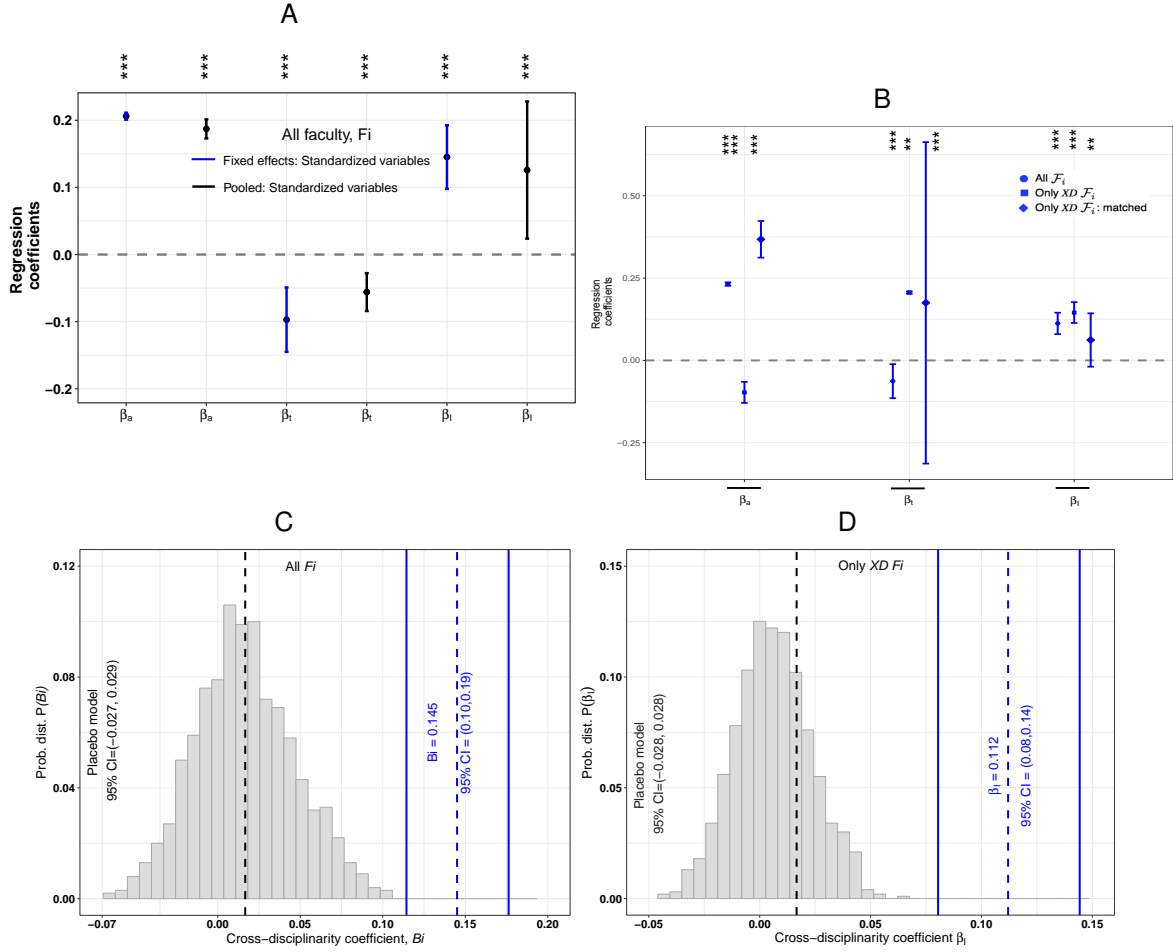


Fig. 5. Career panel regression model. The sub figure A and B reports the parameter estimates for the three principal explanatory variables included in the fixed effects \mathcal{F} career panel model defined in [1]; see table S4 for the complete list of parameter estimates. The sub-figure C and D reports the robustness check of panel regression model. To test the possibility of spurious correlations leading to the significant estimates for the cross-disciplinary variables in the panel model (table S4), we ran this model using a randomized cross-disciplinary indicator variable I_{XD} i;p , implemented by shuffling just that variable across the observations without replacement. (C) For $n = 1000$ shuffled data sets, we do not observe any (0%) coefficient estimates as large as the empirical value $\beta_i = 0.145$ corresponding to the dashed vertical blue line [solid vertical blue lines indicate 95(CI); see table S4, third column cluster]. (D) We repeated the same shuffling method for the panel model applied to only the 1247 \mathcal{F}_i classified with orientation $\mathcal{F}(\mathcal{F}_i) \equiv$, and again, we do not observe any (0%) coefficient estimates as large as the empirical value β_i reported in table S5 (third column cluster). The levels of statistical significance are as follows: $**P \leq 0.01, ***P \leq 0.001$.

See the detailed discussions in after the tables S4, S5 and S6.

Table S4. Career data set: Panel model on all Faculty \mathcal{F} : Each column represents the estimated coefficients for the panel model in which the dependent variable is the normalized citation impact of each of the articles, $z_{i,p}$ of faculty \mathcal{F}_i . The first two columns correspond to the panel regression without \mathcal{F}_i fixed effects and the last two columns correspond to the panel regression with \mathcal{F}_i fixed effects. Estimates in the second and fourth columns are calculated using standardized data, where each of the β coefficients denotes the changes in $z_{i,p}$ associated with a one standard deviation change in the respective independent variable. The model without fixed effects incorporates time-independent author-level characteristics, i.e., adding to the specification of the panel model in [1] the additional terms $[\beta_{\mathcal{C}^{PR}} \ln \mathcal{C}^{PR} + \beta_{\lambda} \ln \lambda_i + D(\mathcal{F}_i)]$. Because of this reason similar to [1] we analyzed 3,900 scholars connected within the network for which \mathcal{C}^{PR} could be estimated; note that these additional variables are absorbed into β_i in the fixed effects model. The additional connectivity variable λ_i is the fraction of the total pollinators that are ‘bridge’ pollinators. Robust standard errors are shown in parenthesis, and X denotes time-independent variables absorbed by the fixed effects model. Y indicates additional fixed effects included in the model.

	No Fixed Effects	No Fixed Effects [Standardized]	Fixed Effects	Fixed Effects [Standardized]
Publication Characteristics				
# of co-author, β_a	0.284*** (0.007)	0.187*** (0.00474)	0.312*** (0.0026)	0.206*** (0.0017)
Career Age, β_{τ}	-0.00547*** (0.000919)	-0.0559*** (0.00156)	-0.00949*** (0.0001)	-0.0971*** (0.0160)
Cross-disciplinary indicator, β_I	0.126*** (0.0341)	0.126*** (0.0341)	0.145*** (0.0158)	0.145*** (0.0158)
Network Parameters				
PageRank Centrality, $\beta_{\mathcal{C}}$	0.044*** (0.0142)	0.0284*** (0.0092)	X	X
Bridge fraction, β_{λ}	0.334*** (0.0256)	0.121*** (0.0093)	X	X
Discipline (\mathcal{F}) dummy	0.0079 (0.0139)	0.0079 (0.0139)	X	X
Constant	0.451*** (0.142)	0.156 (0.102)	-0.293**** (0.046)	-0.293 (0.046)
Year dummy	Y	Y	Y	Y
n	413,565	413,565	413,565	413,565
adj. R^2	0.055	0.055	0.036	0.036

Standard errors in parentheses,

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** denotes overall intercept

Observation and Discussion:

Based on the estimated slope parameters in the table we can see following observations:

1. β_a values for all four models are positive and the p-values are greater than 5%. This denotes that, there is a positive correlation between the number of co-authors and the number of citations, i.e., if there is increased number of authors, the publication is likely to be cited more relative to the one those have small number of citations.

2. β_{τ} values for all four models are negatively correlated and the p-values are greater than 5% meaning if the career age of the author at the time of the publication, the number of citation is likely to be less compared to the ones those has smaller career age. This result is an indication of new researchers entering the field and contributing successful publications.

3. β_I values in all of the four models are positive and the corresponding p-values are all greater than 5%. This is the evidence of the fact that, the number of citation in cross-disciplinary publication is higher than the other ones.

5. Author centrality $\beta_{\mathcal{C}}$ and bridge fraction β_{λ} also have statistically significant positive correlation. This two variables are indication of how much a faculty is connected to the network based on how important

other faculty they worked with and also how many bridge pollinator they co-authored with. This positive correlation indicates that the faculty is more connected or influential in the network.

6. Affiliation of the faculty does not have any statistically significant effect in any of the two models.

Table S5. Career data set: Panel model on the $XD_{\mathcal{F}}$ Faculty: Robustness check of panel model without and with fixed effects, implemented using only the 1,247 \mathcal{F}_i with orientation $\mathcal{O}(\mathcal{F}_i)$.

	No Fixed Effects	No Fixed Effects [Standardized]	Fixed Effects	Fixed Effects [Standardized]
Publication Characteristics				
# of co-author, β_a	0.329*** (0.0123)	0.217*** (0.0081)	0.351*** (0.0039)	0.232*** (0.0026)
Career Age, β_τ	-0.00499*** (0.00018)	-0.051*** (0.0185)	-0.006* (0.0025)	-0.0629* (0.026)
Cross-disciplinary indicator, β_I	0.109*** (0.0328)	0.109*** (0.0328)	0.112*** (0.0163)	0.112*** (0.0163)
Network Parameters				
Author Centrality, $\beta_{\mathcal{C}}$	0.0526* (0.0265)	0.034* (0.0171)	X	X
Bridge fraction, β_λ	0.319*** (0.0493)	0.1157*** (0.0178)	X	X
Discipline (\mathcal{F}) dummy	-0.0383 (0.0256)	-0.0383 (0.0256)	X	X
Constant	0.249 (0.123)	-0.0414 (0.168)	-0.409**** (0.078)	-0.071**** (0.032)
Year dummy	Y	Y	Y	Y
n	166,621	166,621	166,621	166,621
adj. R^2	0.067	0.067	0.049	0.049

Standard errors in parentheses,

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Observation and Discussion:

Based on the estimated slope parameters in the table for robustness check we again notice the following similar observations:

1. β_a values for all four models are positive and the p-values are greater than 5% again. This denotes that, there is a positive correlation between the number of co-authors and the number of citations, i.e., if there is increased number of authors, the publication is likely to be cited more relative to the one those have small number of citations even among the faculties who are XD oriented.

2. β_τ values for all four models are negatively correlated and the p-values are greater than 5% meaning if the career age of the author at the time of the publication, the number of citation is likely to be less compared to the ones those has smaller career age. This result is an indication of new researchers entering the field and contributing successful publications. The result is also consistent with the one in Table S4 among the XD oriented faculties only.

3. β_I values in all of the four models are positive and the corresponding p-values are all greater than 5%. This is the evidence of the fact that, the number of citation in cross-disciplinary publication is higher than the other ones.

5. Author centrality $\beta_{\mathcal{C}}$ and bridge fraction β_λ also have statistically significant positive correlation. This two variables are indication of how much a faculty is connected to the network based on how important other faculty they worked with and also how many bridge pollinator they co-authored with. This positive correlation indicates that the faculty is more connected or influential in the network.

6. Affiliation of the faculty does not have any statistically significant effect in any of the two models. The faculties who are XD oriented, also follows the similar pattern.

Table S6. Career data set: Panel model on the $XD_{\mathcal{F}}$ Faculty with matched pairs: Robustness check of panel model without and with fixed effects, implemented using only the 62 \mathcal{F}_i with orientation $\mathcal{O}(\mathcal{F}_i) = XD_{\mathcal{F}_i}$ have at least 10 matched pairs of publications. Where possible, we matched each p with $I_{i,p}^{XD} = 1$ with a publication with $I_{i,p}^{XD} = 0$ from the same \mathcal{F}_i , having published within two years from each other, and featuring number of co-authors a_p at do not differ more than 20%.

	No Fixed Effects	No Fixed Effects [Standardized]	Fixed Effects	Fixed Effects [Standardized]
Publication Characteristics				
# of co-author, β_a	0.552*** (0.033)	0.407*** (0.024)	0.499*** (0.038)	0.368*** (0.028)
Career Age, β_τ	-0.0056* (0.0026)	-0.058* (0.027)	0.017 (0.024)	0.175 (0.244)
Cross-disciplinary indicator, β_I	0.054 (0.043)	0.054 (0.043)	0.062 (0.041)	0.062 (0.041)
Network Parameters				
Author Centrality, β_ℓ	0.194*** (0.039)	0.128*** (0.026)	X	X
Bridge fraction, β_λ	0.564*** (0.098)	0.159*** (0.028)	X	X
Discipline (\mathcal{F}) dummy	-0.266*** (0.047)	-0.266*** (0.046)	X	X
Constant	0.341 (0.786)	-0.518 (0.718)	-1.291**** (0.684)	0.0127**** (0.167)
Year dummy	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
<i>n</i>	2218	2218	2218	2218
adj. R^2	0.229	0.229	0.096	0.096

Standard errors in parentheses,

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** denotes that the value corresponds to the overall intercept.

Observation and Discussion:

The Table S6 represents the XD_F panel model with matched pairs. Based on the criteria in the paper [1] we could select 62 faculties who have at least 10 matched pairs, i.e., 2218 number of observations. We ran the panel model these observations.

Again the model parameters follow the patterns similar to the Table S4-S5 and [1].

The numeric values does not exactly match with the one in [1].

References

- [1] Alexander M Petersen, Dinesh Majeti, Kyeongan Kwon, Mohammed E Ahmed, and Ioannis Pavlidis. Cross-disciplinary evolution of the genomics revolution. *Science advances*, 4(8):eaat4211, 2018.