Online Appendix

for

Detrimental Collaborations in Creative Work: Evidence from Economics

Keyvan Vakili
London Business School
kvakili@london.edu

Florenta Teodoridis
USC Marshall School of Business
florenta.teodoridis@marshall.usc.edu

Michaël Bikard
INSEAD
michael.bikard@insead.edu

Estimating Collaboration Credit Premium

To estimate the level of credit allocation for each instance of collaboration in our sample, we use the method developed by Bikard et al. (2015). The method assumes that (1) scientists, on average, choose a level of collaboration that is expected to maximize their allocated credit; and (2) the credit allocated to each author on a paper amounts to a fraction of the citations to that paper. The latter assumption simply suggests that the credit for the paper must be shared among collaborators, albeit in a way that might sum up to more or less than 100%.

Following Bikard et al. (2015), we define $\alpha(N)$ as the average fraction of citations attributed to each author listed on a paper, where N is the total number of authors. In the presence of credit premium, $\alpha(N)$ would be larger than $\frac{1}{N}$, and in the presence of credit penalty $\alpha(N)$ would be smaller than $\frac{1}{N}$. In other words, instead of authors each on average obtaining one Nth of the citations to a paper as their credit, they could receive $\alpha(N) > \frac{1}{N}$ of the citations (credit premium) or $\alpha(N) < \frac{1}{N}$ (credit penalty). Thus, to test for the presence of average credit premium in our sample, we need to identify $\alpha(N)$ and compare its value with $\frac{1}{N}$. To do so, we again rely on the Bikard et al. (2015) arguments. Their approach observes that if the presumed level of $\alpha(N)$ reflects the true value of credit allocation in the sample, regressing the calculated fractional credit for each author of a paper on the paper's number of authors should produce an effect of zero. This intuition is driven by the first assumption behind the method—that individuals, on average, are aware of the level of collaboration that maximizes the credit attributed to them. Thus, if individuals optimize, we should see neither a positive nor a negative effect at the population level. A positive effect means that higher levels of collaboration increase the allocated credit, in which case the optimal response is to collaborate more. A negative effect means that the level of collaboration has a negative impact on the allocated credit, in which case the optimal response is to collaborate less. Note that the method does not require every researcher to have a precise estimate of their level of credit allocation. Rather, it requires that, on average, researchers understand whether increasing or decreasing the number of collaborators would increase or decrease their individual credit.

Based on this intuition, we can then try different levels of $\alpha(N)$ to construct the individual credit as the dependent variable and check which one generates an estimated effect of zero for the number of collaborators in the said regression. Including individual, department, and institution-year fixed effects in the regression further controls for idiosyncratic characteristics of individuals, their departments, and their institutions that could affect their collaborative choices. We use the following regression:

$$\ln(1 + Frac\_Credit_{it}) = \beta_0 + \beta_1 NAuthors_{it} + \mu_i + \varphi_{it} + \gamma_{it}.\delta_t + \varepsilon_{it}$$

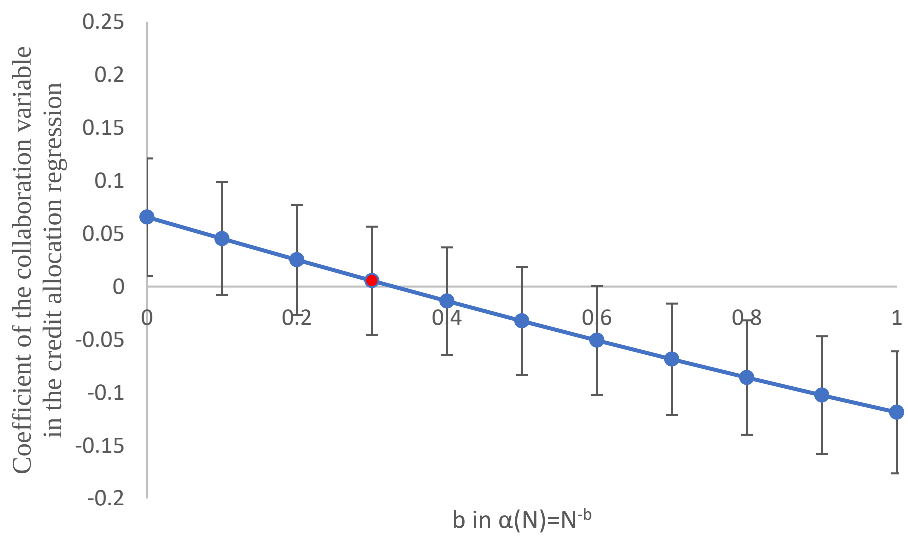
where the dependent variable is the natural log of 1 plus the sum of fractional credits for papers produced by individual i in year t. The i's fractional credit on each paper is calculated as $\alpha(N)$ multiplied by the number of citations received by the paper. Following the standard procedure in the field, we use log normalization to address the heavily skewed citation rates in our sample. $NAuthors_{it}$ is the average number of authors listed on individual i's publications in year t. μ_i, φ_{it}, and $\gamma_i.\delta_t$ each represent individual, department, and institution-year fixed effects, respectively. We use an ordinary least squares (OLS) regression with robust standard errors clustered at the individual level for the estimation.

To construct different levels of $\alpha(N)$ in a systematic way, following Bikard et al. (2015), we define $\alpha(N)$ as N^{-b} where N is the number of authors and b ranges from 0 to 1. Thus, for $b = 0$, $\alpha(N)$ equals 1 which represents the maximum amount of credit premium possible—that is, authors each receive 100% of the number of citations to their papers as their fractional credit. When $b = 1$, $\alpha(N)$ equals $\frac{1}{N}$, which represents the case of zero credit premium—that is, authors each receive on average one Nth of the number of citations to their papers as fractional credit. Thus, values of $0 < b < 1$ represent different levels of credit premium in this range, and values of $b > 1$ represent different levels of credit penalty.

Following this procedure, we ran the above-described regression for all values of b between 0 and 1 with increments of 0.1. Since we find evidence for collaboration credit premium within this range of values for b, there is no need to extend our testing to values of b that are greater than 1. Table A1 below shows the regression results for each value. Figure A1 depicts the estimated value of β_1 (along with its respective 95% confidence interval) with respect to different values of b. The results suggest that the estimated β_1 is closest to zero for $b = 0.34$, which suggests a credit premium level equal to $N^{-0.34}$. This means that for a team of two authors on a paper, each author receives roughly 79% of the citations to the paper as individual credit, an additional 29 percentage points of credit relative to the no-premium scenario. The magnitude of the effect is similar to that reported in Bikard et al.'s (2015) estimated credit premium of $N^{-0.48}$ for their sample of MIT science and engineering faculty members, suggesting that the level of collaboration credit premium among the economists in our sample is slightly larger. Figure A2 replicates Figure A1 for tenured faculty, showing that tenured faculty also experience a similar level of collaboration credit premium in our sample.

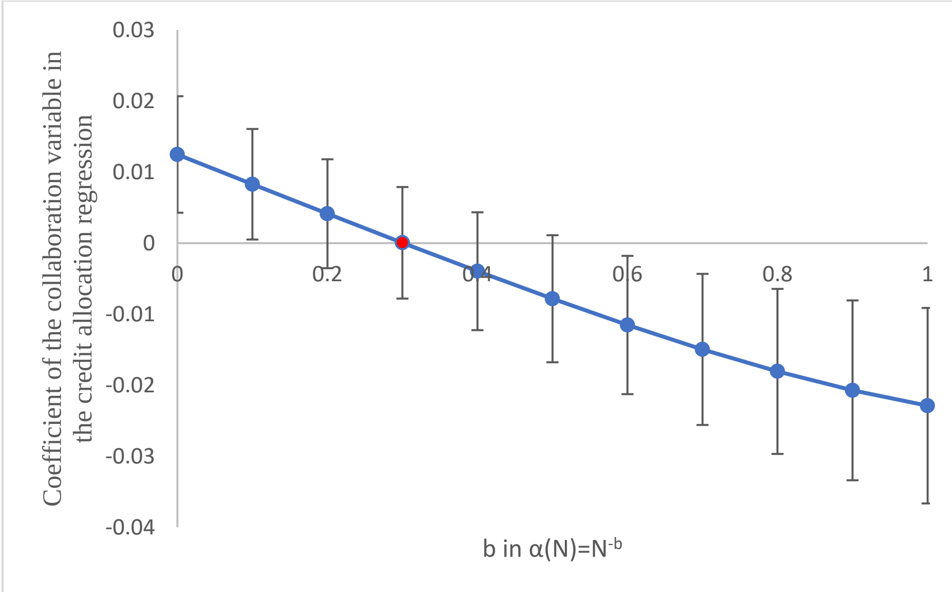
The norm of alphabetical ordering on economics articles means that individuals with a family name that begins with a letter from the beginning of the alphabet should experience a higher collaboration credit premium than their peers whose family names begin with letters from the end of the alphabet. We repeat our estimation of credit premium separately for individuals whose family names begin with letters toward the first half of the alphabet (A to M) and for those whose family names begin with letters toward the second half of the alphabet (N to Z). Figure A3 shows the results for the two groups, suggesting that the former group enjoys a higher level of credit premium ($\alpha(N) = N^{-0.29}$) than the latter ($\alpha(N) = N^{-0.55}$).

Figure A1: Imputing the level of credit allocated to each co-author from scientists' collaboration choices



Note: *α(N)* indicates the level of fractional credit allocated to each author on a publication in our sample and is defined as N^{-b} where *N* is the number of authors and *b* ranges from 0 to 1 (with increments of 0.1). The graph shows the relationship between different assumed levels of *b* and collaboration rate. The *α(N)* for which this relationship is neither positive nor negative reveals the average level of credit allocated to each co-author in our sample. Table A1 in the appendix shows the regression results for each value of *b*.

Figure A2: Replicating figure A1 for the sample of tenured faculty



Note: this graph replicates Figure 2 in the main article for the sample of tenured faculty. *α(N)* indicates the level of fractional credit allocated to each tenured author on a publication in our sample and is defined as N^{-b} where *N* is the number of authors and *b* ranges from 0 to 1 (with increments of 0.1). The graph shows the relationship between different assumed levels of *b* and collaboration rate. The *α(N)* for which this relationship is neither positive nor negative reveals the average level of credit allocated to each co-author in our sample.

Figure 3: Comparing the level of credit premium for scientists whose family name begins with letters A through M versus those whose family name begins with letters N through Z

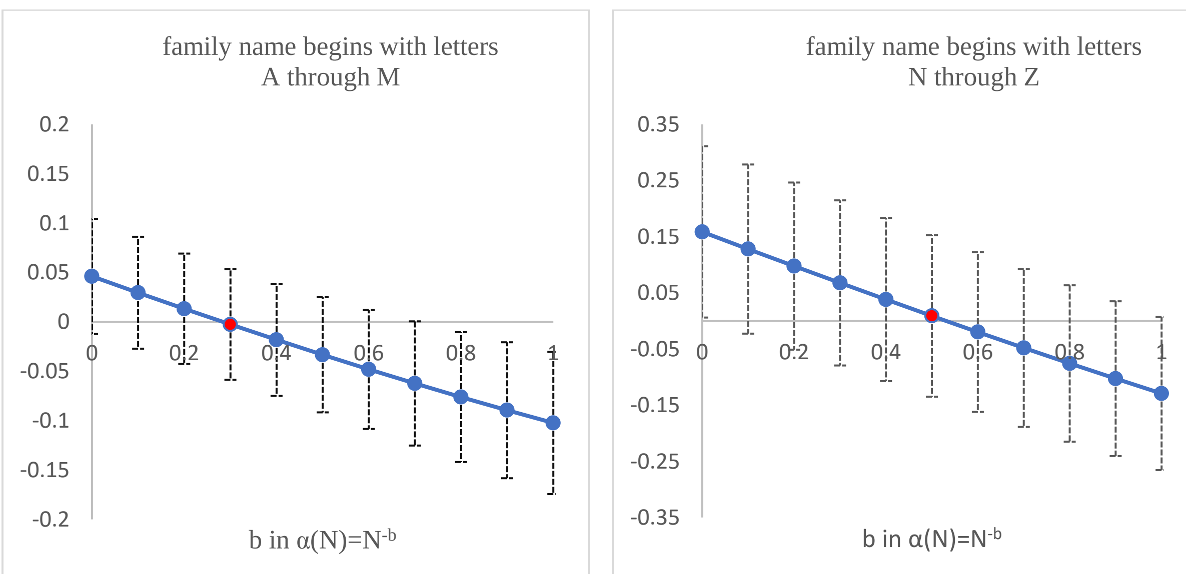


Table A1: Regression results for imputing the level of credit allocated to each co-author from scientists' collaboration choices

Model:	OLS										
Dependent variable:	**Fraction credit (yearly citations × N^{-b}) allocated to each co-author if**										
	b=0	**b=0.1**	**b=0.2**	**b=0.3**	**b=0.4**	**b=0.5**	**b=0.6**	**b=0.7**	**b=0.8**	**b=0.9**	**b=1**
Number of authors	0.065** (0.028)	0.045* (0.027)	0.025 (0.026)	0.005 (0.026)	-0.014 (0.026)	-0.032 (0.026)	-0.051* (0.026)	-0.069** (0.027)	-0.086*** (0.027)	-0.102*** (0.028)	-0.119*** (0.029)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
institution-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,650
Adjusted R-Squared	0.648	0.647	0.646	0.646	0.645	0.644	0.644	0.644	0.643	0.643	0.643

Note: The table presents the results used for creating Figure 2 in the main paper. The dependent variable in each column is the level of fractional credit allocated to each author in our sample and is defined as the total number of citations to that author's publications in any given year multiplied by N^{-b}, where N is the average number of authors on those publications and b is the parameter of interest that determines the level of fractional credit allocation. Each column shows the same regression for different values of b ranging from 0 to 1 with increments of 0.1.

*** p<0.01, ** p<0.05, * p<0.1

Table A2: Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) ln(citations +1)	1.000														
(2) Number of articles	0.717	1.000													
(3) Collaborative articles	0.532	0.669	1.000												
(4) Alphabetical rank	0.018	0.018	0.003	1.000											
(5) ln(cumul. articles+1)	0.289	0.400	0.339	0.004	1.000										
(6) ln(cumul. citations+1)	0.392	0.344	0.292	0.001	0.808	1.000									
(7) Female	-0.036	-0.043	-0.043	-0.022	-0.054	-0.045	1.000								
(8) Asian family name	-0.012	-0.001	0.013	0.065	-0.011	-0.016	0.044	1.000							
(9) N. articles not in alpha. order	0.308	0.587	0.503	0.031	0.250	0.184	-0.016	0.013	1.000						
(10) Current institutions	0.004	0.020	0.008	-0.001	-0.013	-0.012	-0.015	0.018	0.012	1.000					
(11) PhD institutions	-0.038	-0.005	-0.012	0.020	0.008	-0.029	-0.043	0.019	0.000	0.404	1.000				
(12) Type of department	0.016	0.044	0.056	-0.038	0.332	0.227	0.015	-0.028	0.034	-0.137	0.002	1.000			
(13) Year	0.070	0.171	0.182	-0.014	0.424	0.337	0.051	0.064	0.098	-0.040	-0.063	0.197	1.000		
(14) PhD start year	-0.050	-0.001	0.037	0.014	-0.091	-0.058	0.073	0.095	0.025	-0.016	-0.098	-0.177	0.701	1.000	
(15) PhD graduation year	-0.068	-0.016	0.029	0.018	-0.105	-0.074	0.072	0.097	0.018	-0.017	-0.109	-0.177	0.705	0.987	1.000

Table A3: The first-stage results of the instrumental variable method for tenured faculty

Model:	OLS
DV:	Count of collaborative papers
Alphabetical rank	0.000 (0.002)
Ln(cumulative number of articles+1)	0.051** (0.024)
Ln(cumulative number of citations+1)	0.010 (0.010)
Female	-0.010 (0.037)
Asian family name	0.071** (0.029)
Number of articles not in alphabetical order	0.154*** (0.031)
Median author rank on publications	0.049** (0.029)
Additional controls: year dummies, number of articles dummies, current institution dummies, department type dummies, starting year of PhD dummies, finishing year of PhD dummies	Yes
Obs.	6,434
F-statistics	5.870
Adjusted R^2	0.638

Note: This table replicates the results in Table 3 for tenured faculty in our sample. The analysis is at the individual-year level. The estimation is based on OLS regression with robust standard errors dual clustered at the institution and year levels.

*** p<0.01, * p<0.1

Table A4. Clustering at the individual level instead of the institutional level

The Effect of Alphabetical Rank of Scientists on Their Propensity to Collaborate (First-Stage Results of the Instrumental Variable Method)

Model:	**OLS**
DV:	**Count of collaborative papers**
Alphabetical rank	-0.003*** (0.001)
Additional controls: log normalized cumulative number of articles and citations; dummies for the total number of papers; number of articles not in alphabetical order; average author rank on publications, gender dummy; Asian last name dummy; year dummies; current institution dummies; current department dummies; starting year of PhD dummies; finishing year of PhD dummies	Yes
Obs.	16,260
Kleibergen-Paap Wald F-statistic	29.192

Notes. The analysis is at the individual-year level. The estimation is based on OLS regression with robust standard errors dual-clustered at the individual and year levels. The critical value of the Stock-Yogo weak identification test for a bias smaller than 5% for a 5% level Wald test is 16.38.
*** $p < 0.01$.

The Impact of Collaboration (Instrumented) on the Output Quality (Second-Stage Results of the Instrumental Variable Method)

Model:	**OLS**
DV:	**ln(citations+1)**
Count of collaborative papers (instrumented)	-0.642* (0.372)
Additional controls: same as above	Yes
Obs.	16,260
F-statistics	18.71

Note. The analysis is at the individual-year level. The estimation is based on OLS regression with robust standard errors dual-clustered at the individual and year levels.
* $p < 0.1$.

Table A5: The second-stage results of the instrumental variable method based on a two-stage generalized method of moments (GMM)

Model:	**OLS**
DV:	**ln(citations+1)**
Count of collaborative papers (instrumented)	-0.642**
	(0.308)
Ln(cumulative number of articles+1)	-0.417***
	(0.058)
Ln(cumulative number of citations+1)	0.190***
	(0.019)
Female	-0.018
	(0.021)
Asian family name	-0.012
	(0.019)
Number of articles not in alphabetical order	0.156*
	(0.090)
Median author rank on publications	0.089**
	(0.036)
Additional controls: year dummies, number of articles dummies, current institution dummies, department type dummies, starting year of PhD dummies, finishing year of PhD dummies	Yes
Obs.	16,260
F-statistics	24.03

Note: The analysis is at the individual-year level. The estimation is based on OLS regression with robust standard errors dual clustered at the institution and year levels. *** p<0.01, ** p<0.05

Table A6: The second-stage results of the instrumental variable method based on a limited-information maximum likelihood method

Model:	**OLS**
DV:	**ln(citations+1)**
Count of collaborative papers (instrumented)	-0.642** (0.308)
Ln(cumulative number of articles+1)	-0.417*** (0.058)
Ln(cumulative number of citations+1)	0.190*** (0.019)
Female	-0.018 (0.021)
Asian family name	-0.012 (0.019)
Number of articles not in alphabetical order	0.156* (0.090)
Median author rank on publications	0.089** (0.036)
Additional controls: year dummies, number of articles dummies, current institution dummies, department type dummies, starting year of PhD dummies, finishing year of PhD dummies	Yes
Obs.	16,260
F-statistics	24.03

Note: The analysis is at the individual-year level. The estimation is based on OLS regression with robust standard errors dual clustered at the institution and year levels. *** p<0.01, ** p<0.05