Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Dissertation acknowledgments and subsequent collaboration among business scholars

Weilong Bi

Weilong.bi@gmail.com

Luoyang Normal University

Benno Torgler

Queensland University of Technology

Research Article

Keywords: Dissertation acknowledgments, Scientific collaboration, Coauthorship networks, Research careers, Social support, Bibliometric indicators

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11292594/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Dissertation acknowledgments and subsequent collaboration among business scholars

Weilong Bi[1] and Benno Torgler[2,3,4]

[1] School of Business, Luoyang Normal University, Henan, People's Republic of China

[2] School of Economics and Finance, Queensland University of Technology, Brisbane, QLD, Australia

[3] Centre for Behavioural Insights for Technology Adoption, Brisbane, QLD, Australia

[4] Management and the Arts, Zurich, Switzerland

Abstract

Dissertation acknowledgments record support largely absent from publication and coauthorship data. Our core analytical sample comprises 1,343 U.S.-trained business scholars with linked dissertation, career, and bibliometric records. We estimate separate ordinary least squares models for years 1–3, 4–6, and 7–10 after Ph.D. completion. We distinguish new from repeated coauthor ties and examine family acknowledgment counts, text volume, and allocation among mentors. Family counts and total words are positively associated with new collaboration across all intervals; repeated-collaboration associations are positive but less uniformly precise. Mentor acknowledgment levels show a less consistent pattern. An exploratory measure of changes in mentor-text concentration under citation weighting is associated with new collaboration mainly in years 4–6. Secondary analyses show positive associations with the Financial Times 50-journal publication measure in selected specifications. The findings connect doctoral support records to later collaboration and broaden the relational information available for studying scientific careers, without establishing causal effects or personality traits.

Keywords: *Dissertation acknowledgments; Scientific collaboration; Coauthorship networks; Research careers; Social support; Bibliometric indicators*

JEL classification: D23; D91; D64; J24; O31; O32

Corresponding author: Weilong Bi at Weilong.bi@gmail.com

1 Introduction

Scientific collaboration begins before the first joint publication appears. A researcher must find potential partners, establish working relationships and sustain a project through uncertain results and editorial decisions. The growth of team science has made these relationships increasingly consequential (Wuchty et al., 2007; Jones, 2009; Leahey, 2016). Yet the social resources that make collaborative work possible extend beyond the people listed as authors. Partners, relatives, and friends may provide time, financial assistance, and encouragement; supervisors and other academics may supply advice, introductions, and access to research opportunities. Bibliometric records capture the resulting publications more readily than these surrounding relationships. Understanding scientific careers therefore requires attention to both the visible organization of research and the less visible support on which it depends.

Dissertation acknowledgments provide an unusual record of this wider environment at a common career boundary. They bring professional and personal relationships into the same text and allow authors to identify assistance that would otherwise leave little trace in publication databases. Research in that area shows that these texts combine expressions of gratitude with the presentation of a scholarly identity (Hyland, 2003). Studies of doctoral support likewise demonstrate that acknowledgments reveal relationships extending beyond formal supervision (Bitzer & Leshem, 2021). Their value is not that every word transparently measures a resource or disposition. Rather, they preserve authors' accounts of who helped them complete a demanding research project, before a substantial part of their subsequent academic career has unfolded.

This perspective builds on a limited but growing scientometrics literature. Paul-Hus et al. (2017) show how combining authorship and acknowledgments broadens the measurement of scientific collaboration. Rose and Georg (2021) recover informal collaboration and reciprocity from acknowledgments in financial economics, while Baccini and Petrovich (2022) show that normative recognition and strategic association offer complementary explanations of acknowledgment practices in leading economics journals. Smirnova and Mayr (2023) document differences in acknowledgment content across scientific domains. More recent work continues to extend the information available beyond author lists, including McCain's (2024) joint analysis of coauthorship and personal acknowledgments and the integration of acknowledgment relationships into the SOPHIS database (Petrovich et al., 2026). Together, these studies establish acknowledgments as a source for studying the organization of knowledge production. Seckin and Varol (2022) analyze more than 26,000 dissertation acknowledgments and identify academic, administrative, family, friends-and-colleagues, and spiritual support communities. Their productivity analysis considers publications in a window spanning four years before and four years

after graduation. It thus captures work associated with doctoral training as well as the delays involved in publishing it. Related research examines the relationships among acknowledgments, coauthorship and citations in the publications of Nobel laureates (Lou et al., 2023). Building on this evidence, we ask an important career question: how do acknowledgment measures at Ph.D. completion relate to the formation of new coauthor ties and the recurrence of existing ties during the following decade?

A scholar's collaboration record can develop through the formation of new partnerships, further work with existing partners, or both. New partnerships may extend access to ideas and complementary expertise, but also require search and coordination. Repeated partnerships permit collaborators to draw on established knowledge of each other's skills and working practices, although they can also reflect dependence on a narrow set of opportunities. Research on collaboration stability among economists and on strong ties in scientific careers shows why relationship history belongs in the analysis (Cainelli et al., 2015; Petersen, 2015). Our distinction concerns first observed and repeated coauthorship in publications; it provides a tractable account of network development without assuming that every new tie is intellectually novel or every repeated tie is strong.

Our empirical sample comprises 1,343 business scholars trained at U.S. universities, with available dissertation texts, career information, and bibliometric records. Collaboration is examined over years 1–3, 4–6 and 7–10 after Ph.D. completion. The main analysis uses ordinary least squares to estimate associations between acknowledgment measures and new or repeated coauthorship. We describe mentor allocation through the influence-weighting concentration difference (ICD), which records how text concentration changes after weighting mentors by citations. A secondary analysis relates these measures to the reported journal-based and citation-based research outcomes. Observing acknowledgment texts at graduation provides a common temporal reference, but researchers may already have published and collaborated during their doctorate. One issue is that the opportunity to collaborate and the capacity to produce research are jointly determined, as emphasized by Ductor (2015) and Cheng (2022). Thus, our design characterizes career associations rather than identifying a causal return to collaboration.

The findings indicate that several family acknowledgment measures are positively associated with both new and repeated collaboration. Mentor-related level measures have a less uniform pattern across the observation windows. The secondary allocation analysis shows that both nonzero ICD categories are associated with more new collaboration in years 4–6 relative to the zero category, with less consistent evidence for repeated collaboration. In general, associations with research outcomes vary across indicators and specifications.

The contribution is threefold. First, we connect the support relationships recorded at doctoral completion to two distinct forms of subsequent coauthorship over a common first-decade window. Second, we bring nonacademic relationships into a literature that necessarily relies heavily on professionally visible ties, while retaining academic acknowledgment as a comparison. Third, we distinguish acknowledgment volume from its distribution across mentors, treating the latter as an exploratory extension. This approach joins the study of research careers to the expanding use of text as data (Gentzkow et al., 2019; Feuerriegel et al., 2025). Its core implication is that the social environment surrounding a researcher may remain visible in later patterns of collaboration.

2 Acknowledgments and the social organization of collaboration

2.1 Family support and the costs of collaborative research

We distinguish two sources of acknowledgment information. Family-related emphasis brings nonacademic support into view. Collaboration imposes costs that may be difficult to fund or insure: a promising project can fail, a paper can encounter rejection or demanding revisions, and conference travel can require arranging care. Practical and emotional support may make these costs easier to bear. Research on family resources and human-capital investment supplies a general account of how private support can relax constraints on long-term commitments (Belley & Lochner, 2007; Caucutt & Lochner, 2020). The stress-buffering perspective adds a complementary mechanism: supportive relationships may help sustain effort when demands become difficult to manage (Cohen & Wills, 1985). These arguments make nonacademic relationships relevant to the study of research careers even when family members never contribute directly to a publication.

These mechanisms imply potentially different collaboration patterns. Support that makes experimentation less costly could accompany more new ties. Support that helps a scholar honor commitment through setbacks could accompany more repeated ties. Both processes may coexist. Acknowledgment measures could also reflect aspects of how an author attends to and expresses relationships. An individual who routinely recognizes assistance may behave differently in collaborative settings, although the text alone cannot establish that disposition. Equally, mentioning more relatives may simply reflect family composition or a convention of naming everyone individually. The question is therefore whether measured expressions of family recognition are associated with distinct collaboration outcomes, rather than whether a longer acknowledgment proves greater support or a more caring personality.

The broader doctoral literature also gives reasons to avoid expecting identical relationships for all forms of support and all outcomes. Estrada et al. (2019) examine how social support relates to science identity and persistence intentions. Cao et al. (2024) find associations between institutional and supervisory support and doctoral engagement, but no statistically significant association for family support in their sample. These findings make the distinction among support sources useful rather than incidental. Nonacademic support may be more closely related to a researcher's capacity to undertake and sustain particular relationships than to a general engagement measure. Conversely, observed acknowledgment associations may arise from unmeasured circumstances shared by family life and academic careers. Comparing new and repeated collaboration helps characterize the pattern, while leaving these explanations open.

2.2 Mentoring and the allocation of recognition

Mentor-related acknowledgments capture a different set of relationships. Supervisors and other acknowledged academics may supply specialist knowledge, professional introductions and access to established research networks. Studies of advisor matching and early productivity emphasize the importance of this academic environment (García-Suaza et al., 2020). At the same time, opportunities are distributed within institutional hierarchies (Wapman et al., 2022), and research resources contribute to productivity differences between universities (Zhang et al., 2022). Extensive thanks to an influential mentor may consequently reflect substantial assistance, privileged access, customary deference, or a wish to display a valued association. The normative and strategic accounts developed by Baccini and Petrovich (2022) are especially valuable: the same expression can recognize a real contribution and position its author within a professional community. Models of prosocial behavior allow concern for others, social reputation, and self-respect to coexist (Bénabou & Tirole, 2006). This perspective treats acknowledgment as a public act of recognition whose meaning depends on both the relationship and its audience, without assigning a single motive to the observed text.

To examine this domain, we distinguish the amount of mentor-related acknowledgment from its allocation among acknowledged mentors. Two scholars can devote the same number of words to mentors while distributing that recognition very differently. We summarize this allocation with concentration measures and examine how concentration changes when mentor-specific influence weights are applied. We refer to the resulting statistic as the influence-weighting concentration difference (ICD). It describes the relationship between a text allocation and an influence-weighted version of that allocation. Its sign is not a direct measure of preference for prestigious mentors: weighting can increase concentration because mentor influence is unequal even when the author

gives everyone the same amount of text. The statistic is therefore a secondary descriptive measure whose associations are interpreted alongside the simpler acknowledgment counts.

3 Data and measurement

3.1 Sample and sources

The study combines doctoral dissertations, publicly available career information, and Scopus bibliometric records. The sampling frame comprises business scholars associated with the top 150 U.S. universities in the 2002 U.S. News & World Report ranking. Dissertation texts were retrieved from ProQuest, university library archives, and personal websites. Graduation year, doctoral institution, and field of specialization were checked when linking the sources. The initial population contains 4,823 scholars, and the linked acknowledgment and bibliometric dataset contains 1,372. The analytical sample contains 1,343 scholars, 29 fewer than the linked dataset due to mismatches identified during cross-validation, which were excluded from the analysis. The unit of analysis is the individual scholar, and each interval-specific regression contains 1,343 scholar observations. We analyze the three career intervals separately.

Career information was collected from curricula vitae, faculty directories, and professional webpages. It includes recorded demographic characteristics, doctoral institution, employment history, and promotion information. Scholars with incomplete background information or fewer than ten years since Ph.D. completion by 2022 were excluded. The sample also excludes scholars with substantial early industrial experience and specified clinical or practice-oriented appointments. Consequently, the analysis describes an observed group of academic business scholars with sufficiently complete records. It does not represent all business doctoral graduates or include the full range of transitions out of academic employment.

The 2002 ranking defines the institutional sampling frame; the 2022 U.S. News rankings provide the institutional-rank covariates. These dates serve different purposes. In particular, a later ranking assigned to a historical doctoral or employing institution is a retrospective classification, not a contemporaneous measure of the resources available when the scholar studied or worked there. Scopus provides publication, citation, and coauthor information. Author matching used identifiers, affiliations, and manual name disambiguation. Dissertations provide a common observation point at Ph.D. completion, but not necessarily the beginning of a research career: publications, employment, and collaboration may precede graduation. The common follow-up period improves comparability in time since the doctorate without making scholars comparable in accumulated experience or calendar-time conditions.

3.2 Collaboration and research outcomes

New collaboration refers to a coauthor first observed in a scholar's publication history. Repeated collaboration refers to a previously observed coauthor appearing again in subsequent publications. We summarize these outcomes for years 1–3, 4–6, and 7–10 after Ph.D. completion. This distinction captures entry into observed coauthorship and recurrence of an observed tie. It does not directly measure when joint work began, whether collaborators remained in contact between publications, or whether a project was abandoned before publication. Database coverage and name disambiguation also affect whether a tie is classified as new.

New and repeated collaboration have different opportunity sets. A scholar can form new ties without an extensive existing network, whereas repeat collaboration requires an earlier observed tie. More publications also create more opportunities to observe both types of collaboration. The estimates therefore concern publication-based collaboration patterns, not an individual's willingness to cooperate holding research activity constant. Nor is the number of repeated collaborations a retention rate: that interpretation would require a denominator identifying the earlier ties that could have recurred.

Secondary analyses examine a SCImago Journal Rank (SJR) based publication measure, a Financial Times 50-journal (FT-50) publication measure, and a publication-year-weighted, discounted citation measure. They represent different aspects of research output. SJR is a journal-level citation indicator; FT-50 membership identifies a particular set of journals used in business-school research assessment; citations describe recorded scholarly attention. None is a comprehensive measure of the quality of an individual contribution. Journal indicators and rankings can yield different assessments in business and management (Mingers & Yang, 2017), and bibliometric evaluation may disadvantage work whose influence emerges slowly (Wang et al., 2017). Coefficients are interpreted in each outcome's estimation units, not as percentages, standardized effects, or raw counts of additional collaborators, publications, or citations. The final interval covers four years, compared with three in each earlier interval; differences in interval coefficients therefore do not by themselves establish changes in annual rates.

3.3 Acknowledgment extraction

The text extraction proceeded in two rounds. In an initial pilot in early 2022, seven trained research assistants coded acknowledgment texts with the aid of Linguistic Inquiry and Word Count resources. The exercise identified people and relational categories and recorded counts of

individuals, words, and positive-emotion words. Disagreement among coders led to a revised procedure, and the pilot annotations were not retained for the reported analysis. This experience indicates that assigning personal text to relational categories is a substantive measurement task; increasing the number of coders alone does not resolve ambiguous classifications (Lombard et al., 2002).

The subsequent automated natural-language-processing procedure combined LIWC lexical resources with entries from the Cambridge English Dictionary. It assigned acknowledgment text to relational groups and counted people, total words, and positive words within each group. Mentor names were matched to their publication and citation records. To make citation impact comparable across papers of different ages, we converted each paper's total citations into an age-adjusted annual rate. That is, dividing the paper's cumulative citation count by the number of years elapsed between its publication year and the data collection year. Such paper to level annual rates were then summed across all of a mentor's papers to obtain the mentor's yearly weighted citation measure. The resulting dataset contains 4,923 mentor-level observations associated with the 1,372 scholars.

Family and mentor counts describe how many people are identified in the corresponding group. Total word counts describe the volume of text attributed to that group, while positive-word counts describe the number of words classified as positive by the lexicon. A count of positive words is affected by the amount of text available: it is not a length-adjusted sentiment score. Similarly, naming more relatives can reflect family composition or disclosure conventions as well as the breadth of support.

Automated processing makes repeated application feasible, but reproducibility of a classification rule is different from accuracy. Group references, shared sentences, unnamed relatives, and a mentor performing several roles can make attribution ambiguous. The analysis does not provide an independently adjudicated validation exercise quantifying these errors. We therefore treat the measures as recorded features of acknowledgment language and avoid treating them as validated scales of gratitude, closeness, resilience, or personality. This distinction follows the broader emphasis on construct validity in computational text analysis (Gentzkow et al., 2019; Feuerriegel et al., 2025). Appendix Figure 3 presents supplementary distributions of the unweighted, citation-weighted, and Google-weighted measures; its note distinguishes these displays from the analytical sample.

3.4 Allocation of text among mentors

The acknowledgment data link dissertation authors to individuals identified in their acknowledgment sections. The mentor analysis uses the text attributed to each identified mentor within a dissertation. The underlying mentor records and the scholar observations in the regressions are therefore different units: several mentor records contribute to one author-level measure. The same mentor may appear in more than one dissertation.

Let i index the dissertation author, M_i denote that author's set of identified mentors, $j \in M_i$ index a mentor, and $n_i = |M_i|$ denote the number of mentors in the set. Let y_{ij} denote the acknowledgment text attributed by author i to mentor j, measured separately by total-word and positive-word counts. For a given author, these observations describe allocation across mentors within one dissertation; they do not describe how one mentor is acknowledged across different dissertations. Whenever the relevant distribution has a positive total, each Gini coefficient and Theil T index yields one value for that author, which can be linked to the author's subsequent collaboration and research outcomes. The mean allocation is

$$\overline{y}_i = \frac{1}{n_i} \sum_{j \in \mathcal{M}_i} y_{ij} \quad (1)$$

Two concentration measures summarize this distribution. The Gini coefficient is

$$G_i = \frac{\sum_{j \in \mathcal{M}_i} \sum_{k \in \mathcal{M}_i} |y_{ij} - y_{ik}|}{2 n_i^2 \overline{y}_i} \quad (2)$$

and the Theil T index is

$$T_i = \frac{1}{n_i} \sum_{j \in \mathcal{M}_i} \left(\frac{y_{ij}}{\overline{y}_i} \right) \ln \left(\frac{y_{ij}}{\overline{y}_i} \right) \quad (3)$$

Both measures equal zero when each identified mentor receives the same positive allocation. Multiplying every allocation within a dissertation by the same positive constant leaves both indices unchanged. Thus, expressing allocations as shares of that dissertation's mentor text would produce the same concentration values. This property separates concentration from total text volume; it does not make the word-count regressors in Table 1 independent of acknowledgment length. Theil T is more sensitive to high-intensity allocations. Appendix Figure 4 compares the Gini and Theil T summaries graphically.

Zero allocations contribute zero to the Theil sum under the usual limiting convention. A dissertation with only one identified mentor has zero concentration when its allocation is positive; an all-zero distribution is undefined. With n_i mentors, the uncorrected Gini coefficient has a

maximum of $(n_i - 1)/n_i$, while Theil T has a maximum of $\ln(n_i)$. Mentor-group size therefore matters when interpreting concentration differences across scholars.

Let C_j denote the cumulative citations of acknowledged mentor j, rather than those of dissertation author i. These mentor-specific citation counts enter through the influence weight:

$$w_j = \ln(1 + C_j) \quad (4)$$

The corresponding weighted word allocation is

$$\tilde{y}_{ij} = w_j y_{ij} \quad (5)$$

We call the weighted-minus-unweighted concentration difference the influence-weighting concentration difference, abbreviated ICD. The weighted and unweighted indices use the same mentor set, including zero allocations, and each uses its own mean. Writing the original and weighted allocations as vectors, the two versions are:

$$\text{ICD}_i^G = G(\tilde{\mathbf{y}}_i) - G(\mathbf{y}_i), \quad (6a)$$

$$\text{ICD}_i^T = T(\tilde{\mathbf{y}}_i) - T(\mathbf{y}_i). \quad (6b)$$

A positive ICD means that weighting by mentor citations increases concentration. A negative ICD means that weighting decreases concentration. Zero means that concentration is unchanged. These definitions require positive totals in both distributions. Because mentors within a dissertation can have different citation counts, the weighting can change the relative allocations. Applying a single weight to every observation would leave the concentration indices unchanged. Appendix Figure 5 compares the total-word and positive-word constructions under citation and Google weighting.

The ICD has a narrower interpretation than a measure of prestige preference. Equal text allocations can yield a positive ICD if influence weights differ. A positive value can also arise when the less influential mentor receives more text: with allocations (3, 2) and weights proportional to (1, 4), the Gini rises from 0.10 to approximately 0.227. The ICD is then approximately 0.127 even though the first mentor receives more unweighted text. The statistic reflects both mentor composition and text allocation; its sign does not identify the author's motives.

The zero category is also heterogeneous. It can contain single-mentor dissertations, distributions with equal mentor weights, and other cases in which weighting leaves concentration unchanged. It is not a group of authors demonstrably indifferent to status. Furthermore, cumulative mentor citations characterize influence when measured; without a historical citation snapshot they cannot be assumed to represent the influence visible to the student at graduation. This timing limitation concerns the weighted index even though the underlying acknowledgment text precedes the later career outcomes.

3.5 Google-based web visibility

Citations and web visibility capture different forms of recognition. Chan et al. (2014, 2016) use Google Search hit counts to examine scholars' external prominence alongside their standing within academia. Their methodological approach includes searches for quoted publication names and manual checks for namesakes. Chan et al. (2016) also compare repeated searches and results from Google and Bing, and examine counts outside the .edu domain. These checks address attribution and reliability. The counts nevertheless remain estimates of indexed web pages; they do not reveal whether attention is favorable or measure the quality of a scholar's research.

For the supplementary Google construction, let Q_j denote the Google Search hit count for acknowledged mentor j. The alternative weight is $\ln(1 + Q_j)$, which replaces the citation weight in equation (4); the text allocation and concentration differences are then constructed as in equations (5) and (6). These are ordinary web-search counts, distinct from Google Scholar citations and Google Trends search-interest indices. Also here the weights characterize the mentors, not the dissertation authors. The Google panels in Appendix Figures 3 and 5 provide descriptive comparisons of the resulting measures. As with citations, a later web snapshot cannot be assumed to represent visibility at graduation.

4 Empirical approaches

4.1 Conditional associations with collaboration

We estimate separate OLS regressions for each collaboration outcome and career interval. The purpose is to characterize conditional associations between acknowledgment measures and later observed ties. In the level specifications, the family and mentor variables enter together, using either the number of acknowledged individuals, total words, or positive words:

$$C_{it} = \alpha_t + \beta_t F_i + \theta_t H_i + Z_{it}'\delta_t + \varepsilon_{it} \quad (7)$$

Here C_{it} is the collaboration outcome for scholar i in interval t, F_i is the family measure, and H_i is the corresponding aggregate mentor measure. Z_{it} collects the covariates included in the specification. The family and mentor measures in each pair use the same type of count, but their coefficients need not be comparable as measures of substantive importance because their distributions and meanings differ.

The background covariates comprise recorded gender and racial classification, partner or child status, doctoral-institution rank, mentor–student gender difference, and a mentor Nobel indicator. Later intervals additionally include the lagged career variables indicated in the tables, including employment-institution rank and promotion status. Where cohort indicators are included, they group scholars by decade of bachelor's-degree entry. These are broad educational-cohort controls, not year-specific measures of the publication environment. The recorded demographic categories are coarse controls and do not exhaust relevant differences in experience or opportunity.

The distributional specifications replace mentor acknowledgment levels with indicators for negative and positive ICD values, retaining a family text measure:

$$C_{it} = \alpha_t + \beta_t F_i + \gamma_t^- D_i^- + \gamma_t^+ D_i^+ + Z_{it}'\delta_t + \varepsilon_{it} \quad (8)$$

D_i^- and D_i^+ indicate ICD values below and above zero, respectively. The omitted category is ICD equal to zero. Separate regressions use the Gini and Theil constructions and total-word and positive-word allocations. The total-word specifications include family total words; the positive-word specifications include family positive words. Categorization avoids imposing a linear relationship across the ICD distribution, but discards information about the magnitude of the difference. It also makes the composition of the zero group central to interpretation.

OLS provides a common linear summary across the specifications. The tables report heteroskedasticity-robust standard errors. They describe uncertainty under the reported specifications and do not account for uncertainty in text classification or index construction. Shared mentors, doctoral institutions, and coauthorship ties may also generate dependence across scholars that ordinary robust standard errors do not absorb.

4.2 Research outcomes

The secondary analyses relate each research indicator directly to the ICD categories:

$$Y_{it}^k = a_t^k + b_t^{k-} D_i^- + b_t^{k+} D_i^+ + W_{it}' d_t^k + u_{it}^k \quad (9)$$

Y_{it}^k denotes outcome k in interval t. The covariate vector W_{it} includes the background and educational-cohort controls and, in the later intervals, lagged repeated collaboration, employment-

institution rank, and promotion status. The reported outcome specifications use the citation-based Gini ICD calculated from total words or positive words. These regressions examine whether acknowledgment allocation is associated with later bibliometric outcomes after conditioning on the specified career history.

The coefficients summarize associations with each research outcome conditional on the included characteristics. They do not identify an effect independent of collaboration, because prior collaboration, institutional placement, and promotion may themselves reflect earlier support and opportunities. We therefore interpret the research-outcome regressions as exploratory comparisons across indicators, alongside the main collaboration results.

5 Results

5.1 Family and mentor acknowledgment levels

Table 1 presents the level specifications. Each additional acknowledged family member is associated with a 0.081-unit higher new-collaboration measure in years 1–3, 0.109 units in years 4–6, and 0.197 units in years 7–10; robust standard errors are 0.033, 0.051, and 0.087. To illustrate the scale, a difference of two acknowledged family members corresponds to fitted differences of 0.162, 0.218, and 0.394 outcome units, holding the included characteristics constant. The repeated-collaboration coefficients are also positive—0.0545 (0.029), 0.0851 (0.042), and 0.139 (0.051)—with weaker precision in the first interval. This pattern is consistent with family recognition being related to both entry into new partnerships and recurrence of existing ties, without identifying the resources or behavior responsible for the relationship.

Family total-word coefficients are positive across all three intervals, with clearer evidence for new collaboration. The new-collaboration estimates are 0.00353 (0.001), 0.00569 (0.002), and 0.00759 (0.003). The corresponding repeated-collaboration estimates are 0.00209 (0.002), 0.00331 (0.002), and 0.00716 (0.004); the first is imprecise, and the latter two reach only the 10% significance level under the approximate tests reported in the table. Greater text volume may reflect more extensive support, greater willingness to recognize it, or a more expansive writing style. The estimates cannot separate these explanations.

The positive-word specifications qualify this pattern. Family coefficients are positive throughout but are less uniformly precise than the family-count estimates. The clearest positive-word association concerns repeated collaboration in years 7–10, at 0.0572 (0.019). Because positive words are counts rather than a length-adjusted sentiment score, this finding does not establish that

more positive emotional tone is associated with collaboration independently of the amount written. The more consistent result concerns the breadth of family recognition and the volume of family-related text.

Table 1 Acknowledgment counts and word counts and subsequent collaboration

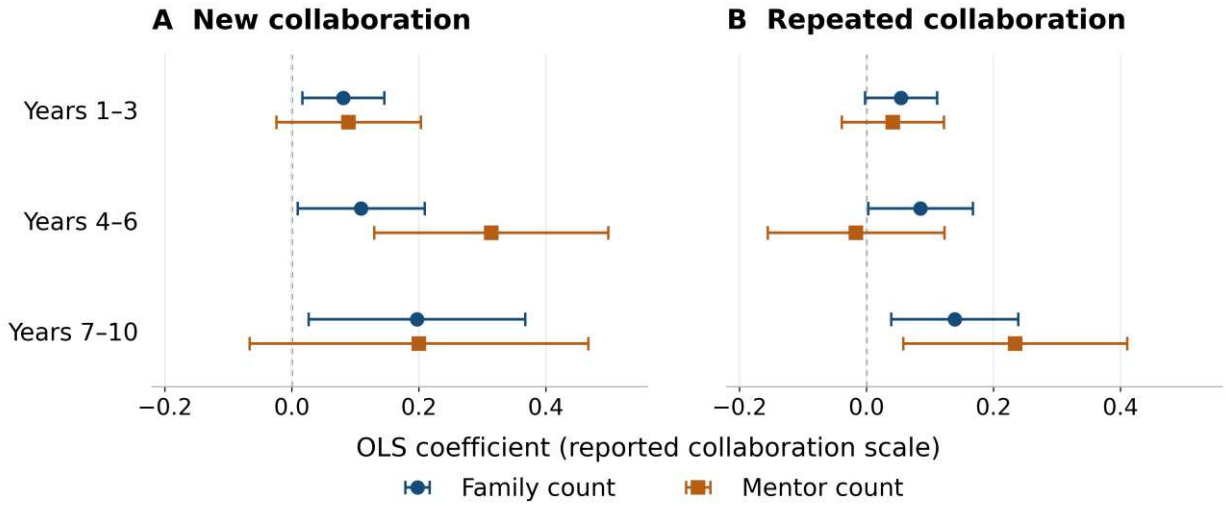
Acknowledgment measure	New collaboration			Repeated collaboration		
Post-Ph.D. years	**1–3**	**4–6**	**7–10**	**1–3**	**4–6**	**7–10**
Panel A Number of acknowledged people						
Acknowledged family members	0.081** (0.033)	0.109** (0.051)	0.197** (0.087)	0.0545* (0.029)	0.0851** (0.042)	0.139*** (0.051)
Acknowledged mentors	0.0893 (0.058)	0.314*** (0.094)	0.2 (0.136)	0.0416 (0.041)	−0.0165 (0.071)	0.234*** (0.090)
Panel B Total word counts						
Family text	0.00353*** (0.001)	0.00569*** (0.002)	0.00759** (0.003)	0.00209 (0.002)	0.00331* (0.002)	0.00716* (0.004)
Mentor text	0.00257** (0.001)	0.0031 (0.002)	0.00249 (0.002)	0.00136 (0.001)	0.00139 (0.001)	0.000493 (0.002)
Panel C Positive-word counts						
Family text	0.022* (0.013)	0.0263 (0.021)	0.0414 (0.034)	0.00503 (0.010)	0.0225* (0.013)	0.0572*** (0.019)
Mentor text	0.0143 (0.017)	0.0593* (0.033)	0.0763 (0.048)	0.0055 (0.012)	−0.0057 (0.026)	0.0229 (0.030)
Scholars (N)	1,343	1,343	1,343	1,343	1,343	1,343
Background controls	Yes	Yes	Yes	Yes	Yes	Yes
Lagged career controls	No	Yes	Yes	No	Yes	Yes

Notes. OLS coefficients; robust standard errors in parentheses. Within each panel, the family and mentor measures enter jointly in a separate regression for each outcome and interval. Panels A, B, and C use person counts, total words, and positive words, respectively. All regressions include the background controls in Section 4.1; lagged career controls are promotion and employment-institution rank. Years are measured after Ph.D. completion. Coefficients retain the reported scales and are not interpreted as probabilities or standardized effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Symbols use approximate two-sided normal tests based on the displayed coefficients and robust standard errors. Rounding can change significance categories. No multiple-testing adjustment is applied.

Mentor acknowledgment levels show selected positive associations rather than a uniform pattern. Each additional acknowledged mentor is associated with a 0.314-unit higher new-collaboration measure in years 4–6 (standard error 0.094) and a 0.234-unit higher repeated-collaboration measure in years 7–10 (0.090). Mentor total words are positively associated with new collaboration in years 1–3, at 0.00257 (0.001), whereas most other word-based estimates are imprecise. These results suggest that mentor counts, text volume, and positive-word counts convey different information.

They do not rank the importance of academic and family support, since the measures have different distributions and capture different relationships. Figure 1 summarizes the count specifications.

Fig. 1 Acknowledged family members and mentors and subsequent collaboration



Notes. Person-count specifications from Table 1, with family and mentor counts entered jointly. Each regression uses 1,343 scholars. Points are reported OLS coefficients; whiskers are approximate 95% confidence intervals calculated from rounded robust standard errors.

5.2 Mentor acknowledgment allocation and new ties

Table 2 reports the citation-weighted ICD category estimates for new collaboration. In years 4–6, the Gini total-word model places the negative category 1.21 outcome units above the zero category and the positive category 0.854 units above it, conditional on the controls. Their robust standard errors are 0.484 and 0.338, giving approximate 95% confidence intervals of 0.26–2.16 and 0.19–1.52. These are differences between categories, not responses to a one-unit increase in the ICD. The Gini positive-word contrasts are 1.49 (0.532) and 0.856 (0.341), and the Theil T constructions show the same positive pattern in this interval. The agreement across constructions makes the middle-interval pattern worth examining, although these are closely related measures of the same acknowledgment records.

Earlier and later estimates are less consistent in sign and less precise. The pattern therefore concerns years 4–6 in these separate regressions; it does not establish that the association is stronger then than in other intervals. Nor does it identify a transition to independence from doctoral supervisors. New coauthors in this interval could come from within or outside the doctoral network. Linking the identities of acknowledged mentors to subsequent coauthors would turn this interpretation into a direct question about how doctoral relationships accompany network expansion.

The fact that both nonzero categories have positive estimates relative to zero is central. It is inconsistent with reading the table as a simple ranking from less to more prestige-oriented behavior. The contrast may instead reflect differences between authors for whom influence weighting changes concentration and a heterogeneous reference group for whom it does not. Mentor-group size, the dispersion of mentor citations, and the relative attention paid to different mentoring roles can all contribute to that distinction.

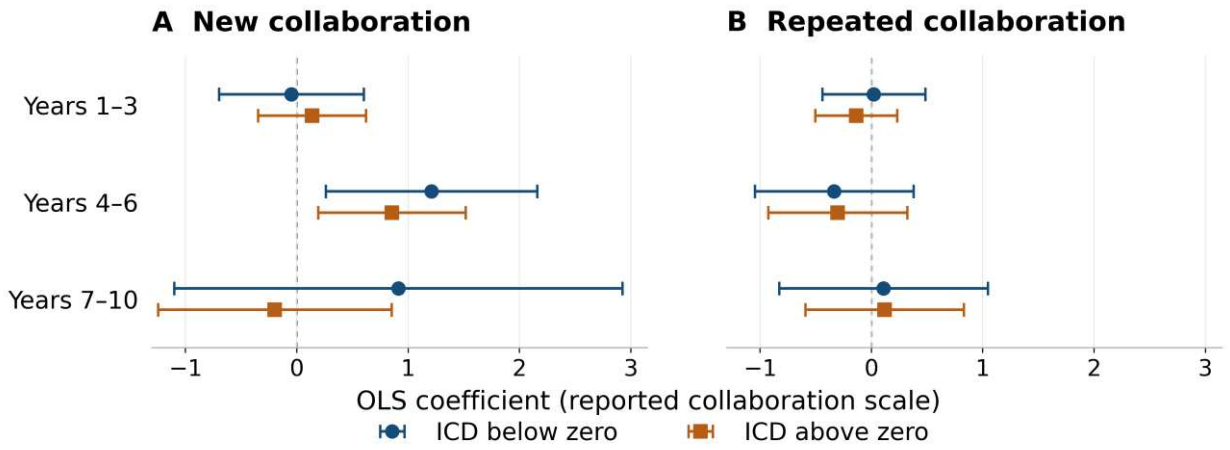
Table 2 Citation-based ICD categories and new collaboration

ICD construction	ICD category	Years 1–3	Years 4–6	Years 7–10
Gini · total words	ICD < 0	−0.0493 (0.332)	1.21** (0.484)	0.912 (1.027)
Gini · total words	ICD > 0	0.137 (0.247)	0.854** (0.338)	−0.198 (0.535)
Theil T · total words	ICD < 0	−0.00656 (0.293)	1.05** (0.492)	−0.534 (0.594)
Theil T · total words	ICD > 0	0.144 (0.249)	0.86** (0.336)	0.0771 (0.550)
Gini · positive words	ICD < 0	0.154 (0.360)	1.49*** (0.532)	1.05 (1.143)
Gini · positive words	ICD > 0	0.127 (0.251)	0.856** (0.341)	−0.15 (0.538)
Theil T · positive words	ICD < 0	−0.0189 (0.301)	1.47** (0.593)	−0.504 (0.630)
Theil T · positive words	ICD > 0	0.155 (0.253)	0.838** (0.336)	0.0686 (0.549)
Scholars (N)		1,343	1,343	1,343
Background controls		Yes	Yes	Yes
Lagged career controls		No	Yes	Yes

Notes. OLS coefficients for new collaboration; robust standard errors in parentheses. ICD is the influence-weighting concentration difference using mentor citations. Both sign categories enter jointly; the reference is ICD = 0 within the same construction. Each construction and interval is a separate regression. Total-word models include family total words; positive-word models include family positive words. Background controls are included throughout, with lagged career controls in later intervals (Section 4.1). Years are measured after Ph.D. completion. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Symbols use approximate two-sided normal tests based on the displayed coefficients and robust standard errors.

Appendix Table A1 presents the corresponding estimates for repeated collaboration. The category coefficients are small relative to their standard errors and vary in sign, providing little precise evidence of an association with repeated ties. This contrast is suggestive of different relationships with tie formation and recurrence, but separate significance results do not establish that the two sets of coefficients differ. Figure 2 places the Gini total-word estimates for both outcomes side by side; the remaining constructions appear in the tables.

Fig. 2 Mentor allocation categories and new and repeated collaboration



Notes. Citation-based Gini total-word ICD specifications from Table 2 and Appendix Table A1; both sign categories are compared with ICD equal to zero. ICD is a concentration difference. Each regression uses 1,343 scholars. Points are reported OLS coefficients; whiskers are approximate 95% confidence intervals calculated from rounded robust standard errors.

5.3 Associations with research outcomes

Tables 3 and 4 relate the ICD categories to research outcomes. In the total-word construction, the positive category is associated with a higher FT-50 publication measure in all three intervals: 0.0556 (0.028), 0.137 (0.044), and 0.0972 (0.038), relative to the zero category. The middle-interval contrast is therefore 0.137 units of the FT-50 measure, with an approximate 95% confidence interval of 0.051–0.223. In that interval, the negative category has a similar point estimate, 0.127 (0.059). The positive-word construction gives closely corresponding estimates of 0.138 (0.044) and 0.126 (0.061). The middle-interval association thus concerns both nonzero categories, while the positive category has the more consistent pattern across intervals; it is not evidence of a monotonic return to influence-oriented acknowledgment.

The other indicators yield a less uniform picture. For SJR, the negative total-word ICD category has a positive early estimate, 0.664 (0.271), whereas the positive total-word category does not show a consistent positive pattern. In the positive-word construction, the positive category has an early SJR estimate of 0.663 (0.270), but subsequent estimates are less precise. The sensitivity of these SJR comparisons to the text construction limits any general interpretation in terms of journal standing.

For the publication-year-weighted citation measure, the positive category is associated with higher outcomes in years 7–10 in both constructions: 1.31 (0.628) for total words and 1.29 (0.626) for positive words. The first contrast has an approximate 95% confidence interval of 0.08–2.54 in weighted-citation units. Earlier estimates are less precise, and the negative-category estimates do not reveal a consistent positive pattern. The late citation association accompanies a less uniform

set of journal-based results, suggesting that different indicators capture different aspects and timing of research recognition. It does not establish that acknowledgment allocation improves research quality or generates attention independently of mentoring and institutional opportunities.

Table 3 Citation-based Gini ICD from total words and research outcomes

Research outcome	ICD category	Years 1–3	Years 4–6	Years 7–10
SJR publication measure	ICD < 0	0.664** (0.271)	0.745 (0.503)	0.265 (0.437)
SJR publication measure	ICD > 0	0.413 (0.377)	−0.0387 (0.618)	−0.472 (0.525)
FT-50 publication measure	ICD < 0	0.0462 (0.041)	0.127** (0.059)	0.0595 (0.047)
FT-50 publication measure	ICD > 0	0.0556** (0.028)	0.137*** (0.044)	0.0972** (0.038)
Year-weighted citations	ICD < 0	−0.566 (1.115)	−0.539 (1.033)	−0.232 (0.688)
Year-weighted citations	ICD > 0	−0.0725 (1.010)	1.25 (0.860)	1.31** (0.628)
Scholars (N)		1,343	1,343	1,343
Background controls		Yes	Yes	Yes
Cohort indicators		Yes	Yes	Yes
Lagged career controls		No	Yes	Yes

Note. OLS coefficients; robust standard errors in parentheses. The explanatory indicators are categories of the citation-based Gini ICD from mentor total-word counts. Both enter jointly; the reference is ICD = 0. Each outcome and interval is a separate regression with background and bachelor-entry cohort controls. Later intervals also include lagged repeated collaboration, employment-institution rank, and promotion. SJR = SCImago Journal Rank; FT-50 = Financial Times 50-journal list. Outcome scales differ, so magnitudes are not comparable across outcomes. Years are measured after Ph.D. completion. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Symbols use approximate two-sided normal tests based on the displayed coefficients and robust standard errors. The citation outcome is discounted by publication year.

Taken together, the research-outcome regressions offer a more specific result than a general relationship with scholarly success. The FT-50 associations are comparatively consistent across text constructions, while the SJR and citation associations depend more on interval and category. For business scholars, placement in a selected journal list, journal-level citation standing, and accumulated article attention need not move together. The results are therefore most informative about the particular outcome being measured; the favorable FT-50 pattern should not be generalized to all dimensions of research performance.

Table 4 Citation-based Gini ICD from positive words and research outcomes

Research outcome	ICD category	Years 1–3	Years 4–6	Years 7–10
SJR publication measure	ICD < 0	0.41 (0.386)	−0.169 (0.623)	−0.455 (0.549)
SJR publication measure	ICD > 0	0.663** (0.270)	0.734 (0.503)	0.252 (0.435)
FT-50 publication measure	ICD < 0	0.0527 (0.045)	0.126** (0.061)	0.0536 (0.047)
FT-50 publication measure	ICD > 0	0.0564** (0.028)	0.138*** (0.044)	0.0996*** (0.038)
Year-weighted citations	ICD < 0	−0.224 (1.164)	−0.517 (1.095)	−0.278 (0.705)
Year-weighted citations	ICD > 0	−0.129 (1.009)	1.21 (0.857)	1.29** (0.626)
Scholars (N)		1,343	1,343	1,343
Background controls		Yes	Yes	Yes
Cohort indicators		Yes	Yes	Yes
Lagged career controls		No	Yes	Yes

Note. OLS coefficients; robust standard errors in parentheses. The explanatory indicators are categories of the citation-based Gini ICD from mentor positive-word counts. Both enter jointly; the reference is ICD = 0. Each outcome and interval is a separate regression with background and bachelor-entry cohort controls. Later intervals also include lagged repeated collaboration, employment-institution rank, and promotion. SJR = SCImago Journal Rank; FT-50 = Financial Times 50-journal list. Outcome scales differ, so magnitudes are not comparable across outcomes. Years are measured after Ph.D. completion. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Symbols use approximate two-sided normal tests based on the displayed coefficients and robust standard errors. The citation outcome is discounted by publication year. Rounding can change significance categories. No multiple-testing adjustment is applied.

6 Discussion

6.1 Support environments and the development of collaboration

The findings connect two parts of a research career that are usually studied separately: the support recognized at doctoral completion and the later pattern of published collaboration. Coauthorship records are indispensable for observing scientific networks, but the assistance that makes participation possible extends beyond people who become coauthors. The family results bring that wider environment into view. Their substantive relevance lies in the possibility that the capacity to enter and sustain research relationships depends partly on resources and relationships outside the formal organization of science.

When a junior scholar decides to join an uncertain project, meetings, preliminary work, and travel may absorb time before the team knows whether the project is viable. Family assistance could reduce the cost of that investment, while emotional support could make it easier to continue after

rejection. These possibilities connect the constraint-based account of family resources (Belley & Lochner, 2007; Caucutt & Lochner, 2020) with research on stress buffering and doctoral persistence (Cohen & Wills, 1985; Estrada et al., 2019). These associations align with these channels but do not measure their relative contribution.

The same acknowledgment can also reflect a relational disposition. A scholar attentive to assistance from family members may be attentive to the contributions of collaborators. Research on social preferences supports investigating this possibility (Carpenter & Seki, 2011; Kosse et al., 2020). Yet expressions of gratitude are produced in particular social settings. Family structure, cultural convention, and expectations about what belongs in a dissertation can generate similar language. Distinguishing a stable disposition from resources and conventions requires independent measures; the current evidence does not justify labelling an author caring, cooperative, or resilient from the acknowledgment alone.

Separating new and repeated ties advances this interpretation. Forming a new tie involves finding a partner and assessing whether knowledge and working practices are compatible. Repeated work can reduce those uncertainties but also reflects earlier selection and opportunities to publish together. This distinction links acknowledgment evidence to work on strong ties and enduring partnerships in scientific careers (Cainelli et al., 2015; Petersen, 2015). Our results suggest that acknowledgment-based studies can benefit from treating collaboration as a sequence of relationships rather than a single stock of coauthors.

6.2 Recognition within unequal academic environments

Mentor acknowledgment brings a different set of relationships into view. Mentors can provide substantive advice, institutional access, recommendations, and models of professional conduct. Their standing may shape both the assistance available and the conventions surrounding its recognition. Research on advisor matching and institutional hierarchy makes these features central to the interpretation of mentor-related text (García-Suaza et al., 2020; Wapman et al., 2022). Acknowledgment allocation thus sits at the intersection of support received and support publicly recognized.

The distinction between normative and strategic accounts of acknowledgments remains useful (Baccini & Petrovich, 2022). Extensive thanks may honor a substantial contribution, comply with a local convention, express affection, or convey an affiliation that the author values. Several motives can coexist. Informal collaboration is itself shaped by reciprocity and patterns of access

that are only partly visible in authorship data (Rose & Georg, 2021). This explains why the text can be informative without being a direct readout of intent.

The distributional analysis adds a measurement distinction to this literature. Overall acknowledgment volume asks how much text is devoted to mentors; unweighted concentration asks how unevenly that text is allocated; the ICD asks how concentration changes when mentor influence is incorporated. These are different empirical objects.

This distinction changes the interpretation of the years 4–6 result. Both ICD sign categories are associated with new ties relative to the zero group. One possibility is that the contrast captures differences in the structure of the doctoral mentoring environment that become visible during subsequent network expansion. Another is that it reflects pre-existing publication activity, mentor-group size, or the special composition of the zero category. The result is informative as an observed pattern, while the explanations remain open to tests that link acknowledged mentors directly to later coauthors.

6.3 Implications for scientometric research and research support

Acknowledgments broaden the observational boundary of scientometrics by representing assistance that is not recorded in author lists. Combining them with publication histories can reveal whether different kinds of recognized support are associated with different features of later scientific activity. This complements the integration of authorship and acknowledgment data in Paul-Hus et al. (2017) and extends the dissertation-network approach of Seckin and Varol (2022) toward the formation and recurrence of ties over the first post-Ph.D. decade. The contribution is a more differentiated account of relational information, rather than a replacement for conventional bibliometric measures.

The results also encourage a broader view of doctoral and early-career support. If family circumstances partly account for differences in the feasibility of collaboration, institutions have reason to examine access to research time, mobility assistance, caring support, and reliable mentoring. These are candidate interventions motivated by a possible constraint, not policies whose effects have been estimated here. The mixed evidence across sources of support in Cao et al. (2024) further cautions against assuming that family, supervisors, and institutions are interchangeable or that additional support produces the same outcome in every setting.

The appropriate research use is to understand environments and opportunities. Acknowledgment language would be a poor basis for evaluating individual merit, hiring, or allocating grants. People

differ in their willingness to disclose private relationships, and some have little family support to acknowledge. Treating those differences as deficiencies could reproduce the inequalities that the analysis seeks to understand. Institutional advantages in research production also persist through mechanisms that acknowledgment text cannot summarize, including access to academic labor (Zhang et al., 2022). A richer measurement of support should improve explanations of careers without turning private expressions of gratitude into performance targets.

6.4 Limitations and directions for future research

Several limitations define the scope of the findings and motivate further research. Acknowledgment language and later collaboration may share unobserved determinants, including intrinsic motivation, personality, prior research experience, and informal mentoring quality. These attributes are not measured by the present text indicators. Later employment and promotion controls also capture circumstances that may develop alongside collaboration, rather than fully accounting for initial differences among scholars. Combining acknowledgments with surveys and pre-graduation histories could help distinguish support received from dispositions and opportunities. Experimental or quasi-experimental evaluations of mentoring and research-support programs could examine whether changes in available support alter collaboration. Such designs would identify the effects of those programs, while the present OLS estimates describe conditional associations.

The sample also covers a particular academic and linguistic setting: U.S.-trained business scholars with sufficiently complete records and an observable academic career. The patterns may differ for graduates who enter industry, leave academia, or have less visible careers. Acknowledgment conventions can also vary across disciplines, countries, and languages (Hyland, 2003; Smirnova & Mayr, 2023). Naming relatives, expressing warmth, and thanking senior academics may carry different meanings across communities. Comparative research could establish which relationships are shared across settings and which depend on local conventions of recognition.

A longitudinal perspective would extend the analysis beyond a single acknowledgment at doctoral completion. As researchers become collaborators, supervisors, team leaders, and managers, they move from receiving support to providing and organizing it. Following acknowledged relationships alongside coauthorship and supervision histories could reveal which ties persist, change function, or give way to new relationships. This would help distinguish continuing reliance on doctoral networks from the development of wider professional roles. It would also contribute to a behavioral

science of science concerned with how support, recognition, and responsibility evolve across scientific careers.

Richer language methods offer a complementary direction. Dictionary counts capture broad positivity but cannot reliably distinguish warmth, respect, deference, or obligation. Contextual natural-language-processing methods, including large language models adapted to scholarly texts and validated against human annotations, could help distinguish these meanings and resolve ambiguous references to people and shared contributions (Feuerriegel et al., 2025). Repeated writing samples could also distinguish relatively stable expression from responses to changing circumstances. Seraj et al. (2021), for example, document language changes around romantic relationship dissolution. Although their setting differs from academic acknowledgment, their longitudinal approach illustrates why language should be studied in relation to changing relationships rather than treated only as a fixed personal characteristic.

A further extension concerns communication beyond written acknowledgments. Public talks and interviews could allow researchers to study verbal content alongside vocal and visual features, and to examine how audiences respond to scholars' presentation. Bi et al. (2020) report field-dependent associations between facial attractiveness and public-speaking fees, raising questions about appearance and social evaluation. Their findings do not measure emotional intent, but they motivate a broader distinction between how scholars communicate and how they are recognized. Connecting these forms of communication to collaboration would extend the study of relational expression while keeping observable behavior separate from assumptions about underlying feelings or personality.

7 Conclusions

Dissertation acknowledgments provide a view of the relationships surrounding researchers before much of their subsequent publication history has unfolded. In this sample of business scholars, family acknowledgment counts and text volume are positively associated with later new collaboration, with positive but less uniformly precise associations for repeated collaboration. Mentor acknowledgment levels and citation-weighted text allocation show different patterns. The latter is associated with new ties mainly in years 4–6 and with selected research indicators, but its sign cannot be interpreted as a preference for prestigious mentors.

The broader contribution is to connect research on invisible support and symbolic recognition with the formation and recurrence of observed coauthorship. The findings encourage scientometric

accounts that examine the conditions under which researchers participate in collaboration alongside the publications produced through it. Their value lies in identifying relationships to explain, rather than estimating causal returns to support. Following those relationships as scholars become collaborators, supervisors, and team leaders offers a constructive next step: a behavioral account of how recognized support at career entry relates to the collective organization of science over time.

References

Baccini, A., & Petrovich, E. (2022). Normative versus strategic accounts of acknowledgment data: The case of the top-five journals of economics. *Scientometrics, 127*(1), 603–635. https://doi.org/10.1007/s11192-021-04185-6

Belley, P., & Lochner, L. (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital, 1*(1), 37–89. https://doi.org/10.1086/524674

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review, 96*(5), 1652–1678. https://doi.org/10.1257/aer.96.5.1652

Bi, W., Chan, H. F., & Torgler, B. (2020). "Beauty" premium for social scientists but "unattractiveness" premium for natural scientists in the public speaking market. *Humanities and Social Sciences Communications, 7*, 118. https://doi.org/10.1057/s41599-020-00608-6

Bitzer, E., & Leshem, S. (2021). The invisible support networks of doctoral candidates: What acknowledgement sections of doctoral theses reveal. *South African Journal of Higher Education, 35*(3), 1–12. https://doi.org/10.20853/35-3-3538

Cainelli, G., Maggioni, M. A., Uberti, T. E., & de Felice, A. (2015). The strength of strong ties: How co-authorship affect productivity of academic economists? *Scientometrics, 102*(1), 673–699. https://doi.org/10.1007/s11192-014-1421-5

Cao, F., Li, H., Chen, X., You, Y., & Xue, Y. (2024). Who matters and why? The contributions of different sources of social support to doctoral students' academic engagement. *European Journal of Education, 59*(3), e12649. https://doi.org/10.1111/ejed.12649

Carpenter, J., & Seki, E. (2011). Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay. *Economic Inquiry, 49*(2), 612–630. https://doi.org/10.1111/j.1465-7295.2009.00268.x

Caucutt, E. M., & Lochner, L. (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy, 128*(3), 1065–1147. https://doi.org/10.1086/704759

Chan, H. F., Frey, B. S., Gallus, J., Schaffner, M., Torgler, B., & Whyte, S. (2014). Do the best scholars attract the highest speaking fees? An exploration of internal and external influence. *Scientometrics, 101*(1), 793–817. https://doi.org/10.1007/s11192-014-1379-3

Chan, H. F., Frey, B. S., Gallus, J., Schaffner, M., Torgler, B., & Whyte, S. (2016). External influence as an indicator of scholarly importance. *CESifo Economic Studies, 62*(1), 170–195. https://doi.org/10.1093/cesifo/ifv010

Cheng, W. (2022). Productivity spillovers in endogenous coauthor networks. *Empirical Economics, 63*(6), 3159–3183. https://doi.org/10.1007/s00181-022-02235-3

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin, 98*(2), 310–357. https://doi.org/10.1037/0033-2909.98.2.310

Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics, 77*(3), 385–407. https://doi.org/10.1111/obes.12070

Estrada, M., Zhi, Q., Nwankwo, E., & Gershon, R. (2019). The influence of social supports on graduate student persistence in biomedical fields. *CBE—Life Sciences Education, 18*(3), ar39. https://doi.org/10.1187/cbe.19-01-0029

Feuerriegel, S., Maarouf, A., Bär, D., Geissler, D., Schweisthal, J., Pröllochs, N., Robertson, C. E., Rathje, S., Hartmann, J., Mohammad, S. M., Netzer, O., Siegel, A. A., Plank, B., & Van Bavel, J. J. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology, 4*(2), 96–111. https://doi.org/10.1038/s44159-024-00392-z

García-Suaza, A., Otero, J., & Winkelmann, R. (2020). Predicting early career productivity of PhD economists: Does advisor-match matter? *Scientometrics, 122*(1), 429–449. https://doi.org/10.1007/s11192-019-03277-8

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature, 57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Hyland, K. (2003). Dissertation acknowledgements. *Written Communication, 20*(3), 242–268. https://doi.org/10.1177/0741088303257276

Jones, B. F. (2009). The burden of knowledge and the "death of the Renaissance man": Is innovation getting harder? *Review of Economic Studies, 76*(1), 283–317. https://doi.org/10.1111/j.1467-937X.2008.00531.x

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy, 128*(2), 434–467. https://doi.org/10.1086/704386

Leahey, E. (2016). From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annual Review of Sociology, 42*(1), 81–100. https://doi.org/10.1146/annurev-soc-081715-074219

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604. https://doi.org/10.1111/J.1468-2958.2002.TB00826.X

Lou, W., He, J., Zhang, L., Zhu, Z., & Zhu, Y. (2023). Support behind the scenes: the relationship between acknowledgement, coauthor, and citation in Nobel articles. *Scientometrics, 128*(10), 5767–5790. https://doi.org/10.1007/s11192-023-04803-5

McCain, K. W. (2024). Collaboration at the phylum level: Coauthorship and acknowledgment patterns in the world of the water bears (phylum Tardigrada). *Scientometrics, 129*(10), 6089–6125. https://doi.org/10.1007/s11192-024-05036-w

Mingers, J., & Yang, L. (2017). Evaluating journal quality: A review of journal citation indicators and ranking in business and management. *European Journal of Operational Research, 257*(1), 323–337. https://doi.org/10.1016/j.ejor.2016.07.058

Paul-Hus, A., Mongeon, P., Sainte-Marie, M., & Larivière, V. (2017). The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements. *Journal of Informetrics, 11*(1), 80–87. https://doi.org/10.1016/j.joi.2016.11.005

Petersen, A. M. (2015). Quantifying the impact of weak, strong, and super ties in scientific careers. *Proceedings of the National Academy of Sciences, 112*(34), E4671–E4680. https://doi.org/10.1073/pnas.1501444112

Petrovich, E., Allevi, S., Fazzini, E., & Gandolfi, L. (2026). SOPHIS: An open-access database for quantitative studies of contemporary philosophy of science. *Scientometrics, 131*, 1947–1978. https://doi.org/10.1007/s11192-026-05591-4

Rose, M. E., & Georg, C.-P. (2021). What 5,000 acknowledgements tell us about informal collaboration in financial economics. *Research Policy, 50*(6), 104236. https://doi.org/10.1016/j.respol.2021.104236

Seckin, O. C., & Varol, O. (2022). Academic support network reflects doctoral experience and productivity. *EPJ Data Science, 11*(1), 57. https://doi.org/10.1140/epjds/s13688-022-00369-z

Seraj, S., Blackburn, K. G., & Pennebaker, J. W. (2021). Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences, 118*(7), e2017154118. https://doi.org/10.1073/pnas.2017154118

Smirnova, N., & Mayr, P. (2023). A comprehensive analysis of acknowledgement texts in Web of Science: a case study on four scientific domains. *Scientometrics, 128*(1), 709–734. https://doi.org/10.1007/s11192-022-04554-9

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy, 46*(8), 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006

Wapman, K. H., Zhang, S., Clauset, A., & Larremore, D. B. (2022). Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature, 610*(7930), 120–127. https://doi.org/10.1038/s41586-022-05222-x

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

Zhang, S., Wapman, K. H., Larremore, D. B., & Clauset, A. (2022). Labor advantages drive the greater productivity of faculty at elite universities. *Science Advances, 8*(46), eabq7056. https://doi.org/10.1126/sciadv.abq7056

Appendix A Additional collaboration estimates

Table A1 reports the citation-based ICD category coefficients for repeated collaboration, using the same total-word and positive-word constructions as Table 2. They provide a comparison with new ties while retaining the exploratory interpretation of the category contrasts.

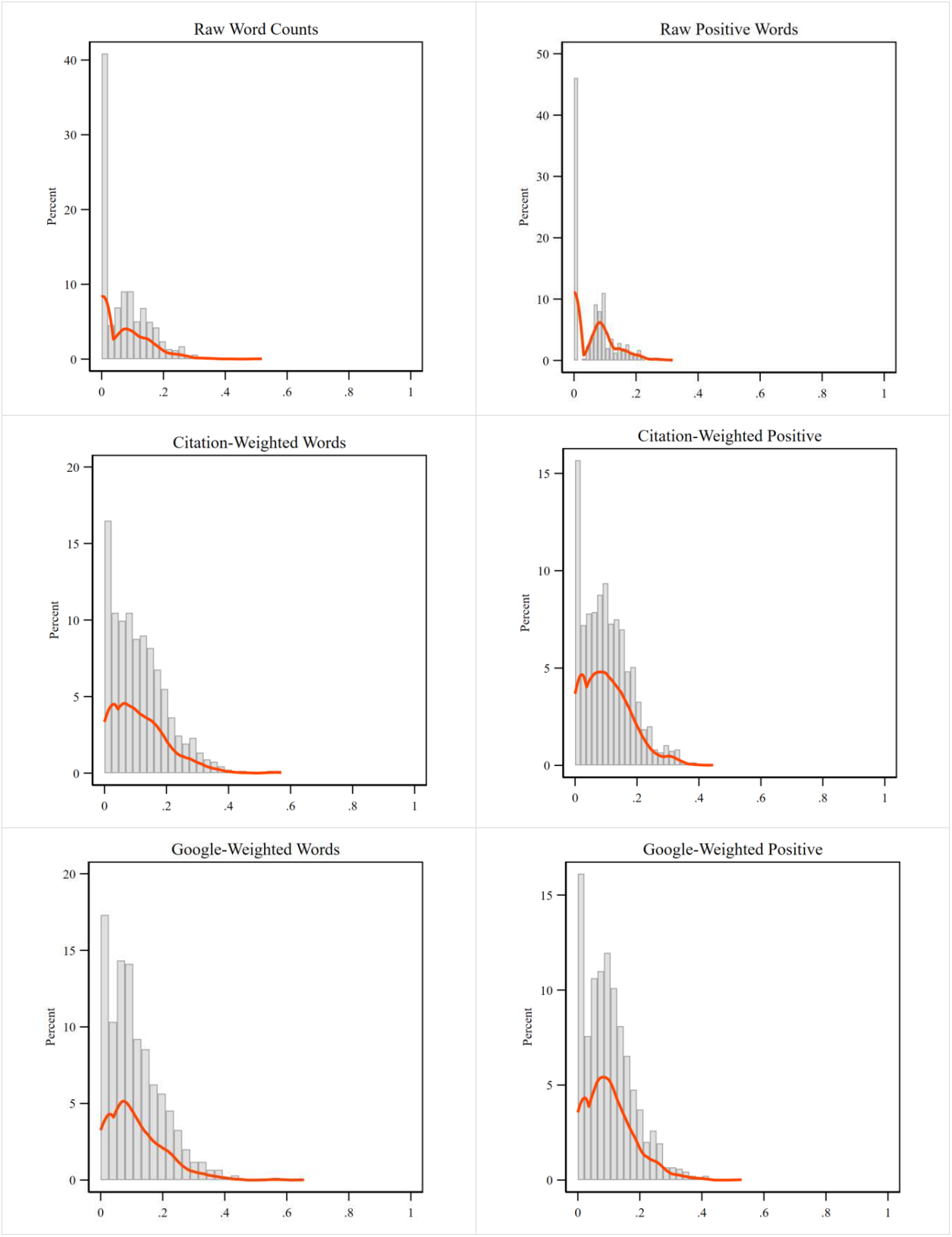
Table A1 Citation-based ICD categories and repeated collaboration

ICD construction	ICD category	Years 1–3	Years 4–6	Years 7–10
Gini · total words	ICD < 0	0.0208 (0.236)	−0.335 (0.364)	0.108 (0.478)
Gini · total words	ICD > 0	−0.137 (0.188)	−0.303 (0.319)	0.119 (0.363)
Theil T · total words	ICD < 0	0.00711 (0.226)	−0.369 (0.341)	0.0267 (0.413)
Theil T · total words	ICD > 0	−0.147 (0.188)	−0.292 (0.321)	0.136 (0.367)
Gini · positive words	ICD < 0	0.00825 (0.227)	−0.34 (0.348)	0.216 (0.474)
Gini · positive words	ICD > 0	−0.129 (0.191)	−0.287 (0.322)	0.116 (0.368)
Theil T · positive words	ICD < 0	0.155 (0.252)	−0.119 (0.365)	0.201 (0.445)
Theil T · positive words	ICD > 0	−0.148 (0.190)	−0.317 (0.321)	0.119 (0.369)
Scholars (N)		1,343	1,343	1,343
Background controls		Yes	Yes	Yes
Lagged career controls		No	Yes	Yes

Note. OLS coefficients for repeated collaboration; robust standard errors in parentheses. ICD is the influence-weighting concentration difference using mentor citations. Both sign categories enter jointly; the reference is ICD = 0 within the same construction. Each construction and interval is a separate regression. Total-word models include family total words; positive-word models include family positive words. Background controls are included throughout, with lagged career controls in later intervals (Section 4.1). Years are measured after Ph.D. completion. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Symbols use approximate two-sided normal tests based on the displayed coefficients and robust standard errors.

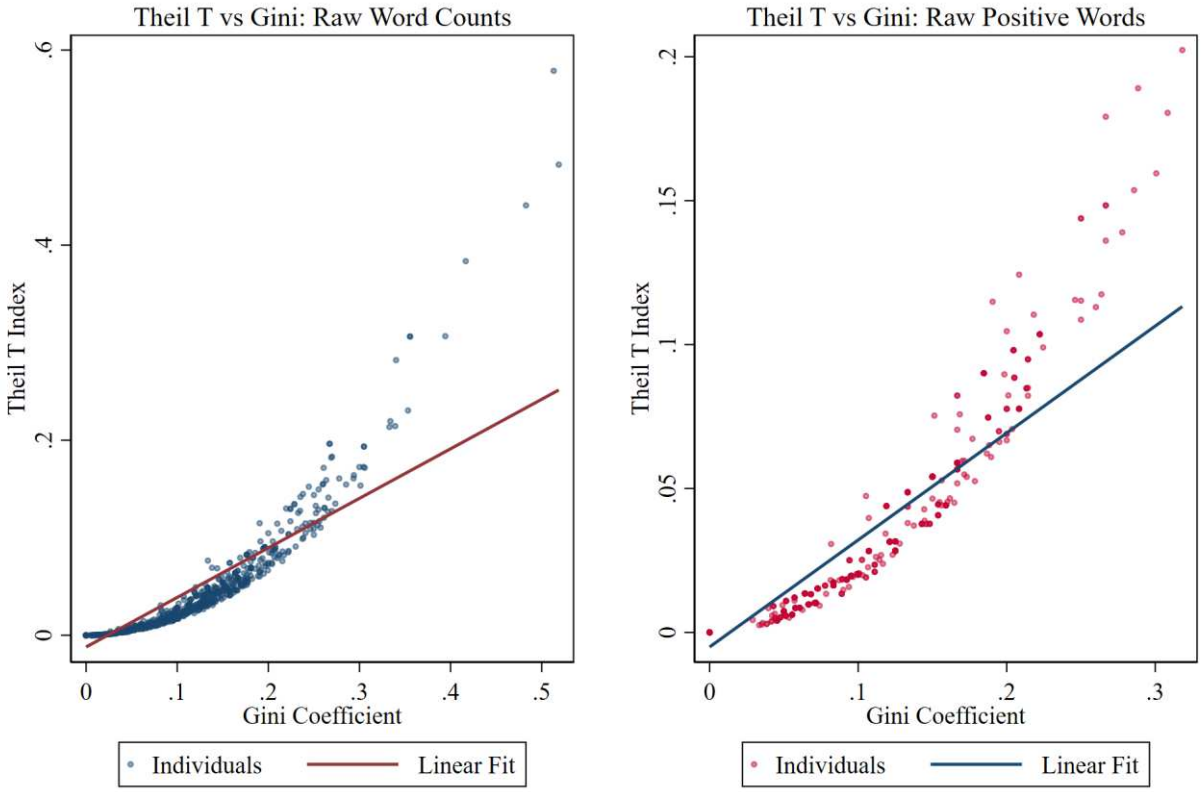
Appendix B Supplementary descriptive figures

Fig. 3 Distributions of acknowledgment-derived measures



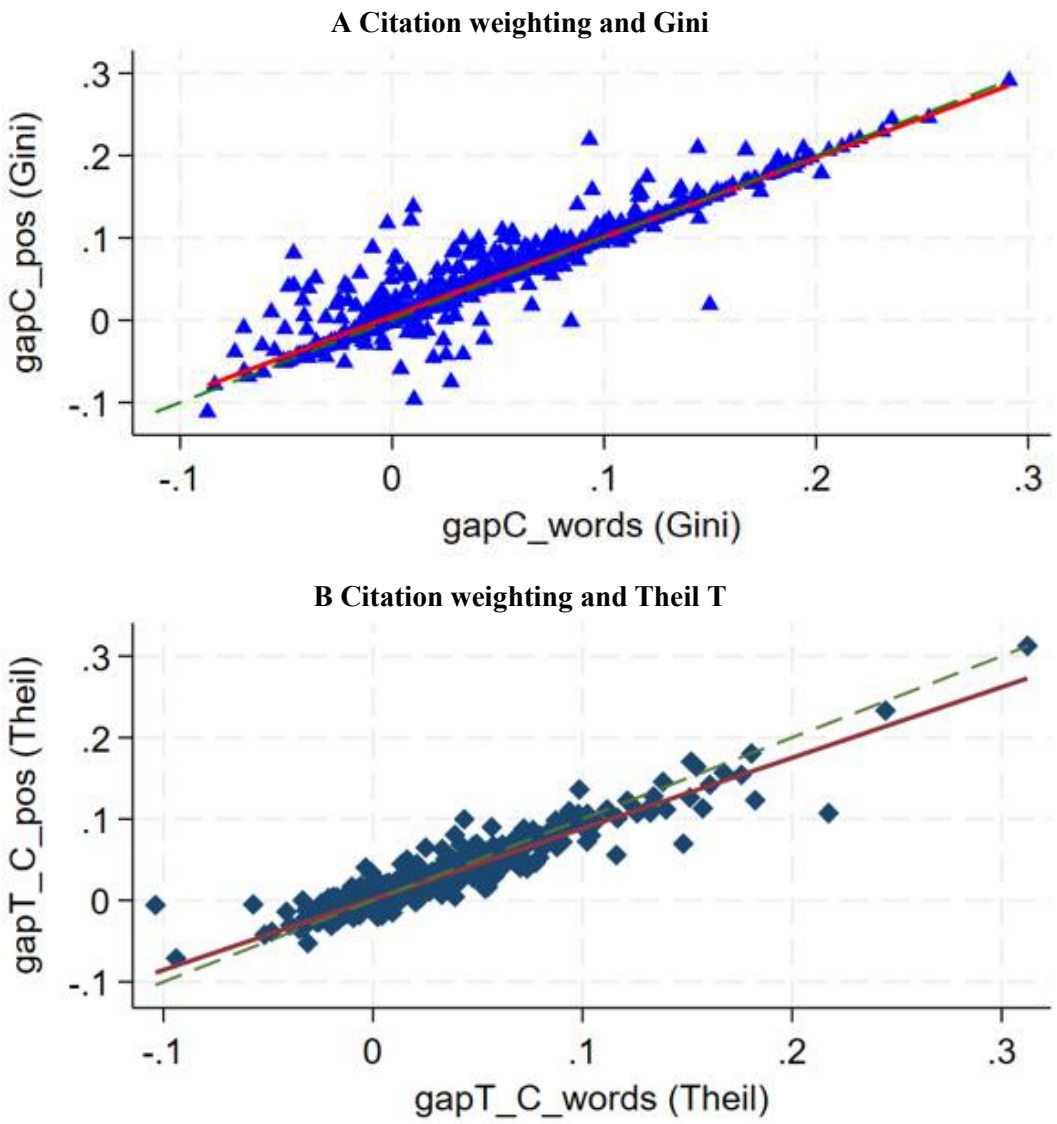
Note. Rows show unweighted, citation-weighted, and Google-weighted measures; columns show total and positive words. Bars and orange overlays retain the original scales, which should not be read as literal word counts. For Figures 3–5, plotted units and sample sizes have not been reconciled with the 1,343-scholar analytical sample. These figures provide descriptive context and do not validate the author-level construction or regression findings. Section 3.5 discusses the Google measures.

Fig. 4 Gini and Theil T measures of acknowledgment concentration



Note. Each point pairs the two concentration statistics in the original graphical output, separately for total words and positive words. The panels use different axis ranges. Solid lines are descriptive linear fits; the observed relationship is nonlinear. Agreement between statistics derived from the same text does not independently validate extraction, establish a psychological trait, or confirm the outcome regressions. The observation-unit and sample qualifications in the note to Figure 3 apply.

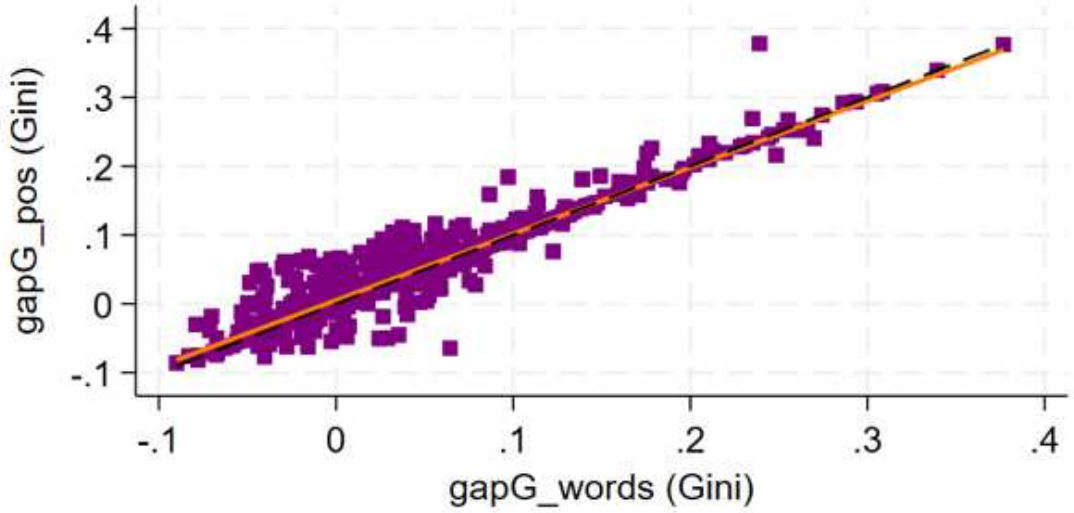
Fig. 5 Concentration differences from total and positive words



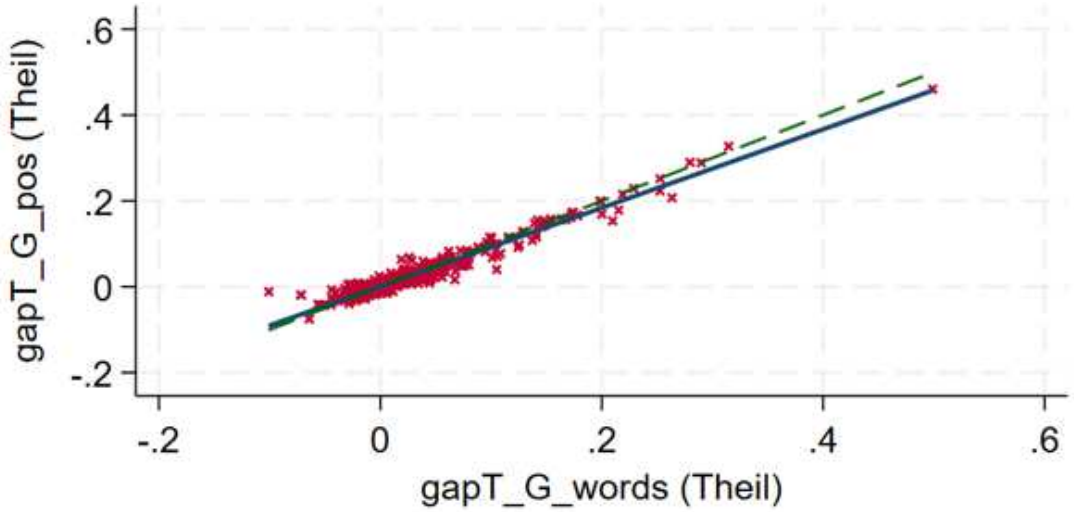
Note. Horizontal axes show differences constructed from total words; vertical axes show the corresponding positive-word differences. Green dashed lines denote numerical equality; solid red lines are fitted relationships. Source labels gapC and gapT_C denote the citation-based Gini and Theil differences. Panels C and D on the next page use Google weights. These ICD differences do not identify prestige preferences. The observation-unit and sample qualifications in the note to Figure 3 apply.

Fig. 5 continued Google-weighted concentration differences

C Google weighting and Gini



D Google weighting and Theil T



Note. Horizontal axes show total-word differences; vertical axes show positive-word differences. Dashed reference lines indicate numerical equality between the two measures; solid lines show fitted linear relationships. Source labels gapG and gapT_G denote the Google-based Gini and Theil differences. The Google measures capture web visibility, subject to the methodological limitations in Section 3.5. These panels do not establish robustness of the outcome regressions. The observation-unit and sample qualifications in the note to Figure 3 apply.