Econometrica Supplementary Material

SUPPLEMENT TO "THE WANDERING SCHOLARS: UNDERSTANDING THE HETEROGENEITY OF UNIVERSITY COMMERCIALIZATION"

JOSH LERNER
Harvard Business School, Harvard University

HENRY MANLEY
Department of Economics, Stanford University

CAROLYN STEIN
Haas School of Business and Department of Economics, UC Berkeley

HEIDI WILLIAMS
Department of Economics, Dartmouth College

This appendix supplements Lerner, Manley, Stein, and Williams (2026). Numbers prefixed with an "S" refer to this appendix; numbers without a prefix refer to the main article.

S1. DETAILS ON SAMPLE CONSTRUCTION

To increase the count of movers in our sample, we follow a multistep procedure illustrated in Figure S1. First, we code 12,235 movers using an author-year panel of corresponding e-mail address stems. Described in Section 2.4, it is a requirement that we observe one's move in the data to code them as a mover. That is, there cannot be any gap in their e-mail-based affiliation between the move. But in cases where there are only one to three years to fill between the move, we regard these authors as *movers* that we would like to add to our sample.[1] We incorporate additional information from the Web of Science (the preferred organization field, described in Section 2.3) to attempt to fill the affiliation gaps for these candidate movers. In some cases, this helps to tell us the exact year in which they leave their origin. If using organization data is not enough to close[2] the gap, we set the candidate mover aside for an additional imputation step described later.

Until this point, an author could only make it into our sample if they were ever listed as the corresponding author on a publication, and that publication shared their e-mail address. This approach to coding movers is helpful in bypassing the noise in the Web of Science organization field and the "multiple affiliations problem" described in Section 2.3. However, it is also quite restrictive and potentially misses out on movers that were never corresponding authors. More generally, one might also worry that using e-mail addresses to code movers is too narrow, since Figure S9 shows that even in 2020 less than a third of all publications link a corresponding email. To address this, and in hopes of adding to our sample, we examine 5,324,663 authors who never appear in our e-mail-based panel. We start by seeing if any of these authors could be added to our sample as-is; meaning, their existent organization record is enough to determine where they resided in each year. There are 1,347 movers and 165,332 non-movers that are added to our sample with this approach. The top right panel of Figure S1 shows this step.

[1] We might worry that an author moves more than once if there are many affiliation-years to impute.

[2] Here, "close" does not necessarily mean every missing affiliation year between the move is filled. Instead, it means the move year is observed in the data. Suppose we observe Josh Lerner at school A in 2012 and at school B in 2016, and so we need to back out where he was 2013-2015. Learning that Josh was at school A in 2015 or that he was at school B in 2013 is enough to know the year he moved in.

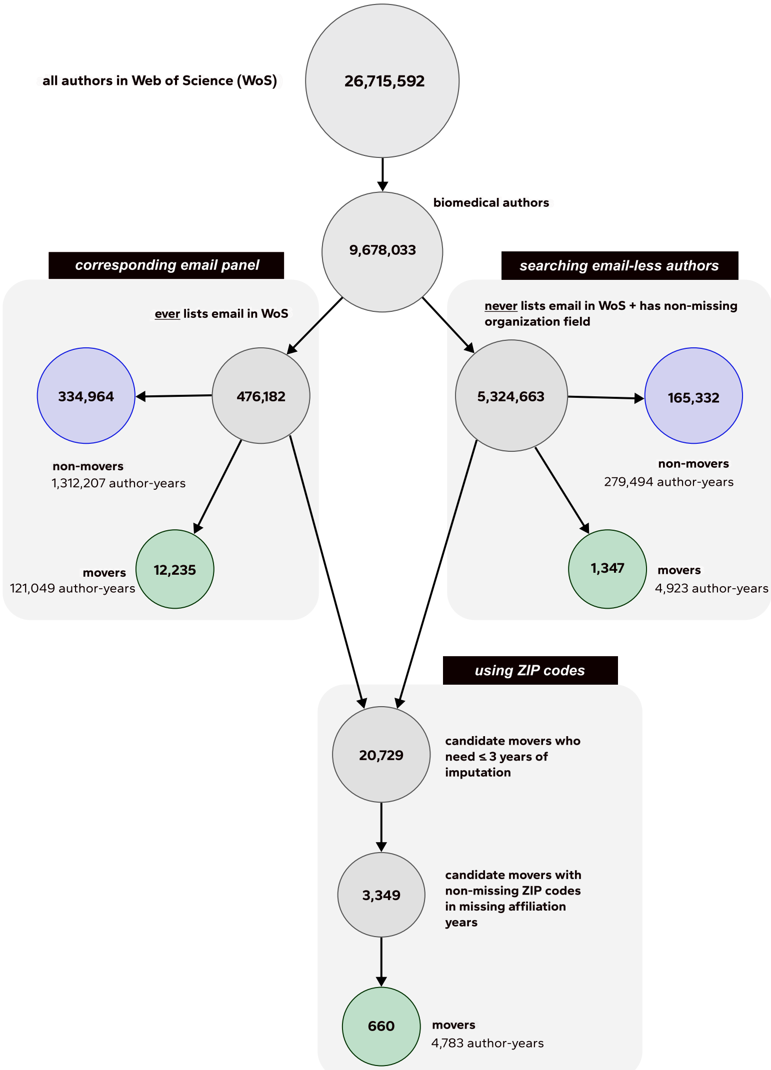


FIGURE S1.—Author-year panel construction

Notes: This diagram tracks the different ways we attempt to identify movers in our sample. This method begins with the author-year email skeleton shown in Table 1. The leftmost region shows the authors we capture from the basic email panel approach described in Section 2.4. The rightmost gray region displays the authors we add to our sample by examining authors that never list corresponding email addresses, and so would never appear in our email panel skeleton. The central gray region takes the "scraps," or authors difficult to code from each preceding procedure, and uses changes in ZIP codes to add additional movers to our sample. These approaches are described in Section S1. Summing across the values in the green nodes generates the total count of movers in our sample, 14,242. Summing across the values in the blue nodes equates the number of non-movers, 500,296. The same statements are true for the count of author-years, which are listed in the periphery of each colored node.

As before with the e-mail panel, there are some authors that we know are movers but cannot add to our sample because there is a gap in their organization record. As a final attempt to include these movers in our sample, we pool them with the candidate movers identified via the e-mail approach who also need additional imputation. Figure S1 shows this as two arrows (one from the top left e-mail based panel and one from the top right "email-less" panel) pointing into the same node in the bottom center panel. There are 20,729 candidate movers that we could add, at maximum.

The final step in our procedure is to use organization ZIP codes to fill affiliation gaps for the remaining 20,729 candidate movers. The idea is that if we know a mover's origin and destination ZIP codes, we can use a *change* in ZIP codes to code the move-year and solve the same imputation problem. In general, the Web of Science ZIP code data is noisy, often identifying multiple codes for an author within the same year. This is especially true around the time of the move, which makes it challenging to confidently discern which ZIP code change is the one that encodes a move. To get around this, we let the first year with non-missing ZIP code data correspond to the mover's origin and the last year of non-missing data correspond to their destination. For each year we need to fill an affiliation for, we then count the number of times there is a match between the origin ZIP codes and the ones listed between the move. We repeat this process for the destination ZIP codes. We then compare these counts. For example, if there are two ZIP code matches to one's origin and zero to one's destination, we code them to still be at their origin. In case of ties, and in accordance with move-timing work described in Section 2.4, we code them to be at their destination. Only 3,349 of the 20,729 candidate movers ever have a ZIP code, of which 660 are coded as movers. Taken together, this procedure codes $12,235 + 1,347 + 660 = 14,242$ movers and $334,964 + 165,332 = 500,296$ non-movers, corresponding to a total of 1,722,456 author-years.

S2. ADDITIVE SEPARABILITY VALIDATION

We perform two exercises to validate our assumption that log commercialization is additively separable, as assumed in Equation (2).

S2.1. *Symmetry (Card, Heining, and Kline 2013)*

Following the logic of Card, Heining, and Kline (2013), if the additive model

$$y_{it} = \alpha_i + \gamma_u + \varepsilon_{it}$$

is correct, then the effect on the outcome y of moving should be symmetric for two researchers moving in opposite directions. To see why, consider researcher i moving from u to u' and i' moving from u' to u:

$$y_{i2} - y_{i1} = (\gamma_{u'} - \gamma_u) + (\varepsilon_{i2} - \varepsilon_{i1})$$

$$y_{i'2} - y_{i'1} = (\gamma_u - \gamma_{u'}) + (\varepsilon_{i'2} - \varepsilon_{i'1})$$

If we average over all of the people moving from u to u' and u' to u, we get (assuming no serial correlation in the ε's):

$$E[\Delta y_{u \to u'}] = \gamma_{u'} - \gamma_u \text{ and } E[\Delta y_{u' \to u}] = \gamma_u - \gamma_{u'}$$

In other words, we get opposite and symmetric changes in the outcomes. Note that this is always true, even if very different groups of people (with different α's) are making the two different kinds of moves, *if additive separability holds*.

Now suppose additive separability does not hold. Instead, suppose the model has match effects between researchers and universities:

$$y_{it} = \alpha_i + \gamma_u + w(\alpha_i, \gamma_u) + \varepsilon_{it}$$

Again, consider worker i moving from u to u' and i' moving from u' to u:

$$y_{i2} - y_{i1} = (\gamma_{u'} - \gamma_u) + (w(\alpha_i, \gamma_{u'}) - w(\alpha_i, \gamma_u)) + (\varepsilon_{i2} - \varepsilon_{i1})$$

$$y_{i'2} - y_{i'1} = (\gamma_u - \gamma_{u'}) + (w(\alpha_{i'}, \gamma_u) - w(\alpha_{i'}, \gamma_{u'})) + (\varepsilon_{i'2} - \varepsilon_{i'1})$$

We can again take averages, but now the α terms no longer cancel. Let S be the set of workers who make $j \rightarrow j'$ moves and S' be the set of workers who make $j' \rightarrow j$ moves. Then taking averages over those sets, we have:

$$E[\Delta y_{u \rightarrow u'}] = (\gamma_{u'} - \gamma_u) + E_S[w(\alpha_i, \gamma_{u'}) - w(\alpha_i, \gamma_u)]$$

$$E[\Delta y_{u' \rightarrow u}] = (\gamma_u - \gamma_{u'}) + E_{S'}[w(\alpha_{i'}, \gamma_u) - w(\alpha_{i'}, \gamma_{u'})]$$

These are now no longer equal but opposite, because the $E_S[\cdot]$ and $E_{S'}[\cdot]$ values average over different sets of individuals with different values of α_i. Therefore, this suggests a natural test of the additive separability assumption: to examine whether equal and opposite moves lead to equal and opposite changes in our outcome measure.

Of course, we do not have enough data to investigate individual, opposite moves (i.e., Berkeley to Stanford versus Stanford to Berkeley). So again following Card, Heining, and Kline (2013), we divide universities into quartiles based on their average commercialization measure. This leaves us with 16 possible types of moves. In Figure S2, we compare the symmetry of researchers making opposite moves, focusing on the most “extreme" of these moves. The results are shown below. The sample sizes for each move type are relatively small (in the 100-200 person range for each move type). Still, the changes in the outcome upon moving are quite symmetric, consistent with additive separability being a reasonably good model.

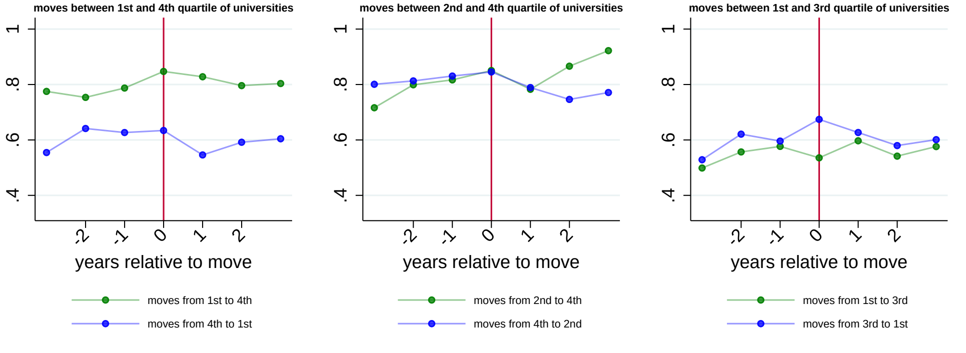


FIGURE S2.—Move symmetry

Notes: This figure shows the mean of IHS of the predicted patent citations by year for individuals who move from different types of universities, following Card, Heining, and Kline (2013). We restrict to a balanced panel.

S2.2. *Estimating complementarities (Bonhomme, Lamadon, and Manresa 2019)*

Next, we try to estimate any complementarities directly, following Bonhomme, Lamadon, and Manresa (2019) (hereafter BLM).

We begin by constructing our sample. BLM use data from the entire working-age population in Sweden. They have 19,557 movers and 580,218 stayers in the 2002-2004 time period. They estimate a two-period model where 2002 corresponds to $t = -1$ and 2004 corresponds to $t = 1$. Recall that we have 14,242 movers in total, but these moves occur across many different years. In an effort to preserve power, we drop the calendar time dimension in our data (in other words, moves in all years in our sample are combined). Still, we can only use movers who we observe for a year before and a year after their move. This leaves us with 8,643 movers. To construct stayers, we randomly assign placebo "move years" to our stayers, drawing from the distribution of real moves. Then we keep stayers that we observe on either side of the placebo move year. This leaves us with 37,790 stayers. Therefore, we have about 44% of the movers and 7% of the stayers that the original authors have in their sample.

We then cluster the universities into 10 groups. Following BLM exactly, we use a k-means algorithm, clustering the universities (using only stayers) according to their outcome distributions (where our outcome is our commercialization measure). We match on 20 points along the distribution, following BLM.

With these clusters in hand, we estimate two regression specifications. For author i moving in relative year t (note that calendar time and relative time are now synonymous), we start by estimating the linear regression:

$$y_{it} = \alpha_i + \gamma_{k(i,t)} + \varepsilon_{it}$$

where α_i is an individual fixed effect and γ_k is a university in group k fixed effect. This is a two-period model, so $t \in \{-1, 1\}$. This is similar to the model that we estimate in the main text, in the sense that it does not allow for researcher-university complementarities. However, it is worth noting that it is exactly not the same: it is estimated on a smaller sample, does not allow for calendar year effects, only uses two periods, and groups universities into ten different groups. Thus, we do not attempt to make direct comparisons between the results of this model and the model estimated in the main text.

Next, we estimate a model that allows for researcher-university interactions:

$$y_{it} = \alpha_i \cdot b_{k(i,t)} + \gamma_{k(i,t)} + \varepsilon_{it}.$$

The key difference is the interaction term $\alpha_i \cdot b_{k(i,t)}$, explicitly allowing for match effects between individuals and university groups. Following the original authors, the model is estimated using limited information maximum likelihood (LIML).

A variance decomposition from both models is shown below in Table S1. The key takeaway is that the two models appear to offer very similar conclusions, suggesting that researcher-university complementarities are not substantively important. The fact that the R^2 values are so similar suggests that the complementarities add very little explanatory power.[3] This is consistent our Card, Heining, and Kline (2013) diagnostics above. It is also consistent with what Bonhomme, Lamadon, and Manresa (2019) find in their setting.

[3]Careful readers might note that the R^2 of the linear model is actually *higher* than that of the interacted model, which seems implausible. However, this happens due to the different estimation methodologies for the two models (OLS vs. LIML). Our overall takeway is simply that the R^2 values are very similar.

TABLE S1

LINEAR VS. INTERACTED MODEL (BLM)

Model	$\frac{Var(\hat{\alpha})}{Var(y)}$	$\frac{Var(\hat{\gamma})}{Var(y)}$	$\frac{2Cov(\hat{\alpha},\hat{\gamma})}{Var(y)}$	$\frac{Var(\varepsilon)}{Var(y)}$	$Corr(\hat{\alpha},\hat{\gamma})$	R^2
Linear	68.41	0.17	1.36	30.06	0.20	0.70
Interacted	65.00	0.22	1.73	33.04	0.23	0.67

Notes: This table compares the variance decomposition from a linear model and an interacted model, following the methodology laid out above and in Bonhomme, Lamadon, and Manresa (2019). The first four columns show the percent of the total variance explained by each constituent part. The fifth column shows the correlation between the estimated researcher and university effects. The sixth column shows the R^2. N = 46,433 researchers, 8,643 of which are movers.

S3. VARIANCE DECOMPOSITION

Like in Finkelstein, Gentzkow, and Williams (2016), we provide results from a variance decomposition of our main outcome: the IHS of predicted five-year patent citations. To begin with, we use the estimates generated by Equation (2)— $\hat{y}_{it}$, $\hat{\gamma}_u$, and $\hat{\alpha}_i$— and average each to the university-year level, and then again to the university level. This produces a dataset with $\hat{y}_u$, $\hat{\gamma}_u$, and $\hat{\alpha}_u$ for each university with a non-zero count of movers ($N = 253$). The share of cross-university variance in commercialization that would be eliminated in a counterfactual where average researcher effects were equalized across universities is

$$S^{var}_{researcher} = 1 - \frac{Var(\hat{\gamma}_u)}{Var(\hat{y}_u)}. \tag{S1}$$

Similarly, the share eliminated if university effects were equalized is

$$S^{var}_{university} = 1 - \frac{Var(\hat{\alpha}_u)}{Var(\hat{y}_u)}. \tag{S2}$$

Note, unlike $S_{researcher}$ and $S_{university}$, $S^{var}_{researcher}$ and $S^{var}_{university}$ are not additive, meaning they will not sum to one so long as $Cov(\hat{\alpha}_u, \hat{\gamma}_u) > 0$. One additional caveat is that because Equation (2) is identified via movers, our ability to precisely estimate γ_u depends on the number of movers observed at u. We want to minimize the extent to which cross-university variance in $\hat{y}_u$ is driven by noise resulting from few movers, or "limited mobility bias." In particular, we know from Equation (S1) that as $Var(\hat{\gamma}_u)$ increases, $S^{var}_{researcher}$ decreases. The same is true for $Cov(\hat{\alpha}_u, \hat{\gamma}_u)$, where for a fixed $Var(\hat{y}_u)$, noise generated by few movers will make $\hat{\gamma}_u$ either an over- or under-estimate of γ_u. And because we take university and researcher effects as additive, if $\hat{\gamma}_u > \gamma_u$ we know $\hat{\alpha}_u < \alpha_u$.

Empirically, Figure S25 shows that the number of movers varies across schools. To minimize noise in $\hat{\gamma}_u$, and in accordance with Andrews, Gill, Schank, and Upward (2012), we restrict the variance decomposition to include only universities with 25 movers.

The results from this exercise are shown in Table 4 and are generally consistent with estimates presented by the event study (Figure 6) and additive (Table 3) analogs. Specifically, it is estimated that about 17% of the cross-university variance in predicted citations is attributable to variance in university effects.

S4. EMPIRICAL BAYES ADJUSTMENT

One additional step we take to minimize noise in our estimate of $\hat{\gamma}_u$ is to employ an Empirical Bayes correction (Morris, 1983, Kline, Saggio, and Sølvsten, 2020). This approach pools information across universities to improve our estimated university fixed effects, while allowing for potential correlation in noise across universities, the latter of which can be important when using mover designs (Walters, 2024).

To begin, we assume that $\hat{\Gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_U)'$ is an unbiased estimate of the true university effects, Γ_u. A $U \times U$ matrix V describes the sampling variance of Γ_u, with squared standard error terms s_u^2 of the individual $\hat{\gamma}_u$'s along the diagonal. $\hat{V}$ is an unbiased estimator of this matrix.

Next, we assume that the latent parameters γ_u are randomly drawn from the distribution G_γ, which is defined in the population of universities and is normally distributed. Together, these assumptions yield the hierarchical model:

$$\begin{aligned} \hat{\gamma}_u | \gamma_u, s_u^2 &\sim N(\gamma_u, s_u^2) \\ \gamma_u | s_u^2 &\sim N(\mu_\gamma, \sigma_\gamma^2). \end{aligned} \tag{S3}$$

This model has two hyperparameters, μ_γ and σ_γ^2, which we estimate, following Walters (2024), as:

$$\hat{\mu}_\gamma = \frac{1}{U} \sum_{u=1}^{U} \hat{\gamma}_u, \qquad \hat{\sigma_\gamma}^2 = \hat{\Gamma}' A_0 \hat{\Gamma} - tr(A_0 \hat{V}). \tag{S4}$$

where

$$A_0 = \frac{1}{U-1} \left(I_U - \frac{1}{U} 1_U 1_U' \right) \tag{S5}$$

Once these two hyperparameters are estimated, the empirical Bayes method produces posteriors as:

$$\hat{\gamma}_u^{EB} = \hat{\mu}_\gamma \left(\frac{s_u^2}{s_u^2 + \sigma_\gamma^2} \right) + \hat{\gamma}_u \left(\frac{\sigma_\gamma^2}{s_u^2 + \sigma_\gamma^2} \right). \tag{S6}$$

Figure S3 shows the result of this adjustment on the distribution of $\hat{\gamma}_u$.

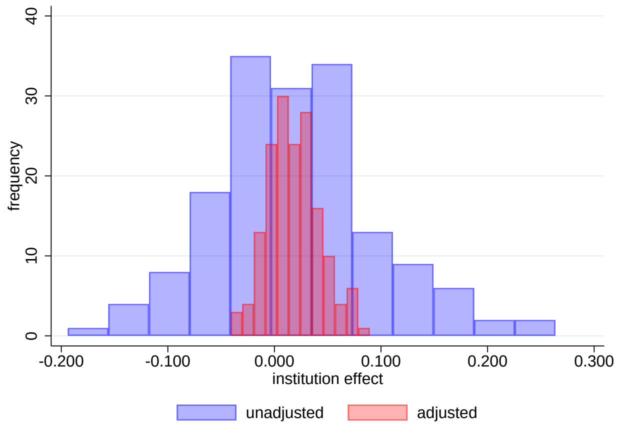


FIGURE S3.—Empirical Bayes: Distribution of university effects

Notes: This figure is based on estimation of Equation (2), where the dependent variable y_{iut} is the IHS of the predicted count of five-year patent citations to author i's papers in year t whose main academic affiliation is at university u. In particular, this figure displays the results of an Empirical Bayesian shrinkage of the causal university effects $(\hat{\gamma}_u)$. The x-axis is the estimated value of the university effect for a given school, and the y-axis is the frequency of schools within an university effect bin. The blue bars show the distribution of the university effects estimated from Equation (2). The red bars show the distribution of the adjusted university effects, following the Bayesian shrinkage $(\hat{\gamma}_u^{EB})$. To center the mean of this distribution near zero, we pre-omit Syracuse University when estimating Equation (2). While our full sample includes 266 universities, we restrict our attention to the 164 with at least 25 movers. This leaves us with 1,668,243 author-years.

S5. SUPPLEMENTAL FIGURES AND TABLES

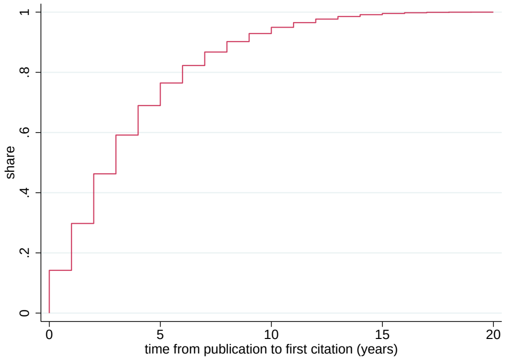
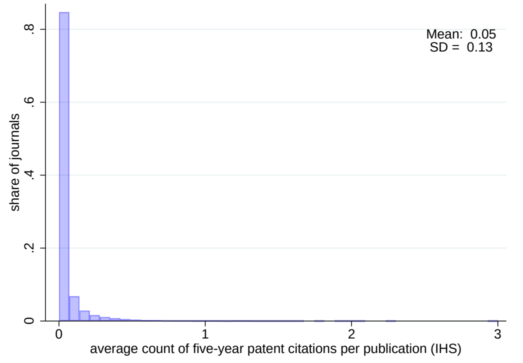


FIGURE S4.—Lag between paper publication and first patent citation

Notes: This plot shows the CDF of the time between a paper's publication and its first citation by a patent conditional on the paper being cited by at least one patent. Patent citations are dated to the year that the patent application was filed. The sample is all publications with at least one patent cite, written by authors in our analysis sample.



(a) Distribution of commercialization across journals

rank	*journal*	*publication count*	*mean patent citations*
1	2008 SID International Symposium, Technical Digest	545	3.00
2	2007 SID International Symposium, Technical Digest	478	2.27
3	Annual Review Of Immunology	270	2.26
4	IEE International Electron Devices Meeting 2000, Technical Digest	201	2.25
5	ACM Transactions On Computer Systems	226	2.09
6	mAbs	1220	2.00
7	IEEE International Electron Devices Meeting 2004, Technical Digest	251	1.97
8	Nature Biotechnology	8493	1.93
9	2007 IEEE International Electron Devices Meeting	240	1.91
10	IBM Systems Journal	502	1.78

(b) Ranked journals by commercialization propensity

FIGURE S5.—Commercialization among the most frequently cited journals

Notes: Panel (a) plots the distribution of average citations within five years of a paper's publication across scientific journals. The sample used to construct this distribution is the universe of papers in the Web of Science database with journal information provided. Panel (b) lists the ten journals whose publications have the highest average count of five-year patent citations.

(a) Original journal article

Programmable probiotics for detection of cancer in urine

Tal Danino,[1]* Arthur Prindle,[2]* Gabriel A. Kwong,[1†] Matthew Skalak,[1] Howard Li,[2] Kaitlin Allen,[1] Jeff Hasty,[2,3,4‡] Sangeeta N. Bhatia[1,5,6,7,8‡§]

[1]Institute for Medical Engineering and Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA. [3]BioCircuits Institute, University of California, San Diego, La Jolla, CA 92093, USA. [4]Molecular Biology Section, Division of Biological Science, University of California, San Diego, La Jolla, CA 92093, USA. [5]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. [6]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. [7]Electrical Engineering and Computer Science and David H. Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA 02139, USA. [8]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.
*Equally contributing lead authors.
†Present address: Wallace H. Coulter Department of Biomedical Engineering, Georgia Tech and Emory School of Medicine, Atlanta, GA 30332, USA.
‡Equally contributing senior authors.
§Corresponding author. E-mail: sbhatia@mit.edu

(b) Web of Science XML file


```
<name seq_no="8" role="author" reprint="Y" addr_no="1 5 6 7 8 9">
<display_name>Bhatia, Sangeeta N.</display_name>
<full_name>Bhatia, Sangeeta N.</full_name>
<wos_standard>Bhatia, SN</wos_standard>
<first_name>Sangeeta N.</first_name>
<last_name>Bhatia</last_name>
<email_addr>sbhatia@mit.edu</email_addr>
```


(c) SQL tables

wos_uid	address_spec_addr_no	organization
WOS:000355179800007	1	Massachusetts Institute of Technology (MIT)
WOS:000355179800007	2	University of California System
WOS:000355179800007	2	University of California San Diego
WOS:000355179800007	3	University of California System
WOS:000355179800007	3	University of California San Diego
WOS:000355179800007	4	University of California System
WOS:000355179800007	4	University of California San Diego
WOS:000355179800007	5	Harvard University
WOS:000355179800007	5	Massachusetts Institute of Technology (MIT)
WOS:000355179800007	5	Broad Institute
WOS:000355179800007	6	Harvard University
WOS:000355179800007	6	Brigham & Women's Hospital
WOS:000355179800007	7	Massachusetts Institute of Technology (MIT)
WOS:000355179800007	8	Massachusetts Institute of Technology (MIT)
WOS:000355179800007	9	Howard Hughes Medical Institute

wos_uid	record_order	reprint	email_addr	full_name
WOS:000355179800007	1			Danino, Tal
WOS:000355179800007	2			Prindle, Arthur
WOS:000355179800007	3			Kwong, Gabriel A.
WOS:000355179800007	4			Skalak, Matthew
WOS:000355179800007	5			Li, Howard
WOS:000355179800007	6			Allen, Kaitlin
WOS:000355179800007	7			Hasty, Jeff
WOS:000355179800007	8	Y	sbhatia@mit.edu	Bhatia, Sangeeta N.

FIGURE S6.—Measuring university affiliations of authors: An example

Notes: Example taken from Danino et al. (2015), Programmable Probiotics for Detection of Cancer in Urine.

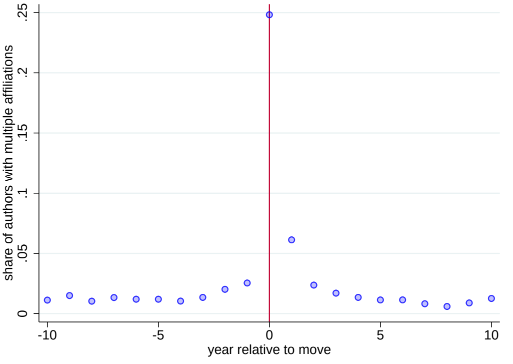


FIGURE S7.—Multiple affiliations in the WoS by relative year

Notes: This figure plots the share of authors whose publications, within a single publication year, list more than one affiliation. This share is aggregated across all authors by relative year. A red vertical line denotes the move year. The sample of all mover-years is used to calculate this share (N =119,705).

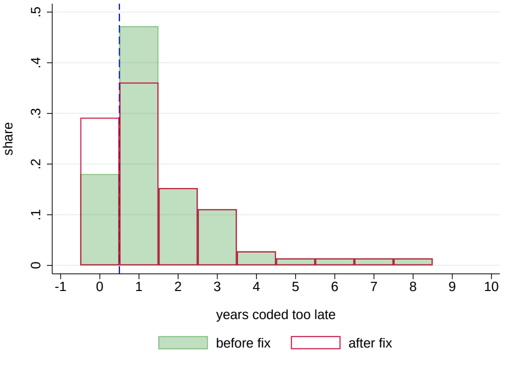
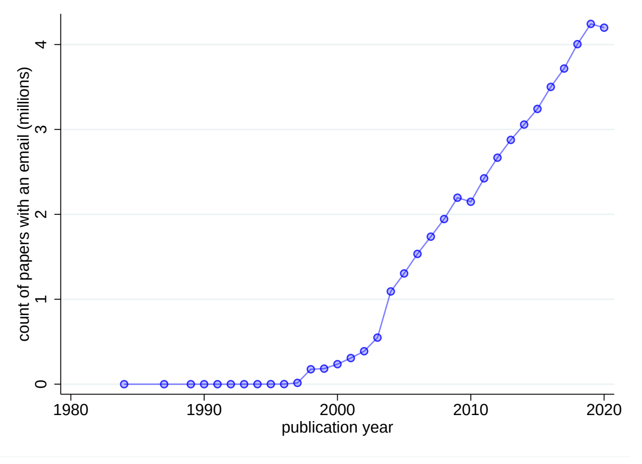
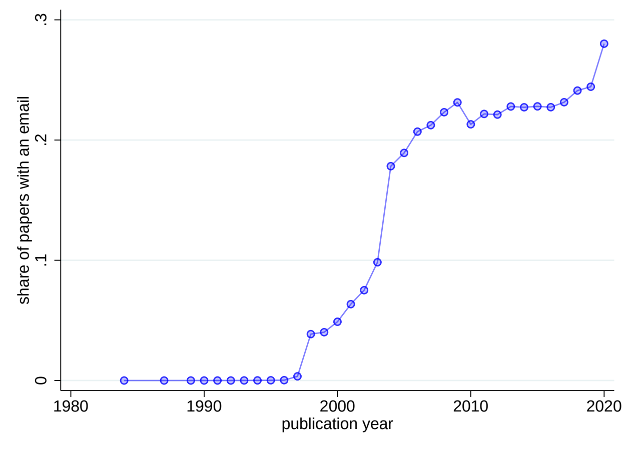


FIGURE S8.—Move timing correction

Notes: This figure plots the distribution of move-timing errors found in a hand-coded, randomly selected sample of 100 movers from our full sample of 14,242. By collecting public information from CVs, researcher LinkedIn profiles, and university lab webpages, we can observe the "ground truth" for the year in which a researcher moved. We use this ground truth and compare it against what we code from the Web of Science publication records. The x-axis is the difference between the year in which a researcher *actually* moves and the year in which we have coded them to move. The y-axis is the share of the hand-coded movers. Per the green bars, the modal move timing error is one year.



(a) Count



(b) Share

FIGURE S9.—Email prevalence over time

Notes: This figure illustrates the change in the propensity for publications recorded in the Web of Science to report a corresponding author e-mail address. Panel (a) plots the count (in millions) of publications with a corresponding email address by publication year (1984-2020). Panel (b) plots the share of publications with a corresponding email address over the same period. The sample used to construct these tabulations is the universe of publications in the Web of Science database from 1984 to 2020.

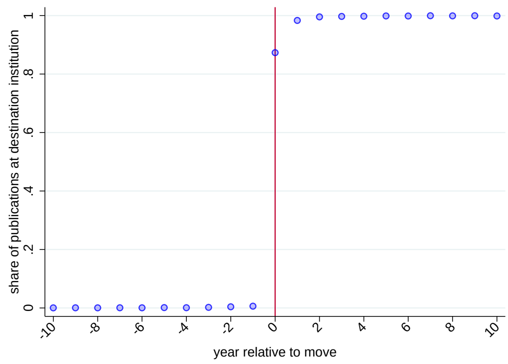
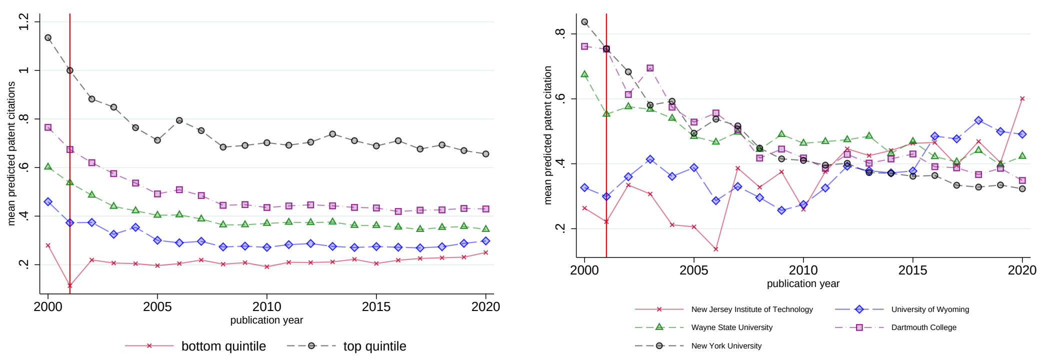


FIGURE S10.—Share of publications associated with destination university by relative year

Notes: This figure plots the average share of a mover's publications per relative year that list their destination university as their academic affiliation, restricting to affiliations that are either the destination or origin. This value is then averaged over all movers, by relative year. A red line is drawn for relative year zero, the move year, to indicate the year when there is expected to be a discontinuous increase in the share of publications associated with a mover's destination. The sample is the full sample of movers (N = 14,242).



(a) Quintiles over time

(b) Sampled university within quintile over time

FIGURE S11.—Stability in commercialization quintiles over time

Notes: Panel (a) plots quintiles of university-level IHS of predicted patent citations over publication years from 2000 to 2020. These quintiles are calculated by estimating the distribution of the 2001 university-level average count of predicted patent citations per publication written by authors at a given TTO. Mean counts of predicted patent citations are then calculated over the author-years assigned to each quintile, by publication year. These quintile means are calculated using both movers and non-movers, totaling 514,538 authors. Panel (b) shows the average count of predicted citations for a single, randomly sampled TTO within each quintile over time. Only TTOs that appear in every year are eligible to be sampled. The quintile for which a TTO is sampled from is denoted by the pattern and color of its time series– for example, the New Jersey Institute of Technology is in the lowest quintile and New York University is in the highest. In each panel, a red vertical line denotes 2001 as the publication year used to construct these quintiles.

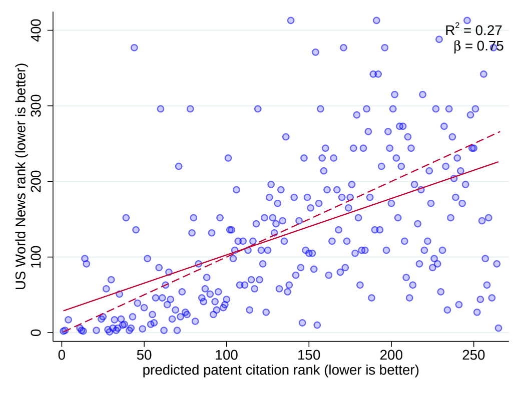


FIGURE S12.—U.S. News and World Report ranking vs. commercialization rank

Notes: This figure plots university-level commercialization rank against *U.S. News and World Report*'s (USNWR) 2025 university rank. To match more TTOs to USNWR ranks, hospitals and institutes are assigned to their parent universities where appropriate. University systems are assigned to their highest-ranking campus. Commercialization rank is calculated as the average count of predicted citations that researchers affiliated with a given university receive, across all publication years in our sample (from 2000 to 2020). A solid red line displays the slope coefficient from a regression of commercialization rank on U.S. News rank; the value of the slope and R^2 are recorded in the top right of the figure. A dashed red line shows the 45^o line along these two axes. There are five schools labeled by name of the graph, each with various ranks. While our full sample spans researchers from 266 universities, only 223 can be linked back to USNWR rankings.

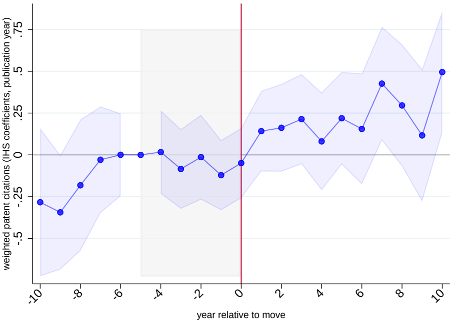
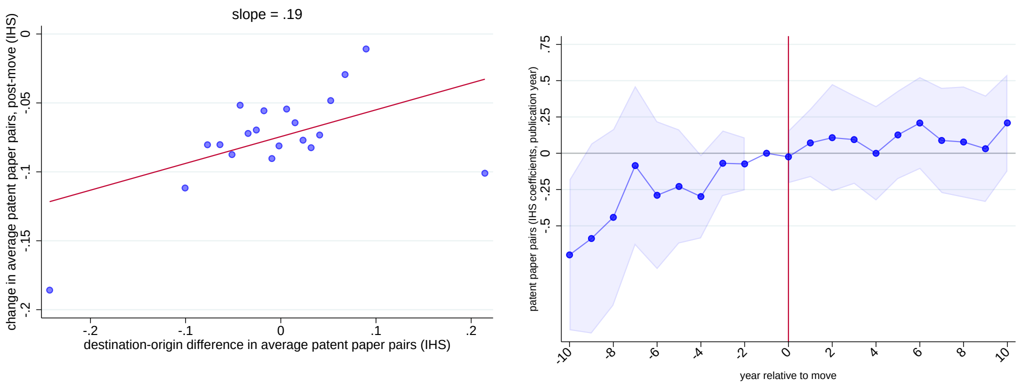


FIGURE S13.—Event study: Weighted observed five-year patent citations

Notes: This figure is identical to Figure 8(b), except that the outcome replaces our patent cite measure with a weighted patent cite measure. Patent citations are weighted by the cumulative count of patent citations they themselves receive following award. In particular, and in accordance with Lerner and Seru (2022), we scale a cited patent's count by the average count of citing patents by WIPO technology class and award year. There are 90,512 author years in our sample, coming from the sample of 12,264 movers for whom we can calculate actual five-year patent citation measures.

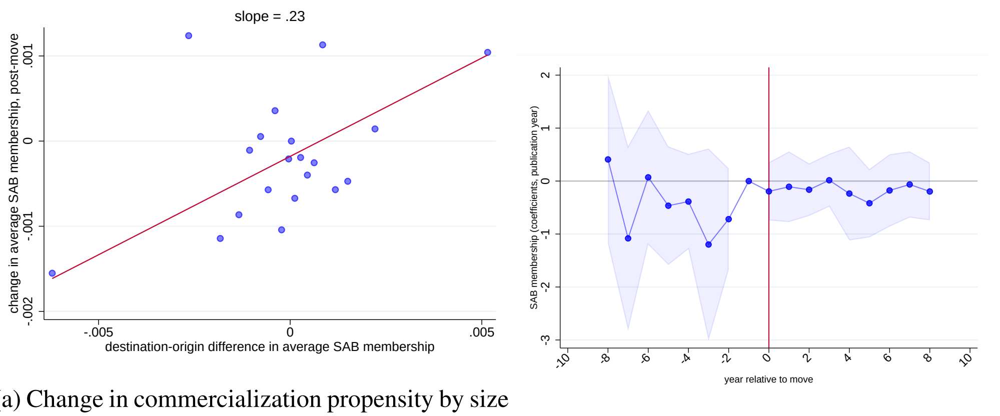


(a) Change in commercialization propensity by size of move

(b) Event study

FIGURE S14.—Patent-paper pairs

Notes: This figure is identical to Figure 8, except that our outcome is patent-paper pairs. There are N = 119,705 author years in our sample, coming from the sample of 14,242 movers.



(a) Change in commercialization propensity by size of move

(b) Event study

FIGURE S15.—Scientific advisory board membership

Notes: This figure is identical to Figure 8, except that our outcome is a 0/1 indicator for whether a researcher joined an SAB in a given year. We restrict the binscatter and event study to ± 8 years around the move, instead of ± 10, since the sparsity of the outcome leads to noisy estimates in extreme years due to few authors joining SABs. As a result, there are 111,368 author years in our sample, coming from the sample of 14,242 movers.

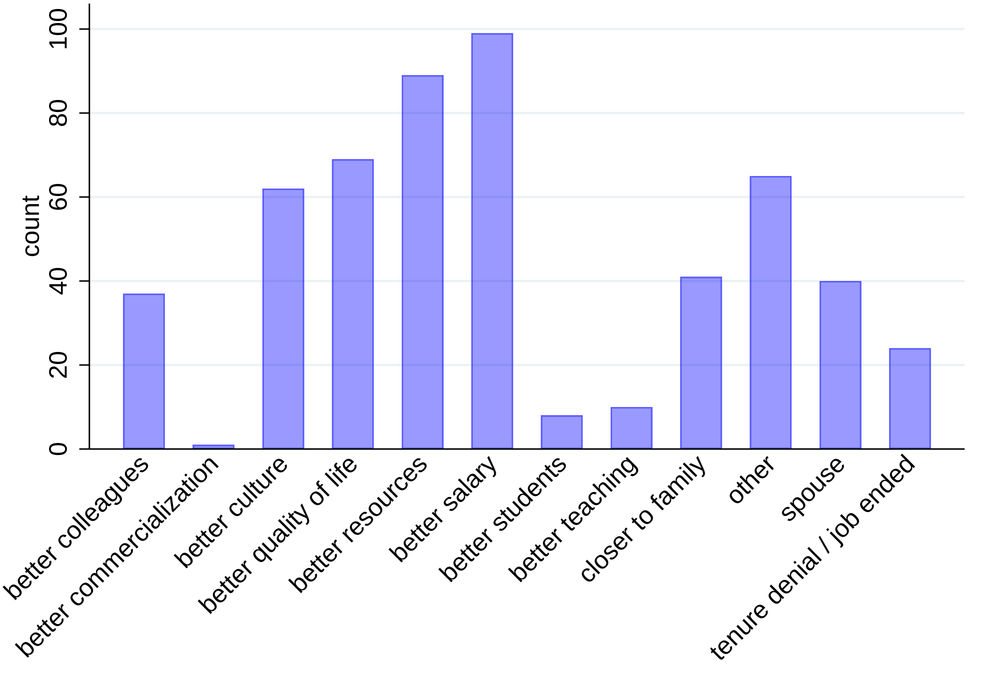


FIGURE S16.—Survey results: reasons for moving

Notes: This figure shows the count of top three reasons that respondents listed for their most recent academic move. $N = 311$. Respondents could provide up to 3 reasons, so the total counts sum to more than 311.

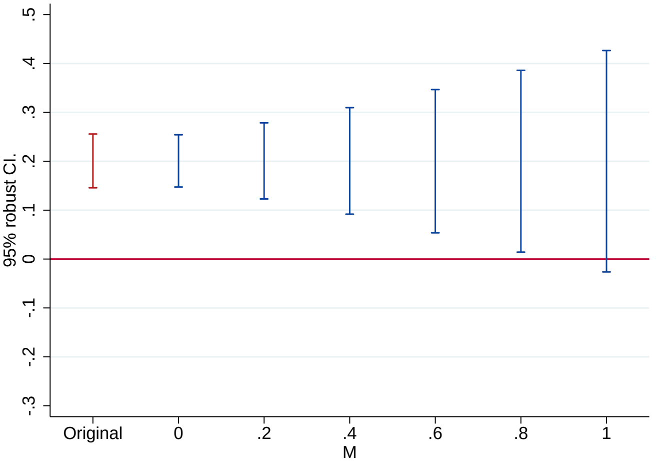
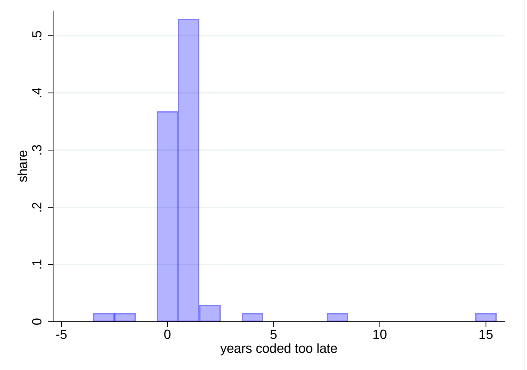
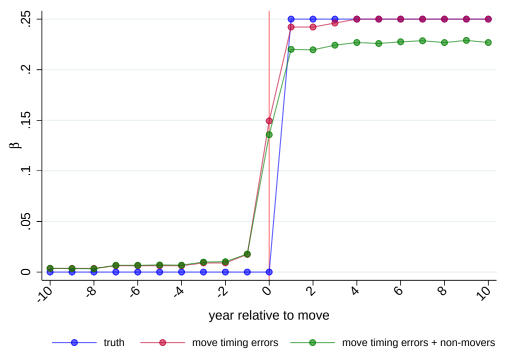


FIGURE S17.—Rambachan-Roth pre-trend sensitivity

Notes: This figure shows the point estimate and standard error for $\hat{\theta}_1$ from Figure 6 in the main text as M increases, following the methodology from Rambachan and Roth (2023). $M = 0$ corresponds to no pre-trends. $M = 1$ corresponds to the worst-case pre-trends that are possible given the observed data.



(a) Move timing errors



(b) Simulated event study

FIGURE S18.—Simulated event study

Notes: For a sample of 100 randomly selected movers, Panel (a) shows the distribution of move years coded too late. Using faculty webpages, CV, and LinkedIn profiles, we are able to determine the actual move year for 68 movers in our sample. The remaining 32 are either untraceable using name and email information (N = 25), or we determine them to have not actually moved (N = 7). Panel (b) shows the results from a simulated event study. The event study uses a panel parametrized to have 14,242 movers observed in the same calendar years ten years pre- and post-move, a constant treatment effect of $\theta_r = 0.25$ for all $r \geq 1$, and a distribution of δ that matches that shown by Figure 4. Because the treatment effect is constant across units and we do not add in a noise term, there are no standard errors. The "truth" coefficient path (blue) plots the estimates from an event study estimation using this panel as-is. The "move timing errors" path (red) shows the results from an estimation where the panel is simulated to have a distribution of move timing errors that matches Panel (a). The "move timing errors + non-movers" path (green) shows the results from a final estimation that incorporates the distribution of timing errors and also recodes 9% of the movers as non-movers. These recoded non-movers are not exposed to the treatment, but remain coded as movers to match the mistake we make in our sample construction.

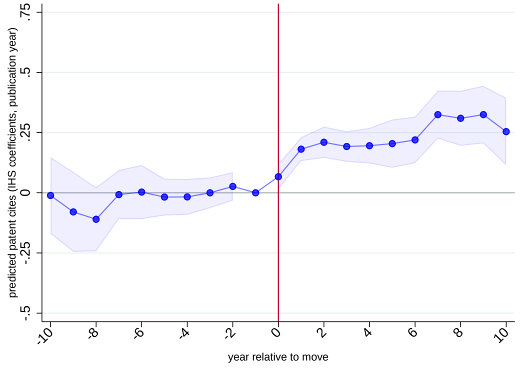
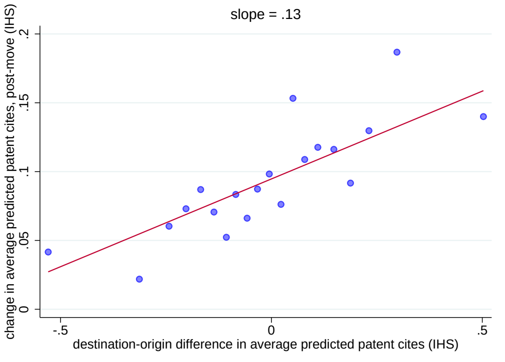
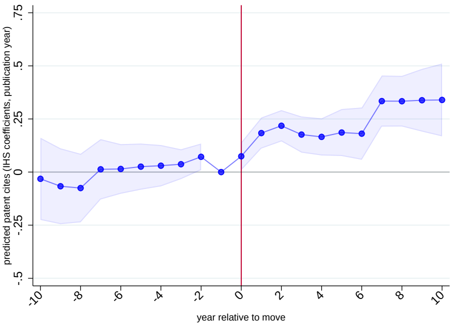


FIGURE S19.—Event study: Predicted five-year patent citations with relative year fixed effects

Notes: This figure is identical to Figure 6, except that it adds uninteracted relative year fixed effects, estimating Equation (11). There are 119,705 author years in our sample, coming from the sample of 14,242 movers.



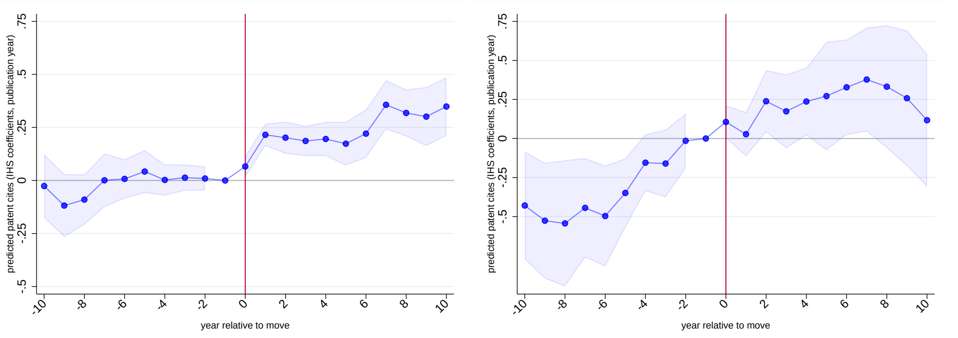
(a) Change in commercialization propensity by size of move



(b) Event study

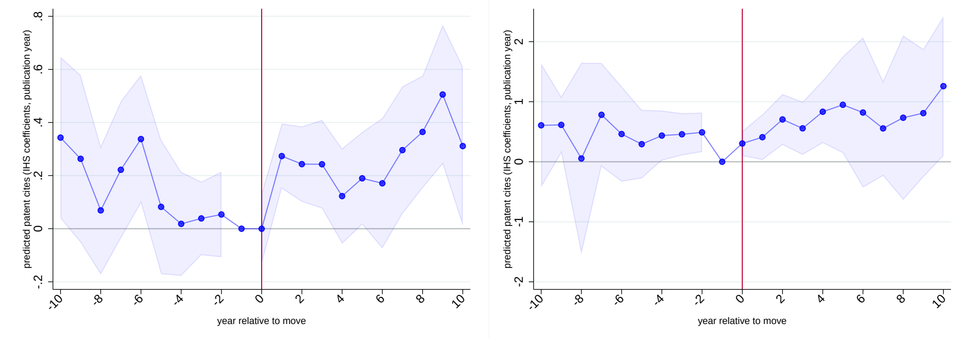
FIGURE S20.—Last authors

Notes: This figure is identical to Figures 5 and 6, except that we restrict to authors who are last authors at both their origin and destination. There are 7,093 last author movers in our sample, which corresponds to a total of 61,895 author-years.



(a) University → university
(12,944 movers)

(b) University → non-university
(718 movers)

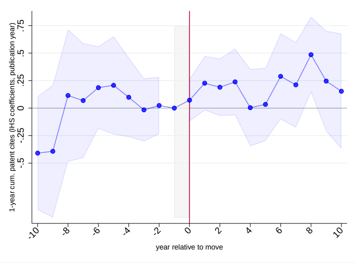


(c) Non-university → university
(532 movers)

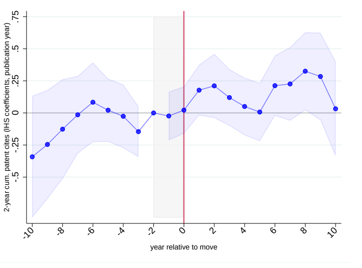
(d) Non-university → non-university
(48 movers)

FIGURE S21.—Event study for different move types

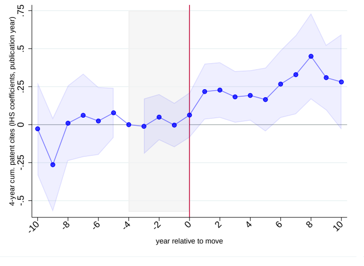
Notes: This figure is identical to Figure 6 from the main text, splitting the movers into four mutually-exclusive groups based on move type. Number of movers is provided in each subpanel.



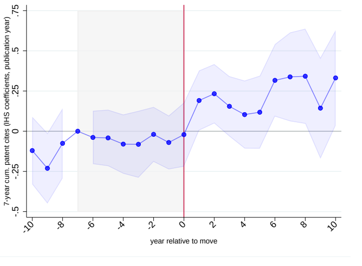
(a) 1-year cites



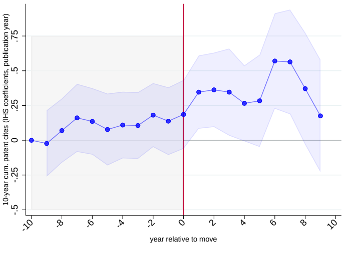
(b) 2-year cites



(c) 4-year cites



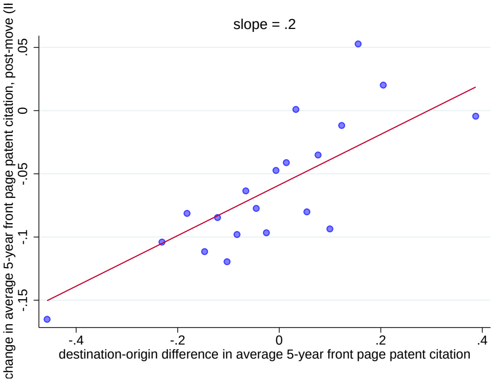
(d) 7-year cites



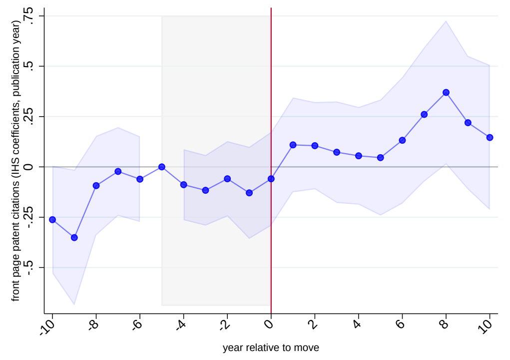
(e) 10-year cites

FIGURE S22.—Citation windows

Notes: These event study plots are identical to Figure 8(b), but the outcomes look at patent citations over different time horizons (one year, two years, etc.) Note, the number of movers, and thus observations, per event study decreases as X increases. This is because only publication years $2020-(X-1)$ and earlier will have consistent measurements of y_{it}, due to truncation. For the ten-year lagged citation outcome, the sample gets so small that it cannot sufficiently identify the $\hat{\theta}_{10}$ coefficient. Thus we do not report it.



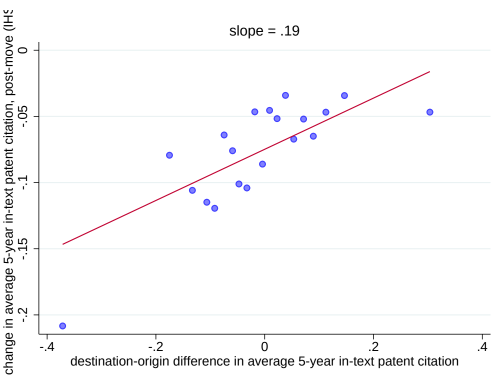
(a) Change in commercialization propensity by size of move



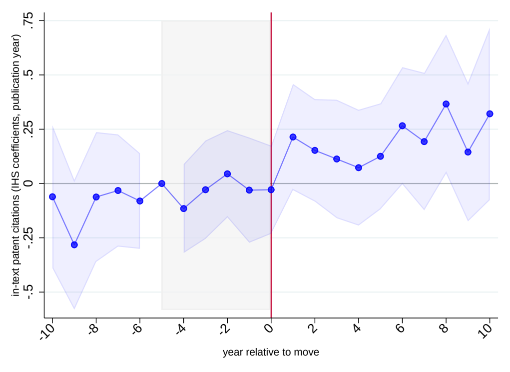
(b) Event study

FIGURE S23.—Five-year front page patent citations

This figure is identical to Figure 8, replacing all five-year patent citations with five-year front page patent citations. Each panel is estimated using 12,264 movers and 90,512 mover-years: this sample is defined by dropping the years between 2017 and 2020 from our main estimation sample with 14,242 movers and 119,705 mover-years.



(a) Change in commercialization propensity by size of move



(b) Event study

FIGURE S24.—Five-year in-text patent citations

This figure is identical to Figure 8, replacing all five-year patent citations with five-year in-text patent citations. Each panel is estimated using 12,264 movers and 90,512 mover-years: this sample is defined by dropping the years between 2017 and 2020 from our main estimation sample with 14,242 movers and 119,705 mover-years.

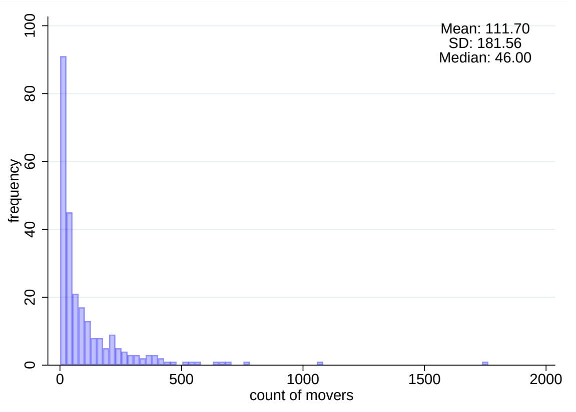


FIGURE S25.—Movers by university

Notes: This figure shows the distribution of the number of movers per university, conditional on observing at least one move. Here, a mover is counted once for their origin and once for their destination. The count of movers is plotted on the x-axis. The number of schools with each mover count is plotted on the y-axis. The mean, median, and standard deviation are shown in the top right corner. Our sample of $14,242$ movers is used; they span 253 universities.

REFERENCES

Andrews, Martyn, Len Gill, Thorsten Schank, and Richard Upward (2012), "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias." *Economics Letters*, 117 (3), 824–827. [6]

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2019), "A distributional framework for matched employer employee data." *Econometrica*, 87 (3), 699–739. [5, 6]

Card, David, Jörg Heining, and Patrick Kline (2013), "Workplace heterogeneity and the rise of West German wage inequality." *Quarterly Journal of Economics*, 128 (3), 967–1015. [3, 4, 5]

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (2016), "Sources of geographic variation in health care: Evidence from patient migration." *Quarterly Journal of Economics*, 131 (4), 1681–1726. [6]

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2020), "Leave-out estimation of variance components." *Econometrica*, 88 (5), 1859–1898. [7]

Lerner, Josh, Henry Manley, Carolyn Stein, and Heidi Williams (2026), "The wandering scholars: Understanding the heterogeneity of university commercialization." *Econometrica*. [1]

Lerner, Josh and Amit Seru (2022), "The use and misuse of patent data: Issues for finance and beyond." *Review of Financial Studies*, 35 (6), 2667–2704. [14]

Morris, Carl N. (1983), "Parametric empirical Bayes inference: Theory and applications." *Journal of the American Statistical Association*, 78 (381), 47–55. [7]

Rambachan, Ashesh and Jonathan Roth (2023), "A more credible approach to parallel trends." *Review of Economic Studies*, 90 (5), 2555–2591. [17]

Walters, Christopher (2024), "Empirical Bayes methods in labor economics." In *Handbook of Labor Economics* (Thomas Lemieux and Christian Dustmann, eds.), Vol. 5, chapter 3, 183–260, Elsevier. [7]