Supplemental Appendix

The Effect of High-Tech Clusters on the Productivity of Top Inventors: Comment

Michael Wiebe

A Computational reproducibility

The cleaning code in Moretti (2021a,b) (henceforth M21) is unreproducible and generates a slightly different dataset for each run of the code. In particular, it uses many-to-many merges with a nonunique sort order. For example, when merging inventors with patent assignees (firms), there are multiple firms per patent, so the patent identifier does not uniquely identify observations. Stata's 'merge' command sorts the data, and 'sort' pseudo-randomly orders the data; hence, different runs of 'merge' produce a different sorting within tied values of the patent identifier. For coauthored patents with multiple firms, the many-to-many merge matches a different firm to different coauthors on separate runs of the code. The same problem arises when there are multiple research fields per patent. Hence, the code generates a different dataset each time it is run.[1]

I change the code to implement a many-to-one merge. I select the first-listed assignee as a patent's unique assignee; the first-listed assignee is more likely to be a firm rather than an individual, which enables more inventors to share a firm fixed effect. The COMETS data includes numeric weights for each listed research field. I select the highest-weighted field; to break ties, I choose randomly. I follow the M21 code in selecting the first-listed class. With only one assignee, field, and class per patent, a reproducible many-to-one merge is achieved. Note that this produces a different sample than M21, so the results are slightly different. For example, the baseline elasticity in M21 is 0.0676; with my sample, it is 0.0588.

A.1 Coding issues

M21's code uses the Stata package 'reghdfe' with the option 'keepsingletons' to not drop singleton groups. The package explicitly warns the user against including singletons, due to

[1]This explains why M21 reports different baseline elasticities throughout the paper: 0.0676 (Table 3 Column 8 and Table A.7 Column 5), 0.0669 (Table 9 Column 3), 0.0662 (in text, p.3369).

biased standard errors.

The cleaning code removes suffixes from inventor names. This creates cases where different inventors are assigned the same name and identifier. For example, patent 3986449 was filed by Paul H. Hamisch, Jr. and Paul H. Hamisch, Sr.; after string cleaning, both are assigned the name "Paul H. Hamisch".

M21 assigns multi-cluster inventors (who patent in different clusters in the same year) to their modal cluster. The code calculates the modal field and modal city independently, which can assign inventors to non-existent clusters. For example, for a cluster (city, field), if an inventor patents in clusters (A,1) and (B,2) in the same year, the code could assign them to (A,2) or (B,1), even if those clusters never occur in the data.

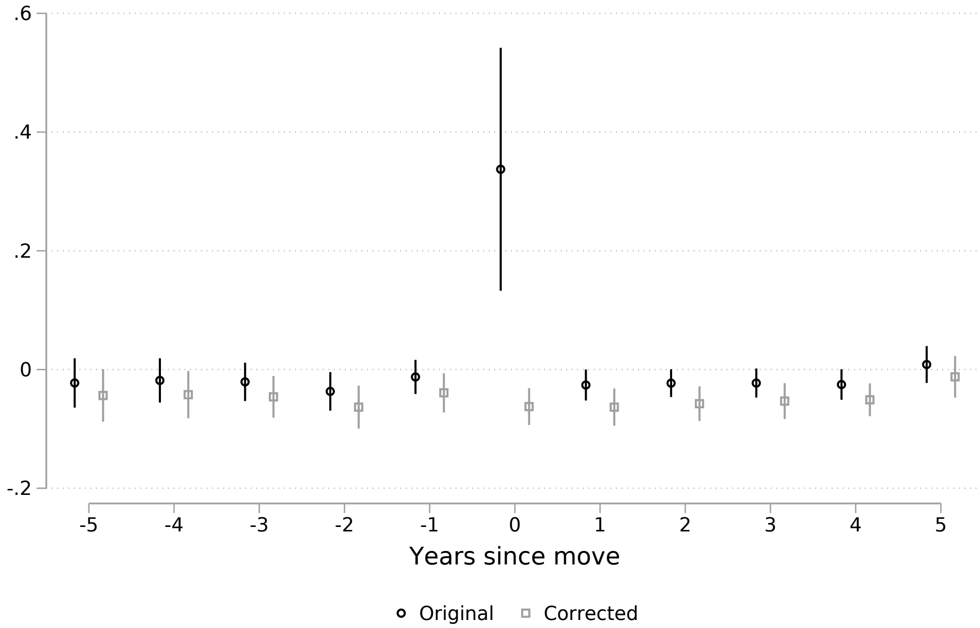
A.2 Data issues

The COMETS dataset appears to not have a unique mapping from cities to numeric BEA codes. For example, Fremont, California is associated with BEA codes 097, 140, and 146 with frequencies 23%, 34%, and 43%. This seems to contradict the description of the BEA variable: "Each county in the U.S. is assigned to a unique BEA area, with multiple counties contained within each BEA area." (p.291, Zucker et al. (2014)).

On p.3334, M21's description of the COMETS data is inaccurate. M21 mentions "the patent assignee/owner" and states that "patents are assigned to one of five main "research fields" and 579 "technology classes."", implying there is a one-to-one mapping between patents and assignees, fields, and classes. This is incorrect: there can be multiple assignees, fields, and classes per patent. M21's code uses many-to-many merges to handle multiple assignees and fields. It keeps the first-listed class, which enables a many-to-one merge; note that this modelling choice is not defended.

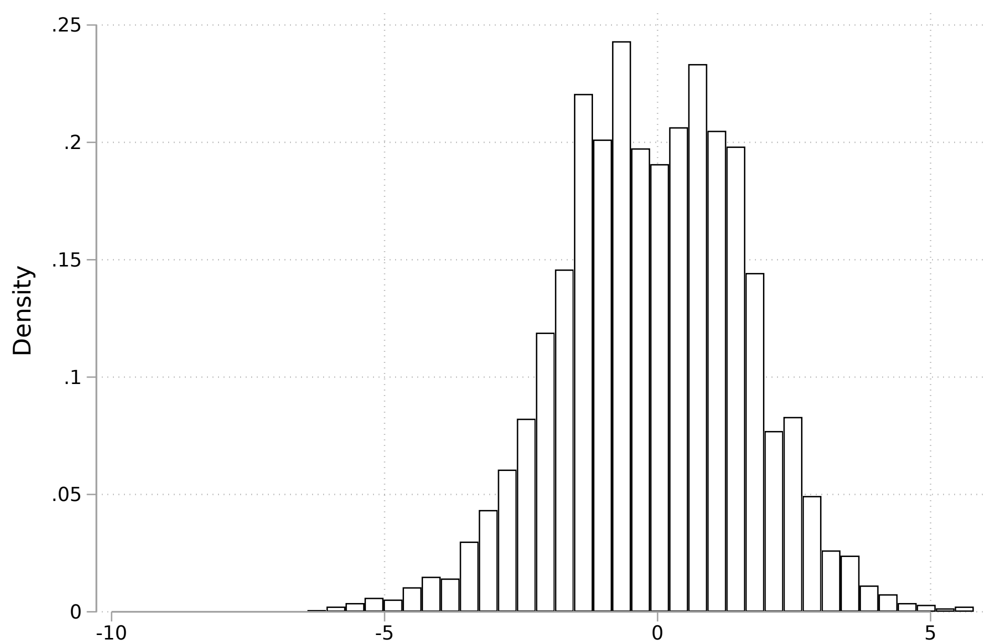
B Event study

Figure A1: Reproduction and correction of M21 Figure 6 (top panel)



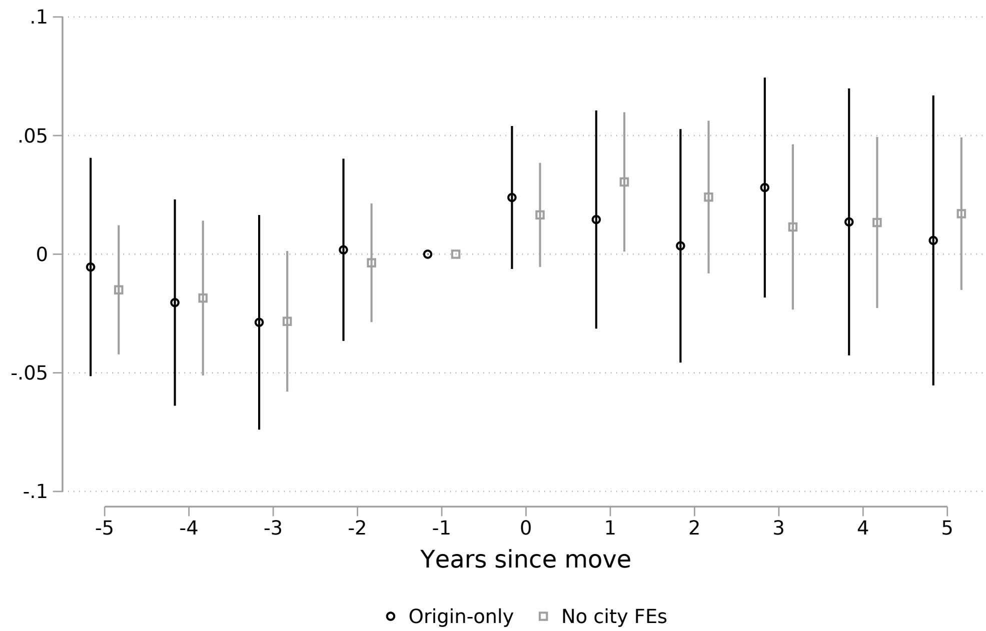
Note: Coefficients and 95% confidence intervals from Equation 1. *Original* follows the M21 code and estimates β_0 using time-varying $Size_{-ifct}$ and without interacting with $\mathbb{1}\{s = 0\}$. *Corrected* estimates β_0 using $Size^{post}_{-ifc} \times \mathbb{1}\{s = 0\}$; following the M21 code, $Size^{post}_{-ifc}$ is calculated over event years 1 to 5, excluding the year of the move ($t = 0$). N= 18,443 in *Original*, N= 18,444 in *Corrected*. Standard errors are clustered by city × research field. M21 claims that the event study includes only "inventors who are in the data for 11 consecutive years" (p.3354). However, the code allows time gaps of any length between observations. My figure does not exactly match M21 Figure 6, because the M21 cleaning code is unreproducible; see Appendix A.

Figure A2: Distribution of changes in cluster size across move



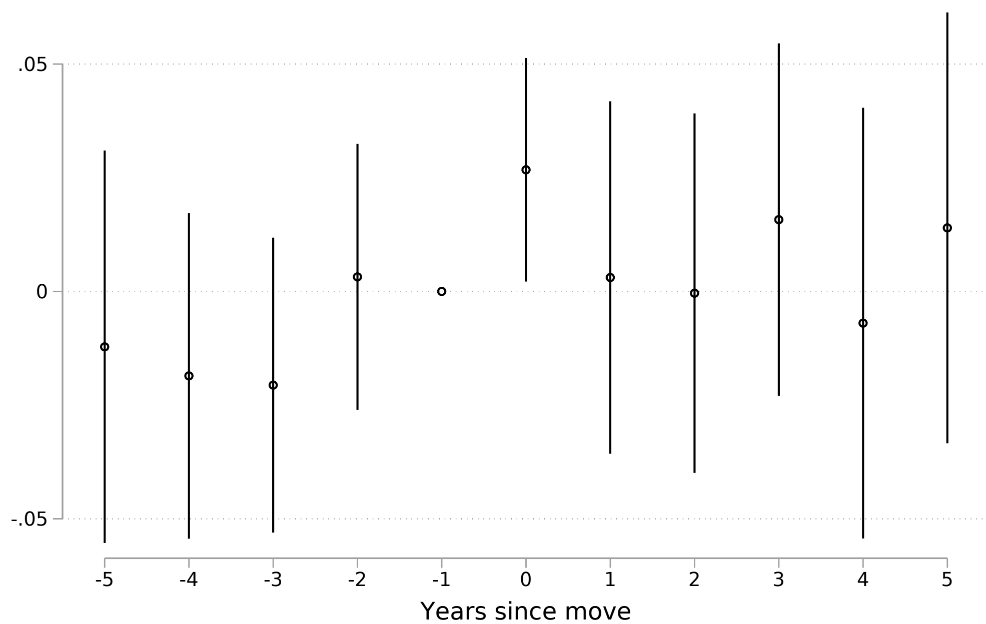
Note: Histogram of changes in cluster size among inventors who move across cities. The change for inventor i is $\Delta\text{Size}_{-i} = \text{Size}_{-i}^{post} - \text{Size}_{-i}^{pre}$, where Size_{-i}^{post} is average log cluster size over the 5 years after the move and Size_{-i}^{pre} is average log cluster size over the 5 years before the move. Inventor i is excluded when calculating cluster size, which is the number of inventors filing a patent. A cluster is a city-research field pair. N = 3,834 inventors. Average log cluster size in the mover sample is -3.40.

Figure A3: Mover event study: alternative city controls



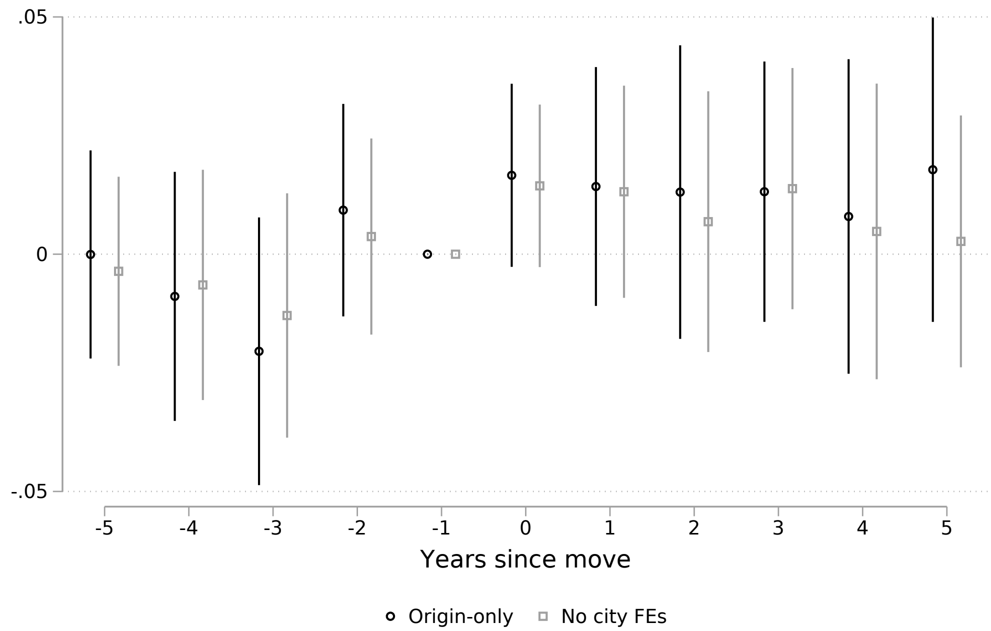
Note: Event study coefficients from Equation 2. Fixed effects: field × year, class × year, inventor, and firm. *Origin-only* also includes fixed effects for origin city × field, origin city × class, and origin city × year; N = 26,486. *No city FEs* omits all city fixed effects; N = 30,967. Standard errors are clustered by city × research field.

Figure A4: Mover event study: including stayers



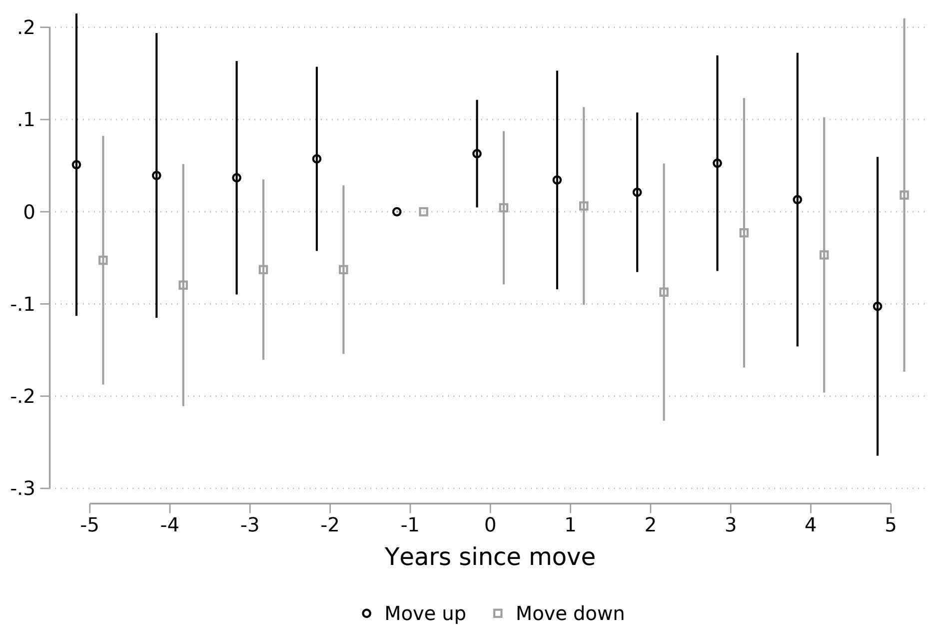
Note: Event study coefficients from Equation 2. The sample includes non-moving inventors as stayers, who are assigned event time -1 and ΔSize = 0. N = 474,575. Fixed effects: origin city $\times$ field, destination city $\times$ field, origin city $\times$ class, destination city $\times$ class, field $\times$ year, class $\times$ year, inventor, origin city $\times$ year, destination city $\times$ year, and firm. Standard errors are clustered by city $\times$ research field.

Figure A5: Mover event study: including stayers, alternative city controls



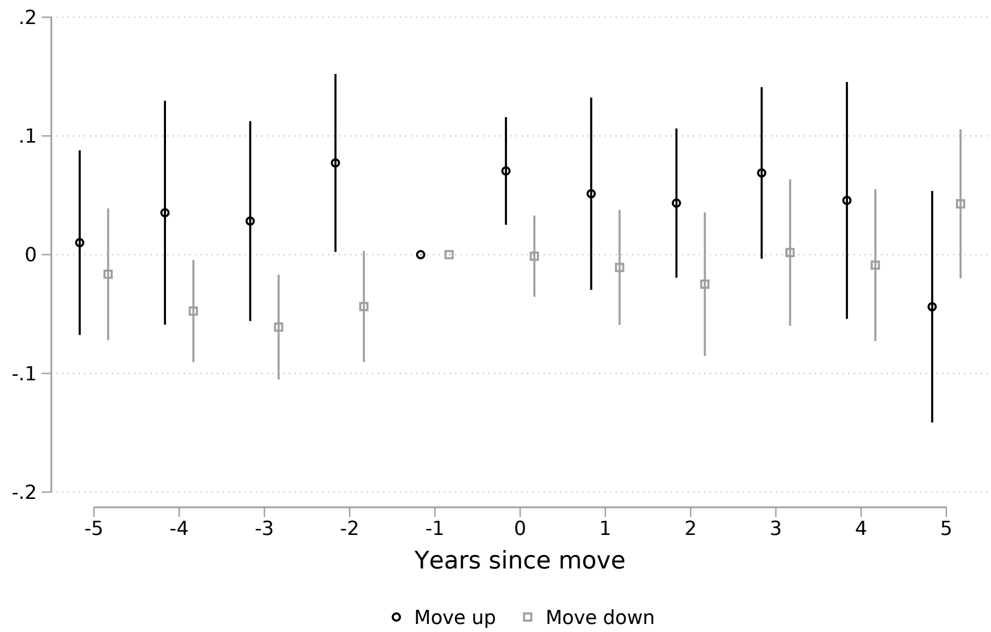
Note: Event study coefficients from Equation 2. The sample includes non-moving inventors as stayers, who are assigned event time -1 and ΔSize = 0. Fixed effects: field $\times$ year, class $\times$ year, inventor, and firm. *Origin-only* includes fixed effects for origin city $\times$ field, origin city $\times$ class, and origin city $\times$ year; N = 476,610. *No city FEs* omits all city fixed effects; N = 483,324. Standard errors are clustered by city $\times$ research field.

Figure A6: Mover event study: heterogeneity by moving-up vs. moving-down



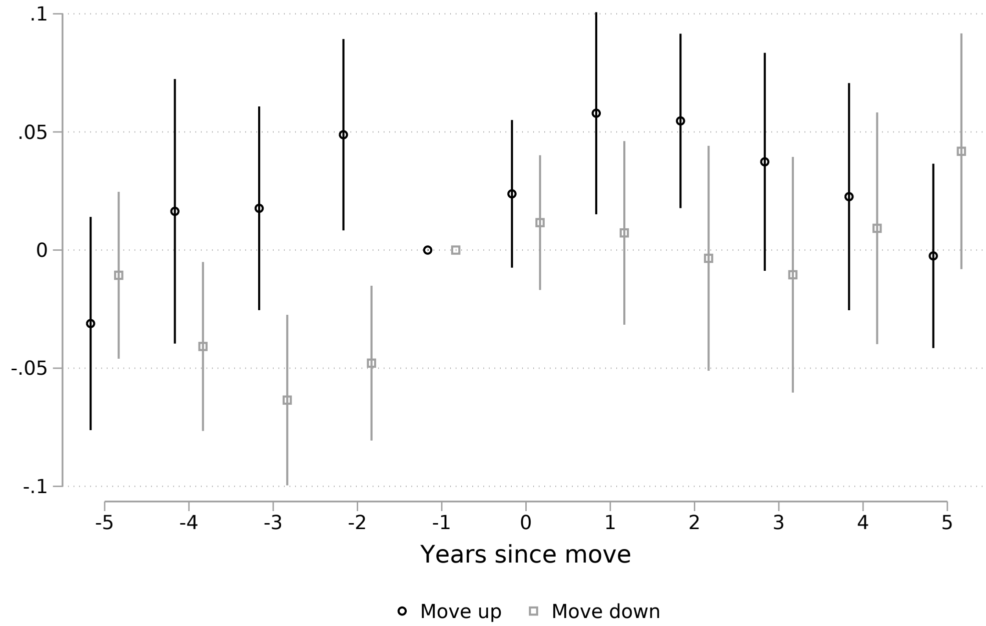
Note: Mover event study with heterogeneous effects by moving to a larger or smaller city. Event study coefficients β_t^{up} and β_t^{down} from $\ln y_{it} = \sum_{s=-5,\, s\neq -1}^{5} (\beta_s^{up} D_{it}^s \times \Delta\text{Size}_{-i} \times \mathbb{1}\{\Delta\text{Size}_{-i} > 0\} + \beta_s^{down} D_{it}^s \times \Delta\text{Size}_{-i} \times \mathbb{1}\{\Delta\text{Size}_{-i} < 0\}) + \text{FEs} + \varepsilon_{it}$. N = 22,155. Fixed effects: origin city × field, destination city × field, origin city × class, destination city × class, field × year, class × year, inventor, origin city × year, destination city × year, and firm. Standard errors are clustered by city × research field.

Figure A7: Mover event study: heterogeneity by moving-up vs. moving-down (origin-city fixed effects)



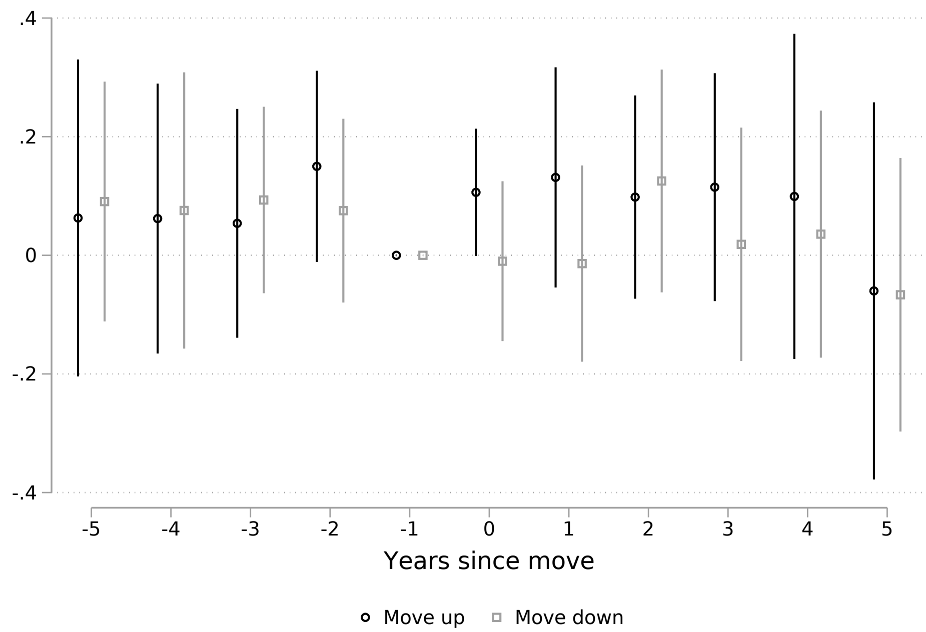
Note: Mover event study with heterogeneous effects by moving to a larger or smaller city. Event study coefficients β_t^{up} and β_t^{down} from $\ln y_{it} = \sum_{s=-5,\, s\neq -1}^{5}(\beta_s^{up} D_{it}^s \times \Delta\text{Size}_{-i} \times \mathbb{1}\{\Delta\text{Size}_{-i} > 0\} + \beta_s^{down} D_{it}^s \times \Delta\text{Size}_{-i} \times \mathbb{1}\{\Delta\text{Size}_{-i} < 0\}) + \text{FEs} + \varepsilon_{it}$. N = 26,486. Fixed effects: origin city × field, origin city × class, field × year, class × year, inventor, origin city × year, and firm. Standard errors are clustered by city × research field.

Figure A8: Mover event study: heterogeneity by moving-up vs. moving-down (no city fixed effects)



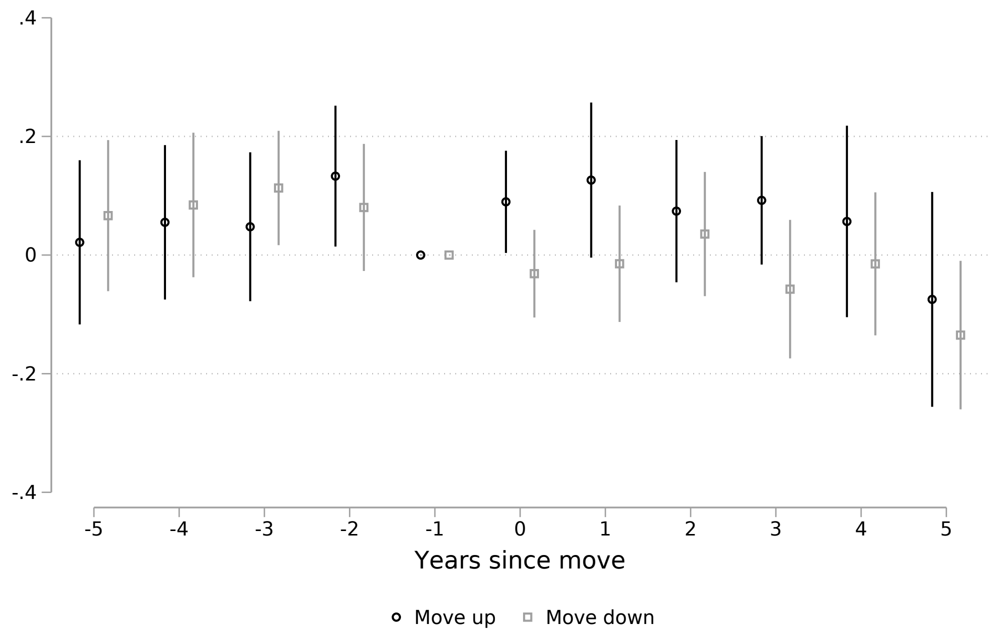
Note: Mover event study with heterogeneous effects by moving to a larger or smaller city. Event study coefficients β_t^{up} and β_t^{down} from $\ln y_{it} = \sum_{s=-5,\, s\neq -1}^{5} (\beta_s^{up} D_{it}^s \times \Delta \text{Size}_{-i} \times \mathbb{1}\{\Delta \text{Size}_{-i} > 0\} + \beta_s^{down} D_{it}^s \times \Delta \text{Size}_{-i} \times \mathbb{1}\{\Delta \text{Size}_{-i} < 0\}) + \text{FEs} + \varepsilon_{it}$. N = 30,967. Fixed effects: field × year, class × year, inventor, and firm. Standard errors are clustered by city × research field.

Figure A9: Binary mover event study: heterogeneity by moving-up vs. moving-down



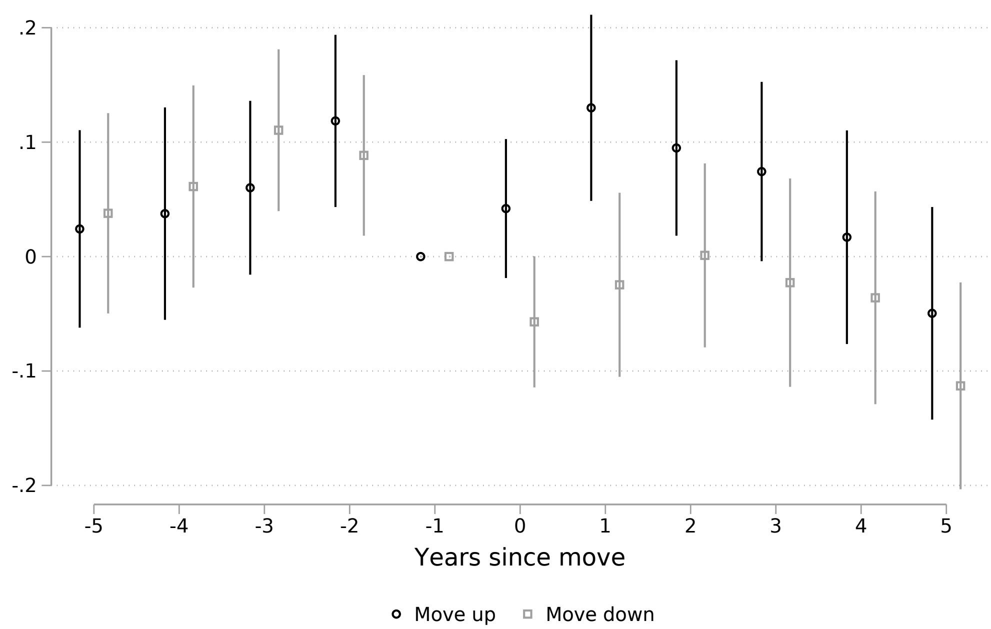
Note: Mover event study with binary treatment variable and heterogeneous effects by moving to a larger or smaller city. Event study coefficients β_t^{up} and β_t^{down} from $\ln y_{it} = \sum_{s=-5,\, s\neq -1}^{5} (\beta_s^{up} D_{it}^s \times \mathbb{1}\{\Delta \text{Size}_{-i} > 0\} + \beta_s^{down} D_{it}^s \times \mathbb{1}\{\Delta \text{Size}_{-i} < 0\}) + \text{FEs} + \varepsilon_{it}$. N = 22,155. Fixed effects: origin city $\times$ field, destination city $\times$ field, origin city $\times$ class, destination city $\times$ class, field $\times$ year, class $\times$ year, inventor, origin city $\times$ year, destination city $\times$ year, and firm. Standard errors are clustered by city $\times$ research field.

Figure A10: Binary mover event study: heterogeneity by moving-up vs. moving-down (origin-city fixed effects)



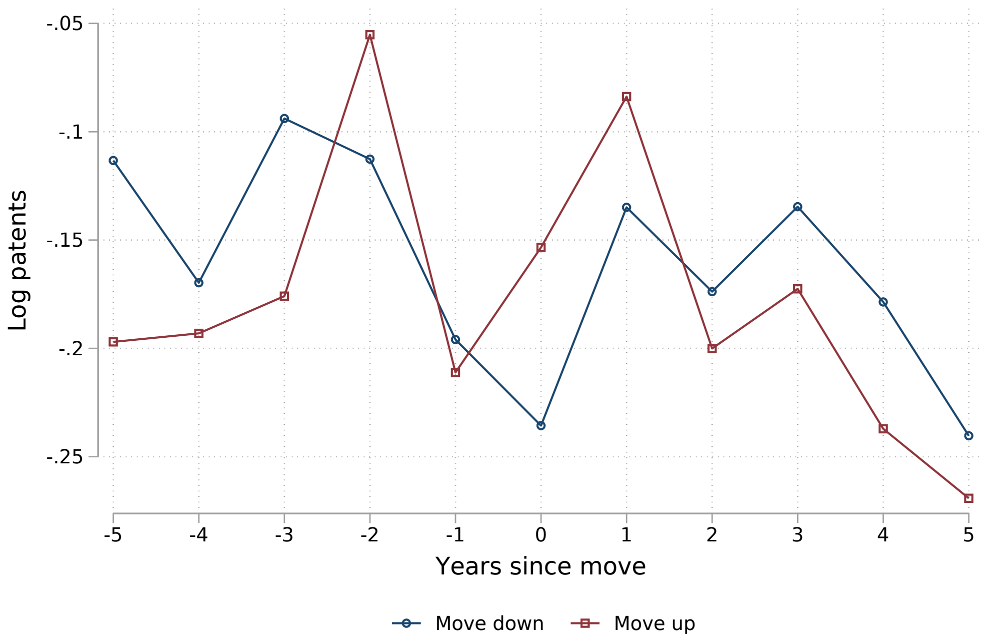
Note: Mover event study with binary treatment variable and heterogeneous effects by moving to a larger or smaller city. Event study coefficients β_t^{up} and β_t^{down} from $\ln y_{it} = \sum_{s=-5,\, s\neq-1}^{5}(\beta_s^{up} D_{it}^s \times \mathbb{1}\{\Delta\text{Size}_{-i} > 0\} + \beta_s^{down} D_{it}^s \times \mathbb{1}\{\Delta\text{Size}_{-i} < 0\}) + \text{FEs} + \varepsilon_{it}$. N = 26,486. Fixed effects: origin city × field, origin city × class, field × year, class × year, inventor, origin city × year, and firm. Standard errors are clustered by city × research field.

Figure A11: Binary mover event study: heterogeneity by moving-up vs. moving-down (no city fixed effects)



Note: Mover event study with binary treatment variable and heterogeneous effects by moving to a larger or smaller city. Event study coefficients β_t^{up} and β_t^{down} from $\ln y_{it} = \sum_{s=-5,\, s\neq -1}^{5}(\beta_s^{up} D_{it}^s \times \mathbb{1}\{\Delta\text{Size}_{-i} > 0\} + \beta_s^{down} D_{it}^s \times \mathbb{1}\{\Delta\text{Size}_{-i} < 0\}) + \text{FEs} + \varepsilon_{it}$. N = 30,967. Fixed effects: field × year, class × year, inventor, and firm. Standard errors are clustered by city × research field.

Figure A12: Average patents by event time



Note: Average log patents by event time, separately by inventors moving to larger ('move up') and smaller ('move down') clusters. The sample is restricted to inventors in the estimation sample from Figure 1b. N = 15,767.

C Citations and patent quality

M21 Table 6 Panel B uses citations per patent as the dependent variable, to test for the effect of cluster size on patent quality.[2] If patent quality is increasing in cluster size, then M21's positive patenting-size elasticity is consistent with agglomeration effects, since larger clusters produce both a higher quantity and quality of patents. However, if patent quality decreases in cluster size, then the positive baseline elasticity could be interpreted as defensive patenting, where firms create low-quality patents to use in intellectual property lawsuits. But this suggests the elasticity could even be *negative* when calculated using quality-adjusted patents, contrary to the standard agglomeration model. Hence, the effect of cluster size on citations is key to understanding the sign and magnitude of agglomeration effects for quality-adjusted patents.

There are two issues affecting the M21 Table 6 results. First, M21 adds 0.00001 when calculating log citations, which assigns a large weight to the extensive margin effect. Second, M21 incorrectly calculates citations per patent using citations per coauthor and patents per coauthor; this nullifies the per-coauthor adjustment and inflates the number of citations for inventors with many coauthors.

According to the text, the dependent variable is the "log of number of subsequent patents that cite patents filed by inventor i in year t divided by the number of patents filed by inventor i in year t." (p.3359) However, this is implemented in the code as $\log(\frac{citations+0.00001}{patents}) = \log(\frac{citations}{patents} + \frac{0.00001}{patents})$. In the code, M21 notes that adding 0.00001 avoids dropping observations with zero citations.[3] But Chen and Roth (2023) reports that the choice of constant c in a $\log(y + c)$ transformation can yield an estimate of any magnitude when there is a nonzero extensive margin. To demonstrate this, I repeat the analysis using log(citations+1) as the dependent variable (while noting that $\log(y + 1)$ and $\log(y + 0.00001)$ are both arbitrary). My preferred approach is Poisson regression, which treats the extensive and intensive margins the same way. That is, moving from creating patents with 0 citations to creating patents with 1 citation contributes to the estimate the same as moving from 1 to 2 citations.[4]

To adjust for the number of coauthors on a patent, M21 uses fractional attribution when calculating the number of patents and citations. Coauthors of a patent are each assigned an equal share of the patent and the citations received.[5] However, M21 uses fractional patents in

[2]The dependent variable in M21 Table 6 Panel A is total citations across all patents created by an inventor in a given year. This does not measure patent quality on a per-patent basis, since an inventor can receive many citations from creating many patents.

[3]When an inventor does not patent in a given year, that observation is missing from the data. $Citations_{it} = 0$ refers to patents created by inventor i in year t that are never cited, as opposed to inventor-year observations with no patents.

[4]Correia et al. (2020) explains that Poisson regression is valid for a non-integer (non-negative) dependent variable.

[5]"If a patent has multiple inventors, I assign equally weighted fractions of the patent and citations to each of

the denominator to calculate citations per patent: $\frac{citations}{coauthors}/\frac{patents}{coauthors}$. For example, consider an inventor i with a single patent that has one other coauthor and receives 10 citations. Since the total number of coauthors is 2, the number of fractional patents for i is $1/2 = 0.5$ and the number of fractional citations is $10/2 = 5$. In the code, fractional citations per patent is calculated incorrectly as $5/0.5 = 10$. But we should divide by the whole number of patents: $5/1 = 5$ fractional citations.

For inventors with a single patent in a year, this is equivalent to not adjusting for the number of coauthors, since the coauthor terms cancel out.[6] I correct the code to calculate citations per patent as fractional citations divided by the whole number of patents. Figure A13 shows a binned scatterplot of log fractional citations per patent against the log number of coauthors, using fractional patents in Panel (a) and whole patents in Panel (b). Since fractional citations is defined as $\frac{citations}{coauthors}$, we expect a negative correlation with the number of coauthors. However, Panel (a) shows that M21's incorrect citations-per-fractional-patent variable is positively related to the number of coauthors. When using the corrected citations-per-whole-patent variable in Panel (b), the correlation is negative, as expected. Hence, M21's code inflates the number of citations for inventors with many coauthors. Since the number of coauthors is positively correlated with cluster size, this overstates the number of citations for inventors in larger clusters, meaning that the estimated effect of cluster size on citations is biased upwards.

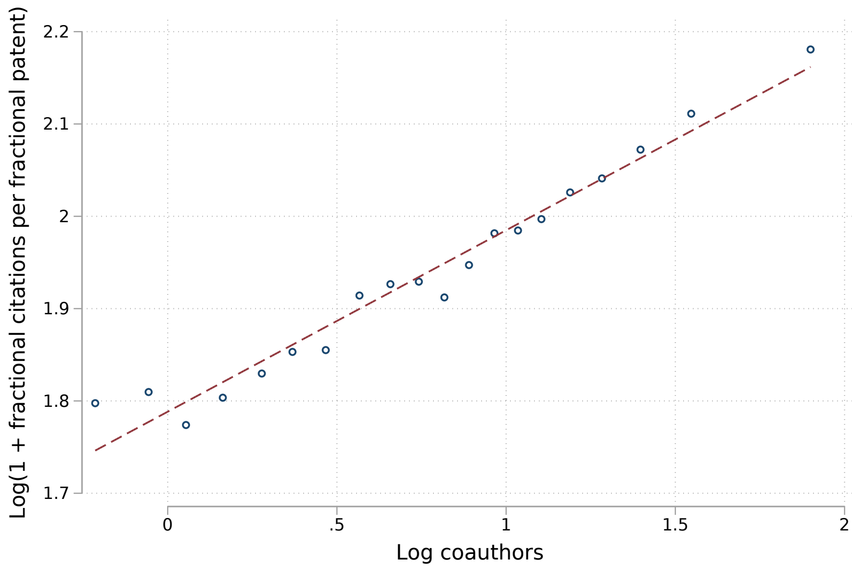
Table A1 presents the M21 Table 6 Panel B results correcting for both issues. First, Panel A reproduces the large positive effect in M21 using $\log(\frac{citations}{patents} + \frac{0.00001}{patents})$, where the denominator is fractional patents. As noted above, when the extensive margin is nonzero, the value of the constant c in $\log(y + c)$ determines the magnitude of the estimate. Table A2 shows that the extensive margin effect is positive, implying that cluster size increases the likelihood of creating a cited patent. Hence, the large effect on citations in M21 could be due to the choice of $c = 0.00001$, which assigns a log-scale value of -11.5 to observations with zero citations.

Panel B presents the results using $\log(\frac{citations}{patents} + 1)$, where the denominator is whole patents. Now the effect is negative and statistically significant. Given the positive extensive margin effect, this implies a negative intensive margin effect, where as cluster size increases, inventors shift from creating patents with some citations to creating patents with fewer citations. Is the negative effect from using c=1 or from correctly defining citations per patent (using whole patents)? Table A3 shows that the Column 6 estimate is -0.0231 (0.0130) when using c=1 and

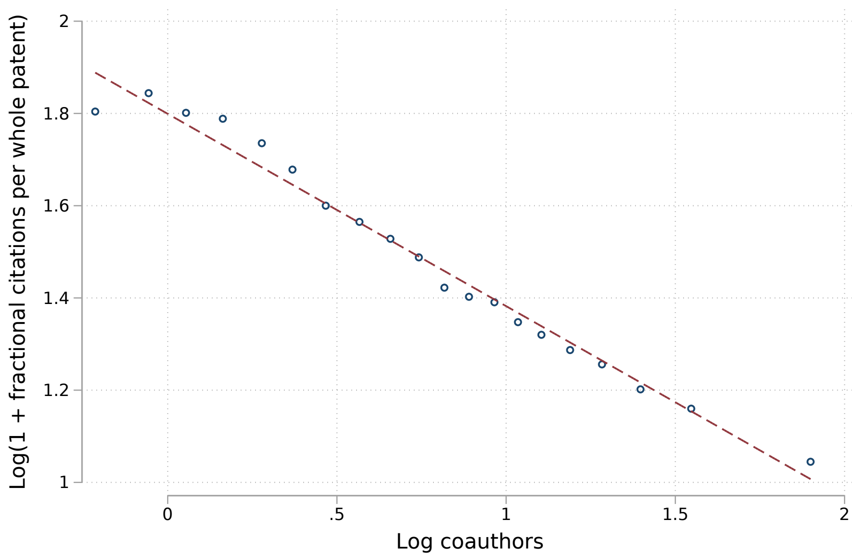
its inventors. For example, if a patent has four inventors, each inventor is credited with one-quarter of a patent and a quarter of the subsequent citations." (p.3334)

[6]For inventors with multiple patents in a year, the equivalence is not exact, since fractional patents and citations are calculated and then aggregated to the inventor-year level, before calculating $\frac{citations}{coauthors}/\frac{patents}{coauthors}$ using the average number of coauthors per patent.

Figure A13: Citations per coauthor is positively correlated with the number of coauthors



(a)



(b)

Note: Panel (a) shows citations per patent calculated incorrectly as the log of fractional citations per fractional patent, so the per-coauthor adjustment cancels out. Panel (b) shows the correct definition as the log of fractional citations per whole patent. Fractional values are calculated by dividing by the number of coauthors. Fixed effects for city, field, and year.

dividing by fractional patents, and is 0.0548 (0.0380) when using c=0.00001 and dividing by whole patents. Hence, the positive effect reported in M21 is mostly due to c=0.00001 putting more weight on the extensive margin.[7] To further explore the impact of the constant c in $\log(y+c)$, Table A5 varies the constant from c=0.00001 to c=1 while following M21 in dividing by fractional patents. The coefficient decreases monotonically as the constant grows larger, from 0.1174 to -0.0231.[8]

Panel C uses Poisson regression with fractional citations per whole patent as the dependent variable. Since Poisson regression treats the extensive and intensive margins the same way, it avoids weighting the extensive margin effect in an unsystematic way. Moreover, in this context the extensive margin does not correspond to a discrete economic choice (such as participating in the labor market), but would seem to represent idiosyncrasies in patent quality, as some patents happen to go uncited. That is, a change between 0 and 1 citations is similar to a change between 1 and 2 citations. Hence, Poisson regression is appropriate. The estimates in Panel C are negative and statistically significant, which implies that the intensive margin effect is negative, and outweighs the positive extensive margin.[9] The overall effect of cluster size on citations is negative.

As a robustness check, I repeat the regressions using the measure of patent importance from Kelly et al. (2023) as the dependent variable. Figure A14 shows that importance is positively correlated with citations per patent from M21's data. Consistent with the results based on citations, Table A6 shows that cluster size also has a negative effect on patent importance, though the estimate loses significance with city-year fixed effects.

So the original M21 finding is reversed: patent quality is in fact decreasing in cluster size. While cluster size increases patent quality along the extensive margin (by raising the probability of creating a cited patent), it decreases quality along the intensive margin (by reducing the number of citations per patent for inventors with cited patents), and the overall effect is negative. This potentially changes the interpretation of M21's main findings. That is, larger clusters produce more patents than smaller clusters, but those patents are of lower quality, possibly because they are used strategically in intellectual property lawsuits rather than

[7]Following Chen and Roth (2023), in Table A4 I calibrate the value of the extensive margin using the transformed outcome $m(y) = \log(y)$ for $y > 0$ and $m(0) = -x$. In this specification, the extensive margin is valued as a $100x$ percent change in the intensive margin. For example, when $x = 0.1$, the extensive margin is valued the same as a 10 log point change along the intensive margin. I match the M21 coefficient using $x = 6$. Hence, the M21 specification using $\log(y + 0.00001)$ can be interpreted as valuing the extensive margin as the equivalent of a 600 log point change (or $100(e^6 - 1)\% \approx 40,000\%$ change) in the intensive margin.

[8]The elasticity for c=0 is -0.0325, representing the intensive margin effect when dropping observations with zero citations.

[9]Since the independent variable is log cluster size and Poisson regression uses a log link function, the Poisson coefficient is an elasticity.

Table A1: Reanalysis of M21 Table 6, Panel B (citations per patent)

	(1)	(2)	(3)	(4)	(5)	(6)	(7) Same city	(8) Different city
Panel A: Original: log(citations per fractional patent + 0.00001/patents)								
Log size	0.1002	0.1086	0.1565	0.1165	0.1131	0.1119	0.2497	-0.0340
	(0.0835)	(0.0427)	(0.0362)	(0.0260)	(0.0368)	(0.0389)	(0.0766)	(0.0296)
Observations	923934	923934	923226	922390	921843	786680	691890	691890
Adjusted R^2	0.449	0.453	0.468	0.514	0.515	0.514	0.287	0.346
Panel B: Log(citations per whole patent + 1)								
Log size	-0.1231	-0.1292	-0.0844	-0.0850	-0.0517	-0.0611	-0.0533	-0.0621
	(0.0173)	(0.0133)	(0.0088)	(0.0105)	(0.0097)	(0.0107)	(0.0108)	(0.0111)
Observations	923934	923934	923226	922390	921843	786680	691890	691890
Adjusted R^2	0.461	0.464	0.487	0.584	0.586	0.605	0.402	0.544
Panel C: Poisson regression (citations per whole patent)								
Log size	-0.1274	-0.1107	-0.0893	-0.0849	-0.0653	-0.0774	-0.0938	-0.0771
	(0.0317)	(0.0152)	(0.0144)	(0.0140)	(0.0153)	(0.0160)	(0.0266)	(0.0165)
Observations	923500	923500	921878	920045	919251	782919	654526	691374
Year	Yes							
City × field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class × year			Yes	Yes	Yes	Yes	Yes	Yes
Inventor				Yes	Yes	Yes	Yes	Yes
City × year					Yes	Yes	Yes	Yes
Firm						Yes	Yes	Yes

Standard errors in parentheses

Notes: Citations are measured as the number of patents that cite any patent filed by inventor i in year t; citations are always calculated per coauthor, dividing by the number of inventors listed on the patent. Citations per patent is measured as citations divided by the number of patents filed by inventor i in year t. Fractional patents are calculated per coauthor, dividing by the number of inventors listed on the patent. Panel A follows the M21 code in calculating citations per patent using fractional patents in the denominator; in this case, the per-coauthor adjustment cancels out. My results in panel A do not exactly match the original Table 6, because the M21 cleaning code is unreproducible; see Appendix A. Panels B and C calculate citations per patent using whole patents in the denominator. The dependent variable in panel C is citations per patent in levels. Column 7 is for citations by inventors located in the same city as the focal inventor. Column 8 is for citations by inventors located in a city different from the focal inventor city. Standard errors are clustered by city × research field.

representing genuine innovation.

C.1 Coding issues

In Columns 7 and 8 of Table 6, M21 tests whether citations come from patents in the same city. The code to join citing and cited patents uses a many-to-many merge, which makes the results unreproducible by producing a different sample each time the code is run. Moreover, in cases where citing patents have multiple coauthors from different cities, this approach assigns different values of 'citing-city' to different cited patents in an unsystematic way. For example, suppose patent p cites patents $q1$, $q2$, and $q3$; the coauthors of p are from cities X and Y. The many-to-many merge will produce the following matching of cited-patent and citing-city: $\{q1{:}X,\ q2{:}Y,\ q3{:}Y\}$. In unreported results, I correct the code by assigning each patent the modal city of its inventors and using a many-to-one merge to join citing and cited patents. The estimates are similar.

C.2 Supplementary tables and figures

Table A2: Reanalysis of M21 Table 6, Panel B: extensive margin effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7) Same city	(8) Different city
Log size	0.0196	0.0210	0.0203	0.0183	0.0120	0.0133	0.0126	-0.0015
	(0.0071)	(0.0034)	(0.0029)	(0.0023)	(0.0028)	(0.0031)	(0.0056)	(0.0016)
Observations	923934	923934	923226	922390	921843	786680	786680	786680
Adjusted R^2	0.348	0.353	0.369	0.401	0.402	0.391	0.180	0.043
Year	Yes							
City × field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class × year			Yes	Yes	Yes	Yes	Yes	Yes
Inventor				Yes	Yes	Yes	Yes	Yes
City × year					Yes	Yes	Yes	Yes
Firm						Yes	Yes	Yes

Standard errors in parentheses

Notes: The dependent variable is an indicator for a strictly positive value of total citations received by an inventor for patents created in a given year. Citations are measured as the number of patents that cite any patent filed by inventor i in year t. Column 7 is for citations by inventors located in the same city as the focal inventor. Column 8 is for citations by inventors located in a city different from the focal inventor city. Standard errors are clustered by city × research field.

Table A3: Reanalysis of M21 Table 6, Panel B (citations per patent): alternative specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7) Same city	(8) Different city
Panel A: Log(citations per fractional patent + 1)								
Log size	-0.0928	-0.0993	-0.0502	-0.0654	-0.0126	-0.0231	-0.0412	-0.0250
	(0.0233)	(0.0139)	(0.0098)	(0.0118)	(0.0123)	(0.0130)	(0.0141)	(0.0131)
Observations	923934	923934	923226	922390	921843	786680	691890	691890
Adjusted R^2	0.483	0.487	0.511	0.601	0.602	0.623	0.421	0.545
Panel B: Log(citations per whole patent + 0.00001/patents)								
Log size	0.0590	0.0662	0.1082	0.0849	0.0516	0.0548	0.1929	-0.0908
	(0.0894)	(0.0414)	(0.0350)	(0.0257)	(0.0357)	(0.0380)	(0.0777)	(0.0293)
Observations	923934	923934	923226	922390	921843	786680	691890	691890
Adjusted R^2	0.466	0.470	0.484	0.527	0.528	0.526	0.280	0.373
Year	Yes							
City × field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class × year			Yes	Yes	Yes	Yes	Yes	Yes
Inventor				Yes	Yes	Yes	Yes	Yes
City × year					Yes	Yes	Yes	Yes
Firm						Yes	Yes	Yes

Standard errors in parentheses

Notes: Citations are measured as the number of patents that cite any patent filed by inventor i in year t; citations are always calculated per coauthor, dividing by the number of inventors listed on the patent. Citations per patent is measured as citations divided by the number of patents filed by inventor i in year t. Fractional patents are calculated per coauthor, dividing by the number of inventors listed on the patent. Panel A follows the M21 code in calculating citations per patent using fractional patents in the denominator; in this case, the per-coauthor adjustment cancels out. Departing from the M21 code, I add 1 instead of $\frac{0.00001}{patents}$. Panel B calculates citations per patent using whole patents in the denominator, and follows the M21 code in adding $\frac{0.00001}{patents}$. Column 7 is for citations by inventors located in the same city as the focal inventor. Column 8 is for citations by inventors located in a city different from the focal inventor city. Standard errors are clustered by city × research field.

Table A4: Reanalysis of M21 Table 6, Panel B: calibrated extensive margin value ($x = 6$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7) Same city	(8) Different city
Log size	0.0932	0.1004	0.1466	0.1094	0.1034	0.1039	0.1155	0.0984
	(0.0858)	(0.0414)	(0.0354)	(0.0253)	(0.0356)	(0.0377)	(0.0572)	(0.0369)
Observations	923934	923934	923226	922390	921843	786680	524513	766583
Adjusted R^2	0.454	0.457	0.472	0.518	0.519	0.517	0.585	0.536
Year	Yes							
City × field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class × year			Yes	Yes	Yes	Yes	Yes	Yes
Inventor				Yes	Yes	Yes	Yes	Yes
City × year					Yes	Yes	Yes	Yes
Firm						Yes	Yes	Yes

Standard errors in parentheses

Notes: The dependent variable is citations per patent, which is measured as the number of subsequent patents that cite patents filed by inventor i in year t divided by the number of patents filed by inventor i in year t. Following the M21 code, I calculate citations per patent is using fractional citations and fractional patents, so that the per-coauthor adjustment cancels out. Following Chen and Roth (2023), I use the transformed outcome $m(y) = \log(y)$ for $y > 0$ and $m(0) = -x$. I first normalize the outcome by dividing by the minimum non-zero number of citations, so that the extensive margin is a move from 0 to 1. I choose x to match the estimates in M21 Table 6 Panel B. Column 7 is for citations by inventors located in the same city as the focal inventor. Column 8 is for citations by inventors located in a city different from the focal inventor city. Standard errors are clustered by city × research field.

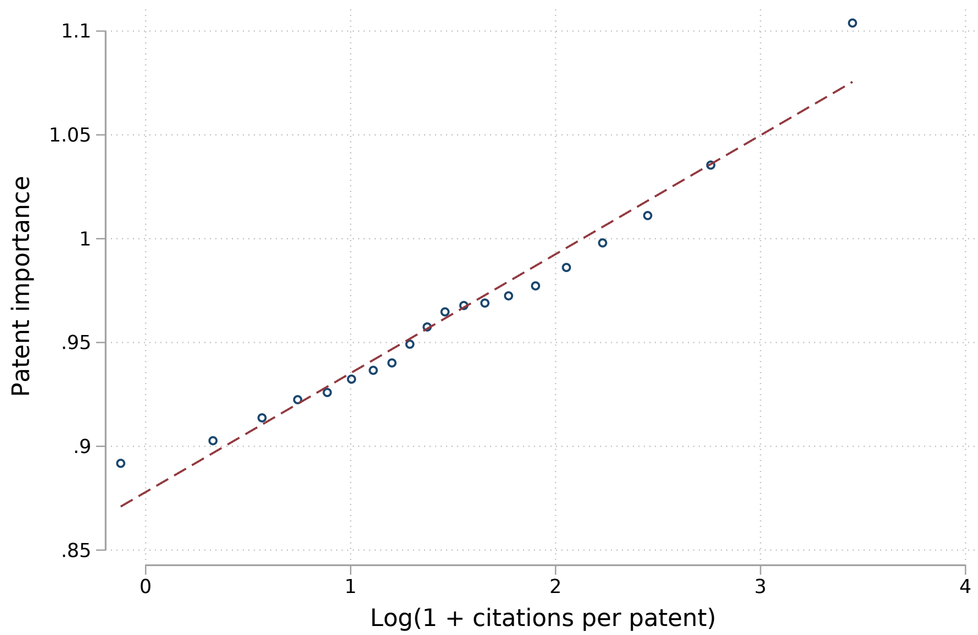
Table A5: Reanalysis of M21 Table 6, Panel B (citations per patent): varying c in $\log(y + c)$

	(1) c=0.00001	(2) c=0.0001	(3) c=0.001	(4) c=0.01	(5) c=0.1	(6) c=1	(7) c=0
Log size	0.1174	0.0869	0.0563	0.0260	-0.0025	-0.0231	-0.0325
	(0.0406)	(0.0341)	(0.0278)	(0.0221)	(0.0172)	(0.0130)	(0.0154)
Observations	786680	786680	786680	786680	786680	786680	691890
Adjusted R^2	0.509	0.530	0.557	0.592	0.626	0.623	0.555

Standard errors in parentheses

Notes: Each column follows the Table 6, Panel B, column 6 specification, with fixed effects for city × field, city × class, field × year, class × year, inventor, city × year, and firm. The dependent variable is citations per patent, which is measured as the number of subsequent patents that cite patents filed by inventor i in year t divided by the number of patents filed by inventor i in year t. Following M21, citations per patent is calculated using fractional patents in the denominator, so that the per-coauthor adjustment cancels out. Column 1 does not match Table A1 Column 6, since it uses $\log(\frac{citations}{patents} + 0.00001)$ rather than $\log(\frac{citations}{patents} + \frac{0.00001}{patents})$. Column 7 drops observations with zero citations. Standard errors are clustered by city × research field.

Figure A14: Patent importance and citations per patent



Note: The dependent variable is average patent importance for patents filed by inventor i in year t. Importance is measured as the log ratio of forward 10-year and backward 5-year aggregate patent similarity; see Kelly et al. (2023). Citations per patent is measured using whole patents in the denominator. Fixed effects for city, field, and year.

Table A6: Reanalysis of M21 Table 6, Panel B: alternative measure of patent quality

	(1)	(2)	(3)	(4)	(5)	(6)
Log size	-0.0268	-0.0330	-0.0126	-0.0176	-0.0048	-0.0048
	(0.0141)	(0.0081)	(0.0042)	(0.0048)	(0.0034)	(0.0034)
Observations	923655	923655	922948	922111	921564	786448
Adjusted R^2	0.598	0.625	0.702	0.773	0.774	0.795
Year	Yes					
City × field	Yes	Yes	Yes	Yes	Yes	Yes
City × class	Yes	Yes	Yes	Yes	Yes	Yes
Field × year		Yes	Yes	Yes	Yes	Yes
Class × year			Yes	Yes	Yes	Yes
Inventor				Yes	Yes	Yes
City × year					Yes	Yes
Firm						Yes

Standard errors in parentheses

Notes: The dependent variable is average patent importance for patents filed by inventor i in year t. Importance is measured as the log ratio of forward 10-year and backward 5-year aggregate patent similarity; see Kelly et al. (2023). Standard errors are clustered by city × research field.

D Aggregate effects of agglomeration

M21 Section V quantifies the aggregate benefits of agglomeration in the United States, using the estimated elasticity of patenting to cluster size. M21 performs an exercise of equalizing cluster size within each field and year, to calculate the decrease in the total number of patents. When agglomeration effects are positive, reallocating inventors from high- to low-productivity places will reduce aggregate output.

To illustrate, consider an example with two cities in the same research field. Inventor i in city c creates patents as a function of cluster size S_c. Given the elasticity α, each inventor produces S_{c}^{α} patents; this functional form is derived from the original regression equation with $\ln(y) = \alpha \ln(S)$.[10] Since there are S_c inventors, total patents created in city c is then $S_{\mathrm{c}} \cdot S_{\mathrm{c}}^{\alpha}$. For cities A and B, initial aggregate output is $Y^1 = S_{\mathrm{A}} \cdot S_{\mathrm{A}}^{\alpha} + S_{\mathrm{B}} \cdot S_{\mathrm{B}}^{\alpha}$.

In the counterfactual, cluster sizes are equalized: $S'_{\mathrm{A}} = S'_{\mathrm{B}} = \frac{S_{\mathrm{A}}+S_{\mathrm{B}}}{2} := \frac{S}{2}$, where S is the total number of inventors. Aggregate output is $Y^2 = S(S/2)^{\alpha}$. The change in output in the case of two cities in a single field and year is $Y^2 - Y^1 = S_{\mathrm{A}} \left[\left(\frac{S}{2}\right)^{\alpha} - S_{\mathrm{A}}^{\alpha} \right] + S_{\mathrm{B}} \left[\left(\frac{S}{2}\right)^{\alpha} - S_{\mathrm{B}}^{\alpha} \right]$. In the general case with many cities, the difference in aggregate output in field f and year t is $Y_{ft}^2 - Y_{ft}^1 = \sum_c S_{cft} \left[\overline{S}_{ft}^{\alpha} - S_{cft}^{\alpha} \right]$, where $\overline{S}_{ft}$ is the resulting cluster size after the population is equalized across cities, and S_{cft} is the actual cluster size. The percentage difference is $(Y_{ft}^2 - Y_{ft}^1)/Y_{ft}^1$. I reproduce the M21 Table 12 results in Columns 1 and 2 of Table A8 below. Similar to the original, aggregate output falls by around 10% when using a constant elasticity, and by roughly 11% when using heterogeneous estimates that allow the elasticity to vary by cluster size quartiles.

There are several problems with the analysis in M21. First, M21 estimates the elasticity using the subsample of the top 10% of inventors (defined using lifetime patents). Hence, the elasticity should be α_{T10} (estimated on the top 10% subsample) rather than α (estimated on the full sample). So per-inventor output is $S_c^{\alpha_{T10}}$ rather than S_c^{α}; note that per-inventor output is based on the total number S_c of inventors in the same cluster, even when restricting to top 10% inventors. More importantly, total patents should be calculated using $S_{\mathrm{T10},c}$, the number of inventors in the top 10% subsample, rather than the total number S_c. Hence, total patents created should be $S_{\mathrm{T10},c} \cdot S_c^{\alpha_{T10}}$ rather than $S_c \cdot S_c^{\alpha_{T10}}$.

Second, M21 describes the exercise as using the number of inventors, but the code uses cluster density, defined as the number of inventors in a city-field-year divided by the field-year total. But when $0 < S < 1$, we have the undesirable behavior that $S^{\alpha} < S^{\beta}$ when $\alpha > \beta$: a larger elasticity is associated with a smaller output. This affects the results in Column 2 of M21 Table 12, which uses heterogeneous elasticities that vary by cluster size.

[10] There is a typo on M21 p.3369: "ln$g(S) = \alpha S$, so that $g(S) = S^{\alpha}$". The first equation should read "ln$g(S) = \alpha \ln S$".

Third, the M21 estimates of heterogeneity by cluster size are affected by a coding error. M21 Table 8 tests whether the elasticity of patenting to cluster size varies with cluster size, to understand whether agglomeration effects are different in small versus large clusters. Specifically, Panel A of M21 Table 8 interacts log cluster size with indicators for cluster size quartiles, to estimate the effect of cluster size on patenting separately for each quartile. There is little evidence that the elasticity varies with cluster size. However, the M21 code incorrectly omits the quartile indicators from the regression, which forces the baseline level of patents to be the same across quartiles, resulting in biased estimates of heterogeneity.[11] I reproduce the original results in Column 1 of Table A7. In Column 2, I show the corrected regression, which includes the quartile indicators. Now there is substantial heterogeneity across cluster size quartiles, with the elasticity in the fourth quartile being seven times larger than the elasticity in the first quartile (0.0192 versus 0.1432).

Table A7: Reanalysis of M21 Table 8: heterogeneity by cluster size quartiles

	(1) Original	(2) Corrected: field-year	(3) Corrected: global
Log size × Q1	0.0601	0.0192	-0.0074
	(0.0136)	(0.0126)	(0.0139)
Log size × Q2	0.0641	0.0890	0.0728
	(0.0139)	(0.0179)	(0.0235)
Log size × Q3	0.0623	0.1483	0.1753
	(0.0143)	(0.0213)	(0.0368)
Log size × Q4	0.0690	0.1432	0.2190
	(0.0159)	(0.0320)	(0.0339)
Observations	786680	786680	786680
Adjusted R^2	0.242	0.242	0.242

Standard errors in parentheses

Notes: The dependent variable is log patents. Regression equation: $\ln y_{it} = \sum_{q=1}^{4} \beta_q Q_q \times \ln S_{-it} + \sum_{q=1}^{3} \lambda_q Q_q + \text{FEs} + \varepsilon_{it}$. Column 1 omits the term $\sum_{q=1}^{3} \lambda_q Q_q$, following the original code. My results in Column 1 do not exactly match the original estimates in Table 8, Panel A, Column 2, because the M21 cleaning code is unreproducible; see Appendix A. Cluster size quartiles are calculated within field-year in columns 1 and 2, and globally (across fields and years) in column 3. Coefficients for the quartile indicators Q_q are not reported. Fixed effects: city × field, city × class, field × year, class × year, inventor, city × year, and firm. Standard errors are clustered by city × research field.

Moreover, while unspecified in the text, the M21 code calculates cluster size quartiles within

[11] M21 Table 8 Panel B tests for heterogeneity by firm productivity instead of cluster size. The code for this regression correctly includes the quartile indicators.

each research field and year, presumably to match the exercise in M21 Table 12, which equalizes cluster size within field-year. But since cluster size varies across fields and years, calculating quartiles within field-year limits the amount of variation across quartiles, and hence understates how much the elasticity varies by cluster size. In Column 3, I repeat the analysis when calculating cluster size quartiles globally, across both fields and years. The heterogeneity is even larger, ranging from -0.0074 in the first quartile to 0.2190 in the fourth quartile.

I repeat the aggregation exercise correcting for these three issues. I multiply per-inventor output by the number of top 10% inventors, instead of the number of all inventors. To avoid the issue of larger elasticities resulting in lower output, I measure cluster size as the number of inventors (so $S \geq 1$) rather than as a density. Finally, I use the corrected heterogeneous estimates from Table A7. The results are shown in Columns 3-5 of Table A8. The reduction in aggregate output is slightly smaller using the constant elasticity (-9.15%, Column 3). When using the corrected heterogeneous estimates based on within-field-year size quartiles from Table A7 Column 2, the reduction in output is twice as large (-23.75%, Column 4).[12] Column 5 uses the elasticities from Table A7 Column 3, where size quartiles are calculated across fields and years to capture the full variation in cluster size. Since the heterogeneity across quartiles is more pronounced, the output losses from equalization are even larger, at -35.11%.

Table A8: Reproduction and correction of M21 Table 12

	(1) Original: constant α	(2) Original: het. α	(3) Corrected: constant α	(4) Corrected: het. α (field-year)	(5) Corrected: het. α (global)
Computer science	-11.80	-12.69	-12.28	-29.93	-43.94
Biology and chemistry	-8.94	-9.98	-9.07	-23.67	-35.74
Semiconductors	-13.24	-14.21	-13.40	-30.28	-44.04
Other engineering	-6.75	-7.45	-7.39	-20.71	-28.74
Other science	-8.57	-9.61	-9.01	-23.03	-34.26
All fields	-9.91	-10.84	-9.15	-23.75	-35.11

Notes: Entries are percentage changes in aggregate output from equalizing cluster size within field-year, for year=2007. Columns 1 and 2 follow the original code: per-inventor output is multiplied by the total number of inventors; cluster size is measured as a density; and the heterogeneous elasticities are from Table A7, Column 1. I use the constant elasticity as estimated on my sample: 0.0588. Columns 3-5 multiply per-inventor output by the number of top 10% inventors and measure cluster size as the number of inventors. Column 4 uses the corrected heterogeneous elasticities from Table A7, Column 2, based on cluster size quartiles calculated within field-year. Column 5 uses the heterogeneous elasticities from Table A7, Column 3, based on cluster size quartiles calculated across fields and years. My results in Columns 1 and 2 do not exactly match the original estimates in M21 Table 12, because the M21 cleaning code is unreproducible; see Appendix A.

[12] When I correct only one issue at a time, the reduction in aggregate output is: -11.66% (using the top 10% population); -9.43% (using the number of inventors); -18.03% (using the corrected heterogeneous estimates). So most of the difference between Column 2 and Column 4 is from using the corrected heterogeneous elasticities.

The above analysis uses elasticities estimated on the subsample of top 10% inventors. M21 does not report estimates using the subsample of bottom 90% inventors. However, M21 Table 9 reports a small elasticity (0.0182) when using the full sample, implying a negative elasticity for bottom 90% inventors. In fact, the baseline elasticity for this subsample is -0.0346 (0.0096). The negative elasticity suggests that cluster size has a crowding-out effect for low-productivity inventors; agglomeration has costs as well as benefits. Table A9 estimates the heterogeneity by cluster size quartiles for the bottom 90% subsample. The negative effect is concentrated in the first quartile, with an elasticity of -0.0681. Hence, it appears that the crowding-out effect is strongest in the smallest clusters.

I repeat the aggregation exercise using the elasticities and inventor populations from the bottom 90% subsample. The negative first-quartile elasticity suggests that equalizing cluster size would amplify the decrease in aggregate output, by reallocating inventors to clusters with agglomeration *dis*economies. However, the fourth quartile elasticity is not as large, so the overall reduction in output is smaller at -5.83% (see Table A10 Column 1). I also equalize cluster size using the full sample of inventors, with separate elasticities for top 10% and bottom 90% inventors.[13] This approach effectively averages the results of the top 10% and bottom 90% exercises, yielding a reduction in aggregate output of -22.08% (see Table A10 Column 2).

Hence, the aggregate productivity gains from agglomeration appear to be several times larger than those reported in M21. However, this conclusion must be tempered in light of my earlier results. First, the null results in the event study and IV regressions leave open the possibility that the elasticity is not causal. Second, the negative effect of cluster size on citations implies that patents created in larger clusters tend to be of lower quality. While reallocating inventors from large to small clusters reduces the total number of patents, the corresponding decrease in quality-adjusted patents is less pronounced.[14]

[13] I formalize the full sample exercise in Appendix D.1.

[14] I repeat the aggregation exercise using citation-weighted patents to adjust for quality. Specifically, I weight each patent by the number of citations received, so the dependent variable is the total number of citations for patents created in a year. To address observations with zero citations, I use Poisson regression, which treats the extensive and intensive margins the same way; the log link function also produces the same functional form in the aggregation exercise. For the top 10% sample, the Poisson estimate of the baseline elasticity (using unweighted patents) is 0.12, and the estimate using citation-weighted patents is 0.08. The corresponding reductions in aggregate patents are -17.75% and -13.07%. In the other specifications, the unweighted and weighted reductions in aggregate patents are -44.94% and -24.08% (using the top 10% sample with heterogeneous elasticities), and -31.95% and -12.16% (using the full sample). Hence, the aggregate productivity gains are 26%, 46%, and 62% smaller when accounting for patent quality.

D.1 Aggregation exercise

The main analysis uses the subsample of top 10% inventors. Here I repeat the aggregation exercise using all inventors.

The size of city c is the sum of its Top10 and Bottom90 inventors:

$$S_c = S_{\mathrm{T10,c}} + S_{\mathrm{B90,c}}$$

We estimate the elasticities separately by subsample: α_{T10} and α_{B90}. Initial aggregate output is:

$$Y^1 = (S_{\mathrm{T10,A}} \cdot S_{\mathrm{A}}^{\alpha_{T10}} + S_{\mathrm{B90,A}} \cdot S_{\mathrm{A}}^{\alpha_{B90}}) + (S_{\mathrm{T10,B}} \cdot S_{\mathrm{B}}^{\alpha_{T10}} + S_{\mathrm{B90,B}} \cdot S_{\mathrm{B}}^{\alpha_{B90}})$$

$S_{\mathrm{T10,A}}$ is the number of Top10 inventors in city A. The per-inventor output for Top10 inventors in city A is $S_{\mathrm{A}}^{\alpha_{T10}}$, who have elasticity α_{T10}. Note that the elasticity is applied to the full population of city A: an inventor's output is determined by the total number of other inventors in the same cluster, regardless of type. To obtain patents created, we multiply the per-inventor output ($S_{\mathrm{c}}^{\alpha_i}$) by the number of inventors in each subgroup ($S_{\mathrm{i,c}}$) for $c =$ A,B and $i =$ T10, B90. In the homogeneous case ($\alpha_{T10} = \alpha_{B90} = \alpha$), we have $Y^1 = S_{\mathrm{A}} \cdot S_{\mathrm{A}}^{\alpha} + S_{\mathrm{B}} \cdot S_{\mathrm{B}}^{\alpha}$.

With inventor types, we equalize across clusters by equalizing the number of inventors in each subgroup. As a result, the clusters have the same size.

$$S'_{\mathrm{A}} = S'_{\mathrm{B}} = \frac{S_{\mathrm{T10}} + S_{\mathrm{B90}}}{2} := \frac{S}{2}$$

where S_{T10} is the total number of Top10 inventors, S_{B90} is the total number of Bottom90 inventors, and S is the total number of inventors (in both cities).

Aggregate output after equalizing is:

$$Y^2 = S_{\mathrm{T10}} \left(\frac{S}{2}\right)^{\alpha_{T10}} + S_{\mathrm{B90}} \left(\frac{S}{2}\right)^{\alpha_{B90}}$$

For homogeneous α, this simplifies to $Y^2 = S(S/2)^{\alpha}$.

The difference in aggregate output is:

$$\begin{aligned} Y^2 - Y^1 = {} & S_{\mathrm{T10,A}} \left[\left(\frac{S}{2}\right)^{\alpha_{T10}} - S_{\mathrm{A}}^{\alpha_{T10}}\right] + S_{\mathrm{T10,B}} \left[\left(\frac{S}{2}\right)^{\alpha_{T10}} - S_{\mathrm{B}}^{\alpha_{T10}}\right] \\ & + S_{\mathrm{B90,A}} \left[\left(\frac{S}{2}\right)^{\alpha_{B90}} - S_{\mathrm{A}}^{\alpha_{B90}}\right] + S_{\mathrm{B90,B}} \left[\left(\frac{S}{2}\right)^{\alpha_{B90}} - S_{\mathrm{B}}^{\alpha_{B90}}\right] \end{aligned}$$

This equation is for two cities within a single field and year. In the general case with many cities, the difference in aggregate output for a field-year is:

$$Y_{ft}^2 - Y_{ft}^1 = \sum_c \left\{ S_{\mathrm{T10,c}} \left[\overline{S}_{ft}^{\alpha_{T10}} - S_{cft}^{\alpha_{T10}} \right] + S_{\mathrm{B90,c}} \left[\overline{S}_{ft}^{\alpha_{B90}} - S_{cft}^{\alpha_{B90}} \right] \right\},$$

where $\overline{S}_{ft}$ is the resulting cluster size in field f and year t after the T10 and B90 populations are equalized across cities, and S_{cft} is the actual cluster size. The percentage difference is $(Y_{ft}^2 - Y_{ft}^1)/Y_{ft}^1$.

D.2 Supplementary tables and figures

Table A9: Reanalysis of M21 Table 8: bottom 90% inventors

	(1) Bottom 90%
Log size × Q1	-0.0681
	(0.0112)
Log size × Q2	-0.0156
	(0.0189)
Log size × Q3	-0.0215
	(0.0241)
Log size × Q4	0.0545
	(0.0336)
Observations	925201
Adjusted R^2	0.359

Standard errors in parentheses

Notes: The dependent variable is log patents. The sample is bottom 90% inventors, defined using lifetime patents. Regression equation: $\ln \mathrm{y}_{it} = \sum_{q=1}^{4} \beta_q Q_q \times \ln S_{-it} + \sum_{q=1}^{3} \lambda_q Q_q + \mathrm{FEs} + \varepsilon_{it}$. Cluster size quartiles are calculated globally (across fields and years). Coefficients for the quartile indicators Q_q are not reported. Fixed effects: city × field, city × class, field × year, class × year, inventor, city × year, and firm. Standard errors are clustered by city × research field.

Table A10: Reproduction and correction of M21 Table 12: additional results

	(1) Bottom 90%: het. α	(2) Full sample: het. α
Computer science	-10.08	-29.29
Biology and chemistry	-7.72	-22.99
Semiconductors	-9.14	-33.54
Other engineering	-2.71	-16.69
Other science	-5.94	-21.14
All fields	-5.83	-22.08

Notes: Entries are percentage changes in aggregate output from equalizing cluster size within field-year, for year=2007. Column 1 uses the sample of bottom 90% inventors (defined using lifetime patents), with heterogeneous elasticities based on global size quartiles from Table A9, Column 1. Column 2 uses the full sample of inventors, with separate elasticities for top 10% and bottom 90% inventors; elasticities are heterogeneous based on global size quartiles (calculated separately by subgroup). Top 10% elasticities are from Table A7 Column 3, and bottom 90% elasticities are from Table A9, Column 1. Aggregate output is calculated by multiplying per-inventor output by the number of inventors in the corresponding subsample, and cluster size is measured as the number of inventors. In Column 2, the counterfactual difference in aggregate output for a field-year is $Y^2_{ft} - Y^1_{ft} = \sum_c \left\{ S_{\mathrm{T10,c}} \left[\overline{S}_{ft}^{\alpha_{T10}} - S_{cft}^{\alpha_{T10}} \right] + S_{\mathrm{B90,c}} \left[\overline{S}_{ft}^{\alpha_{B90}} - S_{cft}^{\alpha_{B90}} \right] \right\}$, where $S_{j,cft}$ is the number of inventors in subgroup j in city c, field f, and year t; α_j is the elasticity for subgroup j; $\overline{S}_{ft}$ is the resulting cluster size after the population is equalized across cities; and S_{cft} is the actual cluster size.

E Distributed lag model

M21 Figure 5 estimates the distributed lag model in Equation A1 to test for selection bias from rising star inventors moving into larger clusters. The regression includes five leads and five lags of cluster size. The lead terms (β_1 to β_5) capture how current patenting is affected by future changes in cluster size, while the lag terms (β_{-5} to β_{-1}) capture lagged effects of past changes in cluster size.

$$\ln y_{ijfkct} = \sum_{s=-5}^{5} \beta_s \ln(\text{Size})_{-ifc(t+s)} + d_{cf} + d_{ck} + d_{ft} + d_{kt} + d_{ct} + d_i + d_j + \varepsilon_{ijfkct} \tag{A1}$$

Since cluster size is a continuous variable and changes every year, this specification represents the most demanding form of a distributed lag model, with multiple treatments of different signs and varying intensities (Schmidheiny and Siegloch, 2023). This is a generalization of an event study, where an individual can experience multiple events and event time indicators are scaled by the treatment magnitude. In this case, a different event occurs in every year: a change in cluster size. As a result, this is a staggered treatment adoption design where two-way fixed effects estimates may be biased. M21 does not discuss the assumptions for identifying a treatment effect of cluster size in this setting.[15]

The M21 code for Figure 5 defines leads and lags incorrectly. To illustrate, it does not implement a one-year lag as $t-1$, referring to the year prior to t. Instead, it uses the *observation* prior to t. Since the inventor-year panel is unbalanced, a one-year lag can hence be defined as any number of years before t. In other words, the code defines a one-year lag as $t - j$ for $j \geq 1$. For example, β_{-5} is supposed to be the effect of cluster size five years in the past on current patents, but actually corresponds to cluster sizes from five to twenty-nine years in the past.[16] M21 states that "for a inventor to be in this sample, the five leads and lags need to be nonmissing, which implies that only inventors observed in 11 consecutive years are included." (p.3352) However, the code does not enforce the requirement that the observations span consecutive calendar years, so the estimation sample also includes inventors with 11 observations in non-consecutive years.

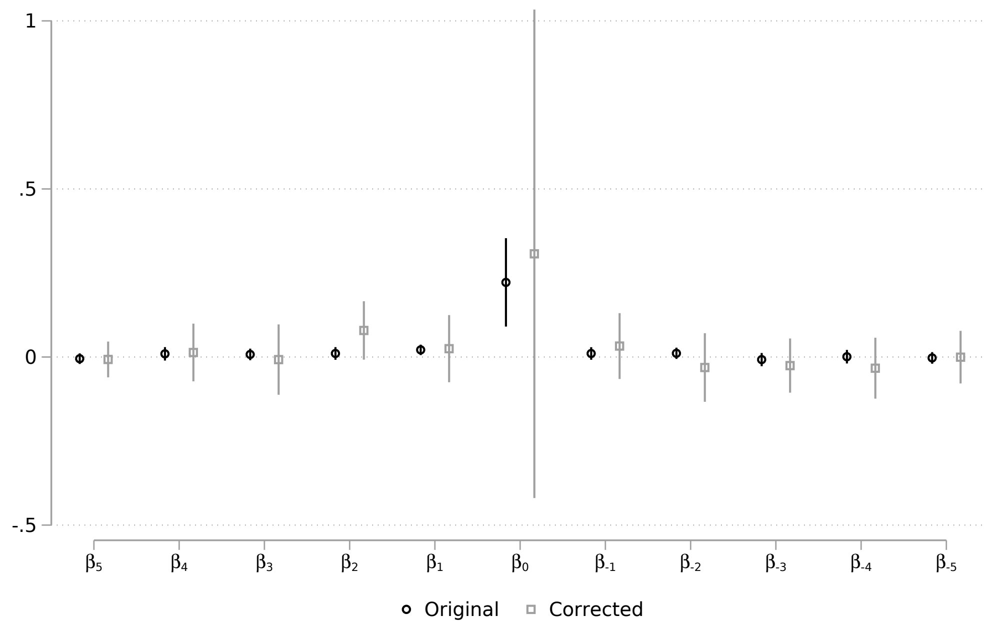
I correct the code to use inventors with observations in 11 consecutive years, and plot the original and corrected versions of M21 Figure 5 in Figure A15. The point estimates are similar, but restricting to observations in consecutive years reduces the sample size, so the confidence

[15]Note that the distributed-lag parameters β_s represent the incremental effect of a treatment, whereas the parameters in an event study specification capture cumulative effects. Hence, we aggregate the β_s terms to present the dynamic treatment effects of a change in cluster size.

[16]The M21 code does not use Stata's panel operators `L.` or `F.`, which require a lag to have a one-year gap. Instead, it defines a lag using Stata's `_n` operator, meaning that a lag is the previous observation, with no restrictions on the length of the gap.

intervals are much wider.[17] As a result, the estimate of β_0 becomes nonsignificant.

Figure A15: Reproduction and correction of M21 Figure 5 (top panel)

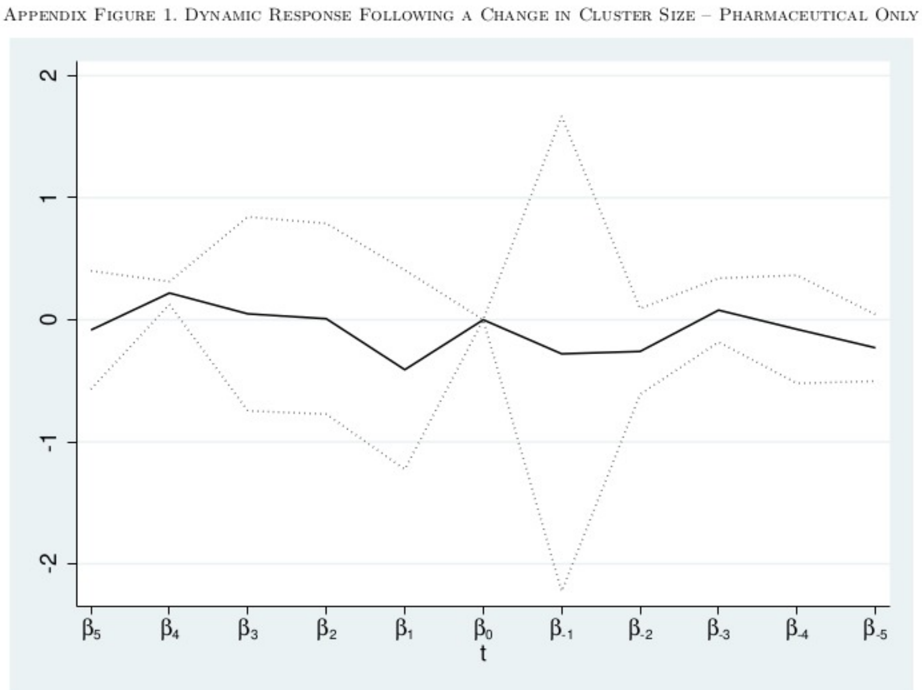


Note: Distributed lag coefficients and 95% confidence intervals from Equation A1, corresponding to the top panel of M21 Figure 5. *Original* includes inventors observed in 11 possibly non-consecutive years. *Corrected* restricts the sample to inventors observed in 11 consecutive years, as stated in the text. N= 56,887 in *Original*, N= 6,326 in *Corrected*. Fixed effects: city × field, city × class, field × year, class × year, inventor, city × year, and firm. Standard errors are clustered by city × research field.

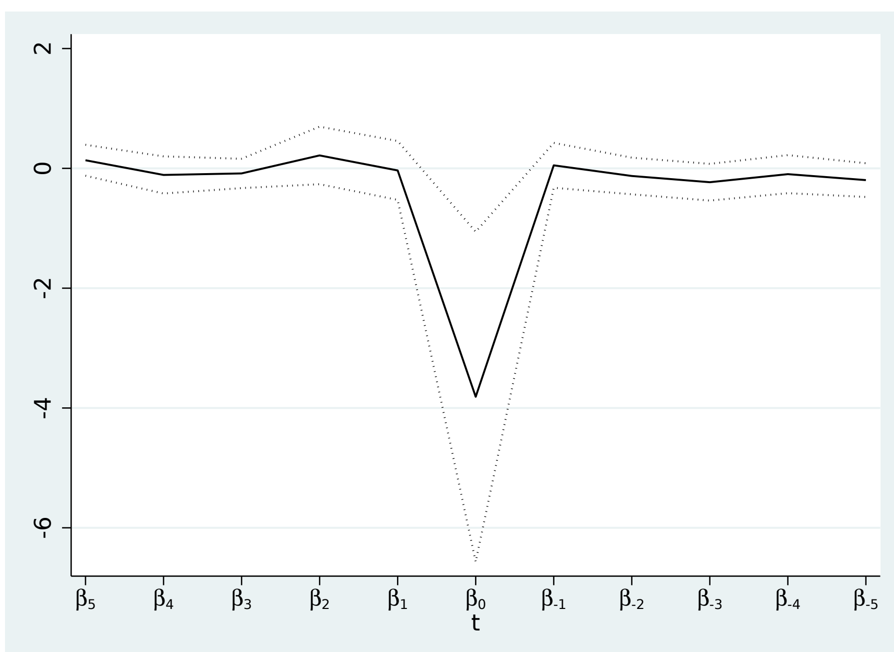
As a placebo test, M21 repeats the M21 Figure 5 analysis using the subset of pharmaceutical patents, where R&D takes longer and cluster size should not have an immediate effect on patenting. I show a screenshot of the published M21 Figure A.1 in Figure A16a. We can see that the output does not match M21 Figure 5, since the β_0 term is incorrectly omitted and has no confidence interval. I run the M21 code and display the output in Figure A16b. In contrast to the published figure, the β_0 term is reported; hence, the M21 code does not produce the same figure that is shown in the published article. As with M21 Figure 5, the code for M21 Figure A.1 does not enforce the requirement that inventors are observed in 11 consecutive years. When I correct the code, the sample size is N= 58 and the β_0 term is dropped as collinear. Since there are very few inventors who create only pharmaceutical patents in 11 consecutive years, the placebo test is not feasible with this sample.

[17] M21 reports a sample size of 21,787. When I run the M21 code, I obtain a sample of 56,887.

Figure A16: Distributed lag model: pharmaceutical patents



(a) M21 Figure A.1



(b) Reproduction of Figure A.1

Note: Panel (a) is a screenshot of M21 Figure A.1 (top panel), which replicates M21 Figure 5 (top panel) using pharmaceutical patents (class code 424). Panel (b) plots the output from running the original code, corresponding to Equation A1. The solid line represents the coefficients, and the dotted lines are 95% confidence intervals. N = 585.

F Cluster quality

In Columns 1-4 of Table A.8, M21 tests whether the elasticity is larger for higher-quality clusters. Cluster quality is defined either as a weighted sum of lifetime patents or citations, or as cluster size measured using inventors with lifetime patents or citations above a threshold. M21 finds that agglomeration effects created by high quality scientists are twice as large.[18] There are two coding issues that affect the results.

First, M21's code calculates cluster size differently than described in the text. For the main results, when inventors patent in different cities or fields in the same year, M21 assigns each inventor the cluster size of their *modal* cluster.[19] For example, when an inventor patents twice in city A and once in city B, that inventor-year observation is assigned city=A and the cluster size of A. In contrast, to calculate cluster size based on inventor quality in Table A.8, M21's code takes the *average* size across clusters for inventors with patents in different cities. That is, M21 aggregates to the inventor-year level, and then assigns the averaged cluster size to each inventor's modal cluster. Continuing the example, the inventor-year observation is assigned city=A and the unweighted average cluster size of A and B. I correct the code by assigning the cluster size of the modal cluster.

Second, in contrast to the main results which measure fractional patents and citations by dividing by the number of coauthors on a patent, M21 uses unadjusted patents and citations to calculate cluster quality. This treats inventors on large teams as higher quality, since they are assigned the full value of a patent or citation, regardless of the number of coauthors making up the team; moreover, team size is positively correlated with cluster size. I repeat the cluster quality analysis using fractional patents and citations.

Table A11, Panel A reproduces the original results, where an inventor's cluster size is averaged across clusters, and cluster size is calculated using unadjusted (not fractional) values of patents and citations. To show how using fractional values affects the results, Panel B uses average cluster size and fractional patents and citations. The estimates tend to be slightly larger. Panel C assigns each inventor the cluster size of their modal cluster, while using unadjusted patents and citations. The elasticity is much smaller for the weighted cluster sizes (Columns 1 and 3), and is close to the baseline estimate when using inventors above a threshold (Columns 2 and 4). The estimates in Panel D are slightly larger compared to Panel C, when using modal cluster size and fractional patents and citations.

Hence, correcting the definition of cluster size produces elasticities about half as large as

[18] Note that this implies a smaller (or negative) effect created by low quality scientists.

[19] "If an inventor has multiple patents with multiple addresses in a single year, I use the modal city for that inventor-year pair. If an inventor has patents in more than one research field or technology class in an year, I use the modal research field or technology class for that inventor-year pair." (p.3334)

those in M21. My results provide mixed evidence for M21's claim that higher quality inventors matter more for agglomeration effects. The estimates using weighted measures of cluster size are smaller than the baseline elasticity estimated on my sample (0.0588), while the estimates measuring cluster size with inventors above a threshold are slightly larger.[20]

Table A11: Reanalysis of M21 Table A.8: cluster quality

	(1) Weighted: patents	(2) Patents $\geq P$	(3) Weighted: citations	(4) Citations $\geq C$
Panel A: Original: average size, unadjusted				
Log size	0.1200	0.1301	0.1097	0.1301
	(0.0430)	(0.0513)	(0.0351)	(0.0529)
Observations	786802	786494	786767	786717
Adjusted R^2	0.244	0.244	0.244	0.244
Panel B: Average size, fractional				
Log size	0.1261	0.1334	0.1149	0.1278
	(0.0446)	(0.0520)	(0.0362)	(0.0494)
Observations	786802	786437	786767	786299
Adjusted R^2	0.244	0.244	0.244	0.244
Panel C: Modal size, unadjusted				
Log size	0.0417	0.0648	0.0153	0.0557
	(0.0107)	(0.0131)	(0.0080)	(0.0139)
Observations	782036	781636	781997	781934
Adjusted R^2	0.243	0.243	0.243	0.243
Panel D: Modal size, fractional				
Log size	0.0500	0.0747	0.0232	0.0595
	(0.0106)	(0.0135)	(0.0080)	(0.0133)
Observations	782036	781563	781997	781396
Adjusted R^2	0.243	0.243	0.243	0.243

Standard errors in parentheses

Notes: Reproduction and reanalysis of Columns 1-4 in M21 Table A.8. The dependent variable is log patents. Cluster size is measured as: the weighted sum of inventors in a given city-field-year cell, with weights reflecting the lifetime number of patents of each inventor (column 1); the number of inventors with a lifetime patent count above a threshold P (column 2); the weighted sum of inventors in a given city-field-year cell, with weights reflecting the lifetime number of citations received (column 3); the number of inventors with a lifetime patent citation count above a threshold C (column 4). Panels A and B follow the original code for Table A.8 and calculate cluster size by averaging across clusters, for inventors with patents in different clusters in the same year. Panels C and D correct the code by assigning each inventor the cluster size from their modal cluster, following the text. Panels A and C follow the original code for Table A.8 and use unadjusted patents and citations, with thresholds in unadjusted values of $P = 3$ and $C = 5$; these correspond to the 75th percentile of patents and the 40th percentile of citations. Panels B and D use fractional patents and citations, with thresholds in fractional values set to the 75th percentiles: $P = 1.5$ and $C = 14$. Fractional values are calculated by dividing by the number of coauthors on a patent. My results in Panel A do not exactly match the original Table A.8 (Columns 1-4), because the M21 cleaning code is unreproducible; see Appendix A. Fixed effects: city × field, city × class, field × year, class × year, inventor, city × year, and firm. Standard errors are clustered by city × research field.

[20]Given my finding in Table A1 that cluster size reduces citations per patent, citations may not be a good proxy for cluster quality.

G Varying the time unit

In Table A.7, M21 varies the time unit of the data, from 1-month to 3-year periods. As with interpolation in M21 Table A.6, the motivation is to address missing observations when inventors do not patent in a given time period. M21 claims that these missing observations create a downward bias, because the extensive margin effect is not estimated when zeros are not observed:

> I expect that when the temporal unit of analysis is short (months), the problem of sample selection and the downward bias are more pronounced. In the extreme, if I were to measure productivity second by second, very few inventor-second pairs would be nonmissing and the selection bias would be large. By contrast, when the temporal unit of analysis is long (two or three years), I expect the problem of sample selection and the downward bias to be less pronounced. In the extreme, if I were to have just one observation per inventor with productivity defined as the number of patents created in all the years in the sample, there would be no selection and both the intensive margin and extensive margin would be reflected in my estimates. (p.3365)

The results in M21 Table A.7 show that the elasticity is negative for 1- and 2-month time units, and grows with the length of the time unit, with the 2- and 3-year elasticities being larger than the baseline.

There are two possible arguments being made. First, that as the time unit shortens, more inventor-time pairs have no patents and are unobserved, so the downward bias is larger; conversely, as the time unit lengthens, fewer inventor-time pairs are unobserved, so the downward bias is smaller. The downward bias here is the difference between the intensive margin effect β_{IM} and the total effect from both the intensive and extensive margins: $\beta_{TOT} = \beta_{IM} + \beta_{EM}$. In other words, the magnitude of the bias is the extensive margin effect $\beta_{EM} > 0$. M21 claims that the bias is increasing as the time unit shortens; that is, β_{EM} is larger when the time unit is shorter. However, β_{EM} can be estimated only if we add zeros to the data by interpolating the unobserved inventor-year pairs with zero patents. Table A.7 does not interpolate, but estimates β_{IM} for different time units. Hence, while it may be true that β_{EM} differs as the time unit changes, this argument does not explain the results in Table A.7, which show estimates of β_{IM}.

A second interpretation is that, as the time unit lengthens, the more the extensive margin is 'reflected' in the estimated elasticity; conversely, as the time unit shortens, the estimate 'reflects' less of the extensive margin. In other words, with a longer time unit, more inventor-time pairs with zero patents (defined under the shorter time unit) are aggregated into a nonzero inventor-time pair (defined under the longer time unit). It is not clear what it means for the estimate to 'reflect' more of the extensive margin in this case. Suppose an inventor patents in 2000 and 2002, but not in 2001 or 2003. Using a 1-year time unit, the observed data is {2000: 1, 2002: 1}. Using a 2-year time unit, we would observe {2000/2001: 1, 2002/2003: 1}. That is, in each case we observe only the inventor-time pairs with nonzero patents. While there is an extensive margin effect from, for example, not patenting in 2001 and patenting in 2002, this is not captured in either case. Since no zeros are observed, no extensive margin can be estimated.

The intensive margin effect does change with the length of the time unit, because the identifying variation is different. For example, with shorter time units, fewer inventors patent at the same time, so measured cluster sizes are smaller, and fewer patents are summed within each inventor-time bin. Hence, the independent and dependent variables differ. Moreover, with shorter time units, the time fixed effects more finely control for variation over time. These changes explain the pattern of growing elasticities in Table A.7.

Furthermore, there is a coding issue with Table A.7. M21 does not recalculate cluster size at the level of the new time unit, but instead uses the baseline 1-year cluster size; M21 assigns the 1-year value to months within a year, and the first 1-year value in a j-year period to that period. For example, with a 1-month time unit, M21 assigns the cluster size for 2000 to each month in that year; with a 2-year time unit, M21 assigns the cluster size for 2000 to the 2-year period covering 2000-2001.

I reproduce the original results in Panel A of Table A12. Panel B recalculates cluster size at the appropriate time unit. While M21 found negative effects for 1- and 2-month time units, I find large negative and significant effects for 1- to 3-month time units. Similar to the original, the effect size continues to grow as the time unit lengthens. In contrast to the original, the coefficients are larger in magnitude: the 1-month effect is -0.0903 (vs. -0.0258) and the 3-year effect is 0.1972 (vs 0.1613).

Why is the elasticity negative when using short time units? If the intensive margin effect is positive, then the estimated elasticity should still be positive when the extensive margin is omitted. The explanation is based on how cluster size changes with the time unit. Because patents are measured on a per-coauthor basis, the negative elasticity is driven by the mechanical negative relationship between patents per coauthor and the number of coauthors in the same cluster. By definition, coauthors patent at the same time as the focal inventor. As the time-unit shortens, coauthors make up a larger share of an inventor's cluster; in the extreme case, with patents measured per second, the only other inventors in a cluster would be the focal inventor's coauthors. So cluster size becomes a proxy for the number of coauthors, which generates a negative correlation between patents per coauthor and cluster size. Table A12 Panel C uses unadjusted patents instead of patents per coauthor as the dependent variable; since coauthors are not included in the denominator, there should not be a negative correlation between total patents and cluster size. As expected, the elasticity is positive in each time unit. Panel D shows the elasticity between the number of coauthors (team size) and cluster size. The elasticity is about twice as large when the time unit is 1-month compared to the baseline of 1-year. Hence, the negative elasticities are explained by the share of coauthors in the cluster.

Table A12: Reanalysis of M21 Table A.7: recalculating cluster size at new time unit

	(1) 1-Month	(2) 2-Month	(3) 3-Month	(4) 6-Month	(5) 1-Year	(6) 2-Year	(7) 3-Year
Panel A: Original							
Log size	-0.0258	-0.0147	-0.0028	0.0229	0.0588	0.1290	0.1613
	(0.0084)	(0.0091)	(0.0096)	(0.0107)	(0.0139)	(0.0144)	(0.0188)
Observations	1286847	1199567	1130305	977635	786680	572459	460922
Adjusted R^2	0.313	0.293	0.282	0.260	0.242	0.211	0.198
Panel B: Corrected							
Log size	-0.0903	-0.0681	-0.0422	0.0045	0.0588	0.1512	0.1972
	(0.0045)	(0.0061)	(0.0073)	(0.0094)	(0.0139)	(0.0179)	(0.0255)
Observations	1242696	1181503	1120045	974633	786680	573500	461906
Adjusted R^2	0.338	0.306	0.290	0.264	0.242	0.212	0.198
Panel C: Log unadjusted patents							
Log size	0.0332	0.0392	0.0537	0.0783	0.1158	0.1826	0.2309
	(0.0024)	(0.0035)	(0.0044)	(0.0064)	(0.0108)	(0.0167)	(0.0240)
Observations	1242696	1181503	1120045	974633	786680	573500	461906
Adjusted R^2	0.142	0.151	0.167	0.202	0.238	0.255	0.254
Panel D: Log team size							
Log size	0.1279	0.1124	0.1025	0.0826	0.0684	0.0475	0.0541
	(0.0032)	(0.0038)	(0.0045)	(0.0054)	(0.0071)	(0.0092)	(0.0105)
Observations	1242696	1181503	1120045	974633	786680	573500	461906
Adjusted R^2	0.475	0.467	0.466	0.472	0.483	0.495	0.499

Standard errors in parentheses

Notes: The dependent variable is log patents per coauthor in panels A and B; log unadjusted patents (not adjusting for coauthors) in panel C; and log team size (average number of coauthors) in panel D. Panel B corrects the original code by calculating cluster size at the level of the corresponding time unit; the M21 code uses the 1-year cluster size for all time units. My results in Panel A do not exactly match the original Table A.7 because the M21 cleaning code is unreproducible; see Appendix A. Fixed effects: city $\times$ field, city $\times$ class, field $\times$ year, class $\times$ year, inventor, city $\times$ year, and firm. Standard errors are clustered by city $\times$ research field.

H Team size

In Table A.8, M21 investigates how team size, measured as the number of coauthors on a patent, affects the relationship between cluster size and patenting. M21 states: "If larger teams are both more productive and more likely to be in larger clusters, team size could be an important omitted variable." (p.3366) The effect size is larger when controlling for team size or excluding coauthors when measuring cluster size. These results are conceptually flawed because patents are already measured on a per-coauthor basis, so adjusting for team size again means controlling for the same variable twice.

When controlling for a quadratic in team size (column 5), the effect of cluster size on patenting is nearly double the baseline magnitude. However, team size is not a confounder for the effect of cluster size on patenting, but a mediator, and should not be controlled for. With agglomeration effects, larger clusters enable collaboration and lead to larger teams, which in turn increase inventors' productivity. The total effect of cluster size on patenting includes the direct effect as well as the indirect effect through team size. Hence, controlling for team size should reduce the estimated elasticity by controlling for the indirect effect. But M21 finds a larger elasticity when controlling for team size, implying that the indirect effect is negative.

The explanation for this puzzle is that M21 measures coauthored patents using fractional attribution, assigning each coauthor an equal share of a patent. Hence, controlling for team size means controlling for the denominator of a rate variable (patents per coauthor). As shown in Table A13, Column 1, there is a mechanical negative relationship between patents per coauthor and the number of coauthors. Moreover, team size is positively related to cluster size, since a patent having more coauthors implies there are more inventors contributing to cluster size (see Table A13, Column 2). Hence, using the formula for omitted variable bias, the bias from omitting team size is negative, which explains the larger elasticity when controlling for team size.

A more fundamental problem is that team size is observed only for patents that are created, which complicates estimation of the effect of team size on patenting. We do not observe the full distribution of team size, since team size is unobserved for teams that did not complete a patent. In this dataset, team size is a characteristic of created patents rather than a factor influencing the creation of patents.

In Column 6 of Table A.8, M21 recalculates cluster size to exclude all members of the focal inventor's team. The M21 code subtracts the inventor's average team size in a given year.[21]

[21] Note that this undercounts the number of coauthors, when the coauthors on different patents are distinct. For example, if inventor A coauthors with B and C on one patent, and separately coauthors with D and E on another, the average number of coauthors is 2, but there are 4 coauthors contributing to A's cluster size. This approach also overcounts by subtracting coauthors from different cities (who do not contribute to A's cluster size).

Table A13: Mechanical correlations of team size

	(1) Log patents	(2) Log team size
Log team size	-0.9049	
	(0.0033)	
Log cluster size		0.0684
		(0.0071)
Observations	787061	786680
Adjusted R^2	0.450	0.483

Standard errors in parentheses

Notes: *Patents* is the total number of patents per coauthor created by an inventor in a year. *Team size* is the number of coauthors listed on a patent, averaged by inventor-year. Fixed effects: city × field, city × class, field × year, class × year, inventor, city × year, and firm. Standard errors are clustered by city × research field.

The estimated elasticity is now over three times larger than the baseline estimate (0.237 vs. 0.0676). As above, adjusting the independent variable for team size is inappropriate when the dependent variable is already normalized by team size.[22]

[22]Furthermore, M21's code for the Column 6 result deviates from the text in the same way as the cluster quality results, as discussed in Section F. To calculate cluster size excluding team members, M21 takes the average size across clusters for inventors with patents in different clusters in the same year. This contrasts with the main results, where M21 assigns each inventor the cluster size of their modal cluster.

I Interpolating missing observations

When inventors do not patent in a year, they have zero patents, but their city and field are unobserved. Since these inventor-year observations are missing, the main estimates do not capture the extensive margin effect of cluster size on patenting.[23] In Table A.6, M21 interpolates missing inventor-year observations when inventors have gaps of length 1 or 2 years, and have the same city, field, class, and firm in the years immediately before and after the gap. Specifically, M21 assigns patent=0 and the values of city, field, class, and firm from the adjacent years. M21 then estimates the effect of cluster size on patenting with the extended sample, using the inverse hyperbolic sine and log(patents + 1) to account for zero values in the dependent variable.

M21's code contains an error that biases the estimates downwards. The code to interpolate two-year gaps fills in only the second year of the gap, leaving the first year as missing. I correct the code to fill in both years of two-year gaps. Table A14 presents the original and corrected interpolation. Since the extensive margin effect is positive, adding more zero observations further increases the elasticity.

Table A14: Reanalysis of M21 Table A.6: interpolation

	(1) Baseline	(2) 1 year	(3) One year (2-year gap)	(4) Two years (2-year gap)
Panel A: Inverse hyperbolic sine				
Log size	0.0470	0.0610	0.0637	0.0672
	(0.0088)	(0.0084)	(0.0083)	(0.0080)
Observations	767636	806612	819379	850036
Adjusted R^2	0.251	0.221	0.217	0.221
Panel B: Log(patents + 1)				
Log size	0.0354	0.0467	0.0489	0.0517
	(0.0066)	(0.0063)	(0.0063)	(0.0060)
Observations	767636	806612	819379	850036
Adjusted R^2	0.254	0.221	0.216	0.221

Standard errors in parentheses

Notes: The dependent variable is patents, as specified in the panel titles. Inventor-year observations are interpolated by assigning patent=0 and the city, field, class, and firm from the year preceding the gap. Column 2 interpolates one-year gaps. Column 3 interpolates one-year gaps and the second year of two-year gaps, following the original code. Column 4 corrects the code by interpolating one-year gaps and both years of two-year gaps. My results in Columns 1-3 do not exactly match the original Table A.6 because the M21 cleaning code is unreproducible; see Appendix A. Fixed effects: city × field, city × class, field × year, class × year, inventor, city × year, and firm. Standard errors are clustered by city × research field.

[23] For this sample of star inventors, the extensive margin should be interpreted as an inventor joining a research project in a given year, rather than a non-inventor becoming an inventor.

References

Chen, J. and J. Roth (2023). Logs with Zeros? Some Problems and Solutions. *The Quarterly Journal of Economics 139*(2), 891–936.

Correia, S., P. Guimarães, and T. Zylkin (2020, March). Fast poisson estimation with high-dimensional fixed effects. *The Stata Journal: Promoting communications on statistics and Stata 20*(1), 95–115.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2023). Measuring technological innovation over the long run: Extended data. Github repository. Downloaded from https://github.com/KPSS2017/Measuring-Technological-Innovation-Over-the-Long-Run-Extended-Data on March 24, 2025.

Moretti, E. (2021a). Data and code for: The effect of high-tech clusters on the productivity of top inventors. `https://doi.org/10.3886/E140662V1`. American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor].

Moretti, E. (2021b). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review 111*(10), 3328–75.

Schmidheiny, K. and S. Siegloch (2023). On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *Journal of Applied Econometrics 38*(5), 695–713.

Zucker, L. G., M. R. Darby, and J. Fong (2014). Communitywide database designs for tracking innovation impact: Comets, stars and nanobank. *Annals of Economics and Statistics* (115/116), 277–311.