Extended Commentary on Moretti (2021)

Michael Wiebe
Independent*

November 5, 2023

*maswiebe@gmail.com

1 Introduction

This document presents additional results from my reanalysis of Moretti (2021). Section 2.3 shows the corrected event study that I refer to in Wiebe (2023). The remainder of the section explores the event study using a DD and DDD with either a binary or continuous treatment variable, and using the imputed sample. Section 3 shows the IV regressions using the instrument defined in levels, also referred to in Wiebe (2023).

Section 4 reproduces the original Table 3, including a regression with a firm-year fixed effect. In Section 5, I show that cluster size is negatively correlated with patents for the bottom 90% of inventors. Using an interaction term instead of a subsample regression, the negative effect disappears, but the positive effect for top 10% inventors is much smaller. Section 6 shows that the effect size is smaller for movers compared to stayers.

In Section 7, I explore whether geographic proximity is the underlying mechanism, by running a horse race of inventors in the same field from own-city vs. other-city. Surprisingly, the correlation is large and negative for the number of other-city inventors, suggesting crowding-out. Section 8 performs a similiar horse race for own-firm vs other-firm inventors. The correlation is negative for inventors from other firms. These results are preliminary, since it is tricky to think about what fixed effects are appropriate.

Moretti counts inventors by counting patents. If an inventor does not patent in a year, they do not contribute to measured cluster size. Moretti addresses this partially by imputing missing observations for gaps of length 1 and 2. In Section 9, I extend this by imputing gaps of all lengths. The effects are slightly larger, plausibly capturing a positive extensive margin. Formally modelling the intensive and extensive margins would be informative.

Section 10 investigates bias from aggregating the data from patent level to inventor-year level. Patents are assigned to an inventor's modal cluster, and modal clusters are larger, generating a small upward bias. In Section 11, I redo the Table A7 results on varying the time unit, since Moretti does not fully reconstruct the data at the new time unit. Moretti finds negative results for shorter time units, but does not provide a compelling explanation.

2 Event study

2.1 Original event study

To test for sorting, Moretti performs an event study using variation in cluster size from inventors who move across cities exactly once. That is, 'stayers' who never move are excluded, so the event study does not have a never-treated group. To generate a treatment-control comparison, Moretti uses average cluster size before and after the move as a continuous treatment variable. Specifically, Moretti interacts pre-move average cluster size with the pre-move event-time indicators, and post-move average cluster size with the post-move event-time indicators.[1] The regression equation is

$$\begin{aligned}\ln y_{ijfct} = &\sum_{s=-5}^{-1} \beta_s \text{Size}_{-ifc}^{pre} \times \mathbb{1}\{t=s\} + \sum_{s=0}^{5} \beta_s \text{Size}_{-ifc}^{post} \times \mathbb{1}\{t=s\} \\ &+ d_{cf} + d_{ck} + d_{ft} + d_{kt} + d_{ct} + d_i + d_j + \varepsilon_{ijfkct}.\end{aligned} \tag{1}$$

Here y is the number of patents by inventor i in firm j, research field f, city c, and year t; k is the research class.

The original event study does not interact the treatment variable with a $t=0$ indicator, but uses time-varying cluster size. Hence, β_0 is estimated using data from all event-years, instead of capturing the effect in $t=0$. I correct the code in Figure 1 by interacting post-move cluster size with the $t=0$ indicator. (This is the graph reported in Wiebe (2023).)

[1] Moretti excludes $t=0$ when calculating post-move average cluster size, despite being the inventor's first year in the new city.

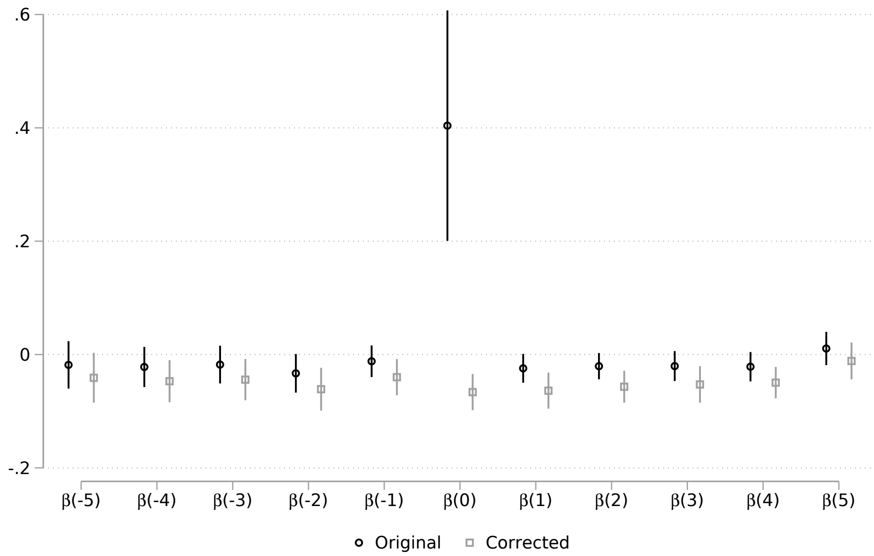


Figure 1: Replication and correction of Figure 6 event study

Note: Event study coefficients from Equation 1. *Original* estimates β_0 using time-varying $Size_{-ifct}$ and without interacting with $\mathbb{1}\{t=0\}$. *Corrected* estimates β_0 using $Size^{post}_{-ifc} \times \mathbb{1}\{t=0\}$; following Moretti, $Size^{post}_{-ifc}$ is calculated excluding $t=0$. N=18,389 in *Original*, N=18,390 in *Corrected*. Standard errors are clustered by city $\times$ research field. Moretti's Figure 6 switches the leads and lags, for example, putting β_{-5} as the last coefficient and β_5 as the first.

2.2 Include $t = 0$ in post-move average cluster size

In Figure 2 I define $Size^{post}_{-ifc}$ to include $t = 0$. The effect in $t = 0$ is smaller, but the overall picture is the same.

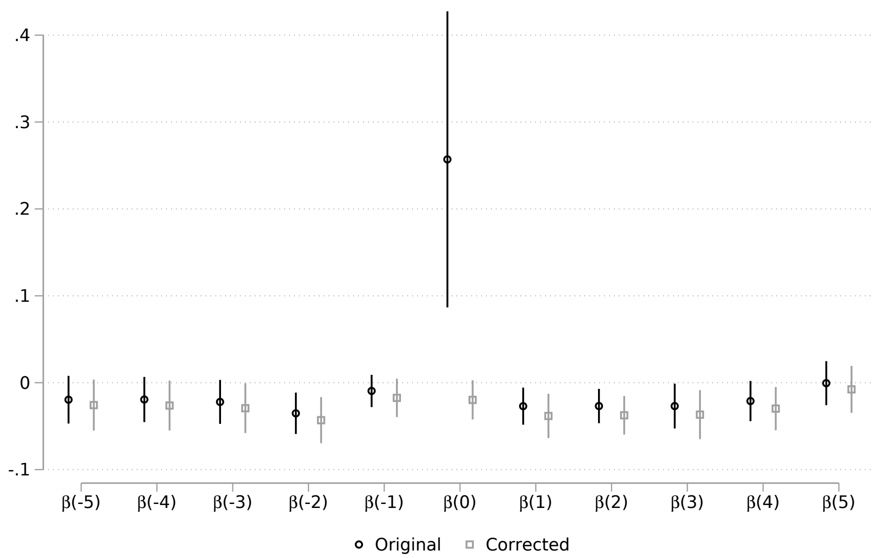


Figure 2: Replication and correction of Figure 6 event study

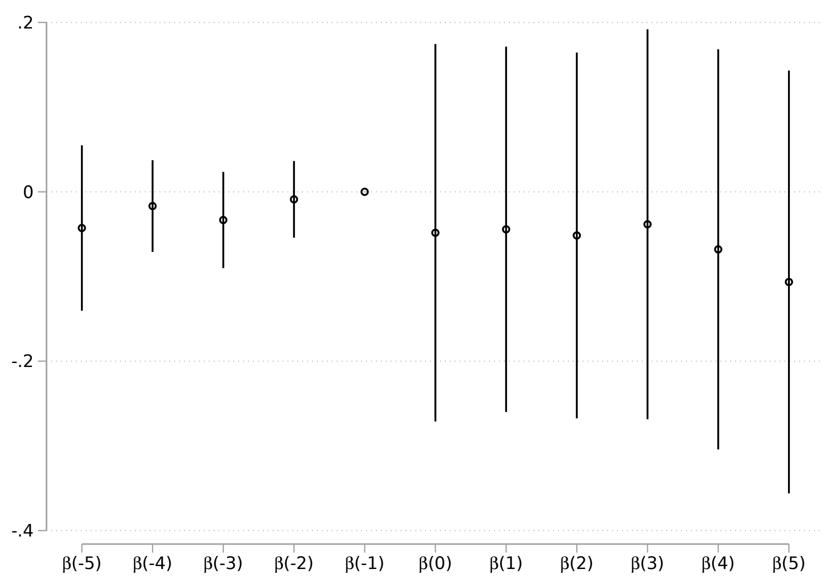
Note: Event study coefficients from Equation 1. *Original* estimates β_0 using time-varying $Size_{-ifct}$ and without interacting with $\mathbb{1}\{t = 0\}$. *Corrected* estimates β_0 using $Size^{post}_{-ifc} \times \mathbb{1}\{t = 0\}$. $Size^{post}_{-ifc}$ is calculated including $t = 0$. N=25,510 in *Original*, N=25,511 in *Corrected*. Standard errors are clustered by city $\times$ research field.

2.3 Corrected event study

I recode the event study in Figure 3 to follow standard practices, but again find null results. I use a constant treatment variable (difference in average cluster size before and after the move: $\text{Size-diff}_{-ifc} = \text{Size}^{post}_{-ifc} - \text{Size}^{pre}_{-ifc}$) where average post-move size is calculated including year 0; interact the treatment variable with all year indicators; omit $t-1$ as a reference year[2]; and restrict the sample to event years [-5,5].

Curiously, the width of the confidence intervals jumps in $t = 0$. This seems to be caused by the combination of using a constant treatment variable and using a fully saturated regression (ie. restricting the [-5,5] and including all interactions). One intuitive explanation is heterogeneous treatment effects by inventors who move to larger vs smaller cities. But this is not heterogeneity, since a constant treatment effect is consistent with a drop in patents for inventors moving to a smaller city, and an increase in patenting for inventors moving to a larger city. Let me know if you can figure out an explanation for this.

Figure 3: Corrected event study



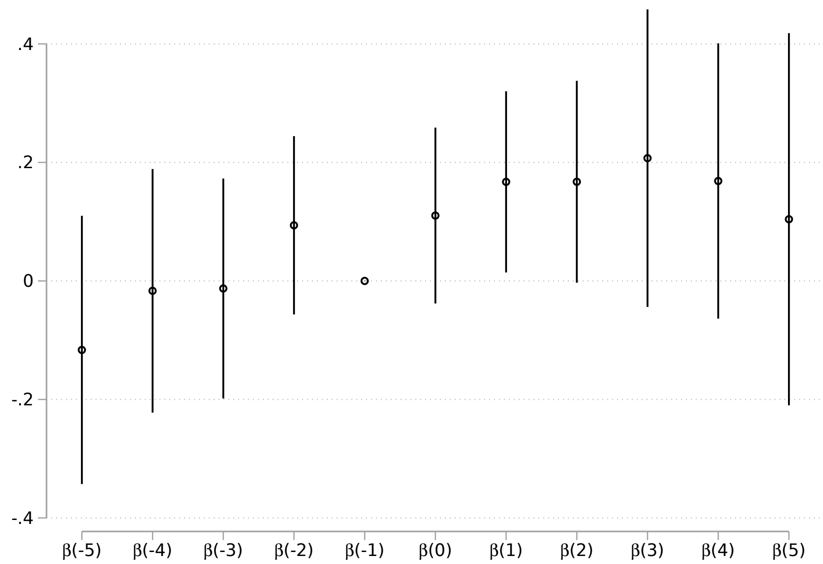
Note: Event study coefficients with five differences from Moretti's Fig. 6: (1) I use a constant treatment variable: the difference in average cluster size before and after the move, defined using event years [-5,5]. (2) Average post-move size is calculated including year 0. (3) I interact the treatment variable with all year indicators, including $t = 0$. (4) I omit $t-1$ as the reference year. (5) I restrict the sample to event years [-5,5]. N=15,156. Standard errors are clustered by city × research field.

[2]In the original specification, years outside of [-5,5] are used as the reference period.

2.4 Binary treatment variable

In Figure 4 I use Move-up = $\mathbb{1}\{\text{Size-diff} > 0\}$ as a binary treatment variable, instead of the continuous Size-diff. This produces a slight positive effect, but there appears to be a pre-trend.

Figure 4: Event study: binary treatment variable



Note: N=15,172. Standard errors are clustered by city × research field.

2.5 DDD including stayers

Moretti's event study uses a sample of movers, so there are no never-treated observations. Here I include stayers (non-movers), which allows me to run a triple-difference:

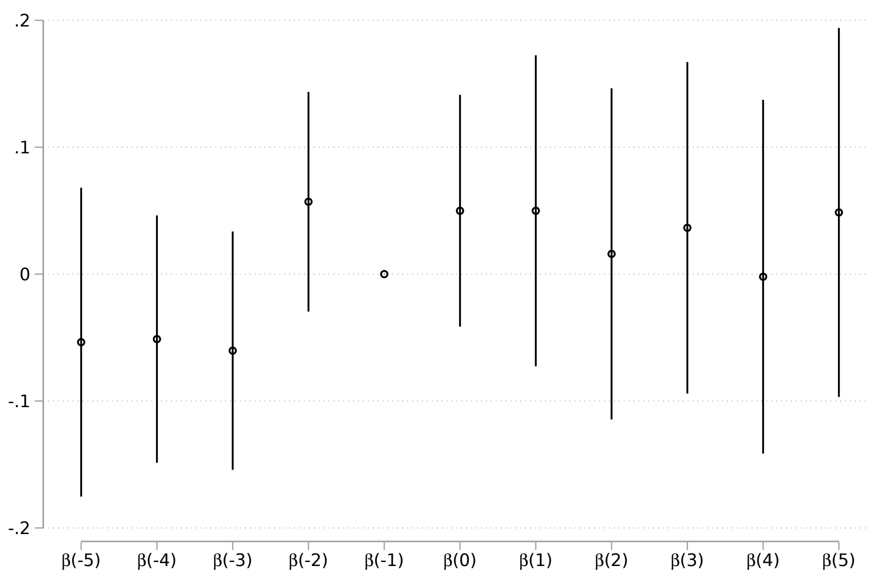
- Binary: Mover × Post + Move-up × Post
- Continuous: Mover × Post + Size-diff × Post

The DD compares move-up and move-down, with no control group, while the DDD compares move-up and move-down relative to stayers. Note that Move-up is a subset of Mover, so Mover × Move-up = Move-up. Also note that Post_t is not collinear with year fixed effects, since it is always 0 for stayers.

2.5.1 Binary treatment

Figure 5 shows the event study coefficients on Move-up interacted with the event time indicators. (The regression also includes Mover interacted with event time indicators.)

Figure 5: Event study: DDD, binary treatment

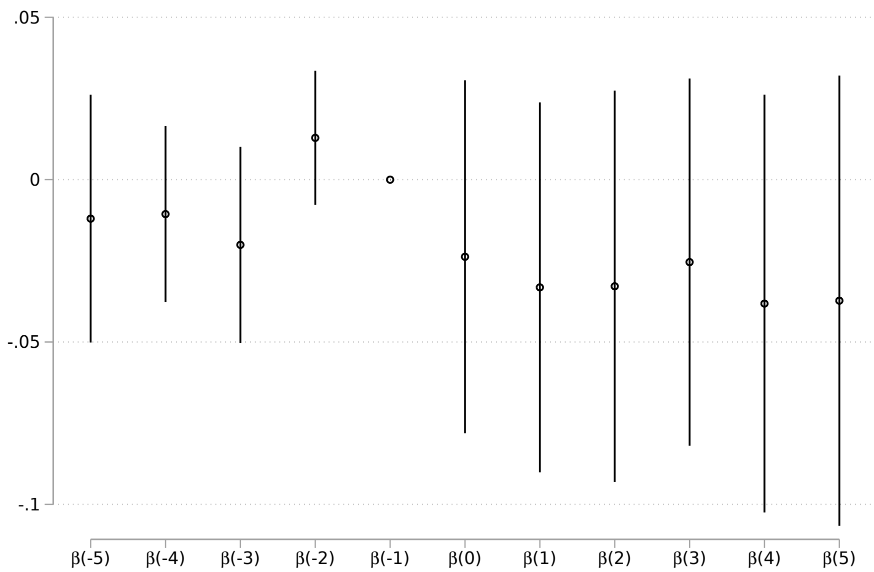


Note: N=465,408. Standard errors are clustered by city × research field. Sample includes stayers.

2.5.2 Continuous treatment

Figure 6 shows the event study coefficients on Size-diff interacted with the event time indicators. (The regression also includes Mover interacted with event time indicators.)

Figure 6: Event study: DDD, continuous treatment



Note: N=465,388. Standard errors are clustered by city × research field. Sample includes stayers.

2.6 Event study using imputed data

In Section 9 I impute the missing observations (where inventors do not patent), using either the origin or destination city. Here I rerun the event study results on the imputed sample, which is more balanced than the original sample. As in Moretti, I include only moves where the timing is identified; ie. we observe the inventor patenting in the origin and destination cities in consecutive years.[3]

Using a DD with a binary treatment, for either origin (Figure 7) or destination (Figure 11) imputation, gives a strange pattern with the coefficients being strongly negative except in $t = 0$. One possible explanation is mechanical: the sample is constructed so that inventors patent in the last year in the origin city ($t = -1$) and in the first year in the destination city ($t = 0$). With imputation, we fill in patent=0 for missing observations, which correspond to years [-5,-2] and [1,5]. This generates a negative correlation for those years relative to $t = -1$. But this pattern holds only for the DD with a binary treatment, and doesn't hold for the continuous DD or the binary DDD. I don't have an explanation for this.

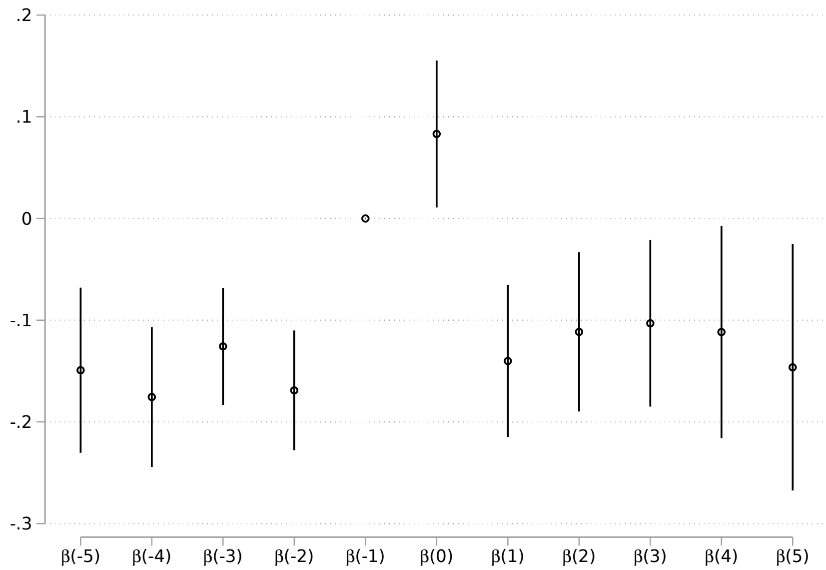
2.6.1 Binary treatment: origin imputation

Figure 7 shows the DD event study coefficients on Move-up interacted with the event time indicators.

Figure 8 shows the DDD event study coefficients on Move-up interacted with the event time indicators.

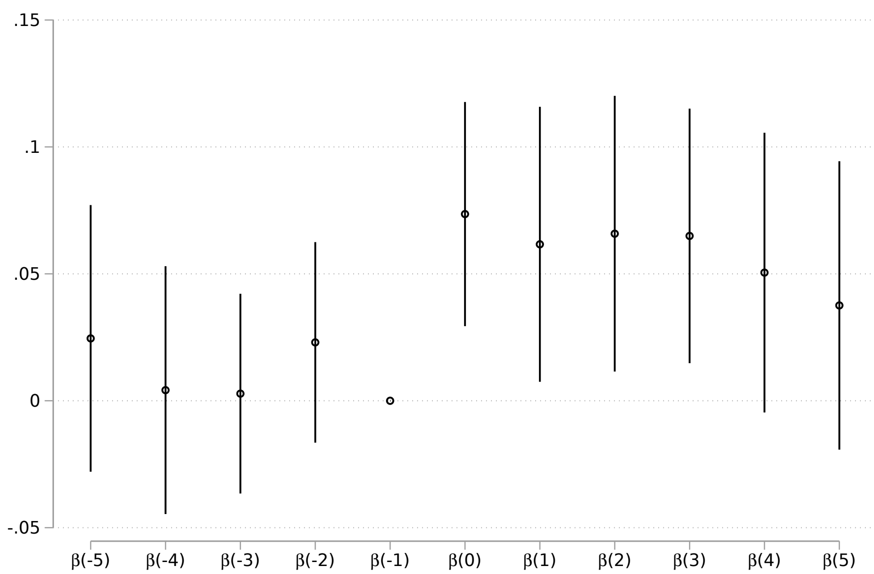
[3]We could also include moves with unidentified timing, where there is a gap between patents in two different cities; this would require using origin or destination imputation, which could lead to mechanical results. For example, with origin imputation, we impute y=0, then a move occurs with y=1 in the destination city. This generates a positive correlation between moving and patenting. Similarly, with destination imputation, y=1 in the origin city, followed by an imputed y=0 for the destination city, generating a negative correlation between moving and patenting. But this seems incomplete, since the event study uses the change in cluster size, not merely moving.

Figure 7: Event study: DD, binary treatment, origin imputation



Note: N=27,670. Standard errors are clustered by city × research field. Sample excludes stayers.

Figure 8: Event study: DDD, binary treatment, origin imputation

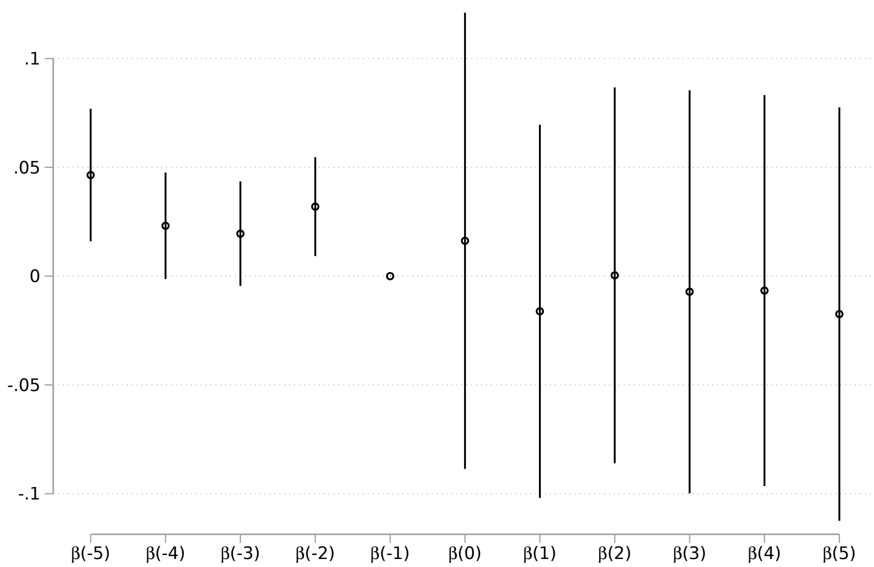


Note: N=921,864. Standard errors are clustered by city × research field. Sample includes stayers.

2.6.2 Continuous treatment: origin imputation

Figure 9 shows the DD event study coefficients on Size-diff interacted with the event time indicators.

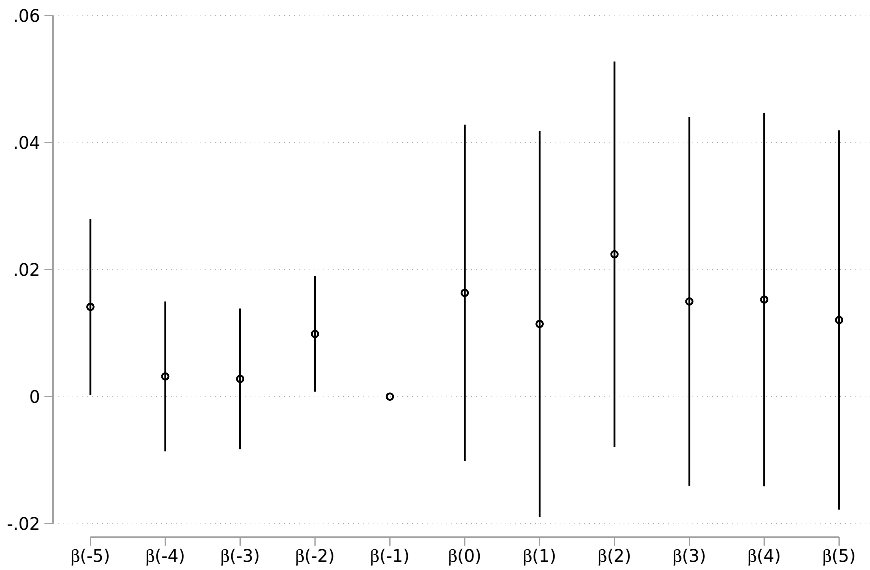
Figure 9: Event study: DD, continuous treatment, origin imputation



Note: N=27,650. Standard errors are clustered by city × research field. Sample excludes stayers.

Figure 10 shows the DDD event study coefficients on Size-diff interacted with the event time indicators.

Figure 10: Event study: DDD, continuous treatment, origin imputation

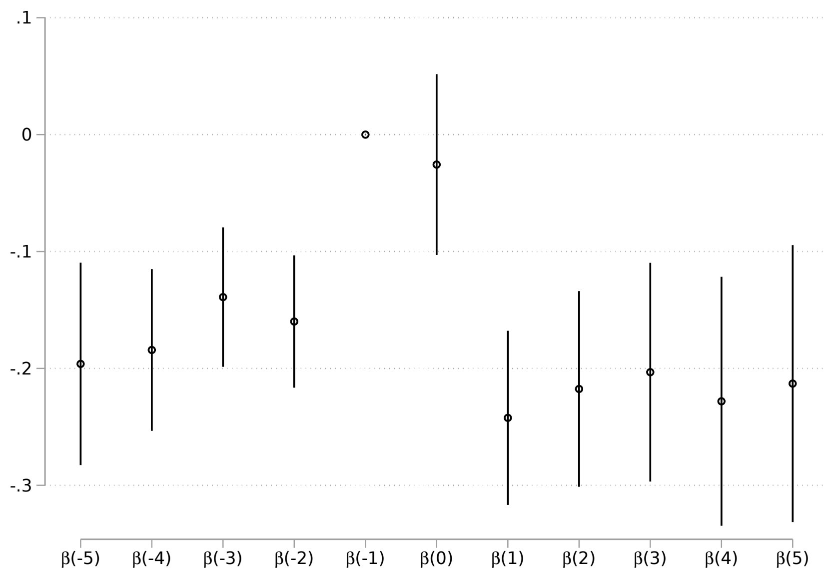


Note: N=921,837. Standard errors are clustered by city × research field. Sample includes stayers.

2.6.3 Binary treatment: destination imputation

Figure 11 shows the DD event study coefficients on Move-up interacted with the event time indicators.

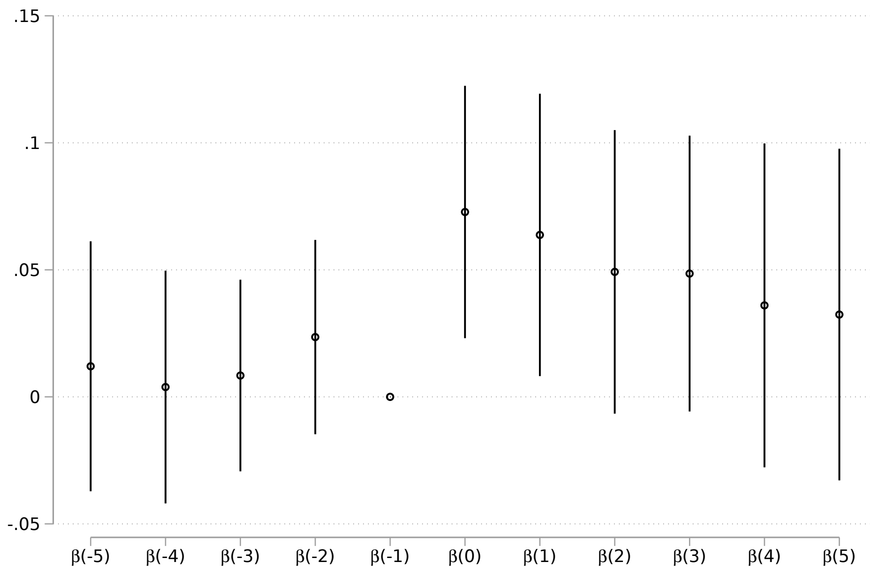
Figure 11: Event study: DD, binary treatment, destination imputation



Note: N=27,367. Standard errors are clustered by city × research field. Sample excludes stayers.

Figure 12 shows the DDD event study coefficients on Move-up interacted with the event time indicators.

Figure 12: Event study: DDD, binary treatment, destination imputation

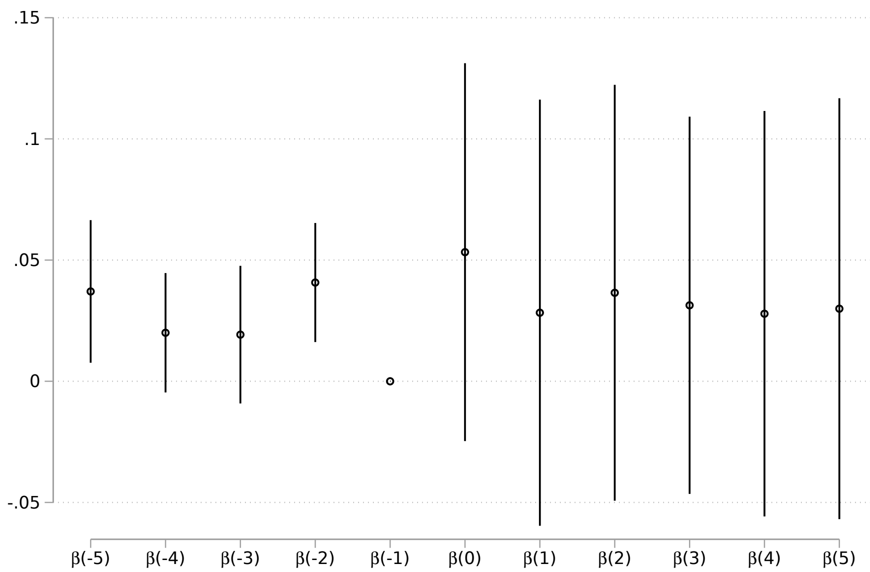


Note: N=907,562. Standard errors are clustered by city × research field. Sample includes stayers.

2.6.4 Continuous treatment: destination imputation

Figure 13 shows the DD event study coefficients on Size-diff interacted with the event time indicators.

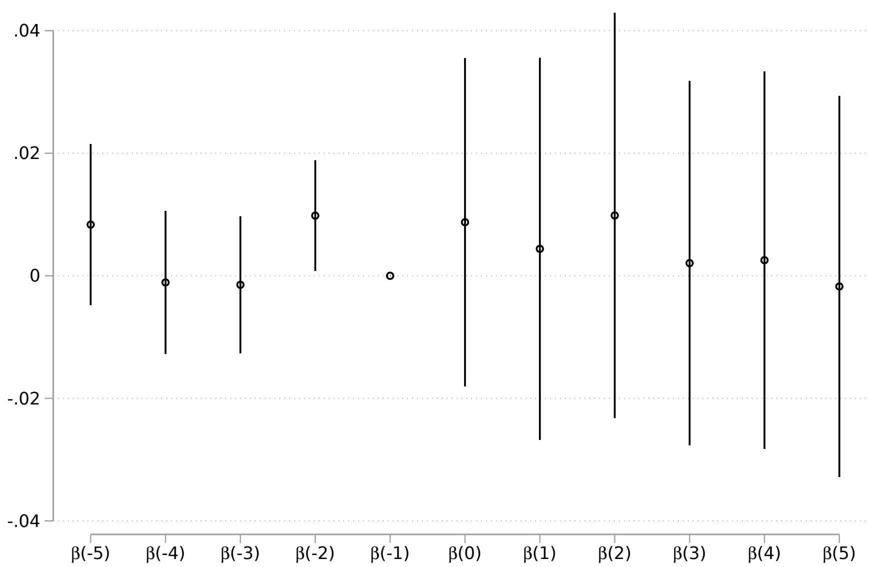
Figure 13: Event study: DD, continuous treatment, destination imputation



Note: N=27,346. Standard errors are clustered by city × research field. Sample excludes stayers.

Figure 14 shows the DDD event study coefficients on Size-diff interacted with the event time indicators.

Figure 14: Event study: DDD, continuous treatment, destination imputation



Note: N=907,534. Standard errors are clustered by city × research field. Sample includes stayers.

2.7 Figure 5 non-event study

In Figure 5, Moretti performs a non-standard event study, including 5 leads and lags of time-varying cluster size. Since there is no single treatment, it is misleading to present an event study graph centered at $t = 0$, and to describe it as an event study with a single treatment ("To interpret this figure, suppose that a change in cluster size takes place at time t = 0." p.3352) On the contrary, cluster size is changing in every time period. Hence, I am unclear how this is informative in testing for pre-trends, with rising star inventors sorting into large cities.

The large effect for β_0 seems partly explained by the effect size increasing in the productivity of the sample. From Table 9 we know that the effect grows as the sample is limited to top inventors. The average lifetime patents for the Figure 5 estimation sample is 23.5, while the 99.5 percentile in the full sample is 15. This makes sense, since we're selecting on inventors who have at least 12 patents (ignoring fractional attribution), since the estimation sample includes inventors with at least 12 observations, to match the 11 regressors; two observations are needed to identify the inventor fixed effect.

One question I have is whether it is appropriate to include leads/lags of X in an unbalanced panel (since years where inventors do not patent are missing). Another issue with leads/lags is inventors changing city, field, class, or firm.[4]

[4] Imputing missing observations would address the unbalanced panel.

3 IV estimates

Moretti runs the IV regressions in first-differences, while the main OLS regressions are in levels. This is not explicitly justified. Here I redo the IV regressions in levels, defining the instrument as the number of inventors in other cities (in the same field), normalized by the number nationally by field. Let $N_{jf(-c)t}$ be the number of inventors at firm j in research field f in all cities excluding c in year t, and let N_{ft} be the number in field f and year t. Let D_{jfct} be an indicator for firm j employing at least one inventor in city c and field f in year t, to capture firms active in c. Then the IV is

$$IV_{jfct} = \sum_{s \neq j} D_{sfct} \frac{N_{sf(-c)t}}{N_{ft}}, \tag{2}$$

where the sum is taken across all firms excluding j.

Table 1 shows that this produces a first stage, in contrast to the results using the corrected first-differenced instrument from Wiebe (2023). However, the 2SLS estimates are again nonsignificant.[5]

[5] It might be interesting to run the first-differenced results on the imputed dataset, since this would avoid restricting the estimation sample to observations where inventors patent in consecutive years.

Table 1: Replication of Table 5: level instrument

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 2SLS								
Log size	0.0376	0.0231	0.0017	0.1270**	0.0830	0.0334	-0.2215	-0.3844
	(0.0362)	(0.1023)	(0.1301)	(0.0564)	(0.0582)	(0.1164)	(0.2395)	(0.2879)
Observations	419540	419495	412824	412824	411639	388526	387587	374840
Panel B: First stage								
IV	20.46***	10.83***	9.50***	9.86***	9.13***	6.06***	2.37***	2.31***
	(3.26)	(2.99)	(2.90)	(2.84)	(2.77)	(1.71)	(0.76)	(0.72)
Observations	419540	419495	412824	412824	411639	388526	387587	374840
F-statistic	39.40	13.14	10.70	12.06	10.86	12.50	9.72	10.42
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Dependent variable is the log number of patents in a year. The instrument is defined in levels instead of first-differences. Fixed effects are the same as Table 3 in Moretti (2021). Standard errors are clustered by city × research field.

4 Original OLS results

I reproduce the main fixed-effects regressions using OLS from Table 3. The results are very similar but not identical, likely due to small changes in sample size.
In Column 9, I control for Firm × Year fixed effects. This reduces the effect size by 3x.

Table 2: Replication of Table 3 fixed-effects regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log size	0.0515***	0.0764***	0.0887***	0.0912***	0.0681***	0.0924***	0.0546***	0.0681***	0.0169
	(0.0081)	(0.0167)	(0.0186)	(0.0092)	(0.0086)	(0.0098)	(0.0114)	(0.0134)	(0.0132)
Observations	931929	931896	923823	923823	923124	922279	921727	786303	685171
Adjusted R^2	0.042	0.044	0.060	0.064	0.070	0.225	0.227	0.250	0.280
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year				Yes	Yes	Yes	Yes	Yes	Yes
Class × year					Yes	Yes	Yes	Yes	Yes
Inventor						Yes	Yes	Yes	Yes
City × year							Yes	Yes	Yes
Firm								Yes	Yes
Firm × Year									Yes

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3, except for Column 9. Sample restricted to top 10% inventors, by lifetime patents. Standard errors are clustered by city × research field.

5 Heterogeneity by top inventors

The main results in Table 3 use the top 10% of inventors by lifetime patents. In Table 9, Moretti reduces the sample from 100%, to top 25% of inventors by lifetime patents, and so on, down to the top 0.5% of inventors. But he doesn't test for the effect for low-productivity inventors. In Table 3, I restrict the sample to the bottom 90% of inventors by lifetime patents. The results are negative, implying that larger clusters reduce patenting. Note that running a separate regression by subgroup means that fixed effects are not shared by above- and below-90% inventors.

Table 3: Main results for bottom 90% inventors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log size	-0.0290***	-0.0593***	-0.0463***	-0.0439***	-0.0516***	-0.0095	-0.0411***	-0.0306***
	(0.0069)	(0.0102)	(0.0102)	(0.0060)	(0.0054)	(0.0066)	(0.0080)	(0.0095)
Observations	1906949	1906946	1898519	1898519	1898150	1138214	1137538	924429
Adjusted R^2	0.139	0.142	0.172	0.174	0.183	0.404	0.404	0.377
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3. Sample restricted to bottom 90% inventors, by lifetime patents. Standard errors are clustered by city × research field.

Instead of running a subsample regression, in Table 4 I use the full sample and interact cluster size with an indicator for being a top 10% inventor. Now the effect for the bottom-90% is 0 or positive, but the Column 8 coefficient for the top-10% is 2.5 times smaller than the coefficient from the original regression. Hence, whether or not the fixed effects are interacted with the Top 10% dummy makes a big difference for the main results. (Note that running separate regressions by subgroup is equivalent to interacting the fixed effects by the subgroup indicator.)

This poses a dilemma for Moretti: either the subsample regressions are correct, in which case the negative effect for the bottom-90% must be explained, or the interaction regression is correct, in which case the effect size for top inventors is much smaller.

Table 4: Main results: heterogeneity by Top 10%

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log size	-0.0099	-0.0269**	-0.0164	-0.0134**	-0.0258***	0.0369***	-0.0028	0.0088
	(0.0064)	(0.0114)	(0.0116)	(0.0062)	(0.0055)	(0.0074)	(0.0091)	(0.0107)
Log size × Top 10%	0.0264***	0.0270***	0.0305***	0.0299***	0.0293***	0.0199***	0.0193***	0.0213***
	(0.0068)	(0.0069)	(0.0070)	(0.0071)	(0.0072)	(0.0042)	(0.0041)	(0.0045)
Top 10%	0.6532***	0.6549***	0.6651***	0.6611***	0.6558***			
	(0.0339)	(0.0340)	(0.0344)	(0.0345)	(0.0349)			
Observations	2838921	2838918	2831024	2831024	2830716	2071999	2071647	1740204
Adjusted R^2	0.198	0.199	0.219	0.221	0.228	0.380	0.381	0.388
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3. Sample includes all inventors. *Top 10%* is collinear with the inventor fixed effect. Standard errors are clustered by city × research field.

6 Heterogeneity by movers and stayers

For the top 10% of inventors (measured by lifetime patents), 67% are movers, meaning they change (modal) research field or city at least once. The remaining 33% are stayers, who are observed only in the same cluster (city-field). Movers makes up 73% of all observations, indicating that movers patent more than stayers, since observations are patents. Stayers are 27% of all observations.

Moretti includes firm fixed effects in Table 3, Column 8. 10% of mover observations and 15% of stayer observations are missing firm identifiers.

6.1 Do inventors move to larger clusters?

To test whether movers move to larger clusters, I compute the cluster size in the years before and after an inventor moves. That is, I use the consecutive observations when an inventor is observed in different clusters; note that the corresponding years may not be consecutive, if there was a gap of more than one year. I find that the average size difference is very small and statistically insignificant. Hence, on average, movers are not moving to larger clusters.

For stayers, I test whether cluster size changes over the course of their time in the data (that is, between their first and last recorded patents). I find that, on average, cluster size decreases by about 3% of the sample mean cluster size. I find almost exactly the same decrease when defining cluster as cities (instead of city-field).

6.2 Heterogeneous treatment effects by movers and stayers

In Table 5 I interact cluster size with an indicator variable for having ever moved. Surprisingly, the effect is smaller for movers.

Table 5: Heterogeneity by movers and stayers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log size	0.0428***	0.0748***	0.0895***	0.0909***	0.0683***	0.1501***	0.1047***	0.1316***
	(0.0082)	(0.0164)	(0.0183)	(0.0091)	(0.0092)	(0.0189)	(0.0232)	(0.0279)
Log size × Mover	0.0005	0.0013	-0.0017	-0.0007	-0.0005	-0.0690***	-0.0571***	-0.0722***
	(0.0043)	(0.0042)	(0.0043)	(0.0042)	(0.0042)	(0.0182)	(0.0201)	(0.0247)
Mover	-0.1363***	-0.1342***	-0.1417***	-0.1373***	-0.1346***			
	(0.0197)	(0.0195)	(0.0187)	(0.0184)	(0.0185)			
Observations	931324	931291	923215	923215	922515	922279	921727	786303
Adjusted R^2	0.048	0.049	0.065	0.069	0.075	0.225	0.227	0.250
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3. *Mover* is an indicator variable for having ever moved across city or field; it is collinear with the inventor fixed effect. Standard errors are clustered by city × research field.

7 Mechanism: geographic proximity?

One natural hypothesis for the mechanism underlying the main results is geographic proximity: when close to other inventors, ideas can be passed around, leading to innovation. In this section, I compare different sources of cluster size: own-city vs other-city (within field).[6] If geographic proximity is the key mechanism, then patenting should be less correlated with the number of inventors in other cities (in the same field).

In Table 6, I regress patents on the number of inventors by city-field and the number of inventors in the same field in other cities. The coefficient on own-city is positive. However, the coefficient on other-city turns from positive to negative with the inclusion of field-year and city-year fixed effects. The effect is Column 8 is more than 15x the effect in Table 3 (and the opposite sign). This is an unexpected result. More inventors in the same field should increase innovation, even if they are not in the same city. (E.g., meeting people at conferences.) One possible explanation for the negative correlation is crowding-out: for a given innovation, more inventors in a field makes it less likely that any one inventor will patent it. (Although a negative coefficient means that the crowding-out effect dominates any agglomeration effect.)

Table 6: Horse race: own- vs other-city (within-field)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log size (city-field)	0.0546***	0.0899***	0.1017***	0.0791***	0.0564***	0.0824***	0.0284**	0.0416***
	(0.0082)	(0.0174)	(0.0192)	(0.0106)	(0.0099)	(0.0118)	(0.0130)	(0.0152)
Log size (other-city, field)	0.0963***	0.0643**	0.0586**	-0.3350*	-0.3108*	-0.3421*	-1.1090***	-1.1455***
	(0.0220)	(0.0257)	(0.0269)	(0.1751)	(0.1684)	(0.1975)	(0.2788)	(0.3367)
Observations	932527	932486	924367	924367	923667	922829	922263	786634
Adjusted R^2	0.043	0.045	0.061	0.064	0.070	0.225	0.227	0.250
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3. Standard errors are clustered by city-field. Independent variables are log counts, instead of log densities as in Table 3. The median number of inventors by own-city is 854, and the median number by other-city is 30152.

[6]Note that I define these variables as counts, instead of normalizing (e.g., by dividing by the national count by field). However, this is irrelevant when including the corresponding fixed effect, since log(a/b) = log(a)-log(b).

8 Mechanism: own-firm cluster size?

Here I compare cluster size from own-firm vs other-firm inventors (within city-field) If patenting is correlated only with the number of inventors in an inventor's own firm, then firm-level strategies like defensive patenting could be driving the results. But if patenting is correlated with the number of inventors at other firms, then we have evidence for spillover effects.

In Table 7, I regress patents on the number of inventors by own-firm and the number of inventors by other-firm, both within city-field. The effect for own-firm is positive, but smaller than in Table 3. As with other-city, the effect for other-firm turns from positive to negative, here with the inclusion of city-year fixed effects. This is again surprising, since inventors at other firms in the same city-field should have positive spillovers. Again, crowding-out is a possible explanation.

Table 7: Horse race: own- vs other-firm (within city-field)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log size (own-firm)	0.0065***	0.0071***	0.0075***	0.0046**	0.0029	0.0063***	0.0016	0.0235***
	(0.0018)	(0.0019)	(0.0020)	(0.0020)	(0.0019)	(0.0017)	(0.0015)	(0.0039)
Log size (other-firm)	0.0418***	0.0485***	0.0609***	0.0227*	0.0039	0.0138	-0.0461***	-0.0313***
	(0.0105)	(0.0165)	(0.0179)	(0.0121)	(0.0111)	(0.0096)	(0.0103)	(0.0105)
Observations	913393	913365	905395	905395	904681	903599	903066	768866
Adjusted R^2	0.042	0.044	0.060	0.063	0.070	0.224	0.226	0.249
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3. Standard errors are clustered by city-field. Independent variables are log counts, instead of log densities as in Table 3. The median number of inventors by own-firm is 15, and the median number by other-firm is 775.

9 Imputing missing observations

Moretti observes patents, and aggregates the patent data to the inventor-year level, by counting patents for inventors with (potentially) multiple patents in the same year. This generates missing observations during years when inventors do not patent, creating an imbalanced panel. These missing observations should have patent=0, but the inventor's city and field during that year is unknown.

I find that 37% of inventor-cluster spells have a gap (of more than one year) between consecutive observations (for example, an inventor patents in 2003 and 2005). The average gap is 3.83 years, and the largest gap is 34 years. For the sample of top 10% of inventors, filling in all missing observations would double the sample size. Decomposing by movers and stayers, 76% of missing observations are from movers and 24% are from stayers.

In Table A6, Moretti imputes missing observations for gaps of length 1 and 2, assigning patent= 0 and the cluster from the last observation before the gap. For stayers, this is the only choice (since cluster is unchanged before and after the gap). For movers, this form of imputation could influence the results. In particular, if movers move to larger clusters, then assigning the (smaller) origin cluster to imputed observations (with patent=0) could generate a positive correlation between cluster size and patenting. Conversely, assigning the destination cluster and patent=0 to imputed observations could generate a negative correlation.

I impute missing observations for gaps of all lengths, assigning either the origin or the destination cluster. This should provide upper and lower bounds on imputation bias. As mentioned above, I find that moving does not change cluster size, so imputation bias should be small. Note that I follow Moretti and calculate cluster size before imputing missing observations (fn 30); so cluster size does not count "inactive" inventors. Proper imputation would count these inventors.

The results are in Tables 8 and 9. The coefficients are slightly larger than in Table 3.[7] The effect in Column 9 is over 4x larger, controlling for firm-year fixed effects. The effects are similar by origin and destination, indicating that imputation bias is negligible. In Table A6, Moretti found that the effect size is increasing in the number of imputed missing observations. So it makes sense that imputing all missing observations would increase the effect even more.[8]

However, I'm not convinced by Moretti's explanation, which is that imputing missing 0s captures more of the extensive margin. For this sample of top inventors, the extensive margin

[7] In unreported results, I show that, on the original dataset (with $y > 0$), using log(1+patents) shrinks the coefficients by one-third compared to log(patents). So the use of log(1+patents) here does not explain the larger effects.

[8] Moretti made an error when imputing gaps of two years; his code only imputes one of the two years in the gap. So Table A6, Column 3 is the effect when interpolating gaps of length 1 and one out of two years for gaps of length 2.

is closer to "an established inventor joining an R&D project in year t" rather than "becoming an inventor". (For comparison, the intensive margin would be something like "filing a second patent from an R&D project, conditional on a first patent".) Perhaps formally modelling the intensive and extensive margins would be helpful. Moreover, most of the imputed observations are from movers, and the effect for movers is smaller, which contradicts the larger effect size observed here.

Table 8: Imputing by origin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log size	0.0491***	0.0909***	0.0899***	0.0913***	0.0807***	0.0827***	0.0970***	0.0980***	0.0766***
	(0.0031)	(0.0065)	(0.0072)	(0.0051)	(0.0046)	(0.0041)	(0.0041)	(0.0043)	(0.0043)
Observations	1855266	1855236	1850670	1850670	1850083	1849419	1849233	1683936	1411384
Adjusted R^2	0.051	0.053	0.071	0.073	0.082	0.229	0.231	0.264	0.303
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year				Yes	Yes	Yes	Yes	Yes	Yes
Class × year					Yes	Yes	Yes	Yes	Yes
Inventor						Yes	Yes	Yes	Yes
City × year							Yes	Yes	Yes
Firm								Yes	Yes
Firm × Year									Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3, plus Column 9. Dependent variable is log(1+patents). Missing observations are imputed by assigning the origin cluster and patent=0. Standard errors are clustered by city × research field.

In unreported results, I test for heterogeneity by movers and stayers on the imputed samples. I find that the negative interaction effect for Movers is slightly smaller compared to using the original sample. The results are similar for origin and destination imputation.

Table 9: Imputing by destination

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log size	0.0535***	0.1044***	0.1051***	0.1053***	0.0909***	0.0892***	0.0963***	0.0925***	0.0730***
	(0.0033)	(0.0065)	(0.0067)	(0.0040)	(0.0038)	(0.0035)	(0.0036)	(0.0037)	(0.0044)
Observations	1854447	1854421	1849928	1849928	1849321	1848656	1848451	1665864	1370730
Adjusted R^2	0.048	0.051	0.069	0.072	0.080	0.226	0.228	0.260	0.296
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × class			Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × year				Yes	Yes	Yes	Yes	Yes	Yes
Class × year					Yes	Yes	Yes	Yes	Yes
Inventor						Yes	Yes	Yes	Yes
City × year							Yes	Yes	Yes
Firm								Yes	Yes
Firm × Year									Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3, plus Column 9. Dependent variable is log(1+patents). Missing observations are imputed by assigning the destination cluster and patent=0. Standard errors are clustered by city × research field.

10 Aggregation bias

To construct the inventor-year panel, Moretti aggregates from patent data to inventor-year level data. If an inventor is in multiple clusters in one year, Moretti assigns them to their modal cluster. Patents from all clusters are assigned to the modal cluster. This aggregation could generate a mechanical bias, if modal clusters are larger. Using inventor-cluster-year data would eliminate this bias, since no patents are reassigned across clusters. Hence, if modal clusters are larger, we should expect smaller effects.[9]

I repeat the main analysis, but aggregate the patent data to the inventor-city-field-year level. This allows inventors to have observations in multiple clusters in the same year. While Moretti assigns the modal city, field, class, and firm to an inventor-year obsevation, here I assign only the modal class and firm to an inventor-city-field-year observation. (So when inventors patent at multiple firms/classes in the same cluster-year, I assign those patents to the modal firm/class.)

I find that modal clusters are about 10% larger than non-modal clusters. Hence, we should expect smaller effects using disaggregated data. And in Table 10 I find effects that are slightly smaller when aggregating to the inventor-cluster-year level.

It would be interesting to use the raw patent data, with inventors in different clusters in the same year, instead of aggregating to the inventor-year level. With finer data, we could include inventor-year fixed effects to make a within-inventor-year comparison (testing whether, for a given year, the same inventor patents more in larger clusters).

10.1 Combining disaggregation and imputation

Since imputing missing observations increases the effect size, and disaggregating reduces the effect size, we should expect that imputing missing observations at the inventor-cluster-year level would lead to effect sizes in between the two. However, this is computationally expensive, since it involves a Cartesian product of inventor, city, field, and year. For inventor-year, rectangularizing creates 48M observations; inventor-city-field-year would create 43B observations. There are more efficient ways to do this, e.g., filling in only clusters from an inventor's history, instead of all possible clusters.

[9]An alternative explanation is that large clusters cause patenting, so inventors who live in large clusters are more likely to patent there (making it their modal cluster), and also more likely to patent in other clusters (making them multi-cluster inventors). On this view, attributing patents to the modal cluster is the correct way of estimating the effect of cluster size on innovation.

Table 10: Aggregating to inventor-cluster-year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log size	0.0631***	0.0374***	0.0475***	0.0512***	0.0331***	0.0630***	0.0473***	0.0633***
	(0.0070)	(0.0130)	(0.0147)	(0.0081)	(0.0075)	(0.0076)	(0.0090)	(0.0100)
Observations	1145819	1145796	1137979	1137979	1137323	1136872	1136382	976873
Adjusted R^2	0.064	0.066	0.086	0.088	0.095	0.247	0.248	0.265
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City × field		Yes	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

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Replication of Table 3. Patent data is aggregated to inventor-cluster-year level. Standard errors are clustered by city × research field.

11 Varying the time unit

In Table A7, Moretti varies the time unit of the data, from 1-month to 3-year periods. As with imputation, the motivation is to address missing observations when inventors do not patent in a given time period. With a very short time unit (one second), most observations would be missing; and with a long time unit (one century), no observations would be missing. Moretti claims that these missing observations create downward bias, because of the missing extensive margin when 0s are not observed. This is consistent with the finding in Table A7, where the effect grows larger with the length of the time unit.

I am not persuaded by this argument. Technically, no zeros are added to the data, so no extensive margin is captured. Instead, the intensive margin is defined as a function of the time-unit. When a different time unit is used, the intensive margin changes. The pattern of effect sizes increasing with the length of the time unit could be explained simply by the variation in patents shrinking to nothing as the time unit shortens (since patent=1 for all observations, in the limit), and increasing as the time unit lengthens. (Although this model does not predict a negative correlation at the shorter time units.)

An alternative explanation for the negative results in short time units is that cluster size is defined using successful inventors in the current period. With shorter time units, this definition may be too restrictive, since inventors may be present in a cluster and contribute to agglomeration effects, but not patent (and hence have a missing row and not be counted in cluster size). That is, in reality, the inventor is present in the cluster, but they are not measured as such in the data. Hence, large clusters (where agglomeration effects create many patents) with a small measured size would generate a negative correlation between measured cluster size and patents. Instead, cluster size should count inactive inventors, or successful inventors from the previous T periods (for some T).

Furthermore, there is a coding issue with Table A7. Moretti does not recalculate cluster size at the level of the new time unit, but uses the baseline 1-year cluster size; he assigns the 1-year value to months within a year, and the first 1-year value in a j-year unit to that unit. (E.g., assign the value for 2000 to the 2-year unit covering 2000-2001.) Note that Table A7 Panel A uses the top 10% of inventors, while Panel B uses the top 1% (defined using lifetime patents).

I recalculate cluster size at the appropriate time unit in Table 11. While Moretti found negative effects for 1- and 2-month time units, I find 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 -0.088 (vs -0.025) and the 3-year effect is 0.20 (vs 0.17).

Table 11: Table A7: correct cluster size

	(1) 1-Month	(2) 2-Month	(3) 3-Month	(4) 6-Month	(5) 1-Year	(6) 2-Year	(7) 3-Year
Log size	-0.0881***	-0.0653***	-0.0388***	0.0092	0.0681***	0.1542***	0.2017***
	(0.0043)	(0.0062)	(0.0073)	(0.0094)	(0.0134)	(0.0178)	(0.0249)
Observations	1242054	1180955	1119547	974191	786303	573182	461584
Adjusted R^2	0.344	0.311	0.295	0.270	0.250	0.224	0.215

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Dependent variable is log(patents) per time unit. Patent data is aggregated to inventor-time level. In constrast to Moretti, I calculate cluster size and time fixed effects using the corresponding time unit. Standard errors are clustered by city $\times$ research field. The corresponding coefficients in Table A7 are -0.0248, -0.0120, 0.000149, 0.0297, 0.0676, 0.134, and 0.171.

12 Other issues

Moretti assigns multi-cluster inventors (who patent in different clusters in the same year) to their modal cluster, and claims to break ties randomly in multi-modal cases. However, he randomizes independently for city, field, class, etc., using `egen mode` separately by dimension. This can assign inventors to non-existent clusters. For example, for a cluster {city, field}, if an inventor is in clusters {A,1} and {B,2} in the same year, Moretti's independent randomization could assign them to {A,2} or {B,1}, even if those clusters never occur in the data. This problem probably biases his results towards zero, because it adds noise. It's also probably a small issue, because the number of ties is small.

Moretti breaks ties 'randomly' by using Stata's `egen mode, maxmode` function, which breaks ties by choosing the largest mode. This is not actually random.

Moretti drops inventors with more than three cities in the same year. This is not justified in the paper. By construction, these are highly productive inventors, since they patent in multiple cities per year.

Moretti implies that he first aggregates the patent data to inventor-year level data (by counting patents and assigning an inventor to their modal cluster), and then calculates cluster size by counting the number of inventors per cluster-year (as a share of all inventors by field-year). However, his code first calculates cluster size on the disaggregated data, and then aggregates the patent data to inventor-year level. This counts multi-cluster inventors as belonging to multiple clusters, and increases the size of non-modal clusters (relative to the method described in the text). The opposite approach would first assign inventors to their modal cluster, and then calculate cluster size; this increases the size of modal clusters.[10]

[10]Note that the aggregation bias discussed above is about patents per inventor: multi-cluster inventors have all of their patents assigned to the modal cluster. The issue here is about inventors per cluster.

13 Coding issues

- The event study in Figure 6 estimates β_0 using time-varying cluster size, instead of interacting with a time indicator for $t = 0$. (reg23.do)
- Figure 6 lists the coefficients in reverse order, from β_5 to β_{-5}. (reg23.do)
- When constructing the instrument used in Table 5, Moretti does not sort the data by city, leading to first-differences taken across cities (instead of within-city and over time). (iv_new.do)
- The code to impute missing observations for gaps of two years only imputes one of the two years. (data_3.do)
- When imputing missing observations, the code should recalculate cluster size. (data_3.do)
- For inventors with multiple modal clusters in the same year, Moretti claims to assign them randomly to one of the modes, but the code is not random. (data_3.do)
- The code to vary the time unit in Table A7 does not recalculate cluster size at the level of the new time unit. (data_4.do)
- When calculating cluster size excluding members of an inventor's team (co-authors on the same patent), the code subtracts total team size (including co-authors in other cities), instead of the number of co-authors in the same city. (density_team.do)

References

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

Wiebe, M. (2023). The effect of high-tech clusters on the productivity of top inventors: Comment. Working paper. `https://michaelwiebe.com/assets/moretti/moretti_comment.pdf`.