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

Michael Wiebe

Independent*

November 16, 2023

Abstract

Moretti (2021b) reports a positive correlation between technology cluster size and patenting, and uses an event study and instrumental variables regressions to justify a causal interpretation. I identify coding errors that overturn the results from these two supporting analyses. First, the event study does not interact the treatment variable with a time indicator in the year of the treatment. Second, the instrumental variable is constructed by incorrectly taking first-differences across cities. Correcting the code produces nonsignificant results for both the event study and instrumental variables regressions.
JEL: D62, J24, L60, O31, O34, R32

---
*maswiebe@gmail.com

1   Introduction

Moretti (2021b,a) studies agglomeration effects for innovation, testing whether the size of technology clusters causes patenting. Agglomeration effects are important for understanding technological progress (Kerr and Robert-Nicoud, 2020), and are affected by constraints on housing supply (Hsieh and Moretti, 2019; Duranton and Puga, 2020). Using US data on patents filed between 1971 and 2007 (Zucker and Darby, 2014), Moretti presents multiple lines of evidence supporting a causal effect of cluster size on patenting. The main results are from OLS regressions of patents on cluster size, controlling for an extensive set of fixed effects. Moretti addresses identification concerns using an event study and an instrumental variables strategy. The event study, using inventors who change cities, tests for selection bias from promising inventors sorting into large clusters. The instrument uses variation in the number of inventors in other cities to control for omitted variable bias from sources like local field-specific subsidies.

In this comment, I identify coding errors that overturn the results from the event study and the instrumental variables regressions. First, the event study does not interact the treatment variable with a time indicator in the year of the treatment. Including the correct interaction produces a nonsignificant effect in $t = 0$. Second, the instrument is constructed by incorrectly taking first-differences across cities, due to not sorting the data by city. When I calculate the first-difference within cities, the first stage and 2SLS estimates are both nonsignificant. Hence, the positive correlations from the OLS regressions could be biased upwards due to sorting and omitted variable bias.

2   Event study

In Figure 6, Moretti uses an event study to address worries about sorting, where 'rising star' inventors (with increasing patent counts) in small clusters are systematically hired by employers in large clusters.[1] If these promising inventors select into larger clusters, then the positive correlation between cluster size and patenting need not represent a causal effect of cluster size. Note that the main results control for inventor fixed effects, so the threat to identification comes only from inventors whose inventiveness is increasing over time. If sorting does occur, we should observe patenting rising in the years before an inventor moves to a new city.

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

---

[1]A cluster is defined as a (city, research field) pair; there are 179 cities and 5 fields. Cluster size is the number of inventors excluding the focal inventor, and inventors are counted if they file a patent in that year. The main analysis focuses on the top 10% of inventors by lifetime patents.

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.[2] The implied regression equation is

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

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 event study results are driven entirely by one coding error. Moretti did not generate an interaction of post-move cluster size with a $t = 0$ indicator, but instead uses the time-varying cluster size variable used in the OLS regressions. Hence, β_0 is estimated using data from all event-years, instead of capturing the effect in $t = 0$. I correct the code by interacting post-move cluster size with the $t = 0$ indicator.[3]

Figure 1 shows the original and corrected event studies.[4] The large effect in $t = 0$ disappears when using the correct specification. Hence, Moretti does not provide evidence against bias in the OLS results from sorting, where promising inventors select into large clusters. Moreover, if agglomeration effects are real, we should detect them using a mover event study. So the absence of an effect in the corrected event study should lead us to discount the main findings.[5] However, since there are roughly 118,000 inventors in the OLS sample and only 3,000 in the event study sample, the size of this discount should be small. Finally, since inventors move in different years, note that this is a staggered adoption design where a two-way fixed-effects estimator may be biased.[6]

---

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

[3]Moretti claims that the event study includes only "inventors who are in the data for 11 consecutive years" (p.3354). However, the code makes no such restriction. The estimation sample includes inventors with an observation count between 2 and 24, and allows time gaps of any length between observations.

[4] My replication does not exactly match the original Figure 6, because Moretti's cleaning code is unreproducible. It uses many-to-many merges with a nonunique sort order, resulting in a different sample for each run of the code.

[5]In Wiebe (2023) I recode the event study to follow standard practices, and find no effect. I use a constant treatment variable (difference in average cluster size before and after the move) 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; and restrict the sample to event years [-5,5].

[6]Since the treatment is continuous, estimators like Sun and Abraham (2021) and Callaway and Sant'Anna (2021) that are defined for binary treatments are not applicable.
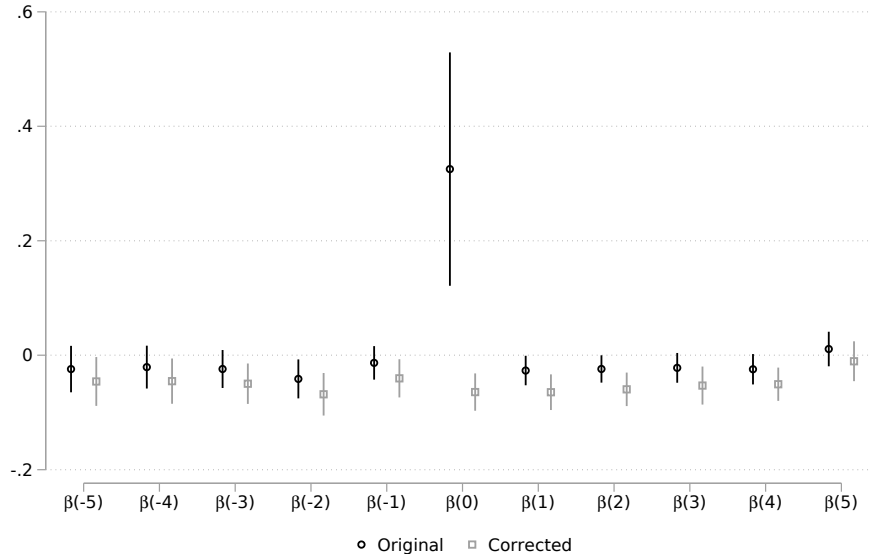
3

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 $Size_{-ifc}^{post} \times \mathbb{1}\{t = 0\}$; following Moretti, $Size_{-ifc}^{post}$ is calculated excluding $t = 0$. N=18,433 in *Original*, N=18,434 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.

3   Instrumental variables estimates

In Table 5, Moretti uses an instrumental variables strategy to address worries about omitted variable bias, where cluster size is correlated with unobserved time-varying shocks at the city-field level. For example, a city subsidizing biotech firms could increase both biotech patents and the size of the biotech cluster.[7] The idea for the instrument is to use variation in the number of inventors in other cities who are employed by firms that are also active in the focal inventor's city.

To illustrate with an example, for the focal inventor i in the field of computer science at Google in San Francisco, we instrument for cluster size using variation in the number of computer science inventors at Microsoft in Seattle, where Microsoft also has inventors in San Francisco. Intuitively, the number of inventors working at Microsoft in Seattle will be correlated with the number of computer science inventors in San Francisco, since Microsoft also has a presence in San Francisco, allowing shocks to Microsoft to propagate to San Francisco. The exclusion restriction is that, controlling for fixed effects, the number of computer science inventors at Microsoft in Seattle is not directly correlated with computer science patents at Google in San Francisco. While this example uses two firms, the

---

[7]Recall that cluster size for a focal inventor is measured as the number of patents filed that year by other inventors in the same research field and city.

instrument uses all firms active in the focal inventor's city, excluding the inventor's own firm.

In contrast to the OLS results, Moretti defines the instrument in first-differences. Let $\Delta N_{jf(-c)t} = N_{jf(-c)t} - N_{jf(-c)(t-1)}$ be the first-difference over time in the number of inventors at firm j in research field f in all cities excluding c, and let ΔN_{ft} be the national change by field. With $D_{jfc(t-1)}$ as an indicator for firm j employing at least one inventor in city c and field f in year $t-1$ (so that the first-difference is well-defined), the IV is

$$IV_{jfct} = \sum_{s \neq j} D_{sfc(t-1)} \frac{\Delta N_{sf(-c)t}}{\Delta N_{ft}}, \tag{2}$$

where the sum is taken across all firms excluding j.

The regression equation is

$$\Delta \ln y_{ijfkct} = \alpha \Delta \ln S_{-ifct} + d_t + d_f + d_k + d_j + d_{ft} + d_{kt} + \varepsilon_{ijfkct}, \tag{3}$$

where y is the number of patents, S is cluster size (excluding the focal inventor i), and k is research class.

There are two coding errors that affect the IV results. First, when calculating the first-differenced term $\Delta N_{jf(-c)t}$, Moretti sorts the data by firm, field, and year, but not by city. This means that the first-difference is taken across different cities, instead of within the focal city. Hence, the code generates an incorrect instrument that does not match the definition in the text. I correct the code to sort by city and take first-differences within city.

Moreover, the code generates unreproducible results. This is because firm-field-year is not a unique sorting order, as there are observations with the same values for firm-field-year but in different cities. Since Stata's `sort` command randomly orders the data before sorting, each run of the code results in a different sorting by city within tied values of firm-field-year. Hence, the first-difference is calculated differently each time. In Figure 2, I show the variance in the original results from 500 iterations of generating the instrument and running the same regression.

The second error is that Moretti's code incorrectly defines the instrument for all observations, including observations for which $\Delta N_{sf(-c)t}$ is missing (because the first-difference is taken across firms, cities, or non-consecutive years). Specifically, Moretti calculates $\sum_{s \neq j} D_{sfc(t-1)} \frac{\Delta N_{sf(-c)t}}{\Delta N_{ft}}$ by first calculating the sum over all firms (by field-city-year), and then subtracting the value for the focal firm j. However, the code assigns the field-city-year total to all firms, and subtracts the summand for firm j *only* if that summand is nonmissing. Hence, firms with a undefined first-difference are assigned the field-city-year total when the instrument should be missing. I correct the code to exclude firms with an undefined first-difference from the estimation sample.

Figure 2: Distribution of coefficients and t-statistics from 2SLS regression
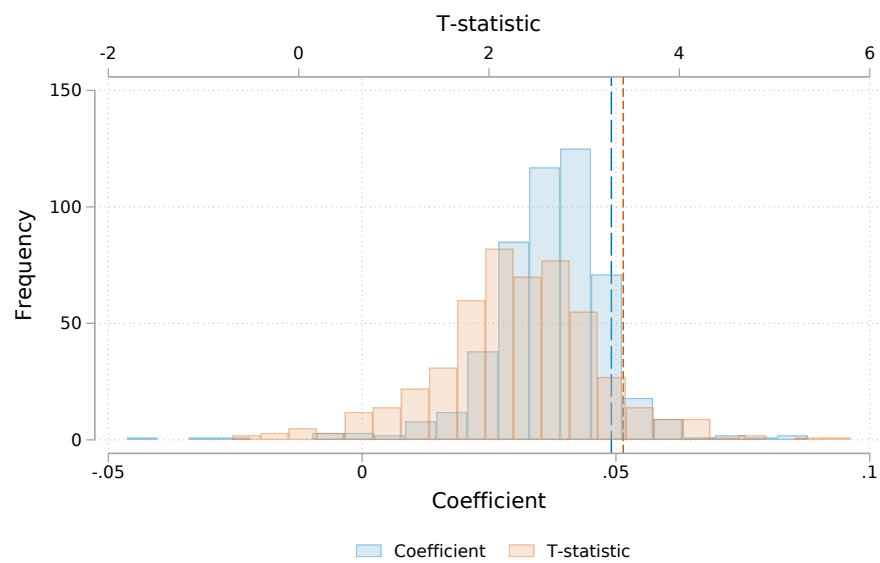
Note: Estimation results from the 2SLS regression in Table 5, Column 6. The distribution is constructed from 500 iterations of generating the instrument and running the regression. The original code sorts by firm-field-year but not city, which leads to a random ordering by city within tied values of firm-field-year. The first-difference in the numerator of the instrument is then (incorrectly) taken across cities, which causes the instrument to be different for each run of the code. The original estimate is 0.0491 (blue long-dashed line) with a t-statistic of 3.41 (orange short-dashed line).

The original results are reproduced in Panels A and B of Table 1, and the results using the corrected instrument are in Panels C and D. Since Moretti's results are not reproducible, my reproductions are similar but not identical to the original estimates in Table 5.[8] The results using the corrected instrument are markedly different, however. The 2SLS estimates are now negative and nonsignificant, likely due to the first stage now being close to zero with an F-statistic around 1.[9] Without a first stage, the IV strategy does not work. This is not evidence against agglomeration effects, but it does mean that Moretti fails to provide evidence against confounding from city-field-year shocks, such as local field-specific subsidies that affect both patenting and cluster size.

4   Conclusion

I reanalyze Moretti (2021b) using the original data and code, and find that the results from the event study and instrumental variables regressions are caused by coding errors. The event study does not interact the treatment variable with a time indicator for the year of the treatment, and the instrument is calculated incorrectly by taking first-differences across cities. While the OLS results for agglomeration effects remain valid, they are vulnerable to confounding from sorting and omitted variable bias.

---

[8]As noted in Footnote 4, Moretti's cleaning code is unreproducible and results in different sample sizes each time. Hence, the sample sizes in my reproduced regressions differ slightly from those in Table 5.

[9]Defining the instrument in levels instead of first-differences produces a significant first stage, but null 2SLS estimates; see Wiebe (2023).

Table 1: Replication and correction of Table 5 IV estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 2SLS (original)						
Δ log size	0.0459	0.0646	0.0504	0.0504	0.0427	0.0420
	(0.0225)	(0.0266)	(0.0253)	(0.0186)	(0.0178)	(0.0162)
Observations	419785	419785	419754	405323	405323	404164
Panel B: First stage (original)						
IV	0.8800	0.8363	0.8518	1.1232	1.1686	1.1804
	(0.1514)	(0.1558)	(0.1539)	(0.1932)	(0.1731)	(0.1709)
Observations	419785	419785	419754	405323	405323	404164
F-statistic	33.79	28.80	30.64	33.80	45.57	47.69
Panel C: 2SLS (corrected)						
Δ log size	-0.5724	-0.4963	-0.4118	-0.8929	-0.4739	-0.0734
	(1.0400)	(0.8266)	(0.6446)	(1.6552)	(1.1969)	(0.9759)
Observations	361401	361401	361370	353888	353887	352649
Panel D: First stage (corrected)						
IV	0.0095	0.0110	0.0133	0.0079	0.0089	0.0100
	(0.0087)	(0.0116)	(0.0114)	(0.0096)	(0.0103)	(0.0107)
Observations	361401	361401	361370	353888	353887	352649
F-statistic	1.18	0.91	1.36	0.68	0.75	0.87
Year	Yes	Yes	Yes	Yes	Yes	Yes
Field		Yes	Yes	Yes	Yes	Yes
Class			Yes	Yes	Yes	Yes
Firm				Yes	Yes	Yes
Field $\times$ year					Yes	Yes
Class $\times$ year						Yes

Standard errors in parentheses

Note: Dependent variable is the change in the log number of patents in a year (2SLS) and the change in log cluster size (first stage). Panels A and B calculate the instrument by incorrectly taking first-differences across cities, while Panels C and D take first-differences within cities. The sample size in Panels A and B is larger due to incorrectly defining the instrument for observations with a missing first-difference; these observations are excluded in Panels C and D. Standard errors are clustered by city. The original code generates unreproducible results, so the estimates in Panels A and B should be interpreted as one realization of a random process.

References

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Duranton, G. and D. Puga (2020). The economics of urban density. *Journal of Economic Perspectives 34*(3), 3–26.

Hsieh, C.-T. and E. Moretti (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.

Kerr, W. R. and F. Robert-Nicoud (2020). Tech clusters. *Journal of Economic Perspectives 34*(3), 50–76.

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.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Wiebe, M. (2023). Extended commentary on Moretti (2021). Working paper. `https://michaelwiebe.com/assets/moretti/moretti_extended.pdf`.

Zucker, L. G. and M. R. Darby (2014). Connecting outcome measures in entrepreneurship technology and science (COMETS) database, public-use version comets 2.0. `https://kauffman.org/entrepreneurship/research/comets/`.