**ONLINE APPENDIX**

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

Enrico Moretti

Appendix Table 1—Largest Patent Assignees by Total Number of Patents 1971-2007

| | |
|---|---|
| INT BUSINESS MACHINES | 155790 |
| GEN ELECT | 69051 |
| MICROSOFT | 43556 |
| INTEL | 42085 |
| EASTMAN KODAK | 41538 |
| MOTOROLA | 40995 |
| XEROX | 35034 |
| MICRON TECHNOLOGY | 31999 |
| TEXAS INSTRUMENTS | 27871 |
| E I DU PONT DE NEMOURS | 25250 |
| HEWLETT-PACKARD DEV LP | 25030 |
| AT&T | 24903 |
| ADV MICRO DEVICES | 21253 |
| DOW CHEM | 19879 |
| GEN MOTORS | 19763 |
| US OF AMER AS REPRESENTED BY SECRETARY OF NAVY | 19680 |
| LUCENT TECHNOLOGIES | 18862 |
| EXXON RES ENGN | 18571 |
| HONEYWELL INT | 18180 |
| PROCTER GAMBLE | 17751 |
| REGENTS OF UNIV OF CALIFORNIA | 16749 |
| APPL MATERIALS | 16449 |
| HEWLETT-PACKARD | 15979 |
| SUN MICROSYSTEMS | 15362 |
| BOEING | 14705 |

Appendix Table 2—Field-Specific Elasticities

| | Comp. Science<br>(1) | Biol. and Chem.<br>(2) | Semiconductors<br>(3) | Other Engineer.<br>(4) | Other Science<br>(5) |
|---|---|---|---|---|---|
| Log Size | 0.187 | 0.145 | 0.262 | 0.104 | 0.0768 |
| | (0.0403) | (0.0208) | (0.0626) | (0.0222) | (0.0299) |
| $N$ | 77208 | 198905 | 38230 | 428029 | 81003 |

Notes: Each column is a separate regression. The level of observation in the regressions is inventor-year. The dependent variable is log of number of patents filed in a year. Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, and Firm effects. City $\times$ Year are not included. Standard errors are clustered by city $\times$ research field.

Appendix Table 3—Some Examples of Firm-Specific Productivity Spillovers – 2007

| Firm | City | Estimated Productivity Spillover |
|---|---|---|
| (A) Computer Science | | |
| MICROSOFT | Seattle-Tacoma-Olympia, WA | .0806 |
| INT BUSINESS MACHINES | Minneapolis-St. Paul-St. Cloud, MN-WI | .0415 |
| CISCO TECHNOLOGY | San Jose-San Francisco-Oakland, CA | .0020 |
| DELL PROD | Austin-Round Rock, TX | .0061 |
| TEXAS INSTRUMENTS | Dallas-Fort Worth, TX | .0184 |
| CATERPILLAR | Peoria-Canton, IL | .1047 |
| MOTOROLA | Chicago-Naperville-Michigan City, IL-IN-WI | .0088 |
| HEWLETT-PACKARD | San Jose-San Francisco-Oakland, CA | .0007 |
| **Average Firm** | **Average City** | .0032 |
| (B) Biology and Chemistry | | |
| E I DU PONT DE NEMOURS | Philadelphia-Camden-Vineland, PA-NJ-DE-MD | .0117 |
| BRISTOL-MYERS SQUIBB | New York-Newark-Bridgeport, NY-NJ-CT-PA | .0054 |
| PROCTER GAMBLE | Cincinnati-Middletown-Wilmington, OH-KY-IN | .0347 |
| AMGEN | Los Angeles-Long Beach-Riverside, CA | .0065 |
| CHEVRON RES | San Jose-San Francisco-Oakland, CA | .0014 |
| 3M | Duluth, MN-WI | .0899 |
| PFIZER | Hartford-West Hartford-Willimantic, CT | .0124 |
| EXXON RES ENGN | Washington-Baltimore-Northern Virginia, DC-MD-VA-WV | .0043 |
| **Average Firm** | **Average City** | .0024 |

Notes: Entries reflect the impact that a specific firm is estimated to have on the productivity of scientists in other firms in the same cluster in 2007. For a given firm $j$, field $f$ and city $c$, entries are obtained as $\hat{\alpha}\Delta S_{-jfct}$ where $\hat{\alpha} = 0.066$ is the estimated elasticity in Table 3, column 8 and $\Delta \ln S_{-jfct}$ is the difference in log cluster size with and without a given firm. In particular, $\Delta \ln S_{-jfct} = [\ln(N_{fct}/N_{ft}) - \ln(N_{jfct}/N_{ft})]$, where $N_{fct}$ in the number of scientists in cluster $fct$; $N_{ft}$ is number of scientists in field $f$ and year $t$; $N_{jfct}$ is number of scientists in firm $j$ in cluster $fct$; and $t = 2007$.

Appendix Table 4—Models in Differences: Effect of Changes in Cluster Size on Changes in Inventor Productivity – OLS and IV Estimates — Single Location Firms

| | OLS<br>(1) | OLS<br>(2) | OLS<br>(3) | 2SLS<br>(4) | 2SLS<br>(5) | 2SLS<br>(6) |
|---|---|---|---|---|---|---|
| $\Delta$ log Size | 0.000657<br>(0.00662) | 0.00105<br>(0.00659) | 0.00157<br>(0.00663) | 0.0828<br>(0.0364) | 0.0778<br>(0.0332) | 0.0529<br>(0.0313) |
| First Stage | | | | 1.264<br>(0.191) | 1.289<br>(0.195) | 1.342<br>(0.180) |
| F stat. | | | | 43.9 | 43.7 | 55.8 |
| $N$ | 53627 | 53624 | 51196 | 53627 | 53624 | 51196 |
| Year | y | y | y | y | y | y |
| Field | y | y | y | y | y | y |
| Class | y | y | y | y | y | y |
| Field × Year | | y | y | | y | y |
| Class × Year | | | y | | | y |

Notes: Each entry is a separate regression. Dependent variable is the change in the log number of patents in a year. The model estimated is equation 3. The instrumental variable for workers in firm $j$ in cluster $fct$ is defined as $IV_{jfct} = \sum_{s\neq j} D_{sfc(t-1)} \frac{\Delta N_{sf(-c)t}}{\Delta N_{ft}}$ where $D_{jcf(t-1)}$ is an indicator equal to 1 if firm $j$ has at least 1 inventor in city $c$ in field $f$ in year $t-1$. $N_{jf(-c)t}$ is the number of inventors that firm $j$ has in field $f$, year $t$ in all the cities excluding city $c$; and $\Delta N_{jf(-c)t} = N_{jf(-c)t} - N_{jf(-c)(t-1)}$ is the change in $N_{jf(-c)t}$ between time $(t-1)$ and $t$. Standard errors are clustered by city.

Appendix Table 5—Cross-Field Spillovers

| | All<br>(1) | Biology and Chem.<br>(2) | Computer Sc.<br>(3) | Other Eng.<br>(4) | Other Sci.<br>(5) | Semicond.<br>(6) |
|---|---|---|---|---|---|---|
| Own Field | 0.0719 | | | | | |
| | (0.0207) | | | | | |
| Mean of Other Fields | 0.0126 | | | | | |
| | (0.0456) | | | | | |
| Biology and Chem. | | 0.173 | -0.0602 | -0.0145 | 0.0268 | -0.0670 |
| | | (0.0395) | (0.0576) | (0.0210) | (0.0478) | (0.0804) |
| Computer Science | | 0.0238 | 0.224 | 0.0119 | 0.00122 | 0.138 |
| | | (0.0141) | (0.0417) | (0.00905) | (0.0276) | (0.0650) |
| Other Eng. | | -0.00385 | 0.0349 | 0.123 | 0.0165 | -0.185 |
| | | (0.0376) | (0.105) | (0.0342) | (0.0710) | (0.206) |
| Other Sci. | | -0.0448 | 0.00740 | -0.0135 | 0.0806 | 0.196 |
| | | (0.0238) | (0.0682) | (0.0145) | (0.0498) | (0.0585) |
| Semicond. | | 0.00754 | -0.0376 | 0.00813 | 0.00220 | 0.218 |
| | | (0.00677) | (0.0319) | (0.00543) | (0.0167) | (0.0644) |
| $N$ | 822320 | 37961 | 174966 | 74825 | 367397 | 74111 |

Notes: Each column is a regression. Column 1 includes all inventors. The sample in columns 2 to 6 includes inventors in the field specified at the top. Entries in a given row show the coefficient on the field-specific cluster size. For example, the entry in row 3, column 3 show the effect of the size of the Biology and Chemistry cluster on the productivity of Computer Scientists. Models include Year, City, Field, Class, City × Field, City × Class, Field × Year, Class × Year, Inventor, and Firm effects.

Appendix Table 6—Interpolation

| | | Interpolation | |
|---|---|---|---|
| | Baseline Sample | 1 Year | 2 Years |
| | (1) | (2) | (3) |
| (A): Inverse Hyperbolic Sine | | | |
| | 0.0543 | 0.0664 | 0.0685 |
| | (0.00860) | (0.00822) | (0.00813) |
| (B): log(patents+1) | | | |
| | 0.0409 | 0.0507 | 0.0525 |
| | (0.00647) | (0.00619) | (0.00612) |
| $N$ | 823375 | 860806 | 873346 |

Models include Year, City, Field, Class, City × Field, City × Class, Field × Year, Class × Year, Inventor, City × Year and Firm effects. Standard errors are clustered by city × research field.

Appendix Table 7—Alternative Units of Time

| | 1 Month | 2 Months | 3 Months | 6 Months | 1 Year (Baseline) | 2 Years | 3 Years |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| (A) All | | | | | | | |
| Log Size | -0.0248 | -0.0120 | 0.000149 | 0.0297 | 0.0676 | 0.134 | 0.171 |
| | (0.00841) | (0.00927) | (0.00970) | (0.0106) | (0.0139) | (0.0148) | (0.0193) |
| $N$ | 1321719 | 1234635 | 1165648 | 1013458 | 823375 | 610136 | 500822 |
| (B) Top 1% Inventors | | | | | | | |
| Log Size | 0.0176 | 0.0456 | 0.0640 | 0.134 | 0.249 | 0.341 | 0.399 |
| | (0.0165) | (0.0204) | (0.0224) | (0.0272) | (0.0405) | (0.0513) | (0.0675) |
| $N$ | 345063 | 304488 | 274684 | 216476 | 155240 | 100759 | 77551 |

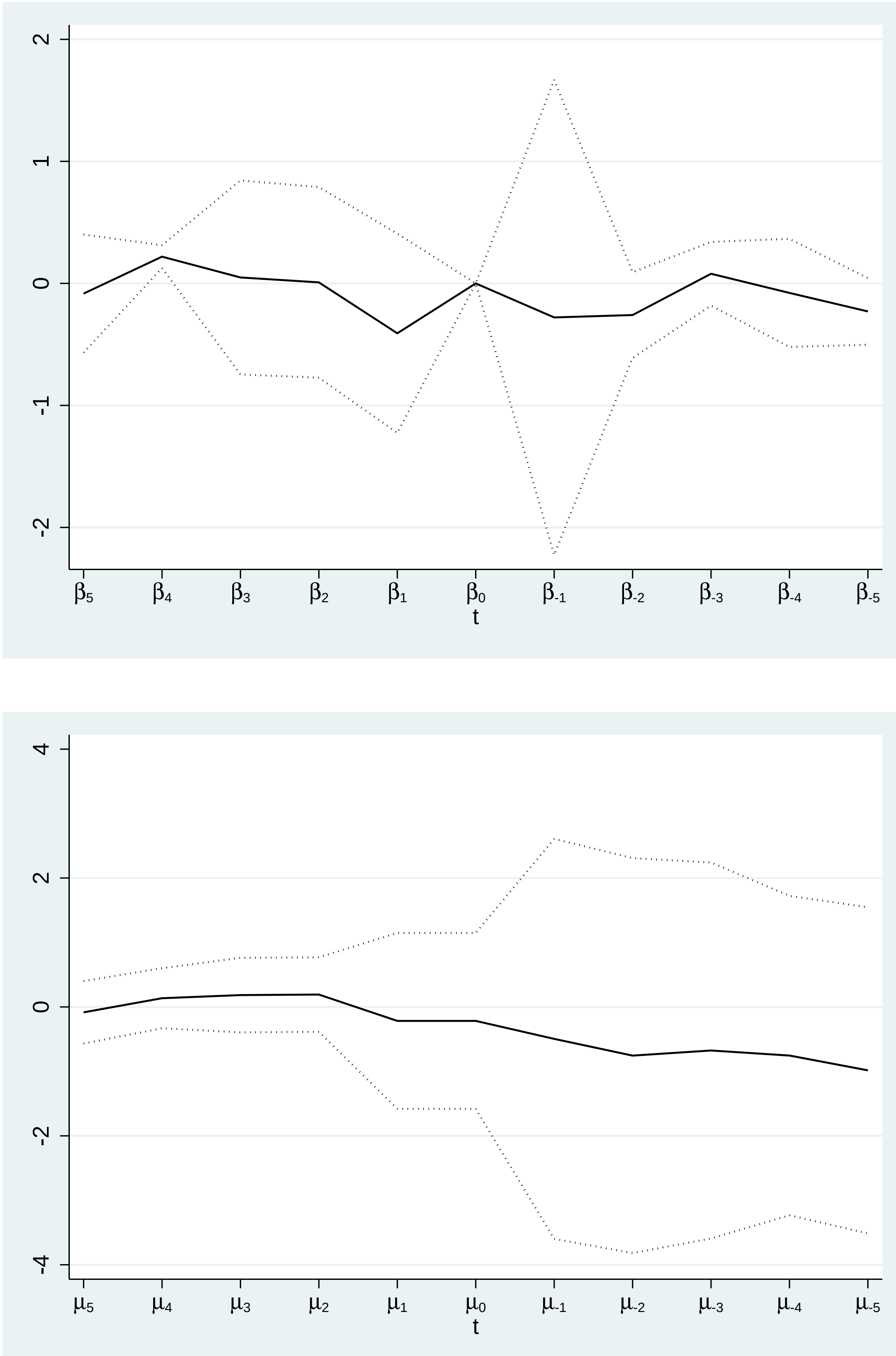
Notes: Dependent variable is log number of patents in unit of time. Baseline in Panel A is the entry in Col 8 of Table 3. Models include Year, City, Field, Class, City × Field, City × Class, Field × Year, Class × Year, Inventor, City × Year and Firm effects. Standard errors are clustered by city × research field.

Appendix Table 8—Cluster Quality and Teams

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| log Size | 0.124 | 0.133 | 0.115 | 0.134 | 0.117 | 0.237 | 0.229 | 0.0886 |
| | (0.0461) | (0.0552) | (0.0372) | (0.0570) | (0.0113) | (0.0446) | (0.0452) | (0.0138) |
| log Size $X$ 1(Team size $\geq$ Median) | | | | | | | 0.00546 | |
| | | | | | | | (0.00370) | |
| log Size $X$ (Solo Inventor) | | | | | | | | -0.0194 |
| | | | | | | | | (0.00404) |
| $N$ | 823536 | 823199 | 823505 | 823434 | 823375 | 822892 | 822747 | 823375 |

Notes: Each entry is a separate regression. Dependent variable is log number of patents in a year. In column 1, 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. In column 2, cluster size is measured as the number of inventors with a lifetime patent count above 3. In column 3, cluster size is measured as the weighted sum of inventors in a given city-field-year cell, with weights reflecting the lifetime number of patent citations. In column 4, cluster size is measured as the number of inventors with a lifetime patent citation count above 5. Column 5 controls for a quadratic in team size. In column 6, cluster size is defined excluding all members of the focal inventor's team. In column 7, cluster size is interacted with a dummy equal to 1 if the focal inventor team has size above median. In column 8, cluster size is interacted with a dummy equal to 1 if the focal inventor share of patents where he is the only inventor is above .9. Models include Year, City, Field, Class, City $\times$ Field, City $\times$ Class, Field $\times$ Year, Class $\times$ Year, Inventor, City $\times$ Year and Firm effects. Standard errors are clustered by city $\times$ research field.

Appendix Figure 1. Dynamic Response Following a Change in Cluster Size – Pharmaceutical Only




Notes: This figure is based on equation 2 in the text. In the top panel I plot the estimated $\beta$ coefficients in equation 2 on the lag and lead terms. For example, $\beta_5$ is the coefficient on the fifth lead term. In the bottom panel, I plot the cumulative response, where the $\mu$'s are defined as: $\mu_n = \beta_5 + \beta_4 + ... + \beta_n$ for $n = -5$ through 5. Standard errors are clustered by city $\times$ research field.