Internet Appendix to "Proximity and the Management of Innovation"

1. Additional tests

1.1. Robustness tests

In this section, we summarize the results of several tests designed to explore potential explanations for our results. A potential concern with the results documented in Table 3 is that they may reflect bargaining between airlines and firms. For example, a firm investing significantly in a research facility may arrange for an airline to provide a direct route from that facility to corporate headquarters. To the extent these investments increase the productivity of local inventors employed by the firm, and these increases are not shared by other inventors living in that location, or by other inventors employed by the firm living elsewhere, this could bias our results. To address this potential concern, we re-estimate Eq. (2) after removing observations where the shortest flight route between the headquarters and the inventor or facility changed from a connecting route to a direct route, or vice versa (i.e., we exclude significant travel time decreases driven by changes in direct flight routes). We report results in Table IA.3. We continue to find that proximity increases innovation production, suggesting our results are not driven by location-specific bargaining between airlines and firms.

Finally, we examine whether our inferences are robust to an alternative functional form assumption. We estimate Eq. (3) after no longer taking the inverse hyperbolic sine transformation of our dependent variables. We find similar results, suggesting our results are not driven by functional form assumptions regarding our dependent variable.[1]

1.2. Proximity and innovation production: Firm-level results

We next estimate a firm-level analysis of how significant travel time decreases relate to innovation production, which, along with our inventor and facility-level analyses, answers the calls of Adler and Chen (2011) and Anderson et al. (2014) for multi-level studies of innovation. We aggregate treatments from the facility-level to the firm-level. We then examine how firm-level innovation outcomes relate to aggregate

[1] All 8 of the estimates on *Significant Flight Time Decrease* are statistically significant and positive (*t*-statistics of 2.24 to 4.31).

treatments, controlling for the number of facilities in the firm. We require firm-years to have at least one facility to appear in the sample for this analysis. We replace our facility state-year fixed effects with headquarters state-year fixed effects, because we conduct this analysis at the firm-level. For the same reason, we replace our firm-year fixed effects with Fama-French 48 industry-year fixed effects.

Because we no longer control for time-varying firm attributes with firm-year fixed effects, we include as controls several changes in firm-level characteristics that we expect may affect innovation production. In particular, we control for the firm's size, or *Market Value*, equity performance, *Equity Return*, monthly volatility of returns, *sigma(Equity Return)*, total R&D spending, *R&D Spending*, accounting performance, *Return on Assets*, whether it has a loss, *Loss Indicator*, an indicator for whether data on R&D spending was missing and replaced with zeroes following the advice in Koh and Reeb (2015), *Missing R&D Indicator*, special items divided by assets, *Special Items*, and the book value of assets to the equity value of assets, *Market to Book Ratio*. We lag controls to avoid bad control problems as suggested by Angrist and Pischke (2008).

Table IA.2 reports results, which suggest that plausibly exogenous decreases in travel time from headquarters to research facilities increases innovation production at the firm level (e.g., three of four coefficients on facility treatments are positive and statistically significant). Together with our earlier individual inventor-level and R&D facility-level results, these firm-level results suggest that proximity to headquarters increases innovation production at all levels.

1.3. Alternative methods of identifying R&D facilities

In this section, we summarize the results of several other potential approaches for identifying facilities. First, in un-tabulated results, we re-estimate our Table 2 and 3 results after separately requiring at least 3, 4, 5, 6… up to 20 unique lead inventors at the facility. Doing so progressively attenuates our non-singleton sample (e.g., when requiring 20 unique inventors to identify a facility the non-singleton sample is about 1/3rd the size of our main sample). We find that none of the 72 estimates that correspond to our Table 2 estimates are statistically insignificant. We do find that 11 of the 72 estimates that correspond to

our Table 3 estimates are statistically insignificant, but do not see a clear or meaningful pattern in when the estimates are statistically insignificant that suggests issues with the aggregation.[2]

Second, we only retain facilities with inventors filing patents over different time horizons. We re-estimate our Table 2 and 3 results after counting the number of unique inventors in the 1, 2, 3, 4, … up to 10 years pre and post year *t*. This progressively truncates the sample, as the patent database ends in 2010. We stop at 10 years as this requires us to end the sample in 2000, for a sample period of 1996-2000. The coefficients of interest are statistically significant in all 80 cases (un-tabulated). The coefficient estimates become slightly larger in magnitude in the most stringent specifications.[3] Third, we use counties, rather than MSAs, to identify facility locations. Counties often cover a smaller area than MSAs, and are more likely to correspond to small clusters of zip codes that contain inventors. The coefficients of interest are statistically significant in all eight specifications.[4] In total, we conclude that our inferences are robust to several alternative methods of identifying R&D facilities, and hope that our method of identifying R&D facility location is useful for future research interested in how the R&D function is internally organized.

[2] We use two or more inventors to identify a facility in our main tests as it best conforms to our conceptualization of a facility as a grouping of inventors likely to be overseen by a local manager, and because our comparison to the Png (2019) dataset suggests that, if anything, this approach is overly conservative (i.e., in the 8.2% of cases where we fail to identify a facility it is because only one lead inventor lived in the MSA).

[3] We measure facilities over three year in our main tests as this corresponds to the adjustment for truncation bias in citation patterns suggested by Hall et al. (2001), and to the average duration between changes in R&D spending and changes in patent filings (Glaeser, 2018).

[4] Absolute value of the *t*-statistics of 2.57 to 10.23. We use MSAs to identify facilities in our main tests because we believe it is possible for inventors live in one county and work in another, but much less likely that inventors live in one MSA and work in another as MSAs.

REFERENCES

Adler, P. S., & Chen, C. X. (2011). Combining creativity and control: Understanding individual motivation in large-scale collaborative creativity. *Accounting, organizations and society*, *36*(2), 63-85.

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of management*, *40*(5), 1297-1333.

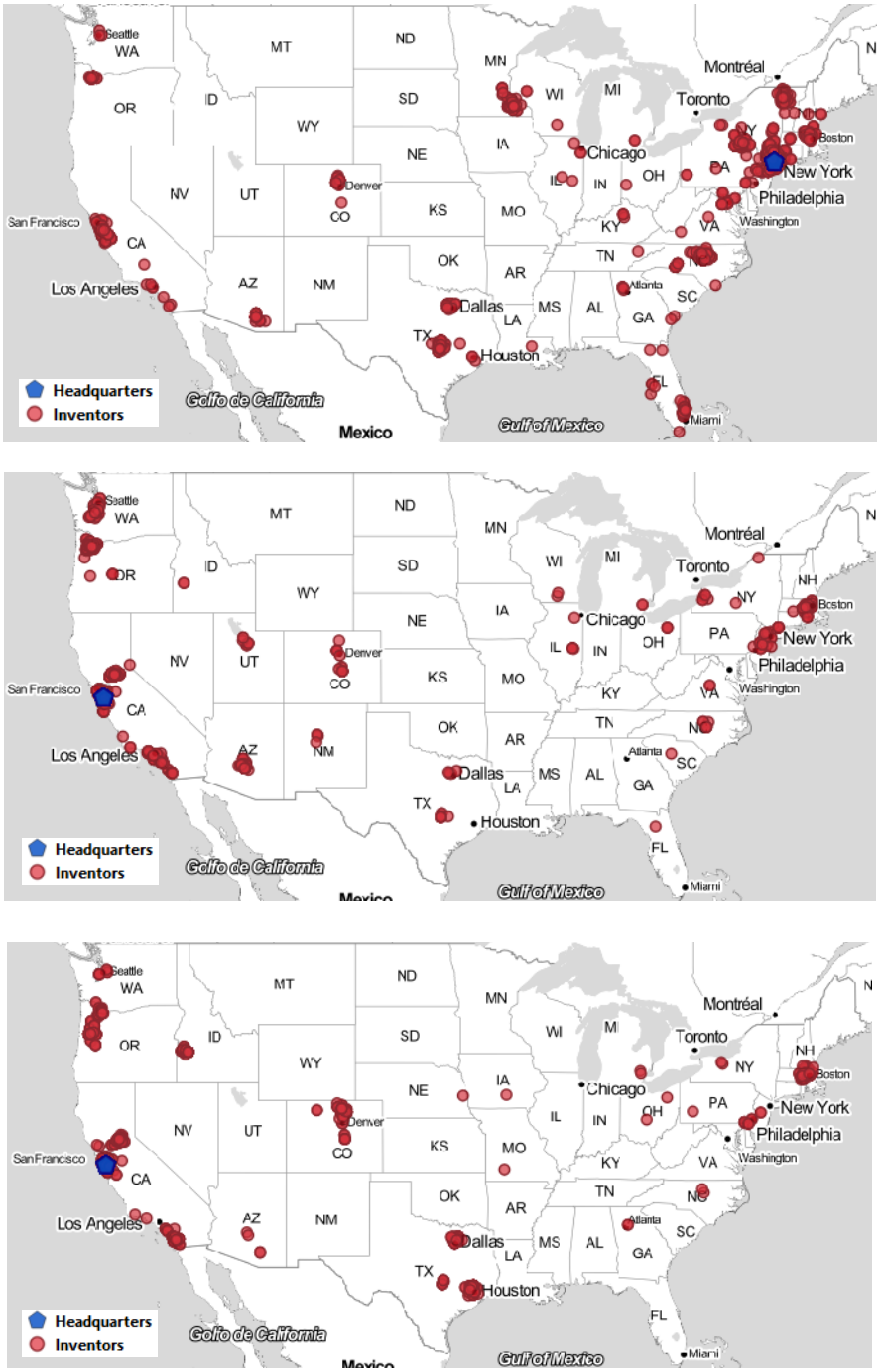
Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton university press.

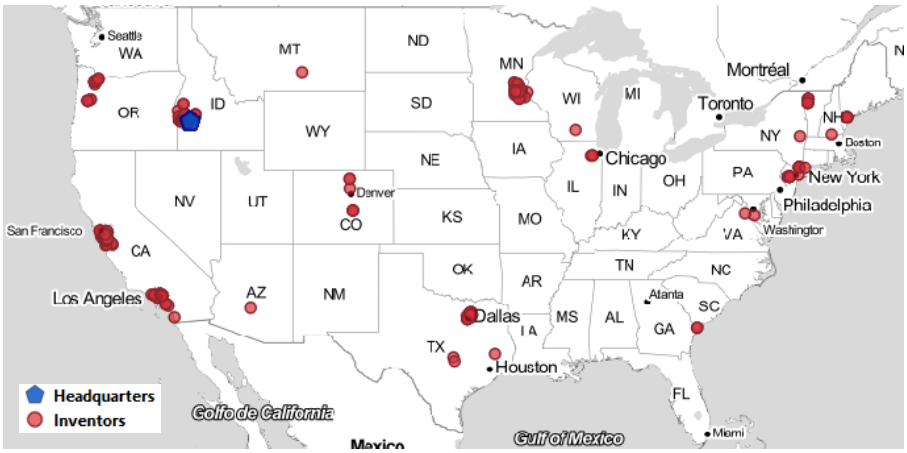
Koh, P. S., & Reeb, D. M. (2015). Missing r&d. *Journal of Accounting and Economics*, *60*(1), 73-94.

Figure IA.1

Heat maps of inventor locations with the firm's headquarters location.

These figures present the location of the firm's headquarters (in blue pentagon) and inventor locations (in red circles) within the United States for the 5 firms with the most observed number of inventors in our data (in order, IBM, Intel, HP, Micron Technology, and Microsoft). The maps exclude Hawaii and Alaska because we observe no inventors located in either state for these five firms.




Seattle
WA
OR
ID
MT
ND
MN
SD
WY
NE
WI
MI
IA
Chicago
IL
IN
OH
PA
Montréal
Toronto
NY
NH
Boston
New York
Philadelphia
Washington
VA
KY
TN
NC
SC
GA
AL
MS
LA
AR
OK
KS
MO
CO
Denver
NV
UT
AZ
NM
TX
Houston
Dallas
Atlanta
Miami
FL
San Francisco
CA
Los Angeles
Headquarters
Inventors
Golfo de California
Gulf of Mexico
Mexico

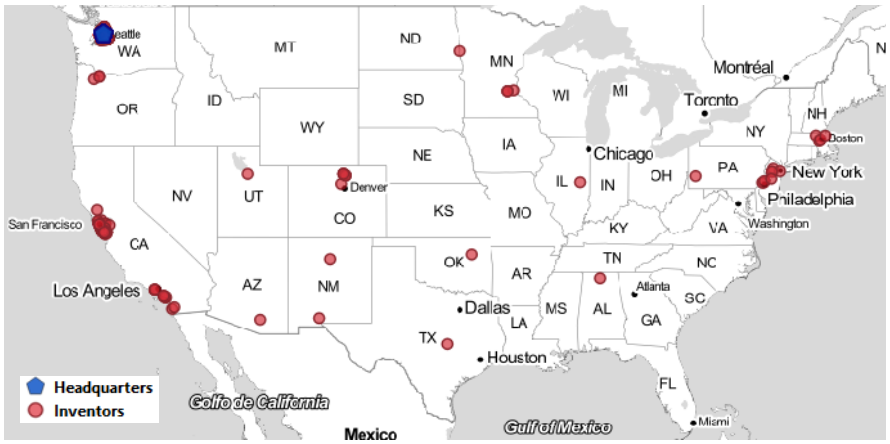

Seattle
WA
OR
ID
MT
ND
MN
SD
WY
NE
WI
MI
IA
Chicago
IL
IN
OH
PA
Montréal
Toronto
NY
NH
Boston
New York
Philadelphia
Washington
VA
KY
TN
NC
SC
GA
AL
MS
LA
AR
OK
KS
MO
CO
Denver
NV
UT
AZ
NM
TX
Houston
Dallas
Atlanta
Miami
FL
San Francisco
CA
Los Angeles
Headquarters
Inventors
Golfo de California
Gulf of Mexico
Mexico

Table IA.1

Proximity and innovation production: Relative size of headquarters inventor team

This Table presents OLS regressions of innovation output as a function of significant flight time decreases between headquarters and the inventor (Panel A) or facility (Panel B), after splitting the sample on whether the firms conducts a majority of R&D at headquarters (majority of inventors living in the headquarters area column(a)) or not (column(b)). All variables are as defined in the Appendix. Innovation output is averaged over three years. The ihs operator refers to the inverse hyperbolic sine function. The Δ operator refers to the difference between the three years pre to the three years post. *t*–statistics appear in parentheses and are based on standard errors clustered by inventor, firm, and date (Panel A) or firm and date (Panel B). ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail). Sample descriptive characteristics are found in Table 1. In our regressions we remove singleton observations, but because these can vary across specifications we report descriptive statistics before their removal.

Panel A: Inventor-level analysis

	(1a)	(1b)	(2a)	(2b)
VARIABLES	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)
	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team
Significant Flight Time Decrease	0.453	0.333**	0.158*	0.094**
	(1.56)	(2.57)	(1.95)	(2.44)
F-stat of the difference	0.152		0.1089	
F-stat p-value	0.697		0.755	
Firm-year fixed effects	Yes	Yes	Yes	Yes
Location-year fixed effects	Yes	Yes	Yes	Yes
Observations	29,577	44,595	30,702	47,082
R-square	0.211	0.181	0.167	0.119

Table IA.1, continued

Proximity and innovation production: Relative size of headquarters inventor team

This Table presents OLS regressions of innovation output as a function of significant flight time decreases between headquarters and the inventor (Panel A) or facility (Panel B), after splitting the sample on whether the firms conducts a majority of R&D at headquarters (majority of inventors living in the headquarters area column(a)) or not (column(b)). All variables are as defined in the Appendix. Innovation output is averaged over three years. The ihs operator refers to the inverse hyperbolic sine function. The Δ operator refers to the difference between the three years pre to the three years post. *t*–statistics appear in parentheses and are based on standard errors clustered by inventor, firm, and date (Panel A) or firm and date (Panel B). ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail). Sample descriptive characteristics are found in Table 1. In our regressions we remove singleton observations, but because these can vary across specifications we report descriptive statistics before their removal.

Panel A, continued: Inventor-level analysis

	(3a)	(3b)	(4a)	(4b)
VARIABLES	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)
	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team
Significant Flight Time Decrease	0.542**	0.280**	0.187**	0.152***
	(2.32)	(2.07)	(1.98)	(2.79)
F-stat of the difference	0.980		0.090	
F-stat p-value	0.322		0.767	
Firm-year fixed effects	Yes	Yes	Yes	Yes
Location-year fixed effects	Yes	Yes	Yes	Yes
Observations	29,577	44,595	29,571	44,587
R-squared	0.138	0.106	0.130	0.100

Table IA.1, continued

Proximity and innovation production: Relative size of headquarters inventor team

This Table presents OLS regressions of innovation output as a function of significant flight time decreases between headquarters and the inventor (Panel A) or facility (Panel B), after splitting the sample on whether the firms conducts a majority of R&D at headquarters (majority of inventors living in the headquarters area column(a)) or not (column(b)). All variables are as defined in the Appendix. Innovation output is averaged over three years. The ihs operator refers to the inverse hyperbolic sine function. The Δ operator refers to the difference between the three years pre to the three years post. *t*–statistics appear in parentheses and are based on standard errors clustered by inventor, firm, and date (Panel A) or firm and date (Panel B). ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail). Sample descriptive characteristics are found in Table 1. In our regressions we remove singleton observations, but because these can vary across specifications we report descriptive statistics before their removal.

Panel B: Facility-level analysis

	(1a)	(1b)	(2a)	(2b)
VARIABLES	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)
	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team
Significant Flight Time Decrease	0.344	1.035***	0.041	0.275**
	(0.75)	(3.13)	(0.20)	(2.54)
Δ*Number Inventors*	0.168***	0.265***	0.140***	0.181***
	(5.80)	(8.03)	(9.87)	(10.51)
F-stat of the difference	1.323		1.188	
F-stat p-value	0.251		0.278	
Firm-year fixed effects	Yes	Yes	Yes	Yes
Location-year fixed effects	Yes	Yes	Yes	Yes
Observations	2,449	5,956	2,572	6,466
R-squared	0.625	0.559	0.700	0.690

Table IA.1, continued

Proximity and innovation production: Relative size of headquarters inventor team

This Table presents OLS regressions of innovation output as a function of significant flight time decreases between headquarters and the inventor (Panel A) or facility (Panel B), after splitting the sample on whether the firms conducts a majority of R&D at headquarters (majority of inventors living in the headquarters area column(a)) or not (column(b)). All variables are as defined in the Appendix. Innovation output is averaged over three years. The ihs operator refers to the inverse hyperbolic sine function. The Δ operator refers to the difference between the three years pre to the three years post. *t*–statistics appear in parentheses and are based on standard errors clustered by inventor, firm, and date (Panel A) or firm and date (Panel B). ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail). Sample descriptive characteristics are found in Table 1. In our regressions we remove singleton observations, but because these can vary across specifications we report descriptive statistics before their removal.

Panel B, continued: Facility-level analysis

	(3a)	(3b)	(4a)	(4b)
VARIABLES	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)
	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team	HQ Team > Non-HQ Team	Non-HQ Team >= HQ Team
Significant Flight Time Decrease	0.751	0.988***	-0.128	0.553***
	(1.53)	(3.97)	(-0.41)	(4.64)
Δ*Number Inventors*	0.050*	0.116***	0.124***	0.175***
	(1.73)	(4.24)	(7.72)	(10.18)
F-stat of the difference	0.212		4.840	
F-stat p-value	0.649		0.030	
Firm-year fixed effects	Yes	Yes	Yes	Yes
Location-year fixed effects	Yes	Yes	Yes	Yes
Observations	2,449	5,956	2,449	5,956
R-squared	0.550	0.438	0.621	0.579

Table IA.2

Proximity and innovation production: Firm-level results

This Table presents OLS regressions of innovation output as a function of significant flight time decreases between the firms' facilities and its headquarters. All variables are as defined in the Appendix. Innovation output is averaged over three years. The ihs operator refers to the inverse hyperbolic sine function. The Δ operator refers to the difference between the three years pre to the three years post. *t*–statistics appear in parentheses and are based on standard errors clustered by firm and date. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail). Sample descriptive characteristics are found in Table 1. In our regressions we remove singleton observations, but because these can vary across specifications we report descriptive statistics before their removal.

	(1)	(2)	(3)	(4)
VARIABLES	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)
Facility-Level Treatments	1.001***	-0.059	0.897***	0.407***
	(4.44)	(-0.66)	(3.04)	(2.73)
Number Facilities	-0.100***	-0.111***	-0.155***	-0.068***
	(-3.50)	(-13.00)	(-7.10)	(-4.36)
ihs(Δ*Market Value*)	0.034**	0.023***	0.047***	0.038***
	(2.25)	(3.38)	(3.63)	(4.37)
ihs(Δ*Equity Return*)	-0.209***	-0.063*	-0.254***	-0.197***
	(-3.05)	(-1.70)	(-2.85)	(-3.17)
ihs(Δ*sigma(Equity Return)*)	0.281	-0.069	0.377	0.279
	(0.47)	(-0.22)	(0.77)	(1.08)
ihs(Δ*R&D Spending*)	0.087**	0.045***	0.087***	0.133***
	(2.13)	(3.38)	(2.63)	(4.61)
ihs(Δ*Return on Assets*)	0.310	0.186	0.236	0.238
	(1.46)	(1.62)	(0.94)	(1.36)
Δ*Loss Indicator*	-0.083	-0.010	-0.087	-0.075
	(-0.69)	(-0.20)	(-0.81)	(-0.88)
Δ*Missing R&D Indicator*	1.115***	0.161	0.622	0.598**
	(2.93)	(0.71)	(1.11)	(2.00)
ihs(Δ*Special Items*)	0.163	0.102	-0.031	0.332***
	(0.55)	(0.62)	(-0.09)	(2.86)
Δ*Market to Book Ratio*	0.094	0.063**	0.127**	0.108**
	(1.37)	(2.04)	(2.04)	(2.12)
Location-year fixed effects	Yes	Yes	Yes	Yes
Industry-year fixed effects	Yes	Yes	Yes	Yes
Observations	11,252	13,642	11,252	11,245
R-squared	0.285	0.271	0.237	0.239

Table IA.3

Proximity and innovation production: natural experiment results excluding changes due to direct route changes

This Table presents OLS regressions of innovation output as a function of significant flight time decreases between headquarters and the inventor (Panel A) or facility (Panel B), after removing observations where the shortest flight route between headquarters and the inventor or facility changed from a connecting route to a direct route, or vice versa. All variables are as defined in the Appendix. Innovation output is averaged over three years. The ihs operator refers to the inverse hyperbolic sine function. The Δ operator refers to the difference between the three years pre to the three years post. *t*–statistics appear in parentheses and are based on standard errors clustered by inventor, firm, and date (Panel A) or firm and date (Panel B). ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two–tail). Sample descriptive characteristics are found in Table 1. In our regressions we remove singleton observations, but because these can vary across specifications we report descriptive statistics before their removal.

Panel A: Inventor-level analysis

	(1)	(2)	(3)	(4)
VARIABLES	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)
Significant Flight Time Decrease	0.452***	0.152***	0.399***	0.168***
	(4.11)	(3.55)	(4.65)	(5.49)
Firm-year fixed effects	Yes	Yes	Yes	Yes
Location-year fixed effects	Yes	Yes	Yes	Yes
Observations	53,105	55,717	53,105	53,098
R-squared	0.209	0.160	0.134	0.127

Panel B: Facility-level analysis

	(1)	(2)	(3)	(4)
VARIABLES	ihs(Δ*Patent Value*$_{avg3}$)	ihs(Δ*Number File*$_{avg3}$)	ihs(Δ*Citations*$_{avg3}$)	ihs(Δ*Scaled Citations*$_{avg3}$)
Significant Flight Time Decrease	1.310	0.317	2.458***	1.017**
	(1.41)	(1.22)	(2.35)	(2.07)
Δ*Number Inventors*	0.234***	0.167***	0.094***	0.159***
	(9.55)	(13.08)	(4.04)	(12.13)
Firm-year fixed effects	Yes	Yes	Yes	Yes
Location-year fixed effects	Yes	Yes	Yes	Yes
Observations	8,195	8,792	8,195	8,195
R-squared	0.558	0.678	0.440	0.576