

# Air travel costs and scientific collaboration

Christian Catalini<sup>1</sup>, Christian Fons-Rosen<sup>2</sup>, Patrick Gaule<sup>3</sup>

<sup>1</sup>MIT <sup>2</sup>Pompeu Fabra <sup>3</sup>CERGE-EI

November 24, 2014

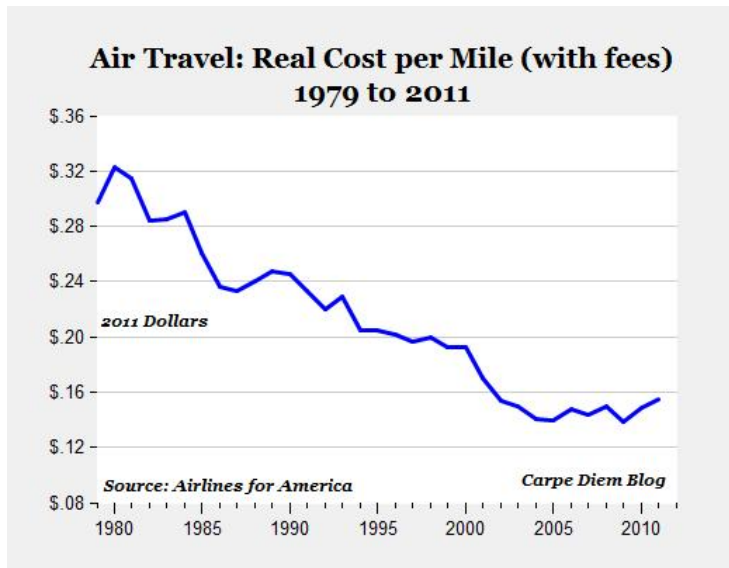
# Scientific knowledge production is changing...

- Knowledge production is increasingly done by teams ([details](#))
- Scientific teams are becoming larger and more likely to include distant collaborators (Adams, Black, Clemmons & Stephan 2005, Wuchty, Jones & Uzzi 2007)
- Decreasing premium from being at a top department (Kim, Morse & Zingales 2008)

# Why?

- Fundamentals of the knowledge production process under knowledge accumulation (Jones 2009)
- Decrease in communication costs with the advent of the internet (Agrawal & Goldfarb 2008, Ding, Levin, Stephan & Winkler 2010)
- Decrease in air travel costs?

# Air travel costs have decreased by half over 30 years



# This paper

- We study the effect of air travel costs on the **incidence of scientific collaboration** between individual scientists
  - ▶ Is collocation still important?
- We are also interested in how a shock to air travel costs affect the **type of collaborations** that occurs



\$\$\$

vs

\$



-> Collaborations?

# Low cost airline entry as source of variation in air travel costs

- The entry of Southwest airlines, the largest U.S. low cost airline, on a given route as a source of variation in prices
  - ▶ Prices decrease on average by about 15% following Southwest entry (details)
  - ▶ Southwest entry has little effect on miles flown or number of connections
- Substantial variation in the timing of Southwest entry

# Empirical strategy: standard difference in differences

$$Y_{ijt} = \alpha + \beta \text{AfterSW}_{ijt} + \mu_t + \gamma_{ij} + \varepsilon_{ijt} \quad (1)$$

Where:

- $Y_{ijt}$  = Number of co-publications of the individual pair  $i, j$  in year  $t$
- $\text{AfterSW}_{ijt}$  = indicator variable that takes value 1 after Southwest entry
- $\gamma_{ij}$  = pair fixed effects.
- $\mu_t$  = year fixed effects



# Threats to identification

- Threat to identification: Southwest may be entering routes based on demand growth, which could itself be correlated with potential determinants of collaboration
- Addressing the threat
  - ▶ Control for (changes in) departmental R&D budgets
  - ▶ Show that the the timing of the effect is consistent with a causal effect
  - ▶ Show that heterogenous effects are consistent with what you would expect if the effect was causal

# Results preview

- Southwest entry leads to a significant and rather large increase in pairwise copublications between faculty (around +50%)
- The effect is stronger for:
  - ▶ pairs of scientists located in department with relatively **low R&D budget**
  - ▶ pairs of **younger** scientists
  - ▶ pairs of scientists that are **more productive** than the average at their respective department
- Stronger effect when weighting publications by cites (proxy for quality)
- Some evidence that the entry of Southwest lead to a decrease in collaborations between pairs that are in the same department or pairs where only one member is affected by entry

# Outline

- 1 Introduction
- 2 Data and institutional context**
- 3 Results
- 4 Conclusion

# Data sources

- **Air travel data.** Bureau of Public Transportation - 10% random sample of itineraries flow within the US
- **Bibliometric data.** Scopus (roughly equivalent to ISI Web of Knowledge)
- **Faculty.** Faculty listings from the American Chemical Society Directory for Graduate Research
- **Misc.** Department rankings from the NRC (1995) assessment of (chemistry) doctoral programs. Department-level *R&D* expenditures from the NSF.

# Faculty data

- Extensive coverage of faculty based in US chemistry, chemical engineering and biochemistry PhD-granting departments anytime between 1991 and 2011 (> 20k individuals)
- Original data cross-sectional (directory issued every two years) but combined to create a longitudinal panel
- Information available: education history, career history, year of birth, gender

# Institutional details about knowledge production in chemistry

- Typical publication lag (from submission to publication) around 3 months
- Median number of authors per paper is 4
- The modal number of faculty authors per paper is 1 - collaborations among faculty are a minority of papers
- Chemistry is mostly done in labs but there are branches of chemistry and chemical engineering that rely on mathematical derivations or computer simulations instead

# Matching universities to airports

- Universities could be located close to several airports (Columbia; JFK, EWR)
- The relevant market is not the nearest airport but all airports relatively close by
- So we match a university to all airports within a 50 miles radius
  - ▶ Some universities do not have any airports in a 50 miles radius, these are dropped from the analysis

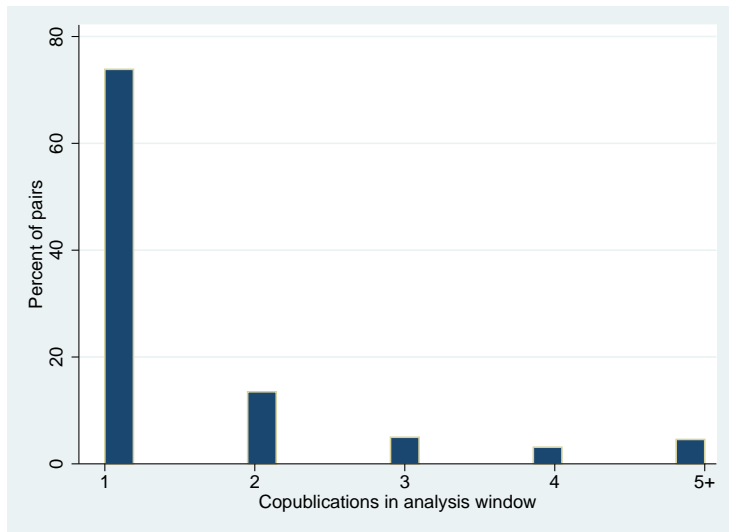
# Sample

We will focus the analysis on 482 pairs of scientists based in US chemistry departments who

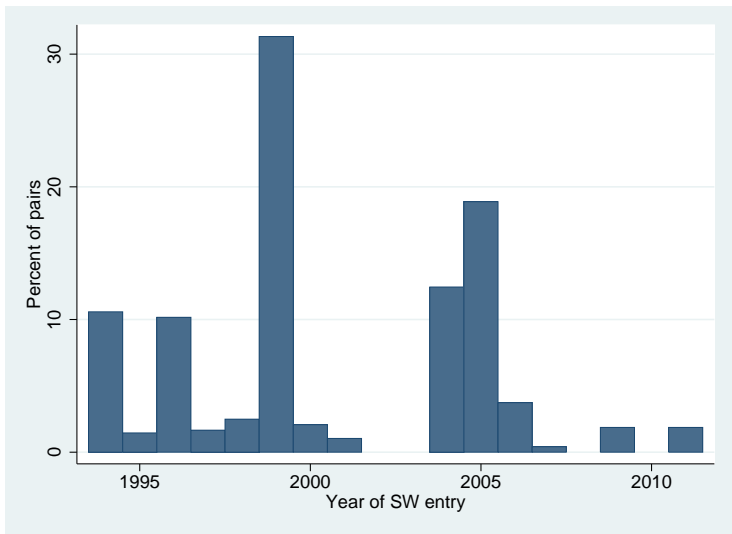
- are not within driving distance ( 300 miles)
- experience Southwest entry between 1993 and 2012
  - ▶ 231 distinct new routes
- co-publish at least once in a 10 years window around the time of Southwest entry
- whose members do not relocate to a different department around the time of Southwest entry



# Copublications in analysis window



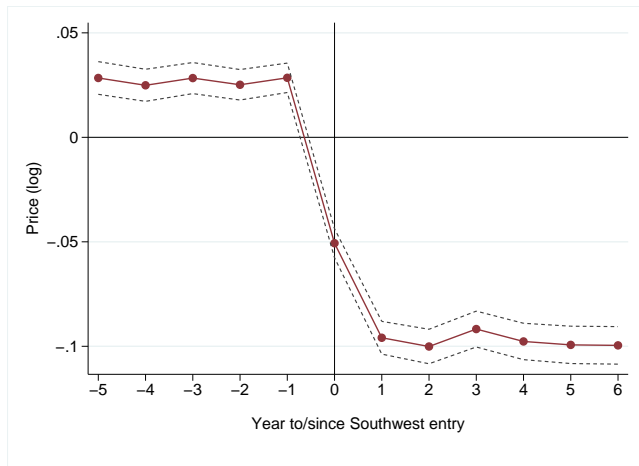
## Southwest entry by year



# Outline

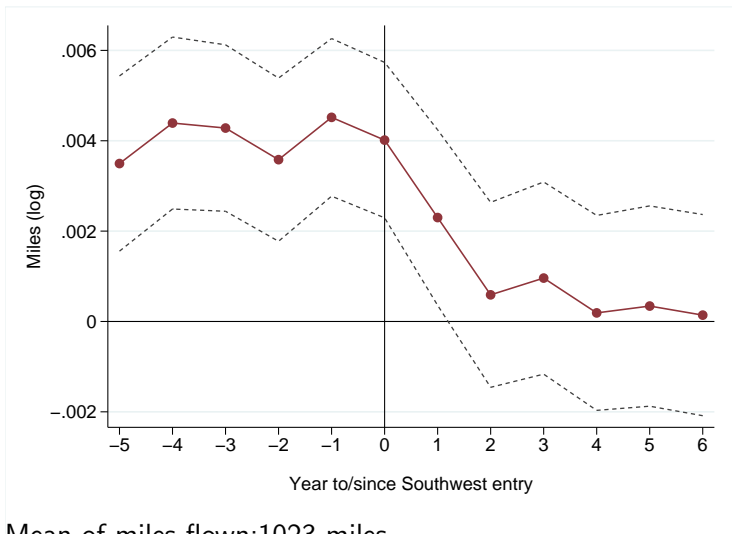
- 1 Introduction
- 2 Data and institutional context
- 3 Results**
- 4 Conclusion

## What does Southwest entry do? - Effect on price



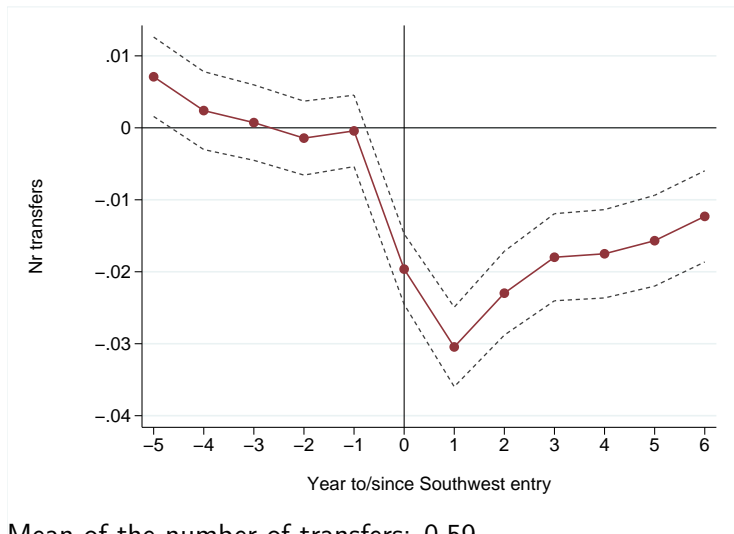
Note: The average prices are for all flights between the endpoints not only flights operated by Southwest.

# What does Southwest entry do? - Effect on miles flown



Notes: Mean of miles flown:1023 miles

# What does Southwest entry do? - Effect on transfers



Notes: Mean of the number of transfers: 0.59.

## Main effect of Southwest entry

	(1)
	Co-publications
Southwest entry	0.510*** (0.170)
Pair Fixed Effects	Yes
Year Fixed Effects	Yes
Mean of dependent variable	0.18
Number of pairs	482
Number of observations	4615

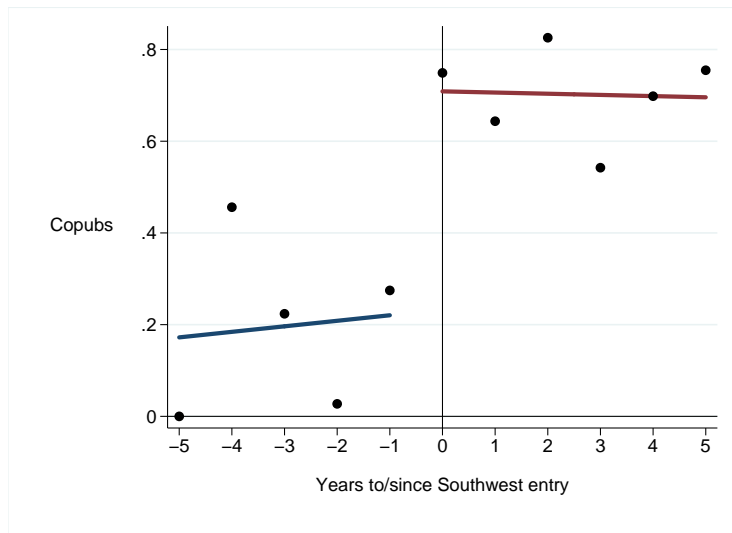
Robust standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

▶ Robustness to serial correlation

▶ Robustness to alternative functional forms

## Dynamics of the effect of Southwest entry





## Extensive or intensive margin?

Column 1: excluding pairs that collaborate before and after SW entry

Column 2: only pairs that collaborate before and after SW entry

	(1)	(2)
Copubs	Extensive	Intensive
Southwest entry	0.411*	0.951***
	(0.220)	(0.252)
Pair Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Mean of dependent variable	0.14	0.48
Number of pairs	427	55
Number of observations	4059	556

Robust standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Extensive or intensive margin weighted by cites

Column 1: excluding pairs that collaborate before and after SW entry

Column 2: only pairs that collaborate before and after SW entry

	(1)	(2)
Copubs weighted by cites	Extensive	Intensive
Southwest entry	0.741** (0.330)	0.693* (0.380)
Pair Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Mean of dependent variable	5.60	19.18
Number of pairs	413	55
Number of observations	3917	556

Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Stronger effect when weighting by cites

	(1) Copubs	(2) Copubs weighted by cites
Southwest entry	0.510*** (0.170)	0.704*** (0.257)
Pair Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Mean of dependent variable	0.18	7.29
Number of pairs	482	468
Number of observations	4615	4473

Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

	(1)	(2)	(3)	(4)
	Copubs	Copubs	Copubs	Copubs
Southwest entry	0.297 (0.190)	0.307* (0.184)	0.071 (0.240)	-0.279 (0.260)
Southwest X both in dept. with below median R&D budget	0.548** (0.217)			0.491** (0.194)
Southwest X both below 50 (at time of entry)		0.576** (0.243)		0.550** (0.222)
Southwest X one more productive than dept average			0.506** (0.216)	0.470** (0.215)
Southwest X both more productive than dept average			0.601** (0.258)	0.554** (0.244)
Pair Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Number of pairs	482	482	482	482
Number of observations	4615	4615	4615	4615

Robust standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Southwest entry and collaboration among local pairs

	(1) copubs
Southwest entry	-0.418* (0.224)
Pair Fixed Effects	Yes
Year Fixed Effects	Yes
Mean of dependent variable	0.24
Number of pairs	164.00
Number of observations	1592

Robust standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Note: pairs in universities with below median R&D budget

What happen to pairs located in the same department after Southwest start operating at least one flight from that location)?

## Indirect effect of Southwest

	(1) copubs
Southwest entry	-0.516** (0.240)
Pair Fixed Effects	Yes
Year Fixed Effects	Yes
Mean of dependent variable	0.18
Number of pairs	198
Number of observations	1765

Robust standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

What happens to pairs where one coauthor experiences Southwest entry (Southwest start operating at least one flight from his/her location) but the other does not?

# Conclusion

- Air travel costs affect the incidence of scientific collaboration
  - ▶ the secular decrease in air travel costs may thus potentially explain part of the increase in distant collaborations
  - ▶ scientists respond to air travel costs!

# The increasing dominance of teams in science and innovation

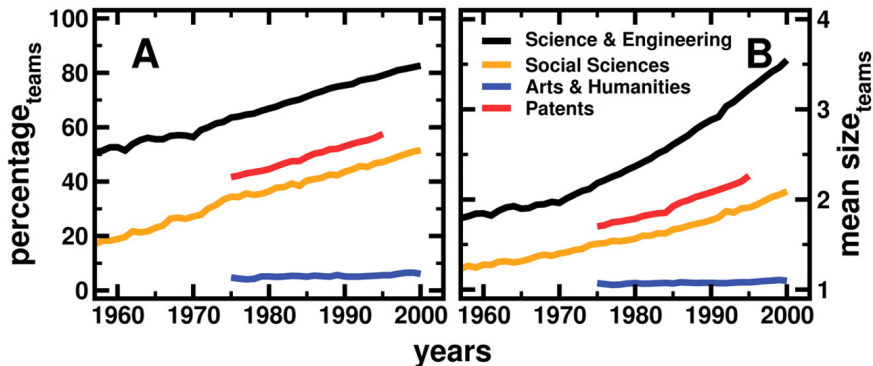


Figure : Source: Wuchty, Jones & Uzzi (2007)



## Robustness to serial correlation

	(1) Residuals
Southwest entry	0.021*** (0.008)
Constant	-0.013** (0.006)
Number of observations	12258
R2	0.00

Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Bertrand and Mullainathan (2004) warn against serial correlation causing spurious inference in difference in differences designs. They suggest to first regress the outcome on cross-sectional fixed effects and pair fixed effects; and then regress the residuals of that regression on an indicator variable for the post period

## Robustness to alternative functional forms

	(1) Poisson QML Copubs	(2) OLS Copubs	(3) OLS Any copub. (0/1)
Southwest entry	0.510*** (0.170)	0.087*** (0.031)	0.063** (0.026)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Mean of dependent variable	0.18	0.18	0.15
Number of pairs	482	482	482
Number of observations	4615	4615	4615

Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

▶ Go back