# Productivity Spillover among Scientific Workers

Wei Cheng

The Ohio State University

ASSA 2016 meeting

# Motivation

- Knowledge and technology is one of the major engines for economic growth.
- A growing trend in scientific collaboration (Wuchty, Jones and Uzzi, 2007).
- This paper investigates how a researcher's productivity is affected by his collaborators' productivities.

# Mechanism

How can a researcher be affected by his collaborators?

1. Collaborative ties serve as conduit for knowledge, skills and ideas. Learning opportunities.
2. Gaining access to tangible resources, such as research funding and editorial goodwill, by the help of collaborators.
3. Peer pressure.

# Literature

- Main challenge in identification: endogenous choice of collaborators
- The literature has taken advantage of events where one's collaborators are exogenously removed. (Azoulay, Graff Zivin and Wang, 2010; Waldinger, 2012; Borjas and Doran, 2014)
- Gap:
  - Formation of collaborator relationships
  - Contemporaneous change in the productivity of one's collaborators

# Econometric model

- Spatial autoregressive model:

$$y_{i,t} = \lambda w_{i,t} Y_t + x_{i,t}\beta + u_{i,t}; \quad (1)$$

$$Y_t = \lambda W_t Y_t + X_t\beta + U_t. \quad (2)$$

  where $w_{i,t}$ is a row vector indicating $i$'s collaborators by value 1.

- If $W_t$ is exogenous,

$$Y_t = (I - \lambda W_t)^{-1}(X_t\beta + U_t) = \sum_{j=0}^{\infty} \lambda^j W_t^j (X_t\beta + U_t); \quad (3)$$

$$W_t Y_t = \sum_{j=0}^{\infty} \lambda^j W_t^{j+1} (X_t\beta + U_t). \quad (4)$$

  Therefore, $(W_t X_t, W_t^2 X_t, \cdots)$ can instrument for $W_t Y_t$.

- If $W_t$ is endogenous, we can first estimate $\hat{W}_t$ and use $(\hat{W}_t X_t, \hat{W}_t^2 X_t, \cdots)$ as instrument for $W_t Y_t$.

# Two-stage strategy

- stage 1. what affect people's choice of collaborators

$$w_{ij,t} = 1\{\gamma_0 + x_{i,t}\gamma_1 + x_{j,t}\gamma_2 + c_{ij,t}\gamma_3 + z_{ij,t}\gamma_4 + \varepsilon_{ij,t} > 0\}$$

- stage 2. whether people are influenced by their collaborators

- Instruments
  - Overlap of research interests (Ductor, 2014)
  - Overlap in collaborative network: the ratio of common collaborators in *i*'s and *j*'s total collaborators (Fafchamps, Leij and Goyal, 2010)

- Choice-based sampling (Cosslett, 1981)

# Data

- National Library of Medicine's Medline publication database
- Torvik and Smalheiser "Author-ity" 2009

- Field: Alzheimer Disease, Dementia and Neurodegenerative Diseases (ADN)
  - *Alzheimer is the fourth largest cause of death for people over 65 years of age. As many as 5 million Americans age 65 and older may have AD, and that number is expected to double for every 5-year interval beyond age 65.*
  - *Alzheimer is the most common form of dementia.*
- Less than 15% of all ADN papers are singled authored. The median of the number of authors on a paper is 5.
- NIH research funding has increased greatly in these three fields, reaching over 3 billion in 2015.

# Sample

- Active authors during the time window 1995-2006, and have at least 5 publications in his whole career.
- The sample consists of 6,437 researchers. 20% female and 60% male.
- Treat three years as one period.
- Define collaborators as previous coauthors.

# Dyad regression

$$w_{ij,t} = 1\{\gamma_0 + x_{i,t}\gamma_1 + x_{j,t}\gamma_2 + c_{ij,t}\gamma_3 + z_{ij,t}\gamma_4 + u_{ij,t} > 0\}$$

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| $common_r$ | **0.028*** ** | **0.023*** ** | **0.022*** ** | **0.019*** ** | **0.019*** ** | **0.018*** ** |
| | **(0.001)** | **(0.001)** | **(0.001)** | **(0.001)** | **(0.001)** | **(0.001)** |
| *mesh* | | | | **0.076*** ** | **0.058*** ** | **0.057*** ** |
| | | | | **(0.001)** | **(0.002)** | **(0.002)** |
| $col_{t-1}$ | | | 6.790*** | | 5.264*** | 5.272*** |
| | | | (0.409) | | (0.411) | (0.411) |
| *degrf.i* | | | | | | 0.496*** |
| | | | | | | (0.087) |
| *degrf.j* | | | | | | 0.738*** |
| | | | | | | (0.085) |
| *degrf.i* $*$ *degrf.j* | | | | | | -0.499*** |
| | | | | | | (0.095) |
| *gender_comF.F* | | 0.292*** | 0.259** | 0.202* | 0.209* | 0.206* |
| | | (0.089) | (0.104) | (0.103) | (0.109) | (0.110) |
| *gender_comM.F* | | 0.238*** | 0.254*** | 0.218*** | 0.235*** | 0.229*** |
| | | (0.044) | (0.050) | (0.051) | (0.053) | (0.054) |
| *gender_comM.M* | | 0.313*** | 0.343*** | 0.338*** | 0.346*** | 0.346*** |
| | | (0.038) | (0.044) | (0.046) | (0.047) | (0.047) |

*N*ote: All regressions except (1) include i's and j's numbers of publications in the previous year and their difference,i's age, j's age and their difference, and year dummies as controls. Constant term is not reported. $common_r$ is scaled up by multiplying 100. Standard errors are reported in the parentheses.

# Productivity spillover

| | Cross section | | | | First difference |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $W_{i,t} \cdot prod_t$ | **0.033*** ** | **0.033*** ** | **0.032*** ** | **0.049*** ** | **0.057*** ** |
| | **(0.001)** | **(0.001)** | **(0.001)** | **(0.007)** | **(0.007)** |
| $prod_{i,t-1}$ | | | | 0.906*** | -0.057*** |
| | | | | (0.006) | (0.016) |
| $W_{i,t} \cdot prod_{t-1}$ | | | | -0.049*** | -0.054*** |
| | | | | (0.007) | (0.006) |
| *age* | | | 0.278*** | 0.022** | |
| | | | (0.015) | (0.009) | |
| $age^2$ | | | -0.004*** | -0.001*** | |
| | | | (0.0003) | (0.0002) | |
| *genderM* | | | 1.085*** | 0.268*** | |
| | | | (0.099) | (0.058) | |
| *genderU* | | | 0.966*** | 0.205*** | |
| | | | (0.117) | (0.068) | |
| *Year F.E.* | No | Yes | Yes | Yes | – |

*N*ote: Standard errors are reported in the parentheses. Constant term is not reported.

- In a period, a researcher's average number of collaborators is 8, average number of publication is 4.

# Conclusion

- Factors that influence people's choice of collaborators:
  - number of common previous coauthors
  - overlap in research interest
  - whether they have coauthored before
- On average a researcher's number of publications increases by around 0.05 unit when his coauthor's total productivity increases by one unit.