Spillovers in Space: Does Geography Matter?

Sergey Lychagin*  Joris Pinkse*  Margaret Slade**  
John Van Reenen***

* Penn State, ** University of British Columbia, *** LSE

CORE, Louvain-la-Neuve, March 5, 2012
Background

R&D spillovers occur when firm $j$’s R&D $\rightarrow i$ more productive

Spillovers are externalities

When they are important, markets fail to allocate resources efficiently

We expect spillovers to occur if firms $i$ and $j$ are close
What does ‘close’ mean?
Two questions of interest:

1) What are the sources of R&D spillovers? Along which dimensions are firms close?

- Vertical
- Horizontal
- Technological
- Geographical
2) Is distance dead?
Does geography matter?

- In a world where information flows electronically (instantaneously) distance might not matter

- If this is true, why do local governments try to attract high-tech firms?
Examples of the two opposing views

Frances Cairncross in her book *The Death of Distance* (1997) states that:

- “The death of distance as a determinant of the cost of communicating will probably be the single most important force shaping society in the first half of the next century”

- “New ideas will spread faster, leaping borders. Poor countries will have immediate access to information that was once restricted to the industrial world”
However:

- President Obama recently proposed increasing US R&D tax credits since US subsidies would lead to US growth.

- Greenstone Hornbeck and Moretti (2010) find that locating a large new plant increases the productivity of other plants in the region.

Perhaps distance is not dead.
Some policy questions:

- Is knowledge **private** or **public**?
  Matters for public policy such as R&D subsidies
  Of interest to public economists

- Are spillovers **local** or **global**?
  Are knowledge flows confined to narrowly defined product markets, technology classes, and/or geographic regions?
  Matters for market and regional growth and convergence
  Of interest to macroeconomists
I will discuss:

- Measures of productivity
- Measures of closeness (or its inverse, distance)
- The empirical R&D–spillover literature to date
- Then concentrate on geographic proximity and present some new measures and new results
Productivity

Productivity growth occurs when there is a shift in the production possibility frontier.

More output can be produced from the same inputs.

More output can be produced at the same cost.

This is different from a growth in output/input holding the production function constant.
Approaches to productivity measurement:

Will consider a production function framework

• The Econometric approach – Estimation

• The Index Number approach – Calculation
Both are based on the relationship

\[ Q = AF(x) \]  \hspace{1cm} (1)

where \( Q \) is output

\( A \) is the level of productivity (assumed to be Hick’s neutral)

and \( x \) is a vector of conventional inputs

In logs

\[ q = \ln(Q) = \ln(A) + \ln(F(x)) = \tilde{a} + f(x) \]  \hspace{1cm} (2)

We are interested in measuring \( \tilde{a} \)
The econometric approach:

Suppose that we have a panel of $n$ firms (regions, industries), $i = 1, \ldots, n$, observed over time, $t$

Let

$$\tilde{a}_{it} = a_{it} + \mu_t + u_{it}$$

(3)

where $a$ is systematic (can depend on own and rival R&D), $\mu$ is an aggregate shock, and $u$ is an idiosyncratic shock.

Then

$$q_{it} = f(x_{it}) + a_{it} + \mu_t + u_{it}$$

(4)
One can estimate (4), \( q_{it} = f(x_{it}) + a_{it} + \mu_t + u_{it} \)

Could use, e.g., firm and time–period fixed effects

Many have done this

There are important problems

I focus on a particular problem, *endogeneity*
Endogenous inputs:

Correlation between $u$ and $x$

Noted by Marschak and Andrews (1944)

A bad shock could cause the firm to use fewer variable inputs

Will causes a bias in the production–function coefficients

Often results in under estimation of returns to scale
Possible solutions:

  A dynamic estimation with fixed factors (K) and variable factors (L)
  Investment in K chosen optimally

- Arrellano and Bond (1991), Blundell and Bond (2000)
  A GMM approach
  Suggest instruments and moment conditions

A continental divide
The index–number approach:

A productivity index is an index of outputs divided by an index of inputs.

One can rewrite equation (1) as

\[
\frac{Q_{it}}{F(x_{it})} = A_{it}
\]  

(5)

Must still estimate production function coefficients.

Alternatively, one can specify a productivity index

\[
TFP(q_{it}, x_{it}) = a_{it} + \mu_t + u_{it}
\]  

(6)

The index \( TFP(q_{it}, x_{it}) \) can be calculated, not estimated.
One must choose a functional form for $TFP$

One possibility, A Tornqvist (for multilateral comparisons)

$$TFP_{it} = \frac{(q_{it} - \bar{q}_t)}{\sum_k \frac{1}{2}(s_{ikt} - \bar{s}_{kt})(x_{ikt} - \bar{x}_{kt})},$$ \hspace{1cm} (7)

where $s$ is a factor cost share and bars denote firm averages

Once the index has been calculated, the equation,

$$TFP_{it} = a_{it} + \mu_t + u_{it},$$ \hspace{1cm} (8)

can be estimated

An advantage: The endogenous variables ($x$) are on the left–hand side
Which should one choose?

An age divide
Spillovers

I now want to put some structure on $a_{it}$

Let

$$S_{it} = (1 - D)S_{it-1} + R_{it-1},$$

(9)

where $S_{it}$ is $i$’s stock of knowledge in period $t$, $D$ is the depreciation rate, $R_{it}$ is firm $i$’s investment in knowledge, and

$$a_{it} = a(S_{it}, S_{-it})$$

(10)

where $S_{-it}$ is a vector of knowledge stocks of other firms
Assume that \( a_{it} \) is a weighted average of the knowledge capital of all firms

\[
a_{it} = \theta S_{it} + \sum_{j \neq i} w_{ij} S_{jt},
\]

(11)

where \( w_{ij} \) is a weight that corresponds to some notion of closeness between \( i \) and \( j \).

Different weighting matrices correspond to different notions of closeness.

Consider four possibilities from the literature.
Types of Spillovers:

- **Vertical**: Learning from suppliers or retailers  
  *Example*: Auto bodies and auto assembly

- **Horizontal**: Learning from product–market rivals  
  *Example*: New drugs

- **Technological**: Learning from technology–market rivals  
  *Example*: Froth flotation

- **Geographic**: Learning from meeting  
  *Example*: Same golf club
Measures of Spillovers:

- **Vertical**
  - Some of the earliest studies
  - Highly aggregate data
  - Input/Output flows (Importance of buying and selling)

- **Horizontal**
  - Firm data
  - 0/1, Same or different product market
  - or Jaffee (1986)–style coefficient

- **Technological**
  - Firm data
  - Jaffe coefficient
The Jaffe (1986) measure (for technology markets)

R&D occurs in $K$ technology classes, indexed by $k$

Let $F_{ik}$ be the fraction of firm $i$’s R&D that is in class $k$

$$w_{ij} = \frac{\sum_{k=1}^{K} F_{ik}F_{jk}}{\sqrt{[\sum_{k=1}^{K}(F_{ik})^2][\sum_{k=1}^{K}(F_{jl})^2]}}.$$  \hspace{1cm} (12)

This is the uncentered correlation coefficient

All classes are treated symmetrically — no class is ‘closer’ to any other

Spillovers occur within but not across classes. In other words, there are no $F_{ik}F_{jl}$ terms with $l \neq k$
• **Geographic**
  Country, industry or firm data
  0/1, Same or different geographic market
  Exponential decline in geographic distance

  Firms are ‘located’ in the region (i.e., country or city) where their **headquarters** are located
Some Results

What has the profession learned about the sources and magnitudes of spillovers?

Will look at some earlier studies in all four classes

Note: Most researchers assess a single source in isolation

A potential bias
Table 1: **Empirical Assessment of Spillovers: Vertical**

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Measure of Productivity</th>
<th>Measure of Closeness</th>
<th>Finding for Spillovers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scherer</td>
<td>1982</td>
<td>Cross section</td>
<td>L Productivity</td>
<td>Tech flow matrix</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US Line of Business</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Griliches&amp;</td>
<td>1984</td>
<td>Cross section</td>
<td>Index number</td>
<td>Input/output table</td>
<td>Weak</td>
</tr>
<tr>
<td>Lichtenberg</td>
<td></td>
<td>US manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goto&amp;Suzuki</td>
<td>1989</td>
<td>Cross section</td>
<td>Index number</td>
<td>Tech flow matrix</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Japanese manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Data</td>
<td>Measure of Productivity</td>
<td>Measure of Closeness</td>
<td>Finding for Spillovers</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------</td>
<td>---------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Bernstein</td>
<td>1989</td>
<td>Panel</td>
<td>Econometric</td>
<td>0/1 Same SIC</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Cost fn)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams &amp; Jaffe</td>
<td>1996</td>
<td>Panel</td>
<td>Index number</td>
<td>0/1 Same SIC</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orlando</td>
<td>2004</td>
<td>Panel</td>
<td>Econometric</td>
<td>0/1 Same SIC</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ornaghi</td>
<td>2006</td>
<td>Panel</td>
<td>Econometric</td>
<td>0/1 Same SIC</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bloom, Schankerman</td>
<td>2008</td>
<td>Panel</td>
<td>Econometric</td>
<td>Jaff Coef</td>
<td>None</td>
</tr>
<tr>
<td>&amp; Van Reenen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Data</td>
<td>Measure of Productivity</td>
<td>Measure of Closeness</td>
<td>Finding for Spillovers</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>-----------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Jaffe</td>
<td>1986</td>
<td>2 Cross sections</td>
<td>Econometric</td>
<td>Jaffe coefficient</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>US firms</td>
</tr>
<tr>
<td>Bloom, Schankeman &amp; Van Reenen</td>
<td>2008</td>
<td>Panel</td>
<td>Econometric</td>
<td>Jaffe coefficient</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>US manu. firms</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Data</td>
<td>Measure of Productivity</td>
<td>Measure of Closeness</td>
<td>Finding for Spillovers</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>---------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Adams &amp; Jaffe</td>
<td>1996</td>
<td>Panel US</td>
<td>Index number</td>
<td>0/1 Close/distant</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orlando</td>
<td>2004</td>
<td>Panel US</td>
<td>Econometric</td>
<td>0/1 Close/distant</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eaton &amp; Kortum</td>
<td>1996</td>
<td>Cross section</td>
<td>L productivity</td>
<td>0/1 3 lead countries</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keller</td>
<td>2002</td>
<td>Panel</td>
<td>Index number</td>
<td>Exponential decay</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Does Geography Matter?

Will now concentrate on geographic closeness

- Some new measures
- Some new results

Will present new results for three measures (no vertical)
Some new measures of geographic proximity:

Previous studies focus on the location of headquarters

Several geographic distance measures have been used:

- 0/1 variable for headquarters in different (same) region

- A declining function of the Euclidean distance between the countries where firm $i$ and $j$ headquarters are located
Is the location of headquarters or research labs more important?

Who communicates with whom? CEOs or researchers?

Their locations are correlated but far from perfectly

Examples:
We assess both **headquarter** and **lab** locations.

We don’t have the location of labs.

However, we have patent data.

Each patent gives the address of its principal inventor.

We construct geographic R&D locational distributions from those addresses.
Suppose that there are $K$ geographic regions, $k = 1, \ldots, K$

Let $F_{ik}$ be the fraction of firm $i$’s inventors that are located in region $k$

Our geographic measures are of the form

$$w_{ij} = \sum_k \sum_l F_{ik}F_{jl}C(d_{kl}), \quad i \neq j, \quad w_{ii} = 0,$$  \hspace{1cm} (13)

where $d_{kl}$ is the Euclidean distance between regions $k$ and $l$
Measures of Geographic Distance:

We use two parametric functions $C(d_{kl})$ of the Euclidean distance between regions

- 0/1, $k \neq l$ or $k = l$
  
  Location matters but not distance (Jaffe)

- Exponential decay in distance, $d_{kl}$
  
  Distance also matters (modified Jaffe plus Keller)

These are referred to as correlation and exponential

We also estimate $C(d_{kl})$ nonparametrically
Note that our measures specialize to a measure of distance based on the location of headquarters.

A headquarters measure is like a firm with one lab located in the same market as its headquarters.
We condition on measures of Horizontal and Technological proximity

Avoid spurious correlation

We use Jaffe measures for both (Location matters but distance in product and technology markets doesn’t)

No vertical data
Data

Two US firm-level data sources are matched

- **Compustat** 1980–2000 – A 21 year panel
  US manufacturing firms
  Data on $Q$, $K$, $L$, $M$, $R&D$, etc.

Create a breakdown of each firm’s activities across
349 3-digit product-market SICs
A product–market distribution
On average each firm reports sales in 4.8 industry codes
• **NBER patents data** (Hall, Jaffe, and Trajtenberg 2005)

  Detailed patenting information for about 2,500 firms with almost 3 million patents

  Create a breakdown of each R&D active firm’s patenting activities in 410 technology classes
  A technology–market distribution
  On average each R&D active firm owns 497 patents in 37 technology classes

  Each patent has the address of the principal inventor
  Create a breakdown of each firm’s patenting activities in 2039 geographic markets (US counties)
  A geographic–market distribution
The matched sample of US manufacturing firms:

812 patenting firms with at least 4 observations on all variables between 1980 and 2001

An unbalanced panel

These firms account for about 90% of all R&D in the larger sample

We also include R&D inactive firms
Will present results based on two approaches

- Production functions
- Index numbers
Production–function estimates

A Cobb-Douglas production function

The dependent variable is $lnQ$, where $Q$ is real sales. $\$ sales are deflated by sector specific price indices.

The explanatory variables are:

Conventional inputs: $lnK$, $lnL$, and $lnM$.

Own stock of knowledge ($\sqrt{OwnS}$).

Product-market spillover variable ($lnSpillSIC$).

Technology-market spillover variable ($lnSpillTech$).

Geographic-market spillover variable ($lnSpillGeog$).

Firm and year fixed effects.
Estimation methods:

We use four methods:

- OLS
- Instrumental variables (IV)
- Semiparametric least squares
- Semiparametric IV
Econometric issues:

- We correct for serial correlation but usually assume that it disappears after $r$ periods.

- We handle endogeneity of $x$ using Blundell and Bond (2000).

- We use basis functions that allow us to impose restrictions on decay (e.g., monotonic or exponential).
Table 1: Correlations Between Spillover Measures

<table>
<thead>
<tr>
<th>Correlation between</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIC–Tech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIC–Geog</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geog–Tech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weights, $W_M$</td>
<td>0.359</td>
<td>0.040</td>
<td>0.038</td>
</tr>
<tr>
<td>$lnSpillM$</td>
<td>0.420</td>
<td>0.162</td>
<td>0.220</td>
</tr>
<tr>
<td>$lnSpillM$, FE removed</td>
<td>0.460</td>
<td>0.129</td>
<td>0.327</td>
</tr>
</tbody>
</table>

Notes:
All correlations are significant at the 1% level
In column one, $M$ denotes SIC, Tech, or Geog
The 3rd row uses the residuals from a regression with firm and year fixed effects

Table 2: Distance Between Inventors and Headquarters

<table>
<thead>
<tr>
<th>Percentile</th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>0 km</td>
<td>0 km</td>
<td>52 km</td>
<td>414 km</td>
<td>1656 km</td>
<td>8097 km</td>
</tr>
</tbody>
</table>
Table 3: **OLS Production Function Estimates, Single Proximity Measures**

Dependent variable is \( \ln Q \)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpillSIC</td>
<td>( \ln K ) 0.061 (3.94) 0.056 (3.73) 0.059 (3.70) 0.045 (3.07)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpillTech</td>
<td>( \ln L ) 0.308 (12.33) 0.318 (13.19) 0.314 (12.66) 0.336 (14.56)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpillGeog</td>
<td>( \ln M ) 0.627 (23.58) 0.606 (24.08) 0.620 (23.56) 0.591 (24.24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OwnS</td>
<td>( \sqrt{OwnS} ) 0.085 (3.55) 0.064 (2.83) 0.089 (4.29) 0.059 (2.79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnSpillSIC</td>
<td>0.293 (5.53) 0.074 (1.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnSpillTech</td>
<td>0.878 (7.58) 0.661 (6.37)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnSpillGeog</td>
<td>1.023 (5.37) 0.808 (4.57)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Based on inventor locations
- Firm and year fixed effects included
- \( t \) statistics are in parentheses (clustering by firm)
- Exponential–decay distance function (\( C \))
- Effect of distance decays at a rate of 50% per 200km
<table>
<thead>
<tr>
<th>Based on locations of:</th>
<th>Inv.</th>
<th>Inv.</th>
<th>Inv.</th>
<th>Head.</th>
<th>Head.</th>
<th>Both</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Fn (C)</td>
<td>Exp</td>
<td>Corr</td>
<td>Both</td>
<td>Exp</td>
<td>Corr</td>
<td>Exp</td>
<td>Corr</td>
</tr>
<tr>
<td>$\sqrt{OwnS}$</td>
<td>0.059</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
<td>0.061</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(2.79)</td>
<td>(2.66)</td>
<td>(2.86)</td>
<td>(2.93)</td>
<td>(2.69)</td>
<td>(2.89)</td>
<td>(2.64)</td>
</tr>
<tr>
<td>$lnSpillSIC$</td>
<td>0.074</td>
<td>0.094</td>
<td>0.070</td>
<td>0.074</td>
<td>0.079</td>
<td>0.071</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>(1.76)</td>
<td>(2.24)</td>
<td>(1.69)</td>
<td>(1.82)</td>
<td>(1.85)</td>
<td>(1.72)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>$lnSpillTech$</td>
<td>0.661</td>
<td>0.723</td>
<td>0.678</td>
<td>0.697</td>
<td>0.791</td>
<td>0.655</td>
<td>0.730</td>
</tr>
<tr>
<td></td>
<td>(6.37)</td>
<td>(6.73)</td>
<td>(6.50)</td>
<td>(6.73)</td>
<td>(7.16)</td>
<td>(6.47)</td>
<td>(6.78)</td>
</tr>
<tr>
<td>$lnSpillGIExp$</td>
<td>0.808</td>
<td>0.954</td>
<td>0.400</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.57)</td>
<td>(4.84)</td>
<td>(2.40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnSpillGICor$</td>
<td>0.183</td>
<td>-0.111</td>
<td>0.186</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.00)</td>
<td>(-2.02)</td>
<td>(2.83)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnSpillGHExp$</td>
<td></td>
<td></td>
<td></td>
<td>0.410</td>
<td>0.272</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.74)</td>
<td>(3.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnSpillGCorH$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.028</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.83)</td>
<td>(1.18)</td>
</tr>
</tbody>
</table>

Conventional inputs included
Firm and year fixed effects included
t statistics in parentheses (clustering by firm)
Table 5: **Production Function Estimates, Covariance Estimation**

<table>
<thead>
<tr>
<th>Dependent variable: $lnQ$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t stat</td>
<td>t stat</td>
<td>$r = 0$</td>
<td>$r = 2$</td>
<td>$r = 21$</td>
<td>Unrestr.</td>
<td></td>
</tr>
<tr>
<td>$lnK$</td>
<td>0.225</td>
<td>29.0</td>
<td>21.5</td>
<td>24.1</td>
<td>14.5</td>
<td>11.9</td>
<td>11.1</td>
</tr>
<tr>
<td>$lnL$</td>
<td>0.614</td>
<td>62.9</td>
<td>45.4</td>
<td>52.1</td>
<td>31.8</td>
<td>25.4</td>
<td>23.6</td>
</tr>
<tr>
<td>$lnOwnS$</td>
<td>0.054</td>
<td>10.2</td>
<td>8.17</td>
<td>9.38</td>
<td>5.42</td>
<td>4.09</td>
<td>4.07</td>
</tr>
<tr>
<td>$lnSpillSIC$</td>
<td>-0.025</td>
<td>-1.42</td>
<td>-1.40</td>
<td>-1.58</td>
<td>-0.88</td>
<td>-0.65</td>
<td>-0.64</td>
</tr>
<tr>
<td>$lnSpillTech$</td>
<td>0.395</td>
<td>10.0</td>
<td>8.07</td>
<td>9.22</td>
<td>4.85</td>
<td>3.1</td>
<td>3.11</td>
</tr>
<tr>
<td>$lnSpillGeog$</td>
<td>0.524</td>
<td>17.9</td>
<td>11.0</td>
<td>13.4</td>
<td>6.86</td>
<td>4.16</td>
<td>4.15</td>
</tr>
</tbody>
</table>

**Notes:**
Based on inventor locations.
Exponential–decay distance function, half-life distance = 200km
Firm and year fixed effects included
Columns (4) – (6) dependence no greater than lag $r$ is assumed
Columns (3) and (7) allow for heteroskedasticity across $t$ and $i$
Columns (4) – (6) allow for heteroskedasticity across $i$, but not $t$
Index–number estimates

Our productivity index number is

\[ PROD_{it} = (q_{it} - \bar{q}_i) / \sum_k \frac{1}{2} (s_{ikt} - \bar{s}_{ik})(x_{ikt} - \bar{x}_{ik}). \]  (14)

We compare firm \( i \) in period \( t \) to an ‘average’ firm \( i \)
Average over time, not firms

Preserves technological heterogeneity across firms

Introduces possible trends
Picked up by \( t \) fixed effects
Advantages

- Endogenous inputs, \( x \), are on left–hand side. Solves input endogeneity problem.

- Exact for a translog technology, which is flexible. Allows for arbitrary substitution patterns.

- Consistent with increasing (decreasing) returns and nonhomotheticity.

- Production–function parameters can differ by firm up to the second order.
Estimation

The dependent variable is \( TFP \)

The explanatory variables exclude the conventional inputs but are otherwise the same

We use the same four estimation techniques
Econometric issues

• We handle endogeneity of $S$ using Blundell and Bond (2000)

• The other issues are handled in the same way
## Table 1: OLS Index Number Estimates, Single Proximity Measures

Dependent variable is $lnQ$

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{OwnS}$</td>
<td>0.082</td>
<td>0.054</td>
<td>0.088</td>
<td>0.043</td>
</tr>
<tr>
<td>$lnSpillSIC$</td>
<td>0.378</td>
<td>0.119</td>
<td></td>
<td>0.119</td>
</tr>
<tr>
<td>$lnSpillTech$</td>
<td>0.980</td>
<td></td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>$lnSpillGeog$</td>
<td></td>
<td>1.149</td>
<td>0.847</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Based on inventor locations
- Firm and year fixed effects included
- $t$ statistics are in parentheses (clustering by firm)
- Exponential-decay distance function
- Effect of distance decays at a rate of 50% per 200km
Table 2: **OLS Index Numbers, Inventors & Headquarters, Different C Functions**

Dependent variable is $lnQ$

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on locations of:</td>
<td>Inv.</td>
<td>Inv.</td>
<td>Inv.</td>
<td>Head.</td>
<td>Head.</td>
<td>Both</td>
<td>Both</td>
</tr>
<tr>
<td>Distance Fn $(C)$</td>
<td>Exp</td>
<td>Corr</td>
<td>Both</td>
<td>Exp</td>
<td>Corr</td>
<td>Exp</td>
<td>Corr</td>
</tr>
<tr>
<td>$\sqrt{OwnS}$</td>
<td>0.043</td>
<td>0.046</td>
<td>0.044</td>
<td>0.045</td>
<td>0.049</td>
<td>0.043</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(2.03)</td>
<td>(2.02)</td>
<td>(2.08)</td>
<td>(2.17)</td>
<td>(2.11)</td>
<td>(2.09)</td>
<td>(2.01)</td>
</tr>
<tr>
<td>$lnSpillSIC$</td>
<td>0.119</td>
<td>0.144</td>
<td>0.117</td>
<td>0.122</td>
<td>0.132</td>
<td>0.116</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>(2.61)</td>
<td>(3.13)</td>
<td>(2.58)</td>
<td>(2.72)</td>
<td>(2.84)</td>
<td>(2.58)</td>
<td>(2.94)</td>
</tr>
<tr>
<td>$lnSpillTech$</td>
<td>0.675</td>
<td>0.748</td>
<td>0.688</td>
<td>0.720</td>
<td>0.835</td>
<td>0.677</td>
<td>0.754</td>
</tr>
<tr>
<td>$lnSpillGIExp$</td>
<td>0.847</td>
<td>0.955</td>
<td></td>
<td>0.478</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.22)</td>
<td>(4.30)</td>
<td></td>
<td>(2.56)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnSpillGICor$</td>
<td>0.216</td>
<td>-0.082</td>
<td>0.204</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.10)</td>
<td>(-1.45)</td>
<td>(3.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnSpillGHExp$</td>
<td></td>
<td></td>
<td>0.412</td>
<td>0.245</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.11)</td>
<td>(2.50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnSpillGCorH$</td>
<td></td>
<td></td>
<td>0.027</td>
<td>0.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.52)</td>
<td>(0.87)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Firm and year fixed effects included
t statistics in parentheses (clustering by firm)
Conclusions

• Location matters

• Distance matters

• Locations of researchers are more important than locations of headquarters but both matter

• Other dimensions of distance (technological and product market) are important but they don’t cause the effect of geographic proximity to disappear
Policy issues

- Spillover effects are Large (economically important and statistically significant) → Public-good

- Geographic markets are very Local → Bad news for convergence
Why does geography matter?

- Social learning and capitalization of complementarities

- Agglomeration effects – e.g., better infrastructure and more skilled workers

- Skilled workers are not very mobile
  When they change jobs many locate in the same region
  (Combes and Duranton 2010 find 75% for France)

- However, even when an inventor leaves a region, the region benefits from the inventor’s ideas
  (Agrawal, Cockburn, and McHale 2003)
  This is especially important across fields
Possible extensions

- Finish what we have started

- We have specified a $C$ function for geographic distance. Could do the same for product and technology markets. For example, cars and trucks are closer than cars and computers. Could use different levels of SICs and technology classes as measures of closeness.

- Our index–number estimates do not separate productivity from returns to scale. Could use methods proposed by Diewert and Fox (2004) to separate the two. However, we can’t reject CRTS.