R&D Policy and Technological Trajectories of Regions: Evidence from the EU Framework Programmes

Wolf-Hendrik Uhlbach ¹ Pierre-Alexandre Ballard ² Thomas Scherngell ³

¹ Copenhagen Business School
² Utrecht University
³ Austrian Institute of Technology GmbH

STI 2017: Location Based Approaches
07.09.2017
Motivation

- Regions can not be too dependent on existing specializations (e.g. vulnerability to shocks)
Motivation

- Regions cannot be too dependent on existing specializations (e.g., vulnerability to shocks)

- Policy efforts to foster the development of new economic activities in regions (e.g., European Cohesion policy)
Motivation

- Regions can not be too dependent on existing specializations (e.g. vulnerability to shocks)
- Policy efforts to foster the development of new economic activities in regions (e.g. European Cohesion policy)
- Importance of availability of different knowledge and capabilities for alternative specialisations
New specializations build on previously acquired knowledge that is transferred to new fields

Regions are constrained in their ability to develop new activities

⇒ Technological Relatedness as a driver of diversification
Diversification patterns also depend on

- Development stage (Petralia et al., 2017)
- Institutions (Boschma & Capone, 2015; Cortinovis et al., 2016)
- Industrial and innovation policy (Rodrik, 2004; Foray, 2009; Mazzucato, 2013)
Diversification patterns also depend on
- Development stage (Petralia et al., 2017)
- Institutions (Boschma & Capone, 2015; Cortinovis et al., 2016)
- Industrial and innovation policy (Rodrik, 2004; Foray, 2009; Mazzucato, 2013)

Increasing policy interest also by means of instruments intending to support diversification capabilities, in particular at a regional level, e.g.: Smart Specialisation Strategy of the EC
Diversification patterns also depend on:
- Development stage (Petralia et al., 2017)
- Institutions (Boschma & Capone, 2015; Cortinovis et al., 2016)
- Industrial and innovation policy (Rodrik, 2004; Foray, 2009; Mazzucato, 2013)

Increasing policy interest also by means of instruments intending to support diversification capabilities, in particular at a regional level, e.g.: Smart Specialisation Strategy of the EC

Also past efforts to stimulate knowledge spillovers to foster innovation capabilities of regions: EU Framework Programmes (FP)
Intended Effects of Collaborative R&D Projects

- Subsidizing collaborative R&D projects in a certain technology leads to higher patenting activity in the respective technology.
Intended Effects of Collaborative R&D Projects

- Subsidizing collaborative R&D projects in a certain technology leads to higher patenting activity in the respective technology
- Additional financial resources
- Potential knowledge spillover from collaboration partners
Intended Effects of Collaborative R&D Projects

- Subsidizing collaborative R&D projects in a certain technology leads to higher patenting activity in the respective technology
- Additional financial resources
- Potential knowledge spillover from collaboration partners
- STI policy can direct undertaken research
Hypotheses

1. Regions are more likely to specialise in technologies for which they receive R&D subsidies.

2. Funding tends to compensate for a lack of related capabilities.

Diagram:

- Incumbent Technology
- Possible Entry
- Expected Entry
- Link indicating relatedness
- R&D subsidy
Hypotheses

$H_1$ Regions are more likely to specialise in technologies for which they receive R&D subsidies

- Incumbent Technology
- Possible Entry
- Expected Entry

Link indicating relatedness → R&D subsidy
Hypotheses

$H_1$ Regions are more likely to specialise in technologies for which they receive R&D subsidies.

$H_2$ Funding tends to compensate for a lack of related capabilities.
## Data Sources

### FP Data (EUPRO*)
- FP5; FP6; FP7
- Patent relevant sub-programmes focusing on collaboration
- 15,983 projects
- Participants classified on NUTS2 regions

*Provision of original data via the RISIS (Research Infrastructure for and Innovation Policy Studies) infrastructure (risis.eu)*

### Patent Data (REGPAT)
- Patent applications, fractionalized by inventor
- 282 NUTS2 Regions, 613 IPC Classes
### Data Sources

#### FP Data (EUPRO*)
- FP5; FP6; FP7
- Patent relevant sub-programmes focusing on collaboration
- 15,983 projects
- Participants classified on NUTS2 regions

#### Technology Fields (Schmoch, 2008)
- Aggregation of IPC classes
- Balanced field sizes
- Distinct field contents

#### Patent Data (REGPAT)
- Patent applications, fractionalized by inventor
- 282 NUTS2 Regions, 613 IPC Classes

*Provision of original data via the RISIS (Research Infrastructure for and Innovation Policy Studies) infrastructure (risis.eu)
Data Sources

FP Data (EUPRO*)
- FP5; FP6; FP7
- Patent relevant sub-programmes focusing on collaboration
- 15,983 projects
- Participants classified on NUTS2 regions

Technology Fields (Schmoch, 2008)
- Aggregation of IPC classes
- Balanced field sizes
- Distinct field contents

Patent Data (REGPAT)
- Patent applications, fractionalized by inventor
- 282 NUTS2 Regions, 613 IPC Classes

*Provision of original data via the RISIS (Research Infrastructure for and Innovation Policy Studies) infrastructure (risis.eu)
Empirical Setting


- \( ENTRY_{i,r,t} \): Emergence of a new specialization in a region

\[
ENTRY_{i,r,t} = \begin{cases} 
0, & RCA_{i,r,t} < 1 \land RCA_{i,r,t-1} < 1 \\
1, & RCA_{i,r,t} \geq 1 \land RCA_{i,r,t-1} < 1
\end{cases}
\]

- \( RD_{i,r,t} \): Relatedness Density
  - Technological relatedness based on co-occurrences on patent files
  - Determine Relatedness Density: For each technology in a region, share of existing related technologies on all related technologies

FP_{z,r,t}: Number of Participations weighted by technologies and periods
Empirical Setting


- $ENTRY_{i,r,t}$: Emergence of a new specialization in a region

\[
ENTRY_{i,r,t} = \begin{cases} 
0, & RCA_{i,r,t} < 1 \land RCA_{i,r,t-1} < 1 \\
1, & RCA_{i,r,t} \geq 1 \land RCA_{i,r,t-1} < 1 
\end{cases}
\]

- $RD_{i,r,t}$: Relatedness Density
  
  1. Technological relatedness based on co-occurrences on patent files
  2. Determine **Relatedness Density**: For each technology in a region, share of existing related technologies on all related technologies
Empirical Setting


- \( ENTRY_{i,r,t} \): Emergence of a new specialization in a region

\[
ENTRY_{i,r,t} = \begin{cases} 
0, & RCA_{i,r,t} < 1 \land RCA_{i,r,t-1} < 1 \\
1, & RCA_{i,r,t} \geq 1 \land RCA_{i,r,t-1} < 1 
\end{cases}
\]

- \( RD_{i,r,t} \): Relatedness Density
  1. Technological relatedness based on co-occurrences on patent files
  2. Determine Relatedness Density: For each technology in a region, share of existing related technologies on all related technologies

- \( FP_{z,r,t} \): Number of Participations weighted by technologies and periods
Empirical Model

\[ ENTRY_{i,r,t} = \beta_1 FP_{z,r,t-1} + \beta_2 RD_{i,r,t-1} + \beta_3 REG + \beta_4 TECH + \]

FP Participation

Relatedness Density

Controls

\[ + \beta_5 FP_{z,r,t-1} \times RD_{i,r,t-1} + \phi_r + \psi_i + \alpha_t + \epsilon_{i,r,t} \]

Interaction Effect

Fixed Effects

i: technology
z: technology field
r: regions
t: time
Differences in the Mean of Entry Probabilities

Legend
- funded
- not funded

Quantiles of Relatedness Density

entry

0.00
0.05
0.10
0.15

0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85
Results 1

<table>
<thead>
<tr>
<th></th>
<th>Pooled FP (1)</th>
<th>Pooled RD (2)</th>
<th>Baseline (3)</th>
<th>Full Model (5)</th>
<th>Full Model F.E. (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>log(FP)</strong></td>
<td>0.0242***</td>
<td>0.0174***</td>
<td>0.0083***</td>
<td>0.0171***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0008)</td>
<td>(0.0010)</td>
<td>(0.0020)</td>
<td></td>
</tr>
<tr>
<td><strong>RD</strong></td>
<td>0.0033***</td>
<td>0.0031***</td>
<td>0.0022***</td>
<td>0.0003***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
<td>(0.00005)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>log(FP) × RD</strong></td>
<td>−0.0010***</td>
<td>−0.0007***</td>
<td>−0.0005***</td>
<td>−0.0005***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.1067***</td>
<td>0.1067***</td>
<td>0.1097***</td>
<td>0.1101***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0006)</td>
<td>(0.0006)</td>
<td>(0.0007)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>284,508</td>
<td>284,508</td>
<td>284,508</td>
<td>212,751</td>
<td>212,751</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.0047</td>
<td>0.0197</td>
<td>0.0220</td>
<td>0.0260</td>
<td>0.0632</td>
</tr>
</tbody>
</table>
## Results Different Levels of RD

<table>
<thead>
<tr>
<th></th>
<th>Low RD (1)</th>
<th>Low RD (2)</th>
<th>Low RD (3)</th>
<th>Mid RD (4)</th>
<th>Mid RD (5)</th>
<th>Mid RD (6)</th>
<th>High RD (7)</th>
<th>High RD (8)</th>
<th>High RD (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FP)</td>
<td>0.0115</td>
<td>−0.0035</td>
<td>−0.0006</td>
<td>0.0133***</td>
<td>0.0089***</td>
<td>0.0164***</td>
<td>−0.0004</td>
<td>−0.0048*</td>
<td>0.0291***</td>
</tr>
<tr>
<td></td>
<td>(0.0087)</td>
<td>(0.0118)</td>
<td>(0.0142)</td>
<td>(0.0021)</td>
<td>(0.0028)</td>
<td>(0.0047)</td>
<td>(0.0022)</td>
<td>(0.0028)</td>
<td>(0.0055)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0449***</td>
<td>0.0466***</td>
<td>0.1276***</td>
<td>0.1283***</td>
<td>0.1648***</td>
<td>0.1628***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0032)</td>
<td>(0.0018)</td>
<td>(0.0022)</td>
<td>(0.0022)</td>
<td>(0.0026)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>5,636</td>
<td>4,400</td>
<td>4,400</td>
<td>33,155</td>
<td>23,400</td>
<td>23,400</td>
<td>27,566</td>
<td>20,023</td>
<td>20,023</td>
</tr>
<tr>
<td>R²</td>
<td>0.0003</td>
<td>0.0057</td>
<td>0.0995</td>
<td>0.0013</td>
<td>0.0048</td>
<td>0.0952</td>
<td>0.00001</td>
<td>0.0018</td>
<td>0.0911</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.0001</td>
<td>0.0044</td>
<td>0.0288</td>
<td>0.0012</td>
<td>0.0046</td>
<td>0.0618</td>
<td>−0.00004</td>
<td>0.0015</td>
<td>0.0530</td>
</tr>
</tbody>
</table>
Conclusions and Discussion

- Participations in collaborative R&D projects weakly associated with an increase the entry probability
  - 10\% increase of project participations will only be associated with a 2\% increase of the mean probability of a technology to enter the region’s portfolio

Impact of R&D subsidies is highest if the level of relatedness density is neither too high nor too low.

Need for more research to investigate exact mechanisms and causality.

Do study on micro level, e.g. using publication data.
Conclusions and Discussion

- Participations in collaborative R&D projects weakly associated with an increase the entry probability
  - 10% increase of project participations will only be associated with a 2% increase of the mean probability of a technology to enter the region’s portfolio
- Participations can to a certain extent compensate for a lack of relatedness

Impact of R&D subsidies is highest if the level of relatedness density is neither too high nor too low

Need for more research to investigate exact mechanisms and causality

Do study on micro level, e.g. using publication data
Conclusions and Discussion

- Participations in collaborative R&D projects weakly associated with an increase the entry probability
  - 10% increase of project participations will only be associated with a 2% increase of the mean probability of a technology to enter the region’s portfolio
- Participations can to a certain extent compensate for a lack of relatedness
- Impact of R&D subsidies is highest if the level of relatedness density is neither too high nor too low

Need for more research to investigate exact mechanisms and causality

Do study on micro level, e.g. using publication data
Conclusions and Discussion

- Participations in collaborative R&D projects weakly associated with an increase the entry probability
  - 10% increase of project participations will only be associated with a 2% increase of the mean probability of a technology to enter the region’s portfolio
- Participations can to a certain extent compensate for a lack of relatedness
- Impact of R&D subsidies is highest if the level of relatedness density is neither too high nor too low
- Need for more research to investigate exact mechanisms and causality
- Do study on micro level, e.g. using publication data
Thank your for your attention!

Wolf-Hendrik Uhlbach
PhD Fellow Copenhagen Business School
Department for Innovation and Organisational Economics
EMAIL
References


European Commission (2014) National/Regional innovation strategies for smart specialization (RIS3)


### Annex I

**Dependent variable:**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FP)</td>
<td>0.0242***</td>
<td>0.0159***</td>
<td>0.0065***</td>
<td>0.0156***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0008)</td>
<td>(0.0009)</td>
<td>(0.0018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>0.0033***</td>
<td>0.0030***</td>
<td>0.0023***</td>
<td>0.0003***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
<td>(0.00005)</td>
<td>(0.00001)</td>
<td>(0.00001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP Dens</td>
<td>0.0018***</td>
<td>0.0011***</td>
<td>0.0006***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
<td>(0.00005)</td>
<td>(0.00001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP/CAP</td>
<td>1.6743***</td>
<td>0.5452***</td>
<td>−3.3205***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0656)</td>
<td>(0.0714)</td>
<td>(0.9437)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pop Dens</td>
<td>−0.00001***</td>
<td>−0.00003***</td>
<td>−0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00001)</td>
<td>(0.00001)</td>
<td>(0.00001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GERD (mio)</td>
<td>−0.00001***</td>
<td>−0.00001***</td>
<td>−0.00001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000001)</td>
<td>(0.000001)</td>
<td>(0.000001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech Grth</td>
<td>0.0001***</td>
<td>0.0001***</td>
<td>0.0001***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00001)</td>
<td>(0.00001)</td>
<td>(0.00002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(FP)×RD</td>
<td>−0.0010***</td>
<td>−0.0007***</td>
<td>−0.0004***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.00001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.1080***</td>
<td>0.1115***</td>
<td>0.1079***</td>
<td>0.1150***</td>
<td>0.1133***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0006)</td>
<td>(0.0007)</td>
<td>(0.0006)</td>
<td>(0.0007)</td>
<td></td>
</tr>
</tbody>
</table>

Observations: 284,508, 284,508, 230,650, 284,508, 230,650, 230,650
Adjusted R²: 0.0047, 0.0197, 0.0212, 0.0220, 0.0274, 0.0654

**Note:** *p<0.1; **p<0.05; ***p<0.01
### Descriptive statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>entry</td>
<td>284,508</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FP</td>
<td>345,732</td>
<td>2.9</td>
<td>11.5</td>
<td>0.0</td>
<td>485.0</td>
</tr>
<tr>
<td>RD</td>
<td>345,732</td>
<td>17.8</td>
<td>12.9</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
Control variables

- $TECH_{i,t-1}$:
  - Funding Density: Share of related industries that receive funding
  - Technology growth: Growth rate of a technology in the previous period

- $REG_{r,t-1}$:
  - GDP per capita
  - Population Density
  - Gross expenditure for research and development (GERD)
Text Classification Strategy

Classify > 1000 projects manually and use as training data

Make a document term matrix

| Used text: Titles + Project Abstract + Objective + Achievements + Title of Subprogramme + Titles of Resulting Documents |
| Preprocessing: Remove short terms (<2), stop words (and, or, etc.), non-alphanumerical terms, weight terms by Tfidf |

Fit a maximum entropy classifier
# Text Classification Strategy

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classify &gt; 1000 projects manually and use as training data</td>
</tr>
<tr>
<td>Make a document term matrix</td>
</tr>
<tr>
<td><strong>Used text:</strong> Titles + Project Abstract + Objective + Achievements + Title of Subprogramme + Titles of Resulting Documents</td>
</tr>
<tr>
<td>Preprocessing: Remove short terms (&lt;2), stop words (and, or, etc.), non-alphanumerical terms, weight terms by Tfidf</td>
</tr>
<tr>
<td>Fit a maximum entropy classifier</td>
</tr>
<tr>
<td>Apply classifier to test data</td>
</tr>
<tr>
<td>Use L2 regularizer to prevent over fitting</td>
</tr>
<tr>
<td>Classify each project to 5 TFs based on probability scores</td>
</tr>
<tr>
<td><strong>External verification using an inventory of 295 patents from FP7 ICT projects</strong></td>
</tr>
</tbody>
</table>
## Logit and Probit Specification


<table>
<thead>
<tr>
<th>Dependent var. is ENTRY</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>is ENTRY</td>
<td>Logit</td>
<td>Probit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Logit</th>
<th>Probit</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FP)</td>
<td>0.443940 ***</td>
<td>0.2361080 ***</td>
</tr>
<tr>
<td>FP Density</td>
<td>−0.009639 ***</td>
<td>−0.0050260 ***</td>
</tr>
<tr>
<td>Density</td>
<td>0.027360 ***</td>
<td>0.0156848 ***</td>
</tr>
<tr>
<td>log(FP) * Density</td>
<td>−0.007595 ***</td>
<td>−0.0042357 ***</td>
</tr>
<tr>
<td>Region F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Technology F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Time F.E.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

| AIC                     | 16130   | 16123   |
| N                       | 18,840  | 18,840  |

***p < 0.001, **p < 0.01, *p < 0.05
## Results Cross Sectional OLS

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(FP)</td>
<td>0.0256***</td>
<td>0.0373***</td>
<td>0.0209***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0021)</td>
<td>(0.0028)</td>
<td></td>
</tr>
<tr>
<td>Relatedness (RD)</td>
<td>0.0035***</td>
<td>0.0038***</td>
<td>0.0011***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>log(FP) × RD</td>
<td></td>
<td>−0.0011***</td>
<td>−0.0006***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.0981***</td>
<td>0.0513***</td>
<td>0.0400***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>140,023</td>
<td>140,023</td>
<td>140,023</td>
<td>140,023</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.0048</td>
<td>0.0213</td>
<td>0.0235</td>
<td>0.0602</td>
</tr>
</tbody>
</table>

**Note:** *p<0.1; **p<0.05; ***p<0.01