Cross-cutting analysis of scientific output versus other STI Indicators

The European Research Area
Context

- The cross-linking of R&D inputs with outputs has increased in the past decades using various methods to study **economies** and **diseconomies of scale** in scientific production to provide policy advice aimed at:
  - improving the allocation and management of resources;
  - ultimately improving the productivity of S&T systems.

- R&D at a regional scale is highly heterogeneous; STI indicators should ideally be relevant and comparable on a regional and national scale to support the development of national/regional policies tailored to the needs of specific countries/regions.

- Yet, large-scale studies on a regional scale remain scarce mainly as a result of the **lack of comprehensive data** on inputs and outputs.

- **This study investigates the factors driving scientific productivity** using the most comprehensive dataset on STI indicators that is currently available for **countries** and **NUTS2 regions** of the **ERA**.
Methods & Results: Selection of Indicators

**Inputs**
- R&D Investment & Expenditure
  - GERD
  - HERD
  - GOVERD
  - BERD
- Human Resources
  - Researchers in the Higher Education Sector
  - HRST with Tertiary Education
  - PhD Students
  - PhD Graduates
  - Foreign Students in Tertiary Education
  - Job-to-Job Mobility of HRST
- Innovation
  - Employment in Technology and Knowledge-Intensive Sectors
  - VCI (Expansion & Replacement)
  - VCI (Buyout)
  - VCI (Early Stage)
- Research Infrastructure
  - Number of new research infrastructures
  - Average lower bound of research infrastructure investment

**Outputs (No. of Papers)**
Methods & Results: Factor Analysis (Country)

- Exploratory Factor Analysis (EFA) for identifying the main dimensions (i.e., factors) explaining patterns of variation among selected STI indicators and the publication output
  - Log Transformation of raw data to meet the assumptions of:
    - Normality
    - Linearity
  - RI indicators dropped
  - GERD dropped
  - Missing data: Pairwise deletion
  - Dataset included repeated measures over time (from 2000 to 2009)

A = Publications (N = 405), B = GERD (N = 339), C = HERD (N = 339), D = BERD (N = 336), E = GOVERD (N = 340), F = HRST with Tertiary Education (N = 218), G = Researchers in the Higher Education Sector (N = 263), H = Foreign Students in Tertiary Education (N = 237), I = PhD Graduates (N = 288), J = PhD Students (N = 262), K = Job-to-Job Mobility of HRST (N = 186), L = Employment in Technology and Knowledge-Intensive Sectors (N = 269), M = VCI (Buyout) (N = 183), N = VCI (Early Stage) (N = 201), O = VCI (Expansion & Replacement) (N = 210), P = Research Infrastructure (N = 31) and Q = Average Lower Bound of RI investment (N = 31)
### Methods & Results: Factor Analysis (Country)

Factor loadings of selected STI indicators on the first factor of the exploratory factor analysis using PCA factoring

<table>
<thead>
<tr>
<th>Indicator</th>
<th>R</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publications (FRAC)</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Job-to-Job Mobility of HRST</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>HERD</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>HRST with Tertiary Education</td>
<td>0.95</td>
<td>0.89</td>
</tr>
<tr>
<td>BERD</td>
<td>0.95</td>
<td>0.91</td>
</tr>
<tr>
<td>GOVERD</td>
<td>0.94</td>
<td>0.89</td>
</tr>
<tr>
<td>PhD Graduates</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>Researchers in the Higher Education Sector</td>
<td>0.92</td>
<td>0.84</td>
</tr>
<tr>
<td>Employment in Technology and Knowledge-Intensive Sectors</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>Foreign Students in Tertiary Education</td>
<td>0.91</td>
<td>0.83</td>
</tr>
<tr>
<td>PhD Students</td>
<td>0.90</td>
<td>0.81</td>
</tr>
<tr>
<td>VCI (Expansion &amp; Replacement)</td>
<td>0.84</td>
<td>0.71</td>
</tr>
<tr>
<td>VCI (Buyout)</td>
<td>0.78</td>
<td>0.61</td>
</tr>
<tr>
<td>VCI (Early Stage)</td>
<td>0.72</td>
<td>0.52</td>
</tr>
</tbody>
</table>

% of Total Variance Explained by the 1<sup>st</sup> Factor  **83%**

- High multicollinearity
- One factor explains 83% of the variance in the dataset
- The output variable is almost perfectly correlated with this factor
Methods & Results: Regression Analysis (Country)

- Even if a single factor can summarise the dataset, there are differences in the way countries allocate R&D spending across sectors (e.g., higher education, government, private) and resources (e.g., HR, infrastructure).

- It is therefore of interest to investigate how the publication output of countries scale relative to individual R&D input indicators using regression analysis.

- In investigating the impact of individual R&D input indicator on the production of countries, the following question arose:
  - Whether the two variables were scaling linearly (i.e., an isometric pattern) or whether one variable was scaling exponentially relative to the other (i.e., an allometric pattern)?
Methods & Results: Regression Analysis (Country)

- Interpreting the pattern of change in the ratio between two variables as one increases (Smith, 2009)

Allometric scaling implies a power law relationship:
\[ y = kx^a \]
or using the logs
\[ \log(y) = a \log(x) + \log(k) \]

- If the 95 CI of the slope did not overlap 1, it was concluded that there were either significant economies of scale \((a > 1)\) or diminishing returns \((a < 1)\)
Methods & Results: Regression Analysis (Country)

Which regression methods to use?

- **Multicollinearity**: Multiple regression could not be used due to potential downsides including spurious conclusions regarding the significance of the regression coefficients and coefficients of unexpected sign.

- **Autocorrelation**: Panel data structure made up of cross-sections (i.e., countries) and time-series (i.e., years = 10).
  - **Implications**: Regression coefficients on the pooled dataset (i.e., all countries and years) would likely compress the confidence intervals of the slopes, increasing the likelihood of falsely concluding that there are either diminishing returns or economies of scale.

- **Potential models**: the fixed-effects model, the between-effects model, the random-effects model and the dynamic panel data model.
Methods & Results: Regression Analysis (Country)

Why the between-effects model?

- **Focus is on the population response means:** the need to account for the within cross-section (i.e., country) variation and autocorrelation does not matter as much as when the analysis aims to investigate subject-specific (i.e., country-specific) effects of explanatory variables.

- In such cases, one can go for **robust inference** using a **between-effects model**, which fits a group-mean regression (Gardiner et al., 2009).

- Group-mean regressions were fitted by means of S-estimators (robust regression) (Rousseeuw & Yohai, 1984). **Ideal when:**
  - Outliers might be present in both the response and explanatory variables;
  - In the presence of violations of the assumptions of normality and homoscedasticity of the residuals.
Methods & Results: Regression Analysis (Country)

- **Diminishing returns** with respect to 5 out of 6 R&D investment indicators (3 specifically on innovation; i.e., VCI)
- Nearly significant and moderate in the case of HERD
Methods & Results: Regression Analysis (Country)

- **Economies of scale** with respect to employment in Tech & KIS
- Potential **slight diminishing returns** with respect to PhD Students

![Graphs showing regression analysis results](image)
Methods & Results: Regression Analysis (NUTS2)

- **Economies of scale** with respect to researchers in higher education
- **Diminishing returns** with respect to R&D Expenditures

Log(Publications) vs. Log(HERD):
- Robust $R^2 = 0.86$
- Slope = 0.84
- 95% CI = [0.81 - 0.88]

Log(Publications) vs. Log(Researchers in the HES):
- Robust $R^2 = 0.82$
- Slope = 1.08
- 95% CI = [1.03 - 1.13]

Log(Publications) vs. Log(GOVERD):
- Robust $R^2 = 0.67$
- Slope = 0.69
- 95% CI = [0.63 - 0.74]

Log(Publications) vs. Log(BERD):
- Robust $R^2 = 0.54$
- Slope = 0.57
- 95% CI = [0.51 - 0.64]
Methods & Results: Regression Analysis (NUTS2)

- Reduced Major Axis (RMA) with CIs of the slopes estimated by bootstrapping:
  - Accounts for errors in both variables which is likely the case in this study
  - Still sensitive to departure from normality and to heteroscedasticity

- Conclusions about **diminishing returns are confirmed with RMA** but are less pronounced than with robust regression (i.e., slopes are closer than they were to $a = 1$; yet no overlap with $a = 1$)

- Conclusions about **economies of scale are confirmed** but are more pronounced (i.e., slopes depart more strongly from $a = 1$)
Discussion: Factor & Regression Analysis

Single factor

- High multicollinearity with a single factor explaining 83% of the variance
  - Not surprising as the 16 indicators are, to a varying extent, intrinsically linked with the GERD of countries.

Economies of scale with respect to human resources (i.e., employment in tech & KIS [country level] as well as researchers in the HES [NUTS2 level])

- Potential mechanisms for explaining the increased productivity of human capital as a country’s or NUTS2 region’s pool of human resources increases include, for example:
  - the diversification and sharing of complementary expertise and competencies;
  - the specialisation and division of labour.
Discussion: Regression Analysis

Diminishing returns with respect to expenditures (i.e., BERD, GOVERD, VCI indicators [country and NUTS2 level] and HERD [NUTS2 level])

- Diminishing returns in terms of publication output were strongest for BERD (and VCI indicators) followed by GOVERD and HERD;
  - This is not surprising, as the tradition to publish results is strongest in the academic sector, followed by the government and private sectors.

- A potential mechanism for explaining the observed reduction in the productivity of countries/regions in terms of publication per euro investment in R&D would be that the number of researchers within entities does not increase as rapidly as their financial resources.
Discussion: Regression Analysis

Diminishing returns with respect to expenditures (i.e., BERD, GOVERD, VCI indicators [country and NUTS2 level] and HERD [NUTS2 level])

Robust $R^2 = 0.65$
Slope = 0.57
95% CI = [0.52 - 0.63]

The population of researchers in the higher education sector scale less rapidly than HERD at the NUTS2 level; as financial resources increase, the maximum production capacity of researchers would be reached.
Discussion: Regression Analysis

Policy implications for increasing productivity

- Findings support a rationale for awarding smaller grants to a larger population of researchers to increase the productivity of a given entity.

- However, research teams operating on larger budgets are more likely to carry out projects that could not be conducted with fewer resources.

- Although the cost of papers from such projects likely exceeds that of papers produced by less expensive projects, they impact a much larger community; thus, entities with larger R&D expenditures might have a higher impact per euro investment in R&D.

- A previous study found that human capital carries more weight in terms of the quantity of academic research, whereas capital accumulation plays a more important role in the impact of academic research (Hung, Lee, and Tsai, 2009).
Discussion: Regression Analysis

Policy Implications for increasing productivity

Robust $R^2 = 0.90$
Slope = 1.05
95% CI = [1.01 - 1.09]

Economies of scale in terms of citations per unit investment in R&D in the higher education sector
Discussion: Regression Analysis

Policy implications for increasing productivity

- Need to account for the complex interplay between the factors that contribute to the efficiency of countries/regions such as funding schemes, disciplinary portfolios, and development stages (Leydesdorff and Wagner, 2009).

- For instance, it was shown that the average cost per publication was higher in the basic medical sciences compared to the humanities (Archambault and Larivière, 2010)
  - Thus, if a larger share of a country’s R&D budget is allocated to the humanities, a country might exhibit stronger productivity in terms of publications per dollar investment in R&D but lesser productivity in terms of received citations per dollar investment in R&D than another country.
Thank you for your time and feedback

CONTACT INFO
David Campbell
Senior Research Analyst | Science-Metrix
514-495-6505 x122
david.campbell@science-metrix.com

ACKNOWLEDGEMENTS
Science-Metrix
Éric Archambault
Julie Caruso
Grégoire Côté

DG Research & Innovation
Carmen Marcus
Matthieu Delescluse

Thanks to the Science-Metrix Team!

This is a study designed, commissioned and owned by the European Commission
(Contract No. RTD/DirC/2009/S 158-229751)

Questions?

Science-Metrix
1335, Mont-Royal E.
Montreal, Quebec H2J 1Y6
Telephone: 514-495-6505
Fax: 514-495-6523
E-mail: info@science-metrix.com

WEB SITE
www.science-metrix.com
References


