Appendices

Our time-series cross-sectional-data (TSCS) capture the total population of trademark filings in Norway. The data are arranged as cross-sections. These groupings or 'clusters' have two dimensions that are fundamental to our strategy: (i) the geographical level and (ii) the industrial sector level. The geographical level consists of two levels: The (NUTS3) level of the county (see Appendix A) and, nested within these, the four centrality regions introduced above (i.e. metropolitan Oslo region, second cities, rural towns, and peripheral areas). Eight industrial categories are in turn nested in this regional grouping. We report the aggregates of industry in terms of 3 broad sectors because they capture what is going on (see Appendix A). The data span the period of 2008-2017. The TSCS data are fitted in Stata using a basic specification of a multi-level mixed-effects negative binomial model using the menbreg command (see Appendix B).

Appendix A. Aggregations

Industries are aggregated into three broad sectors:

1. Manufacturing & production
    a. primary industries, except petroleum
    b. petroleum extraction
    c. manufacture
2. Services
    a. Private-sector Services
3. Other industries
    a. Utilities
    b. Construction
    c. Education, health & other public services

The 19 Norwegian counties (per 2017) are:

1. Østfold
2. Akershus
3. Oslo
4. Hedmark
5. Oppland
6. Buskerud
7. Vestfold
8. Telemark
9. Aust-Agder
10. Vest-Agder
11. Rogaland
12. Hordaland
13. Sogn Og Fjordane
14. Møre Og Romsdal
15. Sør-Trøndelag
16. Nord-Trøndelag
17. Nordland
18. Troms
19. Finnmark

Appendix B. The approach and model

The TSCS data are fitted using a multilevel mixed-effects negative binomial regression model (or menbreg). Details of the menbreg in Stata can be found here:
https://www.stata.com/manuals/memenbreg.pdf.

The menbreg is a log-linear model for the mean of the distribution of y, where

y= domestic trademarks (i,å,r)/unit GDC (i,å,R).

And where:

y*= domestic trademarks
i= industry (3 general sectors)
å= year (2008-2017)
r= region (19 counties)
and
z= GCP (GDP at the level of region and industry).

It takes the following form, where GCP (z) is 'offset' in the model.

$$\ln(y*) = \ln(z) + \beta \ln(x) + \delta i + \varepsilon \quad (1)$$

This transforms to:

$$\ln \frac{y*}{z} = \ln(y) = \alpha + \beta \ln(x) + \delta i + \varepsilon \quad (2)$$

The response variable is thus the rate of trademarks per million GDP at the level of industry and region for a given year.

A negative binomial model is used due to the count nature of our response variable (trademark filings: overdispersion is confirmed). We have repeated measures across 2008-2017.

At the level of regions, we in other words test the effect of key variables on our response variable for each 'group'. In doing so, we want to ascertain whether (how) these effects differ in different regions. We do this because we assume a clustering effect, i.e. that firms localized within a 'group' will be more similar to each other than between them. A test confirms this hypothesis. The "multi-level" models the nesting of clusters: industrial sectors inside centrality inside counties.

The model is used to better understand factors that explain the increase (decrease) in the intensity of trademarking (the number of trademarks/unit GDP) in a given sector in a given region from one year to the next. Here we are interested in shedding light on four effects. We therefore calculate the effect (d) associated with the category (i) of selected categorical covariates (e.g. centrality, sectors, year) on the response variable (trademark 'intensity').

The approach therefore associates changes in the underlying variety of the regional economy for each industry (e.g. in the number of firms, the number of new firms, employees, turnover, etc.) with the diversification/differentiation activities measured by changes in the intensity of trademarking. Does the intensity of trademarking increase if the magnitude of firms (employees, turnover) in manufacturing goes up by 1% in the region? Does this rate differ for other sectors (e.g. Services)? What is the differential between in urban areas and peripheral areas?

The approach is of course distinct to the specific measures in the expanding regional diversification literature. And it is important to emphasize that the set-up is not intended as a separate measure to compete with these measures in the EEG literature (entropy measures of IPC classes in the case of patents, for example). It is intended as a proof of concept that trademarks can be used as a good measure of regional diversification which can complement the other measures.