Innovation, technological variety and income inequality : evidence from EU regions

Innovation, Technological Variety and Income Inequality: Evidence from European regions

Muhammad Yorga Permana*
Carolina Castaldi*
Önder Nomaler*
Floor Alkemade*

*School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, the Netherlands

Paper submitted to the EMAEE conference, Strasbourg, 2017

Work in progress: please do not cite/quote without permission

1. Introduction

In recent years, the relationship between innovation and income inequality has been widely discussed. There is no doubt that innovation plays a key role for long term economic growth. Nevertheless, a concern then has been raised regarding how the benefits of innovation are distributed: whether they are distributed fairly to the entire society or only concentrated in a relatively small number of individuals. If the latter occurred, then innovation may be held responsible for the rise of income inequality. The concern is supported by the fact that income inequality in most advanced countries has been increasing significantly starting from the last technological revolution of digital technology in the beginning of 1980s until now.

Several empirical studies have linked innovation and income inequality, mostly at the country level (Antonelli & Gehringer, 2013), but also at lower levels of aggregation, for instance the state level (Aghion, Akcigit, Bergeaud, Blundell, & Hémous, 2015), the European regional level (Lee, 2011) and United States cities (Florida, 2012). The majority of studies investigates the effect of innovation intensity, often captured by patent counts. However, no study so far has tackled the question of how technological variety in patenting may affect income distributions. Hartmann *et al* (2017) are analyzing the role of variety, but it is economic variety that they capture: they find that higher variety is associated with lower income inequality at the country level.

In this study, we investigate for the first time how regional variety in innovation activities affects regional income distribution: we expect technological variety to reduce income inequality, thanks to its positive role in reducing perverse specialization effects, which are suggested by a complementary literature on technological specialization (Cantwell & Vertova, 2004).

Our main research question is: do regions with higher technological variety experience lower income inequality?

We disentagle different forms of variety, by focusing on related and unrelated variety, as introduced by Frenken, Van Oort, and Verburg (2007). According to the intuition behind Jacobs externatlities, the variety of economic activities in a region provides a potential source of knowledge spillovers and positively impacts employment growth. Several studies have extended the study of the effect of variety into several economic performance indicators. Among them, Castaldi *et al* (2015) are the only authors considering the technological variety of innovation activities. We start from the intuition in that paper to hypothesize that high levels of unrelated, rather than related technological variety are not only beneficial for generating growth and technological breakthroughs, but also for restraining income inequality.

We envision three main contributions to current research.

Firstly, we extend the literature on the relation between innovation and income inequality by focusing on the qualitative nature of innovation specialization. Thereby we not only assess the impact of higher levels of innovation, but also the effect of innovation stemming from different technological fields vs innovation stemming from pure specialization.

Secondly, we contribute to the emerging literature on regional variety (Content & Frenken, 2016), by offering the first study focusing on technological variety and on its effect on income inequality. This contribution is likely to be relevant for questions of regional resilience (Boschma, 2015).

Thirdly, the results from this study can inform policymakers. Innovation policies are evolving from a traditional focus of promoting technological excellence to a focus on inclusive innovation and growth (OECD, 2017), where inclusiveness covers both places (laggard vs top regions) and people (high-skilled high-income vs low-skill low-income). Assessing which types of regional varieties are most associated to regional income inequality can offer warnings for policymakers about the (unintended) effects of promoting innovation specialization.

2. Theory

2.1 Innovation and income inequality

While there is no doubt that innovation and technological specialization are critical drivers of economic growth, concern then is being raised as to how the benefits of innovation-led growth are distributed in society. In brief there are several channels through which innovation is expected to raise inequality. One should note that these channels are not mutually exclusive. Instead, they might overlap in various ways. While the bulk of theories and evidence is at a country level, these channels can operate at a regional level as well and even assume more relevance there (Lee, 2011).

(1) Skill-biased technical change

Many scholars regard skill-biased technical change (SBTC) as the major cause of the rise in wage inequality (see Autor, Katz, and Krueger (1998), Acemoglu (1998), Acemoglu (2002) and Violante (2008) for examples). The theory suggests that technology may not affect all inputs equally but rather may be 'biased': benefit for some but detriment others. It is in contrast to traditional economic model (e.g Solow Neoclassical model) where technology acts as a neutral factor assuming that technology increases productivity for all labor (Violante, 2008). According to the SBTC hypothesis, the introduction of new technologies increases the demand for skilled labor relative to unskilled labor. The more recent version of this theory (Autor, Levy, and Murnane (2003) and Acemoglu and Autor (2011)) focuses on the nature of the tasks performed with the new technologies. Routine tasks such as clerical work, repetitive production, and monitoring jobs are easily codified by machines and consequently the labor input for those tasks decline. On the contrary, the non-routine tasks cannot be replaced by machines and rather complement new technologies. The implication is that the market labor will be polarized, with the middle skilled labor which work for routine jobs are replaced by machine while non-routine jobs at the low and high skills distribution have held up relatively well.

At a regional level, the effect of SBTC might be amplified by the fact that high-skilled workers tend to be more mobile and sort themselves towards more innovative regions, to benefit from higher skill-premiums there (Glaeser, Resseger, and Tobio (2009) and Lee (2011)).

(2) The superstar effect

The rise of total income inequality in last couple decades is going along with the steady increase of top income shares. Atkinson, Piketty, and Saez (2011) provide the long run history of top income inequality in twenty two countries suggesting that in most countries the significant portion of the increases is caused by an increase in top labor income. Indeed skill-biased technical change is unable to explain what is happening at the very top of the distribution, which tends to have very specific tail properties (Castaldi & Milaković, 2007). The superstars theory is introduced (see Rosen (1981), Kaplan and Rauh (2013), and Brynjolfsson and McAfee (2014)). The theory was initially proposed by Rosen (1981), describing the superstars phenomenon as the condition where *"relatively small numbers of people earn enormous amounts of money and dominate the activities in which they engage"*. Due to the rapid technological change at the beginning of 1980s, in particular those in ICT, gains of top performers are extended and pull away from those in the middle. Technological change allows superstars to apply their talent on a larger size of market, to reach larger number of people, and to be supported by greater pools of resources.

It should be noted that superstar effects are not only there in the economic returns of agents adopting new technologies, but also in the very process through which innovations emerge. Specifically, the strong uncertainty surrounding innovation has often been related to the highly skewed nature of innovation outcomes (Verspagen and Silverberg, 2003), where only a tiny share of patents is actually able to create economic value.

(3) Innovation, rents and wages

The role of innovation as a driver of top income inequality can be traced back to Schumpeter's view of growth. According to Schumpeter (1942), the disruptive force of innovation allows the entrepreneurs to enjoy some degree of monopoly rent that is called by Schumpeter (1942) as *"the prizes offered by capitalist society to the successful innovator"*. Market competition occurs because each firm is motivated by the prospect of monopoly rent. However, this monopoly rent is only temporary since it then will be destroyed by the next innovation. Nevertheless this monopoly rent is critical in the current discourse on top income inequality discourse, revived by recent models of Aghion *et al* (2015) and Jones and Kim (2014).

While top income inequality is affected by rents, total income inequality is instead mostly affected by wage dynamics (Autor, Skills, education, and the rise of earnings inequality among the "other 99 percent", 2014). The primary direct effect of innovation is a productivity effect: innovative companies become more productive thereby can pay higher wages to attract the better employees, typically high-skilled ones (Lee, 2011). This effect works towards increasing inequality. On the other hand, innovation also has social returns via knowledge spillovers. These spillovers might counteract the inequality effect of innovation, by stimulating learning for a wider range of economic actors than the innovator itself. There is ample evidence that these spillovers tend to be localized because of the importance of proximity in transfering the tacit components of knowledge (Boschma, Proximity and innovation: a critical assessment, 2005). Thereby, the regional level is a salient one to assess the impact of innovation on inequality.

Additionally, more innovative industries might also show more inequality because the uncertainty in outcomes translates into less regulated industries (Storper and Scott, 1990), where higher wage gaps are possible and jobs are less secure.

In line with the evidence so far, we expect the net effect of regional innovation to be positive. Our first hypothesis reproduces the research question posed first by Lee (2011) for EU regions.

H1: Regional innovation increases regional income inequality

2.2 Technological variety and inequality

Surprisingly, no previous studies has empirically or theoretically directly linked technological specialization and income inequality, despite the natural links between technological specialization, growth and innovation. Cantwell and Vertova (2004) which explore the historical evolution of technological diversification in a 100-year period from 1890 to 1990. The study found that initially more innovative countries tended to be technologically diversified whereas less innovative countries tended to be more specialized. Interestingly, they found the common tendency that all countries increasingly specialize their innovation activities in most recent period. Hartmann *et al* (2017) investigate economic variety of countries and find that more complex economies also show significantly lowers levels of income inequality. They argue that in more diversified economies individuals have more occupational choices and more learning opportunities.

We elaborate on two main channels through which the relation between technological specialization and income inequality is constructed: (1) differences in the demand for skilled labor *between* sectors and (2) differences in the wage premium across skill levels *within* the particular sector. Those mechanism are not mutually exclusive, instead, they complement each other shaping the income distribution.

(1)Technological specialization and inter-industry wage differentials

Many previous studies suggest 'inter-industry wage differential' as the vital contributor of the rise of total income inequality. Thus, to understand how technological specialization affects total income inequality, one has to consider the concept of 'inter-industry wage differential' which refers to the dispersion of wages across sectors. Consider two employees with the same skill and socio-economic characteristics (e.g same in terms of experiences, degree of education, regions, etc) work in different sectors. Hypothetically speaking, in the long run wages are supposed to converge. A basic equilibrium theory could explain why it does so. Initially, when wages between them differ, labor in low wage sector will attempt to move to high wage sector for seeking a new opportunity. This then increases labor supply in high wage sector along with the decrease of labor supply in low wage sector. In equilibrium, it leads to the equalizing level of wages between both sectors.

However, the claim is not followed by empirical facts. Krueger and Summers (1988) for instance investigate the wages differences for equally skilled workers across sectors in U.S. labor market in the period of 1974-1984. After controlled by human capital, demographic background, and working conditions, they found that wages dispersion across sectors is substantial and did not change significantly over time. For instance, they found that the average wages in the petroleum industry, which was indicated as high tech sector at that time, earned between 24 to 37 percent more than average equally skilled labor in all sectors. One should note that the inter-industry wages differences appear across levels of occupations. Osburn (2000) found that even janitors as well as managers commonly tend to receive similar wage differentials comparing to same occupation in other sectors.

The simple explanation for this phenomenon comes from standard competitive theory. It basically still argues that in the long run those wages will reach equilibrium. But in the short run, several factors may determine the differences. Temporary disequilibrium for instance is caused by asymmetrical information or due to the fact that job hunting is costly (Genre, Kohn, & Momferatou, 2011).

On the other hand, wage determination theory proposes different arguments. They rely on the main assumption that in particular sectors firms tend to pay higher wages than those suggested by equilibrium level due to the advantage of innovation or specialization (Genre, Kohn, & Momferatou, 2011). Osburn (2000) also supports this claim by providing empirical evidence that inter-industry wage differentials are positively associated with capital intensity. In other words, jobs in sector with high capital intensity, which representing high level of technology, tend to provide higher income comparing to the equal jobs in less technological sectors.

The explanation is related to the trade and growth theories which act as the connector between specialization and inequality. Since trade specialization and technology specialization co-evolve, high technological sectors tend to carry out higher exports. Higher exports then raise the relative price of goods in the sector as equilibrium is reached with the higher world price (Levinson, 2015). This higher relative price will stimulate demand for labor in the production sector, benefiting those labor relative to other sectors.

Another explanation regarding why wages are higher in industries with higher rates of technological change came from Bartel and Sicherman (1997). According to them, in sectors which use more sophisticated capital, typically more innovative, firms will increase their demand for workers who can more easily learn the new technology and adaptive to change. As the consequence, this sector will employ more skilled workers which then shifted from other sectors. However, higher demands do not push wages downward as accordance to basic equilibrium theory. In contrast, since the capacity of capital is positively correlated with productivity and growth, firms in more innovative sectors are able to pay those high skilled employees much higher than in other lagged sectors. In other words, skill premium in high technological sector is higher because the increase demand for high skilled labor which complements the innovation activities is along with the increase of growth and earnings.

This means that skill-biased technical change is not only about the gap between high and low skilled workers, but also refers to the shift of labor from such low tech to high tech environment (Bartel & Sicherman, 1997). Consequently, if countries tend to be more specialized in only few sectors, demands for high skilled labor will be asymmetric and it then lead to the widening gap of inter-industry wages differences. Reversely, if countries diversify their innovation activities into broad sectors, differences of inter-sectoral wages would be suppressed because the skill premiums are relatively symmetric and more equally distributed across sectors.

(2) Technological specialization and within-sector inequality

Unsurprisingly, as a country tends to be more specialized, it also widens the income gap within sectors particularly in those where innovation activities are more concentrated. Since, trade specialization and technology specialization co-evolve, following the fact that trade volume and growth differ across sectors, effects of specialization on the skill premium must be expected to vary from one sector to another. This hypothesis is supported by the fact that the extent of job polarization differs across industries, it is more obvious in some industries rather than others (Autor, Levy, & Murnane, 2003). For instance, Shim and Yang (2015) found that the decrease in the employment share of middle skilled labor with routine tasks is high in manufacturing, communication, and business related services while the decrease is much lower in transportation and retail trade.

Shim and Yang (2015) moreover found that 'inter-industry wage differentials' are the key source of the different levels of job polarization across sectors. Their finding proposes that in US labor market structure, the progress of job polarization between 1980 and 2009 was more noticeable in sectors that initially paid a high wage premium to workers than in sectors that did not. In other words, high technological sectors suffer a higher gap of inequality comparing to others. The explanation is because firms in a sector with a high wage premium seek alternative ways to minimize production costs by substituting middle skill workers who perform routine tasks with new technology. This is also supported by the evidence that sector with a high growth rate of ICT capital, as measured of technological changes, exhibit more significant job polarization (Michaels, Natraj, & Van Reenen, 2013), while (Levinson, 2015) suggest that trade in technology-intensive is only benefited by high skilled workers by examining trade-inequality relationship in 29 OECD countries through the scope of occupational wages.

Initially, in the first part of this section, we have discussed that technological specialization is correlated to trade specialization. Through the increase of volume of productions and the price of goods, particular sector will generate more money and thus supporting the total growth. In this part the direction to which the benefits are distributed is then elaborated. Apparently, beyond its growth, high technological sectors leave unintended consequences, namely job polarization. The growth of the cake is only benefited by high skilled workers while those low skilled workers have a tendency to suffer as they now face lower relative demand for their skills (Levinson, 2015).

H2: Regional technological variety decreases regional income inequality

2.3 (Un)related variety and income inequality

An emerging literature on Evolutionary Economic Geography has established that only looking at overall variety patterns of regions hides the effects of two underlying types of variety: related and unrelated variety. A region has high related variety if it specializes in fields (economic/technological) that are close to each other. This typifies a situation where regions can benefit most from knowledge spillovers given the cognitive proximity between their activities. Instead, regions with high unrelated variety specialize in fields that are far from each other. The opportunities for spillovers are limited, but the opportunities for innovation, learning and recombination are highest (Boschma, 2005; Castaldi et al, 2015).

Frenken *et al* (2007) already found that unrelated economic variety is negatively correlated to unemployment growth, thus it helps towards dampening the negative effects of external shocks. Instead related variety is positively related to economic growth.

Our hypothesis that variety has a negative correlation with income inequality is constructed based on two main theoretical mechanisms.

(1) Related variety and localized knowledge spillovers

Regions that specialize in a close set of fields have been found to benefit most from localized knowledge spillovers. On the other hand, Gambardella and Giarratana (2010) find evidence for the proposition that higher spillovers are associated with larger differences in skill premiums. This would then result in higher regional inequality by increasing the productivity gap between firms and the resulting wage gaps between workers. These results apply to the case of US cities. The case of European regions might be different given that that higher levels of regulations work towards reducing wage dispersion within industries.

(2) Unrelated variety and breakthroughs

Regions characterized by unrelated variety are expected to have a higher probability of recombining distant pieces of knowledge into breakthrough inventions (Castaldi *et al*, 2015). Breakthrough inventions have been found to impact firms, cities and regions in significant ways. Typically established positions are questioned, and new entrants and new industries appear. We contend that the 'reshuffling' effect that competence-destroying breakthroughs often imply is likely to work as a mechanism restraining income inequality growth. This mechanism exactly opposes all mechanisms mentioned in the previous section to clarify how specialization reinforces inequality instead.

Moreover Castaldi and Los (2017)show that the ability to produce breakthroughs is highly concentrated at the spatial level, even more than innovation in general is. This implies that this mechanisms is likely to critical at the regional level.

H3: Regional unrelated technological variety decreases income inequality more than related variety

3. **Empirical analysis**

3.1 Data and methods

The core of this empirical analysis is carried out at the European NUTS 1 regional level. The dataset covers the period 2004-2011 (8 years) consisting of 84 observation groups. Essentially, there are 98 regions in Europe based on NUTS 1 classifications. However, several regions are merged in order to adjust the data of dependent variable given from Ramos & Royuela (2014). For each country, we merged regions within Netherlands, Finland, and Portugal into one region which initially consists of 4, 2 and 3 NUTS 1 regions respectively. In addition, for the same data, only 5 of 16 regions in Germany are available. List of countries and regions included in this study is presented in **Appendix 1**.

The economic data for the control variables is obtained from Eurostat.

The patent data stems from REGPAT/Patstat. We assign patents to regions according the region of origin of the patent applicants. We choose the applicant instead of the inventor so that the location of patenting is closer to the location where the actual economic value is generated.

3.2 Dependent variables

We use the Gini coefficient as the standard measure of inequality for our baseline models. It ranges from a minimum value of zero, represents all individual income are absolutely equal, to a theoretical maximum value of one where all individual income is zero except one person. This is calculated from the Lorenz curve which represents the function of cumulative percentage of population to its cumulative income. The strong point of Gini index is due to its population symmetry (Lee, 2011). It means that the calculation from a sample of population is generalizable to the population as a whole (Coulter, 1989). Moreover, Gini index is directly comparable between units with different sizes of population (Hale, 2006).

To illustrate the relevance of the regional dimension, we should note that several countries have high variation of inequality between their regions. As an illustration, figure 1 demonstrates the differences of Gini index in a number of NUTS 1 regions within 4 countries. We see that while Spain and Sweden have the low variation of inequality across regions, in contrast, regions within United Kingdom and France have significantly high dispersion of inequality.

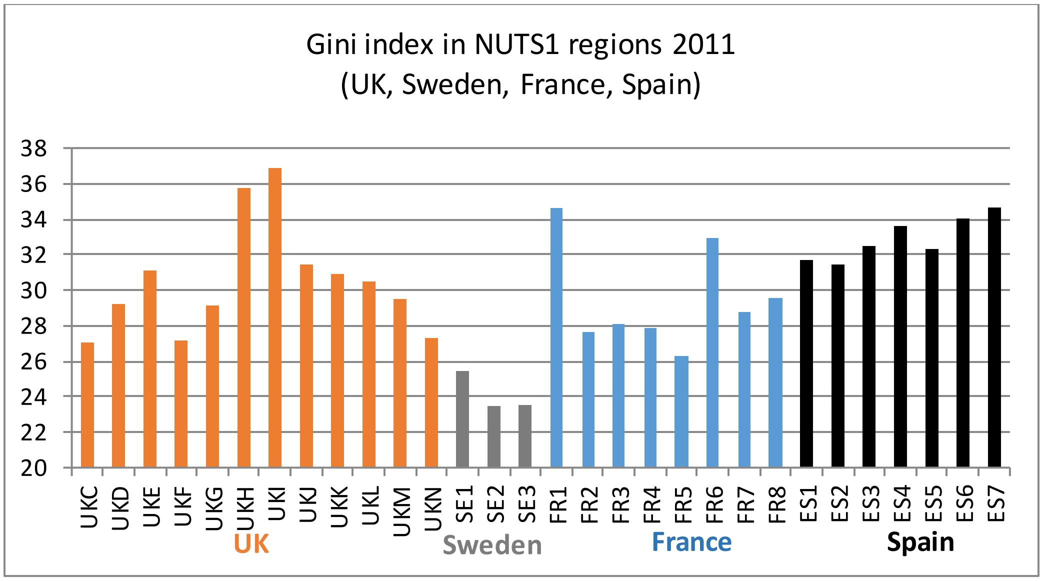

Figure 1 Gini index in NUTS1 regions 2011 (UK, Sweden, France, Spain)

Regions with higher level of innovativeness and prosperity such as London (UKI), East of England (UKH), Paris (FR1) and Centre Est (FR6) for instance, have very high levels of inequality as high as the inequality in emerging countries such as Bulgaria and Lithuania. In contrast, regions with lower level of innovativeness and prosperity such as East Midlands (UKF) and Sud-Ouest (FR5), have low levels of inequality and even to some extent close to the value of regions in Sweden.

Alternative measures of inequality in this study are ratios between income percentiles. The first one is 90-10 ratio which shows the unequal distribution between income at the top of distribution (more than 90th percentile) and at the bottom of the distribution (less than 10th percentile). The second one is 90-50 ratio, measuring the income differences in top half distribution, and the last one is 50-10 ratio which represents the inequality in bottom half of distribution. The data is also provided by Eurostat and Ramos-Royuela (2014).

3.3 Independent variables of theoretical interest

The main independent variables are innovation and technological variety.

This study uses EPO patent application statistics as the proxy of *innovation*. Despite known drawbacks, patent data are commonly used to capture innovation and technological specialization at the regional level. To deal with regional size, we use patents per million of total population, as provided by Eurostat.

Technological variety is constructed as an entropy measure, capturing the 'uncertainty' of the underlying probability distributions. We use the region of origin of the applicant of EPO patent applications as the basis of the entropy calculation and rely on the IPC codes classification for classifying patents into a hierarchical system of technological fields. The advantage of the entropy measure is that total variety can be decomposed into related and unrelated variety.

Total variety is calculated based upon the IPC codes of patents assigned to that region. We can calculate this total variety at different levels, namely based on 1, 3, 4, 8 of full digit IPC codes.

$$TV = \sum_{j=1}^{N} Pj \ln\left(\frac{1}{Pj}\right)$$

Unrelated variety is measured as the entropy of patents over unrelated IPC codes, which describes how diversified each region in their patenting activities across broader and more unrelated technological sections.

$$UV = \sum_{i=1}^{M} Pi \ln\left(\frac{1}{Pi}\right)$$

Instead, *Related Variety* (RV) expresses the diversity of patenting activities within sections in each region. Given the decomposition property of the entropy index, we calculate RV as the difference between total variety and unrelated Variety.

$$RV = \sum_{j=1}^{N} Pj \ln\left(\frac{1}{Pj}\right) - \sum_{i=1}^{M} Pi \ln\left(\frac{1}{Pi}\right)$$

Notice that related and unrelated variety are in principle orthogonal but in practice highly correlated. By including both measures in our analysis we are aiming at capturing which element is dominant in influencing regional income inequality.

The level at which one should measure related versus unrelated variety remains a matter of debate in the literature. In this study we explore several combinations, ranging from the full IPC codes to 1-digit IPC codes.

3.4 Control variables

Several control variables are included in the model to consider the presence of other potential explanatory variables associated with income inequality at the regional level. First of all, GDP per capita is included as the common control for the level of economic development of a region. We use current prices of GDP per capita which are adjusted by purchasing power parities in euro currencies, as provided by the Eurostat.

Additionally, population growth is included as control variable in line with similar studies by Aghion *et al* (2015) and Lee (2011). Population growth is associated with inequality by affecting the density of a region. Previous studies found a negative correlation between population density and inequality. Higher density areas tend to form more diverse and mobile societies which then provide more opportunities (Sylwester, 2004). Hence population growth is expected to decrease income inequality. The data is also available from Eurostat.

Educational attainment levels are also added. The first reason is because level of education is able to distinguish between low skilled and high skilled labor and thus it reflects the distribution of human capital among the population. Secondly, as argued by Goldin and Katz (2009), education and technological change intertwine in determining inequality. The basic idea is that technological changes frequently increase the demand of high skilled educated workers. To restrain the rise of inequality, the supply of high skilled workers must be steady. Hence they called it as the race between education and technology. If the race is won by education, inequality trends to decrease. Reversely, if the race is won by technology, inequality tends to increase. I then distinguish the variables for high and low educational attainment level of labor. Given by Eurostat, high educated workers are those adult population who reach tertiary education level (levels 5 to 8) whereas low educated workers are those adult population whose education level is less than secondary level of education (levels 0 to 2). *EduHigh* and *EduLow* represent the proportion of high educated and low educated labor respectively.

To sum up, all of the variables included in the model are demonstrated in Table 1. The descriptive statistics are also presented in **Table 2**.

Table 1 List of variables

Variable Names	Description	Source
	Measure of inequality	
Gini_rr	Gini index of inequality in regional level	Ramos&Royuela (2014)
S9010	Ratio of income between the richest 10% and the poorest 10%	
S9050	Ratio of income between the richest 10% and the median	Eurostat &
S5010	Ratio of income between the median the poorest 10%	Ramos&Royuela (2014)
	Measure of innovation	
Patent_pop	The number of patent application to the EPO per million population	Eurostat
	Measure of Technological variety	
Tech_variety_entropy	Index of technological variety calculated by entropy index	Calculated by authors
UV	Unrelated variety	
RV	Related variety	
	Control Variables	
GDPpcap	Real GDP per capita in Euro adjusted by Purchasing Power Parity	Eurostat
Popgrowth	Growth of total population	Eurostat
Edu_high	Population of working age with tertiary education degree	Eurostat
Edu_low	Population of working age with lower than secondary education degree	Eurostat

Table 2 Summary statistics at regional level 2003 – 2011

Variable	Obs	Mean	Std. Dev.	Min	Max
Dependent variables					
gini_rr	590	0.294	0.039	0.214	0.465
p9010_rr	589	3.818	0.853	2.537	7.425
p9050_rr	566	1.885	0.192	1.501	2.467
p5010_rr	566	2.001	0.279	1.584	3.455
Independent variables					
A. Innovation					
patent_pop	649	113.654	156.714	0.152	744.922
B. Technological variety					
entr_1d	649	2.402	0.436	0.000	2.851
entr_3d	649	4.271	1.078	0.000	5.816
entr_4d	649	5.408	1.557	0.000	7.693
entr_8d	649	6.509	2.077	0.000	10.016
entr_full	649	7.417	2.500	0.000	12.030
C. (Un)related variety					
UV	649	2.402	0.436	0.000	2.851
RV	649	3.006	1.234	0.000	4.978
Control variables					
edu_high	833	24.796	7.907	7.600	50.600
edu_low	833	27.341	14.499	3.800	80.200
gdp_cap	837	24734.850	10841.990	5100.000	72330.000
pop_growth	837	3.424612	7.317706	-28.9	30.9

3.5 Econometric strategy

We perform a series of panel data regressions to test our hypotheses. Panel data refers to multi-dimensional data on cross-section of countries and regions over time periods. It is selected as the model of analysis considering its advantages which have been proposed by Baltagi (2008). Firstly, panel studies are able to control individual heterogeneity while time-series and cross-section studies are not. Secondly, panel also provide *"more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency"* (Baltagi, 2008). As a final point, panel studies are better to study the dynamics of adjustment, while cross-sectional studies which capture only static duration of time are unable to perform a multitude of changes.

The models are as follows:

Model testing H1:

$$Y_{it} = \propto + \beta_1\ innovation_{i(t-h)} + \ \beta_2\ EduHigh_{it} + \beta_3\ EduLow_{it}\ +\ \beta_6\ GDPpcap_{it} + \beta_7\ Popgrowth_{it} + v_i + \varepsilon_{it}$$

Models testing H2

$$Y_{it} = \propto + \beta_0\ Technological\ Variety_{i(t-h)} + \ \beta_1\ innovation_{i(t-h)} + \ \beta_2\ EduHigh_{it} + \beta_3\ EduLow_{it}\beta_6\ GDPpcap_{it} + \beta_7\ Popgrowth_{it} + v_i + \varepsilon_{it}$$

Models testing H3

$$income\ inequality_{it} = \propto + related\ technology\ variety_{it-h} + unrelated\ technology\ variety_{it-h} + innovation_{it} + \ Edu_{High_{it}} + Edu_{Low\,it}\ +\ GDPpcap_{it} + Popgrowth_{it} + v_i + \varepsilon_{it}$$

where the *s*ubscript *h* in both independent variables represents time lag effect and y, v and ε are unobserved region specific characteristics which are time-invariant effects and idiosyncratic error terms, respectively.

Given the panel nature of the data, we use panel estimation techniques. Fixed effects models are designed to study the causes of change within a unit of analysis (Torres-Reyna, 2007). The Hausman test can be used to test whether the unique errors (idiosyncratic) are correlated with the regressors. If so, the fixed effect model fits better. This test suggests that fixed effects model is the appropriate method of estimation rather than random effects model in our case. Moreover, theoretical reasons justify the use of this approach, which is able to account for idiosyncratic, time-invariant properties of each region, in line with an evolutionary economic interpretation of regional growth as a path-dependent process.

The models include lagged variables for both innovation and technological variety. Since the alteration of income distribution is determined by the alteration of income itself, we account for the time lag with which patents can turn into actual innovations and generate economic value. Ken et al (2008) propose that there is a time lag of 4 to 5 years for patent to give impact to firm profitability in U.S pharmaceutical industry. Meanwhile, using panel analysis of Germany manufacturing industry, Ernst (2001) find patent applications affect sales increases with a time-lag of 2 to 3 years after the priority year. We opt for a 3 years lag after conducting several robustness checks (not reported here, available from the authors).

4. Results

Here we present our preliminary results, which we then compare to the key results in the empirical literature (Table 3).

Table 4 reports the estimation results to test H1 and H2. We find clear support for both hypotheses. H1, stemming from a large literature, is confirmed also in our dataset. It is noteworthy that Antonelli et al (2015) found evidence of a negative relation at the EU country level. Their result might be driven by the econometric strategy that they chose: FGLS models

In the case of H2, the effect of technological variety is significantly negative at all levels of aggregations considered. The dampening effect of variety, as opposed to specialization, is in line with our expectations from theory.

Table 5 investigates the effect of unrelated and related variety. While innovation intensity remains positively related to inequality, our expectation that unrelated variety is instead negatively related is confirmed. We also find that related variety has no significant effect, when conditioning for accounting for unrelated variety as well. The result is robust to different definitions of relatedness. This result suggests that the dampening effect of variety on inequality is in fact due to the unrelated component of total variety, not to its related component.

Table 6 provides a robustness check, using alternative measures of inequality.

In Table 3 we compare our results to the key papers that are tackling the innovation and inequality relation. This table clarifies that this study is the first one tackling the effect of technological variety in its different forms.

Table 3 Findings as compared to previous studies

	Methods & case study	Result
Lee (2011)	Fixed effect panel data model EU regions 1996 – 2001 (6 years)	Innovation increases total income inequality
Antonelli and Gehringer (2013)	FGLS panel data model EU & OECD countries 1996 – 2011 (16 years)	Innovation decreases total income inequality
Aghion et al (2015)	OLS with country and time dummies U.S. state 1975 – 2010 (35 years)	Innovation increases top income inequality but not related to total income inequality
Hartmann et al (2017)	Pooled OLS and fixed effect panel data models Countries worldwide 1996 – 2008 (12 years) and 1963-2008 (45 years)	Economic variety (complexity) limits income inequality
This study (2017)	Fixed effect panel data model EU regions 2004-2011 (8 years)	Innovation increases income inequality Technological variety decreases, but mostly because of the effect of unrelated technological variety

Table 4 Innovation and technological variety (measured by entropy) versus Gini index (based on Region Origin of EPO Patent Applicant)

	(1) Gini Innovation	(2) Gini Entropy 1 digit	(3) Gini Entropy 3 digit	(4) Gini Entropy 4 digit	(5) Gini Entropy 8 digit	(3) Gini Entropy full digit
patent_pop	0.004**	0.004**	0.004**	0.004**	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Entropy_1D		-0.705***				
		(0.265)				
Entropy_3D			-0.425**			
			(0.178)			
Entropy_4D				-0.432***		
				(0.166)		
Entropy_8D					-0.396**	
					(0.162)	
Entropy_fullD						-0.455***
						(0.149)
edu_low	-0.073	-0.061	-0.065	-0.063	-0.066	-0.061
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
edu_high	-0.171**	-0.144*	-0.145*	-0.137*	-0.146*	-0.143*
	(0.075)	(0.075)	(0.075)	(0.076)	(0.075)	(0.075)
gdp_cap	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pop_growth	-0.107***	-0.109***	-0.106***	-0.107***	-0.108***	-0.110***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
R-square	0.0770	0.0927	0.0896	0.0920	0.0903	0.0975
N	481	481	481	481	481	481

Innovation (as measured by *patent_pop*) and entropy index are 3 years lagged.

Time span: 2003-2011 (9 years). *** *p value* < 0.01, ** *p value* < 0.05, **p value* < 0.1. Standard errors are in the brackets.

Table 5 Unrelated and Related Variety versus Gini Index at regional level (based on Region Origin of EPO Patent Applicant)

	(1) Gini D3&DFull	(2) Gini D4&DFull	(3) Gini D3&D8	(4) Gini D4&D8	(5) Gini D1&D4
patent_pop	0.004**	0.004**	0.004**	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
UV	-0.479***	-0.466***	-0.447**	-0.436***	-0.654**
	(0.180)	(0.167)	(0.180)	(0.168)	(0.271)
RV	-0.418*	-0.424	-0.243	-0.069	-0.225
	(0.219)	(0.267)	(0.287)	(0.400)	(0.259)
edu_low	-0.061	-0.061	-0.065	-0.063	-0.061
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)
edu_high	-0.141*	-0.142*	-0.143*	-0.138*	-0.137*
	(0.075)	(0.075)	(0.075)	(0.076)	(0.076)
gdp_cap	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pop_growth	-0.110***	-0.110***	-0.107***	-0.107***	-0.108***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
R-square	0.0976	0.0976	0.0912	0.0921	0.0944
N	481	481	481	481	481

Innovation (as measured by *patent_pop*), UV, and RV are 3 years lagged.

Time span: 2003-2011 (9 years). *** *p value* < 0.01, ** *p value* < 0.05, **p value* < 0.1. Standard errors are in the brackets.

1. Column 1 : Unrelated Variety is 3 Digit IPC Code, Related Variety is Full Digit IPC Code
2. Column 2: Unrelated Variety is 4 Digit IPC Code, Related Variety is Full Digit IPC Code
3. Column 3: Unrelated Variety is 3 Digit IPC Code, Related Variety is 8 Digit IPC Code
4. Column 4: Unrelated Variety is 4 Digit IPC Code, Related Variety is 8 Digit IPC Code
5. Column 5: Unrelated Variety is 1 Digit IPC Code, Related Variety is 4 Digit IPC Code

Table 6 **Robustness Check: Unrelated and Related Variety versus Alternative measures of income inequality at regional level**

	(1) Gini D3&DFull	(2) 90:10 ratio D3&DFull	(3) 90:50 ratio D3&DFull	(4) 50:10 ratio D3&DFull
patent_pop	0.004**	0.001**	0.000**	0.000
	(0.002)	(0.000)	(0.000)	(0.000)
UV	-0.479***	-0.104***	-0.017**	-0.027*
	(0.180)	(0.038)	(0.009)	(0.015)
RV	-0.418*	-0.059	-0.024**	-0.000
	(0.219)	(0.046)	(0.010)	(0.018)
edu_low	-0.061	-0.042***	-0.005	-0.014***
	(0.060)	(0.013)	(0.003)	(0.005)
edu_high	-0.141*	-0.054***	-0.004	-0.017***
	(0.075)	(0.016)	(0.004)	(0.006)
gdp_cap	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
pop_growth	-0.110***	-0.036***	-0.006***	-0.013***
	(0.023)	(0.005)	(0.001)	(0.002)
R-square	0.0976	0.1933	0.1194	0.1657
N	481	481	461	461

Innovation (as measured by *patent_pop*), UV, and RV are 3 years lagged.

Time span: 2003-2011 (9 years). *** *p value* < 0.01, ** *p value* < 0.05, **p value* < 0.1. Standard errors are in the brackets.

Unrelated Variety is 3 Digit IPC Code, Related Variety is Full Digit IPC Code

1. Column 1: Dependent Variable is Gini Index
2. Column 2: Dependent Variable is Ratio Income of top 10% and bottom 10%
3. Column 3: Dependent Variable is Ratio Income of top 10% and median
4. Column 4: Dependent Variable is Ratio Income of median and bottom 10%

5. Conclusions

The preliminary results show a negative correlations between unrelated and income inequality at regional level of EU. Regions tend to have lower levels of income inequality if they have high level of variation in their innovation activities. In other words, diversifying innovation activities into broad sectors is better in restraining the income gap.

Our study is the first one which provides the empirical evidence linking technological variety and income inequality. Many previous studies focus on how the degree of innovation activities affects income inequality. However, none of them discusses the impact of different forms of regional technological diversification.

Our results have important policy implications. This study suggests that high variety in innovation activities would help regions to restrain income inequality. This casts doubts on (innovation) policy initiatives focusing on a few strong sectors in the economy. Instead, our results suggest that cultivating variety might also result in restraining income inequality.

References

Acemoglu, D. (1998). Why Do New Technologies Complement Skills: Directed Technical Change and Wage Inequality. *Quarterly Journal of Economics, 113*, 1055-1089.

Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature, 40*, 7-72.

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics 4*, 1043-1171.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., & Hémous, D. (2015). *Innovation and top income inequality.* (No. w21247). National Bureau of Economic Research.

Antonelli, C., & Gehringer, A. (2013). *Innovation and income inequality.* No. 201324. University of Turin.

Atkinson, A., Piketty, T., & Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature 49.1*, 3-71.

Autor, D. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science, 344(6186)*, 843-851.

Autor, D., Katz, L., & Krueger, A. (1998). Computing Inequality: Have Computers Changed the Labor Market? *Quarterly Journal of Economics 113*, 1169-1213.

Autor, D., Levy, F., & Murnane, R. (2003). The Skill Content of Recent Technical Change: An Empirical Exploration. *Quarterly Journal of Economics 118*, 1279-1334.

Baltagi, B. (2008). *Econometric analysis of panel data.* John Wiley & Sons.

Bartel, A., & Sicherman, N. (1997). Technological change and wages: an inter-industry analysis. *No. w5941. National Bureau of Economic Research*.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies, 39(1)*, 61-74.

Boschma, R. (2015). Towards an evolutionary perspective on regional resilience. *Regional Studies, 49(5)*, 733-751.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & Company.

Cantwell, J., & Vertova, G. (2004). Historical evolution of technological diversification. *Research Policy 33*, 511 - 529.

Castaldi, C., & Los, B. (2017). Geographical patterns in US inventive activity 1977–1998: The "regional inversion" was underestimated. *Research Policy*.

Castaldi, C., & Milaković, M. (2007). Turnover activity in wealth portfolios. *Journal of Economic Behavior & Organization, 63(3)*, 537-552.

Castaldi, C., Frenken, K., & Los, B. (2015). Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. *Regional Studies 49, no. 5*, 767-781.

Content, J., & Frenken, K. (2016). Related variety and economic development: a literature review. *European Planning Studies, 24(12)*, 2097-2112.

Coulter, P. B. (1989). *Measuring inequality: A methodological handbook.* Westview Press.

Ernst, H. (2001). Patent applications and subsequent changes of performance: evidence form time-series cross-section analyses on the firm level. *Research Policy, Vol.30*, 143-157.

Florida, R. (2012). *The Rise of the Creative Class--Revisited.* Basic books.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies 41.5*, 685-697.

Gambardella, A., & Giarratana, M. S. (2010). Localized knowledge spillovers and skill-biased performance. *Strategic Entrepreneurship Journal, 4(4)*, 323-339.

Genre, V., Kohn, K., & Momferatou, D. (2011). Understanding inter-industry wage structures in the euro area. *Applied economics 43.11*, 1299-1313.

Glaeser, E. L., Resseger, M., & Tobio, K. (2009). Inequality in cities. *Journal of Regional Science, 49(4)*, 617-646.

Goldin, C. D., & Katz, L. F. (2009). *The race between education and technology.* Harvard University Press.

Hale, T. (2006). The theoretical basis of popular inequality measures. *University of Texas working paper*.

Hartmann, D., Guevara, M., Jara-Figueroa, C., Aristara, M., & Hidalgo, C. (2017). Linking economic complexity, Institutions and Income Inequality. *World Development, 93*, 75-93.

Kaplan, S., & Rauh, J. (2013). It's the market: The broad-based rise in the return to top talent. *The Journal of Economic Perspectives 27.3*, 35-55.

Ken, Y., Tsai, T., & Ou, Y. (2008). Study of the time lag effect of patent impact on profitability of US pharmaceutical industry from innovation to profit. *PICMET'08-2008 Portland International Conference on Management of Engineering & Technology* (pp. 2588-2596). IEEE.

Krueger, A. B., & Summers, L. (1988). Efficiency wages and the inter-industry wage structure. *Econometrica: Journal of the Econometric Society*, 259-293.

Lee, N. (2011). Are innovative regions more unequal? Evidence from Europe. *Environment and Planning C: Government and Policy 29(1)*, 2-23.

Levinson, T. (2015). Trade Specialization & the Occupational Wage Distribution: Evidence from OECD Countries. *Working papers*.

Michaels, G., Natraj, A., & Van Reenen, J. (2013). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 years. *Review of Economics and Statistics 96(1)*, 60-77.

OECD. (2017). *Making innovation benefit all: policies for inclusive growth.* Paris: OECD.

Osburn, J. (2000). Interindustry wage differentials: patterns and possible sources. *Monthly Lab. Rev. 123*, 34.

Ramos, R., & Royuela, V. (2014). Income inequality in Europe. Analysis of recent trends at the regional level . *University of Barcelona, Research Institute of Applied Economics Working Paper No. 201425.*

Rosen, S. (1981). The Economics of Superstars. *American Economic Review 71(5)*, 845-858.

Shim, M., & Yang, H. (2015). Interindustry wage differentials, technology adoption, and job polarization. *Working paper Monash University Business and Economics*.

Sylwester, K. (2004). A note on geography, institutions, and income inequality. *Economics Letters 85.2*, 235-240.

Torres-Reyna, O. (2007). Retrieved from Panel Data Analysis Fixed and Random Effects using Stata (v. 4.2). Princeton University: https://www.princeton.edu/~otorres/Panel101.pdf

Violante, G. L. (2008). *Skill-biased technical change, .* New York: Palgrave Macmillan.

Appendix

Appendix 1 List of group observation at country and regional level

Table 1 List of EU countries

No	Country	Country code	EU-15	EU-27	EU-28	Included in the Model
1	Austria	AT	v	v	v	v
2	Belgium	BE	v	v	v	v
3	Bulgaria	BG		v	v	v
4	Cyprus	CY		v	v	v
5	Czech Republic	CZ		v	v	v
6	Denmark	DK	v	v	v	v
7	Estonia	EE		v	v	v
8	Finland	FI	v	v	v	v
9	France	FR	v	v	v	v
10	Germany	DE	v	v	v	v
11	Greece	EL	v	v	v	v
12	Hungary	HU		v	v	v
13	Ireland	IE		v	v	v
14	Italy	IT	v	v	v	v
15	Latvia	LV		v	v	v
16	Lithuania	LT		v	v	v
17	Luxembourg	LU	v	v	v	v
18	Malta	MT		v	v	v
19	Netherlands	NL	v	v	v	v
20	Poland	PL		v	v	v
21	Portugal	PT	v	v	v	v
22	Romania	RO		v	v	v
23	Slovakia	SK		v	v	v
24	Slovenia	SI		v	v	v
25	Spain	ES	v	v	v	v

26	Sweden	SE	v	v	v	v
27	United Kingdom	UK	v	v	v	v
28	Iceland	IS				v
29	Norway	NO				v
30	Switzerland	CH				v
31	Turkey	TR				v
32	Croatia	HR			v	

Table 2 List of EU NUTS 1 regions

No	Region code	Region	EU member	Included in the model
1	AT1	Ostösterreich	v	v
2	AT2	Südösterreich	v	v
3	AT3	Westösterreich	v	v
4	BE1	Région de Bruxelles-Capitale	v	v
5	BE2	Vlaams Gewest	v	v
6	BE3	Région wallonne	v	v
7	BG3	Severna i yugoiztochna Bulgaria	v	v
8	BG4	Yugozapadna i yuzhna tsentralna Bulgaria	v	v
9	CY0	Kypros	v	v
10	CZ0	Ceská republika	v	v
11	DE1	Baden-Württemberg	v	v
12	DE2	Bayern	v	v
13	DE3	Berlin	v	v
14	DE4	Brandenburg	v	v
15	DE5	Bremen	v	v
16	DE6	Hamburg	v	v
17	DE7	Hessen	v	v
18	DE8	Mecklenburg-Vorpommern	v	
19	DE9	Niedersachsen	v	
20	DEA	Nordrhein-Westfalen	v	
21	DEB	Rheinland-Pfalz	v	
22	DEC	Saarland	v	
23	DED	Sachsen	v	
24	DEE	Sachsen-Anhalt	v	
25	DEF	Schleswig-Holstein	v	

26	DEG	Thüringen	v	
27	DK0	Danmark	v	v
28	EE0	Eesti	v	v
29	EL1	Voreia Ellada (NUTS 2010)	v	v
30	EL2	Kentriki Ellada (NUTS 2010)	v	v
31	EL3	Attiki	v	v
32	EL4	Nisia Aigaiou, Kriti	v	v
33	ES1	Noroeste (ES)	v	v
34	ES2	Noreste (ES)	v	v
35	ES3	Comunidad de Madrid	v	v
36	ES4	Centro (ES)	v	v
37	ES5	Este (ES)	v	v
38	ES6	Sur (ES)	v	v
39	ES7	Canarias (ES)	v	v
40	FI1	Manner-Suomi	v	v (merged)
41	FI2	Åland	v	
42	FR1	Île de France	v	v
43	FR2	Bassin Parisien	v	v
44	FR3	Nord - Pas-de-Calais	v	v
45	FR4	Est (FR)	v	v
46	FR5	Ouest (FR)	v	v
47	FR6	Sud-Ouest (FR)	v	v
48	FR7	Centre-Est (FR)	v	v
49	FR8	Méditerranée	v	v
50	HR0	Hrvatska	v	v
51	HU1	Közép-Magyarország	v	v
52	HU2	Dunántúl	v	v
53	HU3	Alföld és Észak	v	v

54	IE0	Éire/Ireland	v	v
55	ITC	Nord-Ovest	v	v
56	ITF	Sud	v	v
57	ITG	Isole	v	v
58	ITH	Nord-Est	v	v
59	ITI	Centro (IT)	v	v
60	LT0	Lietuva	v	v
61	LU0	Luxembourg	v	v
62	LV0	Latvija	v	v
63	MT0	Malta	v	v
64	NL1	Noord-Nederland	v	v (merged)
65	NL2	Oost-Nederland	v	
66	NL3	West-Nederland	v	
67	NL4	Zuid-Nederland	v	
68	PL1	Region Centralny	v	v
69	PL2	Region Poludniowy	v	v
70	PL3	Region Wschodni	v	v
71	PL4	Region Pólnocno-Zachodni	v	v
72	PL5	Region Poludniowo-Zachodni	v	v
73	PL6	Region Pólnocny	v	v
74	PT1	Continente	v	v (merged)
75	PT2	Região Autónoma dos Açores (PT)	v	
76	PT3	Região Autónoma da Madeira (PT)	v	
77	RO1	Macroregiunea unu	v	v
78	RO2	Macroregiunea doi	v	v
79	RO3	Macroregiunea trei	v	v
80	RO4	Macroregiunea patru	v	v
81	SE1	Östra Sverige	v	v

82	SE2	Södra Sverige	v	v
83	SE3	Norra Sverige	v	v
84	SI0	Slovenija	v	v
85	SK0	Slovensko	v	v
86	UKC	North East (UK)	v	v
87	UKD	North West (UK)	v	v
88	UKE	Yorkshire and The Humber	v	v
89	UKF	East Midlands (UK)	v	v
90	UKG	West Midlands (UK)	v	v
91	UKH	East of England	v	v
92	UKI	London	v	v
93	UKJ	South East (UK)	v	v
94	UKK	South West (UK)	v	v
95	UKL	Wales	v	v
96	UKM	Scotland	v	v
97	UKN	Northern Ireland (UK)	v	v
98	IS0	Ísland		v
99	CH0	Schweiz/Suisse/Svizzera		v
100	NO0	Norge		v