Paper to be presented at

DRUID15, Rome, June 15-17, 2015

(Coorganized with LUISS)

Diversity and the technological impact of inventive activity: evidence for EU regions

Andrea Morescalchi
European Commission Joint Research Centre
Econometrics and Applied Statistics
andrea.morescalchi@gmail.com

Sjoerd Hardeman
European Commission Joint Research Centre
Econometrics and Applied Statistics
sjoerd.hardeman@jrc.ec.europa.eu

Abstract

Diversity has been considered as a prerequisite for turning prevailing technological trajectories into new and unexpected directions. However, little evidence exists on the exact nature of the more direct relationship between diversity and the impact of technologies. One main contribution of this paper is therefore to investigate the relationship between technological diversity and the impact of inventions across EU regions. Using EPO patent data, a set of measures is created considering different notions of diversity and different levels of technological aggregation, as allowed by the hierarchical structure of the International Patent Classification (IPC). The technological impact of inventions is captured by two citation-based indicators measuring an average and a high impact. For both measures we find that diversity is typically detrimental, or at best neutral, for the impact of new technologies, except when a very fine-grained technological detail is taken into account. However, in the latter case, nearly opposite results are found, namely, positive effects from related variety and, particularly for high technological impact, from combination of relatively distant technologies. Therefore, an important contribution of this paper is to show that these effects are very sensitive to the aggregation level used, and hence that policymakers should gain a very detailed understanding about the relations among technologies before implementing either specialization or diversification strategies.

Jelcodes:O32,-

Diversity and the technological impact of inventive activity: evidence for EU regions

Andrea Morescalchi and Sjoerd Hardeman

Econometrics and Applied Statistics Unit

Joint Research Centre

European Commission

Via E. Fermi 2749, 21027 Ispra (VA), Italy

Abstract: Diversity has been considered as a prerequisite for turning prevailing technological trajectories into new and unexpected directions. However, little evidence exists on the exact nature of the more direct relationship between diversity and the impact of technologies. One main contribution of this paper is therefore to investigate the relationship between technological diversity and the impact of inventions across EU regions. Using EPO patent data, a set of measures is created considering different notions of diversity and different levels of technological aggregation, as allowed by the hierarchical structure of the International Patent Classification (IPC). The technological impact of inventions is captured by two citation-based indicators measuring an average and a high impact. For both measures we find that diversity is typically detrimental, or at best neutral, for the impact of new technologies, except when a very fine-grained technological detail is taken into account. However, in the latter case, nearly opposite results are found, namely, positive effects from related variety and, particularly for high technological impact, from combination of relatively distant technologies. Therefore, an important contribution of this paper is to show that these effects are very sensitive to the aggregation level used, and hence that policymakers should gain a very detailed understanding about the relations among technologies before implementing either specialization or diversification strategies.

Keywords:
Smart specialization; related variety; unrelated variety; specialization; Jacobs externalities; localization; regional innovation systems; European regions.

1. Introduction

Technological Diversity potentially offers the seeds for turning existing technologies into new and unexpected directions and therewith renders major opportunities to un-lock prevailing technological trajectories (Dosi, 1982). It has been widely argued that diversity is key to research, innovation and economic performance (cf. Weitzman, 1998; Olsson, 2000). Acknowledging the potential of diversity, policymakers have called for "smart" policies intended to foster cross-fertilization among existing technologies (cf. European Commission, 2010; OECD, 2013).

However, and notwithstanding the emphasis often being put on the role of diversity in research and innovation, the exact meaning of the notion of diversity itself as well as its use within policy is often left in the midst (Stirling, 2007). First, from a theoretical point of view, different theories stress different aspects of the role played by diversity in steering research, innovation, and economic performance. Second, and with an eye on empirical studies in this field, although various measures of diversity exist, these are hardly ever compared simultaneously within the same analysis. The heterogeneity of diversity measures is further augmented by the use of different levels of technological aggregation in many empirical studies, rendering a comparison of the results of these studies virtually impossible (Beaudry & Schiffauerova, 2009). Therefore, a first objective of this paper is to introduce and compare different theoretical and measurement approaches to assessing the role of diversity in regional research and innovation.

What is more, whilst the relation between diversity and economic growth by now has been extensively addressed (cf. Glaeser et al., 1992; Frenken et al., 2007; Neffke et al., 2011; Van Oort et al., 2014), the more direct relationship between diversity and research and innovation has been somewhat neglected (Boschma, 2013). One issue here concerns the measurement of research and innovation itself and the interpretations attributed to such measurements. That is, whenever patent data are used in empirical studies, it is more appropriate to speak of inventive activity rather than research and innovation at large (Griliches, 1990). However, even when considering inventive activity only by focusing on patent data, still few studies take into account their technological impact and instead most often focus on their quantity only (for an exception see Castaldi et al., 2014). Therefore, a second objective of this paper is to investigate the effects of diversity on invention and in particular on its technological impact, as captured by patent citations.

We focus on the NUTS2 level of EU regions, as this is an important level of aggregation where smart research and innovation policies are implemented. Also, we use patent data to assess the relationship between diversity and technological impact, not only because these are commonly used in the literature, but also because these data allow assessing the relationship at different levels of technological aggregation. In fact, the hierarchical structure of the International Patent Classification (IPC) scheme allows constructing each of our measures for different technological levels of aggregation. Moreover, technological impact is captured by two citation-based indicators. The count of (field-normalized) citations is used to proxy the average technological impact, while the number of highly cited patents is used as a proxy for high technological impact.

The rest of this paper proceeds as follows. The next section offers an overview of the theoretical stances within the literature on the relationship between diversity and technological impact of regional invention. In section 3 we describe the data and methods that we use to address this relationship empirically, focusing in particular on the different ways of measuring diversity. Section 4 presents the results from our analysis, section 5 discusses our main findings and section 6 concludes.

2. Theoretical background

Viewing research and innovation in terms of a process of recombination (cf. Weitzman, 1998; Olsson, 2000) places the notion of technological diversity at the heart of the debate on regional invention (cf. Ejermo, 2005). However, different and to some extent even competing theories exist on the exact specification of the relationship between technological diversity and the technological impact of inventive activity across regions.

On the one hand there are those emphasizing the benefits that might accrue by putting together diverse activities from a limited set of technological backgrounds. Herein, two distinct arguments are important (Frenken et al., 2007); one focusing on specialization, the other on localization. Specialization is about the extent to which an actor, in our case a region, focusses on a few activities within a limited range of technologies only. The argument holds that specialization is beneficial when increases in the division of labor among distinct technologies allows for perfecting the activities being performed on these technologies. The concept of specialization focuses primarily on the activities themselves and, therewith, has no particular spatial connotation, meaning that it says nothing about the spatial distribution of those activities across regions. In contrast, the notion of localization focuses specifically on the spatial distribution of activities across regions and emphasizes the point that, when knowledge flows are geographically localized (Jaffe et al., 1993), benefits accrue due to the concentration of diverse activities that are concerned with the same technologies. In other words, it is not just the specialization of a region in a limited set of technologies that matters for the technological impact of regional invention but also whether those technologies are concentrated in a restricted number of regions only. Though distinct in form, specialization and localization are most often not taken into account separately but instead used interchangeably throughout the literature (cf. Van der Panne, 2004).

On the other hand there are those emphasizing the benefits that might accrue by putting together diverse activities covering a wide set of technologies; that is, diversification. In line with the literature on Jacobs externalities (cf. Glaeser et al., 1992), the emphasis here is on cross-fertilization among different technologies as a source of regional invention. Note then that the emphasis on cross-fertilization here mirrors Page's (2007) argument on "*diversity triumphing ability*" in that it is not so much about the narrow set of technologies that brings about invention rather than the combination of different technologies. Sometimes Jacobs externalities are equated with urbanization economies (cf. Van der Panne, 2004). However, whilst Jacobs externalities are about benefits that accrue due to the availability of a set of different technologies, urbanization economies are about the benefits that arise due to the sheer size of and population density in a region (Frenken et al., 2007). Again, as with the distinction between specialization and localization, the notion of diversification has no particular spatial connotation whilst the notion of urbanization clearly has.

Clearly, the literature stressing the positive effects of specialization and localization seems to be at odds with the literature stressing the positive effects of urbanization and especially diversification. Hence, it should come as no surprise that the effect of specialization and localization on the one hand and diversification and urbanization on the other have often been tested within a single empirical framework as to identify whether the one or the other theory is backed up by empirical evidence (cf. Van der Panne, 2004; Beaudry & Schiffauerova, 2009). Unfortunately, however, the evidence offered from such studies is mixed and, therewith, rather inconclusive (De Groot et al., 2009).

Three reasons can be identified for the evidence being mixed. First, different studies use different conceptualizations to capture specialization, localization, diversification, and urbanization (De Groot et al., 2009). A first way to resolve this issue is to disentangle concepts that have a spatial connotation from those that do not have a spatial connotation. Here, the notions of specialization and diversification with no spatial connotation clearly stand out from the notions of localization and urbanization which have a spatial connotation.

Second, focusing in turn only on specialization and diversification, both concepts have been conceptualized differently within the broader debate on diversity (cf. Palan, 2010; Wagner et al., 2011). A first conceptualization is proposed by Stirling (2007), who points out that diversity concepts employed across different disciplines exhibit some combination of the following three basic attributes of diversity: variety (versus uniformity), evenness (versus imbalance), and disparity (versus similarity). Variety is about the number of categories (i.e. disciplines, sectors, industries, or technologies) that characterize the basic unit of analysis (i.e. individuals, firms, regions, or countries). Within the context of regional invention, variety can thus be thought of in terms of the amount of different technologies available in a region. Evenness is about the distribution of categories characterizing the basic unit of analysis. Again applied to the context of regional invention the notion of evenness translates into the extent to which a number of technologies are equally available in the technological composition of a region. Disparity refers to the heterogeneity of the categories that characterize the basic unit of analysis. Note that in contrast to the notion of imbalance, the notion of disparity is relational and takes into account the distance among categories. Hence, regions characterized by technological disparity – or what can be called unrelated diversification (as opposed to related specialization) – combine distant technologies, for a given number and distribution of themselves. All else being equal, higher variety, higher evenness and higher disparity imply higher diversity, which in turn is expected to have a positive effect on the technological impact of inventive activity (Page, 2007, Stirling, 2007 and Yegros et al., 2013). All these different aspects of diversity need to be incorporated in a comprehensive assessment of the role of diversity in steering the impact of new technologies.

Another conceptualization refers to what Frenken et al. (2007) call related and unrelated variety. Both referring to the basic notion of entropy, related variety is about entropy in the technological composition of regions at a very fine grained level of technological detail whilst unrelated variety is about entropy in the technological composition of regions at a relatively rough level of technological detail.[1] Note then that, like the notion of disparity, in

[1] Building upon the framework of Stirling (2007), the notion of entropy used by Frenken et al. (2007) combines the two attributes of variety and evenness (see Section 3.2).

using different levels of technological detail, the distinction between related and unrelated variety imposes some, albeit crude, notion of distance among technologies That is, based on the specific technological classification system used, unrelated variety refers to entropy among technologies that are more distant from each other whilst related variety refers to entropy among technologies that are closer to each other.

Recently, Castaldi et al. (2014) theorized that whilst related variety is likely to have a positive impact on the quantity of invention in general, unrelated variety raises the likelihood of breakthrough inventions. Using the number of patents as a measure of regional invention quantity and the share of highly cited patents as a measure of breakthrough invention, Castaldi et al. (2014) indeed find empirical evidence supporting their thesis for regional invention at the US state level. It has to be noted though, that the results from their analysis might be specific to the US context only. The main assumption underlying their analysis holds that highly cited patents can be equated with breakthrough inventions. However, whilst citation is generally considered to be an indicator of impact and highly-citedness as a measure of high-impact therewith, these measures in themselves say little to nothing about the nature of inventions being characterized as breakthrough (i.e. radical invention) or not. That is to say, although breakthrough patents are likely to be highly cited, the reverse need not be true: highly cited patents can well be about technologies reflecting incremental inventions. This remark is particularly important within the European context whereas it has been argued from the literature on Varieties of Capitalism (Hall & Soskice, 2001) that whilst liberal market economies like the US have a comparative advantage in radical invention, coordinated market economies like those of most EU member states have a comparative advantage in incremental invention (Boschma & Capone, 2014). In other words, whilst the positive effect of unrelated variety and lack of effect of related variety on the impact of regional invention might hold true for the US context, this might not necessarily be so for the EU context.

Finally, a third reason for why evidence on the relationship between diversity and technological impact is mixed is that different studies use different levels of technological aggregation to test this relationship (De Groot et al., 2009). Building on technological classification systems, it follows that diversity, either in the conceptualization proposed by Stirling (2007), or in the conceptualization proposed by Frenken et al. (2007), can be measured by grouping technologies in a more or less fine grained way. Using different levels of technological detail might be important, whereas at higher levels of technological aggregation it can be expected that regions are more alike, rendering various notions of diversity to have little predictive power for explaining the impact of new technologies.

In sum, we derive three main conclusions from the theoretical literature on the relation between diversity and regional invention. First, arguments can be made supporting opposite theses on the relation between diversity and regional invention. Going from the distinction between specialization and localization on the one hand and diversification and urbanization on the other, a plausible case can be made for both sets of theories.

Second, diversity can be conceptualized in different ways. Following Stirling (2007), it can be decomposed in terms of variety, evenness, and disparity. Herein, the distinction between related and unrelated variety takes an intermediate stance residing in between variety and evenness together on the one hand, and disparity on the other hand. In fact it is based on a decomposition of the notion of entropy whilst taking on board an artificial notion of distance among technologies; that is, one that has been imposed by the structure of technological classification systems. Of the two main conceptualizations, the one proposed by Stirling (2007) offers the most complete perspective on diversity as it takes explicitely into account the distance among technologies (disparity) alongside attributes of quantity (variety) and distributions (evenness).

Third, as technologies can be classified into different levels of technological detail, the exact specification of the relationship between diversity and technological impact might crucially depend on the specific level of technological detail considered. This issue applies to both conceptualization of diversity.

Overall, going primarily from theory and especially for the EU context, the exact nature of the relationship between diversity and regional invention is largely unclear thus far. In what follows, therefore, we will address this relationship empirically.

3. Data, variables, and methods

3.1. Data

The main source of data that we use is REGPAT (Maraut et al., 2008), a database maintained by the OECD collecting information on all patent applications filed with the European Patent Office (EPO). In particular, the data contain information on patent inventors, classes and citations. We construct regional indicators at NUTS2 level exploiting information on regions attached to the patent inventor list.[2] Regional indicators are based on the patent's priority year. The priority year is the year of first filing for a patent and hence it is the closest to the actual date of invention.

We focus on a balanced panel of EU27 regions over 1995-2009. In case of unbalanced panels, some regions may appear or disappear over the sample period causing attrition bias. Perfect balance of the panel is achieved by setting a cutoff c and selecting regions with at least c patents in every year of the selected period. We choose c = 2 trading-off between loss of regions (when c is high) and the risk of zeros or missing values for citation or technological indicators (when c is low).[3] This amounts to a balanced panel of 195 EU27 regions over 1995-2009, representing 22 countries. Regional patent indicators are merged to economic indicators drawn from the Cambridge Econometrics European Regional Database.

3.2. Dependent variables: measuring technological impact

Citation data are used to create measures of the technological impact of inventive activity. First, we use the total count of forward citations of EPO patents (received directly as EPO

[2] The region is based on the inventor's address, which indicates where the invention was made, such as typically the laboratory or research establishment, or the place of residence. In case of multiple inventors from the same NUTS2 region, the patent is counted only once.

[3] In case of very small number of patents, citations counts could be zero. In the estimation sample, 32 cases with zero citations are replaced with the minimum over across the entire sample.

publications, as PCT publications or as national offices publications[4]) as a measure of average impact of regional invention. An EPO publication can correspond to publications from different patenting authorities, covering the same invention; therefore, any version of it can be cited (see Webb et al., 2005). The total count of citations received by each patent is normalized by the average count of citations received by patents in the same field and year (see Appendix A for details). This normalization allows for different citation scales across technological fields (see Squicciarini et al., 2013). To avoid right truncation in forward citations, we consider only citations received by T years, where T is defined as the difference between publication dates of the citing and cited patent. We select $T = 3$ to minimize the loss of yearly observations. This choice necessarily leads to underestimate the impact of patents taking more time to receive citations. However, we checked that the correlation among regional citation counts considering different citation lags is very high. Moreover, remark that the field-normalization removes potential field bias in citation lags.[5]

As a second measure of the technological impact of regional invention, we use the count of highly cited patents. This measure represents the number of top 1% highly cited patents according to field normalized citations. In all, this measure captures a region's high-impact technological inventions. Overall, and in line with Tijssen (2002), whilst average technological impact as measured by field-normalized citation rates can be thought of as research quality, high technological impact as measured by the count of top 1% highly cited patents can be thought of as research excellence.

3.3. Independent variables: measuring technological diversity

We use the list of patent classes to create measures of technological diversity. The hierarchical structure of the International Patent Classification (IPC) is exploited to identify different levels of technological detail. Specifically, the IPC scheme separates the whole body of technological knowledge into the following five levels, in hierarchical descending order: the section (1st digit of the code), the class (first 3 digits), the subclass (first 4 digits), the group (first 10 digits) and the subgroup (the whole code). The first four levels are used in our analysis giving rise to the following self-explanatory labels: ipc1, ipc3, ipc4, ipc10. In the dataset we find respectively 8, 123, 633, and 7209 unique codes. The ipc1 level is very broad and captures very different technologies. The further we move towards ipc10, the more detailed technologies become. We also convert the IPC scheme into technological fields (tec1) and sub-fields (tec2) according to the Schmoch concordance table (Schmoch, 2008). tec1 and tec2 contain respectively 5 and 35 unique codes. Estimates based on these two levels are reported in Appendix B.

As argued before, the literature offers different conceptualizations to capture diversity, in all its attributes, in the organization of regional invention. In order to capture different conceptualizations and attributes of diversity simultaneously, we first employ the distinction between related and unrelated variety made by Frenken et al. (2007), and we

[4] Patent applications under the Patent Co-operation Treaty (PCT) are international applications giving options for future applications to other patent offices in the world, such as the EPO or national offices (see Webb et al., 2005).

[5] By way of example, after removing the within region average, the correlation between regional citation counts with 3-year lag and with no lag restrictions is 61%. Considering field-normalized citations, this correlation reaches 95%.

add, on the one hand, the framework proposed by Stirling (2007), and, on the other hand, two indices considering the spatial dimension of diversity.

First, the distinction between related and unrelated variety is operationalized by exploiting the decomposable nature of the Shannon entropy index (see Frenken et al. (2007)). Specifically, defining A and B as two levels of aggregation, where A is more aggregated than B, the B entropy is equal to the sum of the A entropy and the weighted sum of B entropies within each category of A. Unrelated variety is indicated by A entropy and related variety is indicated by the weighted sum of B entropies. Formally, defining P_k as the share of patents in a category k (k = 1… K) of A, and defining p_l as the share of patents in a category l (l = 1… L) of B (remark that $P_k = \sum_{l \in k} p_l$), we can define

$$uvariety = \sum_{k=1}^{K} P_k \log_2(1/P_k),$$

$$rvariety = \sum_{k=1}^{K} P_k H_k,$$

where

$$H_k = \sum_{l \in k} \frac{p_l}{P_k} \log_2\left(\frac{1}{p_l/P_k}\right).$$

Second, we define the indicators *evenness* and *disparity* as consistent with Stirling's (2007) conceptualization of diversity. Unlike the decomposition between related and unrelated variety, which imposes an artificial notion of distance among technologies, Striling's conceptualization offers the most complete perspective on diversity as it takes explicitly into account the distance among technologies (disparity), for given distribution (evenness) and number of technologies (variety). Note however that ultimately we do not include Stirling's (2007) attribute of variety as it turns out to be highly collinear, especially with the number of patents. In other words, though not necessarily so in theory, empirically the number of technologies is highly correlated with the total number of inventions made in a region.

As a measure of evenness, we use Shannon's entropy index of evenness or equitability (Stirling, 2007):

$$evenness = \left(-\sum_{k=1}^{K} p_k \ln p_k\right) \Big/ \ln K,$$

where K indicates the range of different technologies and p_k indicates the share of patents in technology k. Region and year subscripts are omitted to simplify notation. The standard Shannon entropy index does not contain the division by ln(K). However, we include this division as it removes the likely positive correlation between the Shannon entropy index and the size of a region in terms of the total number of patents. Note then that, in line with the theoretical premise that having different technologies equally available in a region is beneficial for a region's technological impact, this index measures the extent to which different technologies are evenly distributed in a region.

Disparity is defined as a weighted average of the distance among sectors according to the Greenberg-Rao diversity index (see Desmet et al. (2009)). Following Yegros et al. (2013) the distance is defined as 1 minus the similarity cosine index s_{kl}, constructed for any pair of technologies k and l within the region as:

$$s_{kl} = \frac{c_{kl}}{\sqrt{c_k c_l}},$$

where c_{kl} is the count of patents with co-occurrences of k and l, and c_k and c_l are the sum of co-occurrences involving l ($l \neq k$) and k ($k \neq l$), respectively. Remark that c_{kl}, c_k and c_l are defined within the region for every year. Then $(1 - s_{kl})$-s are computed and the resulting distance indicators are aggregated giving rise to

$$disparity = \sum_{k<l} \sum_{l=1}^{K} w_k w_l (1 - s_{kl}),$$

where

$$w_k = \frac{c_k}{\sum_{k=1}^{K} c_k},$$

and K is the number of technologies in the region.[6]

Third, following Frenken et al. (2007) we use the Los-index as a localization index and the log of a region's population to account for urbanization economies. The Los-index is a weighted average of similarity indices for all technology pairs with weights equal to the product of patent counts for the two technologies. Defining C_k and C_l as patent counts for technologies k and l we have

$$localization = \frac{\sum_{k<l} \sum_{l=1}^{K} C_k C_l s_{kl}}{\sum_{k<l} \sum_{l=1}^{K} C_k C_l},$$

where s_{kl} is computed here for every year pooling all regions available in the data set and hence is constant over regions.[7]

Finally, in addition to the afore proposed indicators, the following control variables are included in the specifications: *nocollab*, *collabwr*, *collaborwc*, *collabocweu27*, *collaboeu27*,

[6] Note that while Yegros et al. (2013) use an unweighted disparity index, following Desmet et al. (2009) we use the weighted form.

[7] Remark that $s_{kl} = s_{lk}$ and that s_{kk} is not considered in the index.

log(*gdppc*), log(*pop*), log(*npatgdp*). A set of collaboration share variables is included as collaboration in general (Jones et al., 2008) and international collaboration in particular is considered to render higher impact (Frenken et al., 2010). Specifically, we consider the share of patents with at least one collaboration intra-region (*collabwr*), outside-region but within-country (*collaborwc*), outside-country but within-EU27 (*collabocweu27*) or outside-EU27 (*collabocweu27*). Each patent is assigned to one specific collaboration type giving precedence to higher level collaborations.[8] Considering also the share of patents with only one inventor (*nocollab*), the sum of these shares is therefore equal to 1. One variable is omitted in regressions to avoid perfect multicollinearity. *gdppc* is Gross Domestic Product in per capita terms (millions of euro in 2005 prices per inhabitant) and controls economic cycle effects. *pop* stands for population (in thousands) and captures urbanization effect. As a proxy for propensity to invest in R&D we use the number of patents per unit of GDP, *npatgdp*. Data on regional R&D expenditure from Eurostat are not available before 2000 and have several missing cases afterwards. However, we find that *npat* is highly collinear to R&D expenditure for available data points. The summary statistics for all variables included in the analysis are presented in Table 1. In addition, Table 2 presents the ranking of regions according to the impact of their inventions.

3.4. Model specification

Model parameters are estimated by panel Fixed Effects (FE). The FE method removes the impact of constant variables such geography. In this context, another advantage of panel FE estimation is that multicollinearity in regressors can be importantly reduced by applying the within transformation (Hsiao, 2014). Potential threat of serial correlation in the error is handled by estimating a FE model with an AR(1) disturbance (see Baltagi and Li, 1991; for unbalanced panel see Baltagi and Wu, 1999). This method is a feasible Generalized Least Squares (GLS) estimator in which the error variance-covariance matrix is modeled according to the Prais-Winsten transformation.[9] The FE-AR(1) model is estimated here whenever AR(1) errors in the linear FE are found. The Wooldridge test is used to detect AR(1) in the errors (Wooldridge, 2010; Drukker, 2003).

When *highcit3* is used as outcome variable, a Fixed Effects Negative Binomial (FE-NB) model is estimated to take into account that *highcit3* is a count variable. In fact, linear models assuming normality in the errors can perform poorly for count variables (Cameron and Trivedi, 2013). A NB model is preferred here over a standard Poisson distribution to take into account over-dispersion. The Poisson distribution assumes equal mean and variance, while often count variables are over-dispersed, i.e. exhibit variance larger than the mean. The fixed-effects component applies here to the distribution of the over-dispersion parameter (see Cameron and Trivedi, 2013).

[8] For example, in case inventors from different EU countries(\regions) collaborate with one inventor from another non-EU(\EU) country, the patent counts as an outside-EU27(\outside-country within-EU27) collaboration.

[9] The error variance-covariance matrix is derived assuming AR(1) in the errors. A feasible version of the GLS estimator requires a preliminary estimation of the AR(1) coefficient. See Wooldridge (2010, sec. 10.5.5) for a textbook exposition of FE-GLS models.

Table 1 Summary statistics: Statistics are based on 195 EU27 NUTS2 regions over 1995-2009.

		1995-1999			2000-2004			2005-2009		
Variable	regions	mean	min	max	mean	min	max	mean	min	max
log(*ncit3*)	195	4.358	-0.280	7.951	4.861	-0.280	8.123	5.033	-0.280	8.023
highcit3	195	2.038	0	32	3.186	0.000	43	3.614	0	49
log(*gdppc*)	195	3.020	0.838	4.278	3.132	1.057	4.436	3.211	1.499	4.585
log(*npatgdp*)	195	1.084	-2.666	3.373	1.334	-2.678	3.403	1.391	-1.440	3.513
log(*pop*)	195	7.341	5.484	9.304	7.356	5.512	9.340	7.378	5.550	9.372
nocollab	195	0.287	0	1	0.253	0	1	0.239	0	0.750
collabwr	195	0.210	0	1	0.216	0	0.643	0.228	0	0.706
collaborwc	195	0.307	0	1	0.306	0	0.763	0.313	0	0.763
collabocweu27	195	0.110	0	1	0.124	0	1	0.122	0	0.833
collaboeu27	195	0.085	0	1	0.102	0	0.714	0.098	0	0.529
evenness_ipc1	195	0.826	0	1	0.841	0.494	0.999	0.860	0.618	0.977
evenness_ipc3	195	0.804	0	1	0.797	0.490	0.978	0.821	0.544	0.982
evenness_ipc4	195	0.871	0	1	0.864	0.590	0.988	0.884	0.629	0.995
evenness_ipc10	195	0.950	0	1	0.945	0.839	1	0.947	0.848	1
evenness_tec1	195	0.773	0	1	0.795	0	0.991	0.831	0	0.992
evenness_tec2	195	0.858	0	1	0.860	0.507	0.973	0.875	0.578	0.982
diversity_ipc1	195	0.219	0	0.440	0.243	0	0.441	0.231	0	0.439
diversity_ipc3	195	0.038	0	0.379	0.035	0	0.264	0.028	0	0.243
diversity_ipc4	195	0.028	0	0.333	0.020	0	0.282	0.014	0	0.208
diversity_ipc10	195	0.027	0	0.375	0.018	0	0.250	0.010	0.001	0.189
diversity_tec1	195	0.258	0	0.395	0.282	0	0.389	0.270	0	0.389
diversity_tec2	195	0.080	0	0.402	0.083	0	0.281	0.066	0	0.260
localization_ipc1	195	0.189	0.026	0.481	0.188	0.034	0.469	0.185	0.105	0.490
localization_ipc3	195	0.019	0.007	0.120	0.016	0.007	0.090	0.024	0.012	0.131
localization_ipc4	195	0.039	0.001	0.360	0.040	0.001	0.283	0.040	0.016	0.370
localization_ipc10	195	0.026	0.002	0.159	0.025	0.002	0.221	0.031	0.010	0.267
localization_tec1	195	0.280	0.096	0.436	0.276	0.096	0.399	0.277	0.096	0.416
localization_tec2	195	0.044	0.009	0.329	0.043	0.015	0.279	0.044	0.019	0.386
rvariety_ipc10_ipc4	195	1.525	0	3.314	1.632	0	3.215	1.359	0.087	2.741
rvariety_ipc10_ipc3	195	2.760	0	4.996	2.932	0.333	4.859	2.641	0.475	4.604
rvariety_ipc10_ipc1	195	4.602	0	7.358	4.846	0.667	7.273	4.659	0.675	7.155
rvariety_ipc4_ipc3	195	1.235	0	2.188	1.300	0	2.152	1.282	0.000	2.124
rvariety_ipc4_ipc1	195	3.077	0	4.999	3.213	0	4.858	3.300	0.200	4.911
rvariety_ipc3_ipc1	195	1.842	0	3.027	1.914	0	3.029	2.018	0.133	3.042
rvariety_tec2_tec1	195	2.144	0	2.839	2.235	0	2.841	2.271	0.685	2.848
uvariety_ipc10_ipc4	195	5.406	0	7.498	5.643	0.918	7.575	5.813	1.950	7.605
uvariety_ipc10_ipc3	195	4.171	0	5.764	4.343	0.918	5.755	4.531	1.422	5.792
uvariety_ipc10_ipc1	195	2.328	0	2.906	2.429	0.503	2.892	2.513	0.863	2.932
uvariety_ipc4_ipc3	195	4.171	0	5.764	4.343	0.918	5.755	4.531	1.422	5.792
uvariety_ipc4_ipc1	195	2.328	0	2.906	2.429	0.503	2.892	2.513	0.863	2.932
uvariety_ipc3_ipc1	195	2.328	0	2.906	2.429	0.503	2.892	2.513	0.863	2.932
uvariety_tec2_tec1	195	1.724	0	2.304	1.819	0	2.285	1.914	0.000	2.304

Table 2 Ranking of NUTS2 Regions in Research Outcomes: Reported statistics are yearly averages over 1995-2009.

Rank	Region	*ncit3norm*	Region	*ncit3normpc*	Region	*highcit3*	Region	*highcit3pc*
1	ÎLE DE FRANCE (FR)	2363.7	NOORD-BRABANT (NL)	0.813	NOORD-BRABANT (NL)	26.07	NOORD-BRABANT (NL)	0.0109
2	STUTTGART (DE)	2268.7	RHEINHESSEN-PFALZ (DE)	0.644	OBERBAYERN (DE)	21.87	STOCKHOLM (SE)	0.0078
3	OBERBAYERN (DE)	2183.2	KARLSRUHE (DE)	0.610	DARMSTADT (DE)	21.33	RHEINHESSEN-PFALZ (DE)	0.0062
4	NOORD-BRABANT (NL)	1943.6	STOCKHOLM (SE)	0.590	STUTTGART (DE)	21.13	KARLSRUHE (DE)	0.0058
5	DARMSTADT (DE)	1921.9	STUTTGART (DE)	0.572	ÎLE DE FRANCE (FR)	20.73	DARMSTADT (DE)	0.0057
6	KARLSRUHE (DE)	1652.2	OBERBAYERN (DE)	0.525	KARLSRUHE (DE)	15.67	SYDSVERIGE (SE)	0.0055
7	KÖLN (DE)	1536.2	DARMSTADT (DE)	0.513	KÖLN (DE)	14.73	STUTTGART (DE)	0.0053
8	DÜSSELDORF (DE)	1508.2	TÜBINGEN (DE)	0.482	STOCKHOLM (SE)	14.67	OBERBAYERN (DE)	0.0052
9	LOMBARDIA (IT)	1327.0	FREIBURG (DE)	0.481	DÜSSELDORF (DE)	13.00	HOVEDSTADEN (DK)	0.0052
10	RHEINHESSEN-PFALZ (DE)	1295.4	HOVEDSTADEN (DK)	0.470	LOMBARDIA (IT)	12.73	TÜBINGEN (DE)	0.0050
11	STOCKHOLM (SE)	1099.8	BRABANT WALLON (BE)	0.463	RHEINHESSEN-PFALZ (DE)	12.47	OXFORDSHIRE* (UK)	0.0049
12	RHÔNE-ALPES (FR)	1045.4	SYDSVERIGE (SE)	0.451	OXFORDSHIRE* (UK)	10.47	LÄNSI-SUOMI (FI)	0.0049
13	FREIBURG (DE)	1040.3	MITTELFRANKEN (DE)	0.444	TÜBINGEN (DE)	9.00	MITTELFRANKEN (DE)	0.0045
14	TÜBINGEN (DE)	860.3	VORARLBERG (AT)	0.404	FREIBURG (DE)	8.93	OBERPFALZ (DE)	0.0043
15	OXFORDSHIRE* (UK)	842.8	OXFORDSHIRE* (UK)	0.397	EAST ANGLIA (UK)	8.93	FREIBURG (DE)	0.0041
16	HOVEDSTADEN (DK)	759.2	UNTERFRANKEN (DE)	0.379	RHÔNE-ALPES (FR)	8.67	EAST ANGLIA (UK)	0.0040
17	MITTELFRANKEN (DE)	755.1	VLAAMS-BRABANT (BE)	0.373	HOVEDSTADEN (DK)	8.40	VORARLBERG (AT)	0.0039
18	ARNSBERG (DE)	744.3	OBERPFALZ (DE)	0.370	ZUID-HOLLAND (NL)	7.87	VLAAMS-BRABANT (BE)	0.0038
19	EAST ANGLIA (UK)	725.6	KÖLN (DE)	0.356	ARNSBERG (DE)	7.80	VÄSTSVERIGE (SE)	0.0036
20	EMILIA-ROMAGNA (IT)	697.9	ÖSTRA MELLANSVERIGE (SE)	0.345	MITTELFRANKEN (DE)	7.67	UNTERFRANKEN (DE)	0.0036

*Includes BERKSHIRE, BUCKINGHAMSHIRE AND OXFORDSHIRE (UK).

4. Results

Estimation results are reported in Table 3, Table 4, Table 5 and Table 6. Firstly, estimates of the models for average technological impact (*ncit3norm*) are reported in Table 3 and Table 4. Secondly, estimates of the models for high technological impact (*highcit3*) are reported in Table 5 and Table 6. Different sets of estimates are reported for each technological level of aggregation (ipc1, ipc3, ipc4, ipc10), or for each combination of two levels (ipc10_4, ipc10_3, ipc10_1, ipc4_3, ipc4_1, ipc3_1). Specifically, models in Table 3 and Table 5 include *evenness*, *disparity, localization, and urbanization* as regressors, whose definitions vary according to different levels of technological detail. In Table 4 and Table 6 *evenness* and *disparity* are replaced with *rvariety* and *uvariety*, which vary over combinations of two technological levels, while *localization* is defined according to the lowest level of the two. Alternative results based on technological fields and sub-fields according to the Schmoch concordance table (Schmoch, 2008) are reported in Table 7 of Appendix B. The main findings are qualitatively similar to the ones in Table 3, Table 4, Table 5 and Table 6.

In Table 3 and Table 4 we report estimates of standard FE models and FE models correcting for 1st order autocorrelation in the errors. Under the null of no autocorrelation, the residuals from the regression of the first-differenced variables should have an autocorrelation of -0.5. These residuals are used to test the null hypothesis that the coefficient of the lagged residuals in a regression of the current residuals is equal to -0.5 (Wooldridge, 2010; Drukker, 2003). In Table 3 and Table 4 these coefficients are typically around -.45, signaling some amount of serial correlation. In all these specifications the F test of no AR(1) is significant at 5%; therefore we focus our analysis on FE-AR(1) models. Evidence of multicollinearity in the regressors of interest is found for none of the models in Table 3 and Table 4. In fact, the Variance Inflation Factors (VIFs) are always below reasonable bounds.

First, looking at the effect of diversity measures in Table 3 we find a negative impact of *evenness* on average technological impact for all cases except ipc1. Despite no significant effect is found for the highest level of aggregation, the magnitude of the effect has a stable increasing trend as the level gets more disaggregated. For *disparity*, no significant effect is found, though the coefficient for ipc10 is positive and very close to 10% significance. *Localization* economies are found for ipc10 and, to a lesser extent, for ipc1.[10] *Urbanization* has a positive effect on average technological impact across the board.

Second, looking at Table 4 we find that *related variety* is detrimental to average technological impact in three combinations of technological levels, namely ipc3_1, ipc4_1 and ipc4_3. A negative effect is not found when the lowest level is used in the combination, i.e. ipc10. Moreover, the effect of *related variety* gets positive and significant when ipc10 is combined with ipc4, the second lowest level. *Unrelated variety* is always neutral instead. Similarly to Table 3, localization economies are found when highest level of technological detail is used, corresponding to specifications with ipc10 as the lower level of aggregation. However, a

[10] Remark that *localization* has significant positive impact also for technological fields (tec1), as reported in model (5) in Table 7, corresponding to a very high level of aggregation.

positive effect is found also when ipc4 is used as the lower level of aggregation (i.e. *localization_ipc4* is used). Consistently with Table 3, at the lowest level of technological detail localization has no significant effect in model (7). Also, *urbanization* again has a positive effect on average technological impact of inventive activity across EU regions.

In Table 5 and Table 6 we report FE-NB estimates for the count of highly cited patents (*highcit3*). The likelihood-ratio statistics support the use of a NB model over a Poisson since the null hypothesis of no over-dispersion is always strongly rejected. The test of AR(1) refers here to the log-linear panel case where log(1 + *highcit3*) is used in place of *highcit3*. This test shows clearly that AR(1) is absent in residuals for the transformed variable, suggesting that serial dependence is not worrisome in this case.

Third then, looking at Table 5, we find that *evenness* has generally a detrimental effect also on high technological impact. Unlike for average technological impact in Table 3, the effect on research excellence is not significant for the lowest level of aggregation (ipc10) but it is significant for ipc1. Except for ipc10, the magnitude of the effect increases as the level gets more disaggregated, similarly to Table 3. Similarly to average technological impact, the effect of *disparity* on research excellence is not significant for ipc1, ipc3 and ipc4, but it is positive and becomes strongly significant for the lowest level (ipc10). Results for *localization* are a bit different to Table 3. In fact, while a positive effect for the lowest level is as before, *localization* is found to reduce high technological impact for all the other levels.

Finally, looking at Table 6, estimates of the effect of *related variety* on high technological impact have a remarkably similar pattern to results for average technological impact in Table 4; namely same negative impact for ipc3_1 and ipc4_1, same positive impact for ipc10_4, but the negative impact for ipc4_3 loses significance in Table 6. Interestingly, we notice also that the effect of *related variety* grows monotonically as one of the two levels becomes more detailed. The effect of *unrelated variety* is generally non-significant as in Table 4, with the only exception of a negative effect for ipc4_3. Results for *localization* confirm results of Table 5; namely positive only when the highest level of technological detail (ipc10) is used and negative otherwise. Whilst Table 3 and Table 4 bear witness on the presence of *urbanization* economies concerning average technological impact, when it comes to high technological impact the coefficient is generally non-significant or at best positive and mildly significant (see Table 5 and Table 6). With regards to the effect of collaboration shares, we do find that higher share of collaborations involving non-European inventors improves the performance of EU regions for both citation indicators (see Table 3, Table 4, Table 5 and Table 6). The impact is significant for any collaboration share used for comparison.[11]

[11] As can be understood from Table 3, Table 4, the coefficient on *collabeu27* is significant either considering the baseline share (*nocollab*), or the others. From Table 5 and Table 6 we notice that the coefficient on *collabeu27* is significant with respect to the baseline, however, by running further models using in turn the remaining shares as baseline, the coefficient is still significant in most cases.

Table 3 Diversity and Average Technological Impact: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors in parenthesis. All models are estimated for 195 regions over 1995-2009. FE – AR(1) models correct for 1st order autocorrelation. No autocorrelation corresponds to AR(1) = -0.5 in residuals from the regression of the first-differenced variables. VIFs are variance inflation factors for the specified regressors. VIF < 3 is interpreted here as a sign that the regressor is not collinear.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	FE	FE	FE	FE	FE - AR(1)	FE - AR(1)	FE - AR(1)	FE - AR(1)
log(*ncit3norm*)	ipc1	ipc3	ipc4	ipc10	ipc1	ipc3	ipc4	ipc10
log(*gdppc*)	1.054***	1.022***	1.021***	1.056***	0.877***	0.859***	0.857***	0.904***
	(0.261)	(0.256)	(0.257)	(0.256)	(0.157)	(0.157)	(0.158)	(0.157)
log(*pop*)	0.446	0.403	0.403	0.480	0.650*	0.627*	0.638*	0.683*
	(0.321)	(0.313)	(0.319)	(0.322)	(0.349)	(0.348)	(0.349)	(0.349)
log(*npatgdp*)	1.107***	1.080***	1.066***	1.090***	1.123***	1.100***	1.086***	1.106***
	(0.034)	(0.033)	(0.035)	(0.036)	(0.031)	(0.030)	(0.030)	(0.030)
collabwr	-0.193	-0.245	-0.279	-0.250	-0.284**	-0.356***	-0.401***	-0.378***
	(0.213)	(0.211)	(0.200)	(0.204)	(0.120)	(0.121)	(0.120)	(0.120)
collaborwc	-0.097	-0.109	-0.108	-0.060	-0.134	-0.145	-0.145	-0.104
	(0.165)	(0.170)	(0.164)	(0.159)	(0.116)	(0.116)	(0.115)	(0.115)
collabocweu27	-0.062	-0.079	-0.097	-0.046	-0.117	-0.140	-0.169	-0.108
	(0.214)	(0.214)	(0.208)	(0.209)	(0.120)	(0.120)	(0.119)	(0.118)
collaboeu27	0.346	0.306	0.283	0.315	0.425***	0.402***	0.371***	0.411***
	(0.233)	(0.241)	(0.243)	(0.232)	(0.137)	(0.139)	(0.136)	(0.135)
evenness	0.048	-0.558***	-0.982***	-1.006***	0.067	-0.550***	-0.923***	-1.104***
	(0.197)	(0.210)	(0.222)	(0.282)	(0.133)	(0.165)	(0.212)	(0.296)
disparity	-0.153	-0.290	0.137	0.443**	-0.060	-0.225	0.207	0.262
	(0.275)	(0.311)	(0.300)	(0.196)	(0.184)	(0.203)	(0.204)	(0.160)
localization	0.711	0.577	0.468	2.079	0.521*	0.422	0.591	2.493***
	(0.580)	(1.062)	(0.824)	(1.519)	(0.301)	(0.482)	(0.617)	(0.939)
Observations	2,925	2,925	2,925	2,925	2,730	2,730	2,730	2,730
Number of idnuts2	195	195	195	195	195	195	195	195
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2 within	0.624	0.625	0.628	0.627	0.546	0.548	0.550	0.551
R2 between	0.906	0.901	0.904	0.917	0.947	0.948	0.953	0.955
R2 overall	0.868	0.863	0.867	0.878	0.902	0.904	0.907	0.910
VIF - rvariety	1.11	1.43	1.52	1.25	1.11	1.43	1.52	1.25
VIF - uvariety	1.24	1.22	1.46	1.43	1.24	1.22	1.46	1.43
VIF - localization	1.06	1.26	1.38	1.35	1.06	1.26	1.38	1.35
AR(1) coeff in error	-0.447	-0.446	-0.443	-0.442				
F test no AR(1)	4.616	4.767	5.122	5.644				
Prob > F	0.033	0.030	0.025	0.018				

Table 4 Related\Unrelated Variety and Average Technological impact: *** p<0.01, ** p<0.05, * p<0.1. See notes to Table 3.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	FE	FE	FE	FE	FE	FE	FE - AR(1)	FE - AR(1)	FE - AR(1)	FE - AR(1)	FE - AR(1)	FE - AR(1)
log(*ncit3norm*)	ipc3_1	ipc4_1	ipc4_3	ipc10_1	ipc10_3	ipc10_4	ipc3_1	ipc4_1	ipc4_3	ipc10_1	ipc10_3	ipc10_4
log(*gdppc*)	1.069***	1.068***	1.071***	1.073***	1.087***	1.095***	0.897***	0.881***	0.882***	0.916***	0.932***	0.935***
	(0.268)	(0.262)	(0.259)	(0.252)	(0.251)	(0.250)	(0.160)	(0.160)	(0.160)	(0.159)	(0.158)	(0.158)
log(*pop*)	0.464	0.452	0.421	0.406	0.414	0.417	0.683*	0.665*	0.645*	0.641*	0.643*	0.642*
	(0.314)	(0.309)	(0.309)	(0.313)	(0.313)	(0.306)	(0.350)	(0.349)	(0.348)	(0.348)	(0.348)	(0.347)
log(*npatgdp*)	1.122***	1.154***	1.160***	1.123***	1.114***	1.113***	1.136***	1.160***	1.168***	1.134***	1.126***	1.124***
	(0.037)	(0.042)	(0.043)	(0.047)	(0.047)	(0.045)	(0.032)	(0.034)	(0.034)	(0.040)	(0.040)	(0.040)
collabwr	-0.210	-0.215	-0.219	-0.177	-0.196	-0.216	-0.309**	-0.313***	-0.320***	-0.277**	-0.300**	-0.320***
	(0.218)	(0.211)	(0.207)	(0.210)	(0.210)	(0.210)	(0.120)	(0.119)	(0.119)	(0.119)	(0.119)	(0.119)
collaborwc	-0.104	-0.095	-0.089	-0.043	-0.066	-0.078	-0.144	-0.138	-0.128	-0.085	-0.105	-0.115
	(0.168)	(0.168)	(0.165)	(0.165)	(0.165)	(0.166)	(0.116)	(0.116)	(0.116)	(0.116)	(0.116)	(0.116)
collabocweu27	-0.073	-0.076	-0.070	-0.005	-0.033	-0.048	-0.135	-0.151	-0.139	-0.069	-0.092	-0.111
	(0.214)	(0.205)	(0.208)	(0.210)	(0.213)	(0.212)	(0.120)	(0.120)	(0.120)	(0.118)	(0.119)	(0.119)
collaboeu27	0.311	0.329	0.327	0.379	0.345	0.333	0.396***	0.401***	0.404***	0.460***	0.435***	0.424***
	(0.243)	(0.243)	(0.239)	(0.234)	(0.232)	(0.233)	(0.139)	(0.136)	(0.136)	(0.135)	(0.136)	(0.135)
rvariety	-0.090*	-0.113***	-0.152**	-0.012	0.026	0.082	-0.083**	-0.103***	-0.151***	-0.008	0.025	0.076**
	(0.046)	(0.041)	(0.060)	(0.044)	(0.045)	(0.057)	(0.034)	(0.031)	(0.049)	(0.030)	(0.034)	(0.039)
uvariety	0.006	0.000	-0.053	-0.019	-0.039	-0.042	0.014	0.013	-0.042	0.002	-0.024	-0.030
	(0.053)	(0.052)	(0.032)	(0.051)	(0.036)	(0.035)	(0.041)	(0.040)	(0.027)	(0.043)	(0.031)	(0.030)
localization	0.715	1.147	1.358	2.379*	2.443*	2.180	0.596	1.348**	1.583**	3.023***	3.045***	2.779***
	(1.018)	(0.869)	(0.909)	(1.424)	(1.426)	(1.452)	(0.499)	(0.609)	(0.628)	(1.016)	(1.015)	(1.017)
Observations	2,925	2,925	2,925	2,925	2,925	2,925	2,730	2,730	2,730	2,730	2,730	2,730
Number of idnuts2	195	195	195	195	195	195	195	195	195	195	195	195
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2 within	0.624	0.626	0.626	0.624	0.625	0.626	0.546	0.548	0.550	0.548	0.549	0.551
R2 between	0.908	0.894	0.882	0.888	0.896	0.898	0.952	0.943	0.937	0.943	0.946	0.947
R2 overall	0.870	0.857	0.848	0.852	0.859	0.861	0.906	0.899	0.894	0.898	0.902	0.903
VIF - rvariety	1.38	1.67	1.37	2.54	2.7	2.17	1.38	1.67	1.37	2.54	2.7	2.17
VIF - uvariety	1.49	1.45	1.97	1.57	2.44	2.82	1.49	1.45	1.97	1.57	2.44	2.82
VIF - localization	1.31	1.32	1.41	1.53	1.53	1.54	1.31	1.32	1.41	1.53	1.53	1.54
AR(1) coeff in error	-0.447	-0.449	-0.450	-0.446	-0.447	-0.448						
F test no AR(1)	4.424	4.032	4.202	4.733	4.514	4.341						
Prob > F	0.037	0.046	0.042	0.031	0.035	0.039						

Table 5 Diversity and High Technological impact: *** p<0.01, ** p<0.05, * p<0.1. Analytical standard errors based on the observed information matrix in parenthesis. All models are estimated for 185 regions over 1995-2009; 10 regions are dropped because of all zero outcomes. AR(1) test refers to the log-linear case where log(1 + *highcit3*) is used in place of *highcit3*. See also notes to Table 3. The LR tests the null of no overdispersion; rejection supports a NB over a Poisson.

	(1)	(2)	(3)	(4)
	FE-NB	FE-NB	FE-NB	FE-NB
highcit3	ipc1	ipc3	ipc4	ipc10
log(*gdppc*)	0.394	0.310	0.323	0.583*
	(0.317)	(0.322)	(0.328)	(0.336)
log(*pop*)	0.439	0.348	0.315	0.450
	(0.296)	(0.295)	(0.300)	(0.301)
log(*npatgdp*)	1.087***	1.001***	0.950***	1.079***
	(0.088)	(0.090)	(0.089)	(0.099)
collabwr	0.603	0.576	0.513	0.613
	(0.425)	(0.426)	(0.426)	(0.422)
collaborwc	0.904**	0.852**	0.819**	0.923**
	(0.386)	(0.386)	(0.386)	(0.384)
collabocweu27	0.921*	0.839	0.884*	0.954*
	(0.512)	(0.514)	(0.514)	(0.513)
collaboeu27	1.698***	1.613***	1.513***	1.498***
	(0.531)	(0.538)	(0.538)	(0.537)
evenness	-1.430***	-2.166***	-2.830***	-0.080
	(0.463)	(0.533)	(0.648)	(1.356)
disparity	0.759	0.224	0.134	1.269***
	(0.581)	(0.599)	(0.646)	(0.432)
localization	-3.877***	-7.302***	-5.596**	8.824**
	(1.382)	(2.441)	(2.195)	(3.735)
Observations	2,775	2,775	2,775	2,775
Number of idnuts2	185	185	185	185
Year dummies	Yes	Yes	Yes	Yes
AR(1) coeff in error	-0.486	-0.486	-0.486	-0.487
F test no AR(1)	0.874	0.909	0.958	0.787
Prob > F	0.351	0.341	0.329	0.376
lr overdispersion test (Chi2)	23.340	25.127	23.806	25.127
Prob > Chi2	0.000	0.000	0.000	0.000

Table 6 Related\Unrelated Variety and High Technological impact: *** p<0.01, ** p<0.05, * p<0.1. See notes to Table 5.

	(1)	(2)	(3)	(4)	(5)	(6)
	FE-NB	FE-NB	FE-NB	FE-NB	FE-NB	FE-NB
highcit3	ipc3_1	ipc4_1	ipc4_3	ipc10_1	ipc10_3	ipc10_4
log(*gdppc*)	0.469	0.519	0.479	0.464	0.395	0.385
	(0.323)	(0.328)	(0.327)	(0.325)	(0.327)	(0.327)
log(*pop*)	0.488	0.521*	0.522*	0.408	0.334	0.314
	(0.303)	(0.305)	(0.299)	(0.300)	(0.305)	(0.311)
log(*npatgdp*)	1.148***	1.174***	1.141***	1.076***	1.022***	1.026***
	(0.082)	(0.083)	(0.083)	(0.095)	(0.095)	(0.093)
collabwr	0.619	0.636	0.620	0.693	0.581	0.557
	(0.423)	(0.423)	(0.424)	(0.423)	(0.425)	(0.425)
collaborwc	0.909**	0.939**	0.942**	1.009***	0.921**	0.885**
	(0.384)	(0.384)	(0.384)	(0.384)	(0.385)	(0.386)
collabocweu27	0.907*	0.975*	1.001*	1.035**	0.954*	0.915*
	(0.510)	(0.512)	(0.511)	(0.511)	(0.512)	(0.512)
collaboeu27	1.681***	1.679***	1.677***	1.732***	1.561***	1.536***
	(0.535)	(0.535)	(0.535)	(0.532)	(0.535)	(0.535)
rvariety	-0.342***	-0.295***	-0.178	0.023	0.136	0.206*
	(0.096)	(0.084)	(0.131)	(0.088)	(0.097)	(0.110)
uvariety	-0.150	-0.054	-0.240***	-0.003	-0.044	-0.017
	(0.147)	(0.140)	(0.076)	(0.135)	(0.084)	(0.081)
localization	-6.529***	-4.034*	-4.376**	9.898***	10.337***	9.827***
	(2.455)	(2.173)	(2.206)	(3.548)	(3.538)	(3.530)
Observations	2,775	2,775	2,775	2,775	2,775	2,775
Number of idnuts2	185	185	185	185	185	185
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
AR(1) coeff in error	-0.488	-0.488	-0.488	-0.488	-0.488	-0.488
F test no AR(1)	0.705	0.673	0.672	0.707	0.678	0.682
Prob > F	0.402	0.413	0.413	0.401	0.411	0.410
lr overdispersion test (Chi2)	23.809	24.129	24.422	24.087	23.727	23.553
Prob > Chi2	0.000	0.000	0.000	0.000	0.000	0.000

5. Discussion

The findings documented in Section 4 overall provide mixed evidence on the effect of diversity on the technological impact of inventive activity across EU regions. A general result is that these effects are very sensitive to the aggregation level used to separate technologies.

Concerning the average impact of inventive activities, we clearly observe an advantage of concentrating invention activities in few and relatively closely related technologies. This claim derives from two sets of findings. On the one hand, the less evenly are these activities distributed within a region, the higher is the impact of the region relative to the impact that technologically similar invention activities can have elsewhere (see Table 3). For this effect to be significant, it is necessary that technologies in which inventive activities are concentrated are not defined too widely, as can be understood by the coefficients of *evenness* in various specifications in Table 3.

On the other hand, when the notion of distance is incorporated in the measures, we find evidence that diversity can be beneficial only when it is constrained to very related and detailed technologies. First, related variety results in a positive effect on average technological impact only when it is considered within the boundaries of a very detailed technology, which in turn belongs to a relatively narrow higher level sector (see ipc10_4 in Table 4). Otherwise, related variety can be either neutral or even detrimental when relatively large technological boundaries are considered (see ipc3_1, ipc4_1 and ipc4_3 in Table 4). Second, diversity in the form of combination of more distant technologies (*disparity*) is close to having a positive impact for a highly disaggregated level of analysis (see ipc10 in Table 3).

Concerning the effect of concentration of technologies in a region on the average impact of inventive activity, as captured by the *localization* indicator, we find always a positive and often significant effect. Remark that this indicator is concerned with the *spatial* concentration of sectors rather than specialization *within* a given region. Therefore, and as argued in section 2, the main theoretical justification for the presence of localization economies relates to geographical knowledge spillovers. Our results show that localization economies are strongest at the lowest level of aggregation (see ipc10 in Table 3 and Table 4).

The results for high technological impact are similar to the results for average technological impact, with two important qualifications though. A first distinction concerns the effect of diversity being more beneficial for high technological impact at the lowest level of aggregation (ipc10). In fact, on the one hand, *disparity* becomes strongly significant for ipc10 in Table 5, unlike it is only close to significance in Table 3. On the other hand, looking again at Table 5, while the negative effect of *evenness* on high technological impact increases as the level of aggregation becomes lower, similarly to Table 3, this effect is no more significant for the lowest level. It has to be remarked here that recombinant inventions relating distant domains are more likely to be of a radical nature (Fleming, 2001; Saviotti & Frenken, 2008); therefore they are expected to have higher volatility in impact, and to pay-off in terms of high technological impact rather than in terms of average technological impact. However, evidence that the diversity premium on high impact over the average impact is found only for the most detailed level, suggests that, in the present context, incremental inventions as originating by recombination of very close technologies are more likely to succeed than radical inventions in terms of high impact.

A second distinction concerns the effect of *localization*. While a positive effect on high technological impact is found at ipc10 similarly to average technological impact, a negative impact is found for all the other levels too. Geographical concentration of technologically similar inventive activities is expected to enhance knowledge spillovers, resulting in higher impact. However, it can also induce a lock-in effect in knowledge diffusion so long as knowledge elsewhere potentially building up on that stock of knowledge is too little and sparse. This latter effect can dominate if the citation impact is measured by the capacity of the region to have an outstanding impact. Along the same line of reasoning, a persisting positive effect at the lowest level of aggregation can be interpreted noting that knowledge elsewhere is less likely to be technologically distant when very narrow sectors are considered.

To conclude, the findings can be summarized as follows. On the one hand, the effect of diversity is typically negative, or at best absent, if we do not consider very fine-grained disaggregation of technologies. However, it is also true that diversity can be beneficial when it is constrained to very related and detailed technologies. On the other hand, while localization is always beneficial at the most fine-grained level of aggregation, it can be beneficial too for average technological impact, or detrimental for high technological impact, when less fine-grained levels are considered.

6. Concluding remarks

This paper addressed the relationship between the technological diversity and the technological impact of inventive activity across EU regions. Thus far, this relationship had been left largely unaddressed throughout the literature. What is more, whenever the relationship between the technological composition of regions and the impact of their inventions has been addressed, the evidence offered has been rather inconclusive. From a theoretical point of view, plausible arguments are offered supporting opposite claims, namely that specialization and localization on the one hand and diversification and urbanization on the other have a positive effect on the impact of new technologies. Empirically, we identified two main issues concerning the evidence being inconclusive. One is that the notion of technological diversity itself can be conceptualized in different ways; all leading to different sets of indicators used in the empirical assessment. Another issue is that technological diversity, irrespective of the conceptualization and indicators employed, can apply to different levels of technological detail.

The main conclusion of this paper holds that diversity is typically detrimental, or at best neutral, for the impact of new technologies in Europe, except when a very fine-grained technological detail is taken into account. Specifically, except in the latter case, positive effects are driven by concentration in few and related technologies, as captured by lower evenness and related variety respectively. Moreover, localization is found to be detrimental only for the high impact.

Conversely, nearly opposite conclusions hold if our focus is on to the lowest level of aggregation available. In this case, benefits can arise from (related) variety and, particularly in terms of high technological impact, from combination of relatively distant technologies (disparity). Benefits arise also from localization.

It follows that, in terms of policy implications, two policy options could be considered in order to boost technological impact of new inventions. One option concerns steering specialization in closely neighboring technologies. The other policy option concerns steering diversification among closely related technologies and localization at a coarse-grained technological level. What is crucial here is that policymakers should have a very detailed understanding about the relations among technologies if such specialization/diversification strategies are to succeed. Whenever such detailed understanding is lacking, specialization (diversification) might easily turn out to become counterproductive; that is, decrease instead of increase the technological impact of regional invention in Europe. Therefore, our results emphasize the importance of taking into account (i) the relations among different technologies and (ii) the appropriate level of technological detail along which relations among technologies play out.

It has to be remarked that these implications for policy may derive from the particular reward system for inventions present in the current institutional framework. At the outset, the results found in this paper for the European context seem to be at odds with those found by Castaldi et al. (2014) for the US context. Castaldi et al. (2014) found that in the US diversification across relatively distant technologies rather than specialization in closely neighboring ones has a positive effect on high technological impact of regional invention. Following the literature on Varieties of Capitalism, part of this difference in results might be explained by the US having a comparative advantage in radical invention whilst Europe has a comparative advantage in incremental invention (Boschma & Capone, 2014). It follows that unrelated diversification (related specialization) might be beneficial for the technological impact of regional invention only whenever radical invention is rewarded more (less) than incremental invention. In other words, if diversification across relatively distant technologies and not specialization in relatively close technologies is considered to be a viable policy option in Europe, then the reward system for invention should be drastically revised.

Of course, we would need more research in order to further substantiate some of the claims made in this paper. For example, we could look at whether similar patterns exist for scientific knowledge production by taking into account publication data instead of patent data. Second, and perhaps even more important, it would be interesting to dig deeper into the differences in invention patterns between the US and Europe. Not only in terms of how these research and innovation systems are organized (cf. Hardeman et al., 2014), but also in terms of the characteristics of the inventions that have technological impact. Such an analysis would allow for a more detailed understanding on why nearly opposite effects are found when comparing regional invention in the US with regional invention in Europe.

References

Allison, P. D. (1999). *Multiple Regression: A Primer*. London: Sage.

Allison, P. D. (2009). *Fixed Effects Regression Models*. London: Sage.

Baltagi, B. H., & Q. Li, (1991). A transformation that will circumvent the problem of autocorrelation in an error component model. *Journal of Econometrics*, 48, 385–393.

Baltagi, B. H., and P. X. Wu (1999). Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15: 814–823.

Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, 38(2), 318-337.

Boschma, R. (2013). Constructing Regional Advantage and Smart Specialization: Comparison of Two European Policy Concepts. *Papers in Evolutionary Economic Geography (PEEG)*, No. 1322.

Boschma, R., & Capone, G. (2014). Institutions and Diversification: Related versus Unrelated Diversification in a Varieties of Capitalism framework. *Papers in Evolutionary Economic Geography (PEEG)*, No. 1421.

Cameron, A. C., and Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge University press.

Castaldi, C., Frenken, K., & Los, B. (2014). Related Variety, Unrelated Variety and Technological Breakthroughs: an analysis of US state-level patenting. *Regional Studies*, (ahead-of-print), 1-15.

De Groot, H. L., Poot, J., & Smit, M. J. (2009). 14 Agglomeration externalities, innovation and regional growth: theoretical perspectives and meta-analysis. *Handbook of regional growth and development theories*, 256.

Desmet, K, Ortuño-Ortín, I. & Weber, S (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7(6): 1291-1318.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3), 147-162.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal* (3)2: 168–177.

Ejermo, O. (2005). Technological diversity and Jacobs' externality hypothesis revisited. *Growth and Change*, 36(2), 167-195.

European Commission (2010) *Regional Policy contributing to smart growth in Europe 2020*. COM(2010) 553 final, Brussels: European Commission.

Fleming, L. (2001) Recombinant uncertainty in technological space. *Management Science* 47(1), 117-132.

Frenken, K., Ponds, R., & Van Oort, F. (2010). The citation impact of research collaboration in science-based industries: A spatial-institutional analysis. *Papers in Regional Science*, 89(2), 351-271.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685-697.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126-1152.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature*, 28, 1661-1707.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 16-38.

Hall, P. A., & Soskice, D. (Eds.). (2001). *Varieties of capitalism: The institutional foundations of comparative advantage*. Oxford: Oxford University Press.

Hardeman, S.; Frenken, K.; Nomaler, Ö.; Ter Wal, A.J. (2014). Characterizing and comparing innovation systems by different "Modes" of knowledge production: a proximity approach. *Science and Public Policy*, forthcoming, doi:10.1093/scipol/scu070.

Hsiao, C. (2014). *Analysis of panel data* (Vol. 54). Cambridge University press.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, 108(3), 577-598.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, 322(5905), 1259-1262.

Maraut, S., H. Dernis, C. Webb, V. and D. Guellec (2008). The OECD REGPAT Database: A Presentation. *OECD Science, Technology and Industry Working Papers*, No. 2008/02, Paris: OECD Publishing.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87(3), 237-265.

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41:673–690.

OECD (2013) *Innovation-driven Growth in Regions: The Role of Smart Specialisation*. Paris: OECD Publications.

Olsson, O. (2000). Knowledge as a set in idea space: An epistemological view on growth. *Journal of Economic Growth*, 5(3), 253-275.

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton: Princeton University Press.

Palan, N. (2010). Measurement of Specialization–The Choice of Indices. *FIW- Working Paper*, No. 062, December 2010.

Saviotti, P.P. and Frenken, K. (2008) Trade variety and economic development of countries. *Journal of Evolutionary Economics* 18(2), 201-218.

Schmoch, U. (2008). *Concept of a Technology Classification for Country Comparisons*, Final Report to the World Intellectual Property Organisation (WIPO), http://www.wipo.int/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf

Squicciarini, M., H. Dernis and C. Criscuolo (2013). Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Science, Technology and Industry Working Papers*, No. 2013/03, Paris: OECD Publishing.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15), 707-719.

Tijssen, R. J. (2003). Scoreboards of research excellence. *Research Evaluation*, 12(2), 91-103.

Van der Panne, G. (2004). Agglomeration externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics*, 14(5), 593-604.

Van Oort, F., de Geus, S., & Dogaru, T. (2014). Related variety and regional economic growth in a cross-section of European urban regions. *European Planning Studies*, (ahead-of-print), 1-18.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., ... & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14-26.

Webb, C., H. Dernis, D. Harhoff, K. Hoisl (2005). Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks. *OECD Science, Technology and Industry Working Papers*, No. 2005/09, Paris: OECD Publishing.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, 331-360.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wooldridge, J. M. (2013). *Introductory Econometrics: A modern approach*, South-Western College Publishing.

Yegros, A., D'Este, P., & Rafols, I. (2013). Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity. *Paper presented at 35th DRUID Celebration Conference 2013, Barcelona, Spain, June 17-19*, Retrieved from: http://druid8.sit.aau.dk/druid/acc_papers/54dcxbblnj9vbrlt2v4gku686mcx.pdf

Appendix A. Normalization of the citation measure.

This section describes the main steps used to normalize forward citations. The procedure is based on two main steps (Squicciarini et al., 2013). First each patent $s = 1 \dots S$ is associated to one unique technology k, allowing to define $ncit3_{s,k,t}$ which represents the number of citations received by a patent s filed in year t with technology k (remark that k depends on the level of aggregation used). When several technology fields are allocated to a patent, only the one with the majority of sub-fields occurrences is kept. In case of fields with same number of sub-fields, the field is selected randomly. Citation counts are then aggregated by field and year, giving rise to $ncit3_{k,t} = \sum_s ncit3_{s,k,t}$, where $s = 1 \dots S_{k,t}$, and $S_{k,t}$ is the number of patents in sector k and year t.

Secondly, $ncit3_{s,k,t}$ is divided by a summary statistic of citations received by patents in the same field and year. Specifically, we consider the mean $\underline{ncit3}_{k,t} = \sum_s ncit3_{s,k,t} / S_{k,t}$, giving rise to $ncit3norm_{s,k,t} = ncit3_{s,k,t} / \underline{ncit3}_{k,t}$. Finally, the variable used in the analysis ($ncit3norm_{i,t}$) is obtained aggregating $ncit3norm_{s,k,t}$ by region and year.

Two main choices in the normalization procedure can affect the resulting citation variable. First, the choice of the technological level. Throughout the analysis, the level *tec2* is always used for consistency. The correlation among citation variables based on different technological levels is extremely high. Second, the choice of the statistics for normalization. Different statistics have been considered for the normalization, such as the total, the maximum, the mean and some percentiles. The correlation among the resulting variables is anyway very high.

Appendix B. Estimates with an alternative classification scheme

In this Section we report alternative results using technological fields and sub-fields according to the Schmoch concordance table (Schmoch, 2008), in place of ipc levels. Specifically, Table 7 combines four sets of results corresponding to the ones reported in Table 3, Table 4, Table 5 and Table 6.

Table 7 Variety/Diversity and research impact with technological fields and sub-fields: *** p<0.01, ** p<0.05, * p<0.1. See notes to Table 3, Table 4, Table 5 and Table 6.

	log(*ncit3norm*)						*highcit3*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	FE	FE - AR(1)	FE	FE	FE - AR(1)	FE - AR(1)	FE-NB	FE-NB	FE-NB
	tec2_1	tec2_1	tec1	tec2	tec1	tec2	tec2_1	tec1	tec2
log(*gdppc*)	1.101***	0.916***	1.065***	1.036***	0.882***	0.872***	0.555*	0.371	0.429
	(0.254)	(0.159)	(0.257)	(0.256)	(0.158)	(0.157)	(0.329)	(0.319)	(0.326)
log(*pop*)	0.473	0.697**	0.427	0.422	0.641*	0.658*	0.488	0.458	0.419
	(0.311)	(0.346)	(0.316)	(0.314)	(0.350)	(0.348)	(0.318)	(0.281)	(0.299)
log(*npatgdp*)	1.146***	1.155***	1.104***	1.077***	1.116***	1.099***	1.165***	1.091***	1.046***
	(0.034)	(0.032)	(0.038)	(0.034)	(0.031)	(0.030)	(0.083)	(0.086)	(0.088)
collabwr	-0.198	-0.294**	-0.179	-0.240	-0.260**	-0.357***	0.584	0.714*	0.551
	(0.208)	(0.120)	(0.202)	(0.196)	(0.122)	(0.120)	(0.420)	(0.426)	(0.426)
collaborwc	-0.057	-0.101	-0.078	-0.089	-0.113	-0.133	0.947**	1.007***	0.866**
	(0.168)	(0.116)	(0.165)	(0.161)	(0.117)	(0.116)	(0.383)	(0.388)	(0.385)
collabocweu27	-0.039	-0.100	-0.037	-0.073	-0.095	-0.141	0.941*	1.035**	0.859*
	(0.211)	(0.119)	(0.211)	(0.209)	(0.119)	(0.119)	(0.506)	(0.513)	(0.514)
collaboeu27	0.331	0.411***	0.366	0.313	0.461***	0.393***	1.667***	1.800***	1.601***
	(0.248)	(0.137)	(0.241)	(0.249)	(0.137)	(0.137)	(0.528)	(0.535)	(0.534)
rvariety	-0.136**	-0.122***					-0.553***		
	(0.054)	(0.035)					(0.119)		
uvariety	-0.011	0.010					-0.208		
	(0.062)	(0.042)					(0.146)		
localization	1.148	1.450**	0.390	0.551	0.631**	0.816	-5.950*	-2.564	-7.654**
	(0.970)	(0.639)	(0.412)	(0.928)	(0.302)	(0.583)	(3.223)	(1.624)	(3.001)
evenness			-0.050	-0.621***	0.002	-0.526***		-0.482	-2.140***
			(0.158)	(0.209)	(0.087)	(0.168)		(0.309)	(0.530)
disparity			-0.146	0.058	-0.083	0.128		0.667	0.438
			(0.303)	(0.263)	(0.182)	(0.193)		(0.625)	(0.524)
Observations	2,925	2,730	2,925	2,925	2,730	2,730	2,775	2,775	2,775
Number of idnuts2	195	195	195	195	195	195	185	185	185
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2 within	0.626	0.551	0.623	0.626	0.545	0.549			
R2 between	0.903	0.949	0.900	0.902	0.946	0.951			
R2 overall	0.866	0.904	0.862	0.865	0.901	0.906			
VIF - rvariety	1.34	1.34	1.21	1.28	1.21	1.28			
VIF - uvariety	1.77	1.77	1.24	1.21	1.24	1.21			
VIF - localization	1.45	1.45	1.04	1.20	1.04	1.20			
AR(1) coeff in error	-0.449		-0.444	-0.446			-0.488	-0.486	-0.487
F test no AR(1)	4.411		5.229	4.666			0.706	0.959	0.803
Prob > F	0.037		0.023	0.032			0.402	0.329	0.371
lr overdispersion test (Chi2)							21.612	25.193	24.027
Prob > Chi2							0.000	0.000	0.000