Appendix A: Statistics on specializations gained and lost by the regions

Table A1. Number of new technological specializations at time t+4 (in comparison to time to time t), by period and RCA of the technology at time t

Periods	RCA=0	0<RCA≤0.25	0.25<RCA≤0.5	0.5<RCA≤0.75	0.75<RCA≤1	RCA≥0
2000-2004	2171	102	226	268	312	3079
2001-2005	2222	92	247	267	274	3102
2002-2006	2173	102	243	303	322	3143
2003-2007	2247	111	242	312	308	3220
2004-2008	2141	122	243	279	312	3097
2005-2009	2227	102	243	283	300	3155
2006-2010	2145	104	250	310	310	3119
2007-2011	2099	104	258	329	318	3108
2008-2012	2144	85	267	304	336	3136
2009-2013	2113	109	260	290	353	3125
Total	21682	1033	2479	2945	3145	31284

Source: author's computations

Table A2. Number of technological specializations at time t, in which the region is not specialized in at t+4, by period and RCA of the technology at time t+4

Periods	RCA=0	0<RCA≤0.25	0.25<RCA≤0.5	0.5<RCA≤0.75	0.75<RCA≤1	RCA≥0
2000-2004	1959	105	252	291	350	2957
2001-2005	1983	106	223	282	310	2904
2002-2006	1975	104	258	333	308	2978
2003-2007	1998	111	257	377	337	3080
2004-2008	2013	95	258	296	356	3018
2005-2009	1997	108	283	344	372	3104
2006-2010	2022	103	271	349	364	3109
2007-2011	2075	119	266	345	344	3149
2008-2012	2077	102	245	318	336	3078
2009-2013	2091	117	263	336	347	3154
Total	20190	1070	2576	3271	3424	30531

Source: author's computations

Appendix B: Computation of the technology-adjusted diversification potential

Step 1: Computation of expected number of patents for each region i, technology z and year t

$$E(N_{izt}) = N_{it}\ S_{zt} \qquad (1A)$$

such that:

$$N_{it} = \sum_{z=1}^{n} Nizt \qquad (2A)$$

$$S_{zt} = \frac{Nzt}{Nt} = \frac{\sum_{i=1}^{m} Nizt}{\sum_{i=1}^{m}\sum_{z=1}^{n} Nizt} \qquad (3A)$$

where N represents the number of patents, subscripts i, z and t concern region i, technology z and year t, and m and n are the total number of regions and technologies existing in my sample. $E(N_{izt})$ is the expected number of patents for each region i, technology z and year t, N_{it} represents the total number of patents in region i at time t, and S_{zt} refers to the world (all regions) share of patents for technology z at time t.

Step 2: Computation of expected number of specializations for each region i in year t

$$E(Spec_{it}) = \sum_{z=1}^{n} [\text{E(Nizt) E(SpecRate zt)}]' \qquad (4A)$$

such that:

$$E(SpecRate_{zt}) = \frac{\text{NewSpeczt}}{Nzt} = \frac{\sum_{i=1}^{m} \text{NewSpecizt}}{\sum_{i=1}^{m} \text{Nizt}} \qquad (5A)$$

$$\text{[E(Nizt) E(SpecRate zt)]}' = 1 \text{ if E(Nizt) E(SpecRate zt)} \geq 1$$

(6A)

$$\text{[E(Nizt) E(SpecRate zt)]}' = 0 \text{ if E(Nizt) E(SpecRate zt)} < 1$$

where $E(Spec_{it})$ is the expected number of specializations in region i and year t. $E(SpecRate_{zt})$ is the ratio between the number of specializations in technology z at time t, considering all regions included in the sample, and the number of patents in technology z at time t, again taking into account all regions included in the sample of observations. Condition (6A) is added to take into account the fact that the specialization status of a given region in a given technology z ([E(Nizt) E(SpecRate zt)]') is a dummy variable that takes the value 1 if a given region i is specialized in z, and 0 otherwise.

Step 3: Computation of number of specializations caused by existing differences between the technological structure of each region and the average technological structure prevailing in all regions

$$Dif_{it} = \sum_{z=1}^{n}[\text{Nizt E(SpecRate zt)}]' - \text{E(Specit)} \quad (7A)$$

such that:

$$[\text{Nizt E(SpecRate zt)}]' = 1 \text{ if Nizt E(SpecRate zt)} \geq 1$$

(8A)

$$[\text{Nizt E(SpecRate zt)}]' = 0 \text{ if Nizt E(SpecRate zt)} < 1$$

where Dif_{it} represents the number of specializations in region i and year t that are attributed to difference existing between the technological structure of region i at t, and the average technological structure prevailing in all regions in year t.

Step 4: Computation of number of technology-adjusted specializations in each region i in year t

$$\text{AdjustedSpec}_{it} = \sum_{z=1}^{n} Specizt \text{ - Dif}_{it} \tag{9A}$$

such that:

$$\text{Spec}_{izt} = 1 \text{ if } \text{RCA}_{izt} > 1$$

$$(10A)$$

$$\text{Spec}_{izt} = 0 \text{ if } \text{RCA}_{izt} \leq 1$$

where AdjustedSpec_{it} represents the total number of technology-adjusted specializations in region i and year t, and Spec_{izt} is dummy variable that reflects the specialization status of a given region i in technology z in year t. It takes the value 1 if a given region i has RCA (Revealed Comparative Advantage) greater than 1, and 0 otherwise.

Step 5: Computation of the technology-adjusted diversification potential for each region i in year t

$$\text{AdjustedDivPotential}_{it} = \text{n - AdjustedSpec}_{it} \tag{11A}$$

where $\text{AdjustedDivPotential}_{it}$ is the sector-adjusted diversification potential of a given region i in year t.

Appendix C: Computation of control variables

- **Technological diversity**[1]

To evaluate the existing technological diversity in a given region, this article proposes computing the inverse of the Herfindhal Index:

$$\text{Diversity}_{it} = \frac{1}{\sum_{z=1}^{n} \frac{Nizt Nizt}{Nit\ \ Nit}} \qquad (12A)$$

where N represents the number of patents, the subscripts i, z and t concern region i, technology z and year t, and n is the total number of technologies existing in the sample. Diversity_{it} represents the technological diversity of region i in year t. As this work will conduct inter-regional comparisons over time, and the technological structures of the regions vary considerably, it is useful to normalize (12A) according to the following formula:

$$\text{NormalizedDiversity}_{it} = \frac{\text{Diversityit}-1}{n-1} \qquad (13A)$$

This normalization guarantees that the values for diversity index will always be within the range [0, 1], with 0 meaning no diversity at all, and 1 representing a perfectly diversified technological structure. Total diversity means that in region i and year t, the total number of patents is equally distributed across all possible technological fields (n). Conversely, if diversity for region i and year t takes the value 0, then for that region all patents are concentrated just in one technological field. This analysis uses always the normalized version.

- **EU support: Regional investments from the European Regional Development Fund (ERDF) and the Cohesion Fund (CF)**

[1] *Diversity* is distinct from *diversification*. Whereas the latter is dynamic and evaluates to what extent new varieties emerge, the former is static and measures the existing variety of technologies within a regional technological system, at a given timepoint.

DG Regio provides data on regional ERDF and CF investments 2000–2013. However, such data for each NUTS 2 are disaggregated not by year, but by programming period (2000–2006 and 2007–2013). Moreover, as noted in Ciffolilli et al. (2015)[2] the best way to compare data from the two programming periods is to use investment commitments, not actual expenditures.[3] Therefore, for each region[4] and programming period it is collected data on investment commitments by category,[5] as well as total investment commitments. As the sample covers the period 2000–2013, it is assumed that the EU regional investments from ERDF and CF from each programming period essentially affect a given region in those periods, in which most years coincide with the years included in a given programming period. Here this article assumes that the first five periods of the dataset are affected by the programming period 2000–2006, and the final five periods affected mainly by the programming period 2007–2013. Therefore, the data on EU regional investment commitments for a given region are time-invariant for the first and last five periods included in the analysis. The only source of time variation can be found *between* the first five periods and the five last ones. As data on commitments are available only in nominal terms, this work follows Rodriguez-Pose & Fratesi (2004)[6] and compute the percentage of such nominal commitments as a share of the cumulated GDP PPS at current market prices over each programming period.

- **Other controls: Eurostat regional statistics**

[2] Ciffolilli, A., Condello, S., Pompili, M. & Roemisch, R. (2015). Geography of expenditure. Final report, Work Package 13: Ex post evaluation of Cohesion Policy programmes 2007–2013, focusing on the European Regional Development Fund (ERDF) and the Cohesion Fund (CF). Directorate-General for Regional and Urban Policy, European Commission

[3] This is because 2000–2006 expenditure data were not available, and payments were estimated using the absorption rates by country and fund. I believe that commitments can work as a good proxy for the total EU support received by a given region.

[4] For Denmark, only NUTS 1 data were available. NUTS 2 data were estimated by attributing to each region the same share of the funding by category as attributed for the programming period 2007–2013.

[5] Business Support, Energy Environment and Natural Resources, Human Resources, IT Infrastructure and Services, Other, Research and Technology, Social Infrastructure, Technical Assistance, Tourism & Culture, Transport Infrastructure, and Urban and Rural Regeneration

[6] Rodriguez-Pose, A. & Fratesi, U. (2004). Between development and social policies: The Impact of European Structural Funds in Objective 1 Regions, Regional Studies, 38, 97–113

Although it is possible to use Eurostat data directly, Eurostat Regional Statistics involve some missing values for several years in certain regions. To overcome this data shortcoming, whenever possible I compute missing values using one of the following procedures, in the following order:

- For a given NUTS 2 where data are missing for year t, I compute the ratio between the value for the nearest year after t (t+x) for which data are available at NUTS 2 level and the NUTS 1 value for that year (t+x). This ratio is then multiplied by the NUTS 1 value for the year (t) for which NUTS 2 data are missing;
- For a given NUTS 2 where data are missing for year t, I compute the ratio between the value for the nearest year after t (t+x) for which data are available at NUTS 2 level and the NUTS 0 value for that year (t+x). This ratio is then multiplied by the NUTS 0 value for the year (t) for which NUTS 2 data are missing;
- For a given NUTS 2 where data are for year t, I attribute the same value as t-1;
- For a given NUTS 2 where data are missing for t, I attribute the same value as t+1;
- For a given NUTS 2 where data are missing for year t, I attribute the same value of the NUTS 1 to which the NUTS 2 belongs;
- For a given NUTS 2 where data are missing for year t, I attribute the same value of the NUTS 0 to which the NUTS 2 belongs

Appendix D: Correlation Matrix

Table D1. Correlation Matrix

	N new specializ.	Diversification	CooperationWithin	CooperationBetween	Diversity	EU rsupport	Un_rate	Tertiary_edu	GDPpc	R&D
N new specializ.	1.00									
Diversification	0.91 ***	1.00								
CooperationWithin	-0.13 ***	-0.20 ***	1.00							
CooperationBetween	-0.19 ***	-0.24 ***	0.07 ***	1.00						
Diversity	0.37 ***	0.47 ***	-0.23 ***	-0.22 ***	1.00					
EU support	-0.26 ***	-0.33 ***	0.29 ***	0.18 ***	-0.31 ***	1.00				
Un_rate	-0.15 ***	-0.20 ***	0.24 ***	0.19 ***	-0.23 ***	0.28 ***	1.00			
Tertiary_edu	0.20 ***	0.25 ***	-0.22 ***	-0.12 ***	0.11 ***	-0.30 ***	-0.17 ***	1.00		
GDPpc	0.22 ***	0.33 ***	-0.30 ***	-0.19 ***	0.29 ***	-0.44 ***	-0.42 ***	0.51 ***	1.00	
R&D	0.17 ***	0.29 ***	-0.25 ***	-0.15 ***	0.20 ***	-0.30 ***	-0.20 ***	0.41 ***	0.40 ***	1.00

* p<0.1, ** p<0.05, *** p<0.01

Source: author's computations

Appendix E: Comparison of the Difference GMM estimation results for different number of instruments

Table E1. Estimation results –Difference GMM using all suitable lags as instruments and Difference GMM using only second lag instruments

	Dependent variable = Ln(N new specializations)						Dependent variable = Ln(Diversification/(1-Diversification))					
	Difference GMM (all suitable lags used as instruments)			Difference GMM (only second lag instruments)			Difference GMM (all suitable lags used as instruments)			Difference GMM (only second lag instruments)		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
Coop Within	-1.320*	-3.387**	-5.427***	-2.210**	-4.814**	-8.442***	-1.358*	-3.660**	-6.129***	-2.746***	-5.445**	-10.099***
	(0.727)	(1.445)	(1.751)	(0.882)	(1.994)	(2.640)	(0.764)	(1.677)	(1.939)	(0.912)	(2.275)	(2.930)
Coop Between	-0.472	-0.633	-0.872	-0.761	-1.371	-1.882*	-0.615	-0.800	-1.145**	-1.170	-1.855*	-2.427**
	(0.442)	(0.631)	(0.544)	(0.712)	(0.955)	(1.051)	(0.474)	(0.648)	(0.573)	(0.756)	(1.066)	(1.179)
Diversity	-1.317***	-1.479***	-1.508***	-0.674	-0.937**	-1.027**	-0.594	-0.819**	-0.889**	-0.050	-0.220	-0.383
	(0.369)	(0.346)	(0.340)	(0.426)	(0.445)	(0.441)	(0.403)	(0.404)	(0.380)	(0.488)	(0.531)	(0.523)
EU support	7.157	5.994	6.375	7.462	4.880	4.198	7.895	7.210	7.108	6.857	5.721	4.510
	(5.128)	(4.938)	(4.888)	(4.943)	(4.652)	(4.471)	(5.519)	(5.302)	(5.154)	(5.244)	(5.010)	(4.998)
Unemployment	-0.005	-0.001	-0.001	0.004	0.002	0.006	-0.008	-0.004	-0.004	0.007	0.004	0.008
	(0.007)	(0.006)	(0.006)	(0.007)	(0.008)	(0.009)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.010)
Higher Education	0.016*	0.014*	0.013*	0.007	0.003	0.006	0.018*	0.019**	0.017*	0.006	0.003	0.007
	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)
GDPpc	-0.004	-0.009	-0.009	-0.003	-0.021	-0.020	-0.003	-0.008	-0.005	-0.005	-0.023	-0.022
	(0.012)	(0.013)	(0.012)	(0.012)	(0.016)	(0.018)	(0.013)	(0.015)	(0.012)	(0.013)	(0.019)	(0.020)
R&D	-0.009	-0.009	-0.014	-0.014	-0.012	-0.014	-0.003	0.002	-0.008	-0.012	-0.012	-0.013
	(0.010)	(0.014)	(0.016)	(0.016)	(0.015)	(0.016)	(0.012)	(0.016)	(0.015)	(0.019)	(0.017)	(0.018)
GDPpc*Coop Within		0.189**	0.234***		0.210	0.354**		0.197**	0.253***		0.219	0.402**
		(0.087)	(0.084)		(0.136)	(0.171)		(0.099)	(0.097)		(0.151)	(0.185)
GDPpc*Coop Between		-0.009	-0.012		0.052	0.025		-0.005	-0.012		0.062	0.025
		(0.035)	(0.032)		(0.057)	(0.060)		(0.037)	(0.034)		(0.069)	(0.073)
Coop Within*Between			7.908**			13.710***			9.378**			16.914***
			(3.722)			(5.255)			(3.670)			(5.746)
Lagged Dependent	0.067	0.044	0.031	0.052	0.041	0.019	0.080*	0.053	0.041	0.084*	0.063	0.029
	(0.042)	(0.040)	(0.041)	(0.040)	(0.040)	(0.041)	(0.042)	(0.041)	(0.040)	(0.057)	(0.044)	(0.046)
Time fixed effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
N observations	1808	1808	1808	1808	1808	1808	1808	1808	1808	1808	1808	1808
AR(1) - pvalue	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2) - pvalue	0.362	0.565	0.571	0.308	0.538	0.535	0.254	0.458	0.456	0.120	0.327	0.384
Hansen Test - pvalue	0.199	0.709	0.982	0.445	0.533	0.502	0.276	0.618	0.978	0.499	0.505	0.460
N instruments	165	237	273	45	61	69	165	237	273	45	61	69

* p<0.1, ** p<0.05, *** p<0.01

Source: author's computations

Table E2. Marginal effects of cooperation within regions and between regions, on the dependent variable, for differing levels of GDPpc

	Ln(N new specializations)				Ln(Diversification/(1-Diversification))				
Independent Variable	Diff. GMM (all suitable lags)		Diff. GMM (only second lag)		Diff. GMM (all suitable lags)		Diff. GMM (only second lag)		Assumption for the interaction term
Cooperation within	-2.46	**	-3.78	***	-2.70	**	-4.37	***	GDP pc = min
	0.24		-0.78		0.12		-1.24		GDP pc = Q1
	1.03		0.10		0.95		-0.32		GDP pc = Q2
	1.84		1.01		1.79		0.62		GDP pc = Q3
	12.30	**	12.65		12.71	*	12.76		GDP pc = max
Cooperation between	-0.68		-1.12		-0.82		-1.55	*	GDP pc = min
	-0.81	**	-0.37		-0.89	**	-0.66		GDP pc = Q1
	-0.84	*	-0.16		-0.91	*	-0.40		GDP pc = Q2
	-0.88		0.07		-0.93		-0.14		GDP pc = Q3
	-1.38		2.94		-1.20		3.30		GDP pc = max

* p<0.1, ** p<0.05, *** p<0.01

Source: author's computations

Table E3. Marginal effects of cooperation within regions and between regions, on the dependent variable, for differing levels of interaction terms

	Ln(N new specializations)				Ln(Diversification/(1-Diversification))					
Independent Variable	Diff. GMM (all suitable lags)		Diff. GMM (only second lag)		Diff. GMM (all suitable lags)		Diff. GMM (only second lag)		Assumption for the interaction terms	
Cooperation within	-4.28	***	-6.71	***	-4.89	***	-8.13	***	GDP pc = min	Cooperation between = min
	-0.93		-1.65		-1.27		-2.37		GDP pc = Q1	
	0.05		-0.17		-0.20		-0.68		GDP pc = Q2	
	1.06		1.351		0.88		1.05		GDP pc = Q3	
	14.03	**	20.94	*	14.92	**	23.35	*	GDP pc = max	
Cooperation within	-3.32	***	-5.04	***	-3.74	***	-6.06	***	GDP pc = min	Cooperation between = Q2
	0.03		0.018		-0.12		-0.31		GDP pc = Q1	
	1.02		1.503		0.94		1.382		GDP pc = Q2	
	2.02	*	3.023		2.03		3.113		GDP pc = Q3	
	15.00	***	22.61	*	16.06	**	25.42	*	GDP pc = max	
Cooperation within	2.26		4.636		2.87	*	5.87	*	GDP pc = min	Cooperation between = max
	5.61	**	9.691	**	6.49	**	11.63	**	GDP pc = Q1	
	6.60	**	11.18	**	7.56	***	13.32	**	GDP pc = Q2	
	7.60	***	12.7	**	8.65	***	15.05	**	GDP pc = Q3	
	20.58	***	32.29	**	22.68	***	37.35	**	GDP pc = max	
Cooperation between	0.27		-0.15		0.10		-0.45		GDP pc = min	Cooperation within = min
	3.62	**	4.906		3.72	*	5.301		GDP pc = Q1	
	4.61	**	6.39	*	4.78	**	6.991	*	GDP pc = Q2	
	5.61	**	7.911	*	5.87	**	8.722	*	GDP pc = Q3	
	18.59	***	27.5	**	19.90	**	31.02	**	GDP pc = max	
Cooperation between	0.53		0.288		0.40		0.086		GDP pc = min	Cooperation within = Q2
	3.88	**	5.344	*	4.02	**	5.841	*	GDP pc = Q1	
	4.86	**	6.829	*	5.08	**	7.532	*	GDP pc = Q2	
	5.87	**	8.349	*	6.17	**	9.262	*	GDP pc = Q3	
	18.84	***	27.94	**	20.20	**	31.56	**	GDP pc = max	
Cooperation between	4.23	**	6.705	**	4.79	**	8.002	**	GDP pc = min	Cooperation within = max
	7.58	***	11.76	**	8.41	***	13.76	**	GDP pc = Q1	
	8.56	***	13.25	**	9.47	***	15.45	**	GDP pc = Q2	
	9.57	***	14.77	**	10.56	***	17.18	**	GDP pc = Q3	
	22.54	***	34.36	**	24.59	***	39.48	**	GDP pc = max	

* p<0.1, ** p<0.05, *** p<0.01

Source: author's computations