ON-LINE ONLY

ON LINE APPENDIX A – Extensions and Robustness checks

A.1 Quantile Regression analysis

This extension is based on Quantile Regression (QR) Techniques. As discussed in the paper QR has a number of advantages over standard linear regression analysis. However, when QR is combined with Fixed Effect panel data in order to control for unobserved heterogeneity constant over time (a primary concern in a policy assessment framework), its identification and estimation become very complex (Kato et al. , 2012). In particular, when the number of observations on each individual/region is limited on account of the limited number of available time periods, it is difficult to allow for the effect of an individual FE to change across quantiles in the same way as we can allow for the effects of the X covariates. This difficulty stems from the fact that the standard methods used to cancel out FE are no longer applicable: the quantile of the difference in general is not equal to the difference in quantiles but instead become 'intractable objects' (Ponomareva, 2011).

Most of the literature that studies QR models for panel data with FE tries to deal with this difficulty by assuming that the number of periods t reaches infinity with sample size n and then considers individual heterogeneity a "pure locations shift effect" on conditional quantiles (Canay, 2010; Koenker, 2005) or by allowing it to vary across quantiles (Galvao, 2008). Instead, in relation to a relatively short panel, an attempt to estimate QR has been made by applying correlated random coefficients model (Abrevaya and Dahl, 2008), or by focusing on the identification of the coefficients for a single conditional quantile restriction rather than on the whole set of quantiles (Rosen, 2009) or even by estimating the moment of the conditional distribution of either continuous or discrete covariates (Ponomareva, 2011).

Nevertheless, most empirical QR applications prefer a cross-section framework for analysis (Buchinsky, 1994; Powell, 2011; Powell and Wagner, 2011). In particular, pooled OLS models that regress the regional growth rate at time *t* on the policy and on the other usual covariates at time *t-1* are implemented: the average annual growth rate over the period 2007-2009 is regressed onto the Regional Policy Commitments and onto the other covariates related to the 2000-2006 period and the average annual growth rate over the period 2000-2003 is regressed onto the Regional Policy Commitments and other covariates related to the period 1994-1999. As stated earlier, QR analysis focuses on the 0.10, 0.50 and 0.75 quantiles of the Y distribution.

Table A.1 - Quantile Regression

	Quantile 0.10		Quantile 0.50		Quantile 0.75		Mean regression	
Dependent variable: average regional GDP growth rate 2007-2009								
Initial	-0.0550	0.0320	0.0060	-0.0002	0.0028	-0.0002	0.000	0.0001
condition	(0.0677)	(0.0273)	(0.0058)	(0.0088)	(0.0055)	(0.0092)	(0.0127)	(0.0149)
Regional	0.0211	0.0020	0.0057**	0.0048	0.0050**	0.0048	0.0096*	0.0090
Policy	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Covariates*	No	Yes	No	Yes	No	Yes	No	Yes
Dependent variable: average regional GDP growth rate 2000-2003								

Initial	-0.0072	-0.0144	-0.0021	-0.0149**	-0.0091	-0.0093	-0.0028	-0.0070
condition	(0.0139)	(0.0122)	(0.0059)	(0.0066)	(0.0064)	(0.0087)	(0.0047)	(0.0057)
Regional	0.0051	-0.0038	0.0137***	0.0106**	0.0090**	0.0119*	0.0114***	0.0091**
Policy	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Covariates *	No	Yes	No	Yes	No	Yes	No	Yes

*Covariates included in the model are the variables of the Territorial conditioning factors matrix and of the EU policy matrix plus the Control variables.

** Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

A.2 - Measurement Error in the policy variable and Endogeneity

The additional dataset that makes it possible to test for potential measurement errors in the policy variable includes actual payments for all NUTS-2regions in the EU27 and includes annual Commitments and Payments for the EU Regional Policy and Rural Development (together and not separable) over the period 2000-2009 only. As customary in the economic growth literature the test avoids annual GDP data to measure economic growth and relies on 5-year periods (OECD, 2009a), annual policy data from 2000 to 2009 are aligned to this, and the model was estimated as a cross section, where regional GDP pro capita growth rate over the period 2007-2009 is regressed onto the 'spatially targeted' policies' payments over the 2000-2009 period. The same model was estimated by making use of Commitment data from the previous analysis by linking the regional growth rate of GDP per capita over the period 2007-2009 with 'spatially targeted' policies' Commitments for the last two programming periods (2000-2006 and 2007-2013) in order to maximise comparability.

The results of the model estimated on both datasets are shown in Table A.2 below. In particular, the first column of the table shows the results obtained by running the model on the main dataset of the analysis (and as a consequence policy data refer to the whole period of Commitments). The second column shows the corresponding results obtained by running the model on the actual payments dataset. Finally, column 3 sets out the results obtained by considering Payments as endogenous and, consequently, instrumented by the corresponding Commitments an Instrumental Variable framework.

The check conducted on the main dataset confirms the impact of 'spatially targeted' policies on regional growth: i.e. the coefficient of the 'spatially targeted' policies is positive and significant.

The policy variable coefficient in column 2 is positive but not significant. However, once the endogeneity of actual payment is accounted for in column 3 by means of an appropriate IV strategy the key results of the paper are confirmed. In addition the results of this IV analysis confirm that the key conclusions of the paper are robust to endogeneity bias.

Table A.2 - Measurement Error in the Policy Variable and Endogeneity

Dependent variable: GDP per capita Average Growth Rate			
	1 - OLS	2 - OLS	3 - IV
Spatially Targeted Policies	0.0004***	0.00001	0.000001*
	(0.000)	(0.000)	(0.000)
Constant	-0.732***	-0.026***	-0.026***
	(0.1563)	(0.003)	(0.003)
N of Regions	139	198	198
R-squared	0.149	0.170	0.170
Prob>F	0.000	0.000	0.000

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

A.3 - Measurement Error in the outcome variable

Table A-3.a - Measurement Error in the outcome variable (Total EU Expenditure)

Dependent variable: GDP per capita Average Growth Rate		
	1	2
Total EU Funding	0.0093*	0.0042
	(0.0047)	(0.0034)
Ln of initial GDP p.c.	-0.1484***	-0.1493***
	(0.0270)	(0.0284)
Social Filter Index	0.0014	-0.0038
	(0.0040)	(0.0037)
R&D Activities	0.0030	0.0020
	(0.0051)	(0.0044)
Infrastructural endowment	0.2808	-0.0188
	(0.3082)	(0.3340)
Spatially Lagged Social Filter		0.0046
		(0.0041)
Spatially lagged R&D Activities		0.0225***
		(0.0037)
Spatially lagged Infrastructure		0.1500
		(0.5357)
National Growth Rate	0.0356***	0.0410***
	(0.0050)	(0.0046)
Krugman Index	0.0010	-0.0020
	(0.0082)	(0.0092)
Population Density	0.0001**	0.0001**
	(0.0000)	(0.0000)
Constant	1.3974***	1.3861***
	(0.2636)	(0.2751)
Obs	242	242
R squared	0.870	0.899
Prob>F	0.000	0.000

Robust and clustered standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.10

Table A.3.b - Measurement Error in the outcome variable (Regional Policy, Rural Development Policy and CAP).

Dependent variable: GDP per capita Average Growth Rate

	1	2	3	4	5
Regional Policy	0.0277***	0.0223**	0.0038***	0.0355***	0.0003
	(0.0066)	(0.0091)	(0.0085)	(0.0070)	(0.0206)
Rural Development Policy	0.0093	0.0031	-0.0028	-0.0078	0.0412**
	(0.0085)	(0.0066)	(0.0113)	(0.0120)	(0.0201)
CAP	-0.0198***	-0.0115*	-0.0120	-0.0173	-0.0303***
	(0.0056)	(0.0068)	(0.0095)	(0.0070)	(0.0089)
Social Filter Index*Regional Policy		-0.0128**			
		(0.0054)			
Social Filter Index*Rural Development Policy		0.0215*			
		(0.0115)			
Social Filter Index*CAP		-0.0048*			
		(0.0026)			
R&D Activities*Regional Policy			-0.0096		
			(0.0085)		
R&D Activities*Rural Development Policy			0.0104		
			(0.0083)		
R&D Activities*CAP			-0.0055		
			(0.0057)		
Infrastructure*Regional Policy				-1.2217***	
				(0.3197)	
Infrastructure*Rural Development Policy				1.7491***	
				(0.4157)	
Infrastructure*CAP				-1.4444***	
				(0.2981)	
Regional Policy* Rural Development Policy					0.0145
					(0.0221)
Regional Policy*CAP					0.0111**
					(0.0051)
Rural Development Policy*CAP					-0.0187
					(0.0133)
Log of Initial GDP, 'Territorial Conditioning Factors', 'Spatially Lagged terms', Controls and constant	X	X	X	X	X
Period Dummies	X	X	X	X	X
Obs	242	242	242	242	242
R squared	0.919	0.927	0.921	0.939	0.924
Prob>F	0.000	0.000	0.000	0.000	0.000

Robust and clustered standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 - The log of initial GDP, 'territorial conditioning factors' (Social Filter Index, R&D Activities, Infrastructural endowment Spatially Lagged Social Filter, Spatially lagged R&D Activities, Spatially lagged infrastructure) and the same control variables (Constant; National Growth Rate; Krugman Index and Population Density) reported in Table 2.a are included in all regressions but not reported in the table.

A.4 - ***Spatial Heterogeneity and Spatial Panel Data Analysis***

The recent spatial econometrics literature has identified three different types of interaction effects that could affect local economic phenomena and consequently their analysis: endogenous interaction effects linked to the dependent variable (Y), exogenous interaction effects among independent variables (X) of the units of analysis, and interaction effects among error terms. Hitherto, by inserting spatially-lagged independent variables (spatially-lagged variables matrix) into the key model specification the regressions only included controls for the second types of these spatial interactions. However, the analysis implemented in the paper has not yet fully controlled for the spatial dependence of the dependent variable and error terms. These forms of spatial dependence can be treated in a panel data framework. By accounting for the unobservable spatial and time-period specific effects, the panel data and spatial econometric literatures offer a common setting, enabling us to account for the cross-sectional and state dependence of the Y and the Xs, while at the same time controlling for unknown heterogeneity. We can also account for them simultaneously through 'Spatial Dynamics Panel Data Models' (SDPDM). Such models can easily identify the dynamic responses over time and space of the space-time diffusion of policy impacts through cross-partial derivatives related to changes in the explanatory variables and in the dependent variables (Elhrost, 2005). Once the need to account for spatial dynamics has been identified, the most serious issue seems to be the identification, among the Spatial Panel Data Models, of that model that can best capture and represent the spatial dependence of the data. Some analyses of European regional convergence processes have found evidence of model misspecification if the spatial interdependencies of regional growth are ignored. The most common approaches that address the issue of spatial dependence (Anselin, 2006) adopted in the existing literature refer to 'spatial error autocorrelation' (Arbia and Piras, 2009) and 'spatial lag' models. The latter, often considered a spatial autoregressive model, would seem to be more appropriate for quantifying how a region's growth rate is affected by the growth rate in surrounding regions (Anselin, 2006). The addition of a spatially-lagged dependent variable ('spatial lag; models), however, causes simultaneity and endogeneity problems that GMM (Badinger et al., 2004) and maximum likelihood (Elhorst, 2005) methods can address.[1] As in classical panel data literature, a fixed-effects model is largely preferred (Elhrost, 2005) because the unobserved component is allowed to depend on the other regressors included in the model.

Within this FE spatial panel data framework, this section extends the main analysis of the paper by allowing the model (1) to account, in addition to the spatial dependence of the Xs, for Y and for error-term dependencies. For this purpose, model (1) will assume three additional specifications (SAR, DURBIN and SEM) and the results provided by the estimation (via maximum likelihood) of each of them will be analysed in a comparative sense in order to a) decide the best way to model the spatial dependence of the phenomena analysed and b) test if the results of the main analysis are robust, given the overall spatial dependence of the phenomena under analysis.

In this manner the Spatial Autoregressive (SAR and DURBIN) specifications of model (1) will account for the spatial dynamics of the dependent variable that estimates the spatially lagged Y (Spatial lag models) coefficient. The Spatial Error Model (SEM) will, instead, account for the dependence determining the

1 In this sense, a variety of estimators have been recently proposed by the literature: Yu et al. (2008) and Lee and Yu (2010) provide the asymptotic properties of a quasi-maximum likelihood for an SDPD model with exogenous explanatory variables. More recently, Korniotis (2010) proposed a solution based on the Least Square Dummy Variable and instrumental methods (Anderson and Hsiao, 1982) extended to allow for the spatial effect.

spatially inter-correlation between the error terms (LeSage & Pace, 2009). Among the Spatial Autoregressive models, DURBIN could be understood as a special case of SAR as besides including the spatially lagged Y it also includes other exogenous spatially-lagged regressors. The choice of the regressors is unconstrained: both Xs and additional variables could be inserted in their spatial lag version. On the basis of the results reported in the literature, the DURBIN version of the model is considered the most appropriate and informative for regional analysis insofar as it is a "Spatial lag" specification that, moreover, makes it possible to control for Xs spatial dependence (LeSage & Pace, 2009).

As already seen for the previous robustness checks, both the data's and the model's structure needs to be adjusted to take due account of the setting in which the robustness check is to be performed, which, in this case, is the framework provided by the spatial panel data model.

In this sense, the panel was reset to comprise two periods: for the first period the independent variables refer to the first period of the main analysis (policy programming period 1994-99) whereas the dependent variable is the GDP Growth rate in the second period of the main analysis (2000-06). For the second period, the data used for the regressors refer to the period 2000-06 whereas the outcome variable is that used in the third period of the main analysis (2007-13). By performing the analysis on such a panel, we deploy explanatory variables with a one-period-lag with respect to the dependent variable, even if the SPDM framework lags prevents us from taking lags directly into account in estimating a model.

Results from the SAR, DURBIN and SEM models, presented in Table A.4 below, refer to the version of model (1) estimated by considering Regional Policy, Rural Development Policy and CAP separately. The estimated models includes all regional conditioning factors, spatially lagged terms and controls included in the main results presented in Tables 2 and 3 of the paper. The analysis was carried out by implementing the STATA routine "XSMLE" (Hughes et al., 2012) and using a "Rook Contiguity" matrix as a spatial weight.

Table A.4 shows that the spatially lagged Y coefficient is never significant. Spatial influences on regional growth rates seem to be fully accounted for by the spatial correlation among the explanatory values (already included in the main specification of the model reported in tables 2 and 3 in the paper) while the endogenous spatial dependence in terms of Y seems to be irrelevant. This robustness check highlights that the main analysis has already accounted for the overall spatial dependence characterising regional growth. Even by accounting for the additional and potentially strong source of spatial dependence related to Y, the results obtained by the main model do not change. The findings on the main coefficient of interest (Regional Policy) are all confirmed. Moreover, the signs of the other explanatory variables are also generally confirmed, albeit with a different level of significance.

The results from the three different models (SAR, DURBIN and SEM) are coherent with each other. For each variable the coefficients used always have the same signs. By making comparisons between them, the different ways of modelling spatial dependence are shown to lead to similar conclusions. The SEM model, which accounts for the spatial dependence affecting the regression's residuals, leads to very similar results with respect to those (SAR and DURBIN) provided by directly accounting for the spatial dependence of Y.

Table A.4 - Spatial Panel data models

Dependent variable: GDP per capita Average Growth Rate			
	SAR	DURBIN	SEM
Spatially lagged Y	-0.139646	-0.1973682	
	(0.1807)	(0.1630)	
Ln of initial GDP p.c.	0.0706273	0.1195099*	0.076037
	(0.1041)	(0.0677)	(0.0617)
Regional Policy	0.0001165**	0.0001183***	0.0001288***
	(0.0000)	(0.0000)	(0.0000)
Rural Development Policy	0.0000284	0.0000352	0.0000141
	(0.0000)	(0.0000)	(0.0000)
CAP	-0.0000312	-0.0000361	-0.0000378
	(0.0000)	(0.0000)	(0.0000)
Social Filter Index	-0.01765	-0.0177591	-0.0160019
	(0.0143)	(0.0112)	(0.0109)
R&D Activities	0.0378418	0.0390929**	0.0321667*
	(0.0309)	(0.0193)	(0.0189)
Infrastructural endowment	2.713875*	3.237553**	2.989705**
	(1.4307)	(1.3450)	(1.3417)
Spatially lagged Social Filter Index		-0.0610526*	
		(0.0372)	
Spatially lagged R&D Activities		-0.1299134	
		(0.1006)	
Spatially lagged Infrastructure		4.652283	
		(6.1640)	
National Growth Rate	0.1716416***	0.1646388***	0.1663109***
	(0.0231)	(0.0128)	(0.0122)
Krugman Index	0.1649753***	0.1686406***	0.1768534***
	(0.0501)	(0.0317)	(0.0319)
Population Density	0.0000243***	0.0000231***	0.000027***
	(0.0000)	(0.0000)	(0.0000)
Obs	242	242	242
R squared	0.157	0.108	0.144

** Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

A.5 - *Additional References cited in the Appendix*

Abrevaya, J., and C. M. Dahl (2008): "The Effects of Birth Inputs on Birthweight: Evidence From Quantile Estimation on Panel Data," *Journal of Business and Economic Statistics*, 26(4), 379-397.

Anderson, T.W. and C. Hsiao (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.

Anselin, L. (2006): "Spatial econometrics," in *Palgrave handbook of econometrics*, ed. by T. Mills and K. *Patterson, vol. 1, 901-969. Palgrave Macmillan, Basingstoke.*

Arbia, G. and Piras, G. (2009): "A new class of spatial concentration measures," *Computational Statistics & Data Analysis, Elsevier*, 53(12), 4471-4481.

Badinger, H., W. Muller and G. Tondl (2004): "Regional convergence in the European Union 1985-1999: A spatial dynamic panel analysis," *Regional Studies*, 38(3), 241-253.

Canay, I. (2010): "A Note on Quantile Regression for Panel Data Models," *working paper, Northwestern University*.

Chernozhukov, V., and C. Hansen (2004): "The Impact of 401(K) on Savings: An Instrumental Quantile Regression Analysis," *Review of Economics and Statistics* 86(3), 735 - 751.

Elhrost, J. P. (2005): "Models for dynamic panels in space and time. An application to regional unemployment in the eu," *Paper prepared for 45th meetings of the European Regional Science Association* in Amsterdam, 23-27 August 2005.

Galvao, A. (2008): “Quantile Regression for Dynamic Panel Data with Fixed Effects," *working paper*, *University of Wisconsin-Milwaukee*.

Hughes G., Piano Mortari A. and F. Belotti (2012): “Implementing procedures for spatial panel econometrics in Stata.” *Mimeo*.

Jaffe, A. B. (1989): “Characterizing the "technological position" of firms with application to quantifying technological opportunity and research spillovers," *Research Policy*, Elsevier, 18(2), 87-97.

Kato, K., Galvao, A. and G.V. Montes-Rojas (2012): "Asymptotics for panel quantile regression models with individual effects," *Journal of Econometrics* 170 (1), 76-91.

Koenker, R. (2005): Quantile Regression. Econometric Society Monograph, Cambridge University Press, Cambridge.

Korniotis, G. (2010): “Estimating Panel Models With Internal and External Habit Formation,” *Journal of Business & Economic Statistics*, 28 (1), 145–158.

Lee, L-F. and J. Yu (2010): “A Spatial Dynamic Panel Data Model with both Time and Individual Fixed effects," *Econometric Theory*, 26, 564-597.

LeSage, J. P. and K. R. Pace (2009): Introduction to spatial econometrics. CRC Press/Taylor and Francis Group, London.

Rosen, A. (2009): “Identification via Quantile Restrictions in Short Panels," *working paper, UCL*.

Yu J, de Jong R and L-F. Lee (2008): “Quasi Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects when both n and T Are Large,” *Journal of Econometrics*, 146, 118-134.

ON LINE APPENDIX B – Data

B.1 - *Description of Variables*

Regional GDP Growth rate per capita (dependent variable): the growth rate of regional GDP is the dependent variable and is used as a proxy for regional economic performance. It is computed as the logarithmic ratio between average GDP per capita (expressed in PPP - Purchase Power Parity) for the first three years of the period t and the correspondent value for the period t-1. As is customary in growth analyses, GDP growth rate is hence computed over multiannual periods rather than on a yearly basis in order to minimize the influence of external macro trends and shocks (the robustness checks confirm that results are robust to different methodological choices in this regard).

'EU Policy matrix'

The role of EU policies in regional growth dynamics is captured by examining the corresponding expenditure in each region for the EU budget programming periods 1994-99; 2000-06 and 2007-13 for Regional Policy, Rural Development Policy ('spatially targeted' policies) and CAP ('spatially blind' policy with territorial implications), covering approximately 80% of total EU expenditure. With respect to the resources of the EU Regional Policy, the funds allocated to each region in each programming period depend on strict eligibility criteria that earmark a relevant share of the funds to the regions whose GDP per capita is below 75% of the EU average (named 'Objective 1 Regions' until 2006; 'Convergence Regions' in the 2007-2013 period; 'Less Developed Regions' in the 2014-2020 period).

'Territorial conditioning factors matrix'

This matrix aims to include the key territorial features that shape policy success and failure under the strong constraint of data availability for all EU regions. It includes structural socio-economic conditions in terms of demographics, productive structure and the labour market as well as regional innovative capacity and infrastructural endowment.

In particular, socio-economic conditions are captured by a Social Filter Index - a composite index extensively used in existing studies on innovation and regional growth (Crescenzi and Rodríguez-Pose 2009 and 2011) combining a set of proxies for territorial structural preconditions conducive to favourable environments for the genesis of innovation and its translation into economic growth. The Social Filter Index covers two main domains: educational achievements (Crescenzi, 2005; Iammarino, 2005; Lucas, 1988; Lundvall, 1992; Malecki, 1997; Rodríguez-Pose and Crescenzi, 2008) and the productive employment of human resources (Fagerberg et al., 1997; Rodríguez-Pose, 1999).

With reference to the first domain, the index accounts for human capital accumulation (share of tertiary educated population in relation to the population aged 15+) and skilled labour force (share of tertiary educated employees in relation to total employees). For the second domain, employment in agriculture is included in order to account for the composition of the local productive structure. The long-term component of regional unemployment (long-term unemployment percentage) is included in the index in order to account for the rigidity of local labour markets and the stratification of inadequate skills (Gordon, 2001) that hamper innovation and economic growth[2].

Other two important features influencing policy impacts are: a) the level of R&D activities (Share of R&D in Regional GDP) that "captures the existence of a system of incentives (in the public and the private sector) for intentional innovative activities" (Crescenzi and Rodríguez-Pose, 2011, p. 14); b) the level of regional infrastructural endowment (regional kilometers of motorways standardized by 'total regional surface'[3]) as a proxy for a region's existing physical capital endowment.

Interactions matrix:

This matrix includes two key types of interactions: interactions between the individual components of the 'EU policy matrix' – in order to capture synergies or trade-offs between different EU policies – and interactions between the 'policies' and the 'conditioning factors' matrices in order to identify factors conditioning policy impacts. The elements of this 'interactions matrix' can capture the existence of synergetic/countervailing forces able to influence policy impacts by augmenting or diminishing its magnitude. In particular, in line with the conditioned impact literature (Ederveen et al., 2002; Ederveen et al., 2006), the overall impact of the policy is evaluated by assessing the sign and joint significance of the coefficient of the policy itself (i.e. the coefficient of the variable of interest indicated in the 'policy matrix') and the coefficient of the term of interaction with the identified conditioning factors (i.e. the 'interaction matrix')[4]

Spatially lagged variables

In order to account for interactions between neighbouring regions, this additional matrix introduces the spatially lagged values of 'conditioning factors'. These values enable us to explicitly model spatially-mediated inter-regional spillovers while, at the same time, minimising the spatial autocorrelation of the residuals. In particular, the spatially lagged variables included in the model are calculated by multiplying each territorial variable by a spatial matrix computed with the k-nearest neighbours (with k=4) criterion,

[2] The index is calculated by using Principal Component Analysis (PCA) and accounts (considering only its first component) for around 50% of the total variance in the single variables that it synthetizes (Tables B.1 and B.2, Appendix B). It prevents collinearity problems potentially generated by the simultaneous inclusion of all the variables in the model (Duntenam, 1989; Esposti et al., 2013). The four variables considered enter the composite index with the expected sign: human capital and skilled labour force – which also displays the greatest relative weighting – have a positive sign, while long-term unemployment and the agricultural share of employment, by contrast, figure in the social filter index with a negative sign. The Index is computed for each year (time variant indicator) holding constant the PCA coefficients (computed on the longitudinal dataset). The stationarity of the variables was preliminarily tested: The tests confirmed the stationarity of the series, allowing us to implement the PCA analysis on the panel dataset and assure the comparability of the index across programming periods.

[3] The standardisation proposed is used in order to purge potential biases linked to the different geographical sizes of the EU regions. Even if this is the customary proxy used in the existing literature, it should be noticed that it is uninformative on the quality and condition of the infrastructures themselves and nor does it reflect differences in construction and maintenance costs.

[4] In the paper, in line with the existing literature on conditioned impact, we focus on the sign and significance of coefficients, rather than on the size of specific point estimates. In general, following Wooldridge (2003), the magnitude of the overall effect can be computed by plugging in interesting values of the interacted variable (e.g. the mean or the lower and upper quartiles in the sample) to obtain the partial effect.

which can minimize not only 'endogeneity' induced by travel-time distance weighting but also potential bias due to differences in the number of neighbours as between central and peripheral European regions. In particular, the 'spatially lagged matrix' includes the spatially lagged value of the social filter index, spatially lagged R&D activities and the spatially lagged infrastructural endowment.

These spatially lagged indicators place each region in the broader European space, making it possible to assess their interactions with neighbouring regions. They can capture spillovers of various kinds influenced by geographical accessibility or peripherality. Favourable socio-economic conditions in neighbouring regions (spatially lagged social filter index) influence indigenous economic performance through imitative effects and the mobility/movement of human capital/skills facilitated by geographical proximity. Accessibility to extra-regional innovative activities (spatially lagged R&D variable) can also influence internal economic performance through localised knowledge spillovers while the infrastructural endowment of neighbouring regions insures adequate accessibility to the region and the lack of transport bottlenecks.

<u>Control matrix</u>:

The 'initial conditions' of the regions are controlled for by including in the model the log-level of GDP per capita (Eurostat) at the beginning of each period (OECD, 2009a): this is the term Y on the right-hand side of Equation 1. The 'control matrix' is included in all specifications of the model and contains a set of additional control variables. The national annual growth rate accounts for the link between the national economic context and regional economic performance (Monastiriotis, 2014) while minimizing the effect of spatial autocorrelation by accounting for some of the common trends that characterize groups of territorial units; the Krugman index of specialization controls for the specialisation in local employment (Midelfart-Knarvik and Overman, 2002) by giving territorial unit *i* a zero rating if it has an industrial structure identical to other units, and by attributing a maximum value of 2 if it has no industries in common with other territorial units,[5] and finally population density controls for the local economy's degree of agglomeration.

B.2 - *Computation of the Social Filter Index*

Table B.1 Principal component Analysis. Eigen analysis of the Correlation Matrix.

Component	Eigenvalue	Difference	Proportion	Cumulative
Component 1	2.35352	1.37195	0.5884	0.5884
Component 2	0.981569	0.319494	0.2454	0.8338
Component 3	0. 662075	0.659236	0.1655	0.9993
Component 4	0. 002839	-	0.0007	1.0000

[5] The Krugman Index is computed – as customary in the existing literature - following Midelfart-Knarvik and Overman (2002). We compute: a) for each region, the share of industry k in that region's total employment: $v_i^k(t)$; b) the share of the same industry in the employment of all other regions: $\bar{v}_i^k(t)$; and c) the absolute values of the difference between these shares, added over all industries: $K_i(t) = \sum_k abs(v_i^k(t) - \bar{v}_i^k(t))$ with $v_i^k(t) = \sum_{j \neq i} x_i^k(t) / \sum_k \sum_{j \neq i} x_i^k(t)$. The index takes the value zero if region *i* has an industrial structure identical to the rest of the EU regions, and takes the maximum value of two if it has no industries in common with the rest of the EU

Table B.2 Principal component Analysis. Principal Components' Coefficients.

Variable	Comp 1	Comp 2	Comp 3	Comp 4
Agricultural share of employment	-0.3963	0.4757	0.7852	-0.0094
Long term unemployment	-0.3132	0.7339	-0.6026	0.0105
Human Capital	0.6103	0.3407	0.1101	0.7066
Skilled labour force	0.6102	0.3449	0.0905	-0.7074