Regional productivity growth in Europe: a Schumpeterian perspective

Roberto Basile (Second University of Naples)

September 19, 2014

Outline

Background

- ***Ertur and Koch (2011, EK11)*** propose an extension of the multi-country Schumpeterian growth model (Aghin and Howitt, 1998; Howitt, 2000) which incorporates the force of ***technology transfer***:
    - the further away an economy is from the technological frontier, the higher its productivity in the research sector (***advantage of backwardness*** conferred to technological laggards)
- While in Howitt all economies share the same global technological frontier, in EK11 the technological frontier becomes *local* (i.e. it is specific to each economy), being defined as a geometric weighted average of knowledge levels in all economies, with weights given by some measure of bilateral distance between the countries (***spatial technological interdependence***)

Schumpeterian growth equation: linear SDM

$$\begin{aligned}\gamma_i &= \beta_0 + \beta_1 \ln y_{0,i} + \beta_2 \ln \frac{s_{k,i}}{n_i + 0.05} + \beta_3 \ln s_{A,i} + \beta_4 \ln n_i \\ &\quad + \theta_1 \sum_{j \neq i}^{N} w_{ij} \ln y_{0,j} + \theta_2 \sum_{j \neq i}^{N} w_{ij} \ln \frac{s_{k,j}}{n_j + 0.05} + \rho \sum_{j \neq i}^{N} w_{ij} \gamma_j + \varepsilon_i \\ &\quad \varepsilon_i \sim iid\mathcal{N}\left(0, \sigma_\varepsilon^2\right) \qquad i = 1, ..., n\end{aligned}$$

- In matrix form, we have

$$\mathbf{y} = \mathbf{X}\beta + \theta \mathbf{W}_N \mathbf{Z} + \rho \mathbf{W}_N \mathbf{y} + \varepsilon$$

- Reduced form:

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W}_N)^{-1} (\mathbf{X}\beta + \theta \mathbf{W}_N \mathbf{Z} + \varepsilon)$$

Direct and indirect effects in SDM

- *Direct effect* of a change in X_{ik}on region i:

$$\partial E\left[y_i\right]/\partial X_{ik} = \left(\mathbf{I}_n - \widehat{\rho}\mathbf{W}_n\right)^{-1}_{ii}\left(\mathbf{I}_n\widehat{\beta}_k + \mathbf{W}_n\widehat{\theta}_k\right) \neq \widehat{\beta}_k$$

  - It includes the effect of feedback loops where observation i affects observation j and observation j also affects i. Its magnitude depends upon: 1) the position of the economies in space, 2) the degree of connectivity among economies which is governed by W, 3) the parameters β_k, θ_k, ρ
- *Indirect effect* of a change in X_{jk} on economy *i:*

$$\partial E\left[y_i\right]/\partial X_{jk} = \left(\mathbf{I}_n - \widehat{\rho}\mathbf{W}_n\right)^{-1}_{ij}\left(\mathbf{I}_n\widehat{\beta}_k + \mathbf{W}_n\widehat{\theta}_r\right) \neq 0$$

where $\left(\mathbf{I}_n - \widehat{\rho}\mathbf{W}_n\right)^{-1}_{ij}$ represents the ij th element of the matrix $\left(\mathbf{I}_n - \widehat{\rho}\mathbf{W}_n\right)^{-1}$

Motivation

- Other potential sources of misspecification bias in regional growth analysis:
  - **Functional form mis-specification**:
    - *Threshold effects* and *club convergence* (multiple regimes) in regional growth (Ertur and Gallo, 2009):
    - A *semiparametric framework* which recognizes the uncertainty in the functional form is often recommended (e.g. Liu and Stengos, 1999; Durlauf, Kourtellos, and Minkin, 2001; Kalaitzidakis, Mamuneas, Savvides, and Stengos, 2001)
  - **Unobserved spatial heterogeneity**:
    - Failing to control for it can introduce *omitted-variable biases* and preclude causal inference
    - Spatial interdependence may simply be the consequence of *(spatially correlated) omitted variables* rather than being the result of spillovers
    - If this is the case, there are no compelling reasons for using traditional parametric models (SDM, SAR or SEM). A simple semiparametric model, with a smooth interaction between latitude and longitude (*Geoadditive Model*), can remove unobserved heterogeneity (McMillen, 2012)
    - However, in economic growth analysis it is relevant to assess the impact of spillover effects (for example the global effect of a localized shock in R&D investment) rather than simply compensate for unobserved heterogeneity

Aim of the paper

- In this paper, we show that spatial interdependence effects (spatial spillovers) can only be correctly identified by using a flexible semiparametric approach which allows us to control for nonlinearities and spatial heterogeneity
- In particular, we show that ***Penalized-Spline Spatial Autoregressive and Spatial Durbin models*** (PS-SAR and PS-SDM) (Basile et al., 2014) prove to be very powerful tools to solve the identification problem
- Using data for the European regions at NUTS-2 level, we apply PS-SDM and PS-SAR models to empirically test the predictions of multi-country endogenous growth models with technological interdependence without imposing a functional form and controlling for the effect of unobserved spatial heterogeneity

A Penalized-spline Spatial Durbin Schumpeterian Model (PS-SDM-Schumpeterian)

$$\begin{aligned}\gamma_i &= \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + f_3(\ln s_{A,i}) + f_4(\ln n_i) \\ &\quad + m_1\left(\sum_{j \neq i}^{N} w_{ij} \ln y_{0,j}\right) + m_2\left(\sum_{j \neq i}^{N} w_{ij} \ln \frac{s_{k,j}}{n_j + 0.05}\right) + h(no_i, e_i) \\ &\quad + \rho \sum_{j \neq i}^{N} w_{ij}\gamma_j + \varepsilon_i\end{aligned}$$

$$\begin{aligned}\mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \Sigma_{q_1}\beta_{1q_1}b_{1q_1}(x_1) + \Sigma_{q_2}\beta_{2q_2}b_{2q_2}(x_2) + \ldots + \boldsymbol{\varepsilon} \\ &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\delta + \boldsymbol{\varepsilon}\end{aligned}$$

A Penalized-spline Spatial Durbin Schumpeterian Model (PS-SDM-Schumpeterian)

- This model reflects the notion of spatial dependence made of two parts:
    - a ***spatial trend*** due to unobserved regional characteristics, which is modeled by the smooth function of the coordinates
    - global ***spatial spillover effects***, which are modeled by including the spatial lag terms

Direct, indirect and total smooth effects

- The direct smooth effect can be computed as:

$$\widehat{f}_k^D(x_k) = [\mathbf{I}_n - \widehat{\varrho}\mathbf{W}_n]_{ii}^{-1} x_k \widehat{\beta}_k$$

- The indirect (spillover) smooth effect of x_k ca be written as

$$\widehat{f}_k^I(x_k) = [\mathbf{I}_n - \widehat{\varrho}\mathbf{W}_n]_{ij}^{-1} \left[x_k \widehat{\beta} + W x_k \widehat{\delta}_k\right]$$

- Finally, the total smooth effect is:

$$\widehat{f}_k^T(x_k) = \widehat{f}_k^D(x_k) + \widehat{f}_k^I(x_k)$$

Estimation

- Any semiparametric PS model can be expressed as a mixed model and its parameter can be estimated using **REML** (Ruppert, Wand, and Carroll, 2003; Wood, 2011)
- A complication with the PS-SAR and the PS-SDM is the presence of the endogenous spatial lag term $\mathbf{W}_n\mathbf{y}$
- Basile et al. (2014) show how the REML methodology can be extended to estimate the parameters of PS-SAR and PS-SDM either in a single-step or in a 2-step "**control function**" (CF) approach
- In the present study, we use the 2-step approach

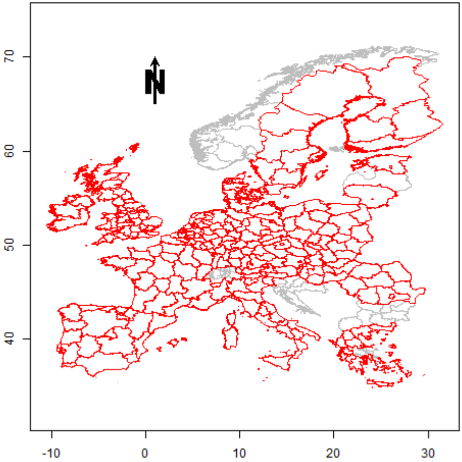
Instruments

- To identify the causal effect of $\mathbf{Wy}$, Kelejian and Prucha (1997) suggest to use $[\mathbf{X} \quad \mathbf{WX} \quad \dots \quad \mathbf{WX}^g]$ as instrumental variables
- Gibbons and Overman (2012) criticize this choice and point out the difficulty in justifying the exclusion restrictions on WX
- We use the spatial lag of the ***quality of regional governance*** developed by Charron et al. (2014) and the spatial lags of various measures of ***social capital***
- We argue that there are no direct impacts on the growth rate of a region from the policy intervention in its neighbors, but the policy does have effects via its influence on neighboring growth rates. In other words, the quality of the regional governance of its neighbors may affect the growth performance ($\mathbf{y}$) of region i only through $\mathbf{Wy}$
- Similar considerations can be made to justify the exclusion restrictions for the (spatial lags of) social capital measures

European regional data

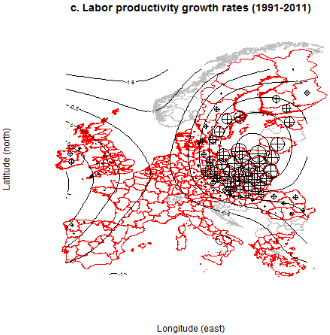
- Cambridge Econometrics data
- 248 NUTS2 regions for the 1991-2011 period
- Income per worker (y_{it}) of region i at time t: ratio between gross value added at constant prices 2000 and total employment
- Income levels are normalized with respect to the EU-27 average
- Average annual productivity growth rates:
  $$\gamma_y = \frac{\ln y_T - \ln y_0}{T}$$

Map of the European regions included in the sample



Spatial distribution of growth rates

- Eastern regions show the highest growth rates. Also Western (Ireland and UK) regions show higher growth rates
- If regional observable characteristics are unable to fully account for this spatial trend, unobserved spatial heterogeneity would contribute to generate spatial dependence in the residuals



The W matrix

$$w_{ij} = \begin{cases} e^{-d_{ij}} / \sum_{j \neq i} e^{-d_{ij}} & if \quad 100km < d_{ij} < 320km \\ 0 & if \quad i = j \quad or \quad if \quad d_{ij} < 100km \quad or \quad if \quad d_{ij} > 320km \end{cases}$$

d_{ij} is the great-circle distance between the centroids of regions i and j, expressed in kilometers (km)

The W matrix: next step

- Alternative interaction matrices:
  - $v_{ij} = H_i w_{ij}$, where H_i is the *human capital stock* of the receiving region i (EK11)
  - $v_{ij} = Q_i w_{ij}$, where Q_i is the *quality of regional governance* in the receiving region i (Coe et al., 2009)
  - Both H and Q are supposed to reflect the *capacity of absorption of new knowledge*
  - In order to capture West-East spillovers and international and intra-national spillovers, we (will) use block-triangular structures of the above mentioned interaction matrices

Linear models

- Linear a-spatial neoclassical (Solow) model:

$$\gamma_i = \beta_0 + \beta_1 \ln y_{0,i} + \beta_2 \ln \frac{s_{k,i}}{n_i + 0.05} + \varepsilon_{S,i}$$

- Linear a-spatial Schumpeterian (Howitt, 2000) model:

$$\gamma_i = \beta_0 + \beta_1 \ln y_{0,i} + \beta_2 \ln \frac{s_{k,i}}{n_i + 0.05} + \beta_3 \ln s_{A,i} + \beta_4 \ln n_i + \varepsilon_{H,i}$$

- Linear Spatial Durbin neoclassical (Erthur and Koch, 2007) model:

$$\begin{aligned}\gamma_i = {} & \beta_0 + \beta_1 \ln y_{0,i} + \beta_2 \ln \frac{s_{k,i}}{n_i + 0.05} \\ & + \theta_1 \sum_{j \neq i}^{N} w_{ij} \ln y_{0,j} + \theta_2 \sum_{j \neq i}^{N} w_{ij} \ln \frac{s_{k,j}}{n_j + 0.05} + \rho \sum_{j \neq i}^{N} w_{ij} \gamma_j + \varepsilon_{EK07,i}\end{aligned}$$

- Linear Spatial Durbin Schumpeterian (Erthur and Koch, 2011) model:

$$\begin{aligned}\gamma_i = {} & \beta_0 + \beta_1 \ln y_{0,i} + \beta_2 \ln \frac{s_{k,i}}{n_i + 0.05} + \beta_3 \ln s_{A,i} + \beta_4 \ln n_i \\ & + \theta_1 \sum_{j \neq i}^{N} w_{ij} \ln y_{0,j} + \theta_2 \sum_{j \neq i}^{N} w_{ij} \ln \frac{s_{k,j}}{n_j + 0.05} + \rho \sum_{j \neq i}^{N} w_{ij} \gamma_j + \varepsilon_{EK11,i}\end{aligned}$$

Nonlinear models

- Nonlinear a-spatial neoclassical model:

$$\gamma_i = \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + \varepsilon_{NLS,i}$$

- Nonlinear a-spatial Schumpeterian model:

$$\gamma_i = \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + f_3(\ln s_{A,i}) + f_4(\ln n_i) + \varepsilon_{NLH,i}$$

- Nonlinear SDM neoclassical model:

$$\begin{aligned}\gamma_i &= \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + \\ &+ m_1\left(\sum_{j\neq i}^{N} w_{ij} \ln y_{0,j}\right) + m_2\left(\sum_{j\neq i}^{N} w_{ij} \ln \frac{s_{k,j}}{n_j + 0.05}\right) + \rho \sum_{j\neq i}^{N} w_{ij}\gamma_j + \varepsilon_{NLEK07,i}\end{aligned}$$

- Nonlinear SDM Schumpeterian model:

$$\begin{aligned}\gamma_i &= \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + f_3(\ln s_{A,i}) + f_4(\ln n_i) \\ &+ m_1\left(\sum_{j\neq i}^{N} w_{ij} \ln y_{0,j}\right) + m_2\left(\sum_{j\neq i}^{N} w_{ij} \ln \frac{s_{k,j}}{n_j + 0.05}\right) + \rho \sum_{j\neq i}^{N} w_{ij}\gamma_j + \varepsilon_{NLEK11,i}\end{aligned}$$

Geoadditive models

- Geoadditive SDM neoclassical model:

$$\begin{aligned}\gamma_i &= \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + \\ &+ m_1\left(\sum_{j\neq i}^{N} w_{ij}\ln y_{0,j}\right) + m_2\left(\sum_{j\neq i}^{N} w_{ij}\ln \frac{s_{k,j}}{n_j+0.05}\right) + h(no_i, e_i) + \rho\sum_{j\neq i}^{N} w_{ij}\gamma_j + \varepsilon_{GeoNLEK07,i}\end{aligned}$$

- Geoadditive SDM Schumpeterian model:

$$\begin{aligned}\gamma_i &= \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + f_3(\ln s_{A,i}) + f_4(\ln n_i) \\ &+ m_1\left(\sum_{j\neq i}^{N} w_{ij}\ln y_{0,j}\right) + m_2\left(\sum_{j\neq i}^{N} w_{ij}\ln \frac{s_{k,j}}{n_j+0.05}\right) + h(no_i, e_i) + \rho\sum_{j\neq i}^{N} w_{ij}\gamma_j + \varepsilon_{GeoNLEK11,i}\end{aligned}$$

- Geoadditive SAR neoclassical model:

$$\gamma_i = \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + h(no_i, e_i) + \rho\sum_{j\neq i}^{N} w_{ij}\gamma_j + \varepsilon_{GeoNLEK07,i}$$

- Geoadditive SAR Schumpeterian model:

$$\begin{aligned}\gamma_i &= \beta_0 + f_1(\ln y_{0,i}) + f_2\left(\ln \frac{s_{k,i}}{n_i + 0.05}\right) + f_3(\ln s_{A,i}) + f_4(\ln n_i) \\ &+ h(no_i, e_i) + \rho\sum_{j\neq i}^{N} w_{ij}\gamma_j + \varepsilon_{GeoNLEK11,i}\end{aligned}$$

Model comparison

Model	BIC	Spatial trend	Spatial dependence	Rho
Neoclassical a-spatial linear	5.230	Trend	Yes	
Schumpeterian a-spatial linear	5.065	Trend	Yes	
Neoclassical a-spatial nonlinear	5.210	Trend	Yes	
Schumpeterian a-spatial nonlinear	5.061	Trend	Yes	
Neoclassical SDM linear	5.089	Trend	No	0.882
Schumpeterian SDM linear	4.971	Trend	No	0.736
Neoclassical SDM nonlinear	4.925	Trend	No	0.809
Schumpeterian SDM nonlinear	4.841	Trend	No	0.707
Neoclassical SDM Geoadditive	4.942	No trend	No	0.463
Schumpeterian SDM Geoadditive	4.817	No trend	No	0.324
Neoclassical SAR Geoadditive	4.867	No trend	No	0.450
Schumpeterian SAR Geoadditive	4.767	No trend	No	0.303

Results linear models

Variable	Solow (1956)	Howitt (2000)	Ertur-Koch (2007)	Ertur-Koch (2011)
Intercept	-0.099*	-2.362***	-0.059	-1.236***
$\ln y_0$	-0.960***	-1.310***	-1.015***	-1.130***
$\ln \frac{s_k}{n+0.05}$	0.141*	0.660***	0.166**	0.655***
$\ln s_A$		0.510***		0.261***
$\ln n$		0.538**		0.488**
$W \ln y_0$			0.936***	0.706***
$W \ln \frac{s_k}{n+0.05}$			-0.167**	-0.187***
$W\gamma$			0.882***	0.736***
Diagnostic tests				
Weak instruments			11.189***	8.378***
Wu-Hausman			13.786***	7.555***
Sargan			12.848	9.621

Results linear models

- A 1% increase in R&D investments (s_A) would result in a 1% increase in growth. Around 4/10 of this impact comes from the direct effect, and 6/10 from the indirect or spatial spillover impact

Variable	Solow (1956)	Howitt (2000)	Ertur-Koch (2007)	Ertur-Koch (2011)
ADE of $\ln y_0$	-0.960	-1.310	-0.931	-1.236
AIE of $\ln y_0$			0.263	-0.369
ATE of $\ln y_0$			-0.668	-1.604
ADE of $\ln \frac{s_k}{n+0.05}$	0.141	0.660	0.124	0.904
AIE of $\ln \frac{s_k}{n+0.05}$			-0.132	0.868
ATE of $\ln \frac{s_k}{n+0.05}$			-0.008	1.772
ADE of $\ln s_A$		0.510		0.423
AIE of $\ln s_A$				0.565
ATE of $\ln s_A$				0.987
ADE of $\ln n$		0.538		0.792
AIE of $\ln n$				1.057
ATE of $\ln n$				1.849

CF estimates of the Schumpeterian PS-SAR Model

First stage	Second stage	
Parametric terms	*Estimate (Bootstrap p-value)*	
(Intercept)	0.0686	-1.212
	(0.052)	(0.000)
$W\gamma$		0.298
		(0.000)
$\ln y_0$		-0.941
		(0.000)
$\ln s_A$		0.253
		(0.000)
Smooth terms	*edf*	*edf*
$f_1(\ln y_0)$	1.000	
$f_2\left(\ln \frac{s_k}{n+0.05}\right)$	1.455	3.020
$f_3(\ln s_A)$	1.000	
$f_2(\ln n)$	1.000	2.542
$h(no, e)$	14.674	13.720
$l(res(first.step))$		1.000
$g(W(gov))$	2.505	
$s_1(W(SC_1))$	2.500	
$s_2(W(SC_2))$	2.965	
$s_3(W(SC_3))$	3.075	
$s_4(W(SC_4))$	1.814	
$s_5(W(SC_5))$	1.854	
$s_6(W(SC_6))$	2.968	
$s_7(W(SC_7))$	3.629	
$s_8(W(SC_8))$	2.222	

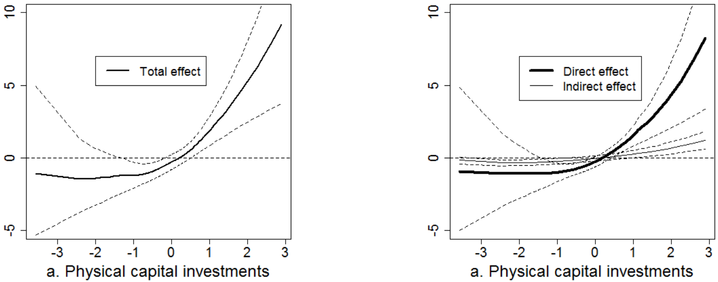
Control function estimates of the Schumpeterian Geoadditive SAR Model

- A 1% increase in R&D investments (s_A) would result in a 0.36% increase in growth. Around 7/10 of this impact comes from the direct effect, and only 3/10 from the indirect or spatial spillover impact

	First stage	Second stage
Parametric terms	*Estimate (Bootstrap p-value)*	
Average direct effect of $\ln y_0$		-0.991
		(0.000)
Average indirect effect of $\ln y_0$		-0.351
		(0.000)
Average total effect of $\ln y_0$		-1.342
		(0.000)
Average direct effect of $\ln s_A$		0.266
		(0.000)
Average indirect effect of $\ln s_A$		0.094
		(0.000)
Average total effect of $\ln s_A$		0.361
		(0.000)

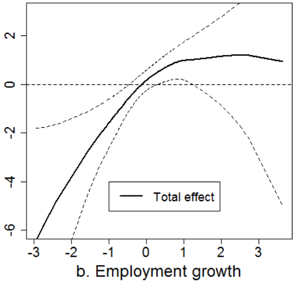
Direct, indirect and total smooth effects from nonlinear SAR EK11 model

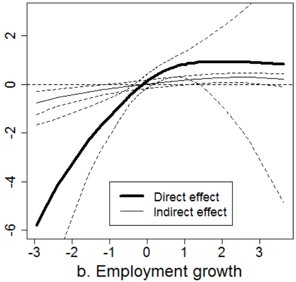
- An increase in $\ln \frac{s_k}{n+0.05}$ is associated with an increase in growth rates only when $\ln \frac{s_k}{n+0.05}$ is above the EU average



Direct, indirect and total smooth effects from nonlinear SAR EK11 model

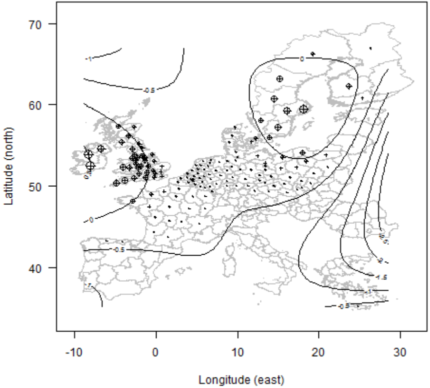
- The influence of the employment growth rate is positive, albeit not homogeneous across the sample





Spatial trend surface

- Ceteris paribus, the growth rates are significantly higher in North-Eastern and North-Western regions (UK and Ireland)



Conclusions

- We use European regional data to compare several competing linear and nonlinear (both neoclassical and Schumpeterian) growth equations
- An accurate model selection strategy leads us to select the nonlinear geoadditive Schumpeterian SAR model as the best specification
- **Our results corroborate the theoretical prediction of EK11 according to which R&D investment and international and interregional R&D spillovers are important drivers of regional growth**
- **However, spillover effects are much lower after controlling for spatial unobserved heterogeneity**
- Moreover, important nonlinearities in the effect of physical capital investments emerge, putting into question the strong homogeneity assumption and suggesting a threshold effect in growth behavior