Measuring Efficiency in a Spatial Context Through Quantile Regression

Benedetti R.[1], Billé AG.[2] , Piersimoni F.[3] , and Salvioni C.[1]

[1] DEC, University of Chieti-Pescara, Pescara, Italy
[2] DEF, Tor Vergata University, Rome, Italy
[3] Istat, Istituto Nazionale di Statistica, Rome, Italy

July 2016

Purpose of the Study

Primary Purpose: To identify the technical efficiency of olive-growing farms observed in the Italian Farm Accountancy Data Network (FADN), yearly survey carried out by the Member States of the European Union in 2012, by using Quantile Regression

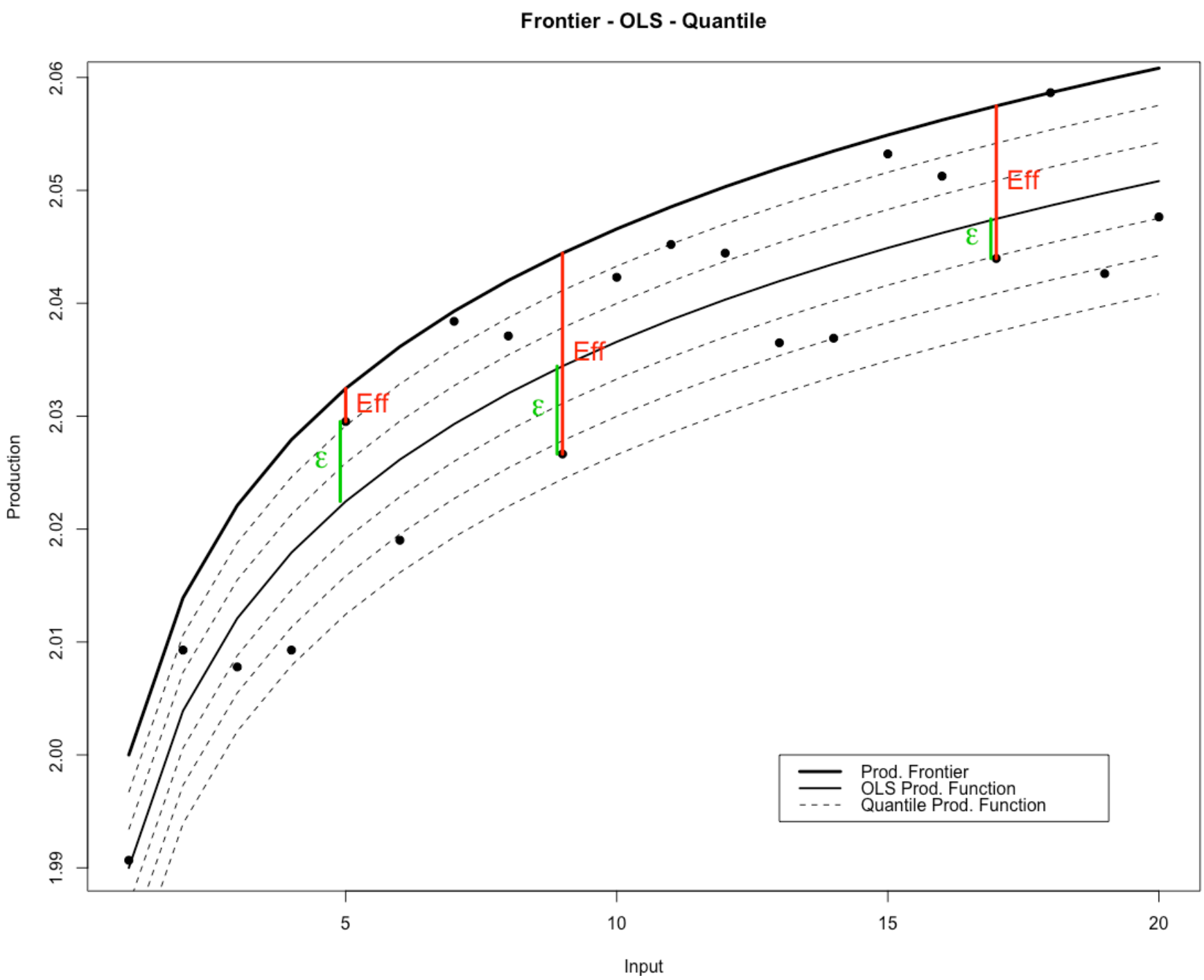
Secondary purposes:

- Explicit introduction of spatial autocorrelation. We consider spatial spillover effects through the use of spatial autocorrelation parameters.
- To account for inherent spatial heterogeneity in production functions in terms of spatially varying relationships through the use of Spatially weighted regression models.
- To identify spatial regimes in production functions at the farm level. We try to use farms' geographical coordinates to proxy for the effects of the interplay among a variety of latent unobserved factors, which gives rise to structural differences across space.

Efficiency Analysis

- Efficiency analyses focus on the efficiency of some production process in transforming inputs into outputs.
  - DEA (Data Envelopment Analysis) is a *non-parametric approach* that uses mathematical programming to identify the efficient frontier.
  - Frontier methods use a frontier to identify the efficiency of individual organizations relative to a reference set of organizations SFA (Stochastic Frontier Analysis) is a parametric *approach* that hypothesizes a functional form and use the data to econometrically estimate the parameters of that function
  - Quantile regression (M-Quantile) : either DEA and SFA do not use the distribution of the data to define the efficiency measure.

  P Kokic, R Chambers, J Breckling, S Beare, (1997) A measure of production performance, Journal of Business and Economic Statistics 15, 445-451
- The measure of efficiency is normally one of either:
  - The distance between observed and maximum possible output for given inputs (output efficiency)
  - The distance between observed and minimum possible input for given outputs (input efficiency)
  - τ_q : the closest quantile regression to each unit *I* among the Q *a priori* defined


Frontier - OLS - Quantile
Production
Input
1.99
2.00
2.01
2.02
2.03
2.04
2.05
2.06
5
10
15
20
Eff
ε
Eff
ε
Eff
ε
Prod. Frontier
OLS Prod. Function
Quantile Prod. Function

$$\log(y_i) = \log(X_i)\beta + \varepsilon_i \quad \Big| \quad \min_{\beta} \left[\sum_i (y_i - X_i\beta)^2 \right]$$

OLS ε can not be considered an efficiency measure

SFA, Stochastic Frontier Analysis

log (y) = βlog (X) + v – u

where:

v ~ iid $N(0,\sigma^2_v I)$;

u ~ $N^+(0 , \sigma^2_u I)$ this is an inefficiency measure;

v and u are independent of each other and of the regressors;

Quantile Regression (τ not necessarily =0.5, usually for several τs)

$$y_i = X_i\beta_\tau + \varepsilon_i$$

$$\min_{\beta_\tau}\left[\sum_{\{i|y_i \geq X_i\beta_\tau\}} \tau\left|y_i - X_i\beta_\tau\right| + \sum_{\{i|y_i < X_i\beta_\tau\}} (1-\tau)\left|y_i - X_i\beta_\tau\right|\right]$$

Negative residuals

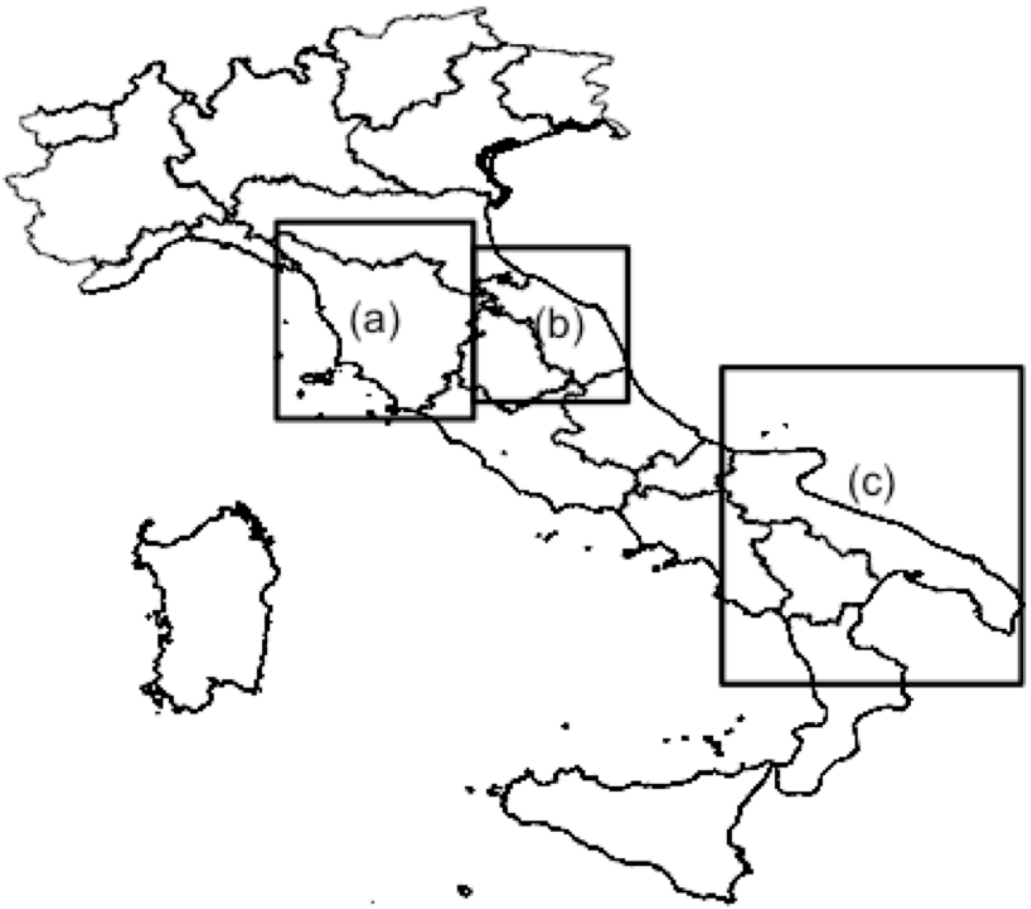
Positive residuals

For each unit *i* an efficiency measure can be represented by the τ corresponding to the regression line closest to *i*

AN APPLICATION TO OLIVE FARMS IN ITALY

Italian FADN: 3 Regional samples of olive-growing farms:

Tuscany (a - 317) – Marche (b - 268) - Apulia (c - 270).



AN APPLICATION TO OLIVE FARMS IN ITALY

Dependent variable: olive production (kg)

Explanatory variables:

- **Land** grown to olive-tree cultivation (ha);
- **Labor**: hired and family labor (hours);
- **Capital**: proxied by mechanical work (hours);
- **Other inputs:** water, fertilizers, pesticides, fuel and electric power and other miscellaneous expenses, augmented with expenses for contract work (euros);

We use the **Cobb–Douglas** (CD) functional form:

- Coefficients are easy to interpret;
- CD avoids the multicollinearity problem that arises with more flexible functional forms;
- flexibility is not an issue in our case given our coefficients are local specific .

Potential endogeneity of labor, capital and intermediate inputs : Instrumental Variables

The set of variables used as instruments contains

- lagged values (Pindyck and Rotenberg, 1983);
- prices of nutrients (PF) and pesticides (PP);
- opportunity cost of labor (OCL) and capital (OCK) provided in the Italian FADN;
- a proxy of the investments in machinery (I) obtained as the variation in horsepower at the farm level observed between 2013 and 2012.

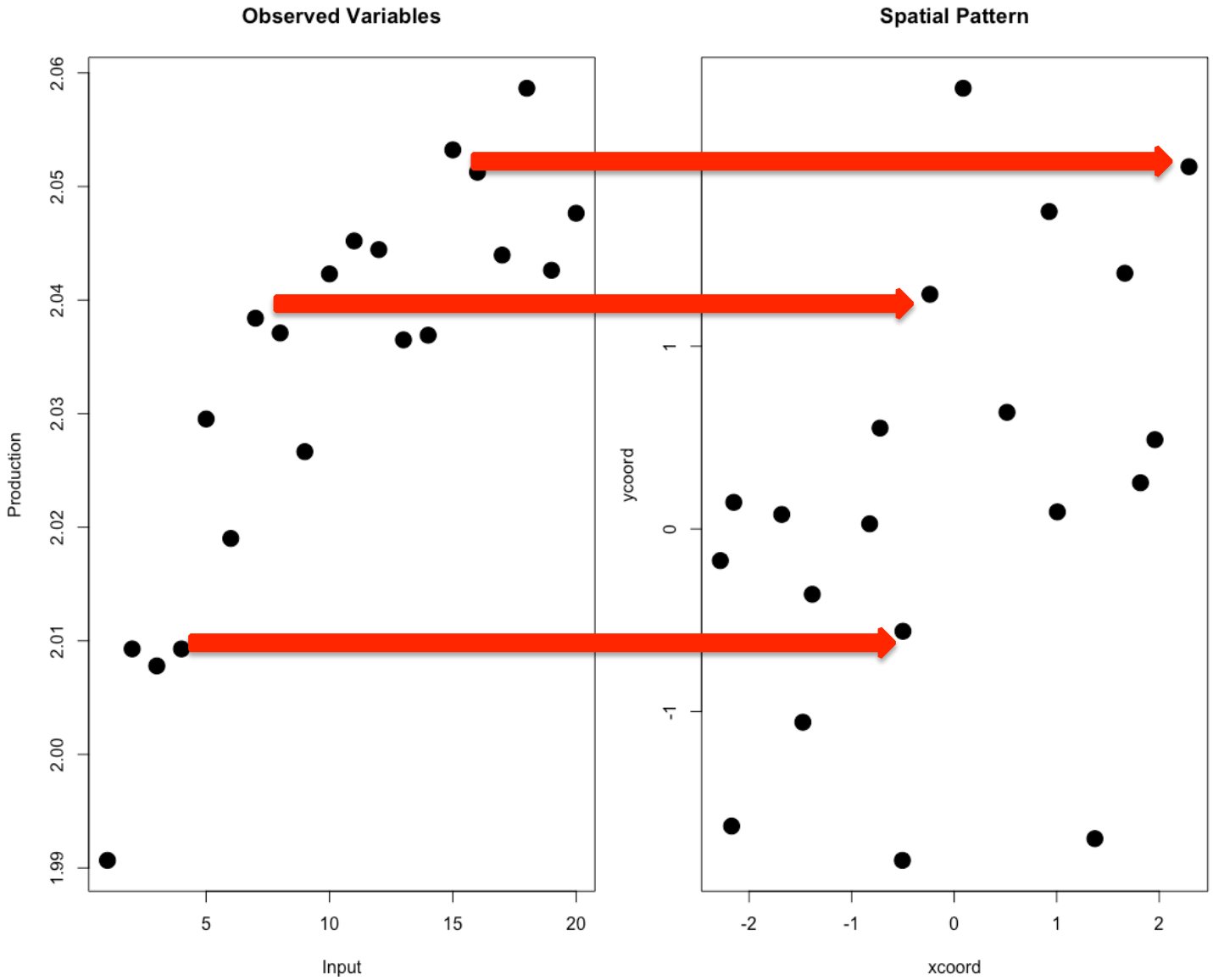
After testing for the correlation among the endogenous variables and IV by using simple linear regression models, we find that

- In Tuscany and Marche, the lagged values of inputs are valid instruments of the three variables suspected of being endogenous,
- in Apulia, the valid instruments are the lagged value and opportunity cost in the case of labor (OCL); lagged value, opportunity cost and investment in machinery for capital (OCK, I); and finally, lagged value and price of fertilizers (PF) in the case of intermediate inputs.

Quantile Regression FADN data 2012 (τ= 0.10,0.25,0.50,0.75,0.90)

		Intercept		Land (UAA)		Labor		Capital		Other inputs	
Region	Quantile	beta	se	beta	se	beta	se	beta	se	beta	se
TOS	q10	-0,025	0,416	0,333	0,361	0,076	0,161	0,119	0,289	0,338	0,373
TOS	q25	1,069	0,101	0,538	0,057	0,080	0,035	0,154	0,048	0,197	0,066
TOS	q50	1,505	0,097	0,566	0,071	0,139	0,019	0,125	0,058	0,134	0,077
TOS	q75	1,836	0,084	0,527	0,054	0,221	0,018	0,103	0,038	0,066	0,082
TOS	q90	3,336	0,075	0,706	0,054	0,064	0,034	0,125	0,040	-0,026	0,049
MAR	q10	1,460	0,882	0,679	0,582	0,062	0,303	-0,340	0,366	0,465	0,447
MAR	q25	1,249	0,140	0,655	0,098	-0,026	0,046	-0,347	0,051	0,653	0,082
MAR	q50	1,529	0,208	0,599	0,164	-0,140	0,083	-0,086	0,092	0,551	0,142
MAR	q75	2,320	0,172	0,695	0,162	-0,123	0,084	0,010	0,095	0,370	0,151
MAR	q90	2,756	0,215	0,658	0,154	-0,084	0,051	0,084	0,060	0,232	0,119
PUG	q10	-0,717	0,742	0,543	0,799	0,448	0,602	0,085	0,478	0,075	0,370
PUG	q25	-0,702	0,146	0,625	0,103	0,430	0,079	0,026	0,058	0,175	0,050
PUG	q50	-0,073	0,243	0,590	0,225	0,286	0,170	0,041	0,148	0,271	0,085
PUG	q75	0,725	0,164	0,646	0,106	0,040	0,078	0,050	0,069	0,428	0,034
PUG	q90	1,969	0,173	0,723	0,146	-0,054	0,095	0,005	0,085	0,401	0,080

Geocoded FADN data



Spatial dependence…

the existence of a functional relationship between what happens at one point in space and what happens elsewhere

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\lambda}\mathbf{W}\mathbf{y} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \propto N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

Tobler's (1970) First Law of Geography, according to which
"everything is related to everything else, but near things are more related than distant things".
Goodchild (1992) also defines spatial dependence as
"the propensity for nearby locations to influence each other and to possess similar attributes".

SSFA, Spatial Stochastic Frontier Analysis

The classic SFA model is based on the hypothesis of independence among the productive units: therefore, this specification ignores the role of any spatial effects that may be present in the data.

$\log(y) = \beta \log(X) + v - (I - \rho W)^{-1} \mathbf{u}$

where:

$V \sim \text{iid } N(0, \sigma^2_v I)$;

$\mathbf{u} \sim N^+(0, [(I - \rho W)^{-1} (I - \rho W^T)^{-1}] \sigma^2_{\bar{u}} I)$;

v and u are independent of each other and of the regressors;

Spatial Quantile Regression

(several τs)

$$y_i = X_i \beta_\tau + \rho_\tau \mathbf{W}\mathbf{y} + \varepsilon_i$$

Estimation method

Chernozhukov, Victor and Christian Hansen, "Instrumental Quantile Regression Inference for Struc- tural and Treatment Effect Models," *Journal of Econometrics* 132 (2006), 491-525.

Kim, Tae-Hwan and Christophe Muller, "Two-Stage Quantile Regression when the First Stage is Based on Quantile Regression, *Econometrics Journal* 7 (2004), 218-231.

The procedure has two stages. In the first stage, an instrumental variable is constructed for *WY* using the predicted values from a quantile regression of *WY* on a set of instruments, *Z*. The second stage is a quantile regression of *Y* on *X* and the predicted values of *WY*.

Spatial Quantile Regression FADN data 2012 (τ= 0.10,0.25,0.50,0.75,0.90)

		Intercept		Land (UAA)		Labor		Capital		Other inputs		ρ	
Region	Quantile	beta	se	beta	se	beta	se	beta	se	beta	se	beta	se
TOS	q10	-0,953	0,820	0,347	0,044	-0,005	0,050	0,122	0,041	0,284	0,023	**0,481**	**0,227**
TOS	q25	0,451	0,341	0,540	0,021	0,074	0,011	0,148	0,013	0,161	0,016	**0,246**	**0,098**
TOS	q50	0,825	0,249	0,567	0,008	0,095	0,009	0,124	0,009	0,148	0,008	**0,236**	**0,070**
TOS	q75	1,326	0,258	0,561	0,012	0,190	0,009	0,095	0,011	0,057	0,011	**0,212**	**0,073**
TOS	q90	2,666	0,274	0,691	0,015	0,070	0,021	0,127	0,020	-0,038	0,007	**0,187**	**0,080**
MAR	q10	-0,064	0,759	0,736	0,026	0,130	0,037	-0,292	0,034	0,311	0,029	**0,684**	**0,269**
MAR	q25	-0,722	0,532	0,642	0,019	0,240	0,032	-0,254	0,034	0,262	0,041	**0,853**	**0,198**
MAR	q50	-0,359	0,591	0,699	0,020	0,066	0,028	-0,168	0,026	0,299	0,028	**0,941**	**0,217**
MAR	q75	1,025	0,611	0,710	0,017	-0,037	0,033	-0,116	0,032	0,360	0,032	**0,509**	**0,223**
MAR	q90	2,485	0,836	0,659	0,031	-0,106	0,048	0,118	0,046	0,203	0,032	**0,148**	**0,307**
PUG	q10	-5,476	1,458	0,608	0,039	0,205	0,084	0,174	0,036	0,241	0,048	**0,933**	**0,263**
PUG	q25	-2,748	0,903	0,575	0,022	0,544	0,043	0,003	0,015	0,105	0,033	**0,405**	**0,181**
PUG	q50	-2,126	0,768	0,672	0,022	-0,023	0,055	0,128	0,019	0,381	0,026	**0,552**	**0,143**
PUG	q75	-1,100	0,712	0,638	0,016	-0,155	0,036	0,102	0,018	0,533	0,024	**0,412**	**0,136**
PUG	q90	0,110	0,822	0,691	0,026	-0,270	0,052	0,165	0,034	0,443	0,028	**0,433**	**0,158**

Heterogeneity in technologies

in most empirical literature a **global** production function is proposed that **assumes production technology is invariant** across firms

...all individuals come from a population with a **single slope β**

Firms do not operate on a **common production function** (homogeneous technology), … values of β are different over firms.

Assuming a common production function encompassing every sample observation, i.e. failing to recognize the geographical variations in technology, leads to **biased estimates**.

Modeling observed heterogeneity in technologies

Classify sample observations into categories defined on the basis of *a priori* sample separation (e.g. regions) … **estimate a production function for each group** …

…. the use of single or even multiple characteristics to split a sample of observations can only be **incomplete proxy** for technologies, since differences in technologies can be the result of **both observed** and **unobserved factors**.

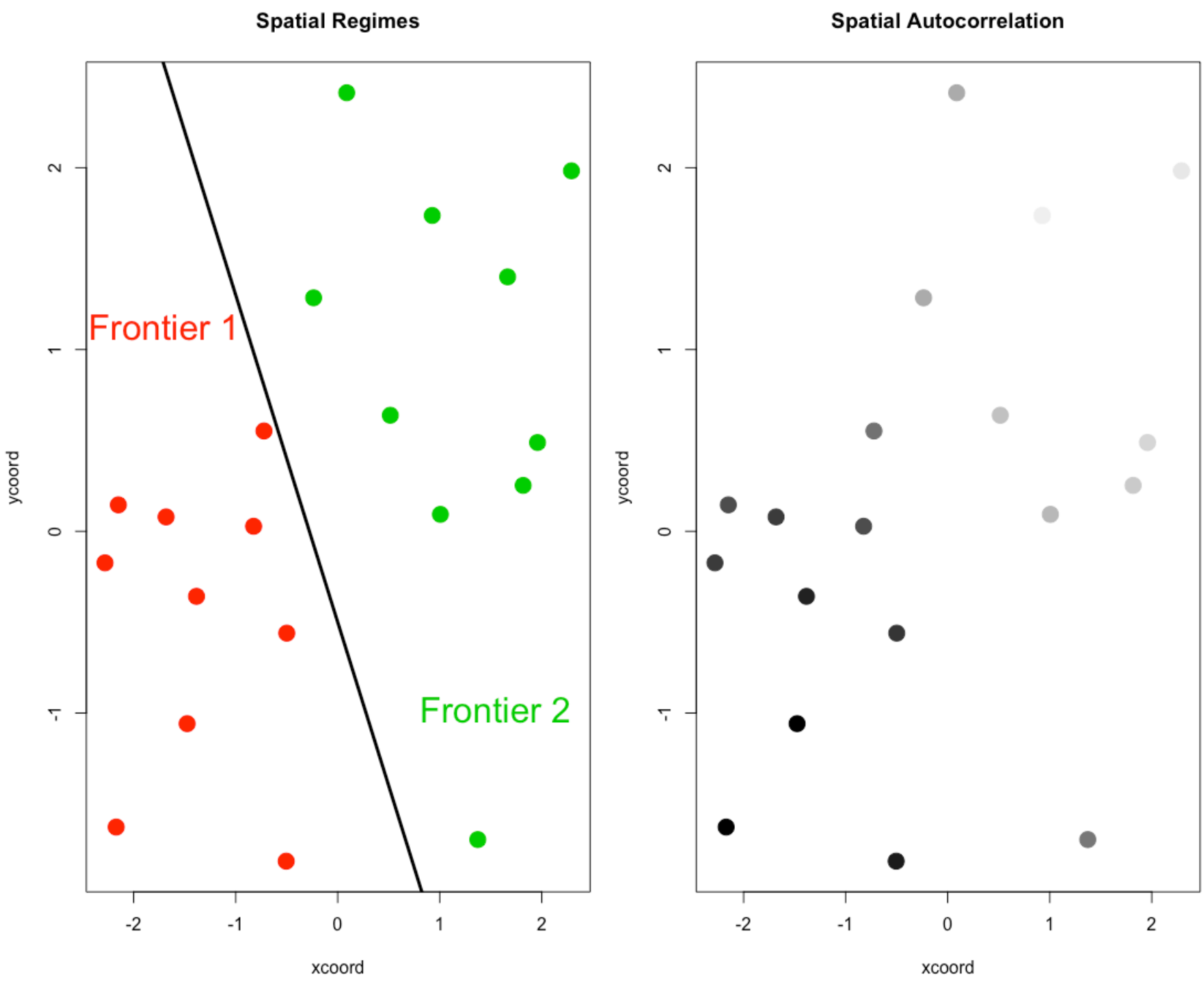
Spatial heterogeneity…

exists when the mean, variance, covariance structure “drifts” over a mapped process.
Typified by regional differentiation.
Reflects “spatial continuities” of physical-social-economic processes which, “taken together help bind space into recognizable structures” a “mosaic of homogeneous (or nearly homogeneous)” areas in which each is different from its neighbors (Haining 1990)
“The term spatial heterogeneity refers to variation in relationships over space.”

James P. LeSage
Spatial Econometrics

2 Approaches to deal with heterogeneity

- Geographically Weighted Regression
  (continuous spatial heterogeneity)
- Spatial Regimes
  (discrete spatial heterogeneity)


Spatial Regimes
Frontier 1
Frontier 2
ycoord
xcoord
-2
-1
0
1
2
Spatial Autocorrelation
ycoord
xcoord

Spatially Weighted Quantile Regression
(several τs)

$$y_i = X_i \beta_\tau + \varepsilon_i$$

$$\min_{\beta_\tau} \left[\sum_{\{i | y_i \geq X_i \beta_\tau\}} w_i \tau \left| y_i - X_i \beta_\tau \right| + \sum_{\{i | y_i < X_i \beta_\tau\}} w_i (1-\tau) \left| y_i - X_i \beta_\tau \right| \right]$$

Negative residuals

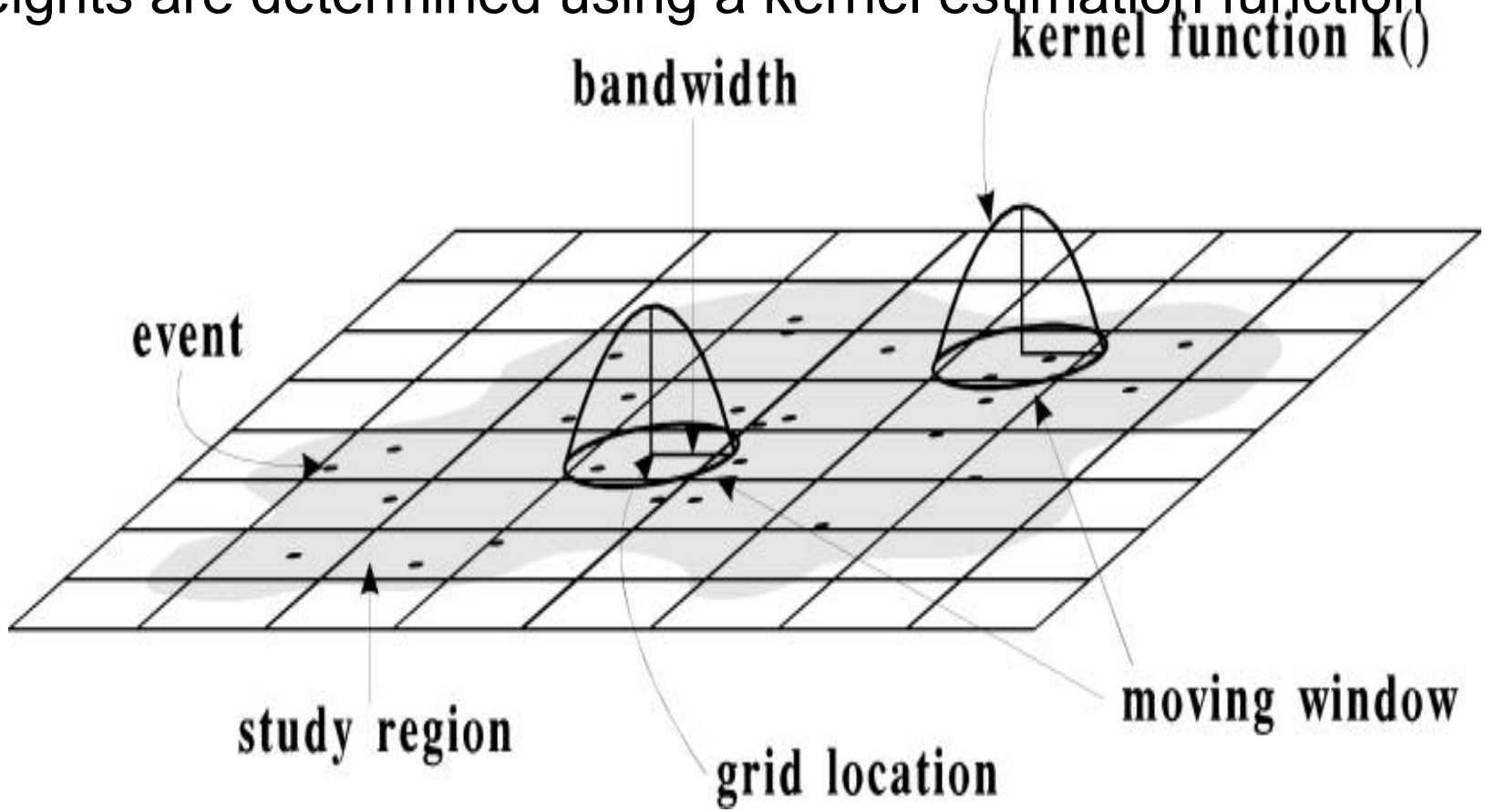
Positive residuals

Spatially Weighted Quantile Regression

- Productivity is non-stationary
- We have come to accept the reality that phenomena vary depending on where they are measured
  - Certainly for single variables
  - But multivariate relationships?
- Specifically, GWR is a tool for exploring and identifying variation in statistical *relationships* over space
- It's a way of exploring "spatial heterogeneity" ("spatial non-stationarity"); i.e., where the same stimulus provokes a different response in different parts of the study region

Spatially Weighted Quantile Regression

Weights are determined using a kernel estimation function



	Tuscany	Marche	Apulia
BW	67	126	111

Spatially Weighted Quantile Regression

Strengths of

- Potentially important tool when exploring spatial data. Nothing is the same everywhere. Helps you to understand spatial heterogeneity in your data.
- Provides better understanding global model. Serves as a device for possibly identifying specification errors in global model (e.g., important interaction effects). Thus, GWR and local analysis becomes a potential model-building procedure

Some Faults

- Regression undertaken at each “regression point” without much care regarding regression assumptions
- Data with spatially autocorrelated residuals fit with OLS rather than spatial regression model
- Results are very difficult to present in a table. It is usually better to show them on a map

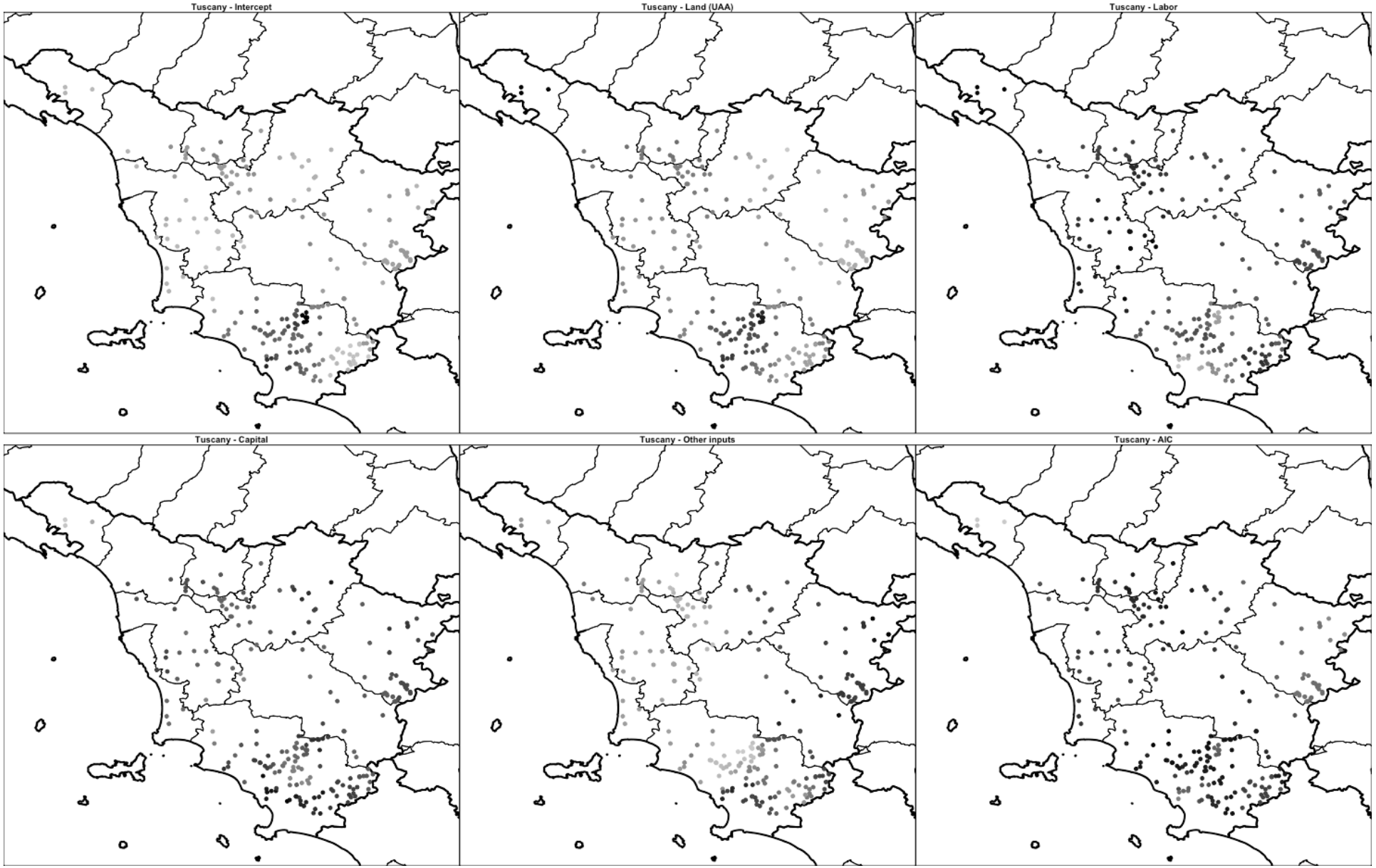
Spatially Weighted Quantile Regression

- Usual rule in regression: n observations, k parameters; $n >> k$
- GWR fits n x k parameters with only n observations
- Some unusual results arising from GWR are not yet fully understood. For example, it has been observed and commented upon that GWR can sometimes generate high (negative) correlations among estimated parameters

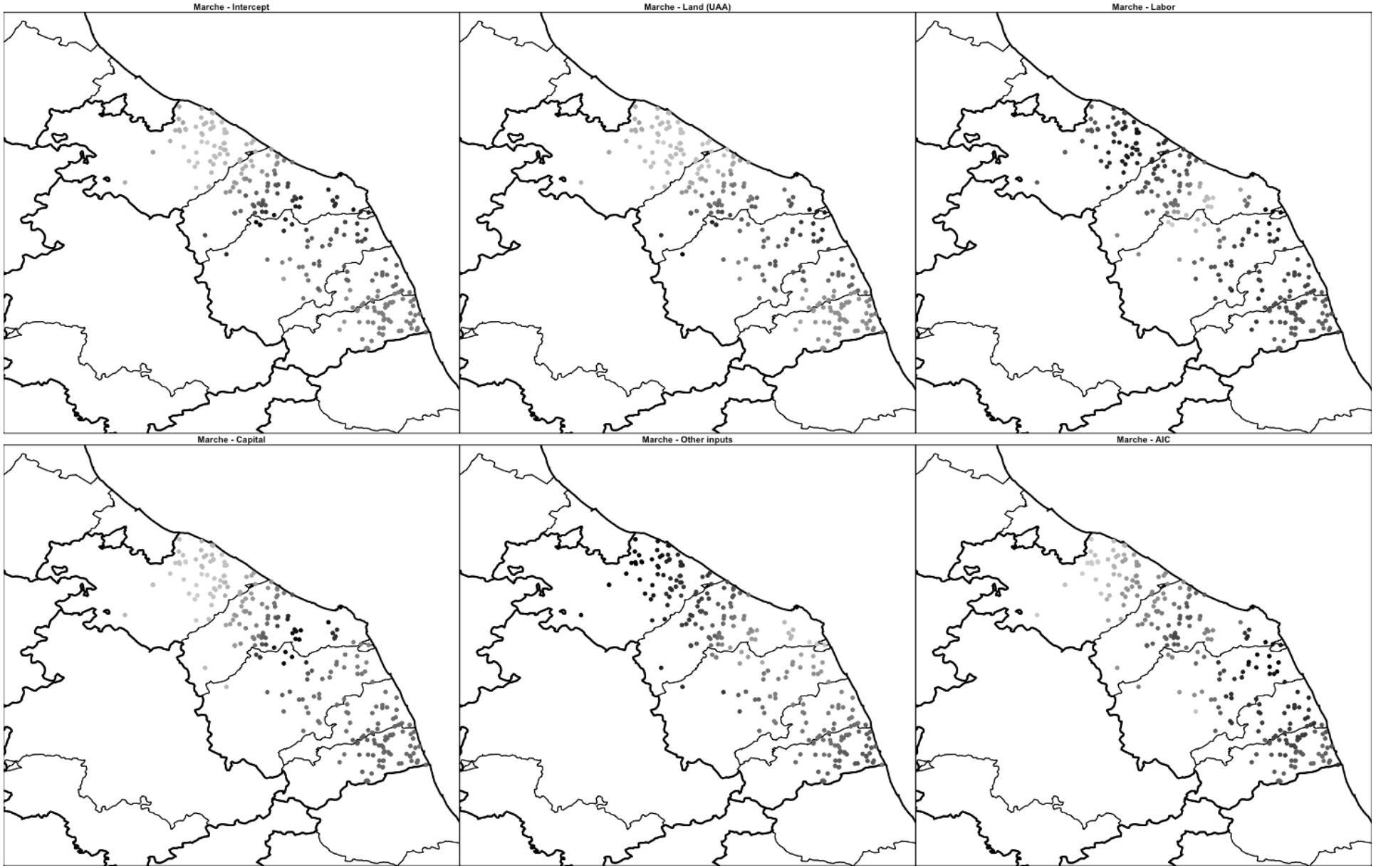
Spatial Weighted Quantile Regression FADN data 2012 (τ= 0.10,0.25,0.50,0.75,0.90)

		Intercept			Land (UAA)			Labor			Capital			Other inputs		
Reg	Q	Q1	Med	Q3	Q1	Med	Q3	Q1	Med	Q3	Q1	Med	Q3	Q1	Med	Q3
TOS	q10	0,061	0,868	2,883	0,274	0,593	0,897	-0,082	0,029	0,334	-0,152	0,047	0,100	0,009	0,200	0,298
TOS	q25	0,676	1,925	2,679	0,372	0,703	0,854	-0,048	0,054	0,257	-0,023	0,059	0,153	0,005	0,159	0,248
TOS	q50	1,171	2,314	2,918	0,474	0,668	0,828	-0,045	0,093	0,190	0,023	0,109	0,161	-0,014	0,096	0,194
TOS	q75	1,962	2,510	3,259	0,524	0,679	0,722	-0,048	0,124	0,211	0,083	0,119	0,142	0,011	0,051	0,104
TOS	q90	2,346	3,231	3,643	0,607	0,685	0,773	-0,050	0,088	0,215	-0,010	0,098	0,164	-0,033	0,015	0,078
MAR	q10	0,061	0,868	2,883	0,274	0,593	0,897	-0,082	0,029	0,334	-0,152	0,047	0,100	0,009	0,200	0,298
MAR	q25	0,676	1,925	2,679	0,372	0,703	0,854	-0,048	0,054	0,257	-0,023	0,059	0,153	0,005	0,159	0,248
MAR	q50	1,171	2,314	2,918	0,474	0,668	0,828	-0,045	0,093	0,190	0,023	0,109	0,161	-0,014	0,096	0,194
MAR	q75	1,962	2,510	3,259	0,524	0,679	0,722	-0,048	0,124	0,211	0,083	0,119	0,142	0,011	0,051	0,104
MAR	q90	2,346	3,231	3,643	0,607	0,685	0,773	-0,050	0,088	0,215	-0,010	0,098	0,164	-0,033	0,015	0,078
PUG	q10	0,061	0,868	2,883	0,274	0,593	0,897	-0,082	0,029	0,334	-0,152	0,047	0,100	0,009	0,200	0,298
PUG	q25	0,676	1,925	2,679	0,372	0,703	0,854	-0,048	0,054	0,257	-0,023	0,059	0,153	0,005	0,159	0,248
PUG	q50	1,171	2,314	2,918	0,474	0,668	0,828	-0,045	0,093	0,190	0,023	0,109	0,161	-0,014	0,096	0,194
PUG	q75	1,962	2,510	3,259	0,524	0,679	0,722	-0,048	0,124	0,211	0,083	0,119	0,142	0,011	0,051	0,104
PUG	q90	2,346	3,231	3,643	0,607	0,685	0,773	-0,050	0,088	0,215	-0,010	0,098	0,164	-0,033	0,015	0,078

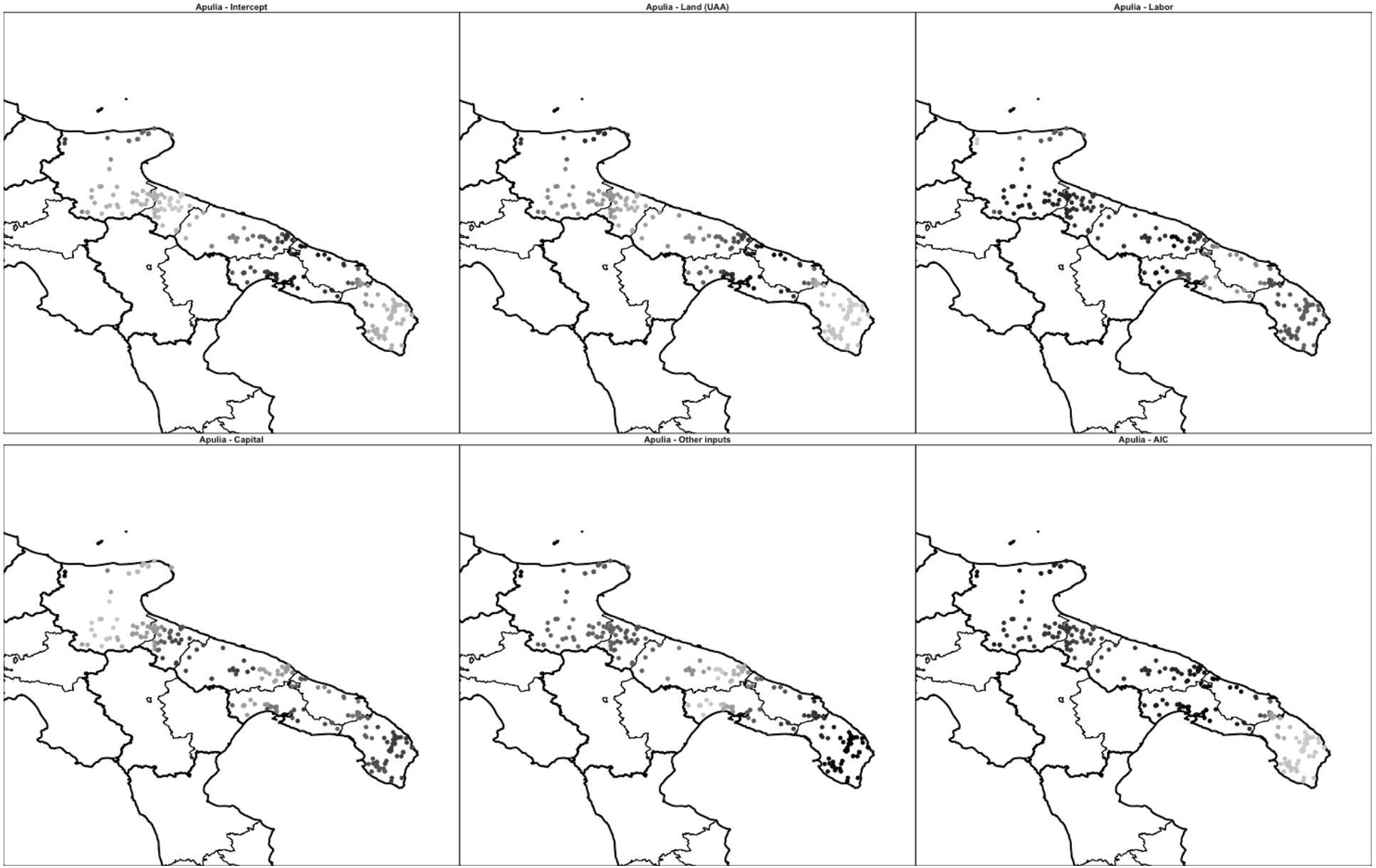
GWQR – Tuscany τ=0.5



GWQR – Marche τ=0.5



GWQR – Apulia τ=0.5



A brief note on M-Quantile

Spatial M-Quantile and Geographically Weighted M-Quantile (mainly for SAE):

1 Salvati, N.; Pratesi, M.; Tzavidis, N.; and Chambers, R., "Spatial M-quantile Models for Small Area Estimation", Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 15-08, 2008, 16p.

2 Molina I., Salvati N. & Pratesi M. (2009). Bootstrap for estimating the mean squared error of the Spatial Eblup. Computational Statistics, 24, 441-458.

3 Salvati, N.; Tzavidis, N.; Pratesi, M.; and Chambers, R., "Small Area Estimation via M-Quantile Geographically Weighted Regression", CCSR Working Paper 2007-09 & TEST, March 2012, Volume 21, Issue 1, pp 1-28

4 Chandra H., Salvati N. & Chambers R. (2015). A Spatially Nonstationary Fay-Herriot Model for Small Area Estimation. Journal Of Survey Statistics And Methodology, 3, 109-135.

Some Output Analysis

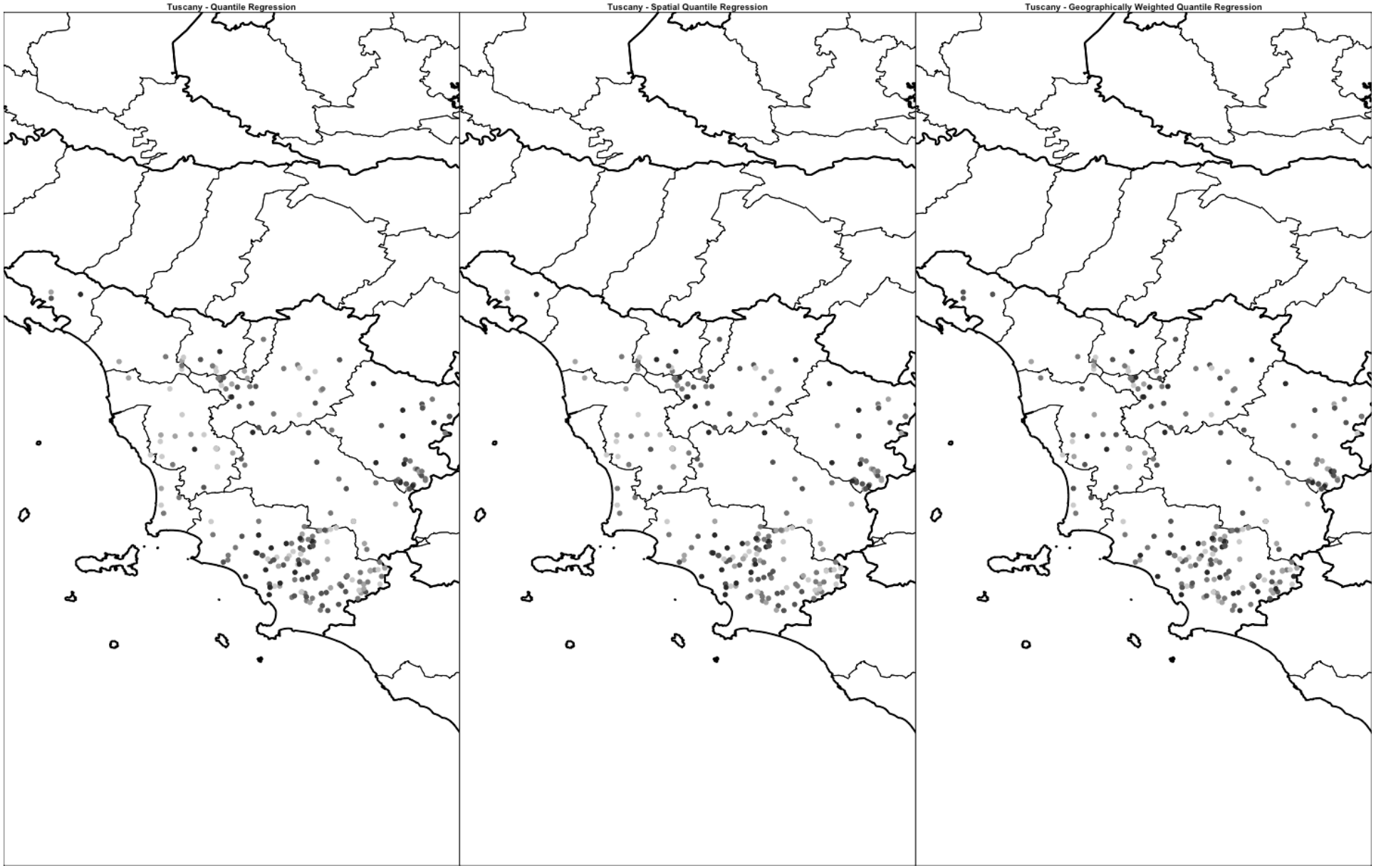
AIC for each model, region and quantile

	Tuscany			Marche			Apulia		
	QR	SQR	GWQR	QR	SQR	GWQR	QR	SQR	GWQR
q10	598,980	589,662	587,926	468,199	465,042	463,289	589,563	576,111	574,382
q25	468,621	464,420	462,684	350,109	343,776	342,024	494,399	482,227	480,498
q50	384,714	375,864	374,128	284,295	267,418	265,666	441,482	426,657	424,928
q75	388,650	377,928	376,192	300,409	297,302	295,550	431,798	422,045	420,316
q90	436,877	432,996	431,260	375,335	377,623	375,871	462,946	465,743	464,014

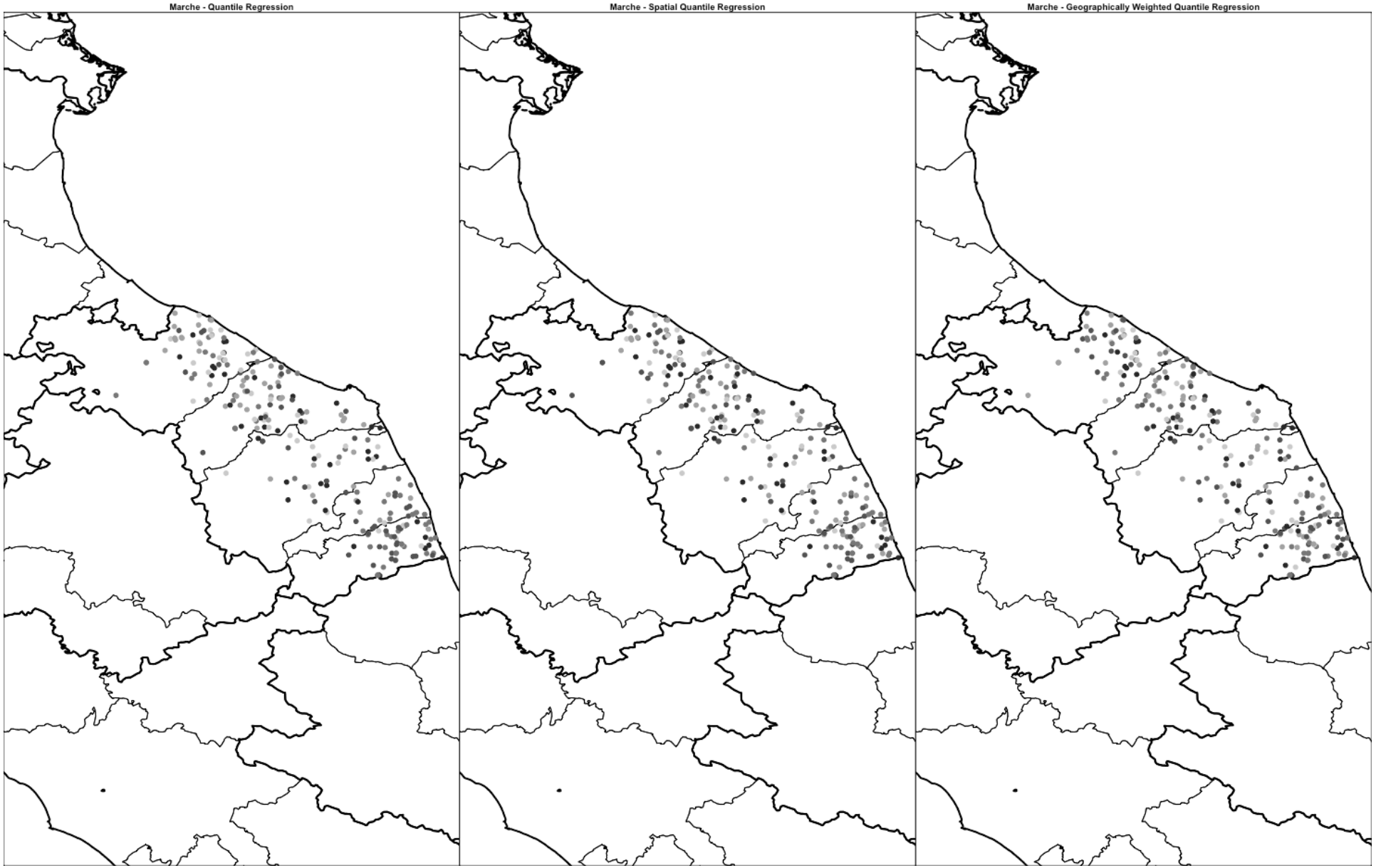
Comparison between the 3 models

	% Tace of Cont. Table			Corr. Coeff.			Moran Spat. Autoc. Coeff.		
	QR - SQR	QR - GWQR	SQR - GWQR	QR - SQR	QR - GWQR	SQR - GWQR	QR	SQR	GWQR
TOS	0,708	0,549	0,489	0,906	0,711	0,707	0,068	0,007	-0,006
MAR	0,766	0,581	0,601	0,914	0,819	0,818	0,031	-0,003	-0,005
PUG	0,678	0,493	0,467	0,878	0,763	0,783	0,203	0,110	-0,014

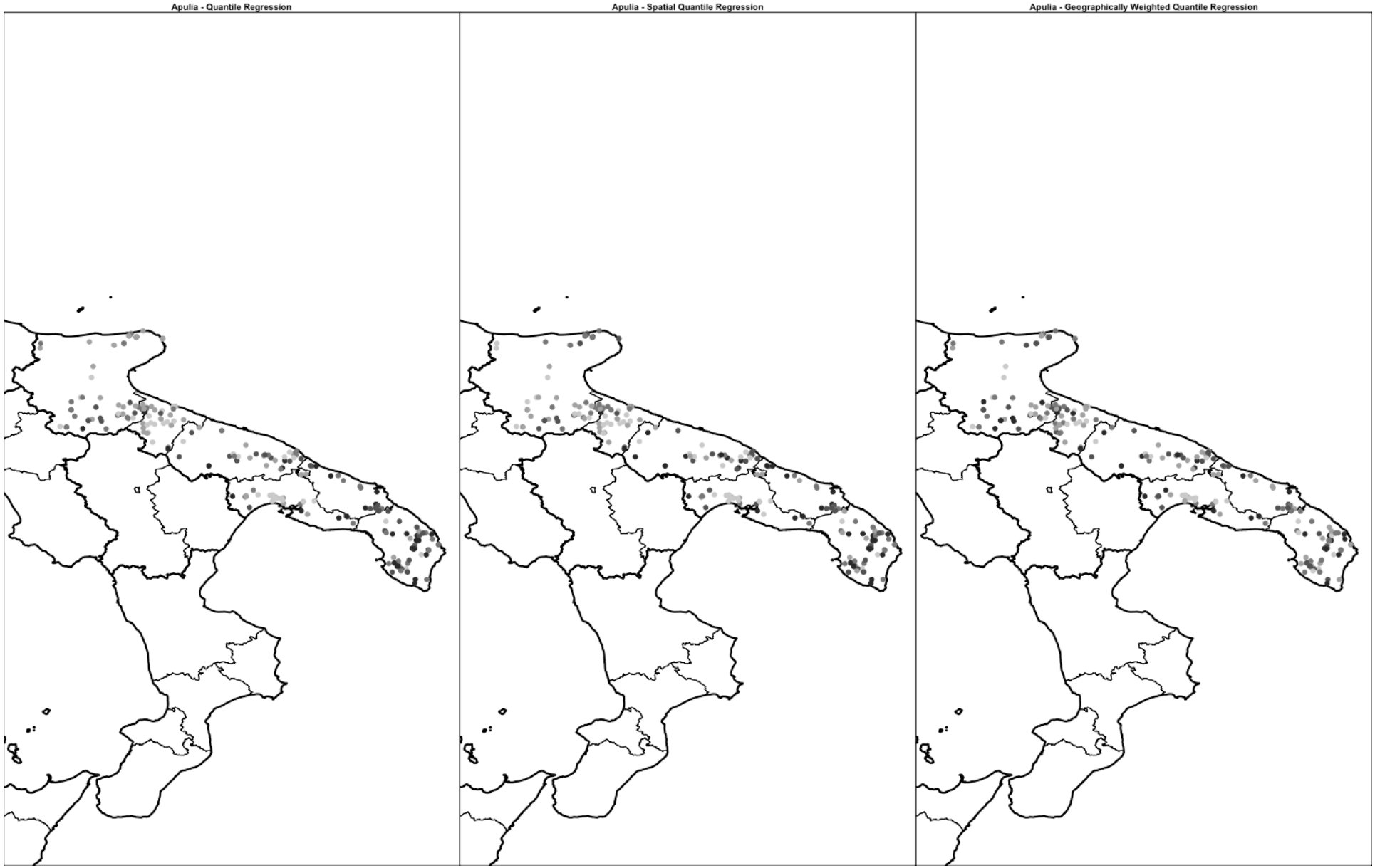
Spatial Pattern Eff. Measures Tuscany



Spatial Pattern Eff. Measures Marche



Spatial Pattern Eff. Measures Apulia



Spatial heterogeneity and spatial regimes in agriculture

The same stimulus has different response in different parts of the study region … values of β are different over space (spatially varying parameters) …. structural breaks in space (or ***spatial regimes***).

Why are we interested in **spatial regimes**?

Because in **agriculture** long **tradition** of studies aimed at identifying **homogeneous farming areas** rusulting from the long term, dynamic interactions among site-specific environmental variables (e.g., landscape, soil type, climate) and farmer decision making about technology (e.g., cultivar and management choices).

homogeneous farming areas = spatial regimes (in technologies)

Spatial regimes in agriculture

Recent studies strengthened the assumption that **geographical origin** does **leave a footprint** in food products and that both soil and substrate, in interaction with climate and cultural choices, influence the shaping of crop phenology and agricultural product quality (Vardour et al., 2015).

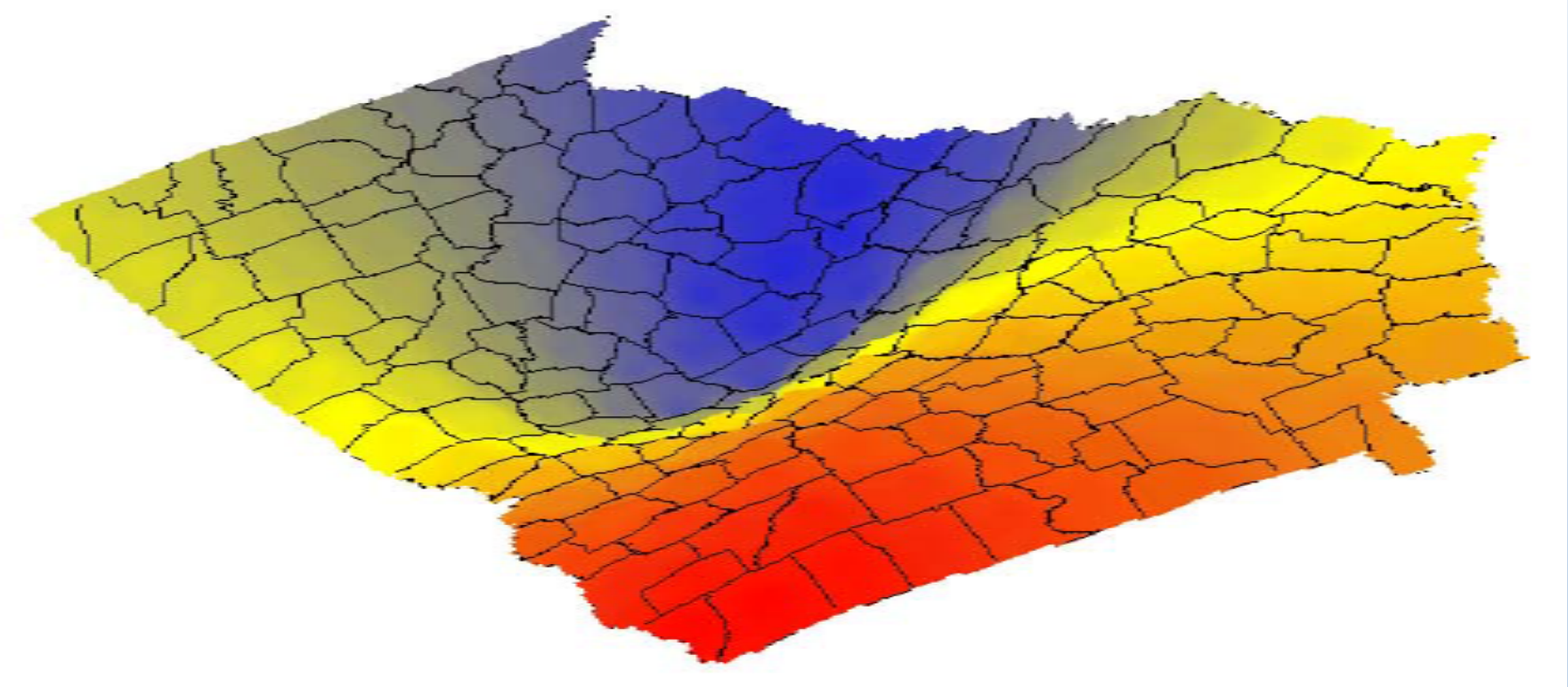
These issues **motivate** us to propose a **novel** approach to **endogenously identify homogeneous farming areas** , i.e. spatial clusters of farms that are homogeneous in terms of production technologies, based on a method, which exploits the information incorporated in the **geographic coordinates of farms**.

The geographical coordinates **subsume** the effects of a variety of complex, dynamic interactions among site-specific environmental variables and farmer decision making about technology that are often **not observed** at the farm level.

Output from GWR

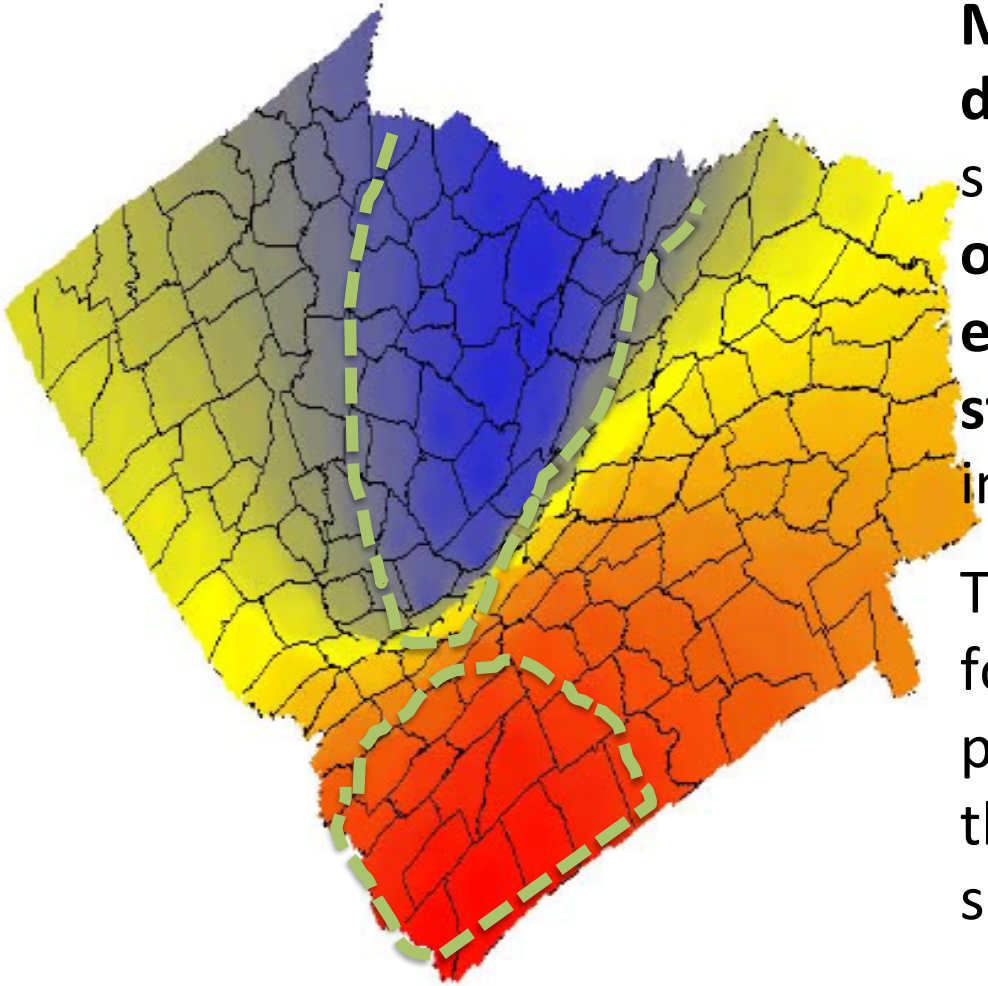
GWR allows **coefficient estimates to vary over space** by *calibrating* the global model separately for each spatial unit to produce *n* sets of parameter estimates, with the basic idea that local econometric models fit better the data in small geographic areas.

Main output from GWR is a set of **location-specific parameter estimates** which can be mapped and analysed to provide information on spatial non-stationarity. **GWRs interpret the spatially varying parameter problem** as a **smooth change over space (**continuous heterogeneity**)**



Spatial Regimes

Our aim: to identify **spatial clusters of farms** in which a single local econometric model is justified (*spatial regimes*), i.e. fix the borders



Motivation: Production technology **does not change smoothly** over space; rather, we expect that **farms operating in a proximal environment will not exhibit statistically significant differences** in production elasticities.

This expectation suggests that a form of spatially varying parameters can be interpreted as the presence of structural breaks in space.

Spatial Regimes vs mixture analysis

LCM accommodate for unobserved heterogeneity in technologies but **do not account for spatial heterogeneity**. In other words firms are separated in **not necessarily spatially contiguous** classes.
Regimes classifies farms in spatial regimes. The **advantage** is that the geographic **coordinates** of farms **incorporate the effects of a variety complex, dynamic interactions** among site-specific **environmental variables** and farmer **decision making about technology** often not observed at the farm level.

For example, agricultural surveys usually do not carry information about cultivars grown, which are key to correctly associate the climatic variables collected in local meteorological stations with farms in consideration of the differences in the timing of the phenological stages of cultivars.

The technology sets available to farms depend on the characteristics of the physical, social and economic environment in which production takes place.

In other words, the underlying production technology is not the same for all farms, rather it is location specific, and the group of farms sharing the same technology can be defined as **local technology cluster.**

Identification of K (fixed) Spatial Regimes : Simulated Annealing

Algorithm:

Step 1: generate a random code from 1 to K for every unit, it's a random partition. Estimate the Quantile regression in each region.
Step 2: for every unit, draw a random region code {1,2,…,K} different from the old one. Estimate again the Quantile regression in each region. If the objective function J(S), for a fixed configuration S of regions, is smaller that unit will belong to the new region otherwise accept the change with probability given by

$$P=\mathrm{Exp}\{ -[J(S_1)-J(S_2)]/t \}$$

where $J(S_1)$ is the obj. function at step 1 and $J(S_2)$ at step 2. t is the temperature that decreases with the augmenting of the iterations, i.e. at the beginning there is a high probability to accept worsenings of the objective function; at the end, when $t \rightarrow 0$ only reductions of the objective function are accepted.

Step 3….: step 2 is repeated many times until the objective is minimized (max number of iterations, no unit changes of region code, etc.).

Identification of K (fixed) Spatial Regimes : Simulated Annealing

$$J(S_j) = \hat{E}_j(\hat{\theta}, \mathbf{X}, k) - \alpha V_j(k) \cdots \left(1-\alpha\right) \hat{E}_j(\hat{\theta}, \mathbf{X}, k) + \alpha V_j(k)$$

where

$$\hat{E}_j(\hat{\theta}, \mathbf{X}, k) = \sum_{q=1}^{Q} \left(\sum_{\{i | y_i \geq X_i \beta_{\tau_q}\}} \tau_q \left| y_i - X_i \beta_{\tau_q} \right| + \sum_{\{i | y_i < X_i \beta_{\tau_\tau}\}} \left(1-\tau_q\right) \left| y_i - X_i \beta_{\tau_q} \right| \right)$$

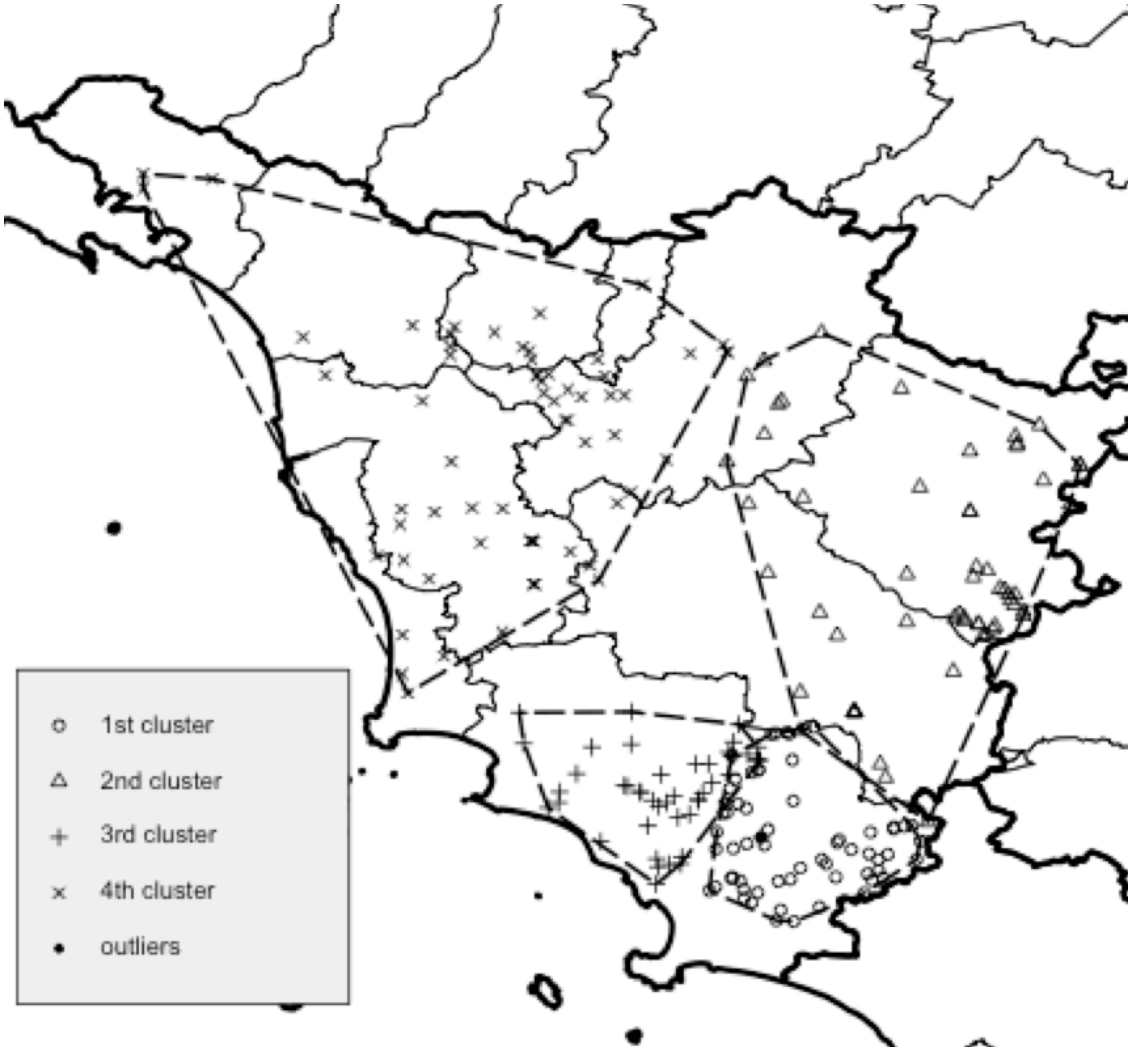
$$V_j(k) = \sum_{s=1}^{n} \sum_{v=1}^{n} c_{sv} \mathbf{1}\left(k(j)_s = k(j)_v\right)$$

This model discourages configurations with high sum of residuals and spatial regimes composed by not contiguous units.

Benedetti R., Pratesi M. and Salvati N., "Local Stationarity in Small Area Estimation Models", Statistical Methods and Applications, 22: 81-95, 2013

Postiglione P., Andreano S.M. and Benedetti R. "Using Constrained Optimization for the Identification of Convergence Clubs", Computational Economics, 42: 151-174, 2013

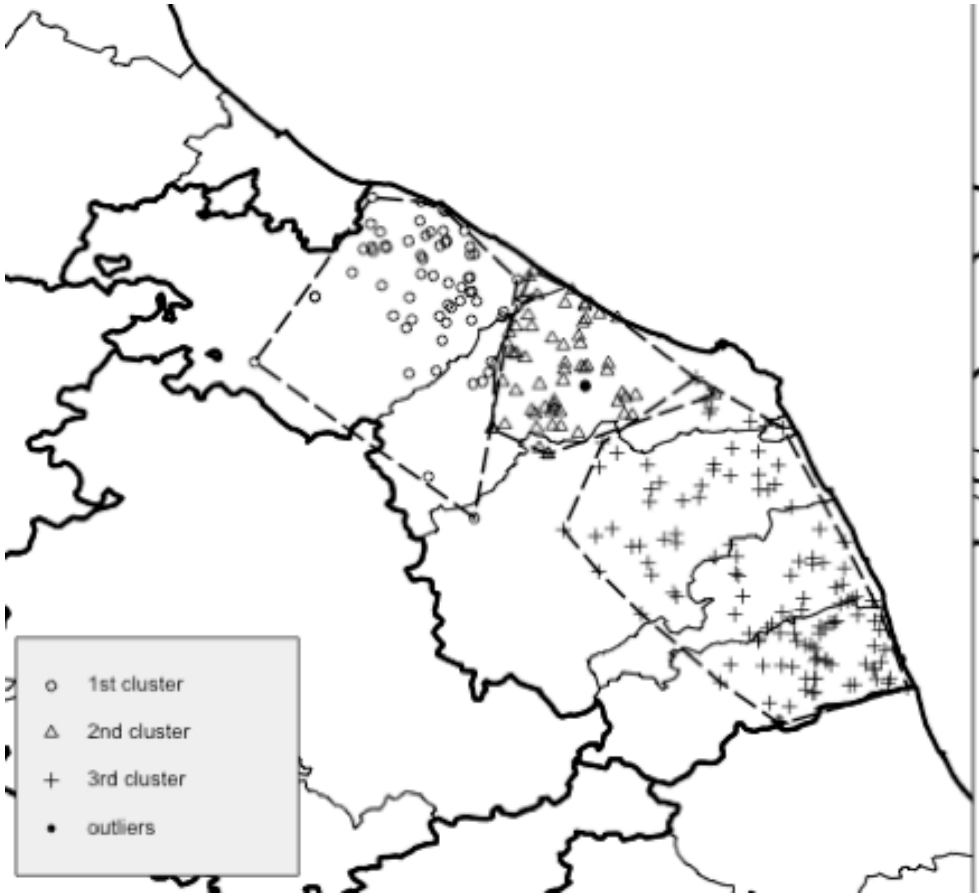
Spatial Regimes – Tuscany



Production regimes are **related** to the existence of some **latent,** not observed**, factors** that are closely **related to the spatial position** of the observed farms …. **climate, soil type, cultivars and** social factors.

high degree of overlapping between
- our clusters,
- map of varieties
*(available only at the **territorial level**)*

Spatial Regimes – Marche & Apulia



Empirical results

- Confirmed the existance of spatial regimes related to the existence of some **latent, not observed, factors that are closely related to the spatial position** of the observed farms (**climate, soil type, cultivars and** social factors).
- therefore empirical analysis that fails to incorporate parameter heterogeneity can produce misleading results.
- Once partitioned the study area we observed that the spatial interaction between farms belonging to the same cluster is not anymore significant, giving rise to the hypothesis that the effects of farm localization on production technology are mainly represented by heterogeneity, and not by spatial dependence.

Conclusions

Results are encouraging even if some additional research is needed

- the proposed strategy performs very well with large datasets, but we **need to explore the computational burden when huge sample sizes are used**

Future development

- from cross sectional to panel data model.


THANK YOU
FOR YOUR KIND ATTENTION