Spatial Econometric Modeling of Farm Data

1. Spatial Modeling of Farm Data

Since the publication in 2002 of the Special Issue in Agricultural Economics entitled "*Spatial Analysis for Agricultural Economists: Concepts, Topics, Tools and Examples*" the number of applications of spatial econometric methods in agricultural economics has steadily increased over time.

Spatial explicit models have been largely used to analyse the drivers of technology adoption (Case 1992; Florax, Voortman, and Brouwer 2002; Anselin, Bongiovanni, and Lowenberg-DeBoer 2004; Abdulai and Huffman 2005; Krishna and Qaim 2012; Krishnan and Patnam 2013; Lapple and Kelley 2015; Fang and Richards 2018; Lapple et al. 2017), with specific attention to organic farming (Lapple and Kelley 2015; Schmidtner et al. 2012; Wollni and Andersson 2014), and spatial regimes in technologies (Billé, Salvioni, and Benedetti 2018).

Another relevant stream of literature estimates spatial models that address both spatial dependence and spatial heterogeneity to explain variation in farmland values (Nickerson and Lynch 2001; Patton and McErlean 2003; Cavailhès and Wavresky 2003; Livanis et al. 2006; Kostov 2009; Maddison 2009; Wang 2018; Yang, Odening, and Ritter 2019).

The importance of neighbourhood effects has been recognized in studies explaining market participation (Holloway and Lapar 2007), farm household diversification (Corral and Radchenko 2017), farm survival (Saint-Cyr et al. 2018), land rental intention (Skevas, Skevas, and Swinton 2018), pesticide use (Aida 2018) and in the analysis of policy intervention in agriculture (Grogan

and Goodhue 2012; Storm, Mittenzwei, and Heckelei 2014; Marconi, Raggi, and Viaggi 2015; Feichtinger and Salhofer 2016; Fruh-Muller et al. 2019).

Finally, an increasing group of studies analyses the issue of spatial dependence in the analysis of technical efficiency in the context of stochastic frontier (Druska and Horrace 2004; Areal, Balcombe, and Tiffin 2012; Fusco and Vidoli 2013; Tsionas and Michaelides 2016; Pede et al. 2018).

In section 2 we focus on the frontier production function. Before presenting the R codes to fit and test spatial production and spatial frontier functions, we describe the general formulation of the production function and that of the frontier production function.

1.1.   Production function and the frontier production function

In agriculture there is a long tradition of statistics-gathering and it is one of the economic sectors with the deepest supply of micro data set on input and output from the production process. This explains the large amount of studies that estimate agricultural production functions, analyse productivity and measure the efficiency of farms.

A production function describes the technical relationship that transforms inputs into output. For each level of input use, the function assigns a unique output level. A general way of writing a production function is

$$y = f(x_1, \dots, x_n)$$

where y is an output and x_i are the productive inputs that can include labour, capital, knowledge (human capital), energy consumption, raw materials, natural resources (land, water, minerals), and others. It is usually assumed that production functions fulfil some properties: essentiality of

inputs, positive returns, diminishing returns and/or proportional returns to scale (see Chambers 1998, 9). Many mathematical specifications can be used to estimate the production function (Griffin, Montgomery, and Rister 1987). The choice of functional form brings a series of implications with respect to the shape of the isoquants and the values of elasticities of factor demand and factor substitution. The simplest specification is the linear production function:

$$y = \alpha + \sum_i \beta_i x_i$$

Despite its mathematical simplicity, this linear form is rarely used since it violates the property of essentiality.[1]

A widely used form is the Cobb Douglas production technology

$$y = A \prod_{i=1}^{N} x_i^{\alpha_i}$$

where A represents the Hicksian neutral efficiency level of firm i, which is unobserved by the econometrician. The natural logs transformation of the previous equation can be expressed as a linear equation in the form:

$$ln(y) = \alpha_0 + \sum_{i=1}^{N} \alpha_i \, ln(x_i)$$

---

[1] Essentiality of inputs: If at least one $x_i = 0$, then $y = 0$, *i.e.*, production is not possible without any of the inputs.

The Cobb Douglas function is often chosen because it has economic properties (diminishing returns to each input and constant returns to scale) superior to the simple linear function and because its parameters are easy to obtain from real data. Another largely used functional form is the translog for one output and K inputs

$$ln(y) = \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln(x_i) + \frac{1}{2} \sum_i \sum_j \alpha_{ij} \ln(x_i)\ln(x_j)$$

This specification is a more flexible extension of the Cobb-Douglas function, it fulfils a set of desirable characteristics and it is easy to derive and allowing the imposition of homogeneity.

Once a functional form is chosen, a key issue in the estimation of production functions is that direct OLS estimation of the production function is problematic because of endogeneity (Marschak and Andrews 1944; Ackerberg, Caves, and Frazer 2015).

Endogeneity can arise when observed inputs are correlated with unobserved shock. Under such circumstances OLS will yield biased and inconsistent estimates. In a linear framework, the standard approach for addressing the potential endogeneity bias is to use instrumental variables or fixed effects.

The estimated model of production is the means to the objective of measuring inefficiency. This is because the production function represents the maximum output attainable given a set of inputs. Measurement of (in)efficiency is, then, the empirical estimation of the extent to which observed agents (fail to) achieve the production frontier as originally argued by Debreu (1951) and Farrell (1957).

One of the main approaches to study productivity and efficiency is the Stochastic Frontier Model (SFM), independently proposed by Aigner, Lovell, and Schmidt (1977) and Meeusen and Broeck

(1977). The SFM is motivated by the theoretical idea that no economic agent can exceed the frontier and the deviations from this extreme represent the individual inefficiencies. For the description, implementation and testing of the non-spatial and spatial Stochastic Frontier Analysis models see section 2.

For a review some of the most important developments in the econometric estimation of the stochastic frontier models (*e.g.* endogeneity issues, recent advances in generalized panel data stochastic frontier models, etc.) see for example Greene (2008) and Kumbhakar, Parmeter, and Zelenyuk (2017).

In the stochastic production frontier estimation, the endogeneity of input problem has been neglected until recently. In such environment the endogeneity issue can be solved for example by using the semi-parametric approach proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003) (see for example Shee and Stefanou (2014), and Latruffe et al. (2017) for an application to agriculture).

2. Fitting and Testing Spatial Stochastic Frontier Models

In this section we explain how to fit and test spatial models with a specific focus on farm data. In particular, we define a spatial frontier model by making use of the stochastic frontier approach. Log-log transformations of the Cobb-Douglas production function can be used to implement the well-known spatial linear model specifications. For details on these types of models in the context of regional data, the reader is referred to Chapter 12. Note that in agricultural production function/stochastic frontier models a problem of potential endogenity of some inputs (regressors) is often present. In this section we assume the exogeneity of all the regressors in our

model specifications. For details on the use of the spatial linear production function and potential endogeneity see *e.g.* Billé, Salvioni, and Benedetti (2018).

In the following subsection, we consider the simulation setup for the definition of the true values of the parameters, the sample size, the generation of the spatial coordinates and the spatial weighting matrix used to assume a particular spatial process. The assumed weighting matrix/matrices is/are based on a k-nearest neighbour approach (k-nn), *i.e.* we define a Boolean matrix with the same number $k \in \mathbb{N}$ of nearest neighbours for each random variable in space. Let $W = \{w_{ij}\}$ be the spatial weighting matrix with elements equal to the weights among pairs of random variables (y_i, y_j) for $i, j = 1, \ldots, n$, with n the sample size, then

$$\begin{cases} w_{ij} = 1 & \Leftrightarrow \quad y_j \in \mathcal{N}_k \\ w_{ij} = 0 & otherwise \end{cases}$$

where $\mathcal{N}_k$ is the set of nearest random variables y_j to y_i defined by k. Finally, W is row-normalized such that $\sum_j w_{ij} = 1, \forall i$. Discussions on the definition on different spatial weighting matrices as well as model specifications can now be found in several spatial book references, see *e.g.* Anselin (1988), LeSage and Pace (2009), Elhorst (2014), Arbia (2014), Kelejian and Piras (2017).

To set the simulation setup, consider the following codes. We first load the following package

```
library(spdep)
```

in order to use some useful spatial functions inside, see Bivand and Piras (2015). Then we set the seed, the number of nearest neighbours k, the true values of the vector of parameters $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$ for the constant, Land, Labour and Capital inputs, the true values of the vector of parameters $\theta = (\theta_1, \theta_2, \theta_3)'$ for spatially-lagged Land, Labour and Capital inputs, and the true

values of the autoregressive coefficients, ρ and λ, in the dependent variables and among the error terms (or the inefficiencies), respectively, as follows

```
set.seed(3)
k       <- 30
beta    <- c(10, 0.5, 0.3, 0.2)
theta   <- c(0.6,0.2,0.2)
rho     <- 0.6
lambda <- 0.4
```

We set the sample size n and we generate the longitude and latitude coordinates by using two Uniform distributions from 0 to 50 and from -70 to 20, *i.e.* $\mathcal{U}(0,50)$ and $\mathcal{U}(-70,20)$, respectively,

```
n       <- 500
coords <- cbind(long = runif(n,0,50), lat = runif(n,-70,20))
head(coords)

##             long       lat
## [1,]   8.402076 -43.33272
## [2,]  40.375820 -49.23735
## [3,]  19.247118  12.29245
## [4,]  16.386716 -40.08144
## [5,]  30.105034 -19.92414
## [6,]  30.219703 -17.64876
```

and we generate an n-dimensional Identity matrix and the weighting matrix W by using some functions into the `spdep` package as follows

```
I_n  <- as(diag(n), "CsparseMatrix")
nb   <- knn2nb(knearneigh(coords, k = k, longlat=TRUE))
W    <- as(nb2mat(nb, style="W"), "CsparseMatrix")
```

In particular, the function `knearneigh` provides a list of class *knn* with the information into the first member of the region number ids to define the nearest neighbours for each random variable. The argument `longlat=TRUE` selects the Great Circle geographical distances among pairs of units in space. The `knn2nb` function transforms the object of class *knn* into an object of class *nb* (neighbour list), while the `nb2mat` function transforms an object of class *nb* into an n-dimensional weighting matrix. The argument `style = "W"` directly row-normalizes the weights.

Both the Identity matrix I_n and the weighting matrix W are sparse by using the function `as(,"CsparseMatrix")`.

Finally, we set the matrix of regressors which include the constant, Land, Labour and Capital inputs by drawing numbers from $\mathcal{U}(1.5,4)$

```
X <- cbind(constant = 1, A = runif(n,1.5,4), L = runif(n,1.5,4), K =
runif(n,1.5,4))
```

This section is devoted to the description, implementation and testing of the Stochastic Frontier Analysis models (SFA) and of a recent extension of SFA, called Spatial Stochastic Frontier Analysis (SSFA).

A caveat, however, is mandatory: the number of different techniques proposed in the literature for estimating production (or cost) efficiency is wide, differentiating among parametric (as SFA, Aigner, Lovell, and Schmidt (1977) and Meeusen and Broeck (1977), R packages `Benchmarking`, `frontier`), non-parametric (as Data Envelopment Analysis (DEA), Farrell (1957) and Charnes, Cooper, and Rhodes (1978), R packages `Benchmarking`, `nonparaeff` or `FEAR` Wilson (2008)) or semi-parametric techniques (Park and Simar (1994), Kuosmanen and Kortelainen (2012) and Ferrara and Vidoli (2017), R package `semsfa`).

If, at first glance, non-parametric techniques seem to be particularly flexible and generalizable, the main disadvantage lies precisely in their deterministic nature, since it is not even possible to recognize if the difference in terms of efficiency among units is caused by technical inefficiency or by exogenous/accidental effects (Fried and Lovell 2008). The parametric model of stochastic frontier overcomes the main limits associated with deterministic models, providing a detailed analysis of the inefficiency sources that are not directly associated to farm policy and/or random

disturbances, too. On the other hand, the most significant disadvantage associated with the SFA approach is the lack of of flexibility associated with the specification of a given functional form.

The SFA approach implies the construction of the stochastic optimum frontier, based on an underlying production/cost function, identified through the relative comparison of the firm performance in a set economic system. The observed deviations from the optimum frontier may be split into the combination of two effects: the effect caused by the random noise and the technical/cost inefficiency. More formally, the stochastic frontier model can be written as:

$$y_i = f(x_i; \beta) + v_i - u_i, i = 1, \dots, n$$

where $Y_i \in R_+$ is the single output of unit i, $X_i \in R_+^p$ is the vector of p inputs, $f(\cdot)$ defines a production (frontier) relationship between inputs X and the single output Y, v_i is a symmetric two-sided error representing random effects, usually assumed Normal $v \sim N(0, \sigma_v^2)$, and $u_i > 0$ is one-sided error term which represents technical inefficiency, usually assumed Half-Normal ($u \sim N^+(0, \sigma_u^2)$). Please note that *(i)* the inefficiency and the error terms must both be orthogonal to the input, output, or to the other variables used in the functional specification, and *(ii)* it is usually assumed that v and u are each identically independently distributed (*i.i.d.*).

Starting from the DGP designed in the previous chapter (same simulation setup, sample size, spatial coordinates and spatial weighting matrix), the next step is merely to set up the inefficiency (positive) term u using the `rsn` function (`sn` package) that generates a random sample of n units from a skew-normal distribution with 1 as location parameter and 4 as the scale parameter.

```
u   <- abs(rsn(n, 0, 1, 4))
```

This hypothesis, namely the independence among production units, can be often violated especially in the regional applications where exogenous conditional factors or the different resiliencies of specific territories may lead to comparative advantages or disadvantages for units within the same region. For this reason, and to test different models of stochastic frontier, the inefficiency term u has been simulated as dependent by two distinct effects (which in practical applications can be substitutive or complementary): *(i)* a global spatial spillover effect and *(ii)* a drift effect.

From an economic point of view, therefore, u_2 may catch the spillover effect from which neighbouring units may benefit, influencing each other and benefiting from the agglomeration or Marshallian atmospheric externalities or may grasp the different resilience of the territories as a consequence of an economic shock. More formally, u_2 can be defined as: $u_2 = (I_n - \lambda W)^{-1} * u$.

Practically in R, u_2 can be expressed as:

```
u2 <- solve(I_n - lambda * W)%*%u
```

The second effect, can be found in some regional or agricultural applications, is the drift effect, that is linked not so much to the neighbourhood as to the physical position, in terms of latitude and longitude, of the units in the analysed region. A different impact on land productivity, for example, may be the result of a different spatial distribution of temperatures or rainfall. Obviously, the presence of the two spatial effects may or may not overlap depending on the application problem.

Given these premises, we modify u_2 in such a way that the inefficiency of the units is function of their absolute position in the region, too, multiplying u_2, previously obtained, with drift normalized value function of latitude and longitude.

```
drift = 1 - 0.5*(coords[,1] + mean(coords[,1]))*(coords[,2] + mean(coords[,2]))
u2  <- u2 * drift/mean(drift)
```

We have, therefore, all the elements to calculate a simulated y (named *y_sfa*) that takes into account the covariates X (Land, Labour and Capital inputs as defined in the previous section), the random error term, but also an inefficiency term that depends on the neighbourhood and the relative position of the single unit in the region.

```
y_sfa  <- as.matrix(X%*%beta + eps - u2)
```

It is so possible to assess how the standard SFA model help to estimate the simulated DGP and whether the basic assumptions regarding errors (*i.i.d.*) are respected even in the presence of a strong component linked to the economic space within which farms produce. Different packages are available on CRAN; in this application exercise, benchmarking package has been chosen for the large number of options and functions complementary to the main function sfa. Please note that, in this package, X (first parameter) and y (second parameter) must be passed as matrix.

```
library(Benchmarking)
x_sfa = as.matrix(cbind(df$A, df$L, df$K))
y_sfa = as.matrix(y_sfa)
fit.sfa<- sfa(x_sfa,y_sfa)
summary(fit.sfa)

##             Parameters   Std.err   t-value   Pr(>|t|)
## (Intercept)    10.7183   0.43005    24.924    0.000
## x1              0.5646   0.08844     6.384    0.000
## x2              0.1849   0.08545     2.164    0.030
## x3              0.1223   0.08579     1.426    0.154
## lambda          2.9148   0.43275     6.736    0.000
## sigma2          5.1373
## sigma2v =   0.5409906 ;   sigma2u =   4.596338
## log likelihood =   -887.1153
## Convergence =   4 ; number of evaluations of likelihood function 24
## Max value of gradien: 7.733437e-06
## Length of last step: 0
## Final maximal allowed step length: 0.33075
```

Standard diagnostics of the SFA function (obtained by the usual function *summary*) reports a good adaptation of the model to the data (the estimated β are very similar to the simulated ones,

the intercept is higher - as it must be for the estimation of the productive frontier); different from other non-frontier models two key values of this analysis are reported here: σ_v^2 (equal to 0.719) and σ_u^2 (equal to 3.367). They represent, respectively, the estimated variance of the random component and the inefficiency; if compared to the total estimated variance ($\sigma^2 = 4.087$) it is possible to evaluate how the part of inefficiency ($\sigma_u^2/\sigma^2 = 3.367/4.087 = 82\%$) is greater than the random one ($\sigma_v^2/\sigma^2 = 0.719/4.087 = 18\%$), as proof of a good adaptation of the model to the data and a good differentiation of the units in terms of estimated inefficiency. Standard diagnostics, therefore, does not report any warning in the estimated model nor is there any standard test available on the basic assumptions. In other terms, the spatial dependence of inefficiency is not grasped and this issue may be all the more serious as environmental and contextual factors are important in explaining efficiency differentials among different territories.

To test the spatial autocorrelation among the estimated efficiency, Geary test (through the function `geary.test` of the `spdep` package) can be used; please note that, in this simulation, testing the estimated efficiencies (estimated by the `eff` function, `benchmarking` package) or residues leads to very similar results since the simulated random part (v) is not spatially autocorrelated.

```
Wnb <- nb2listw(nb)
geary.sfa.eff <- geary.test(as.vector(eff(fit.sfa)),listw = Wnb)
geary.sfa.eff

##
##   Geary C test under randomisation
##
## data:   as.vector(eff(fit.sfa))
## weights: Wnb
##
## Geary C statistic standard deviate = 21.235, p-value < 2.2e-16
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
## Geary C statistic       Expectation           Variance
##      0.7468869459      1.0000000000        0.0001420774
```

Geary test (equal to 0.743) and Figure 1 show, as expected, a strong spatial autocorrelation in efficiency estimates: in particular, the first two figures at the top show respectively the spatial distribution and the kernel distribution by quadrant of the efficiencies highlighting a strong difference among the quadrants themselves; the graph at the bottom shows the Moran plot.
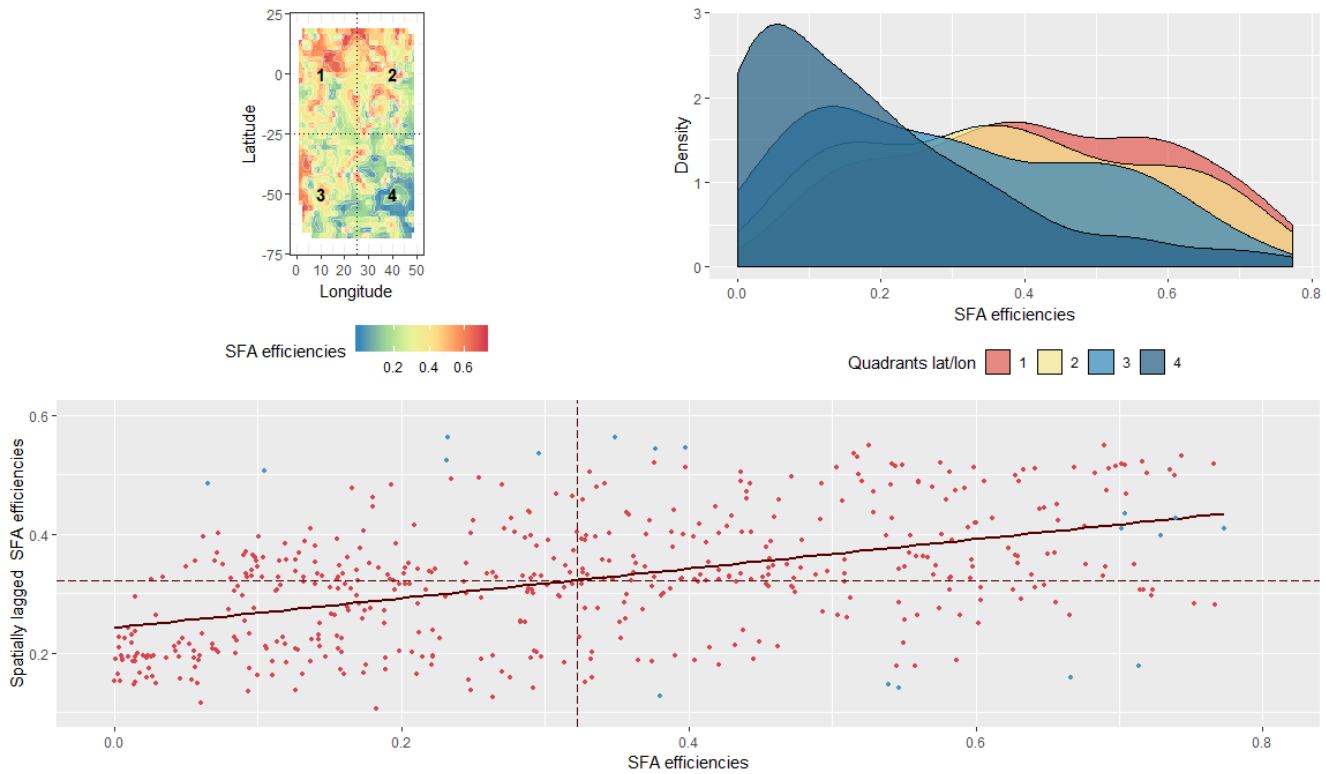

Figure 1 SFA efficiencies

Once it has been demonstrated that it is the external conditions that led to greater or lesser efficiency and not so much the allocation of the single producer, what method can be used to split the territorial from the individual effect? An effective solution for cross-section data - and a generalization of the SFA - has been proposed by Fusco and Vidoli (2013) with a method called "*Spatial Stochastic Frontier Analysis*" (SSFA); they proposed to "*split the inefficiency term into three parts: the first related to spatial peculiarities of the territory in which each single unit operates, the second related to the specific production features and the third the random error*

term". More in deep, Fusco and Vidoli (2013) propose a maximum likelihood solution to this model:[2]

$$y_i = f(x_i; \beta) + v_i - (1 - \rho \sum_i w_{i.}) \hat{u}_i, i = 1, \dots, n$$

where $\hat{u}_i$ and v_i are independently distributed of each other and of the regressors. R package `ssfa` provides the `ssfa` function that allows to estimate this specific model; please note, in particular, the need to indicate the spatial weight matrix in the `data_w` parameter.

```
library(ssfa)
fit.ssfa <- ssfa(y_sfa ~ A+L+K , data_w=W, data=df, form = "production",
par_rho=TRUE, intercept = TRUE)
summary(fit.ssfa)

## Spatial Stochastic frontier analysis model
##
##                 Estimate   Std. Error   z value Pr(>|z|)
## Intercept    1.04612e+01 1.98829e-01 52.61399  < 2e-16 ***
## A            8.34315e-01 1.57974e-01  5.28134  < 2e-16 ***
## L            4.76402e-01 1.51928e-01  3.13572 0.001714 **
## K            8.18766e-01 1.90999e-01  4.28675  1.8e-05 ***
## sigmau2_dmu 9.89902e+00 4.80903e-01 20.58426  < 2e-16 ***
## sigmav2     2.38643e-06 1.58278e-06  1.50774 0.131620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pay attention:
## 1 - classical SFA sigmau2 = sigmau2_dmu + sigmau2_sar: 10.312152 where
sigmau2_sar: 0.413129
## 2 - sigma2 = sigmau2_dmu + sigmau2_sar + sigmav2: 13.767977
##
## Inefficiency parameter Lambda = sigmau_dmu/sigmav: 4148044.355609
##
## Spatial parameter Rho: 0.303932
##
## Moran I statistic: 0.141358
```

[2] For homogeneity with Fusco and Vidoli (2013) paper and the SSFA diagnostic, the autoregressive coefficient λ for the error term will be from now on renamed as ρ.

```
## 
## Mean efficiency: 0.07876
## 
## LR-test: sigmau2_dmu = 0 (inefficiency has no influence to the model)
## H0: sigmau2_dmu = 0 (beta_ssfa = beta_ols)
## 
##        Value Log-Lik
## ssfa    -1041.3802
## ols      -914.6648
## 
## Value LR-Test: -253.431 p-value 0
## 
## AIC: 2110.76, (AIC for lm: 1837.33)
```

Even in this case, the β parameters of the simulations have been estimated correctly[3], but the most interesting issue to consider concerns the opportunity to split the inefficiency variance σ_u^2 into two terms: the individual part (named by authors σ_{dmu}^2) and a neighbourhood related one (named σ_{sar}^2).

In the specific simulation, please note how the individual effect ($\sigma_{dmu}^2 = 1.417$) is more important than the the spatial one ($\sigma_{sar}^2 = 0.353$), but this part still represents about 25% of the individual effect. Another key parameter to take into account is the spatial parameter ρ equal to 0.857; in this case, therefore, SSFA diagnostic shows a strong spatial correlation between the inefficiency estimates (please note that the function `eff.ssfa` estimate the efficiency of each producer without spatial effects). Replicating the Geary test on the SSFA efficiency (equal to 1.033) and analysing Figure 2, it can be noted how the SSFA model has succeeded in isolating the

---

[3] Please note the higher value of the intercept; as stated by Vidoli et al. (2016) it may occur a *"shift in the production curve with respect to the SFA as a consequence of the isolation of the spatial effect, transforming the average value of β_0 into a multiplicity of individual effects"*.

territorial effect from the efficiency estimates; of course, the differences between the SFA and SSFA estimates may be of interest in the study of the determinants of local development.

```
geary.ssfa.eff <- geary.test(as.vector(eff.ssfa(fit.ssfa)),listw = Wnb)
geary.ssfa.eff
```

```
##
##   Geary C test under randomisation
##
## data:  as.vector(eff.ssfa(fit.ssfa))
## weights: Wnb
##
## Geary C statistic standard deviate = 2.2942, p-value = 0.01089
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
## Geary C statistic        Expectation            Variance
##      0.9443271280       1.0000000000        0.0005888975
```
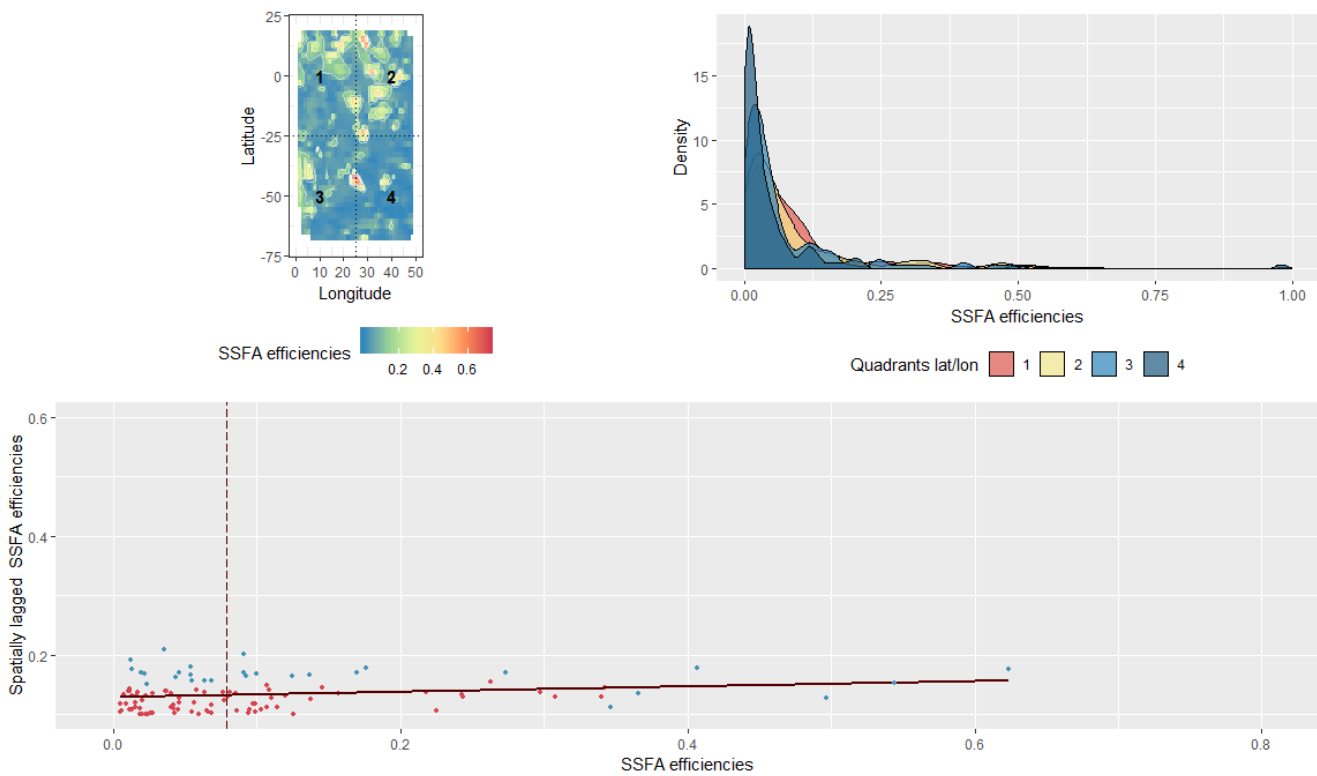

Figure 2 SSFA efficiencies

3. Controlling for Unobserved Spatial Heterogeneity

In this chapter we explain how to fit and test spatial production models when farm data exhibit spatial heterogeneity. The form of spatial heterogeneity is typically unknown, *i.e.* unobserved, and it can be ascribed either to mean instability or to heteroskedasticity. The omission of such spatial effects leads to biased inference (Anselin 1988; LeSage and Pace 2009; Le Gallo 2014).

Mean instability may imply local clustering of the values of a variable. In the case of agricultural production, mean instability implies that the territory can be divided into clusters of farms each one using the same production technology. In other words, the production function coefficients differ according to a number of distinct unknown spatial production regimes (goupwise heterogeneity), see *e.g.* Anselin (2010). As explained in Billé, Salvioni, and Benedetti (2018), these local clusters in technology can emerge as a result of dynamic interactions among site-specific environmental variables and farmer decision making about technology. For example, farmers usually choose to grow those varieties that are best suited to the environmental conditions in which the farm operates. In turn, the choice of a specific variety is often connected to specific management systems such as water management or timing of harvesting. The technology prevailing in each local technology cluster is the efficient solution to the specific techno-economic problem faced by farms operating in that portion of the territory. This view is consistent with evolutionary theories (Nelson and Winter 1982; Dosi 1988) according to which firms cannot be assumed to operate using a single common production function.

The zoning of spatial regimes in farm technologies is largely unknown and their identification needs comprehensive spatial modelling of soil, agronomical and climatic properties, including their changes through time, hence the processing of large quantities of data acquired at a very

fine spatial resolution, while researchers can usually rely only on a few control variables. An alternative way to identify the spatial regimes in farm technologies relies on the use of the Earth coordinates of latitude and longitude to proxy the micro-geographic determinants of production that are unknown to the econometrician. In subsection 3.1 we show a possible way to identify unobserved spatial regimes and then explain how to properly estimate more flexible spatial models, see Billé, Benedetti, and Postiglione (2017).

Spatial heterogeneity may also arise due to heteroskedasticty, that is when the variances of error temrs vary over space causing the instability of the functional form. One solution to this problem has been recently proposed by Chasco, Le Gallo, and López (2018). Also in the case of heteroskedasticity, the territory can then be divided into clusters of farms characterized by significantly different groupwise variances, i.e. unobserved spatial groupwise heteroskedasticity. The identification of such clusters in the error terms is not considered in Billé, Salvioni, and Benedetti (2018).

In subsection 3.2 we show how another possible solution to the problem of the identification of contiguous spatial clusters is to use a graph-based approach, which captures the adjacency relations between object, farms in our case. Differently from the previous case, in this section spatial homogeneous areas are not defined according to a functional relationship, rather on some farm or territorial characteristics likewise to the standard cluster methods.

In both the following subsections we basically assume the same sample size $n = 500$, generation of the spatial coordinates, spatial weighting matrices, generation of the regressors X and of the innovations ϵ of section 3. By setting the same seed we can include in our new database the original X. We also consider the same simulation setup for the identification of the clusters as in

the following. Let first define the dataframe with the coordinates of section 2 and with an indicator variable that associates each point to a cluster

```
dataset   <- as.data.frame(coords)
set.seed(3)
dataset$A <- runif(n,1.5,4)
dataset$L <- runif(n,1.5,4)
dataset$K <- runif(n,1.5,4)
dataset$clu <- ifelse(dataset$long < 20 & dataset$lat < -20 , 1, 0)
dataset$clu <- ifelse(dataset$long > 20 & dataset$lat < -40 , 2, dataset$clu)
dataset$clu <- ifelse(dataset$long < 20 & dataset$lat > -40 , 3, dataset$clu)
dataset$clu <- ifelse(dataset$long > 20 & dataset$long < 40 & dataset$lat > -40 &
dataset$lat < 10, 4, dataset$clu)
dataset$clu <- ifelse(dataset$long > 20 & dataset$lat > 0   , 5, dataset$clu)
dataset$clu <- ifelse(dataset$long > 40 & dataset$lat > -40 , 5, dataset$clu)
```

Figure 1 shows the generated clusters/regimes in space. Note that these spatial regimes can be also not geographically well-defined, *i.e.* points that belong to the same cluster can be also sparsely-distributed in space. For each of them we need to assume a different spatial production function.
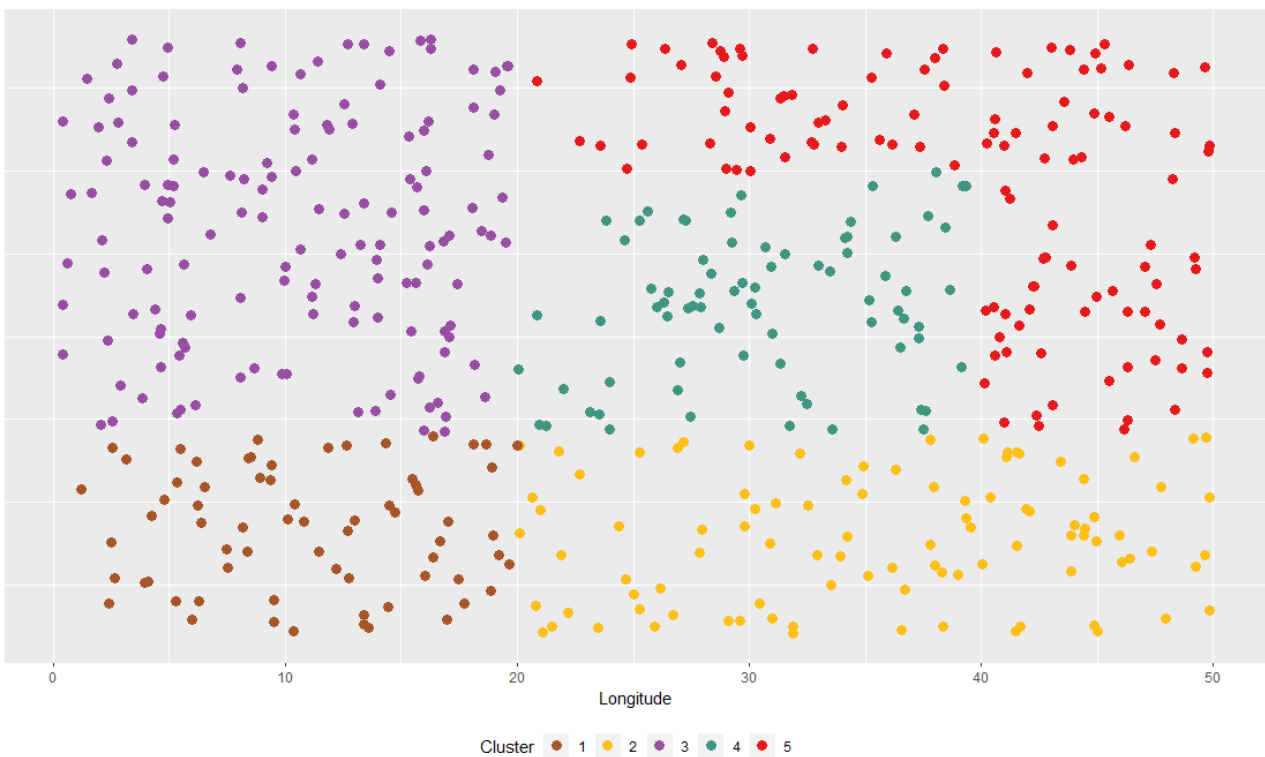
Figure 1 Example of Spatial Regimes

Suppose we have two different true DGPs: *(i)* an OLS model with regimes, *(ii)* a SARAR/SAC model with regimes. Then, the following codes generate different production functions for different clusters

```r
beta1 <- c(13,0.5,0.3,0.2)
beta2 <- c(11,0.8,0.1,0.1)
beta3 <- c(9,0.3,0.2,0.5)
beta4 <- c(7,0.4,0.3,0.3)
beta5 <- c(5,0.2,0.6,0.2)

dataset$y_ols <- ifelse(dataset$clu==1,13 + 0.5*dataset$A + 0.3*dataset$L +
0.2*dataset$K + eps, 0)
dataset$y_ols <- ifelse(dataset$clu==2,11 + 0.8*dataset$A + 0.1*dataset$L +
0.1*dataset$K + eps, dataset$y_ols)
dataset$y_ols <- ifelse(dataset$clu==3,9 + 0.3*dataset$A + 0.2*dataset$L +
0.5*dataset$K + eps, dataset$y_ols)
dataset$y_ols <- ifelse(dataset$clu==4,7 + 0.4*dataset$A + 0.3*dataset$L +
0.3*dataset$K + eps, dataset$y_ols)
dataset$y_ols <- ifelse(dataset$clu==5,5 + 0.2*dataset$A + 0.6*dataset$L +
0.2*dataset$K + eps, dataset$y_ols)
dataset$y_sac <- ifelse(dataset$clu==1,as.matrix(A_rho_inv%*%(X%*%beta1 +
B_lambda_inv%*%eps)), 0)
dataset$y_sac <- ifelse(dataset$clu==2,as.matrix(A_rho_inv%*%(X%*%beta2 +
B_lambda_inv%*%eps)), dataset$y_sac)
dataset$y_sac <- ifelse(dataset$clu==3,as.matrix(A_rho_inv%*%(X%*%beta3 +
B_lambda_inv%*%eps)), dataset$y_sac)
dataset$y_sac <- ifelse(dataset$clu==4,as.matrix(A_rho_inv%*%(X%*%beta4 +
B_lambda_inv%*%eps)), dataset$y_sac)
dataset$y_sac <- ifelse(dataset$clu==5,as.matrix(A_rho_inv%*%(X%*%beta5 +
B_lambda_inv%*%eps)), dataset$y_sac)
head(dataset)

##         long       lat        A        L        K clu     y_ols      y_sac
## 1   8.402076 -43.33272 1.920104 2.240758 1.540459   1  12.80053  37.03793
## 2  40.375820 -49.23735 3.518791 2.076740 1.891127   2  12.94834  33.57304
## 3  19.247118  12.29245 2.462356 3.785901 1.758021   3  11.45821  29.45562
## 4  16.386716 -40.08144 2.319336 2.331071 2.664323   1  15.58018  39.85193
## 5  30.105034 -19.92414 3.005252 2.890996 2.832992   4  10.10911  24.67326
## 6  30.219703 -17.64876 3.010985 2.954201 2.368872   4  10.01893  24.28573
```

where the column "clu" refers to the specific cluster and the columns "y_ols" and "y_sac" to the generated dependent variables in the linear and spatial case, respectively. In the following

subsection we formally define a SARAR/SAC model with regimes. Alternative spatial models with regimes can be straightforwardly specified.

3.1.   Spatial Regimes

In this subsection we explain how to estimate spatial autoregressive models with pre-identified production regimes; the way in which we can identify unobserved spatial regimes can be made with different approaches, like *e.g.* mixture models, see *e.g.* Greene (2005) and Emvalomatis (2012), or iteratively locally weighted regressions, see Billé, Benedetti, and Postiglione (2017) and Billé, Salvioni, and Benedetti (2018). The procedure in Billé, Benedetti, and Postiglione (2017) and Billé, Salvioni, and Benedetti (2018) is based on a continuous smoothing updating algorithm of the weights used to repeated local estimations, by making use of the Wald test statistics to simultaneously compare vector of beta coefficients. The algorithm is available upon request.

Let's assume we know the form of the clusters in space. Then, a SARAR/SAC model with regimes can be defined in the following way

$$\dot{y} = \rho W_1 \dot{y} + \dot{X} \dot{\beta} + \dot{u} \quad \dot{u} = \lambda W_2 \dot{u} + \dot{\varepsilon} \quad \dot{\varepsilon} \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$$

where $\dot{y} = \{\dot{y}_j\}$ is a partitioned n-dimensional column vector of dependent variables with generic vector element $\dot{y}_j$ for $j = 1, \dots, c$ and c is the total number of regimes, $\dot{X} = \{\dot{X}_j\}$ is an $n \times (k \times c)$ block-diagonal matrix with generic matrix element $\dot{X}_j$ for $j = 1, \dots, c$, $\dot{\beta} = \{\dot{\beta}_j\}$ is $(k \times c)$ partitioned column vector of parameters with generic vector element $\dot{\beta}_j$ for $j = 1, \dots, c$, $\dot{\varepsilon} = \{\dot{\varepsilon}_j\}$ is a partitioned n-dimensional column vector with generic element $\dot{\varepsilon}_j \sim \mathcal{N}(0, \sigma_{\varepsilon_j}^2 I_{n_j})$ for $j =$

$1, ..., c$, while (ρ, λ) and (W_1, W_2) are defined in section 3. By letting $\rho = \lambda = 0$ the model is defined as a linear model (OLS) with regimes.

By using the clusters generated in Figure 1, we now estimate the OLS model with and without regimes and the SARAR/SAC model with and without regimes in the following way

```
CLU <- as.factor(dataset$clu)
fit.ols   <- lm(dataset$y_ols ~ dataset$A + dataset$L + dataset$K)
fit.ols.r <- lm(dataset$y_ols ~ (CLU:(dataset$A + dataset$L + dataset$K + CLU)) +
0)
fit.sac <- sacsarlm(dataset$y_sac ~ dataset$A + dataset$L + dataset$K,
                data=dataset,listw=listw,type="sac",method="eigen")
fit.sac.r <- sacsarlm(dataset$y_sac ~ (CLU:(dataset$A + dataset$L + dataset$K +
CLU))
                      + 0,data=dataset,listw=listw,type="sac",method="eigen")
```

where $fit.ols.r$ and $fit.sac.r$ are the linear and spatial models with regimes, respectively. By using the function "summary" we can see *e.g.* the difference in the estimates between the spatial model with regimes (correct specification) and the spatial model without regimes as follows

```
summary(fit.sac)

##
## Call:
## sacsarlm(formula = dataset$y_sac ~ dataset$A + dataset$L + dataset$K,
##     data = dataset, listw = listw, type = "sac", method = "eigen")
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -8.081323 -1.064716  0.097549  1.097462  7.355392
##
## Type: sac
## Coefficients: (asymptotic standard errors)
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 18.46596   14.81133  1.2467   0.2125
## dataset$A   -0.82323    0.58894 -1.3978   0.1622
## dataset$L   -2.53457    2.13584 -1.1867   0.2354
## dataset$K   -0.18637    0.14083 -1.3234   0.1857
##
## Rho: 0.68998
## Asymptotic standard error: 0.28374
##     z-value: 2.4317, p-value: 0.015026
## Lambda: 0.83349
## Asymptotic standard error: 0.18597
##     z-value: 4.4819, p-value: 7.3987e-06
```

```
## 
## LR test value: 463.07, p-value: < 2.22e-16
## 
## Log likelihood: -1116.847 for sac model
## ML residual variance (sigma squared): 4.8025, (sigma: 2.1915)
## Number of observations: 500
## Number of parameters estimated: 7
## AIC: 2247.7, (AIC for lm: 2706.8)
```

summary(fit.sac.r)

```
## 
## Call:sacsarlm(formula = dataset$y_sac ~ (CLU:(dataset$A + dataset$L +
##     dataset$K + CLU)) + 0, data = dataset, listw = listw, type = "sac",
##     method = "eigen")
## 
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.331635 -0.714640   0.022138   0.723094   2.782272
## 
## Type: sac
## Coefficients: (asymptotic standard errors)
##                    Estimate Std. Error z value  Pr(>|z|)
## CLU1             36.5398258  3.6448095 10.0252 < 2.2e-16
## CLU2             32.8433967  3.6672349  8.9559 < 2.2e-16
## CLU3             29.2626066  2.8784764 10.1660 < 2.2e-16
## CLU4             28.3887092  3.7131831  7.6454 2.087e-14
## CLU5             18.6113983  3.5142146  5.2960 1.183e-07
## CLU1:dataset$A  -0.1747229  0.5781469 -0.3022   0.76249
## CLU2:dataset$A  -0.0168871  0.3791063 -0.0445   0.96447
## CLU3:dataset$A  -0.1388111  0.4762281 -0.2915   0.77068
## CLU4:dataset$A  -0.0886247  0.6568061 -0.1349   0.89267
## CLU5:dataset$A   0.4818576  0.4812195  1.0013   0.31667
## CLU1:dataset$L   0.9166597  0.9939641  0.9222   0.35641
## CLU2:dataset$L   0.5052904  0.9412775  0.5368   0.59140
## CLU3:dataset$L  -0.0339779  0.4444444 -0.0765   0.93906
## CLU4:dataset$L  -1.3998376  0.6727553 -2.0808   0.03746
## CLU5:dataset$L  -0.2251711  0.5023752 -0.4482   0.65400
## CLU1:dataset$K   0.0816411  0.1743796  0.4682   0.63966
## CLU2:dataset$K  -0.0144752  0.1528647 -0.0947   0.92456
## CLU3:dataset$K   0.0105306  0.1235978  0.0852   0.93210
## CLU4:dataset$K   0.0034255  0.1809109  0.0189   0.98489
## CLU5:dataset$K  -0.3080458  0.1398329 -2.2030   0.02760
## 
## Rho: 0.017777
## Asymptotic standard error: 0.054341
##     z-value: 0.32714, p-value: 0.74356
## Lambda: 0.71
## Asymptotic standard error: 0.079216
##     z-value: 8.9629, p-value: < 2.22e-16
## 
## LR test value: 21.358, p-value: 2.3021e-05
## 
```

```
## Log likelihood: -730.6234 for sac model
## ML residual variance (sigma squared): 1.0628, (sigma: 1.0309)
## Number of observations: 500
## Number of parameters estimated: 23
## AIC: 1507.2, (AIC for lm: 1524.6)
```

As it can be observed, the significance of the estimated beta coefficients may change with respect to the type of cluster we consider. In this particular case, according to the results of the spatial model without regimes, only the variable Capital K is significant in all the clusters considered, but the significance level depends on the different clusters. Interestingly, one or more spatial autocorrelation coefficients might be no more statistically significant, as ρ in this case. Details on this aspect have been shown in Billé, Benedetti, and Postiglione (2017).

To test if the partition of the spatial data, *i.e.* clusters/regimes, is statistically significant one can use the Chow test, see Chow (1960), and the Spatial Chow test, see Anselin (1990), for the presence of (spatial) structural breaks. Indeed, identifying clusters in space, by comparing *e.g.* the significance of the beta coefficients, is simply a way to econometrically finding structural breaks in a spatial process. The statistics of the Chow test and the Spatial Chow test are defined, respectively, as follows

$$C = \frac{(e_r'e_r - e_u'e_u)/k}{e_u'e_u/(n-2k)} \sim F_{k,n-2k}$$

$$C_s = \frac{(e_r'\Psi^{-1}e_r - e_u'\Psi^{-1}e_u)}{\sigma^2} \sim \chi_k$$

where e_r is the vector of residuals from the restricted model (model without regimes), e_u is the vector of residuals from the unrestricted model (model with regimes), Ψ is the variance-covariance matrix of the spatial model and σ^2 is the error variance for either the restricted model, the unrestricted model, or both.

The functions of the Chow test and the Spatial Chow test in R, see Anselin (2005), are written as follows

```
chow.test <- function(rest,unrest)
{
  er <- residuals(rest)
  eu <- residuals(unrest)
  er2 <- sum(er^2)
  eu2 <- sum(eu^2)
  k <- rest$rank
  n2k <- rest$df.residual - k
  c <- ((er2 - eu2)/k) / (eu2 / n2k)
  pc <- pf(c,k,n2k,lower.tail=FALSE)
  list(c,pc,k,n2k)
}
spatialchow.test <- function(rest,unrest)
{
  lrest <- rest$LL
  lunrest <- unrest$LL
  k <- rest$parameters - 2
  spchow <- - 2.0 * (lrest - lunrest)
  pchow <- pchisq(spchow,k,lower.tail=F)
  list(spchow,pchow,k)
}
```

Both of them compare the restricted model (without regimes) with the unrestricted model (with regimes). The spatial version takes the spatial dependence into account.

```
Ct.ols <- chow.test(fit.ols,fit.ols.r)[[2]]
Ct.ols

## [1] 3.837563e-126

SCt.sac <- spatialchow.test(fit.sac,fit.sac.r)[[2]]
SCt.sac

## [1] 1.05556e-164
```

As we can observe from the $p-values$, both the tests reject the null hypothesis of absence of spatial regimes.

3.2.　　Spatially Constrained Clustering

In this section the *Skater* procedure (Spatial K'luster Analysis by TreeEdgeRemoval, Assuncao et al. (2006)) for the identification of homogeneous contiguous areas according to a spatial neighborhood is described; differently from the previous one, in this section a regressive relationship between the output and the inputs is not a priori assumed and therefore spatial homogeneous areas are not defined according to a functional relationship, but to some characteristics likewise to the standard cluster methods.

In other terms, the described method can be useful in determining contiguous spatial clusters in which some contextual variables, exogenous to a given production process, describe a similar level of demand and/or supply of the territory.

From an application point of view, starting from the DGP simulated in the previous section, we must therefore construct some contextual exogenous variables, correlated with *E(Y|X)*, that represent some factors external to the farm (such as weather or different production areas).

Given that, the variables *Z1* and *Z2* have been calculated according to the following steps: *(1) Z1* and *Z2* are generated, *(2)* the contextual variables are centered and scaled, *(3)* the correlation among variables is removed according the Cholesky matrix transformation,[4] *(4)* a variance-covariance matrix is set and then *(5)* it's used in order to transform the original *Z1* and *Z2*; finally *(6)* the operation of centering and scale is removed.

---

[4] Cholesky matrix transformation let to transform uncorrelated variables into correlated ones according to a set variance-covariance matrix, but it is also useful for reverse operation.

```
### Step 1
ZZ <- cbind(dataset$y_sac /dataset$L,
            Z1=rnorm(dim(dataset)[1]),
            Z2=rnorm(dim(dataset)[1]))
### Step 2
mns <- apply(ZZ, 2, mean)
sds <- apply(ZZ, 2, sd)
ZZ2 <- sweep(ZZ, 2, mns, FUN="-")
ZZ2 <- sweep(ZZ2, 2, sds, FUN="/")
### Step 3
v.obs <- cor(ZZ2)
ZZ3 <- ZZ2 %*% solve(chol(v.obs))
### Step 4
r <- cbind( c(1, 0.7, 0.3),
            c(0.7, 1, 0.03),
            c(0.3, 0.03, 1))
### Step 5
ZZ4 <- ZZ3 %*% chol(r)
### Step 6
ZZ4 <- sweep(ZZ4, 2, sds, FUN="*")
ZZ4 <- sweep(ZZ4, 2, mns, FUN="+")
dataset$Z1 = ZZ4[,2]
dataset$Z2 = ZZ4[,3]
```

Several analytical regionalisation methods (also known as "*spatially constrained clustering*") have

been proposed in literature (see *e.g.* (Murtagh 1985) and (Duque, Ramos, and Suriñach 2007));

Skater procedure, proposed in the `spdep` package, offers many advantages, including simplicity

of application, the hierarchical nature of the method and the possibility of binding the minimum

number of units within each cluster.

So let's use the simulated exogenous variables *Z*:

```
datasetx <- dataset[,c("Z1","Z2")]
```

and load the coordinates using the `spdep` library; from these coordinates, the neighbours list

have been calculated thanks to `tri2nb` function.

```
library(spdep)
coords = coordinates(cbind(dataset$lon,dataset$lat))
neighbours = tri2nb(coords, row.names = NULL)
```

We, therefore, have all the objects that allow us to practically estimate the contiguous homogeneous areas according to the *Skater* algorithm; first of all, a cost scheme has to be set given the neighbourhood using the `nbcosts` function

```
lcosts <- nbcosts(neighbours, datasetx)
```

After computed the cost of each edge - as the distance between it nodes - the neighbours list with spatial weights has to be set as:

```
nb.w <- nb2listw(neighbours, lcosts, style="B")
```

The weighted neighbours list (plotted in Figure 2, left plot) is often - in the practical applications - too complex to evaluate; the "*minimum spanning tree*" algorithm - as stated before - allows to simplify the structure of the initial graph with the aim of achieving a minimum path among units/nodes.

```
mst.bh <- mstree(nb.w,5)
```

Figure 2 (right plot) highlights how all units are always connected and how these links are reduced compared to the original neighbourhood ones.

```
par(mar=c(0,0,0,0))
par(mfrow=c(1,2))
plot(nb.w, coords)
plot(mst.bh, coords, col=2, cex.lab=.5, cex.circles=0.035, fg="blue", bty="n")
```
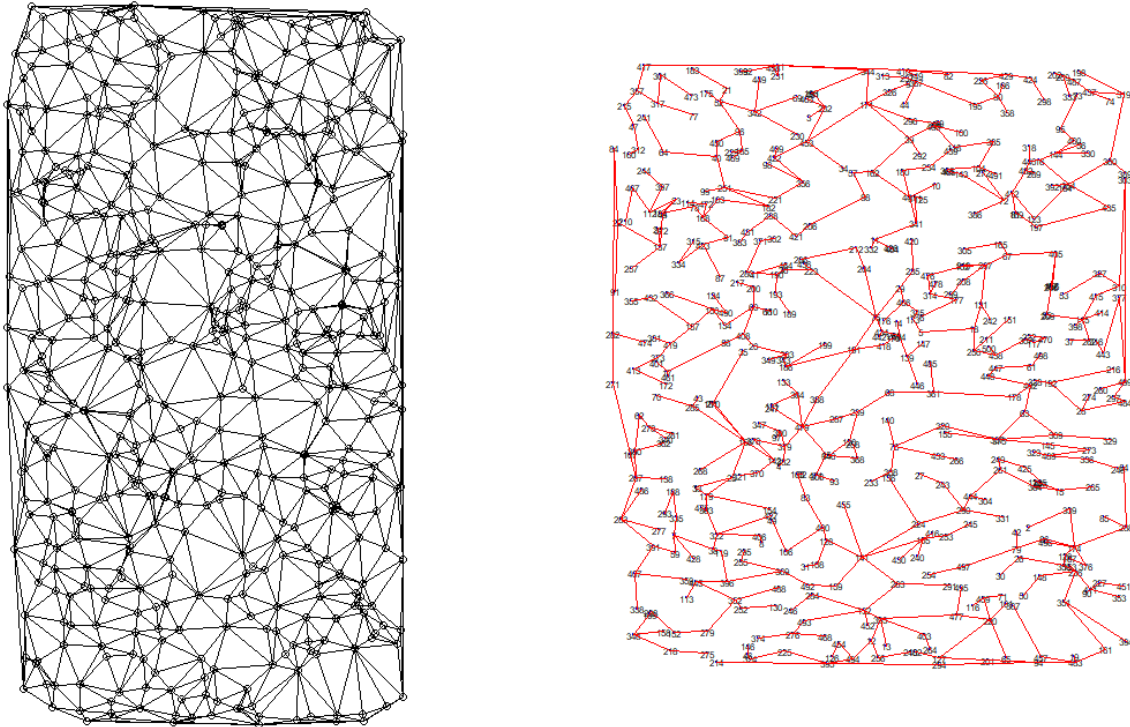
Figure 2 Full weighted neighbours and minimum spanning tree

It is now possible to apply the *Skater* clustering procedure; being a hierarchical procedure derived from the *k-means* algorithm, - in addition to the demand/supply variables and the neighbourhood structure - it requires the indication of the preferred number of homogeneous clusters in the `ncuts` argument. Other important options are available in the `skater` function: the most useful one - in our opinion for application purposes - is `crit` that allows to indicate the minimum number of units that must be present in each cluster in order to do not identify clusters too small or affected by outlier data in the demand/supply variables. For the sake of clarity, three thresholds/cuts have been chosen.

```
ska1 <- skater(mst.bh[,1:2], datasetx, ncuts=4, crit=30)
ska2 <- skater(mst.bh[,1:2], datasetx, ncuts=5, crit=30)
ska3 <- skater(mst.bh[,1:2], datasetx, ncuts=6, crit=30)
dataset$cluster_SKATER1 = ska1$groups
dataset$cluster_SKATER2 = ska2$groups
dataset$cluster_SKATER3 = ska3$groups
```

Figure 3 shows the units attribution varying parameter k; a reasonable stability of the solutions can be noted.
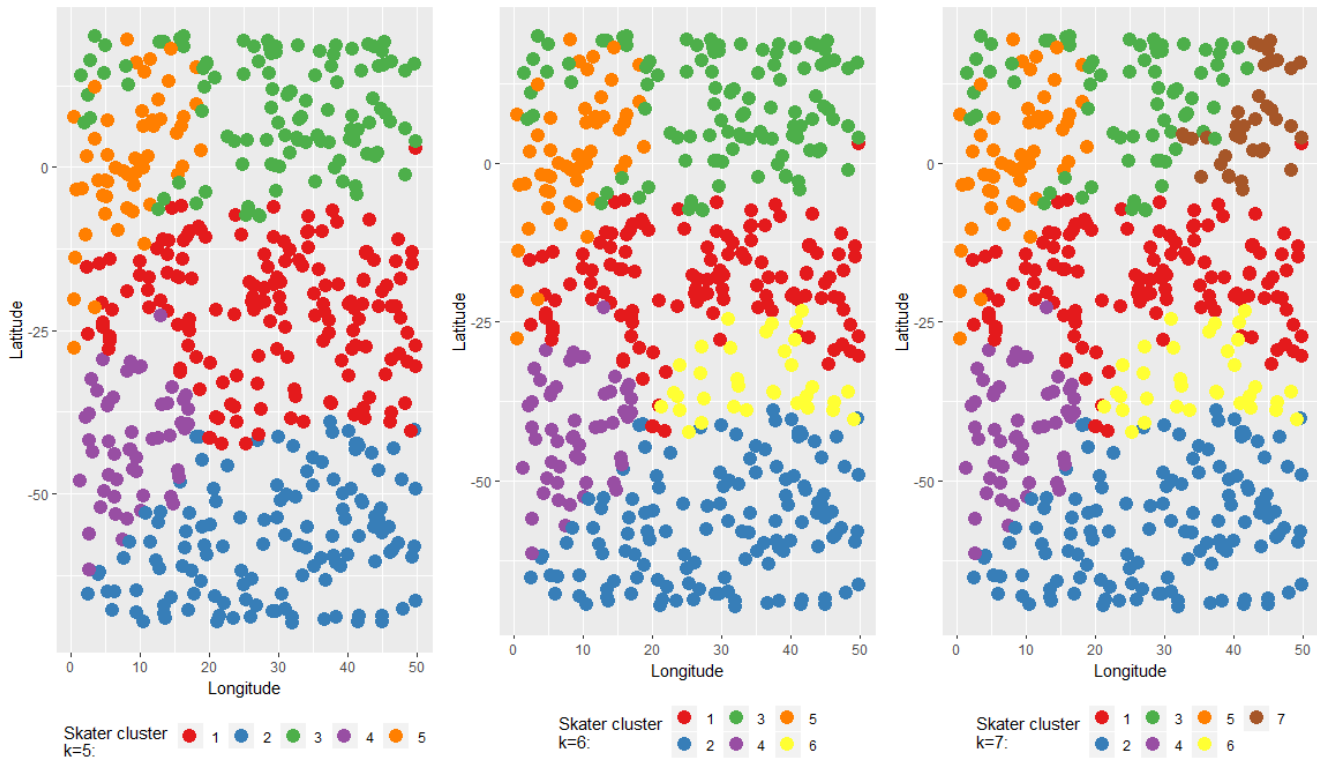

Figure 3 Skater cluster varying k

The obtained spatial clusters can therefore be used to describe homogeneous demand areas or - for example - be prodromal to subsequent regressive or causal analyses. Similarly to the previous section, we have chosen to use these spatial clusters in a regressive model to better explain the spatial differences in the relationship between y and X among units.

```
SKAT <- as.factor(dataset$cluster_SKATER1)
ly = log(dataset$y_sac)
lA = log(dataset$A)
lL = log(dataset$L)
lK = log(dataset$K)
fit.ols     <- lm(ly ~ lA + lL + lK)
fit.ols.sk <- lm(ly ~ (SKAT:(lA + lL + lK + SKAT)) + 0)
```

Tables 1 and 2 and their relative fitting measures (Adjusted R2 respectively equal to 0.6494513 and 0.9993578) show how - for construction - the augmented model can better describe the relationships between inputs and output.

Table 1 OLS

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.217	0.037	112.457	0.000
lA	-0.378	0.021	-17.592	0.000
lL	-0.527	0.021	-25.184	0.000
lK	0.004	0.022	0.193	0.847

Table 2 OLS con cluster

	Estimate	Std. Error	t value	Pr(>|t|)
SKAT1	3.722	0.087	42.625	0.000
SKAT2	3.715	0.052	71.245	0.000
SKAT3	3.636	0.186	19.507	0.000
SKAT4	4.292	0.113	37.920	0.000
SKAT5	3.291	0.206	15.989	0.000
SKAT1:lA	-0.544	0.029	-19.015	0.000
SKAT2:lA	-0.105	0.034	-3.071	0.002
SKAT3:lA	-0.468	0.035	-13.275	0.000
SKAT4:lA	0.013	0.101	0.129	0.898
SKAT5:lA	0.046	0.105	0.435	0.664
SKAT1:lL	0.025	0.070	0.360	0.719
SKAT2:lL	-0.068	0.060	-1.127	0.260
SKAT3:lL	-0.025	0.135	-0.188	0.851
SKAT4:lL	-0.929	0.120	-7.731	0.000
SKAT5:lL	0.048	0.176	0.270	0.787
SKAT1:lK	0.020	0.026	0.785	0.433
SKAT2:lK	0.011	0.027	0.401	0.688
SKAT3:lK	0.043	0.030	1.417	0.157
SKAT4:lK	-0.032	0.046	-0.699	0.485
SKAT5:lK	0.020	0.048	0.415	0.678

What we are interested in is not only a better adaptation of the model to the data - as it is tautological to expect from our simulation - but also a lower spatial correlation in the estimation residuals. The Moran's I test confirms that the spatial autocorrelation in residuals of the augmented model reduces by half respect to the linear production model (function `lm.morantest`).

```
lm.morantest(fit.ols, nb2listw(neighbours, style="W"))

##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = ly ~ lA + lL + lK)
## weights: nb2listw(neighbours, style = "W")
##
## Moran I statistic standard deviate = 24.608, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Observed Moran I        Expectation           Variance
##     0.6320120827     -0.0060215765      0.0006722517

lm.morantest(fit.ols.sk, nb2listw(neighbours, style="W"))

##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = ly ~ (SKAT:(lA + lL + lK + SKAT)) + 0)
## weights: nb2listw(neighbours, style = "W")
##
## Moran I statistic standard deviate = 11.918, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Observed Moran I        Expectation           Variance
##     0.2744713564     -0.0228614607      0.0006223683
```

4. References

Abdulai, A., and W.E. Huffman. 2005. "The Diffusion of New Agricultural Technologies: The Case of Crossbred-Cow Technology in Tanzania." *American Journal of Agricultural Economics* 87 (3): 645–59.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–51.

Aida, T. 2018. "Neighbourhood Effects in Pesticide Use: Evidence from the Rural Philippines." *Journal of Agricultural Economics* 69 (1): 163–81.

Aigner, D., C. A. K. Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6: 21–37.

Anselin, L., R. Bongiovanni, and J. Lowenberg-DeBoer. 2004. "A Spatial Econometric Approach to the Economics of Site-Specific Nitrogen Management in Corn Production." *American Journal of Agricultural Economics* 86 (3): 675–87.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Vol. 4. Springer Science & Business Media.

———. 1990. "Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis." Journal of Regional Science 30 (2): 185–207.

———. 2005. "Spatial Regression Analysis in R: A Workbook." Urbana 51.

———. 2010. "Thirty Years of Spatial Econometrics." Papers in Regional Science 89 (1): 3–25.

Arbia, Giuseppe. 2014. *A Primer for Spatial Econometrics: With Applications in R*. Springer.

Areal, F.J., K. Balcombe, and R. Tiffin. 2012. "Integrating Spatial Dependence into Stochastic Frontier Analysis." *Australian Journal of Agricultural and Resource Economics* 56 (4): 521–41.

Assuncao, R.M., M.C. Neves, G. Camara, and C. Da Costa Freitas. 2006. "Efficient Regionalization Techniques for Socio-Economic Geographical Units Using Minimum Spanning Trees." International Journal of Geographical Information Science 20(7): 797–811.

Billé, A. G., R. Benedetti, and P. Postiglione. 2017. "A Two-Step Approach to Account for Unobserved Spatial Heterogeneity." Spatial Economic Analysis 12 (4): 452–71.

Billé, A. G., C. Salvioni, and R. Benedetti. 2018. "Modelling Spatial Regimes in Farms Technologies." *Journal of Productivity Analysis* 49 (2): 173–85.

Bivand, Roger, and Gianfranco Piras. 2015. "Comparing Implementations of Estimation Methods for Spatial Econometrics." In. American Statistical Association.

Case, A. 1992. "Neighborhood Influence and Technological Change." *Regional Science and Urban Economics* 22: 491–508.

Cavailhès, J., and P. Wavresky. 2003. "Urban Influences on Periurban Farmland Prices." *European Review of Agricultural Economics* 30 (3): 333–57.

Chambers, G. 1998. *Applied Production Analysis: A Dual Approach*. Cambridge University Press, UK.

Charnes, A., W. Cooper, and W. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2(4): 429–44.

Chasco, Coro, Julie Le Gallo, and Fernando A López. 2018. "A Scan Test for Spatial Groupwise Heteroscedasticity in Cross-Sectional Models with an Application on Houses Prices in Madrid." Regional Science and Urban Economics 68: 226–38.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." Econometrica: Journal of the Econometric Society, 591–605.

Corral, P., and N. Radchenko. 2017. "What's so Spatial About Diversification in Nigeria?" *World Development* 95: 231–53.

Debreu, G. 1951. "The Coefficient of Resource Utilization." *Econometrica* 19: 273–92.

Dosi, G. 1988. "Sources, Procedures, and Microeconomic Effects of Innovation." J Econ Lit 26: 1120–71.

Druska, V., and W.C. Horrace. 2004. "Generalized Moments Estimation for Spatial Panel Data: Indonesian Rice Farming." *American Journal of Agricultural Economics* 86 (1): 185–98.

Duque, Juan Carlos, Raúl Ramos, and Jordi Suriñach. 2007. "Supervised Regionalization Methods: A Survey." International Regional Science Review 30 (3): 195–220.

Elhorst, J Paul. 2014. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Vol. 479. Springer.

Emvalomatis, Grigorios. 2012. "Adjustment and Unobserved Heterogeneity in Dynamic Stochastic Frontier Models." Journal of Productivity Analysis 37 (1): 7–16.

Fang, Di, and T.J. Richards. 2018. "New Maize Variety Adoption in Mozambique: A Spatial Approach." *Canadian Journal of Agricultural Economics*.

Farrell, M. J. 1957. "The Measurement of Productive Efficiency." *Journal of the Royal Statistic Society* 120 (III): 253–81.

Feichtinger, P., and K. Salhofer. 2016. "The Fischler Reform of the Common Agricultural Policy and Agricultural Land Prices." *Land Economics* 92 (3): 411–32.

Ferrara, G., and F. Vidoli. 2017. "Semiparametric Stochastic Frontier Models: A Generalized Additive Model Approach." *European Journal of Operational Research* 258 (2): 761–77.

Florax, R.J., R.L. Voortman, and J. Brouwer. 2002. "Spatial Dimensions of Precision Agriculture: A Spatial Econometric Analysis of Millet Yield on Sahelian Coversands." *Agricultural Economics* 27: 425–43.

Fried, H. O., and C. A. K. Lovell. 2008. *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press, USA.

Fruh-Muller, A., M. Bach, L. Breuer, S. Hotes, T. Koellner, C. Krippes, and V. Wolters. 2019. "The Use of Agri-Environmental Measures to Address Environmental Pressures in Germany: Spatial Mismatches and Options for Improvement." *Land Use Policy* 84: 347–62.

Fusco, Elisa, and Francesco Vidoli. 2013. "Spatial Stochastic Frontier Models: Controlling Spatial Global and Local Heterogeneity." *International Review of Applied Economics* 27 (5): 679–94.

Greene, William. 2005. "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model." Journal of Econometrics 126 (2): 269–303.

Greene, William H. 2008. "The Econometric Approach to Efficiency Analysis." *The Measurement of Productive Efficiency and Productivity Growth* 1 (1): 92–250.

Griffin, R.C., J.M. Montgomery, and M.E. Rister. 1987. "Selecting Functional Form in Production Function Analysis." *Western Journal of Agricultural Economics* 12 (2): 216–27.

Grogan, K.A., and R.E. Goodhue. 2012. "Spatial Externalities of Pest Control Decisions in the California Citrus Industry." *Journal of Agricultural and Resource Economics*, 156–79.

Holloway, G., and M.L.A. Lapar. 2007. "How Big Is Your Neighbourhood? Spatial Implications of Market Participation Among Filipino Smallholders." *Journal of Agricultural Economics* 58 (1): 37–60.

Kelejian, Harry, and Gianfranco Piras. 2017. *Spatial Econometrics*. Academic Press.

Kostov, P. 2009. "Spatial Dependence in Agricultural Land Prices: Does It Exist?" *Agricultural Economics* 40 (3): 347–53.

Krishna, V.V., and M. Qaim. 2012. "Bt Cotton and Sustainability of Pesticide Reductions in India." *Agricultural Systems* 107: 47–55.

Krishnan, P., and M. Patnam. 2013. "Neighbors and Extension Agents in Ethiopia: Who Matters More for Technology Adoption?" *American Journal of Agricultural Economics* 96 (1): 308–27.

Kumbhakar, Subal C., Christopher F. Parmeter, and Valentin Zelenyuk. 2017. "Stochastic Frontier Analysis: Foundations and Advances." Working Papers 2017-10. University of Miami, Department of Economics.

Kuosmanen, T., and M. Kortelainen. 2012. "Stochastic Non-Smooth Envelopment of Data: Semi-Parametric Frontier Estimation Subject to Shape Constraints." *Journal of Productivity Analysis*, 1–18.

Lapple, D., G. Holloway, D.J. Lacombe, and C. O'Donoghue. 2017. "Sustainable Technology Adoption: A Spatial Analysis of the Irish Dairy Sector." *European Review of Agricultural Economics* 44 (5): 810–35.

Lapple, D., and H. Kelley. 2015. "Spatial Dependence in the Adoption of Organic Drystock Farming in Ireland." *European Review of Agricultural Economics* 42: 315–37.

Latruffe, Laure, Boris E Bravo-Ureta, Alain Carpentier, Yann Desjeux, and Victor H Moreira. 2017. "Subsidies and Technical Efficiency in Agriculture: Evidence from European Dairy Farms." *American Journal of Agricultural Economics* 99 (3): 783–99.

Le Gallo, J. 2014. "Cross-Section Spatial Regression Models." Handbook of Regional Science, 1511–33.

LeSage, JP, and R Kelley Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman &Hall/CRC.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2): 317–41.

Livanis, G., C.B. Moss, V.E. Breneman, and R.F. Nehring. 2006. "Urban Sprawl and Farmland Prices." *American Journal of Agricultural Economics* 88 (4): 915–29.

Maddison, D. 2009. "A Spatio-temporal Model of Farmland Values." *Journal of Agricultural Economics* 60 (1): 171–89.

Marconi, V., M. Raggi, and D. Viaggi. 2015. "Assessing the Impact of Rdp Agri-Environment Measures on the Use of Nitrogen-Based Mineral Fertilizers Through Spatial Econometrics: The Case Study of Emilia-Romagna." *Ecological Indicators* 59: 27–40.

Marschak, J., and W. Andrews. 1944. "Random Simultaneous Equations and the Theory of Production." *Econometrica,* 12 (3/4): 143–205.

Meeusen, W., and J. van den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18: 435–44.

Moran, Patrick AP. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1/2): 17–23.

Murtagh, Fionn. 1985. "A Survey of Algorithms for Contiguity-Constrained Clustering and Related Problems." The Computer Journal 28 (1): 82–88.

Nelson, R., and S.G. Winter. 1982. An Evolutionary Theory of Economic Change. Harvard Business School Press, Cambridge.

Nickerson, C.J., and L. Lynch. 2001. "The Effect of Farmland Preservation Programs on Farmland Prices." *American Journal of Agricultural Economics* 83 (2): 341–51.

Olley, G., and A. Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–97.

Park, B. U., and L. Simar. 1994. "Efficient Semiparametric Estimation in a Stochastic Frontier Model." *Journal of the American Statistical Association* 89 (427): 929–36.

Patton, M., and S. McErlean. 2003. "Spatial Effects Within the Agricultural Land Market in Northern Ireland." *Journal of Agricultural Economics* 54 (1): 35–54.

Pede, V.O., F.J. Areal, A. Singbo, J. McKinley, and K. Kajisa. 2018. "Spatial Dependency and Technical Efficiency: An Application of a Bayesian Stochastic Frontier Model to Irrigated and Rainfed Rice Farmers in Bohol, Philippines." *Agricultural Economics* 49 (3): 301–12.

Saint-Cyr, L.D., H. Storm, T. Heckelei, and L. Piet. 2018. "Heterogeneous Impacts of Neighbouring Farm Size on the Decision to Exit: Evidence from Brittany." *European Review of Agricultural Economics* 46 (2): 237–66.

Schmidtner, E., C. Lippert, B Engler, A.M. Hearing, J. Aurbacher, and S. Dabbert. 2012. "Spatial Distribution of Organic Farming in Germany: Does Neighbourhood Matter?" *European Review of Agricultural Economics* 39: 661–83.

Shee, Apurba, and Spiro E Stefanou. 2014. "Endogeneity Corrected Stochastic Production Frontier and Technical Efficiency." *American Journal of Agricultural Economics* 97 (3): 939–52.

Skevas, T., I. Skevas, and S.M. Swinton. 2018. "Does Spatial Dependence Affect the Intention to Make Land Available for Bioenergy Crops?" *Journal of Agricultural Economics* 69 (2): 393–412.

Storm, H., K. Mittenzwei, and T. Heckelei. 2014. "Direct Payments, Spatial Competition, and Farm Survival in Norway." *American Journal of Agricultural Economics* 97 (4): 1192–1205.

Tsionas, E.G., and P.G. Michaelides. 2016. "A Spatial Stochastic Frontier Model with Spillovers: Evidence for Italian Regions." *Scottish Journal of Political Economy* 63 (3): 243–57.

Vidoli, Francesco, Concetta Cardillo, Elisa Fusco, and Jacopo Canello. 2016. "Spatial Nonstationarity in the Stochastic Frontier Model: An Application to the Italian Wine Industry." *Regional Science and Urban Economics* 61: 153–64.

Wang, H. 2018. "The Spatial Structure of Farmland Values: A Semiparametric Approach." *Agricultural and Resource Economics Review* 47 (3): 568–91.

Wilson, P.W. 2008. "FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R." *Socio-Economic Planning Sciences* 42: 247–54.

Wollni, M., and C. Andersson. 2014. "Spatial Patterns of Organic Agriculture Adoption: Evidence from Honduras." *Ecological Economics* 97: 120–28.

Yang, X., M. Odening, and M. Ritter. 2019. "The Spatial and Temporal Diffusion of Agricultural Land Prices." *Land Economics* 95 (1): 108–23.