CONTRIBUTI DI RICERCA CRENOS

CLUSTERING SPACE-TIME SERIES: A FLEXIBLE STAR APPROACH

Edoardo Otranto
Massimo Mucciardi

WORKING PAPERS

2017/07

CENTRO RICERCHE ECONOMICHE NORD SUD
(CRENOS)
UNIVERSITÀ DI CAGLIARI
UNIVERSITÀ DI SASSARI

CRENOS was set up in 1993 with the purpose of organising the joint research effort of economists from the two Sardinian universities (Cagliari and Sassari) investigating dualism at the international and regional level. CRENoS' primary aim is to improve knowledge on the economic gap between areas and to provide useful information for policy intervention. Particular attention is paid to the role of institutions, technological progress and diffusion of innovation in the process of convergence or divergence between economic areas. To carry out its research, CRENoS collaborates with research centres and universities at both national and international level. The centre is also active in the field of scientific dissemination, organizing conferences and workshops along with other activities such as seminars and summer schools.
CRENoS creates and manages several databases of various socio-economic variables on Italy and Sardinia. At the local level, CRENoS promotes and participates to projects impacting on the most relevant issues in the Sardinian economy, such as tourism, environment, transports and macroeconomic forecasts.

www.crenos.it
info@crenos.it

CRENoS – CAGLIARI
VIA SAN GIORGIO 12, I-09100 CAGLIARI, ITALIA
TEL. +39-070-6756406; FAX +39-070- 6756402

CRENoS - SASSARI
VIA MURONI 23, I-07100 SASSARI, ITALIA
TEL. +39-079-213511

Title: CLUSTERING SPACE-TIME SERIES: A FLEXIBLE STAR APPROACH

First Edition: September 2017

Clustering Space-Time Series: A Flexible STAR Approach

Edoardo Otranto

Department of Economics and CRENoS,

University of Messina, Piazza Pugliatti, 1 - 98122 Messina, Italy (*eotranto@unime.it*)

Massimo Mucciardi *

Department of Economics,

University of Messina, Piazza Pugliatti, 1 - 98122 Messina, Italy (*mucciard@unime.it*)

*Corresponding author: Phone +39 0906413784; Fax:+39 0906413784

Abstract

The STAR model is widely used to represent the dynamics of a certain variable recorded at several locations at the same time. Its advantages are often discussed in terms of parsimony with respect to space-time VAR structures because it considers a single coefficient for each time and spatial lag. This hypothesis can be very strong; we add a certain degree of flexibility to the STAR model, providing the possibility for coefficients to vary in groups of locations. The new class of models is compared to the classical STAR and the space-time VAR by simulations and an application.

Keywords: clustering; forecasting; space–time models; spatial weight matrix.

Jel Codes: C30, C38, C50, J11

1 Introduction

Since the mid-seventies the research in statistical models describing the space–time evolution of real series has been widely diffused with several methodological contributions and applications, devoted to capturing both the dynamics along time and the correlations based on spatial relationships. After the seminal paper of Cliff and Ord (1975), the space-time models were extended from Pfeifer and Deutsch (1980), who propose the Space–Time ARIMA (STARIMA) class of models, an extension of the ARIMA class developed for time series to include the linear dependencies in both space and time dimensions; spatial dependencies are imposed by means of a spatial weight matrix, which incorporates spatial features such as distances between locations and neighboring sites. The consideration of the spatial structure in economic, social and environmental data sets is present in several papers; excellent reviews can be found in Anselin (1988) and Haining (1990).

The importance of considering the presence of spatial dependencies in the forecasting performance of the models was verified by Giacomini and Granger (2004) and Arbia, Bee and Espa (2011). They utilized Monte Carlo experiments to show that the use of separate univariate forecasts for each region, ignoring spatial dependence (in other words, considering only the time dimension), leads to highly inaccurate forecasts. They used the most widespread space-time

model, the so–called STAR(1,1) (the acronym STAR stands for Space–Time AutoRegressive), where the same coefficients, referred to the time and spatial lag, are used for the full time span and the full location set.

The success of the STAR model is due to its simple (linear) form and the possibility of including the effects of spatial autocorrelation in forecasting, because the spatial effects are considered with a time lag, differently from the purely Spatial AutoRegressive (SAR; Besag, 1974) model, where the spatial relationships are considered only simultaneously. Moreover, LeSage and Pace (2009) showed that the STAR model implies a long–run steady–state equilibrium model equivalent to the SAR model.

The STAR(1,1) model is a very parsimonious representation of space–time series, but imposes strong constraints in the behavior of the spatial units over time. An unconstrained model would consider different spatial and time coefficients for each spatial unit (call it Unconstrained STAR–USTAR model); as the spatial dimension increases, the estimation becomes unfeasible, falling under the so–called curse of dimensionality, which causes inaccuracy and uncertainty in the estimation of the model (see Giacomini and Granger, 2004). A good compromise would be a more flexible STAR(1,1) model, where the coefficients change only for groups of spatial units and not for each spatial unit, as in the USTAR model. The groups could be detected on the basis of information about the similarity of the locations (for example, geographical aggregations), but this might be subjective.

In the light of the above, in this paper we propose a procedure to detect these groups, based on a clustering agglomerative algorithm, which has some similarity to the method developed by Otranto (2010) to detect financial assets with similar conditional dynamic correlation structure. Following the classification of the methods for time series clustering proposed by Liao (2005), our approach belongs to the class of model-based clustering, where the time series are considered similar when the models characterizing them are similar. More precisely, there is a strong connection with methods devoted to classify panel data, given the two involved dimensions (space and time). A nice review of model-based methods to classify panel data is present in Fruhwirth-Schnatter (2011). As noted by this author, the purpose in the classification of time series is to assign each series to a latent class. All time series belonging to the same class are

characterized by the same data generating process. Following this idea, the density of a multivariate time series is a mixture of densities. In our approach we follow a similar idea, where the time series are referred to different spatial units and the purpose is to detect what spatial units follow the same data generating process. In other words the model we proposed is a sort of finite mixture STAR model. Again, Fruhwirth-Schnatter (2011) underlines as the model-based clustering approach is preferable in time series and panel data (and, as a consequence, in space-time series) with respect to the distance–based approach, difficult to be extended to time series. Our approach is model-based but uses also some characteristics of the distance–based approaches: the classification of the time series is based on p-values of statistic tests to verify the hypothesis of equal coefficients in two STAR models. The p–value is used as a distance measure to provide the classification; in fact it is a measure of similarity, satisfying properties of a semi-metric (see Maharaji, 1999). This idea was used in other model–based clustering approaches for time series, such as Maharaji (1999) and Otranto (2008, 2010).

The performance of this procedure is evaluated in terms of simulation experiments, using different time spans and number of locations, considering uncorrelated and correlated disturbances and different spatial weights. We investigate the ability of the proposed procedure in identifying the correct clusters and the forecasting performance of our model (called the Flexible STAR–FSTAR), compared with respect to STAR and USTAR models. The same is made with real data, where also a Sparse Vector AutoRegressive (SVAR) model is added to the comparison.

The paper is organized as follows: section 2 recalls the STAR and USTAR models and discusses the new FSTAR model. Section 3 describes the algorithm for the identification of the locations with similar STAR structure, leaving the technical details for its implementation to the final Appendix. Section 4 discusses the Monte Carlo experiments to assess the performance of the method proposed, whereas section 5 contains a comparison among the competitive models by using demographic data. Some final remarks complete the article.

2 The Flexible STAR Model

Let us consider a set of space–time observations, relative to N locations at T different times, collected in a matrix $\mathbf{Y} = \{\mathbf{y_{i,t}}\}$ ($i = 1 \ldots, N$ and $t = 1, \ldots, T$).

The classical STAR(1,1) model follows a particular autoregressive dynamics with one time lag and one spatial lag:

$$y_{i,t} = \phi y_{i,t-1} + \psi \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \varepsilon_{i,t} \tag{2.1}$$

where ϕ represents the coefficient of the time–lagged effect, whereas ψ is the coefficient of the spatial lagged effect. The number w_{ij} is the $ij-$th element of a $N \times N$ weight matrix $\mathbf{W}$ representing the spatial link between location i and location j ($w_{ij} = 0$ when $i = j$, nonnegative otherwise); the matrix $\mathbf{W}$ is normalized to have each row summing up to one. This matrix is generally fixed a priori and reflects the geographical characteristics of the spatial locations in terms of neighboring, distance, etc.; anyway it can represent other characteristics, such as economic distance (see. for example, Pinkse and Slade, 1998, Otranto, Mucciardi and Bertuccelli, 2016). It is evident that, fixing $\mathbf{W}$ in advance, the estimation of the model will be subject to possible misspecification of the weight matrix (see Stetzer, 1982); for this reason, in our successive applications, we experimented with several weight matrices. A necessary, but not sufficient, stationarity condition is given by $|\phi + \psi| < 1$ (see, for example, Arbia, Bee and Espa, 2011); we restrict the discussion to this constraint. Moreover we assume that process (2.1) is isotropic, so that the joint distribution of the random variables $y_{1,t}, \ldots, y_{N,t}$ depends on the location distance only and not on orientation (see Anselin, 1988).

The variable $\varepsilon_{i,t}$ is a zero mean white noise. It is possible to relax this hypothesis in several ways; a common specification considers the presence of correlation in the spatial dimension with the following parameterization (see Anselin, 1988):

$$\varepsilon_t = \rho \mathbf{W} \eta_{\mathbf{t}} \tag{2.2}$$

where ε_t is a vector containing the disturbances relative to the N locations at time t, η_t is a vector of uncorrelated white noise disturbances and ρ is a scalar ranging in $[-1; 1]$.

Collecting the spatial observations at time t in a $(N \times 1)$ vector $\mathbf{y}_t$, model (2.1) can be expressed in a VAR form (see Lütkepohl, 1993):

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{\mathbf{t-1}} + \varepsilon_{\mathbf{t}} \tag{2.3}$$

where $\mathbf{A} = \phi\mathbf{I}_\mathbf{N} + \psi\mathbf{W}$.

Notice that this representation is feasible for the isotropic assumption, which implies that the out–of–diagonal elements of $\mathbf{A}$ are $\psi_{ij} = \psi w_{ij}$, depending only on the inter–location distance and not on orientation (Arbia, Bee and Espa, 2011). Model (2.1) is very parsimonious, adopting the same pair of coefficients (ϕ and ψ) for all the locations and time periods, but it has good forecasting properties, as shown in Giacomini and Granger (2004) and Arbia, Bee and Espa (2011).

The basic assumption of the STAR model is that each observation y_{it} depends on its lagged value $y_{i,t-1}$ and on the lagged values of the neighboring locations; the corresponding coefficients are constant across the locations and over time. This assumption could be too restrictive, particularly in terms of fitting of the STAR(1,1) model (for a discussion about the presence of heterogeneity in the coefficients of spatial models, see Aquaro, Bailey and Pesaran, 2015, and LeSage and Chih, 2016). A more flexible model could be obtained allowing for changing in the parameter values corresponding to different locations. In a vectorial form, this model (call it Unconstrained STAR–USTAR) would assume the form:

$$\mathbf{y_t} = \phi \odot \mathbf{y_{t-1}} + \psi \odot (\mathbf{W}\mathbf{y_{t-1}}) + \varepsilon_{\mathbf{t}} \tag{2.4}$$

where $\odot$ is the Hadamard (element–wise) product, $\phi' = (\phi_1, \ldots, \phi_N)'$ and $\psi' = (\psi_1, \ldots, \psi_N)'$, with $|\phi_i + \psi_i| < 1$ (for $i = 1, \ldots, N$).

Model (2.4) is clearly more flexible but, at the same time, requires a large number of coefficients, $2(n-1)$ more than model (2.1), involving a curse of dimensionality problem. This problem could be avoided if we were able to identify a small number of groups of locations following the same STAR dynamics. In practical terms, the idea of a parsimonious but flexible STAR(1,1) model is to detect G groups of locations so that the vector of coefficients ϕ and ψ

assume the form:

$$\phi = \begin{bmatrix} \phi_1 \iota_{n_1} \\ \vdots \\ \phi_G \iota_{n_G} \end{bmatrix}; \qquad \psi = \begin{bmatrix} \psi_1 \iota_{n_1} \\ \vdots \\ \psi_k \iota_{n_G} \end{bmatrix} \tag{2.5}$$

where n_k ($k = 1, \ldots, G$) is the size of the k–th group ($\sum_{k=1}^{G} n_k = N$) and ι_j a vector of ones of dimension j. We call this model the FSTAR(1,1) model, where F stands for Flexible.

A crucial problem is the correct assignment of each location to each group. This problem can be seen as a problem of time series classification, referring to different spatial units; considering the FSTAR model as a mixture of STAR models, we want to assign each series to a class characterized by the same coefficients of the STAR model (same data generating process). For this purpose we propose a test–based agglomerative algorithm to provide a clustering of locations.

3 Identifying Locations with Similar STAR Structure

As underlined, the purpose is to group in the same class the spatial units following the same STAR(1,1) process (see Fruhwirth-Schnatter, 2011); in practice this purpose can be reached by identifying groups of units with similar coefficients ϕ and ψ. We develop a hierarchical clustering algorithm based on iterative statistical tests, using in particular the p–value as a distance measure; this approach can be performed because the p-value is a measure of similarity, satisfying the properties of a semi–metric (Maharaji, 1999). For example, it has been successfully employed in time series clustering in Maharaji (1999) and Otranto (2008, 2010).

The idea of the algorithm is as follows: we compare the coefficients of all the univariate STAR models (one model for each spatial unit); then we group together the two spatial units with more similar (not significantly different) coefficients. Then we estimate a new STAR model with this pair of spatial units and verify if other series have a similar (not significantly different) STAR structure. This procedure is iteratively applied by adding spatial units with similar coefficients to the same group until no series present similar characteristics; when this happens we consider this group completed and restart with the same procedure on the remaining

series.

To verify the similarity of the coefficients, we adopt a Wald statistic. The validity of this approach in time series clustering has been corroborated by Otranto (2008 and 2010) with simulation experiments and it is based on the theory developed by Steece and Wood (1985).

The clustering algorithm requires the testing of several hypotheses and multiple testing problems could arise[1]. For this reason we correct the p–value by adopting the Holm (1979) method. This correction improves the performance of the procedure by reducing the risk of overfitting, related to the type I error of hypothesis testing. Due to the fact that we adopted the Holm correction, the correct detections of the number of groups G (shown in the next section) increased by 5% on average.

Refer to the final appendix for technical details about the clustering procedure adopted in this work.

4 Simulation Study

In order to verify the capability of the previous algorithm in detecting the correct groups of locations, we have performed several simulation experiments, with the purpose of evaluating its validity under several scenarios:

- different sizes of time T hypothesizing to know the weight matrix $\boldsymbol{W}$: this experiment has the purpose of evaluating the dependence of the algorithm on the length of the time series; of course we expect that, in a model–based clustering, the performance of the algorithm is linked to the efficiency of the estimation procedure, which is strictly linked to the number of observations;

- fixed T but wrong $\boldsymbol{W}$: the aim of this experiment is to evaluate the robustness of the algorithm in the presence of uncertainty about the spatial structure;

- forecasting performance: the purpose is to compare the forecasts obtained with an FS-TAR model (based on the clustering algorithm proposed) against the STAR and USTAR

[1] We would like to thank an anonymous referee and the Associate Editor who called this problem to our attention.

models described in Section 2.

We simulated data by hypothesizing that the spatial patterns are represented by two alternative regular lattices of size 3×3 ($N = 9$) and 5×5 ($N = 25$); moreover we have considered six different number of groups ($G = 1, 2, 3, 4, 5, 6$) and the weight matrix is based on the classical rook contiguity criterion for regular lattices. The data are generated from multivariate standard Normal distributions; we consider both the cases of uncorrelated disturbances ε_t and spatially correlated disturbances, transforming the generated ε_t by (2.2) with $\rho = 0.7$.

The coefficients used to generate data from the FSTAR model (2.5) are chosen to combine different time and spatial effects in different groups; in detail, they are:

– when $G = 1$ (STAR model), $\phi_1 = 0.5$, $\psi_1 = 0.3$;

– when $G = 2$, $\phi_1 = 0.5$, $\psi_1 = 0.3$, $\phi_2 = 0.3$, $\psi_2 = 0.6$;

– when $G = 3$, $\phi_1 = 0.5$, $\psi_1 = 0.3$, $\phi_2 = 0.3$, $\psi_2 = 0.6$, $\phi_3 = 0.8$, $\psi_3 = 0.1$.

– when $G = 4$, $\phi_1 = 0.5$, $\psi_1 = 0.3$, $\phi_2 = 0.3$, $\psi_2 = 0.6$, $\phi_3 = 0.8$, $\psi_3 = 0.1$, $\phi_4 = 0.1$, $\psi_4 = 0.7$.

– when $G = 5$, $\phi_1 = 0.5$, $\psi_1 = 0.3$, $\phi_2 = 0.3$, $\psi_2 = 0.6$, $\phi_3 = 0.8$, $\psi_3 = 0.1$, $\phi_4 = 0.1$, $\psi_4 = 0.7$, $\phi_5 = 0.4$, $\psi_5 = 0.4$.

– when $G = 6$, $\phi_1 = 0.5$, $\psi_1 = 0.3$, $\phi_2 = 0.3$, $\psi_2 = 0.6$, $\phi_3 = 0.8$, $\psi_3 = 0.1$, $\phi_4 = 0.1$, $\psi_4 = 0.7$, $\phi_5 = 0.4$, $\psi_5 = 0.4$, $\phi_6 = 0.7$, $\psi_6 = 0$.

The spatial units are equally divided among the groups, with at least a difference of one unit in favor of the first groups. For example, when $N = 25$ and $G = 5$, each group contains 5 spatial units; when $N = 25$ and $G = 6$, the first group contains 5 spatial units and the others 4 spatial units; when $N = 9$ and $G = 5$, the first four groups contains 2 spatial units and the last group only one.

The number of replications is 1000 for each case.

Finally, we also replicated the simulation experiments here on series with the characteristics of the real data set used in Section 5. This is made to consider a more realistic spatial pattern, with irregular boundaries, depending on distance–based weights. This experiment provides some indication about the degree of reliability of the results illustrated in Section 5.

4.1 Effect of time length

We have generated space–time data with three different time spans ($T = 100, 500, 1000$). The performance evaluation is conducted: 1) recording for each replication the number of groups detected; 2) evaluating the similarity of the composition of the detected group with the true one. The second point is made because the number of groups could be correctly detected but the composition of the groups is not equal to the true one; moreover we are interested in evaluating the magnitude of the differences. For this purpose we adopt the adjusted Rand index (Rand, 1971; Hubert and Arabie, 1985):

$$r = \frac{\sum_{i=1}^{G}\sum_{j=1}^{G^*}\binom{\hat{n}_{ij}}{2} - [\sum_{i=1}^{G}\binom{n_i}{2}][\sum_{j=1}^{G^*}\binom{\hat{n}_j}{2}]/\binom{N}{2}}{[\sum_{i=1}^{G}\binom{n_i}{2} + \sum_{j=1}^{G^*}\binom{\hat{n}_j}{2}]/2 - [\sum_{i=1}^{G}\binom{n_i}{2}][\sum_{j=1}^{G^*}\binom{\hat{n}_j}{2}]/\binom{N}{2}} \tag{4.1}$$

where G and G^* represent the number of groups in the true and detected clustering respectively; n_i and $\hat{n}_j$ are the number of locations belonging to the group i of the true and group j of the detected clustering respectively, whereas $\hat{n}_{ij}$ is the number of locations belonging to the group i in the true pattern and assigned to the group j in the detected clustering. We can use r to evaluate the performance of the proposed method because $r \in [0, 1]$, assuming value 0 when the differences between the groups are at their maximum (worst performance) and 1 in the case of coincidence between the true and the detected clustering.

In Tables 1–2 we show the performance evaluation relative to the first point; in practice we show the distribution of the number of groups detected (G^*) for the 3×3 and 5×5 lattices respectively. In the 3×3 lattice (Table 1) we notice that the number of groups is correctly detected for each time span with a large percentage when the number of groups G is 1 or 2; when G increases the performance rapidly decreases for $T = 100$, whereas it increases for $T = 500$ and $T = 1000$ until $G = 4$. The presence of spatial correlation favors the correct detection of the number of groups, with a general increase of the number of correct cases, in particular for high G. When $G = 5$ or 6 only the simulations with a large T seem to provide a satisfactory result with more than 770 cases where $G^* = G$.

Considering the 5×5 lattice (Table 2), the algorithm performs better for $T = 100$ with respect to the 3×3 lattice (apart the case $G = 1$). This is explained for the small number of

Table 1: Simulation results for a 3×3 lattice, with uncorrelated ($\rho = 0$) and correlated ($\rho = 0.7$) disturbances, and number of groups G: number of groups detected in correspondence to different time length T.

	$\rho = 0$			$\rho = 0.7$		
	$T = 100$	$T = 500$	$T = 1000$	$T = 100$	$T = 500$	$T = 1000$
G^*			$G = 1$			
1	832	849	872	894	906	872
2	168	151	127	106	94	127
3	0	0	1	0	0	1
G^*			$G = 2$			
1	225	0	0	150	0	0
2	737	878	871	817	905	908
3	38	120	126	33	93	89
4	0	2	3	0	2	3
G^*			$G = 3$			
1	1	0	0	0	0	0
2	469	0	0	315	0	0
3	517	894	905	668	935	932
4	13	103	95	16	59	59
5	0	3	0	1	6	9
G^*			$G = 4$			
2	195	0	0	143	0	0
3	727	21	0	738	3	0
4	78	933	956	118	953	973
5	0	45	43	1	41	26
6	0	1	1	0	3	1
G^*			$G = 5$			
2	222	0	0	179	0	0
3	694	12	0	705	2	0
4	83	692	247	113	602	223
5	1	296	752	3	390	773
6	0	0	1	0	6	4
G^*			$G = 6$			
2	285	0	0	219	0	0
3	599	9	0	647	2	0
4	114	181	2	128	42	0
5	2	697	327	6	662	206
6	0	113	667	0	332	781
7	0	0	4	0	2	9
8	0	0	0	0	0	4

Note: Data generated using the rook contiguity matrix, that is supposed known. G^* indicates the number of groups detected by the algorithm described in Section 3. The number of replications is 1000 for each case study.

Table 2: Simulation results for a 5×5 lattice, with uncorrelated ($\rho = 0$) and correlated ($\rho = 0.7$) disturbances, and number of groups G: number of groups detected in correspondence to different time length T.

	$\rho = 0$			$\rho = 0.7$		
	$T = 100$	$T = 500$	$T = 1000$	$T = 100$	$T = 500$	$T = 1000$
G^*			$G = 1$			
1	712	682	705	734	778	754
2	281	316	290	263	222	243
3	7	2	5	3	0	3
G^*			$G = 2$			
1	106	0	0	46	0	0
2	794	711	692	852	793	761
3	100	277	286	102	198	229
4	0	12	22	0	9	10
G^*			$G = 3$			
2	251	0	0	158	0	0
3	688	724	700	780	814	784
4	61	252	272	61	170	197
5	0	22	28	1	16	19
6	0	2	0	0	0	0
G^*			$G = 4$			
2	70	0	0	38	0	0
3	697	1	0	643	0	0
4	224	763	712	313	835	805
5	9	221	252	6	149	177
6	0	14	35	0	15	17
7	0	1	1	0	1	1
G^*			$G = 5$			
2	96	0	0	38	0	0
3	714	2	0	697	0	0
4	188	478	40	247	293	15
5	2	474	785	18	640	844
6	0	46	166	0	64	126
7	0	0	9	0	3	14
8	0	0	0	0	0	1
G^*			$G = 6$			
2	83	0	0	25	0	0
3	611	0	0	525	0	0
4	290	11	0	392	2	0
5	16	592	53	55	337	21
6	0	382	805	3	611	848
7	0	15	135	0	50	124
8	0	0	6	0	0	7
9	0	0	1	0	0	0

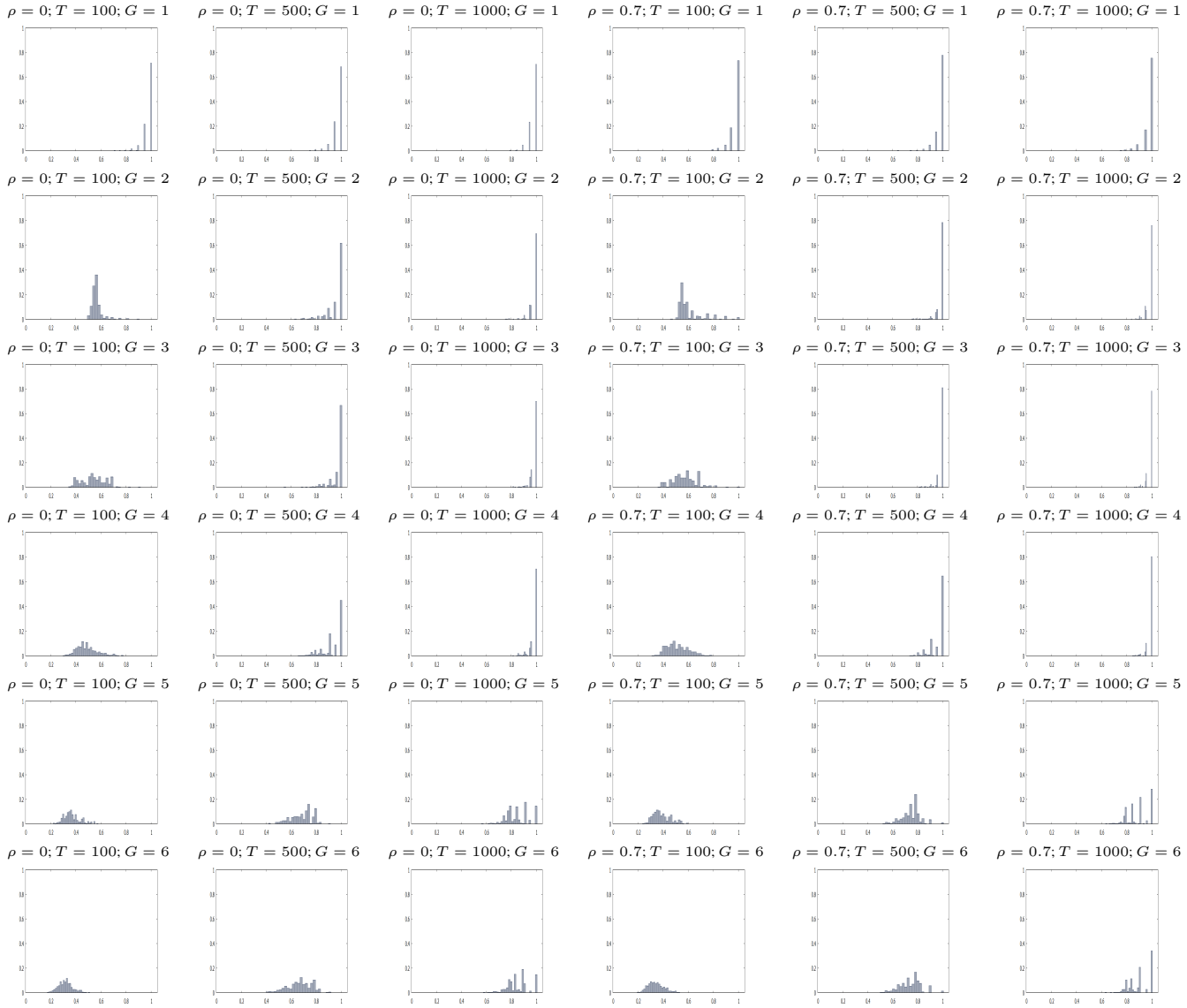
Note: Data generated using the rook contiguity matrix, that is supposed known. G^* indicates the number of groups detected by the algorithm described in Section 3. The number of replications is 1000 for each case study.

Figure 1: Simulation results for a 3×3 lattice, with uncorrelated ($\rho = 0$) and correlated ($\rho = 0.7$) disturbances, number of groups G and different time length T: distribution of the Rand index.

elements present in a large number of groups when the lattice is relative to only 9 spatial units. For $T = 500$ and 1000 the performance increases when G is 5 or 6. Again, the presence of spatial correlation favors the correct detection of the number of groups.

The evaluation of the second point (correct assignment of the spatial units to the group) is conducted in graphical terms, plotting the empirical distribution of the Rand index for each case. These distributions are illustrated in Figures 1 (lattice 3×3) and 2 (lattice 5×5). When the data generating process is the STAR(1,1) ($G = 1$), the detection of the true model is very frequent independently of the time length: the Rand index (4.1) is equal to 1 in more than

Figure 2: Simulation results for a 5×5 lattice, with uncorrelated ($\rho = 0$) and correlated ($\rho = 0.7$) disturbances, number of groups G and different time length T: distribution of the Rand index.



83% of cases for the 3×3 lattice and this value increases with spatial autocorrelation of the disturbances. This percentage decreases sensitively in the 5×5 lattice, with a higher percentage of 2 groups detected, but the difference with respect to the true pattern is not relevant; in fact the percentage of cases with $r \geq 0.85$ is always around 96%. The procedure seems to fail when a FSTAR(1,1) model is the true data generating process and T is small; we can notice the small number of cases with the correct pattern detection, whereas the performance is good when T increases and $2 \leq G \leq 4$, with percentages of $r \geq 0.85$ near to 100%. When $G = 5, 6$, the performance of the clustering algorithm strongly depends on the time dimension, reaching

sufficiently good performances only when $T = 1000$ for the 3×3 lattice (Figure 1), in particular in presence of spatial autocorrelation of disturbances; these cases show a poor performance in the 5×5 lattice (Figure 2).

In practice the algorithm does work for large T, whereas the correct pattern is difficult to be detected when T is small and $G > 1$. Larger numbers of groups require a longer length of time; also the increase of the number of spatial units implies a longer length of time, consistently with the findings of Otranto and Gallo (1994), investigating balanced space–time patterns to avoid estimation problems in STARMA models.

4.2 Effect of the weight matrix

Previous experiments were performed supposing that the weight matrix was known; in practical cases it is not fixed a priori. In order to understand how important the knowledge of the true matrix $\boldsymbol{W}$ is, we run the same set of simulations of the previous subsection, fixing $T = 1000$ to avoid time dependence, but using a *wrong* contiguity matrix in the estimation (i.e. data generated using one type of spatial weight matrix, but estimated using another type of spatial weight matrix). In particular we used the queen contiguity matrix, which considers more neighbours with respect to the rook case.[2] In Table 3 we show the results of this experiment. Comparing the results of Table 3 with the corresponding results of Tables 1 (for the 3×3 lattice) and 2 (for the 5×5 lattice), it is clear that the identification of the number of groups depends on the adoption of the correct weight matrix; anyway this difference is less relevant in the case of the absence of spatial correlation among disturbances, whereas the decline is dramatic when $\rho = 0.7$; in this case we generally obtain an overestimation of the number of groups. The result is not surprising: the absence of spatial correlation between disturbances implies that the weight matrix enters more weakly in the data generating process, so that its correct identification is less relevant in the estimation process. Vice versa, when the spatial correlation among disturbances increases, the correct detection of the weight matrix provides a more accurate classification of the spatial units.

Figure 3 shows the distribution of the Rand index for the simulated series. It confirms the

[2]We have also used the bishop contiguity matrix and the results are similar and in many cases worse than the queen case. Results available on request.

Table 3: Simulation results: Number of groups detected (G^*) for 3×3 and 5×5 lattices, with uncorrelated ($\rho = 0$) and correlated ($\rho = 0.7$) disturbances, time length $T = 1000$ and different number of groups G, adopting a wrong weight matrix.

	3×3 lattice		5×5 lattice		3×3 lattice		5×5 lattice	
	$\rho = 0$	$\rho = 0.7$	$\rho = 0$	$\rho = 0.7$	$\rho = 0$	$\rho = 0.7$	$\rho = 0$	$\rho = 0.7$
G^*	$G = 1$				$G = 2$			
1	847	10	680	247	0	0	0	0
2	153	833	311	676	846	13	647	9
3	0	157	9	76	149	559	329	356
4	0	0	0	1	5	410	24	547
5	0	0	0	0	0	18	0	85
6	0	0	0	0	0	0	0	3
G^*	$G = 3$				$G = 4$			
3	873	18	622	244	0	0	0	0
4	121	872	334	519	884	295	245	22
5	6	104	43	213	113	631	574	396
6	0	6	1	24	3	74	168	454
7	0	0	0	0	0	0	13	123
8	0	0	0	0	0	0	0	5
G^*	$G = 5$				$G = 6$			
4	274	181	219	4	78	1	0	0
5	708	613	633	452	446	20	197	0
6	18	193	142	457	456	166	594	198
7	0	13	6	84	0	777	193	611
8	0	0	0	3	0	36	15	177
9	0	0	0	0	0	0	1	13
10	0	0	0	0	0	0	0	1

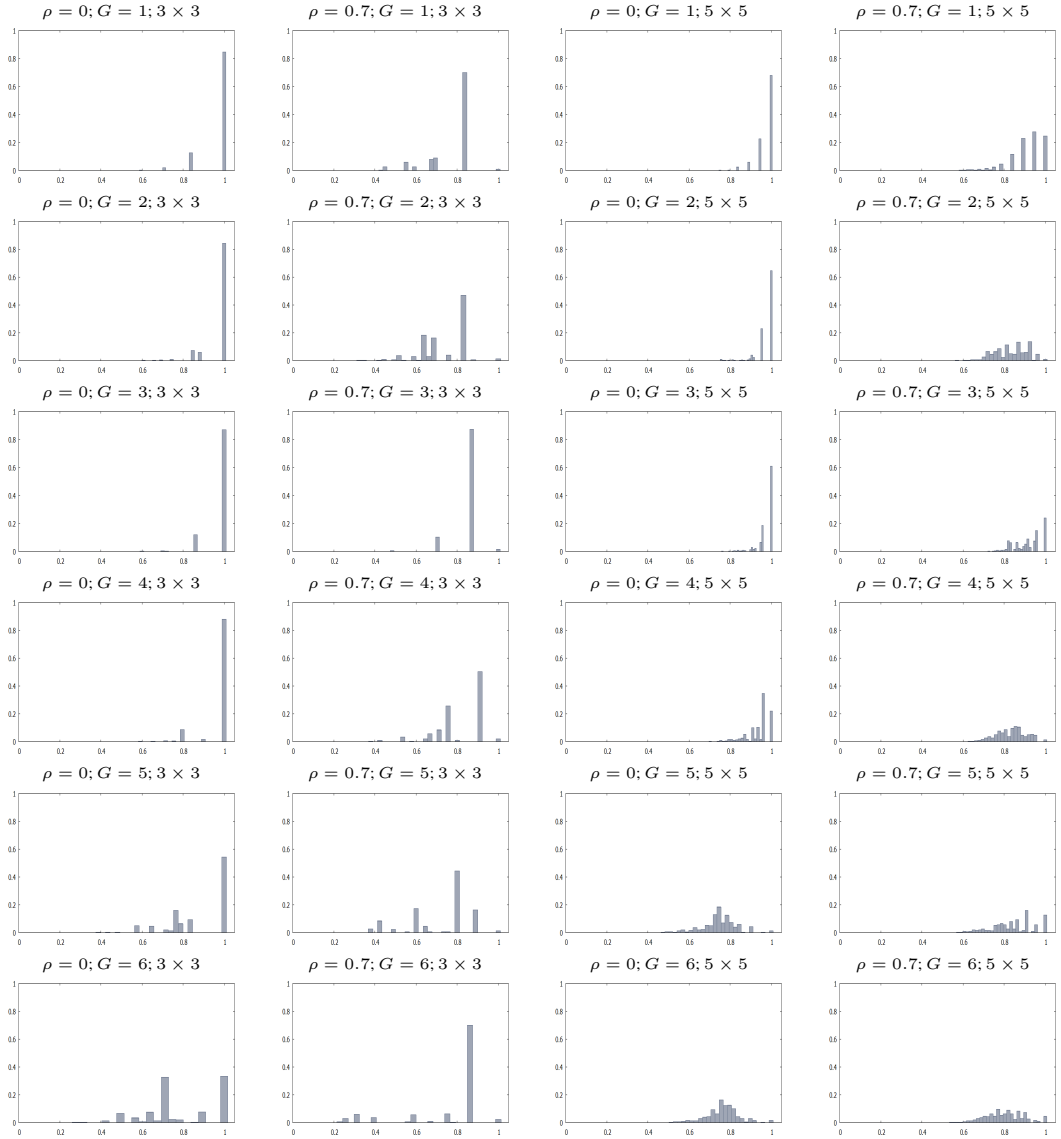
Note: Data generated using the rook contiguity matrix, but estimated using the queen contiguity matrix. G^* indicates the number of groups detected by the algorithm described in Section 3. The number of replications is 1000 for each case study.

previous comments, underlying as we have good chances to detect the correct groups when $\rho = 0$ and the lattice is 3×3, whereas it is more difficult to have success in the 5×5 lattice. In general the groups are similar to the true ones for $1 \leq G \leq 4$, given a Rand index generally higher than 0.8.

4.3 Forecasting performance

In the previous experiments we have evaluated the capability of the algorithm to identify the groups of spatial units with similar STAR structure. Another important task is to compare the FSTAR model with respect to alternative models, in particular in terms of out–of–sample forecasting performance. This could be a good point in favor of the FSTAR model if its estimation,

Figure 3: Simulation results for 3×3 and 5×5 lattices, with uncorrelated ($\rho = 0$) and correlated ($\rho = 0.7$) disturbances, time length $T = 1000$ and different number of groups G, adopting a wrong weight matrix: distribution of the Rand index.



based on an acceptable detection of groups of spatial units, provides a better forecasting performance with respect to a classical STAR model and a similar performance with respect to an overparameterized USTAR model.

We follow the lines of the experiment performed in Frühwirth-Schnatter and Kaufmann (2008), who evaluate a model–based clustering procedure for multiple time series. We start from the previous simulated space–time series, adding 4 observations for each spatial unit; then, using the first T observations for estimation, we forecast the successive 4 observations for each spatial unit; call them $\hat{y}_{i,T+h}$. Finally we calculate the Root Mean Squared Error

Table 4: Simulation results: Average RMSE of out—of sample forecasts for alternative models.

	G=1		G=2		G=3		G=4		G=5		G=6	
	$\rho=0$	$\rho=0.7$	$\rho=0$	$\rho=0.7$	$\rho=0$	$\rho=0.7$	$\rho=0$	$\rho=0.7$	$\rho=0$	$\rho=0.7$	$\rho=0$	$\rho=0.7$
						$T=100$						
STAR	0.998	0.398	1.008	0.404	1.037	0.416	1.060	0.424	1.048	0.419	1.046	0.421
USTAR	1.008	0.402	1.007	0.402	1.008	0.403	1.007	0.403	1.007	0.403	1.008	0.403
FSTAR	0.998	0.399	1.007	0.403	1.014	0.405	1.015	0.405	1.013	0.405	1.014	0.406
						$T=500$						
STAR	0.997	0.401	1.006	0.406	1.038	0.419	1.057	0.427	1.046	0.421	1.046	0.423
USTAR	0.999	0.402	0.999	0.401	0.999	0.401	0.999	0.402	0.999	0.402	0.999	0.401
FSTAR	0.997	0.401	0.997	0.401	0.997	0.401	0.998	0.401	0.999	0.401	0.999	0.401
						$T=1000$						
STAR	0.998	0.402	1.008	0.407	1.038	0.419	1.058	0.427	1.047	0.422	1.046	0.424
USTAR	0.999	0.402	0.999	0.402	0.999	0.402	0.999	0.402	0.999	0.402	0.999	0.402
FSTAR	0.998	0.402	0.998	0.402	0.998	0.402	0.998	0.402	0.999	0.402	0.999	0.402
						$T=1000$ and wrong $\boldsymbol{W}$						
STAR	1.005	0.404	1.023	0.412	1.046	0.423	1.075	0.434	1.061	0.431	1.055	0.430
USTAR	1.006	0.404	1.016	0.406	1.012	0.407	1.024	0.411	1.020	0.410	1.015	0.409
FSTAR	1.006	0.404	1.015	0.406	1.011	0.407	1.023	0.411	1.020	0.410	1.015	0.409

Note: Data generated using the rook contiguity matrix; wrong $\boldsymbol{W}$ means that the models are generated using the rook contiguity matrix, but estimated using the queen contiguity matrix. The number of replications is 1000 for each case study.

(RMSE):

$$RMSE = \sqrt{\frac{1}{4N}\sum_{i=1}^{N}\sum_{h=1}^{4}\left(y_{i,T+h} - \hat{y}_{i,T+h}\right)^2} \tag{4.2}$$

We evaluate (4.2) for the alternative models STAR, USTAR and FSTAR for all the combinations of T, N and ρ described in this section, adopting the true weight matrix (rook contiguity matrix) and also in the case of wrong weight matrix (queen contiguity matrix).[3] The average of the RMSE obtained in the 1000 replications of each case are showed in Table 4.

When $T = 100$ the overparameterized USTAR model performs better than FSTAR and STAR if the number of groups G is higher than 2 and $\rho = 0$; in the other cases the FSTAR has a similar behavior, confirming that the presence of spatial correlation between disturbances and the use of the true weight matrix help in the identification of the clusters. When $G = 1$ STAR and FSTAR outperform USTAR. Increasing the number of observations the performance of USTAR and FSTAR is very similar, with a clear superiority with respect to the STAR model (excluding the case $G = 1$, where STAR is the data generating process). The same holds when the weight matrix is wrong; of course there is a higher RMSE with respect to the case of true weight matrix, but this wrong choice affects all three models in a similar way.

This experiment seems to suggest that the FSTAR model has a similar forecasting performance with respect to an overparameterized model as USTAR and it is clearly better than the

[3]In the wrong weight matrix case, we fix $T = 1000$ as in the previous subsection.

classical STAR, when the data generating process includes groups of spatial units with similar STAR structure. Of course the performance of FSTAR depends on the capability of the clustering algorithm to detect the correct groups, and this result is linked to the length of the time series. The choice of the correct weight matrix affects the goodness of forecasting of all the models, not altering the ranking of the forecasting performance.

4.4 Distance–based weight matrices

The previous simulations were based, as usual, on regular lattices, adopting the typical binary weight matrices derived from contiguity criteria. We complete the simulation study reproposing the previous exercises for a more realistic pattern, derived from the space–time structure of the data set used in the successive section. The data set contains $T = 576$ observations along the time and $N = 12$ spatial units; we generate 1000 space–time series of this length with the FSTAR coefficients shown in Table 7 for $G = 2$ and a weight matrix depending on the geographical distance between each pair of spatial units. It is worthwhile to mention that the coefficients of the two distinct groups are close (in particular the ψ coefficients). The data are generated by a multivariate Normal distribution with constant variance equal to 0.093 and $\rho = 0.157$.[4]

The weight matrix used to generate data is a particular *kernel matrix*, where the weight w_{ij} is a continuous and monotonic decreasing function of the (Euclidean) distance d_{ij} (Fotheringham, Brunsdon and Charlton, 2002). The choice of kernel functions is particularly appropriate because the bandwidth h provides a control for the circular area of influence of each observation i. We adopt Gaussian distance-decay-based functions as:

$$w_{ij} = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right] \tag{4.3}$$

with $h = max(d_{ij})$ and call it the $Kmax$ weight matrix. As in the previous experiments, we will consider both the cases of knowing the weight matrix and adopting wrong weight matrices; in particular we will consider another kernel matrix with a lower area of influence

[4]The value of the variance is the estimated variance of the FSTAR model shown in Table 7; the coefficient ρ is the average of the 12 Moran indices (see Anselin, 1988) of each spatial unit.

Figure 4: Simulation results for series generated from the FSTAR model of Table 7: distribution of the Rand index using three different weight matrices.

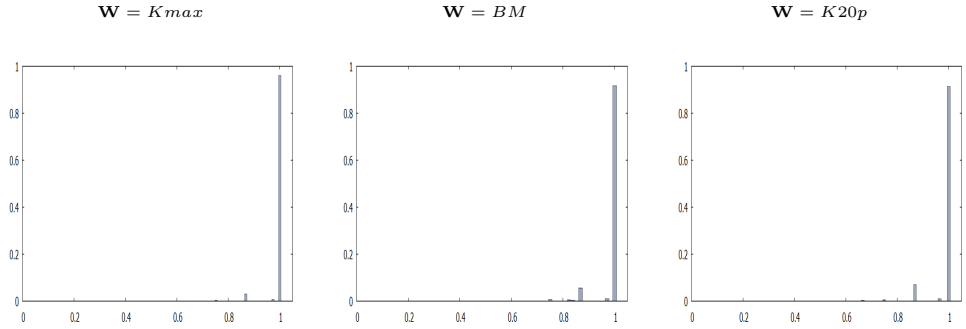


Table 5: Simulation results for series generated from the FSTAR model of Table 7: Average RMSE (multiplied by 1000) of out–of sample forecasts for alternative models in correspondence with different weight matrices .

	$Kmax$	BM	$K20p$
STAR	13.744	13.764	13.754
USTAR	13.694	13.715	13.704
FSTAR	13.678	13.701	13.689

Note: Data generated using the Kmax matrix. The number of replications is 1000 for each case study.

($h = 0.2max(d_{ij})$), called $K20p$, and a *Boundary matrix* (BM). The latter is a binary matrix, with weight equal to 1 if the corresponding units have a common boundary, 0 otherwise.

In Figure 4 we show the distribution of the Rand index for the 1000 replications and the three alternative weight matrices. The results seem very positive with 96% correct detections of groups in the case of true weight matrix and more than 90% in the wrong cases.

We have also compared the forecasting performance of the STAR, USTAR and FSTAR models, as in the previous subsection, under the new experimental design (Table 5). The differences are small, but FSTAR always shows the lowest MSE.

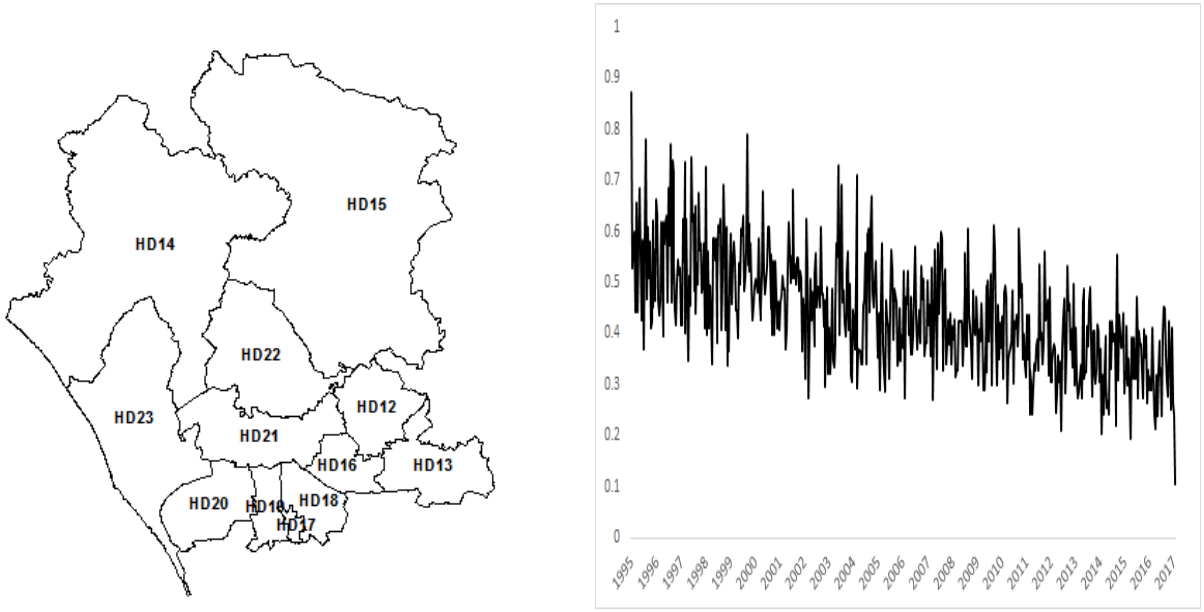
5 Application

Let us consider the data set of the Crude Birth Rate (henceforth CBR) relative to the 12 health districts of the city of Caserta (Italy) from January 1^{st} 1995 to January 28^{th} 2017; the data are recorded every two weeks, for a total of $T = 576$ time observations.[5] On the left side of Figure 5 the map of the geographical area considered is shown, distinct in the 12 health

[5]We are very grateful to Francesco Giorgianni and Gianluca Trifirò who have produced and made available this data set.

districts, labeled HD12, . . . , HD23. On the right side of Figure 5, we show the time dynamics of one of the 12 spatial units;[6] the presence of a decreasing linear trend is clear. In order to obtain a stationary space–time series we have subtracted a linear trend (of the form $a_i + b_i t$, $t = 1, \ldots, T$, $i = 1, \ldots, N$) from each time series.

Figure 5: Map of the 12 health districts of Caserta (left side) and time series of the district HD21 (right side).



As previously explained, the choice of the spatial weight matrix could play a crucial role in the inference of the model. It seems to lose importance in the space–time structure considered in this application, as shown in the simulation results of subsection 4.4. So in order to obtain the robustness of the results with respect to the $\mathbf{W}$ spatial matrix chosen, we chose eight different weight specifications (for a review see Cliff and Ord, 1973, 1981; Getis and Ord, 1992; Cressie, 1993), distinguishing between binary and kernel spatial weight matrices:

- **Binary matrix**: the weight w_{ij} is equal to 1 if i and j are *neighbors*, 0 otherwise. The interconnection system is defined in terms of boundaries or distance under a certain threshold.

  1. *Boundary Matrix* (BM): it is the classical contiguity binary matrix based on the

[6] The other time series have a very similar behaviour; they are available on request.

common boundary; in practice the weight is 1 when the two locations share a common boundary; by convention $w_{ii} = 0$ for $i = 1, \ldots, N$.

2. *Maxmin matrix* (Mm): the neighbors are detected by the Maxmin distance (Mucciardi and Bertuccelli, 2012). It is defined as

$$d_{Mm} = max(e_1, \ldots, e_N)$$

where e_i ($i = 1, \ldots, N$) represents the minimum Euclidean distance of the generic location i and the other locations j ($i \neq j$). As a consequence all locations have at least one connection and the neighbors of location i are the locations with Euclidean distance lower than d_{Mm}.

- **Kernel matrix**: the weight w_{ij} is obtained adopting the Gaussian distance-decay-based functions (4.3). For our experiments we have selected the following kernel weight matrices:

3. $Kmin$: with $h = min(d_{ij})$ (in our application it corresponds to 3.77 Kilometers) ;
4. $K10p$: with $h = 0.1max(d_{ij})$ (4.80 KM);
5. $K20p$: with $h = 0.2max(d_{ij})$ (9.61 KM);
6. KMm: with $h = d_{Mm}$ (21.20 KM);
7. $Kmea$: with $h = mean(d_{ij})$ (22.32 KM);
8. $Kmax$: with $h = max(d_{ij})$ (48.05 Km)

In Figure 6 the behavior of the Maxmin distance and the six kernel functions are illustrated. The Maxmin binary criterion is a step function which assigns weight 1 to the locations with Euclidean distance lower than the Maxmin distance; the Kernel weights decrease with the distance (in Kilometers) and the area of influence increases with h. Each weight matrix is successively standardized by row. The curves relative to $Kmin$ and $K10p$ as well those relative to KMm and $Kmea$ are very closed, given the similar bandwidth adopted.

We applied the clustering algorithm using alternatively each weight matrix, obtaining different clustering. Using the Rand index (4.1) to measure the similarity of each pair of classifi-

Figure 6: Weights of the spatial matrix derived from the Maxmin criterion and five Kernel functions.

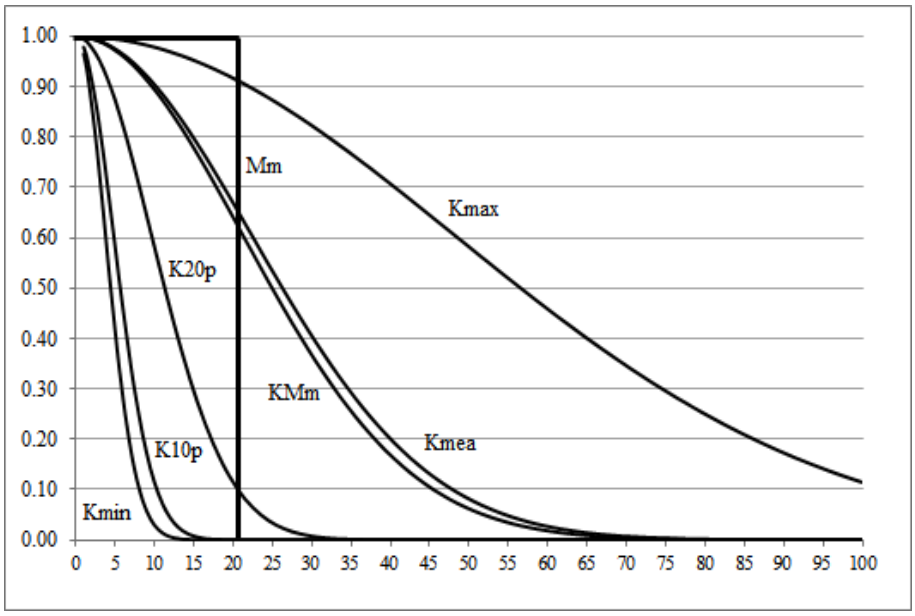


Table 6: CBR data set: Rand index for each pair of classification derived from the hierarchical algorithm with different spatial weight matrices.

	Mm	$Kmin$	$K10p$	$K20p$	KMm	$Kmea$	$Kmax$
BM	0.480	0.759	0.645	0.677	0.444	0.444	0.444
Mm		0.430	0.396	0.455	0.275	0.275	0.275
$Kmin$			0.430	0.445	0.380	0.380	0.380
$K10p$				0.899	0.551	0.551	0.551
$K20p$					0.435	0.435	0.435
KMm						1.000	1.000
$Kmea$							1.000

cations, we obtain a synthetic representation of the different results derived from the different $\mathbf{W}$ matrices, as shown in Table 6. It seems clear that the kernel matrices with a large bandwidth (KMm, $Kmea$ and $Kmax$) provide the same classification (Rand index equal to 1), whereas the classification seems sensitive to changes in the bandwidth when it is small, showing different clusters when comparing $Kmin$ and $K10p$. The smallest Rand index is equal to 0.275, obtained comparing Mm and the kernel matrices with high bandwidth h. As expected, the BM matrix provides results more similar to kernel matrices with the smallest h.

We have estimated, separately for each weight matrix, the STAR(1,1), the USTAR(1,1) and the FSTAR(1,1) models. Moreover, given the relationship between STAR and VAR models, as shown in equation (2.3), we estimate also a Sparse VAR (SVAR) model, shrinking some coefficients toward 0. We adopt a Lasso-penalized approach, using the coordinate descent

algorithm for its estimation (see Friedman, Hastie and Tibshirani, 2008). The amount of penalty is driven by a scalar tuning parameter; for this purpose we use the Elastic Net approach of Zou, Hastie and Tibshirani (2004).[7] The estimated SVAR model involves 84 nonzero coefficients in the $\boldsymbol{A}$ matrix in (2.3); to save space we do not show these coefficients, that are available on request.

As an example, in Table 7 we show the estimates of USTAR and FSTAR models using the $Kmax$ weight matrix, which, in terms of Mean Squared Error, show the best results (the estimates of all the other models are available on request). The horizontal line separates the two groups identified by the hierarchical algorithm. It is evident as the coefficients of the USTAR model are not significant; this is the typical problem of this model which makes it not efficient in real cases. On the other hand, the coefficients of the two groups of the FSTAR model are significantly different from zero, with the exception of the ϕ coefficient of the second group, which is equal to zero.

In the last two columns of Table 7 we show the average of the weights w_{ij} of each row i (referred to the spatial unit in the first column) corresponding to the spatial units j belonging to the same cluster of i, and the average of the weights of the spatial units j which do not belong to the same cluster of i. This is a useful exercise to understand if spatial units sharing a large weight tend to be clustered together into one group, and if spatial units sharing a small weight are clustered into different groups. The two averages are very similar and, in five cases, the second average is higher than the first one. These results are similar for all the eight weight matrices adopted in this work. In practice the clustering depends on the weight matrix adopted, as shown in this application and in the previous simulations; in fact this choice affects the estimates of the coefficients because it enters directly into the equation (2.4), averaging the lagged observations of the spatial units; but a direct relationship between the composition of the cluster and the magnitude of the weight does not seem to exist.

Comparing the goodness of the fit of the four alternative models we find that the lowest RMSE is in SVAR, followed by USTAR, FSTAR and STAR. This ranking is the same for all the estimations with a different weight matrix. It is an expected result; in fact it is well

[7] We utilize the sparcevar package (ver. 0.0.10) in R language for the estimation of this model.

Table 7: CBR data set: Estimation (standard errors in parentheses) of the coefficients of FSTAR and USTAR models, using the $Kmax$ weight matrix, and in correspondence to each spatial unit, the average of the weights of the units belonging to the same cluster and the average of the weights of the units belonging to the other cluster.

Health District	ϕ **FSTAR**	ϕ **USTAR**	ψ **FSTAR**	ψ **USTAR**	Av. w_{ij} in cluster	Av. w_{ij} out of cluster
HD13	0.121	0.026	0.484	0.461	0.092	0.088
	(0.048)	(0.170)	(0.094)	(0.280)		
HD15		0.064		0.355	0.088	0.099
		(0.180)		(0.270)		
HD16		0.212		0.381	0.092	0.088
		(0.159)		(0.302)		
HD17		0.041		0.385	0.092	0.087
		(0.133)		(0.268)		
HD18		0.241		0.364	0.092	0.088
		(0.145)		(0.300)		
HD19		0.135		0.624	0.092	0.088
		(0.138)		(0.306)		
HD20		0.125		0.546	0.092	0.089
		(0.124)		(0.285)		
HD22		0.057		0.647	0.090	0.093
		(0.144)		(0.289)		
HD23		0.127		0.576	0.090	0.093
		(0.148)		(0.282)		
0						
HD12	0.000	0.046	0.412	0.336	0.085	0.092
	(0.107)	(0.228)	(0.158)	(0.280)		
HD14		0.004		0.374	0.093	0.090
		(0.189)		(0.273)		
HD21		-0.034		0.529	0.087	0.092
		(0.166)		(0.284)		

Note: The horizontal lines separate the groups identified by the procedure described in section 3. The parameters estimated with the STAR(1,1) model are $\phi = 0.103$ (0.044) and $\psi = 0.407$ (0.081). *Av.* w_{ij} *in cluster* is the average of the weights of row i (corresponding to the spatial unit in the first column of the table) corresponding to the spatial units belonging to the same cluster of i, whereas *Av.* w_{ij} *out of cluster* is the average of the weights of row i (corresponding to the spatial unit in the first column of the table) corresponding to the spatial units belonging to the cluster which does not contain i.

Table 8: Out–of–sample forecasting of the CBR data: percentage variation in RMSE of STAR, USTAR, and SVAR models with respect to the FSTAR model.

W	STAR	USTAR	SVAR
BM	15.42	15.38	61.95
Mm	20.32	30.64	68.68
$Kmin$	20.39	20.32	71.90
$K10p$	24.82	26.89	77.97
$K20p$	17.93	27.54	66.59
KMm	15.65	23.12	60.33
$Kmea$	15.62	22.90	60.09
$Kmax$	15.96	22.37	58.59

known that overparameterized models show a better in–sample performance than simpler models, which could be not confirmed in out–of–sample terms (see, for example, Hansen, 2010, who provides an analytical proof of this result). For this reason the out–of–sample evaluation is particularly interesting. We re–estimate the models on a reduced data set, excluding the last 4 observations for each unit, re–applying the clustering algorithm to estimate the FSTAR model; then we perform the one–step ahead forecasts for the out–of–sample period (12×4 space–time observations). In Table 8 we show the percentage variations in RMSE of STAR, USTAR and SVAR with respect to FSTAR with the eight different weight matrices.[8] It is clear that the FSTAR model performs better than the alternatives and a bad SVAR performance with a very high RMSE. The STAR performs better than USTAR in six cases, but all its RMSEs are larger than 15% with respect to the RMSE of FSTAR.

6 Final Remarks

The use of a parsimonious model, as the STAR(1,1), to represent space–time series is a common practice in statistical modeling because the alternative VAR model (we called it USTAR model) causes inefficiency in the estimation and overparameterization. We have proposed a flexible model (FSTAR) with a reduced number of parameters, which allows spatial units with similar dynamics to have the same coefficients. The identification of the locations can be done through a hierarchical clustering algorithm, based on a Wald test which verifies the similarity of the coefficients of different locations. Consequently, the advantage of the FSTAR model is twofold:

[8]The RMSE of the SVAR model is constant, not depending on the weight matrix.

the model is a good compromise between the parsimony of the STAR model and the flexibility of the overparameterized USTAR model; the model building procedure identifies groups of locations with a similar space–time behavior and this result can be used for spatial aggregation or clustering analysis.

The simulation results show a good capability for identifying the true patterns for high time dimension, whereas this result is poor when the time length is short. Anyway, the out–of–sample performance of the FSTAR model, in our practical application, seems to provide results outperforming the overparameterized USTAR model and the simple STAR model, but also a Sparse VAR model involving several coefficients.

One of the crucial problems of space–time models is the choice of the spatial weight matrix; in our application we have used eight exogenous weight matrices, which provide some differences in clustering for the FSTAR model; but the results, in terms of in–sample and out–of–sample performance, seem sufficiently robust. Alternatively, in the case of kernel spatial matrices, we could estimate the bandwidth h with the other unknown coefficients, considering an exogenous $\mathbf{W}$ matrix (see Otranto, Mucciardi and Bertuccelli, 2016).

Furthermore, the clustering and identification procedure can be extended to larger time and spatial lags, implying the consideration of a larger number of coefficients and a Wald test with several constraints to be jointly verified. Similar considerations could be made extending the methodology to STARMA models; in fact the clustering procedure is based on the Steece and Wood (1985) equivalence test, which was developed for general ARMA models. We leave these extensions to future research.

References

[1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

[2] Aquaro M., N. Bailey, H.M. Pesaran (2015). Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients, CESifo Working Paper N. 5428.

[3] Arbia, G., Bee, M. and Espa, G. (2011). Aggregation of regional economic time series with different spatial correlation structures. *Geographical Analysis* **43**, 78–103.

[4] Besag, J. (1974). Spatial Interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

[5] Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation.* Pion, London .

[6] Cliff, A.D. and Ord, J.K. (1975). Space–time modeling with an application to regional forecasting. *Transactions of the Institute of British Geographers* **64**, 119–128.

[7] Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes, Models and Applications.* Pion, London .

[8] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

[9] Fotheringham, A.S., Brunsdon, C., and Charlton, ME. (2002). *Geographically Weighted Regression: the Analysis of Spatially Varying Relationship.* Wiley, New York.

[10] Friedman, J., Hastie, T., and Tibshirani, R. (2008), Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics* **9**, 432–441.

[11] Frühwirth-Schnatter, S. (2011). Panel data analysis – a survey on model based clustering of time series. *Advances in Data Analysis and Classification* **5**, 251–280.

[12] Frühwirth-Schnatter, S. and S. Kaufmann (2008). Model–based clustering of multiple time series. *Journal of Business & Economic Statistics* **26**, 78–89.

[13] Getis, A. and Ord, J. K. (1981). The analysis of spatial association by use of distance statistics. *Geographical Analysis* **24**, 189–206.

[14] Giacomini, R. and Granger, C.W.J. (2004). Aggregation of space–time processes. *Journal of Econometrics* **118**, 7–26.

[15] Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge.

[16] Hansen, P.R. (2010). A Winner's curse for econometric models: on the joint distribution of in–sample fit and out–of–sample fit and its implications for model selection. *Mimeo*, Stanford University.

[17] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70.

[18] Hubert, L., Arabie, P. (1985). Comparing partitions. *Journal of Classification* **2**, 193–218.

[19] LeSage J., Y.-Y. Chih (2016). Interpreting heterogeneous coefficient spatial autoregressive models, *Economic Letters*, **142**, 1–5.

[20] LeSage, J. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall, Boca Raton.

[21] Liao, T. (2005). Clustering time series data: A survey. *Pattern Recognition* **38**, 1857–1874.

[22] Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin.

[23] Maharaj, E.A. (1999). Comparison and classification of stationary multivariate time series. *Pattern Recognition* **32**, 1129–1138.

[24] Mucciardi, M. and Bertuccelli, P. (2012). The impact of the weight matrix on the local indicators of spatial association: an application to per–capita value added in Italy. *International Journal of Trade and Global Markets* **5**, 133–141.

[25] Otranto, E. (2008). Clustering heteroskedastic time series by model–based procedures. *Computational Statistics and Data Analysis* **52**, 4685–4698.

[26] Otranto, E. (2010). Identifying financial time series with similar dynamic conditional correlation. *Computational Statistics and Data Analysis* **54**, 1–15.

[27] Otranto, E. and Gallo, G.M. (1994). Regression diagnostic techniques to detect space-to-time ratios in STARMA models. *Metron* **52**, 129–145.

[28] Otranto, E., Mucciardi, M. and Bertuccelli, P. (2016). Spatial effects in dynamic conditional correlations. *Journal of Applied Statistics* **43**, 604–626.

[29] Pfeifer, P.E. and Deutsch, S.J. (1980). A Three–stage iterative procedure for space–time modeling. *Technometrics* **22**, 35–47.

[30] Pinkse, J. and Slade, M.E. (1998). Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* **85**, 125–154.

[31] Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**, 846–850.

[32] Steece, B. and Wood, S. (1985). A test for the equivalence of k ARMA models. *Empirical Economics* **10**, 1–11.

[33] Stetzer, F. (1982). Specifying weights in spatial forecasting models: the results of some experiments. *Environment and Planning A* **14**, 571–584.

[34] Zou, H., Hastie, T. and Tibshirani, R. (2004). Sparse principal component analysis. *Technical Report*, Statistics Dept. Stanford University

Appendix: The clustering algorithm

The algorithm we propose can be synthesized in the following steps:

1. estimate all the N univariate STAR(1,1) models and put $N^* = N$ and $G = 0$;

2. verify the joint null hypothesis $\phi_h = \phi_i$ and $\psi_h = \psi_i$ for each h and i and, if at least one hypothesis is not rejected, select the two series with maximum p-value (corrected with the Holm, 1979, procedure) and put $G = G + 1$ and $N^* = N^* - 2$; otherwise stop the algorithm and put $G = G + N^*$;

3. estimate the STAR(1,1) model with the series selected in group G; call the coefficients ϕ_G and ψ_G;

4. verify the joint null hypothesis $\phi_G = \phi_i$ and $\psi_G = \psi_i$ for each remaining location i and, if at least one hypothesis is not rejected, select the series i which provides the maximum corrected p-value, add it to the previous group and put $N^* = N^* - 1$;

5. if at least one hypothesis is not rejected, repeat steps 3. and 4.; if all the null hypotheses are rejected, the series selected until step 4. form a group of locations with the same coefficients.

6. repeat steps 2.–5. with the remaining series until no series remain.

It is important to emphasize some points.

There are two *counters* (N^* and G) to record the number of series remaining and the number of groups identified respectively. Notice that the number of groups G is identified by the algorithm and does not need to be fixed a priori. When, in step 2., we reject all the null hypotheses, we will obtain N^* groups of size 1.

If all the elements of the row i of matrix $\mathbf{W}$ are equal to zero, the previous hypotheses, involving location i, are relative only to the ϕ coefficient because the ψ coefficient is not identified and is a nuisance parameter.

In step 2. the p–value is adopted to establish the order in which the locations enter into the groups. The Holm correction consists in ordering the unadjusted p–values such that $p_1 \leq p_2 \leq \cdots \leq p_m$ and then applying:

$$\tilde{p}_j = min[(m - j + 1)p_j, 1]$$

The corrected p–value $\tilde{p}_j$ is used as measure of similarity between STAR models in the algorithm

The test of the hypothesis in steps 2. and 4. can be performed by the Wald statistic. The validity of this approach in time series clustering has been supported by the simulation experiments of Otranto (2008 and 2010) and based on the theory developed by Steece and Wood (1985). More in detail, let us suppose that, in step 4., we compare the parameters of a group of

locations with the parameters of a location i; let us consider:

$$\theta = (\phi_{\mathbf{G}}, \psi_{\mathbf{G}}, \phi_{\mathbf{i}}, \psi_{\mathbf{i}})'; \qquad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_G} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_i} \end{bmatrix} \tag{6.1}$$

where $\mathbf{\Sigma_G}$ is the covariance matrix of the parameters estimator of group G, $\mathbf{\Sigma_i}$ the covariance matrix of the parameters estimator of the location i and $\mathbf{0}$ a matrix with all the elements equal to 0. Each matrix has dimension (2×2). Moreover let:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

be a constant matrix; the Wald statistic is expressed by:

$$\Xi = (\mathbf{C}\theta)' \left(\mathbf{C}\mathbf{\Sigma}\mathbf{C}'\right)^{-1} (\mathbf{C}\theta) \tag{6.2}$$

and follows a chi–squared distribution with 2 degrees of freedom.

In step 2., the same statistic is adopted, but in (6.1) we have to substitute ϕ_G, ψ_G and Σ_G with ϕ_h, ψ_h and Σ_h, respectively.

Ultimi Contributi di Ricerca CRENoS

I Paper sono disponibili in: http://www.crenos.it

17/06 *Simone Franceschini, Gerardo Ettore Marletto,* "The dynamics of social capital during public participation: new knowledge from an on-going monitoring"

17/05 *Luca G. Deidda, Ettore Panetti,* "Banks' Liquidity Management and Systemic Risk"

17/04 *Luca Frigau, Tiziana Medda, Vittorio Pelligra,* "From the Field to the Lab An Experiment on the Representativeness of Standard Laboratory Subjects"

17/03 *William Addessi, Manuela Pulina,* "Sectoral Composition of Consumption Expenditure: A Regional Analysis"

17/02 *Claudio Detotto, Marta Meleddu, Marco Vannini,* "Cultural identity and willingness to protect and preserve art"

17/01 *Adriana Di Liberto, Laura Casula,* "Teacher assessments versus standardized tests: is acting "girly" an advantage?"

16/16 *Massimo Del Gatto, Carlo S. Mastinu,* "Sequentiality and Distance(s) in Cross-Border Mergers and Acquisitions: Evidence from Micro Data"

16/15 *Angelo Antoci, Simone Borghesi, Gianluca Iannucci*, "Green licenses and environmental corruption: a random matching model"

16/14 *Vittorio Pelligra, Tommaso Reggiani, Daniel John Zizzo,* "Responding to (un)reasonable requests"

16/13 *Pasqualina Arca Gianfranco Atzeni, Luca Deidda,* "Asset exemption in entrepreneurs' bankruptcy and the informative role of collateral"

16/12 *Miguel Casares, Luca Deidda, Jose E. Galdon-Sanchez*, "Loan production and monetary policy"

16/11 *Manuela Pulina, Valentina Santoni,* "An analysis on the Italian agricultural firms: effects of public subsidies"

16/10 *Tiziana Medda, Vittorio Pelligra, Tommaso Reggiani,* "Does Experience Affect Fairness, Reciprocity and Cooperation in Lab Experiments?"

16/09 *Bianca Biagi, Maria Giovanna Brandano, Manuela Pulina,* "The effect of tourism taxation: a synthetic control approach"

16/08 *Edoardo Otranto*, *Massimo Mucciardi* , "A Flexible Specification of Space–Time AutoRegressive Models"

16/07 *Leonardo Becchetti, Vittorio Pelligra, Francesco Salustri,* "Testing for Heterogeneity of Preferences in Randomized Experiments: A Satisfaction-Based Approach Applied to Multiplayer Prisoners' Dilemmas"

16/06 *Anna Bottasso, Maurizio Conti, Giovanni Sulis,* "Firm Dynamics and Employment Protection: Evidence from Sectoral Data"

16/05 *Oliviero A. Carboni, Giuseppe Medda,* "R&D, Export, and Investment Decision"

16/04 *Leonardo Becchetti, Germana Corrado, Vittorio Pelligra, Fiammetta Rossetti,* "Happiness and Preferences in a Legality Social Dilemma Comparing the Direct and Indirect Approach"

www.crenos.it