Using R for Spatial Shift-Share Analysis

Gian Pietro Zaccomer
zaccomer@dss.uniud.it

Luca Grassetti
grassetti@dss.uniud.it

Department of Statistics
University of Udine

13 august 2008

Talk Outline

- The spatial shift-share analysis
- Our specific decomposition
- Some code-lines
- Results
- Concluding remarks and ongoing

Talk Outline

- The spatial shift-share analysis
- Our specific decomposition
- Some code-lines
- Results
- Concluding remarks and ongoing

Talk Outline

- The spatial shift-share analysis
- Our specific decomposition
- Some code-lines
- Results
- Concluding remarks and ongoing

Talk Outline

- The spatial shift-share analysis
- Our specific decomposition
- Some code-lines
- Results
- Concluding remarks and ongoing

Talk Outline

- The spatial shift-share analysis
- Our specific decomposition
- Some code-lines
- Results
- Concluding remarks and ongoing

The main purpose

The study we are presenting is about the development of a spatial shift-share decomposition model in R.

The presented application is about the spatial shift-share analysis of the labor data collected in the Italian Statistical Register of Active Enterprises (called ASIA) for the Friuli Venezia Giulia.

In particular, we concentrate on the occupation growth rate (g) of the manufacturing sector.

The "traditional" model

The classical model formulation (with 3 components) is generally referred to Dunn (1960). The growth rate in a Δt can be written as:

$$g_{r.} = \frac{\Delta x_{r.}}{x_{r.}} = g_{..} + \sum_{i=1}^{I} (g_{.i} - g_{..}) \frac{x_{ri}}{x_{r.}} + \sum_{i=1}^{I} (g_{ri} - g_{.i}) \frac{x_{ri}}{x_{r.}}$$

where:

X the variable investigated (economic phenomenon)
r the territorial unit (NUTS-5 classification) $r = 1, \ldots, R$
i the economic activity (NACE classification) $i = 1, \ldots, I$

The NH spatial model

Nazara and Hewings (2004) proposed to replace the national sector growth rate $g_{.i}$ with the equivalent neighboring growth rate $\check{g}_{ri}$ to obtain:

$$g_{r.} = g_{..} + \sum_{i=1}^{I} (\check{g}_{ri} - g_{..}) \frac{x_{ri}}{x_{r.}} + \sum_{i=1}^{I} (g_{ri} - \check{g}_{ri}) \frac{x_{ri}}{x_{r.}}$$

wherethe neighbouring growth rates may be written as:

$$\check{g}_{ri} = \frac{\sum_{s=1}^{R} \check{w}_{rs} x_{si}^{(t+1)} - \sum_{s=1}^{R} \check{w}_{rs} x_{si}^{(t)}}{\sum_{s=1}^{R} \check{w}_{rs} x_{si}^{(t)}}$$

and the row-standardized matrix W represents the spatial weight system.

The spatial model for the Italian Register of Businesses

The model proposed by Zaccomer (2006, 2007a) for the IRB data uses two decomposition factors: economic activity and enterprise legal status. This model is based on 6 components:

$$\begin{aligned} g_{r..} \quad &= \quad g_{...} + (\check{g}_{r..} - g_{...}) + \textstyle\sum_{f=1}^{F} (\check{g}_{r.f} - \check{g}_{r..}) \frac{x_{r.f}}{x_{r..}} \\ &\quad + \textstyle\sum_{i=1}^{I} (\check{g}_{ri.} - \check{g}_{r..}) \frac{x_{ri.}}{x_{r..}} + C_r + \textstyle\sum_{i=1}^{I} \sum_{f=1}^{F} (g_{rif} - \check{g}_{rif}) \frac{x_{rif}}{x_{r..}} \end{aligned}$$

where f identifies the enterprises' legal status, the component Cr is due to the presence of association between the two decomposition factors and can be written as:

$$C_r = \sum_{i=1}^{I} \sum_{f=1}^{F} (\check{g}_{rif} - \check{\delta}_{rif}) \frac{x_{rif}}{x_{r..}}$$

with $\check{\delta}_{rif} = \check{g}_{ri.} + \check{g}_{r.f} - \check{g}_{r..}$

The components of the IRB model

The growth rate $g_{r..}$ is then decomposed in

- (1) National component NAZ: the same in the classical model
- (2) Component CFR is related to the gap between the selected unit's neighbourhood and the national growth rate
- Intra-neighbourhood components: (3) by economic activity; (4) by legal status; (5) Cr (is null in presence of independence between industry mix and firm's legal status).
- (6) National (or regional) component LOC: based on the difference between unit and neighbouring rates, as in the NH model.

Spatial weight systems W

There are many methods to construct a spatial weight system. In this work, we classify them into three main groups:

- G1 based on the physical contiguity of any order (usually the first);
- G2 distance-based matrices;
- G3 based on a territorial reorganization (or "economic contiguity").

G1: contiguity matrices

The contiguity matrix is a symmetric square binary matrix defined by

$$w_{rs} = \begin{cases} 1 & \text{if } s \in V(r) \\ 0 & \text{if } s \notin V(r) \end{cases}$$

where $V(r)$ is the neighborhood of r-spatial unit. the neighborhood is built on two choices: the first is related to the criterion (i.g. rook or queen criterion) while the second to the spatial contiguity order.

G2: distance-based matrices (1)

- binary matrices with threshold

$$w_{rs} = \begin{cases} 1 & \text{if } d_{rs} \leq D_m \\ 0 & \text{if } d_{rs} > D_m \end{cases}$$

- simple inverse distance

$$w_{rs} = \frac{1}{d_{rs}^{\alpha}} = d_{rs}^{-\alpha}$$

- Cliff and Ord (1981) weights

$$w_{rs} = \frac{p_{rs}^{\beta}}{d_{rs}^{\alpha}}$$

- negative exponential (with threshold, Stetzer, 1982)

$$w_{rs} = \frac{1}{\exp(\alpha d_{rs})} = \exp(-\alpha d_{rs})$$

and

$$w_{rs} = \begin{cases} \exp(-\alpha d_{rs}) & \text{if } d_{rs} \leq D_m \\ 0 & \text{if } d_{rs} > D_m \end{cases}$$

G2: distance-based matrices (2)

- "economic distances" of Case, Rosen and Hines (1993) and Boarnet (1998) where E is an economic variable (e.g. *export*)

$$w_{rs} = \frac{1}{|E_r - E_s|} \text{ and } w_{rs} = \frac{\frac{1}{|E_r - E_s|}}{\sum_{s=1}^{R} \frac{1}{|E_r - E_s|}}$$

- Molho (1995) and Mitchell, Bill and Juniper (2005)

$$w_{rs} = \frac{E_s \exp(-\alpha d_{rs})}{\sum_{h \neq r}^{R} E_h \exp(-\alpha d_{rh})} \text{ and}$$

$$w_{rs} = \begin{cases} \frac{E_s \exp(-\alpha d_{rs})}{\sum_{h \neq r}^{R} E_h \exp(-\alpha d_{rh})} & \text{if } d_{rs} \leq D_m \\ 0 & \text{if } d_{rs} > D_m \end{cases}$$

G3: "Economic contiguity"-based W

Zaccomer (2006) proposes a new criterion to build the neighbourhood on a well-known spatial reorganization of the macro-area. This reorganization must be related to the economic phenomenon investigated. For example:

Industrial Districts: neighbourhood $\equiv$ quasi-ID
Labour Local Systems: neighbourhood $\equiv$ quasi-LLS

"Quasi" means that the study is based on the usual principle (for W based on the physical contiguity or distance) that a single territorial unit is not incorporated in its neighbourhood. This implies that all diagonal elements are $w_{rr} = 0$.

R implementation

The software used to carry out all decompositions, plots and prints functions is *R*.

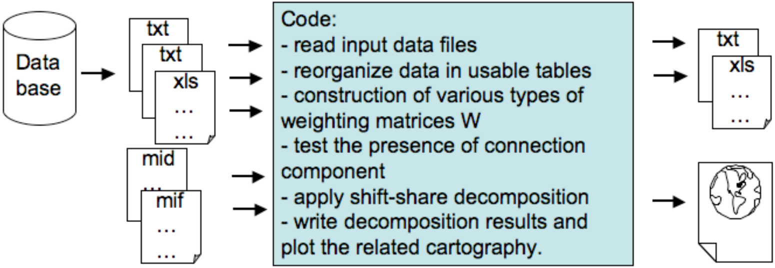
Firsts steps were developed in Zaccomer and Mason (2007), but now the *R* program takes all information directly from the *GIS* system and it is not necessary to use the software *GeoDa* (L. Anselin) for building W matrices.

By now each kind of spatial weight system can be constructed by this program (i.g. Cliff and Ord).

Finally, physical distances are now calculated on geographic coordinates of the town hall, and not on the simple polygon centroid.

The code structure

The procedure presents a hierarchical structure of nested micro functions. The use of the produced routine results is a sequence of preliminary actions, the call for the decomposition algorithm and a sequence of plot functions.

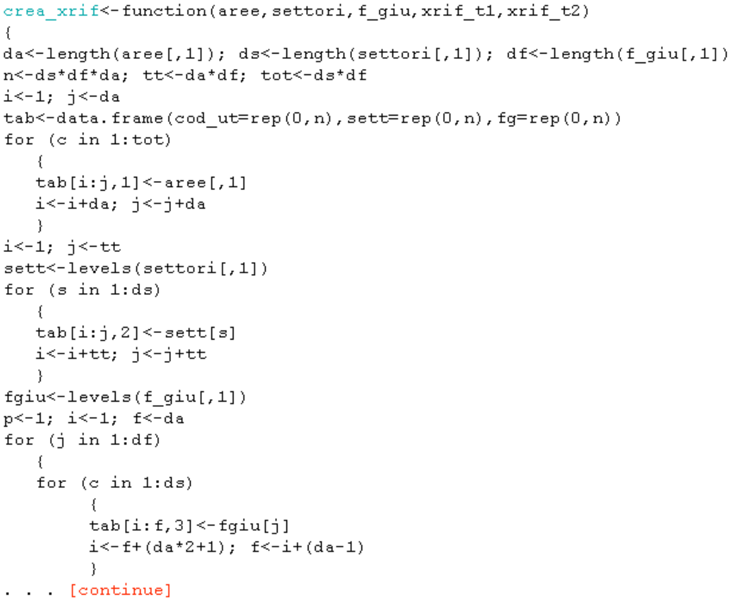


Code:
- read input data files
- reorganize data in usable tables
- construction of various types of weighting matrices W
- test the presence of connection component
- apply shift-share decomposition
- write decomposition results and plot the related cartography.

Some code-lines – Preliminary Phases - 1

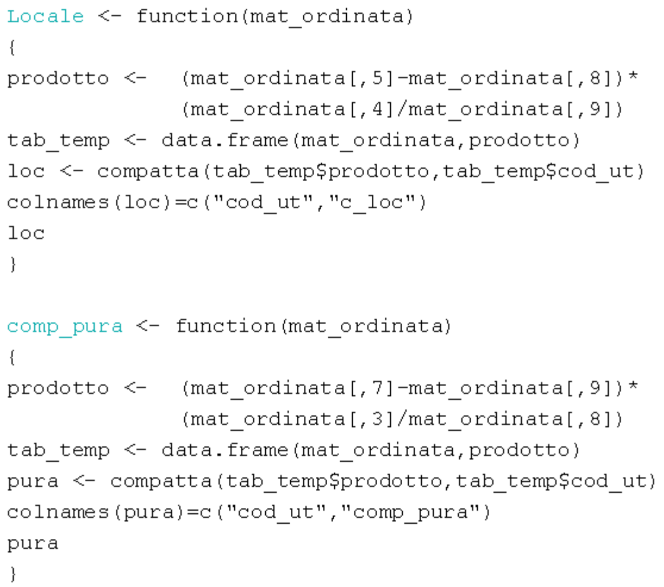
```
molho_distance<-function(dati,mat_w,alfa)
{
m<-exp_distance(mat_w,alfa)
l<-length(dati[,1])
w<-matrix(0,l,l)
rownames(w)<-rownames(m)
colnames(w)<-colnames(m)
for (j in 1:l)
  {
  for (i in 1:l)
      {
      if (i!=j) {w[i,j]<-dati[j,2]*m[i,j]}
      else w[i,j]<-0
      }
  }
mo_w<-standard_w(w)
mo_w
}
```

Some code-lines – Preliminary Phases - 2



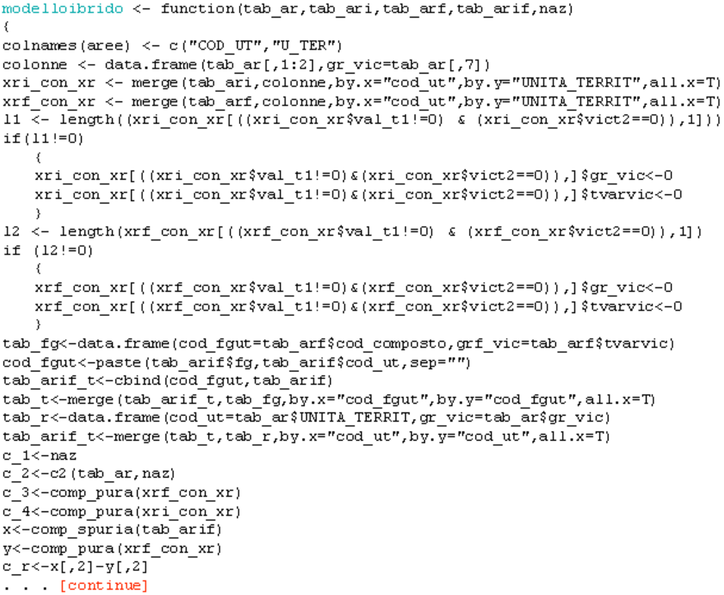
```
crea_xrif<-function(aree,settori,f_giu,xrif_t1,xrif_t2)
{
da<-length(aree[,1]); ds<-length(settori[,1]); df<-length(f_giu[,1])
n<-ds*df*da; tt<-da*df; tot<-ds*df
i<-1; j<-da
tab<-data.frame(cod_ut=rep(0,n),sett=rep(0,n),fg=rep(0,n))
for (c in 1:tot)
   {
   tab[i:j,1]<-aree[,1]
   i<-i+da; j<-j+da
   }
i<-1; j<-tt
sett<-levels(settori[,1])
for (s in 1:ds)
   {
   tab[i:j,2]<-sett[s]
   i<-i+tt; j<-j+tt
   }
fgiu<-levels(f_giu[,1])
p<-1; i<-1; f<-da
for (j in 1:df)
   {
   for (c in 1:ds)
        {
        tab[i:f,3]<-fgiu[j]
        i<-f+(da*2+1); f<-i+(da-1)
        }
. . . [continue]
```

Some code-lines – The SSS Decomposition - 1



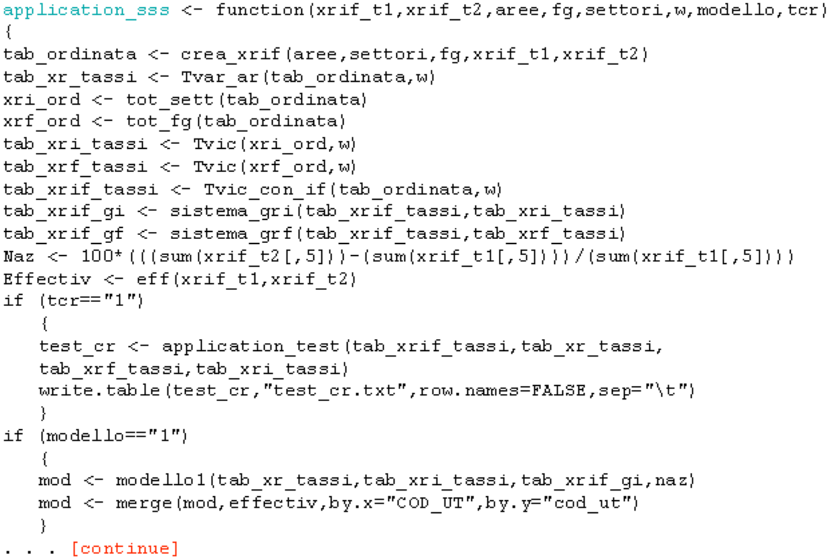
```
Locale <- function(mat_ordinata)
{
prodotto <-   (mat_ordinata[,5]-mat_ordinata[,8])*
              (mat_ordinata[,4]/mat_ordinata[,9])
tab_temp <- data.frame(mat_ordinata,prodotto)
loc <- compatta(tab_temp$prodotto,tab_temp$cod_ut)
colnames(loc)=c("cod_ut","c_loc")
loc
}

comp_pura <- function(mat_ordinata)
{
prodotto <-   (mat_ordinata[,7]-mat_ordinata[,9])*
              (mat_ordinata[,3]/mat_ordinata[,8])
tab_temp <- data.frame(mat_ordinata,prodotto)
pura <- compatta(tab_temp$prodotto,tab_temp$cod_ut)
colnames(pura)=c("cod_ut","comp_pura")
pura
}
```

Some code-lines – The SSS Decomposition - 2



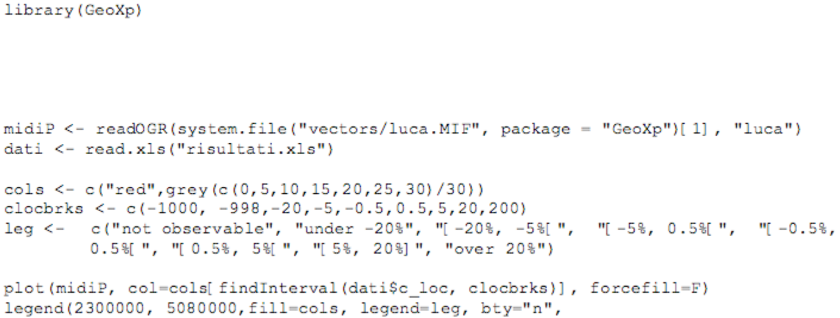
```
modelloibrido <- function(tab_ar,tab_ari,tab_arf,tab_arif,naz)
{
colnames(aree) <- c("COD_UT","U_TER")
colonne <- data.frame(tab_ar[,1:2],gr_vic=tab_ar[,7])
xri_con_xr <- merge(tab_ari,colonne,by.x="cod_ut",by.y="UNITA_TERRIT",all.x=T)
xrf_con_xr <- merge(tab_arf,colonne,by.x="cod_ut",by.y="UNITA_TERRIT",all.x=T)
l1 <- length((xri_con_xr[((xri_con_xr$val_t1!=0) & (xri_con_xr$vict2==0)),1]))
if(l1!=0)
   {
   xri_con_xr[((xri_con_xr$val_t1!=0)&(xri_con_xr$vict2==0)),]$gr_vic<-0
   xri_con_xr[((xri_con_xr$val_t1!=0)&(xri_con_xr$vict2==0)),]$tvarvic<-0
   }
l2 <- length(xrf_con_xr[((xrf_con_xr$val_t1!=0) & (xrf_con_xr$vict2==0)),1])
if (l2!=0)
   {
   xrf_con_xr[((xrf_con_xr$val_t1!=0)&(xrf_con_xr$vict2==0)),]$gr_vic<-0
   xrf_con_xr[((xrf_con_xr$val_t1!=0)&(xrf_con_xr$vict2==0)),]$tvarvic<-0
   }
tab_fg<-data.frame(cod_fgut=tab_arf$cod_composto,grf_vic=tab_arf$tvarvic)
cod_fgut<-paste(tab_arif$fg,tab_arif$cod_ut,sep="")
tab_arif_t<-cbind(cod_fgut,tab_arif)
tab_t<-merge(tab_arif_t,tab_fg,by.x="cod_fgut",by.y="cod_fgut",all.x=T)
tab_r<-data.frame(cod_ut=tab_ar$UNITA_TERRIT,gr_vic=tab_ar$gr_vic)
tab_arif_t<-merge(tab_t,tab_r,by.x="cod_ut",by.y="cod_ut",all.x=T)
c_1<-naz
c_2<-c2(tab_ar,naz)
c_3<-comp_pura(xrf_con_xr)
c_4<-comp_pura(xri_con_xr)
x<-comp_spuria(tab_arif)
y<-comp_pura(xrf_con_xr)
c_r<-x[,2]-y[,2]
. . . [continue]
```

Some code-lines – The SSS Decomposition - 3



```
application_sss <- function(xrif_t1,xrif_t2,aree,fg,settori,w,modello,tcr)
{
tab_ordinata <- crea_xrif(aree,settori,fg,xrif_t1,xrif_t2)
tab_xr_tassi <- Tvar_ar(tab_ordinata,w)
xri_ord <- tot_sett(tab_ordinata)
xrf_ord <- tot_fg(tab_ordinata)
tab_xri_tassi <- Tvic(xri_ord,w)
tab_xrf_tassi <- Tvic(xrf_ord,w)
tab_xrif_tassi <- Tvic_con_if(tab_ordinata,w)
tab_xrif_gi <- sistema_gri(tab_xrif_tassi,tab_xri_tassi)
tab_xrif_gf <- sistema_grf(tab_xrif_tassi,tab_xrf_tassi)
Naz <- 100*(((sum(xrif_t2[,5]))-(sum(xrif_t1[,5])))/(sum(xrif_t1[,5])))
Effectiv <- eff(xrif_t1,xrif_t2)
if (tcr=="1")
   {
   test_cr <- application_test(tab_xrif_tassi,tab_xr_tassi,
   tab_xrf_tassi,tab_xri_tassi)
   write.table(test_cr,"test_cr.txt",row.names=FALSE,sep="\t")
   }
if (modello=="1")
   {
   mod <- modello1(tab_xr_tassi,tab_xri_tassi,tab_xrif_gi,naz)
   mod <- merge(mod,effectiv,by.x="COD_UT",by.y="cod_ut")
   }
. . . [continue]
```

Some code-lines – The Cartography



```
library(GeoXp)



midiP <- readOGR(system.file("vectors/luca.MIF", package = "GeoXp")[ 1] , "luca")
dati <- read.xls("risultati.xls")

cols <- c("red",grey(c(0,5,10,15,20,25,30)/30))
clocbrks <- c(-1000, -998,-20,-5,-0.5,0.5,5,20,200)
leg <-   c("not observable", "under -20%", "[ -20%, -5%[ ",  "[ -5%, 0.5%[ ",  "[ -0.5%,
         0.5%[ ", "[ 0.5%, 5%[ ", "[ 5%, 20%] ", "over 20%")

plot(midiP, col=cols[ findInterval(dati$c_loc, clocbrks)] , forcefill=F)
legend(2300000, 5080000,fill=cols, legend=leg, bty="n",
```

Empirical results

The application was carried out on regional industrial employment data for 2001-04. These refer to the Italian Business Statistical Register (ASIA) for 214 municipalities (LAU2 level) and 5 municipalities are omitted because they do not present any manufacturing enterprise.

The dataset structure counts:

- 12 LLS of the FVG (NUTS2 level)
- 10 manufacturing sectors are obtained from NACE Rev. 1.1 [The enterprises entering sector D are grouped in 10 clusters.]
- 3 legal status
  - sole
  - limited
  - unlimited

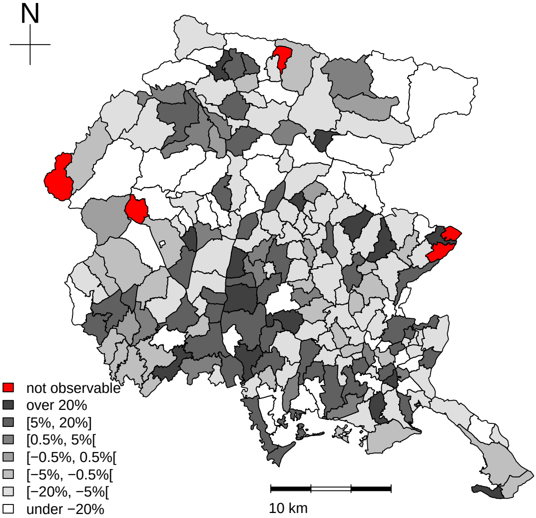
Example 1: details

Some interesting results regard the municipalities of the *LLS* number **176**. In these results one can observe:

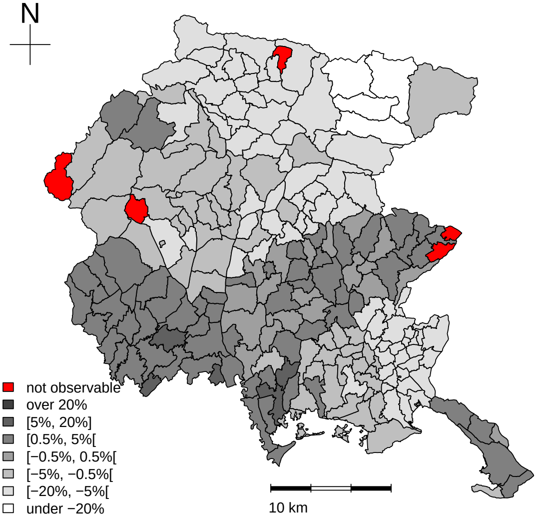
1. the national component is negative
2. the structural component capture the effect of the economic context
3. the intra-neighborhood component assumes very different patterns given the characteristics of the considered enterprises
4. the local component works as a residual effect

					Intra.Neig.		
Micro Area	**Growth Rate**	**Nat. Comp.**	**Str. Comp.**	**Loc. Comp.**	**L.S. Comp.**	**E.A. Comp.**	**Con. Comp.**
93002	99.17	-4.06	-5.26	-6.51	-10.40	-4.89	130.30
93003	-8.11	-4.06	4.43	-4.30	4.32	-4.10	-4.40
93004	-2.57	-4.06	4.47	-0.77	-2.79	0.54	0.04
93005	-4.10	-4.06	4.64	-3.45	-4.35	2.34	0.77
93006	0.00	-4.06	-4.28	-7.64	-4.46	-29.56	50.00
93007	-1.56	-4.06	4.58	0.40	0.23	-0.15	-2.56
93008	-2.14	-4.06	4.43	0.94	-0.43	-0.73	-2.30
93009	-24.10	-4.06	4.79	-1.17	1.68	0.18	-25.51
93010	-4.66	-4.06	4.46	-2.16	-4.71	1.27	0.55

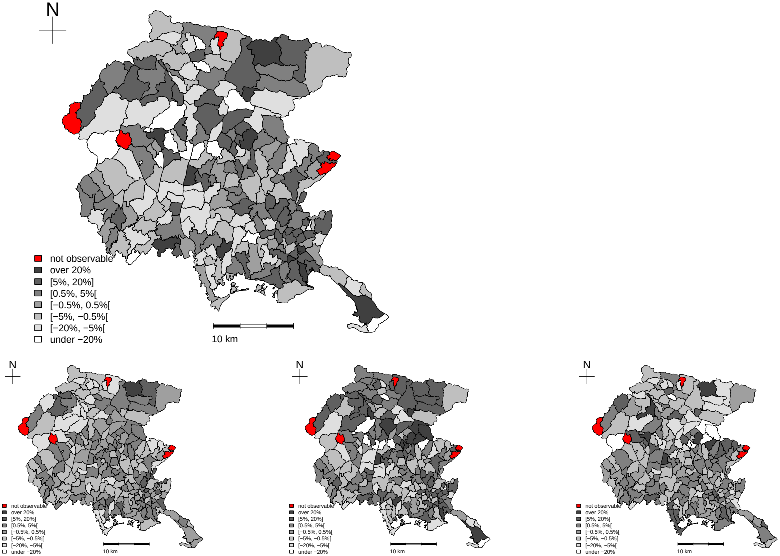
Example 1: cartography - growth rates



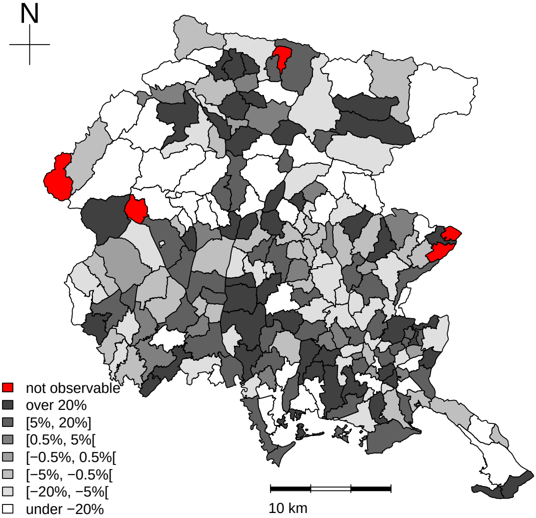
Example 1: cartography - structural component



Example 1: cartography - intra-neigh. component



Example 1: cartography - local component

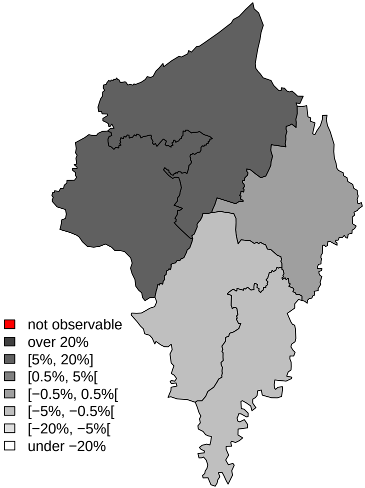


Example 2: details

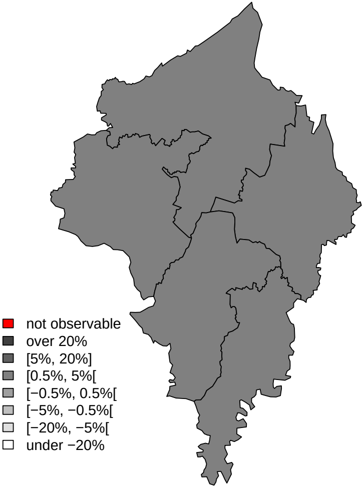
The decomposition considering the macro area of LLS brings to the following results.

Area	Growth Rate	Nat. Comp.	Str. Comp.	Loc. Comp.	L.S. Comp.	Intra.Neig. E.A. Comp.	Con. Comp.
156	1.93	-4.06	4.39	-0.30	-1.71	1.10	2.51
166	-5.36	-4.06	-3.83	-0.96	4.65	2.92	-4.07
167	-0.91	-4.06	-4.56	-0.37	2.49	0.37	5.21
168	-4.56	-4.06	1.17	0.52	0.08	1.19	-3.47
169	1.18	-4.06	3.51	-0.12	-0.54	2.01	0.38
170	-28.44	-4.06	-5.12	0.00	1.71	0.60	-21.58
171	-9.18	-4.06	-2.36	0.00	3.57	0.06	-6.38
172	-0.71	-4.06	-0.69	-0.27	0.36	0.61	3.34
173	-19.98	-4.06	2.48	-0.94	0.01	1.28	-18.75
174	-2.91	-4.06	-15.92	-2.32	38.00	-6.09	-12.53
175	-7.11	-4.06	3.34	-0.80	0.99	-0.05	-6.52
176	0.86	-4.06	2.28	1.36	2.04	1.03	-1.78

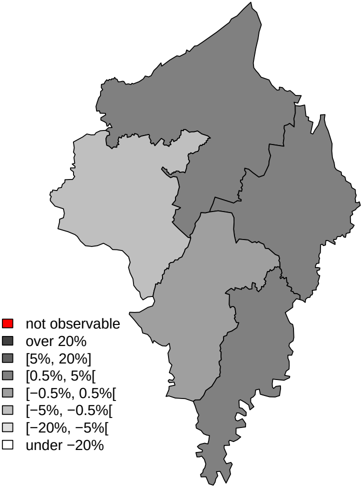
Detailed cartography - growth rates



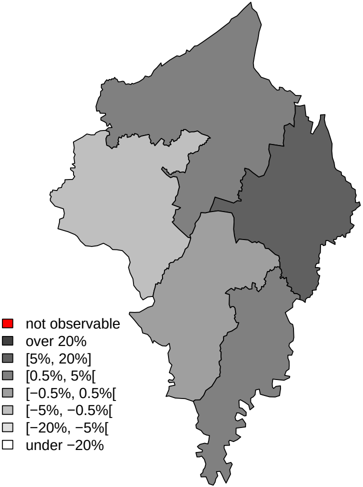
Detailed cartography - structural component



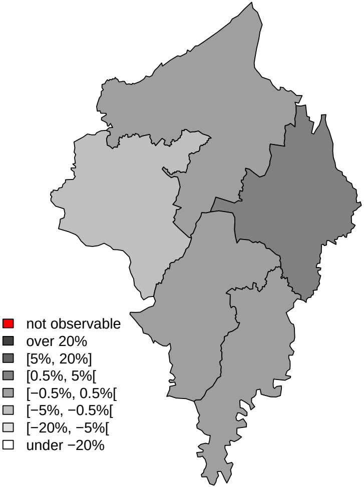
Detailed cartography - legal status component



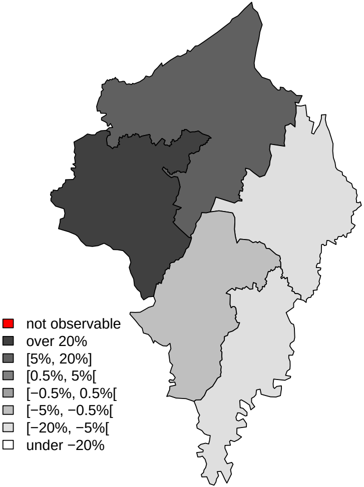
Detailed cartography - econ.act. component



Detailed cartography - connection component



Detailed cartography - local component



Concluding remarks and ongoing

Till now we developed

- a full 6 component shift-share decomposition
- the code for data reorganization and preliminary analysis
- the distance calculation (considering all possible distances)
- an integrated cartography adopting the package "GeoXp" and all correlated packages.

And now its time for

- some necessary code refinement