sptpsar: an R package for spatio-temporal semiparametric models

International Conference on Regional Science. Valencia

Román Mínguez, University of Castilla-La Mancha (Spain); Roberto Basile University of L'Aquila (Italy); María Durbán, University Carlos III (Spain)

21-23 Noviembre (2018)

Abstract

A collection of R functions for estimation and inference of spatial and spatio-temporal semiparametric models including spatial or spatio-temporal non-parametric trend, parametric and non-parametric covariates and, possibly, a spatial lag for the dependent variable. The spatio-temporal trend can be decomposed in ANOVA way including main and interaction components for the trend.

---

1. Description of the model

2. Technicalities

3. Examples

4. Conclusions & work to do

5. References

Description of the general specification modeled in `sptpsar`

Spatio-Temporal Semiparametric model with a spatial lag and AR(1) disturbances.

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2(\mathbf{I}_N \otimes \mathbf{R}_T))$$

This model tries to capture:

- spatial heterogeneity bias
- omitted time-related factors bias
- spatial dependence bias
- functional form bias
- Details in Mínguez, R.; Basile, R. and Durbán, M., 2018: An Alternative Semiparametric Model for Spatial Panel Data. Under evaluation in *Regional Science and Urban Economics*

Technicalities…

Non-parametric terms are modeled using P-Splines

Univariate

$$f(\mathbf{x}) = \sum_{j=1}^{c} \mathbf{B}_j(\mathbf{x})\theta_j, \quad j = 1, \ldots, c$$

with $\mathbf{B}_j$ a B-spline basis function, and θ_j a component of a vector of regression coefficients of length c (the number of knots used to construct the basis).

Penalized regression problem (assuming *no spatial lag* and *no temporal autocorrelation*):

$$\|\mathbf{y} - \mathbf{B}\boldsymbol{\theta}\|^2 + \lambda\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$$

with $\mathbf{P} = \mathbf{D}'\mathbf{D}$

$\mathbf{D}$ is the matrix of second-order differences and λ is the *smoothing parameter*.

Bivariate

$$f_{1,2}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^{c_i}\sum_{l=1}^{c_k} \mathbf{B}_j(\mathbf{x}_i)\mathbf{B}_l(\mathbf{x}_k)\theta_{jl}, \quad j = 1, \ldots, c_1 \quad l = 1, \ldots, c_2$$

Penalty term (allowing anisotropy):

$$\mathbf{P} = \lambda_1\mathbf{D}_1'\mathbf{D}_1 \otimes \mathbf{I} + \lambda_2\mathbf{I} \otimes \mathbf{D}_2'\mathbf{D}_2$$

- Details in Eilers, P. and Marx, B.; 1996: Flexible Smoothing with B-Splines and Penalties. *Statistical Science* 11, 89-121.

ANOVA decomposition of spatio-temporal trend.

Decompose spatio-temporal trend in main and interaction terms.

$$f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) = f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_\tau) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_\tau) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_\tau) + f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau)$$

- Details in Lee, D. and Durbán, M.; 2011: P-Spline ANOVA Type Interaction Models for Spatio-Temporal Smoothing. *Statistical Modelling* 11, 49-69.

Nested basis for interaction terms in ANOVA decomposition of spatio-temporal trend.

Idea: use a matrix $\breve{\mathbf{B}}$ in the interaction terms, such that the space spanned by this matrix is a subset of the space spanned by $\mathbf{B}$.

$$\texttt{ndx}^* \text{ of } \breve{\mathbf{B}} = \frac{\texttt{ndx} \text{ of } \mathbf{B}}{\texttt{div}} \Rightarrow \mathsf{span}(\breve{\mathbf{B}}) \subset \mathsf{span}(\mathbf{B})$$

- Details in Lee, D. J.; Durban, M. and Eilers, P.; 2013: Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Computational Statistics & Data Analysis* 61, 22-37.

Express the semiparametric model as a mixed model and apply REML.

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \quad \sim N(\mathbf{0}, \mathbf{G}),$$

where matrices $\mathbf{X}^*$, $\mathbf{Z}$ and $\mathbf{G}$ are computed using an appropriate decomposition on the penalty term. This approach allows the simultaneous estimation of all parameters in the model.

- Wood, S.N.; 2017: *Generalized Additive Models: An Introduction with R, Second Edition* CRC-Press.
- Fahrmeir, L., Kneib, Th., Lang, S., Marx, B.; 2013: *Regression: Models, Methods and Applications*, Springer

Use *SAP* to compute smoothing parameters without apply numerical optimization.

Separation of Anisotropic Penalties, *SAP*, algorithm maximizes REML much faster than numerical optimization procedures can. This algorithm can be applied to any mixed model in which the precision matrix of random effects $(\mathbf{G}^{-1})$ can be expressed as a linear combination over the variance components. SAP is computationally very efficient and the estimates produced match those obtained using REML.

- Details in Rodriguez-Alvarez, M. X.; Kneib, T.; Durban, M.; Lee, D.J.; and Eilers, P.; 2015: Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm. *Statistics and Computing* 25 (5), 941-957.

Use of numerical optimization to estimate ρ and ϕ parameters.

These parameters correspond to the spatial lag and the AR(1) temporal correlation, respectively. They are estimated using numerical techniques (no closed solutions).

SAP and estimation of ρ and ϕ are iterated until convergence.

- Details in Mínguez, R.; Basile, R. and Durbán, M., 2018: An Alternative Semiparametric Model for Spatial Panel Data. Under evaluation in *Regional Science and Urban Economics*.

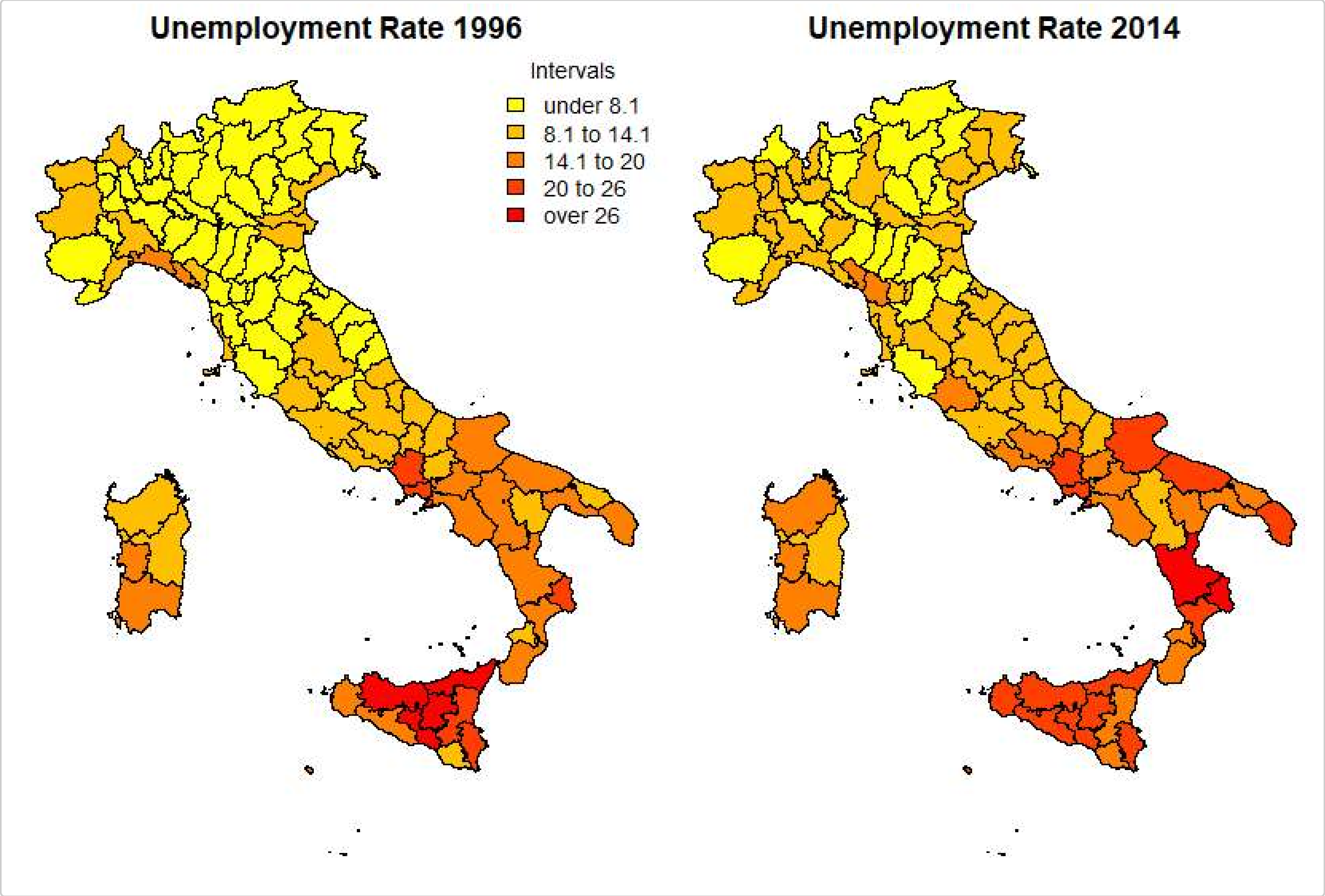
Examples

Data sets in `sptpsar`

Panel data of rate of unemployment for 103 Italian provinces in period 1996-2014

```r
library(sptpsar)
 data(unemp_it); head(unemp_it)
```

```
##     prov   name      reg year    area    unrate     agri      ind     cons
## 1      1 Torino Piemonte 1996 6830.25 12.348965 2.692673 29.55477 5.304503
## 104    1 Torino Piemonte 1997 6830.25 13.210734 2.615366 29.50610 5.355169
## 207    1 Torino Piemonte 1998 6830.25 12.510946 2.418538 30.03541 5.342118
## 310    1 Torino Piemonte 1999 6830.25 10.897371 2.283917 28.87175 5.466022
## 413    1 Torino Piemonte 2000 6830.25  9.696943 2.245436 27.28693 5.623352
## 516    1 Torino Piemonte 2001 6830.25  7.924167 2.272278 26.12132 5.710334
##         serv   popdens partrate   empgrowth   long     lat South
## 1   62.44805 0.3214042 45.15605  1.39978225 377217 5000171     0
## 104 62.52337 0.3203764 45.24058 -1.11505704 377217 5000171     0
## 207 62.20393 0.3194230 44.17415 -1.86285651 377217 5000171     0
## 310 63.37832 0.3184794 45.07623  3.61708405 377217 5000171     0
## 413 64.84428 0.3176521 45.75648  2.60942492 377217 5000171     0
## 516 65.89607 0.3170161 44.92868 -0.08196475 377217 5000171     0
##     ln_popdens
## 1    -1.135056
## 104  -1.138259
## 207  -1.141239
## 310  -1.144197
## 413  -1.146799
## 516  -1.148803
```



Spatial semiparametric model without spatial trend (GAM-SAR).

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$$

```
form1 <- unrate ~ partrate + agri + cons +
                  pspl(serv,nknots=15) +
                  pspl(empgrowth,nknots=20)
Wsp <- Wsp_it
gamsar <- psar(form1,data=unemp_it,sar=TRUE,Wsp=Wsp_it)
```

```
##
## Fitting Model...
##
##  Time to fit the model:  50.06 seconds
##  Time to compute covariances:  0.02 seconds
```

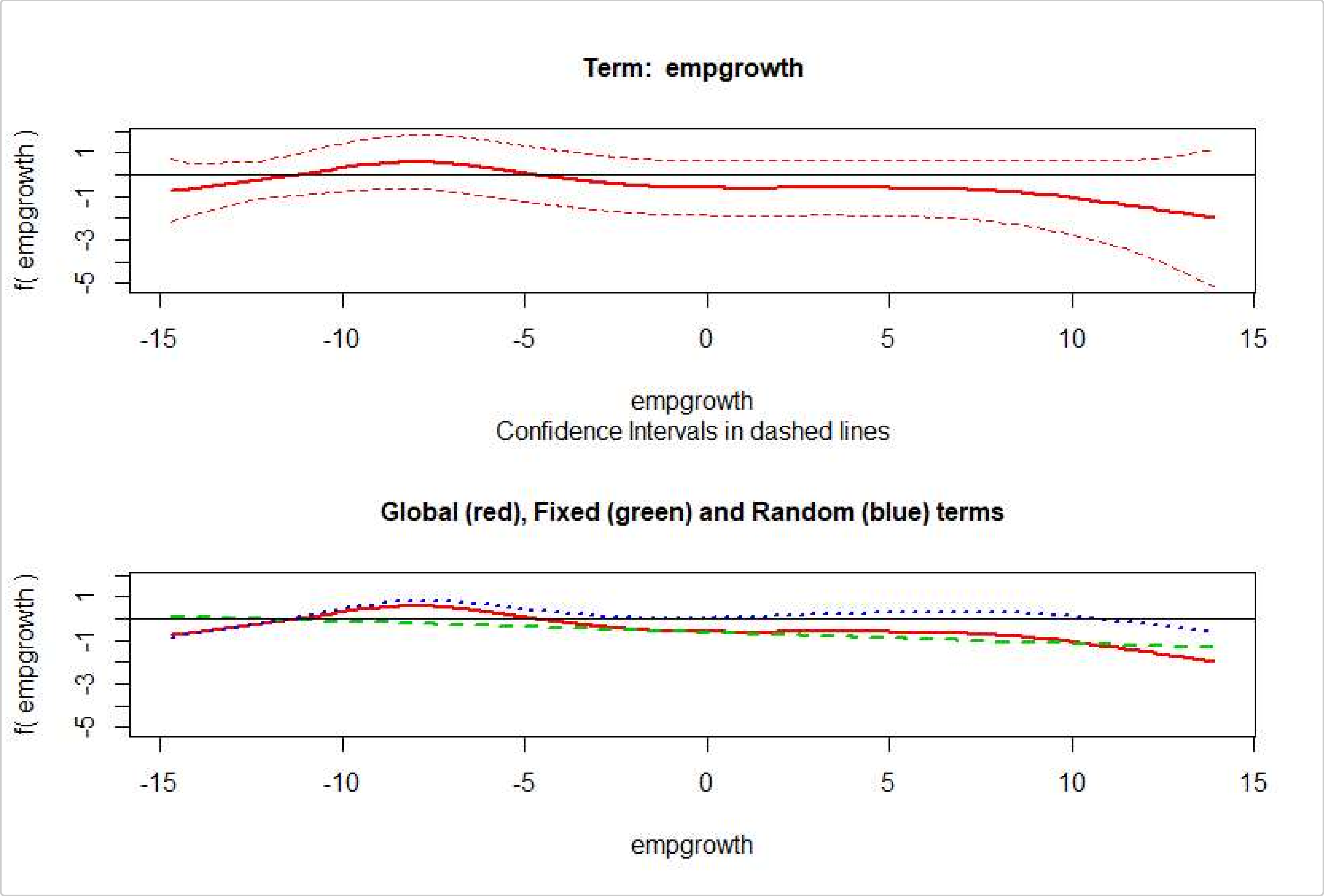
```
summary(gamsar)
```

```
## Call:
## psar(formula = form1, data = unemp_it, Wsp = Wsp_it, sar = TRUE)
##
##  Parametric Terms
##            Estimate Std. Error  t value  Pr(>|t|)
## Intercept 14.771556   1.076340  13.7239 < 2.2e-16 ***
## partrate  -0.267609   0.017120 -15.6311 < 2.2e-16 ***
## agri       0.049098   0.014750   3.3287 0.0008889 ***
## cons      -0.033630   0.037508  -0.8966 0.3700385
## rho        0.615791   0.015873  38.7954 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Non-Parametric Terms
##                                 EDF
## pspl(serv, nknots = 15)      7.6258
## pspl(empgrowth, nknots = 20) 5.8075
##
##  Goodness-of-Fit
##
## EDF Total: 20.4333  Sigma: 2.53269
## AIC:       5968.21 BIC:  6082.21
```

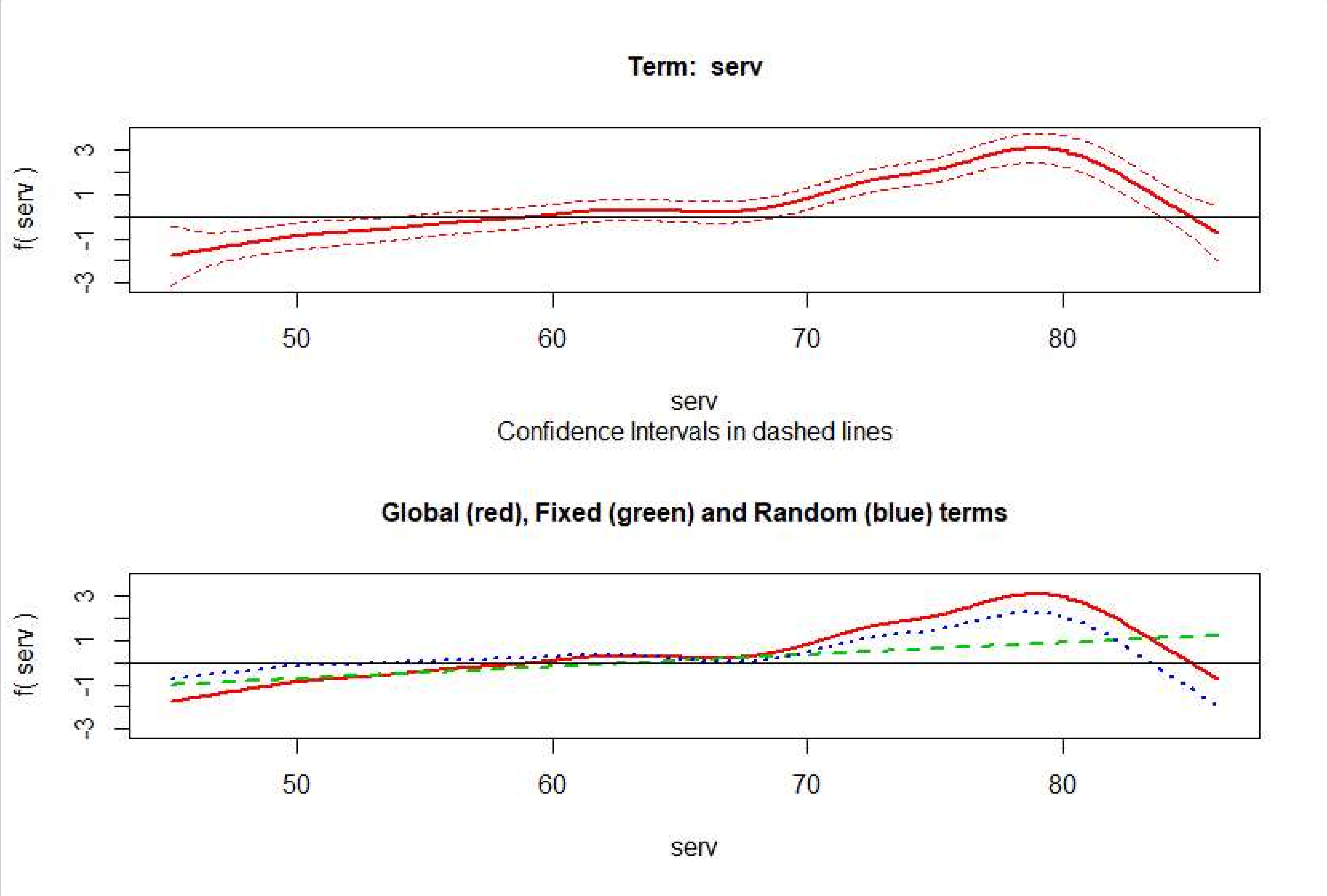
Plot of Non-Parametric Terms

```
######################  Fit non-parametric terms
 list_varnopar <- c("serv","empgrowth")
 terms_nopar <- fit_terms(gamsar,list_varnopar)
######################  Plot non-parametric terms
 par(mar=c(1,1,1,1))
 plot_terms(terms_nopar,unemp_it)
```

Using mixed model representation, non-parametric terms can be decomposed in (unpenalized) fixed term and (penalized) random term.


Term: empgrowth
f(empgrowth)
1
-1
-3
-5
-15
-10
-5
0
5
10
15
empgrowth
Confidence Intervals in dashed lines
Global (red), Fixed (green) and Random (blue) terms
f(empgrowth)
1
-1
-3
-5
-15
-10
-5
0
5
10
15
empgrowth

[1]



[2]

Parametric Total, Direct and Indirect Effects

$$(\mathbf{I}_{NT} - (\rho \mathbf{W}_N \otimes \mathbf{I}_T))^{-1} \mathbf{X}\boldsymbol{\beta} = \mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta}$$

```
list_varpar <- c("partrate","agri","cons")
eff_parvar <- eff_par(gamsar,list_varpar)
```

```
summary(eff_parvar)
```

```
##
##  Total Parametric Effects
##             Estimate Std. Error    t value Pr(>|t|)
## partrate  -0.699406   0.054293 -12.882115   0.0000
## agri       0.127786   0.040755   3.135495   0.0017
## cons      -0.084396   0.096587  -0.873782   0.3822
##
##  Direct Parametric Effects
##             Estimate Std. Error    t value Pr(>|t|)
## partrate  -0.318452   0.021133 -15.068921   0.0000
## agri       0.058191   0.018456   3.152995   0.0016
## cons      -0.038374   0.043896  -0.874194   0.3820
##
##  Indirect Parametric Effects
##             Estimate Std. Error    t value Pr(>|t|)
## partrate  -0.380954   0.035703 -10.670023   0.0000
## agri       0.069594   0.022459   3.098767   0.0019
## cons      -0.046023   0.052750  -0.872465   0.3830
```

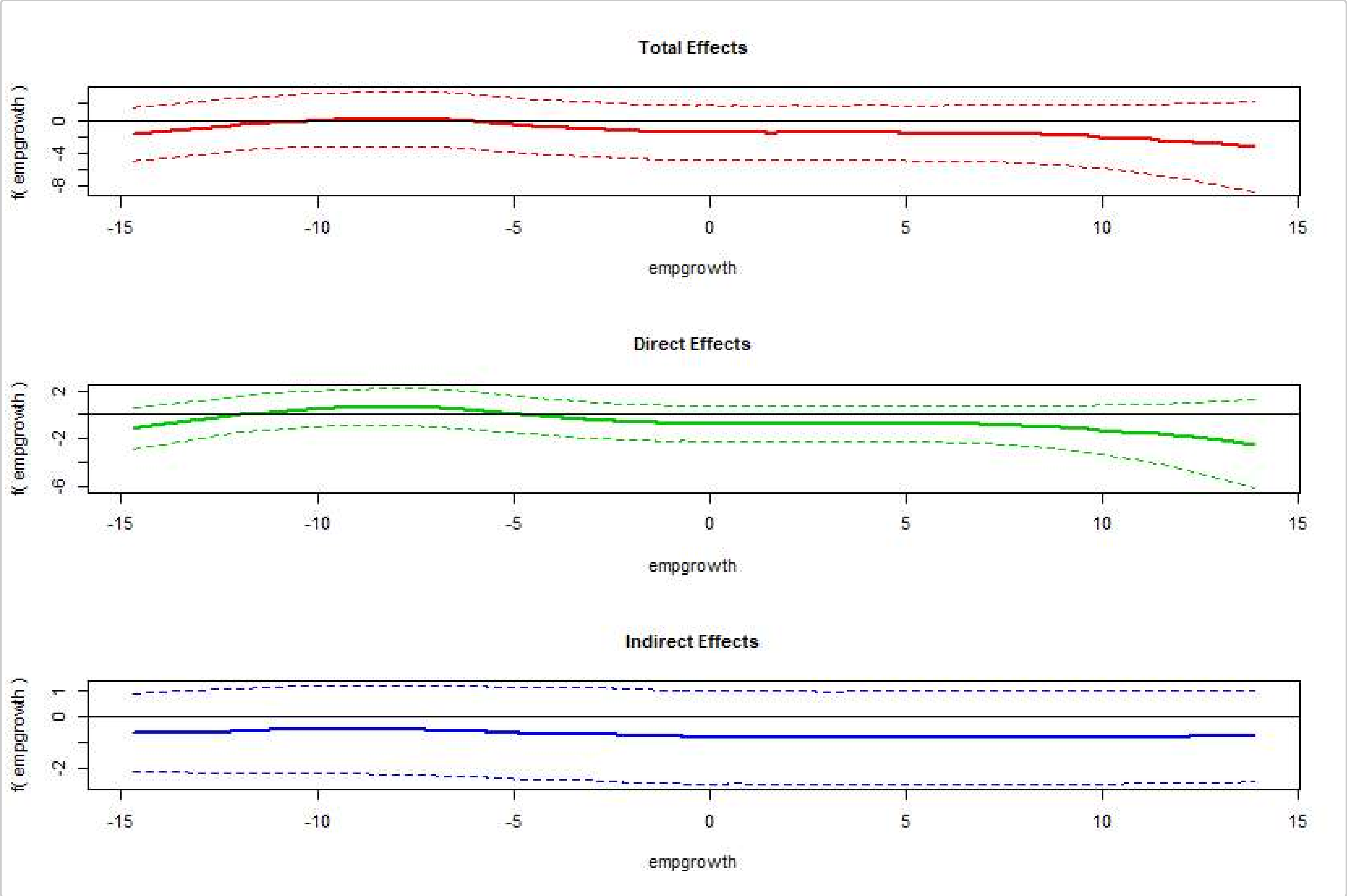
Plot of Non-Parametric Total, Direct and Indirect Effects

$$(\mathbf{I}_{NT} - (\rho \mathbf{W}_N \otimes \mathbf{I}_T))^{-1} \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) = \mathbf{A}^{-1} \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta)$$

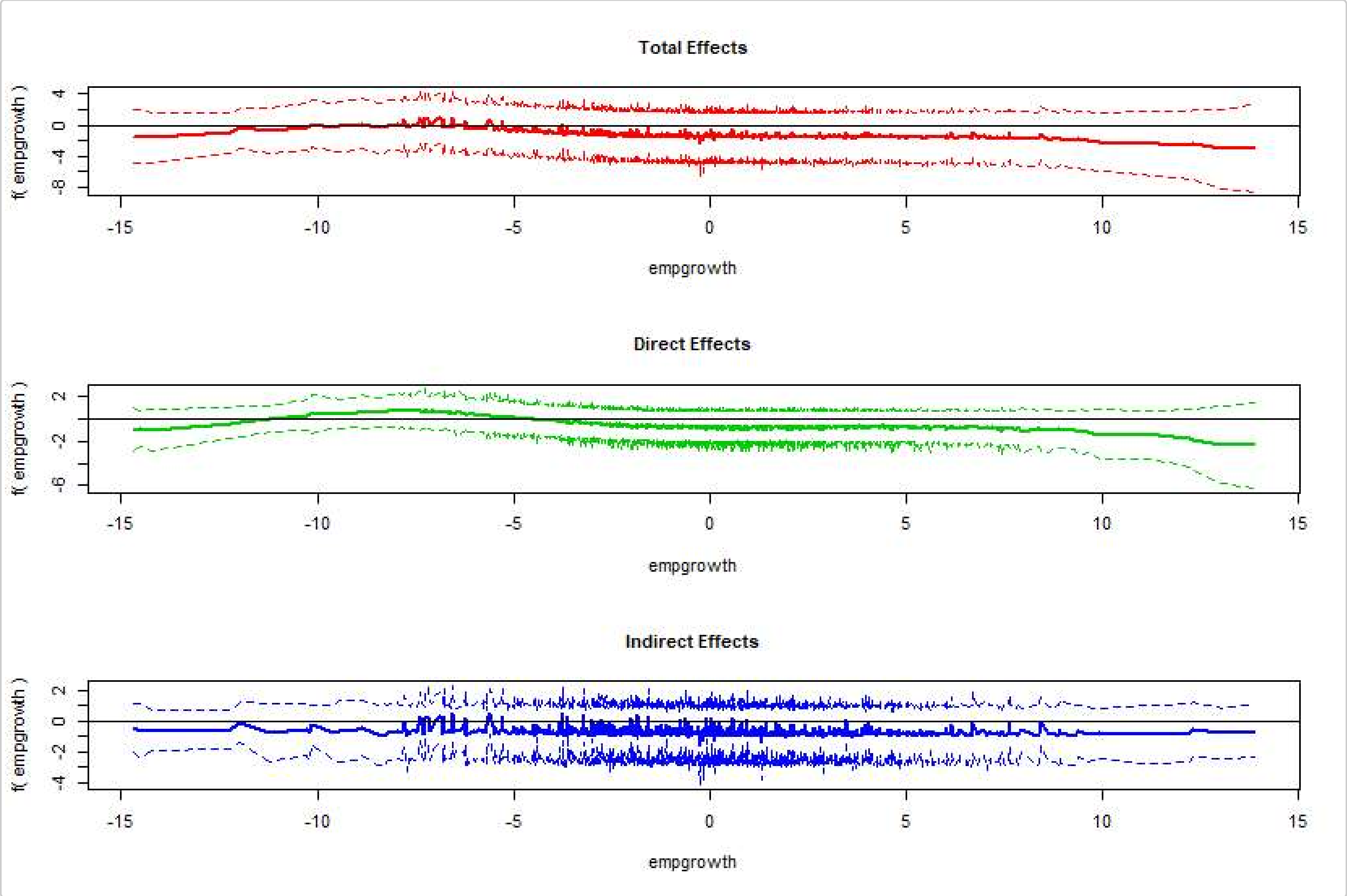
The non-parametric functions are fitted using the mixed model representation and premultiplied by $\mathbf{A}^{-1}$ matrix.

Especially for the indirect effects, the estimates could be unstable and it is convenient to smooth them to visualize better.

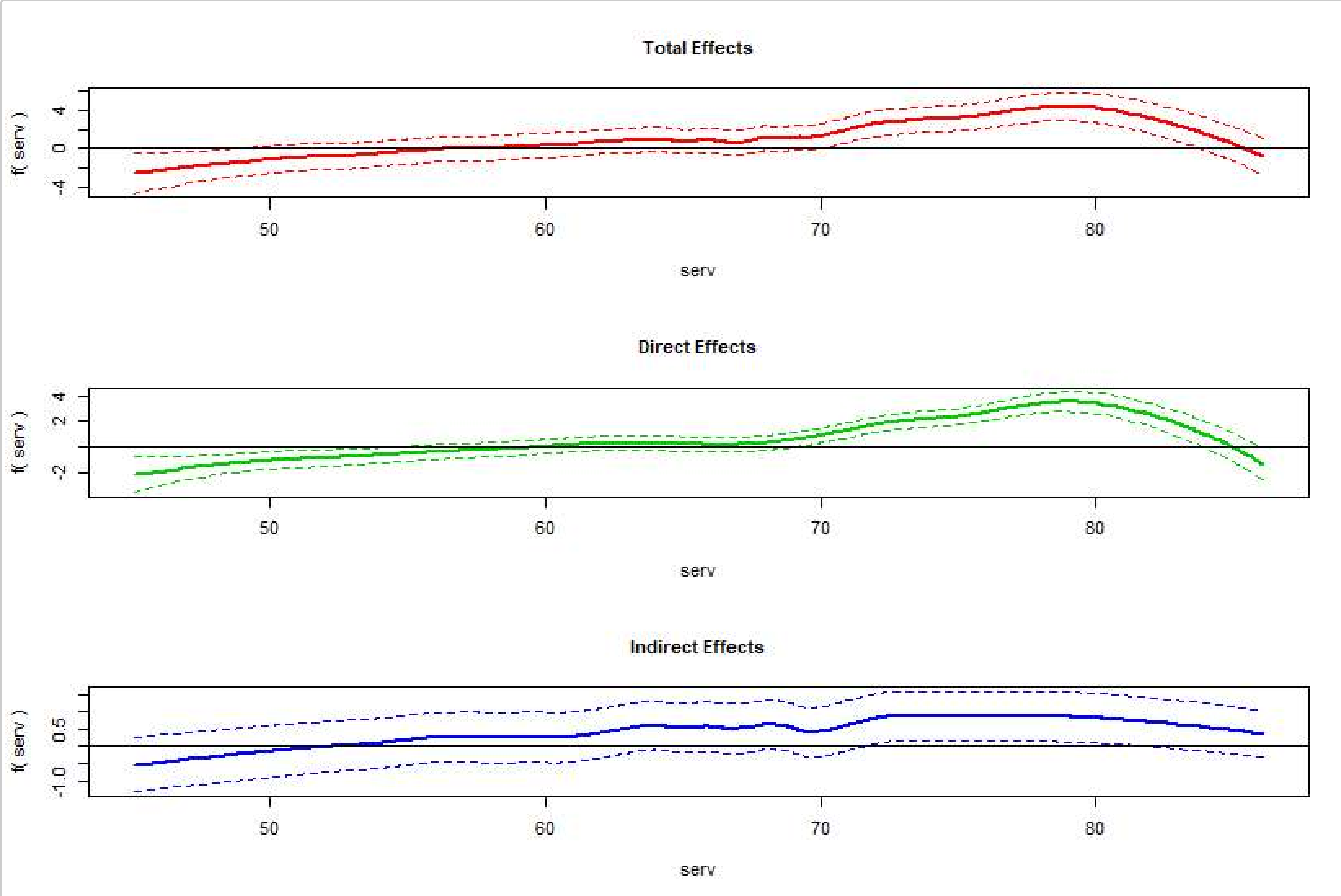
```
eff_nparvar <- eff_nopar(gamsar,list_varnopar)
plot_effects_nopar(eff_nparvar,unemp_it,smooth=TRUE)
plot_effects_nopar(eff_nparvar,unemp_it,smooth=FALSE)
```


Total Effects
f(empgrowth)
0
-4
-8
-15
-10
-5
0
5
10
15
empgrowth
Direct Effects
f(empgrowth)
2
-2
-6
-15
-10
-5
0
5
10
15
empgrowth
Indirect Effects
f(empgrowth)
1
0
-2
-15
-10
-5
0
5
10
15
empgrowth

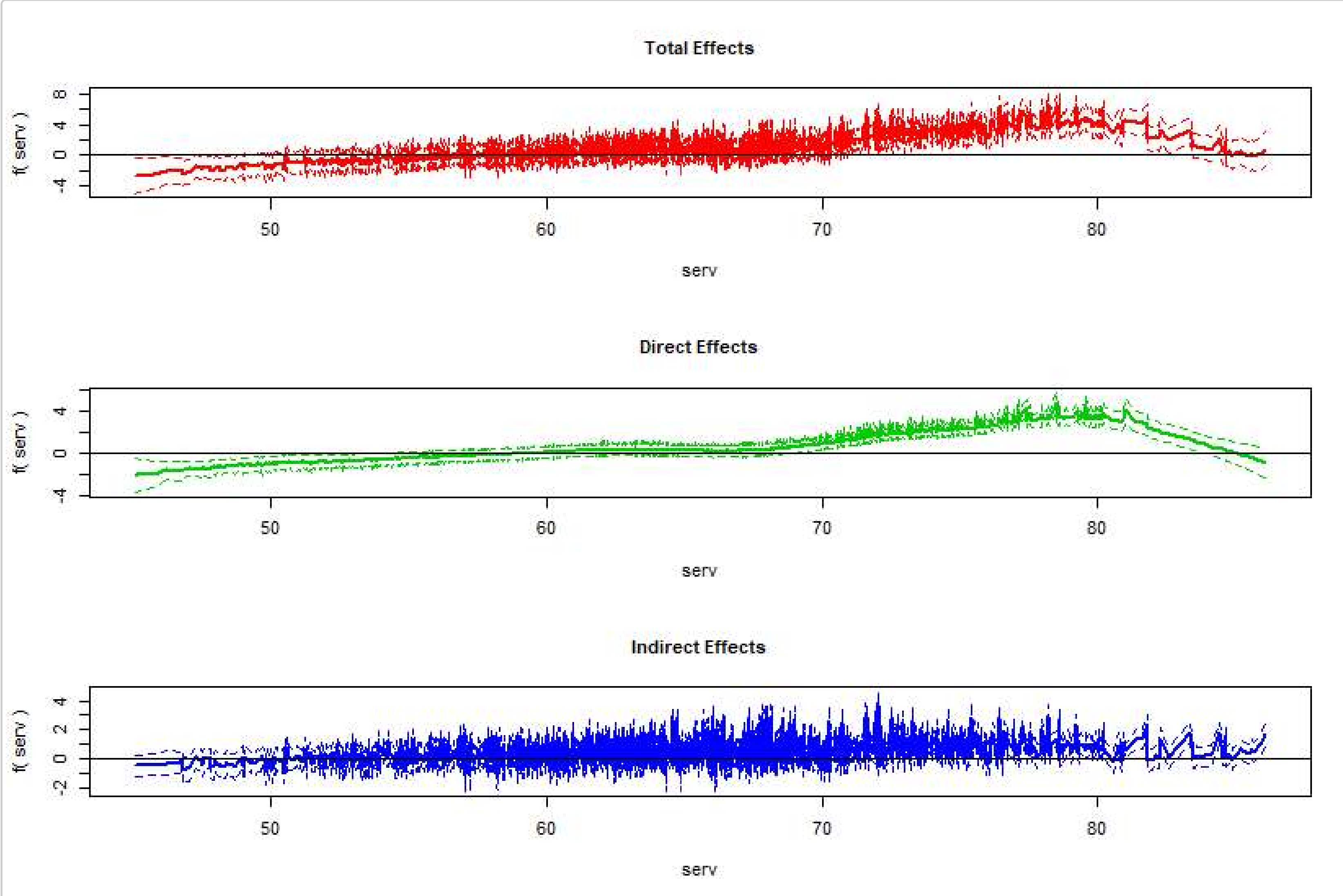
[1]


Total Effects
f(empgrowth)
empgrowth
Direct Effects
f(empgrowth)
empgrowth
Indirect Effects
f(empgrowth)
empgrowth

[2]


Total Effects
f(serv)
-4
0
4
50
60
70
80
serv
Direct Effects
f(serv)
-2
2
4
50
60
70
80
serv
Indirect Effects
f(serv)
-1.0
0.5
50
60
70
80
serv

[3]



[4]

Spatial semiparametric model with spatial trend (no ANOVA decomposition)

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$$

```r
form2 <- unrate ~ partrate + agri + cons +
                 pspl(serv,nknots=15) + pspl(empgrowth,nknots=20) +
                 pspt(long,lat,nknots=c(20,20),psanova=FALSE)
geospsar <- psar(form2,data=unemp_it,Wsp=Wsp_it,sar=TRUE)
```

```
##
## Fitting Model...
##
##  Time to fit the model:  105.56 seconds
##  Time to compute covariances:  108.98 seconds
```

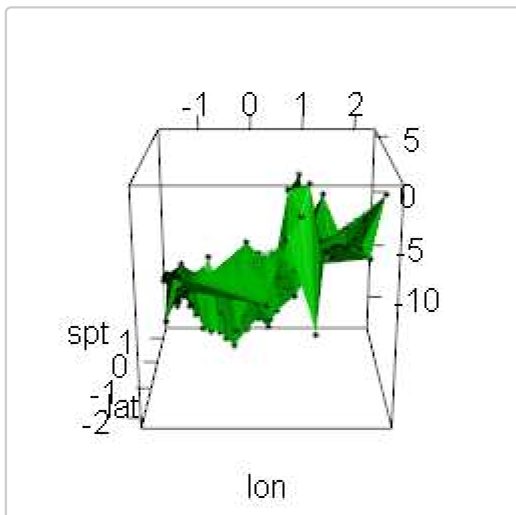
```r
summary(geospsar)
```

```
## Call:
## psar(formula = form2, data = unemp_it, Wsp = Wsp_it, sar = TRUE)
##
##  Parametric Terms
##           Estimate Std. Error  t value Pr(>|t|)
## partrate  0.377818   0.027566  13.7058   <2e-16 ***
## agri     -0.032237   0.026008  -1.2395   0.2153
## cons     -0.503668   0.044994 -11.1941   <2e-16 ***
## rho       0.527784   0.018595  28.3838   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Non-Parametric Terms
##                                  EDF
## pspl(serv, nknots = 15)      7.5737
## pspl(empgrowth, nknots = 20) 7.0116
##
##  Non-Parametric Spatio-Temporal Trend
##        EDF
## sp1 45.394
## sp2 37.841
##
##  Goodness-of-Fit
##
```

```
## EDF Total: 107.82  Sigma: 1.88408
## AIC:       5300.58 BIC:  5902.13
```

Plot of Spatial Trend

```
sptrend <- fit_terms(geospsar,"spttrend")
n <- dim(unemp_it)[1]
nT <- 19
sp_seq <- seq(from=1,to=n,by=nT)
lon <- scale(unemp_it[sp_seq,c("long")])
lat <- scale(unemp_it[sp_seq,c("lat")])
spt <- sptrend$fitted_terms[sp_seq]
require(rgl); require(akima)
akima_spt <- akima::interp(x=lon, y=lat, z=spt,
                           xo=seq(min(lon), max(lon), length = 1000),
                           yo=seq(min(lat), max(lat), length = 1000))
open3d(); plot3d(x=lon,y=lat,z=spt)
persp3d(akima_spt$x,akima_spt$y,akima_spt$z,
        xlim=range(lon),ylim=range(lat),
        zlim=range(spt),col="green3",aspect="iso",add=TRUE)
movie3d(spin3d(axis = c(0,0,1), rpm = 10), duration=6,  type = "png")
```



Spatial semiparametric model with spatial trend (ANOVA decomposition).

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$$

In this case we nest the B-Spline basis for interaction term $f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$ dividing by 2 the number of knots.

```
form3 <- unrate ~ partrate + agri + cons +
                   pspl(serv,nknots=15) + pspl(empgrowth,nknots=20) +
                   pspt(long,lat,nknots=c(18,18),psanova=TRUE,
                   nest_sp1=c(1,2),nest_sp2=c(1,2))
geospsaranova <- psar(form3,data=unemp_it,Wsp=Wsp_it,sar=TRUE)
```

```
##
## Fitting Model...
##
##  Time to fit the model:  63.52 seconds
##  Time to compute covariances:  66.22 seconds
```

```
summary(geospsaranova)
```

```
## Call:
## psar(formula = form3, data = unemp_it, Wsp = Wsp_it, sar = TRUE)
##
##  Parametric Terms
##              Estimate Std. Error t value Pr(>|t|)
## Intercept -1.8867454  9.7448780 -0.1936   0.8465
## partrate   0.3142102  0.0278713 11.2736   <2e-16 ***
## agri       0.0054901  0.0231111  0.2376   0.8123
## cons      -0.3911695  0.0447932 -8.7328   <2e-16 ***
## rho        0.5142088  0.0194390 26.4525   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Non-Parametric Terms
##                                  EDF
## pspl(serv, nknots = 15)      7.4418
## pspl(empgrowth, nknots = 20) 6.6225
##
```

```
##  Non-Parametric Spatio-Temporal Trend
##     EDF sp1 EDF sp2
## f1   10.488   0.000
## f2    0.000   9.401
## f12  21.378  22.166
##
##  Goodness-of-Fit
##
## EDF Total: 87.4986  Sigma: 1.99318
## AIC:       5446.57 BIC:  5934.74
```

Semiparametric model with spatio-temporal trend (ANOVA decomposition)

$$\mathbf{y} = f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_{\tau}) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_{\delta}(\mathbf{x}_{\delta}) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$$

$$f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_{\tau}) = f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_{\tau}) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_{\tau}) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_{\tau}) + f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_{\tau})$$

Nested B-Spline bases are used for second-order interaction terms $f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$; $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_{\tau})$; $f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_{\tau})$ (dividing by 2) and third-order interaction term $f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_{\tau})$ (dividing by 3 the number of spatial knots and by 2 the number of temporal knots).

```
form4 <- unrate ~ partrate + agri + cons +
                  pspl(serv,nknots=15) + pspl(empgrowth,nknots=20) +
                  pspt(long,lat,year,nknots=c(18,18,8),psanova=TRUE,
                  nest_sp1=c(1,2,3),nest_sp2=c(1,2,3),
                  nest_time=c(1,2,2),ntime=19)
sptanova <- psar(form4,data=unemp_it,sar=FALSE,
                 control=list(thr=1e-2,maxit=200))
```

```
##
## Fitting Model...
##
##  Time to fit the model:  34.34 seconds
##  Time to compute covariances:  0.73 seconds
```

```
summary(sptanova)
```

```
## Call:
## psar(formula = form4, data = unemp_it, sar = FALSE, control = list(thr = 0.01,
##     maxit = 200))
##
##  Parametric Terms
##            Estimate Std. Error t value Pr(>|t|)
## Intercept -1.934183   9.337761 -0.2071   0.8359
## partrate   0.308579   0.027622 11.1715   <2e-16 ***
## agri      -0.015665   0.020901 -0.7495   0.4537
## cons      -0.050344   0.042065 -1.1968   0.2315
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Non-Parametric Terms
##                                 EDF
## pspl(serv, nknots = 15)      7.8057
## pspl(empgrowth, nknots = 20) 7.9829
##
##  Non-Parametric Spatio-Temporal Trend
##      EDF sp1 EDF sp2 EDF time
## f1    11.016   0.000    0.000
## f2     0.000  10.068    0.000
## ft     0.000   0.000    7.535
## f12   20.729  23.794    0.000
## f1t    7.192   0.000    0.142
## f2t    0.000  10.404   10.964
## f12t  13.052  13.118   21.349
##
##  Goodness-of-Fit
##
## EDF Total: 178.152  Sigma: 1.46465
## AIC:       4621.31 BIC:  5615.25
```

Spatial Semiparametric model with spatio-temporal trend (ANOVA decomposition)

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$$

$$f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) = f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_\tau) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_\tau) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_\tau) + f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau)$$

```
sptsaranova <- psar(form4,data=unemp_it,sar=TRUE,Wsp=Wsp_it,
                  control=list(thr=1e-2,maxit=200))
```

```
##
## Fitting Model...
##
##  Time to fit the model:  106.36 seconds
##  Time to compute covariances:  0.75 seconds
```

```
summary(sptsaranova)
```

```
## Call:
## psar(formula = form4, data = unemp_it, Wsp = Wsp_it, sar = TRUE,
##     control = list(thr = 0.01, maxit = 200))
##
##  Parametric Terms
##            Estimate Std. Error t value Pr(>|t|)
## Intercept -1.667097   9.343482 -0.1784   0.8584
## partrate   0.308830   0.027599 11.1898   <2e-16 ***
## agri      -0.017002   0.020866 -0.8148   0.4153
## cons      -0.048900   0.042022 -1.1637   0.2447
## rho       -0.036996   0.031108 -1.1893   0.2345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Non-Parametric Terms
##                                 EDF
## pspl(serv, nknots = 15)      7.7934
## pspl(empgrowth, nknots = 20) 8.0070
##
##  Non-Parametric Spatio-Temporal Trend
```

```
##      EDF sp1 EDF sp2 EDF time
## f1    11.035   0.000    0.000
## f2     0.000  10.087    0.000
## ft     0.000   0.000    7.687
## f12   20.717  23.796    0.000
## f1t    6.967   0.000    0.140
## f2t    0.000  10.535   11.103
## f12t  13.547  13.695   22.781
##
##  Goodness-of-Fit
##
## EDF Total: 181.891  Sigma: 1.45967
## AIC:       4626.97 BIC:  5641.77
```

Spatial Semiparametric model with spatio-temporal trend (ANOVA decomposition) and AR(1) disturbances.

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma^2(\mathbf{I}_N \otimes \mathbf{R}_T)\right)$$

$$f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) = f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_\tau) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_\tau) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_\tau) + f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau)$$

```
sptsaranovaar1 <- psar(form4,data=unemp_it,sar=TRUE,ar1=TRUE,Wsp=Wsp_it,
                    control=list(thr=1e-1,maxit=200))
```

```
##
## Fitting Model...
##
##  Time to fit the model:  116.6 seconds
##  Time to compute covariances:  0.84 seconds
```

```
summary(sptsaranovaar1)
```

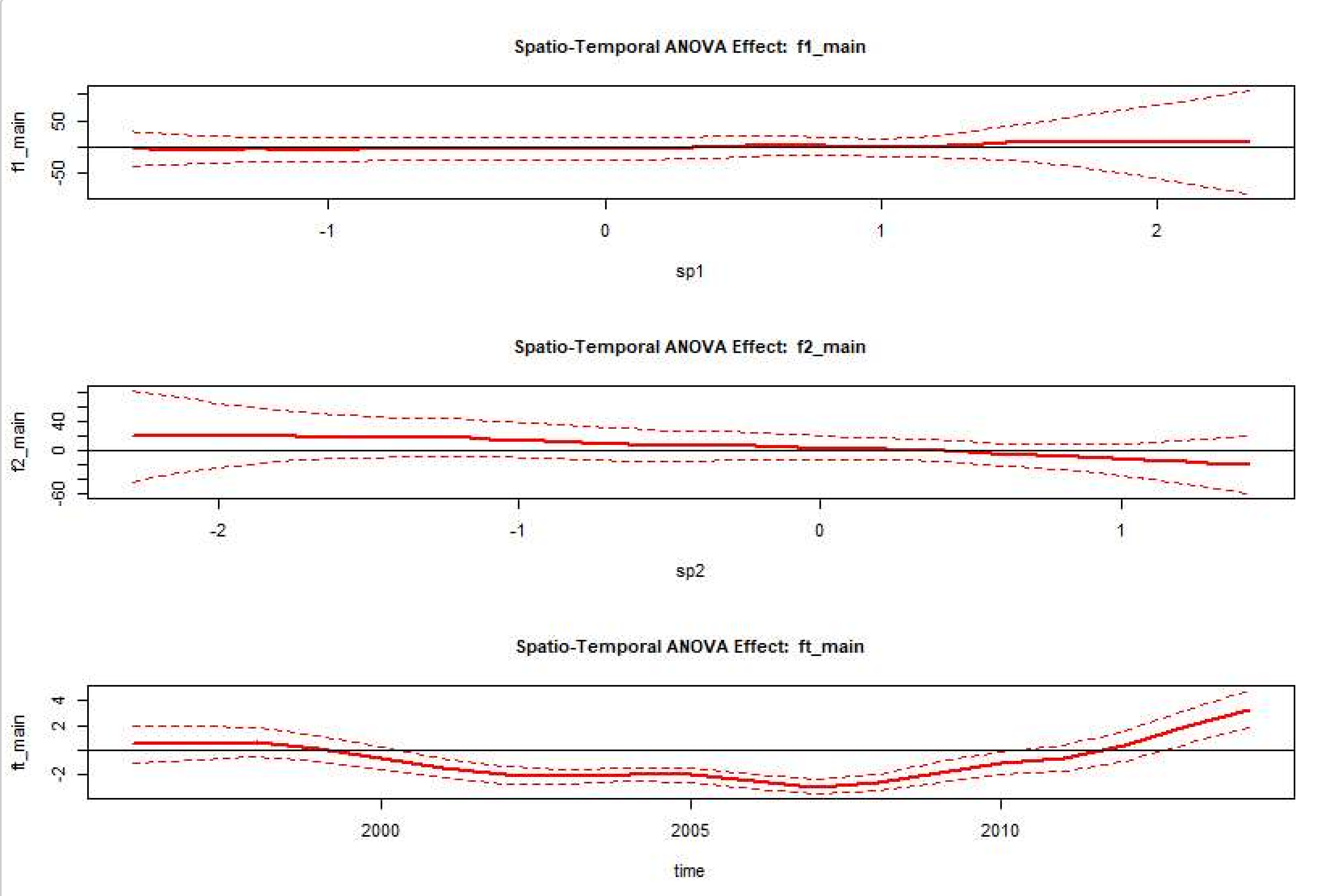
```
## Call:
## psar(formula = form4, data = unemp_it, Wsp = Wsp_it, sar = TRUE,
```

```
##     ar1 = TRUE, control = list(thr = 0.1, maxit = 200))
##
##  Parametric Terms
##             Estimate Std. Error t value Pr(>|t|)
## Intercept -11.101764   9.024544 -1.2302   0.2188
## partrate    0.596526   0.029154 20.4609   <2e-16 ***
## agri       -0.036253   0.027779 -1.3051   0.1920
## cons       -0.012260   0.045054 -0.2721   0.7856
## rho         0.043580   0.027346  1.5937   0.1112
## phi         0.683961   0.019346 35.3548   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Non-Parametric Terms
##                                 EDF
## pspl(serv, nknots = 15)      6.9776
## pspl(empgrowth, nknots = 20) 9.0430
##
##  Non-Parametric Spatio-Temporal Trend
##      EDF sp1 EDF sp2 EDF time
## f1    10.070   0.000    0.000
## f2     0.000   9.786    0.000
## ft     0.000   0.000    8.368
## f12   18.483  17.888    0.000
## f1t    0.000   0.000    0.000
## f2t    0.000   6.740    6.795
## f12t   8.085   6.464    4.859
##
##  Goodness-of-Fit
##
## EDF Total: 128.557  Sigma: 1.66939
## AIC:       3365.21 BIC:  4082.45
```

Plot of main effects in spatio-temporal trend

```
spttrend <- fit_terms(sptsaranovaar1,"spttrend")
lon <- scale(unemp_it$long); lat <- scale(unemp_it$lat)
time <- unemp_it$year
```

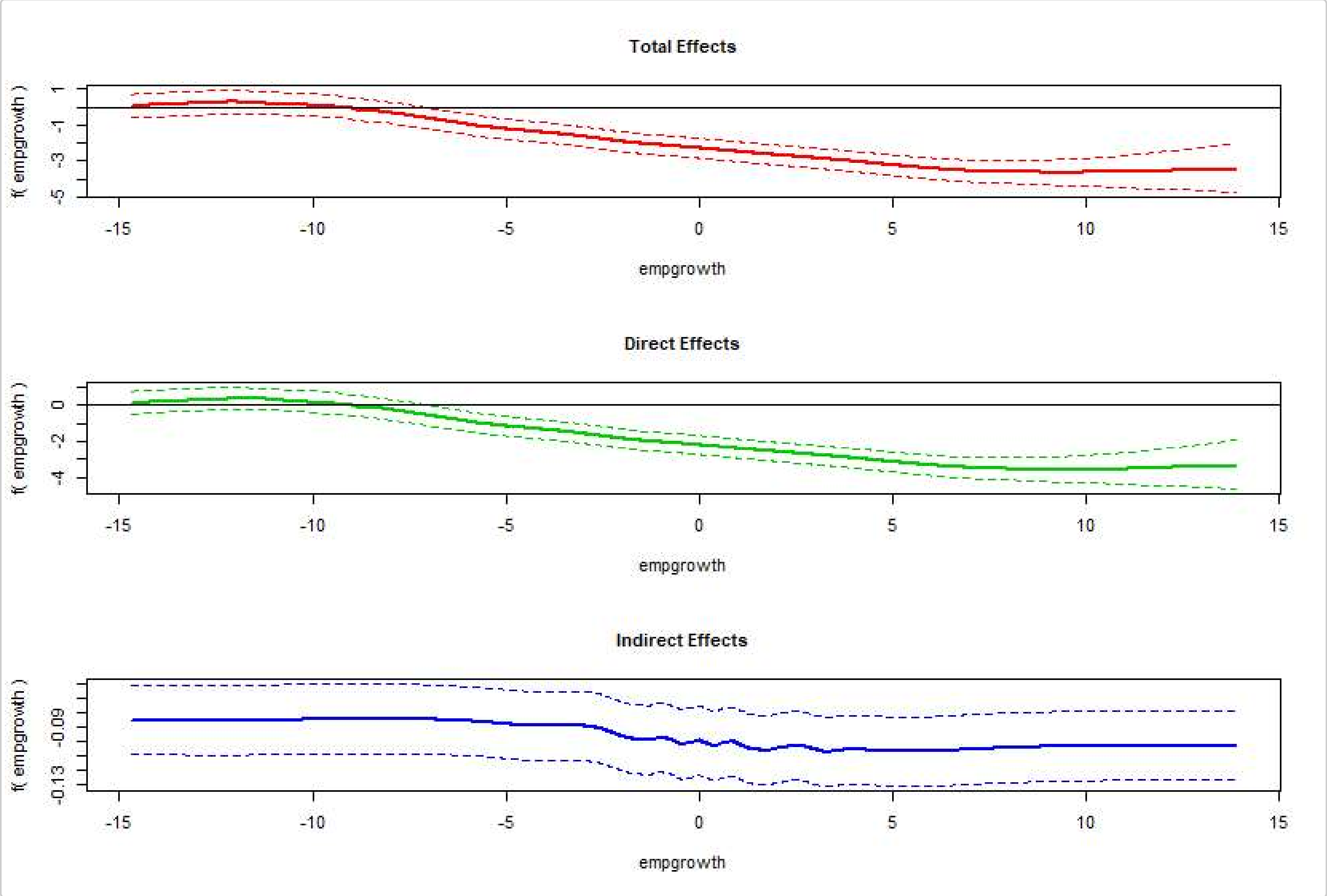
```
### Plot main effects
plot_main_spt(spttrend,sp1=lon,sp2=lat,time=time,nT=19)
```



Non-Parametric Total, Direct and Indirect Effects

```
list_varnopar <- c("serv","empgrowth")
eff_nparvar <- eff_nopar(sptsaranovaar1,list_varnopar)
```

```
plot_effects_nopar(eff_nparvar,unemp_it,smooth=TRUE)
```



Spatial Semiparametric model with restricted spatio-temporal trend (ANOVA decomposition) and AR(1) disturbances.

In this case we set to 0 the $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_\tau)$ term.

$$\mathbf{y} = (\rho \mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) + \mathbf{X}\boldsymbol{\beta} + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \sigma^2(\mathbf{I}_N \otimes \mathbf{R}_T)\right)$$

$$f(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau) = f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_\tau) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_\tau) + f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_\tau)$$

```
form5 <- unrate ~ partrate + agri + cons +
                  pspl(serv,nknots=15) + pspl(empgrowth,nknots=20) +
                  pspt(long,lat,year,nknots=c(18,18,8),psanova=TRUE,
                  nest_sp1=c(1,2,3),nest_sp2=c(1,2,3),
                  nest_time=c(1,2,2),ntime=19,f1t_int=FALSE)
sptsaranovaar1_rest <- psar(form5,data=unemp_it,sar=TRUE,ar1=TRUE,Wsp=Wsp_it,
                   control=list(thr=1e-1,maxit=200))
```

```
##
## Fitting Model...
##
##  Time to fit the model:  92.69 seconds
##  Time to compute covariances:  0.58 seconds
```

```
summary(sptsaranovaar1_rest)
```

```
## Call:
## psar(formula = form5, data = unemp_it, Wsp = Wsp_it, sar = TRUE,
##     ar1 = TRUE, control = list(thr = 0.1, maxit = 200))
##
##  Parametric Terms
##             Estimate Std. Error t value Pr(>|t|)
## Intercept -10.968690   9.128623 -1.2016  0.22969
## partrate    0.586193   0.029078 20.1595  < 2e-16 ***
## agri       -0.034349   0.027464 -1.2507  0.21121
## cons       -0.015049   0.044884 -0.3353  0.73744
## rho         0.046720   0.027391  1.7057  0.08823 .
## phi         0.667820   0.019463 34.3130  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##  Non-Parametric Terms
##                                EDF
## pspl(serv, nknots = 15)      6.9929
## pspl(empgrowth, nknots = 20) 8.9812
##
##  Non-Parametric Spatio-Temporal Trend
##      EDF sp1 EDF sp2 EDF time
## f1    10.252   0.000    0.000
## f2     0.000   9.877    0.000
## ft     0.000   0.000    8.327
## f12   18.588  18.190    0.000
## f1t    0.000   0.000    0.000
## f2t    0.000   7.551    7.082
## f12t   7.923   6.579    4.700
##
##  Goodness-of-Fit
##
## EDF Total: 129.043  Sigma: 1.66017
## AIC:       3379.44 BIC:  4099.39
```

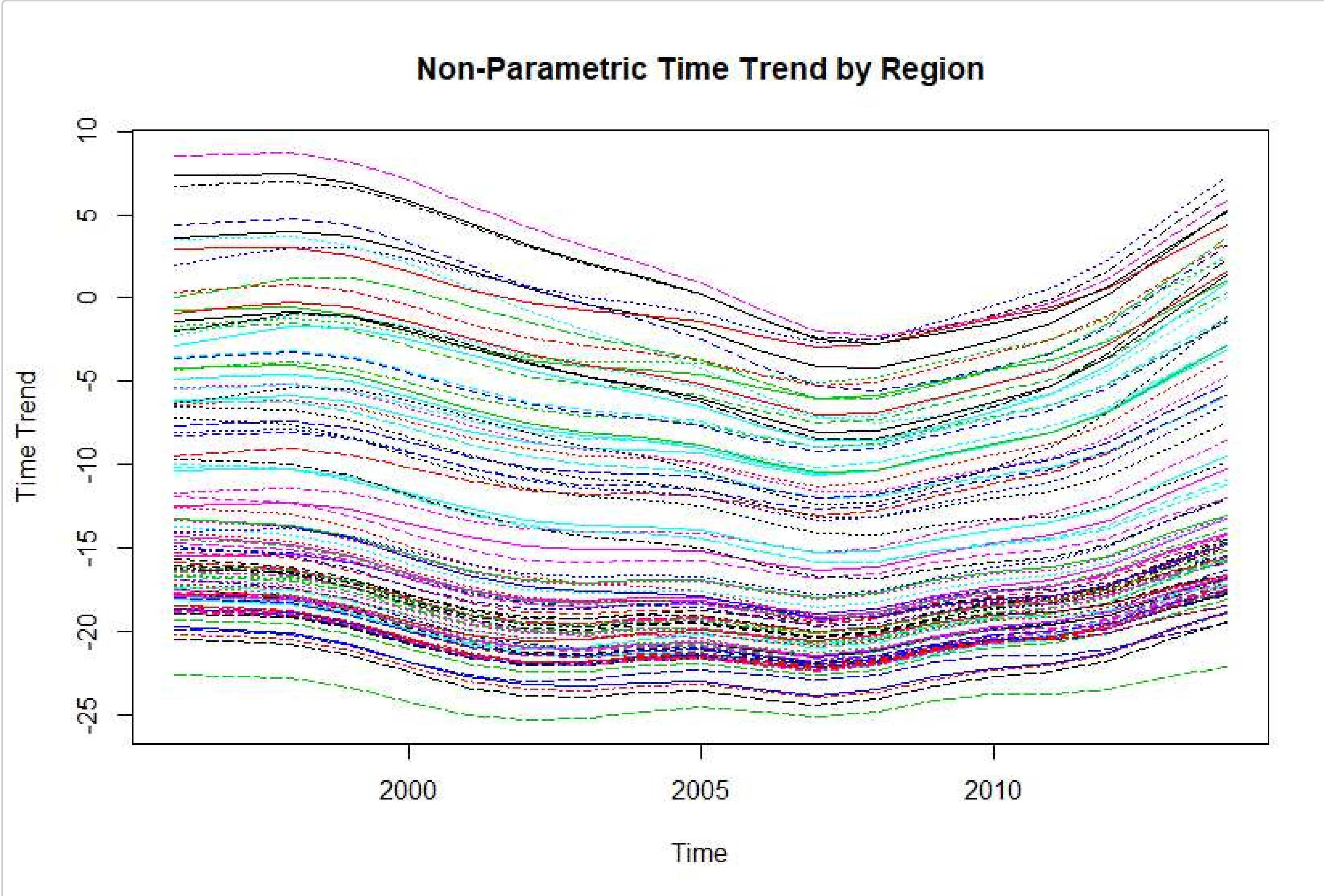
Comparison of models

```
anova(geospsaranova,sptanova,sptsaranovaar1,sptsaranovaar1_rest)
```

```
##         EDF    logLik   RlogLik      AIC      BIC   Res.Df Sigma.Sq
## 1  87.49863 -2635.785 -2636.130 5446.567 5934.737 1869.501 3.972775
## 2 178.15184 -2132.505 -2129.641 4621.313 5615.252 1778.848 2.145196
## 3 128.55741 -1554.048 -1556.118 3365.210 4082.453 1828.443 2.786877
## 4 129.04272 -1560.675 -1562.919 3379.436 4099.387 1827.957 2.756159
```

Plot of time trend by region

```
time_trend <- plot_timetrend(sptsarfit=sptsaranovaar1_rest,data=unemp_it,
                             time="year",spat="name",
                             xlab="Time",ylab="Time Trend",
                             title="Non-Parametric Time Trend by Region")
```

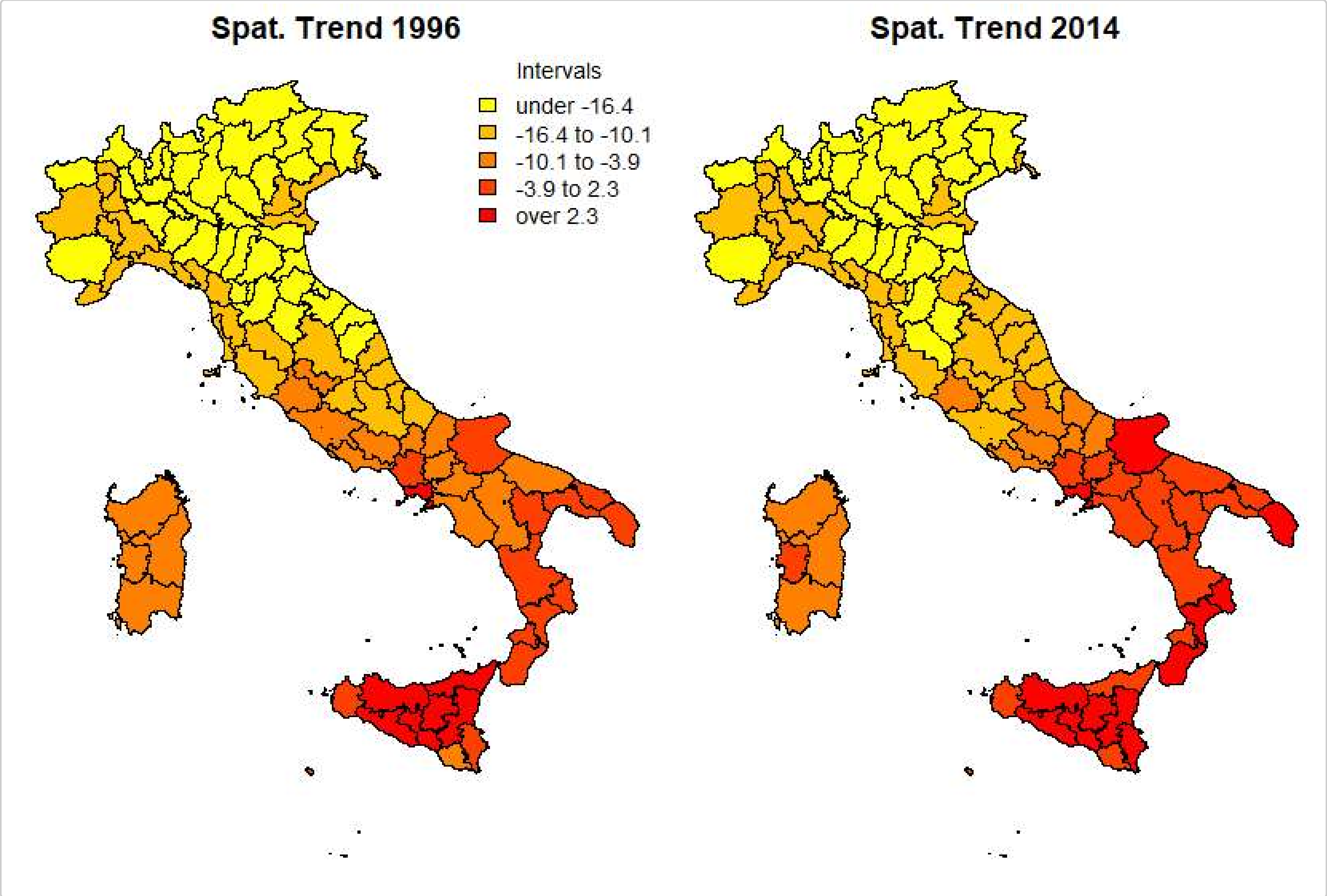


Plot of spatial trend in 1996 and 2014

```
spttrend_all <- fit_terms(sptsaranovaar1,"spttrend")
unemp_it$spttrend_anovasarar1 <- spttrend_all$fitted_terms[,c("spttrend")]
spttrend_t1 <- unemp_it$spttrend_anovasarar1[unemp_it$year=="1996"]
spttrend_tf <- unemp_it$spttrend_anovasarar1[unemp_it$year=="2014"]
nint <- 5
```

```
int_spttrend_t1 <- classIntervals(spttrend_t1,nint,"equal")
int_spttrend_tf<- classIntervals(spttrend_tf,nint,style="fixed")
library(classInt); library(maptools)
breaks_int_spttrend_t1 <- round(int_spttrend_t1$brks,1)
multicol <- colorRampPalette(c("yellow","red"))
colores <- multicol(nint)
col_t1 <- findColours(int_spttrend_t1,colores)
par(mfrow=c(1,2))
par(mar=c(0.2,0.2,1.5,0))
x  <- readShapePoly("~/Data/Prov2001_WGS84",IDvar="COD_PRO")
# Read polygon shape files into SpatialPolygonsDataFrame objects
coord <- coordinates(x)
plot(x, border="black",col=col_t1,xlim=NULL,axes=FALSE,
     main="Spat. Trend 1996")
leyenda <- leglabs(breaks_int_spttrend_t1,
                   under="under",
                   over="over",between="to")
legend('topright',legend=leyenda,fill=colores,cex=0.9,
       pt.cex=cex,title="Intervals",bty="n")

col_tf <- findColours(int_spttrend_tf,colores)
plot(x, border="black",col=col_tf,xlim=NULL,axes=FALSE,
     main="Spat. Trend 2014")
```



CD tests

```
library(plm)
```

```
## Loading required package: Formula
```

```
pdataframe <- pdata.frame(unemp_it,index=c("prov","year"))
pdataframe$resids_geospsaranova <- geospsaranova$residuals
pdataframe$resids_sptanova <- sptanova$residuals
pdataframe$resids_sptsaranova <- sptsaranova$residuals
pdataframe$resids_sptsaranovaar1 <- sptsaranovaar1$residuals
pcdtest(pdataframe$resids_geospsaranova)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  pdataframe$resids_geospsaranova
## z = 60.954, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

```
pcdtest(pdataframe$resids_sptanova)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  pdataframe$resids_sptanova
## z = 1.9339, p-value = 0.05313
## alternative hypothesis: cross-sectional dependence
```

```
pcdtest(pdataframe$resids_sptsaranova)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  pdataframe$resids_sptsaranova
## z = 1.9793, p-value = 0.04779
## alternative hypothesis: cross-sectional dependence
```

```
pcdtest(pdataframe$resids_sptsaranovaar1)
```

```
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  pdataframe$resids_sptsaranovaar1
## z = -1.584, p-value = 0.1132
## alternative hypothesis: cross-sectional dependence
```

Serial Correlation tests

```
rho.H0 <- 0
my.pwartest(pdataframe$resids_geospsaranova,N,id1,id2,rho.H0)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:
## chisq = 2002.6, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

```
my.pwartest(pdataframe$resids_sptanova,N,id1,id2,rho.H0)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:
## chisq = 482.07, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

```
my.pwartest(pdataframe$resids_sptsaranova,N,id1,id2,rho.H0)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:
```

```
## chisq = 466.17, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

```
my.pwartest(pdataframe$resids_sptsaranovaar1,N,id1,id2,rho.H0)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:
## chisq = 1849.8, p-value < 2.2e-16
## alternative hypothesis: serial correlation
```

```
pdataframe$residsnorm_sptsaranovaar1 <- sptsaranovaar1$residuals_norm
my.pwartest(pdataframe$residsnorm_sptsaranovaar1,N,id1,id2,rho.H0)
```

```
##
##  Wooldridge's test for serial correlation in FE panels
##
## data:
## chisq = 17.444, p-value = 2.959e-05
## alternative hypothesis: serial correlation
```

Conclusions & work to do

- `sptpsar` can be a useful R-package to model in a flexible way spatio-temporal and spatial panel data especially from a spatial econometrics perspective.
- Available in GitHub [https://github.com/rominsal/sptpsar/]
- There are other well-known and powerful alternatives R-packages for semiparametric models: `mgcv`,`BayesX`,`GAMLSS`,… Nevertheless they are not focused on spatial econometrics perspective.
- Working on more functionalities and estimation algorithms
  - SEM specifications.
  - Extension to generalized linear models.
  - Dynamic spatial panel data.
  - New algorithms for big datasets.

- .....

References

- Eilers, P. and Marx, B.; 1996: Flexible Smoothing with B-Splines and Penalties. *Statistical Science* 11, 89-121
- Fahrmeir, L., Kneib, Th., Lang, S., Marx, B.; 2013: *Regression: Models, Methods and Applications*, Springer
- Lee, D. and Durbán, M.; 2011: P-Spline ANOVA Type Interaction Models for Spatio-Temporal Smoothing. *Statistical Modelling* 11, 49-69.
- Lee, D. J.; Durban, M. and Eilers, P.; 2013: Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Computational Statistics & Data Analysis* 61, 22-37.
- Mínguez, R.; Basile, R. and Durbán, M., 2018: An Alternative Semiparametric Model for Spatial Panel Data. Under evaluation in *Regional Science and Urban Economics*
- Rodriguez-Alvarez, M. X.; Kneib, T.; Durban, M.; Lee, D.J.; and Eilers, P.; 2015: Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm. *Statistics and Computing* 25 (5), 941-957.
- Wood, S.N.; 2017: *Generalized Additive Models: An Introduction with R, Second Edition* CRC-Press.