Depressive symptoms and urban residential greenness: Effects of measurement errors of the mean normalised difference vegetation index (NDVI) on its association with depressive symptoms in spatial regression

Dany Djeudeu[1], Katja Ickstadt[1], Susanne Moebus[2]

[1]Fakultät Statistik, TU Dortmund
[2]IMIBE, Uniklinikum Essen

technische universität dortmund

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Heinz Nixdorf recall study

- Ongoing prospective study conducted in Bochum, Essen, and Mülheim/Ruhr
- Baseline 2000-2003 including 4814 participants between 45 and 75 years old
- Participants randomly selected from population registries
- **Individuals eligible if their address was valid**
- First (5-year) follow-up in 2006

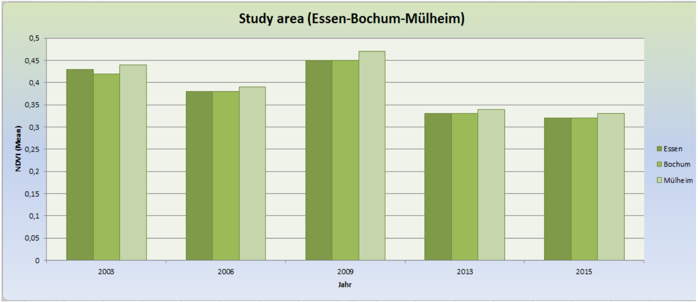
Depressive symptoms and greenness

- Depressive symptoms assessed using a 15-item short-form questionnaire of the CES-D
- Depression scores range from 0 to 45, higher score $\Rightarrow$ more depressive symptoms
- Access to green spaces may be beneficial for mental health [*Gascon et al.*]
- Greenness defined using the Normalized Difference Vegetation Index (NDVI), **calculated from satellite imagery**
- Patients without depressive symptoms at baseline
- Association depressive symptoms and greenness adjusted for a binary variable, health status

Sources of measurement error

- Error of geometry
- Atmospheric correction
- Change of measurement instruments: different satellites $\Rightarrow$ different resolutions
- The values of NDVI are changing over time (Figure 7)

Changes in NDVI values



	2003	2006	2009	2013	2015
Essen	0,43	0,38	0,45	0,33	0,32
Bochum	0,42	0,38	0,45	0,33	0,32
Mülheim	0,44	0,39	0,47	0,34	0,33

Contents

Fitting the (independent) Poisson model

- $E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z)$
- Link-funktion: g = log
- $y \rightarrow$ dep. score $x \rightarrow$ NDVI and $z \rightarrow$ health status

Table: Fitting the Poisson Model, classical (frequentist) approach

Var. name	Coef. estimate	Std. Error	Conf. intervalle	p-value
Intercept	2.4982	0.0297	[2.4401,2.5563]	<.0001
NDVI	-0.3239	0.0813	[-0.4833, -0.1645]	<.0001
Health status	0.5110	0.0158	[0.4800, 0.5420]	<.0001

Table: Fitting the Poisson Model, Bayesian approach with R-INLA

Var. name	mean	sd	0.025quant	0.5quant	0.975quant	mode
Intercept	1.9865	0.0273	1.9328	1.9865	2.040	1.9865
NDVI	-0.3200	0.0805	-0.4780	-0.3200	-0.162	-0.3200
Health status	0.5113	0.0157	0.4804	0.5113	0.542	0.5113

Contents

Classical measurement error model using INLA, [*Muff et al.*]

- Glmm with linear predictor: $E(y) = g^{-1}(\beta_0 1 + \beta_x x + \beta_z z)$
- Three level hierarchical model
- **First level**: The observational model, $y|v, \theta_1$
- v:(latent) unknown, θ_1: hyperparameters
- **Second level**: Describe latent model $v|\theta_2$, θ_2: Hyperparameters
- **Third level**: Define hyperpriors $\theta = (\theta_1, \theta_2)$

Classical measurement error model using INLA

- Exposure model:
  - $x|z \sim N(\lambda_0 + \lambda_z z, \frac{1}{\tau_x} I)$
  - $w = x + u, \quad u \sim N(0, \frac{1}{\tau_u} I)$, u independent of x and y
- Unknows: $v = (x^T, \beta_0, \beta_z, \lambda_0, \lambda_z)^T, \theta = (\beta_x, \tau_u, \tau_x)^T$
- general model:

$$\begin{aligned} E(y) &= g^{-1}(\beta_0 1 + \beta_x x + \beta_z z) \\ 0 &= -x + \lambda_0 1 + \lambda_z z + \epsilon_x, \quad \epsilon_x \sim N(0, \frac{1}{\tau_x} I) \\ w &= x + u, \quad u \sim N(0, \frac{1}{\tau_u} I) \end{aligned}$$

Classical measurement error model using INLA

- Exposure model:
  - $x|z \sim N(\lambda_0 + \lambda_z z, \frac{1}{\tau_x} I)$
  - $w = x + u, \quad u \sim N(0, \frac{1}{\tau_u} I)$, u independent of x and y
- Unknows: $v = (x^T, \beta_0, \beta_z, \lambda_0, \lambda_z)^T$, $\theta = (\beta_x, \tau_u, \tau_x)^T$
- general model:

$$\begin{aligned} E(y) &= g^{-1}(\beta_0 1 + \beta_x x + \beta_z z) \\ 0 &= -x + \lambda_0 1 + \lambda_z z + \epsilon_x, \quad \epsilon_x \sim N(0, \frac{1}{\tau_x} I) \\ w &= x + u, \quad u \sim N(0, \frac{1}{\tau_u} I) \end{aligned}$$

Classical measurement error model using INLA

- Exposure model:
  - $x|z \sim N(\lambda_0 + \lambda_z z, \frac{1}{\tau_x} I)$
  - $w = x + u, \quad u \sim N(0, \frac{1}{\tau_u} I)$, u independent of x and y
- Unknows: $v = (x^T, \beta_0, \beta_z, \lambda_0, \lambda_z)^T$, $\theta = (\beta_x, \tau_u, \tau_x)^T$
- general model:

$$
\begin{aligned}
E(y) &= g^{-1}(\beta_0 1 + \beta_x x + \beta_z z) \\
0 &= -x + \lambda_0 1 + \lambda_z z + \epsilon_x, \quad \epsilon_x \sim N(0, \frac{1}{\tau_x} I) \\
w &= x + u, \quad u \sim N(0, \frac{1}{\tau_u} I)
\end{aligned}
$$

Prior specification

- λ_0, λ_z, β_0 and $\beta_z \sim N(0, 10^2)$ priors
- w_1, w_2 independent $N(x, \frac{1}{\tau_u} I)$,
- Prior for τ_u:

$$\tau_u | w_1, w_2 \sim G\left(n, \frac{1}{2}\sum_i [(w_{i1} - \bar{w}_i)^2 + (w_{i2} - \overline{w}_i)^2]\right)$$

- $\bar{w}_i = \dfrac{w_{i1} + w_{i2}}{2}$
- prior for τ_x:

$$\tau_x \sim G(100, 99/a),$$

- a: estimated (empirical) value of τ_x

Prior specification

- λ_0, λ_z, β_0 and $\beta_z \sim N(0, 10^2)$ priors
- w_1, w_2 independent $N(x, \frac{1}{\tau_u} I)$,
- Prior for τ_u:

$$\tau_u | w_1, w_2 \sim G\left(n, \frac{1}{2} \sum_i [(w_{i1} - \bar{w}_i)^2 + (w_{i2} - \overline{w}_i)^2]\right)$$

- $\bar{w}_i = \frac{w_{i1} + w_{i2}}{2}$
- prior for τ_x:

$$\tau_x \sim G(100, 99/a),$$

- a: estimated (empirical) value of τ_x

Prior specification

- λ_0, λ_z, β_0 and $\beta_z \sim N(0, 10^2)$ priors
- w_1, w_2 independent $N(x, \frac{1}{\tau_u} I)$,
- Prior for τ_u:

$$\tau_u | w_1, w_2 \sim G\left(n, \frac{1}{2} \sum_i [(w_{i1} - \bar{w}_i)^2 + (w_{i2} - \overline{w}_i)^2]\right)$$

- $\bar{w}_i = \dfrac{w_{i1} + w_{i2}}{2}$
- prior for τ_x:

$$\tau_x \sim G(100, 99/a),$$

- a: estimated (empirical) value of τ_x

Prior specification

- λ_0, λ_z, β_0 and $\beta_z \sim N(0, 10^2)$ priors
- w_1, w_2 independent $N(x, \frac{1}{\tau_u} I)$,
- Prior for τ_u:

$$\tau_u | w_1, w_2 \sim G\left(n, \frac{1}{2} \sum_i [(w_{i1} - \bar{w}_i)^2 + (w_{i2} - \overline{w}_i)^2]\right)$$

- $\bar{w}_i = \dfrac{w_{i1} + w_{i2}}{2}$
- prior for τ_x:

$$\tau_x \sim G(100, 99/a),$$

- a: estimated (empirical) value of τ_x

Prior specification

- λ_0, λ_z, β_0 and $\beta_z \sim N(0, 10^2)$ priors
- w_1, w_2 independent $N(x, \frac{1}{\tau_u} I)$,
- Prior for τ_u:

$$\tau_u | w_1, w_2 \sim G\left(n, \frac{1}{2} \sum_i [(w_{i1} - \bar{w}_i)^2 + (w_{i2} - \overline{w}_i)^2]\right)$$

- $\bar{w}_i = \dfrac{w_{i1} + w_{i2}}{2}$
- prior for τ_x:

$$\tau_x \sim G(100, 99/a),$$

- a: estimated (empirical) value of τ_x

Prior specification

- λ_0, λ_z, β_0 and $\beta_z \sim N(0, 10^2)$ priors
- w_1, w_2 independent $N(x, \frac{1}{\tau_u} I)$,
- Prior for τ_u:

$$\tau_u | w_1, w_2 \sim G\left(n, \frac{1}{2} \sum_i [(w_{i1} - \bar{w}_i)^2 + (w_{i2} - \overline{w}_i)^2]\right)$$

- $\bar{w}_i = \dfrac{w_{i1} + w_{i2}}{2}$
- prior for τ_x:

$$\tau_x \sim G(100, 99/a),$$

- a: estimated (empirical) value of τ_x

Fitting model with R-INLA

Table: Fixed effects

Var. name	mean	sd	0.025quant	0.5quant	0.975quant	mode
beta.0	2.0390	0.0371	1.9720	2.0340	2.1132	2.0328
beta.z	0.5111	0.0157	0.4802	0.5111	0.5419	0.5112
alpha.0	0.3442	0.0015	0.3412	0.3442	0.3473	0.3442
alpha.z	-0.0027	0.0040	-0.0106	-0.0027	0.0052	-0.0027

Table: Hyperparameters

Var. name	mean	sd	0.025quant	0.5quant	0.975quant	mode
Prec. G. obs.[2]	167.6429	4.4046	159.1197	167.5977	176.4156	167.5036
Prec. G. obs.[3]	776.9589	11.1075	755.2085	776.9278	798.8515	776.8979
Beta for beta.x	-0.4698	0.1036	-0.6736	-0.4698	-0.2667	-0.4696

Contents

Characterize spatial autocorrelation using GAM, classical

- Semi-parametric Model

$$E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z + s(S_1, S_2))$$

- (S_1, S_2): Physical location (coordinates)
- s: smoothing spline
- Approach: find h that minimizes

$$\sum_{i=1}^{n}(y_i - h(x_i))^2 + \lambda \int h''(t)dt$$

- λ: nonnegative tuning parameter, h: smooth function, y: observed data

Significant residual spatial variation

Table: Parametric coefficients:

Var. name	Coef. estimate	Std. Error	z value	p-value
Intercept	1.96285	0.02898	67.725	$< 2e-16$
NDVI	-0.25395	0.08564	-2.965	0.00302
Health status	0.50500	0.01579	31.975	$< 2e-16$

Table: Significance of smooth terms

Var. name	edf	Ref.df	Chi.sq	p-value
$s(S_1, S_2)$	25.66	28.32	130.1	3.5e-15

$s(S_1, S_2)$ significant $\Rightarrow$ possible spatial autocorrelation

Modeling the spatial effect in the residual

- $E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z)$, g: log-link
- $Var(y) = R$
- spatial dependency can be modeled in R
- Choice of spatial effect: semivariogram analysis of the residuals of the independent model
- Gaussian semivariogramm model:

$$\gamma(h) = c_n + \sigma_0^2(1 - exp(-\frac{|h|^2}{a_0^2}))$$

Modeling the spatial effect in the residual

- $E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z)$, g: log-link
- $Var(y) = R$
- spatial dependency can be modeled in R
- Choice of spatial effect: semivariogram analysis of the residuals of the independent model
- Gaussian semivariogramm model:

$$\gamma(h) = c_n + \sigma_0^2(1 - exp(-\frac{|h|^2}{a_0^2}))$$

Modeling the spatial effect in the residual

- $E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z)$, g: log-link
- $Var(y) = R$
- spatial dependency can be modeled in R
- Choice of spatial effect: semivariogram analysis of the residuals of the independent model
- Gaussian semivariogramm model:

$$\gamma(h) = c_n + \sigma_0^2(1 - exp(-\frac{|h|^2}{a_0^2}))$$

Modeling the spatial effect in the residual

- $E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z)$, g: log-link
- $Var(y) = R$
- spatial dependency can be modeled in R
- Choice of spatial effect: semivariogram analysis of the residuals of the independent model
- Gaussian semivariogramm model:

$$\gamma(h) = c_n + \sigma_0^2(1 - exp(-\frac{|h|^2}{a_0^2}))$$

Modeling the spatial effect in the residual

- $E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z)$, g: log-link
- $Var(y) = R$
- spatial dependency can be modeled in R
- Choice of spatial effect: semivariogram analysis of the residuals of the independent model
- Gaussian semivariogramm model:

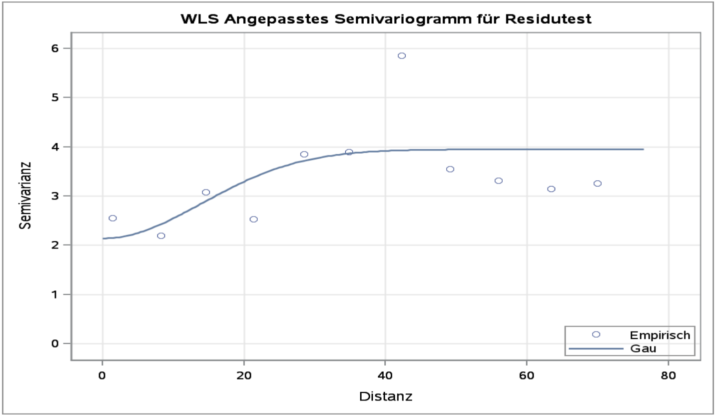
$$\gamma(h) = c_n + \sigma_0^2(1 - exp(-\frac{|h|^2}{a_0^2}))$$

Semivariogramm analysis

1

Prozedur VARIOGRAM

Abhängige Variable:Residutest



Accounting for the spatial dependency, Gaussian semivariogramm model

Table: Effect of spatial autocorrelation

Var. name	Coef. estimate	Std. Error	p-value
Intercept	2.4982	0.05736	<.0001
NDVI	-0.3239	0.1573	< 0.0395
Health status	0.5110	0.03059	<.0001

- Coefficient estimates are identical to the case of independence assumption
- Standard errors of the model accounting for spatial autocorrelation are larger
- The p-value associated to the NDVI is larger when spatial variation considered.

Contents

The simple linear model

Table: The simple linear model

$$\log(y+1) = \beta_0 + \beta_1 x + \epsilon$$

Var. name	Coef. estimate	Std. Error	t-values	p-value
Intercept	2.00062	0.05372	37.244	<2e-16
NDVI	-0.37758	0.15904	-2.374	0.0176

$$\log(y+1) = \beta_0 + \beta_1 x + \tilde{s}(S_2) + \epsilon$$

Var. name	Coef. estimate	Std. Error	t-values	p-value
Intercept	2.00218	0.05375	37.249	<2e-16
NDVI	-0.38235	0.15917	-2.402	0.0164

Table: Significance of smooth terms

Smooth term	edf	Ref.df	F	p-value
$\tilde{s}(S_2)$	3.446	4.329	3.737	0.00417

$\tilde{s}(S_2)$ significant $\Rightarrow$ possible spatial autocorrelation

Semiparametric approach, [*Huque et al.*]

- x_i = true covariate of interest, $y_i = \beta_0 + \beta_1 x_i + G_1(S_i) + \epsilon_i$
  - $\epsilon = \epsilon_i \sim N(0, \sigma_\epsilon^2)$, $\{G_1(S_i), S_i \in \mathbb{R}^2\}$ unknow and captures the spatial correlation
  - $\epsilon_i, G_1(S_i)$ independent of each other and of the x_i
- Error model: w observed, $w_i = x_i + u_i$, $u_i \sim N(0, \sigma_u^2)$

Semiparametric approach, [*Huque et al.*]

- x_i = true covariate of interest, $y_i = \beta_0 + \beta_1 x_i + G_1(S_i) + \epsilon_i$
  - $\epsilon = \epsilon_i \sim N(0, \sigma_\epsilon^2)$, $\{G_1(S_i), S_i \in \mathbb{R}^2\}$ unknow and captures the spatial correlation
  - $\epsilon_i, G_1(S_i)$ independent of each other and of the x_i
- Error model: w observed, $w_i = x_i + u_i$, $u_i \sim N(0, \sigma_u^2)$

Semiparametric approach, [Md Hamidul Huque, 2016]

- Assume the true covariate x is smooth, modeled by $G_2(S_i)$
- $\{G_j(S_i) = B_j^T(S_i)\theta_j\}$, $B_j(S_i)$ thin splines basis functions
- (Y, W) fitted to two sets of spline basis functions $B_1(S_i)$, $B_2(S_i)$ by penalized least square
- Minimizing the sum of squares plus roughness penalties
- Consistent estimate of β_1
- Estimate standard errors for β_1

Semiparametric approach, [Md Hamidul Huque, 2016]

- Assume the true covariate x is smooth, modeled by $G_2(S_i)$
- $\{G_j(S_i) = B_j^T(S_i)\theta_j\}$, $B_j(S_i)$ thin splines basis functions
- (Y, W) fitted to two sets of spline basis functions $B_1(S_i)$, $B_2(S_i)$ by penalized least square
- Minimizing the sum of squares plus roughness penalties
- Consistent estimate of β_1
- Estimate standard errors for β_1

Semiparametric approach, [Md Hamidul Huque, 2016]

- Assume the true covariate x is smooth, modeled by $G_2(S_i)$
- $\{G_j(S_i) = B_j^T(S_i)\theta_j\}$, $B_j(S_i)$ thin splines basis functions
- (Y, W) fitted to two sets of spline basis functions $B_1(S_i)$, $B_2(S_i)$ by penalized least square
- Minimizing the sum of squares plus roughness penalties
- Consistent estimate of β_1
- Estimate standard errors for β_1

Semiparametric approach, [Md Hamidul Huque, 2016]

- Assume the true covariate x is smooth, modeled by $G_2(S_i)$
- $\{G_j(S_i) = B_j^T(S_i)\theta_j\}$, $B_j(S_i)$ thin splines basis functions
- (Y, W) fitted to two sets of spline basis functions $B_1(S_i)$, $B_2(S_i)$ by penalized least square
- Minimizing the sum of squares plus roughness penalties
- Consistent estimate of β_1
- Estimate standard errors for β_1

Semiparametric approach, [Md Hamidul Huque, 2016]

- Assume the true covariate x is smooth, modeled by $G_2(S_i)$
- $\{G_j(S_i) = B_j^T(S_i)\theta_j\}$, $B_j(S_i)$ thin splines basis functions
- (Y, W) fitted to two sets of spline basis functions $B_1(S_i)$, $B_2(S_i)$ by penalized least square
- Minimizing the sum of squares plus roughness penalties
- Consistent estimate of β_1
- Estimate standard errors for β_1

Semiparametric approach, [Md Hamidul Huque, 2016]

- Assume the true covariate x is smooth, modeled by $G_2(S_i)$
- $\{G_j(S_i) = B_j^T(S_i)\theta_j\}$, $B_j(S_i)$ thin splines basis functions
- (Y, W) fitted to two sets of spline basis functions $B_1(S_i)$, $B_2(S_i)$ by penalized least square
- Minimizing the sum of squares plus roughness penalties
- Consistent estimate of β_1
- Estimate standard errors for β_1

Joint effect of measurement error and spatial variation

Table: Comparison table

Var. name	Coef. Estimate	St. Error
NDVI with ME and spatial variation	-0.5456091	0.3803308
NDVI naive simple linear model	-0.37758	0.15904

- Coefficient estimate larger in presence of measurement error and spatial variation
- The standard error of the model adjusted for measurement error and spatial variation is larger:

Contents

Summary tables

Table: Summary table

$E(y) = g^{-1}(\beta_0 + \beta_1 x + \beta_2 z), \quad g$: log-link

	Ind. assumption	Spatial autocorrelation
Estimate (no ME)	-3239	-0.3239
Std. error(no ME)	0.0813	0.1573
p-value(no ME)	<0.0001	0.0395
mean(INLA) with ME	-0.4698	
Std. Error(INLA) with ME	0.1036	

$log(y + 1) = \beta_0 + \beta_1 x + \epsilon$

	Ind. assumption	Spatial autocorrelation
Estimate (no ME)	-0.37758	
Std. error(no ME)	0.15904	
p-value(no ME)	0.0176	
Estimate(Huque et al.) with ME		-0.5456
Std. Error(Huque et al.) with ME		0.3803

Perspectives

- Examine the join effect of spatial dependency and measurement errors of covariates in the glmm
- Introduce these spatial dependency in the hierarchical model for measurement error analyss in R-INLA (Bayesian)
- The Idea of Huque in the case of generalized linear mixed model (frequentist)

Literatur I

Md Hamidul Huque, Howard D. Bondell, R. J. C. L. M. R. (2016).
Spatial regression with covariate measurement error: A semi-parametric approach.
Biometrics.

Mireia Gascon, Margarita Triguero-Mas, D. M. P. D. J. F. A. P. and Nieuwenhuijsen, M. J. (2015).
Mental health benefits of long-term exposure to residential green and blue spaces: A systematic review.
Int. J. Environ. Res. Public Health, 12:43544379.

Stefanie Muff, Andrea Riebler, H. R. P. S. L. H. (2014).
Bayesian analysis of measurement error models using inla.
Journal of the Royal Statistical Society, Series C (Appl. Statist.) 64, Part 2:231 – –252.