Spatial Regression

Mark Green

DASC507 – Advanced Biostatistics II

Analysis Methods for Complex Data Structures

Outline

- Linear regression – a recap
- Spatial regression
  - Spatial lag
  - Spatial error
  - Approaches that combine both
- Criticisms
- Case study

Linear regression – a recap

Why do we use regression?

- To try and explain the world (note: correlation ≠ causation)
- To predict an outcome of interest

Linear regression – a recap

Outcome of interest

Coefficient applied to X

$$y = \alpha + \beta X + \mathrm{e}$$

Constant

Explanatory variables

Error term

Linear regression – a recap

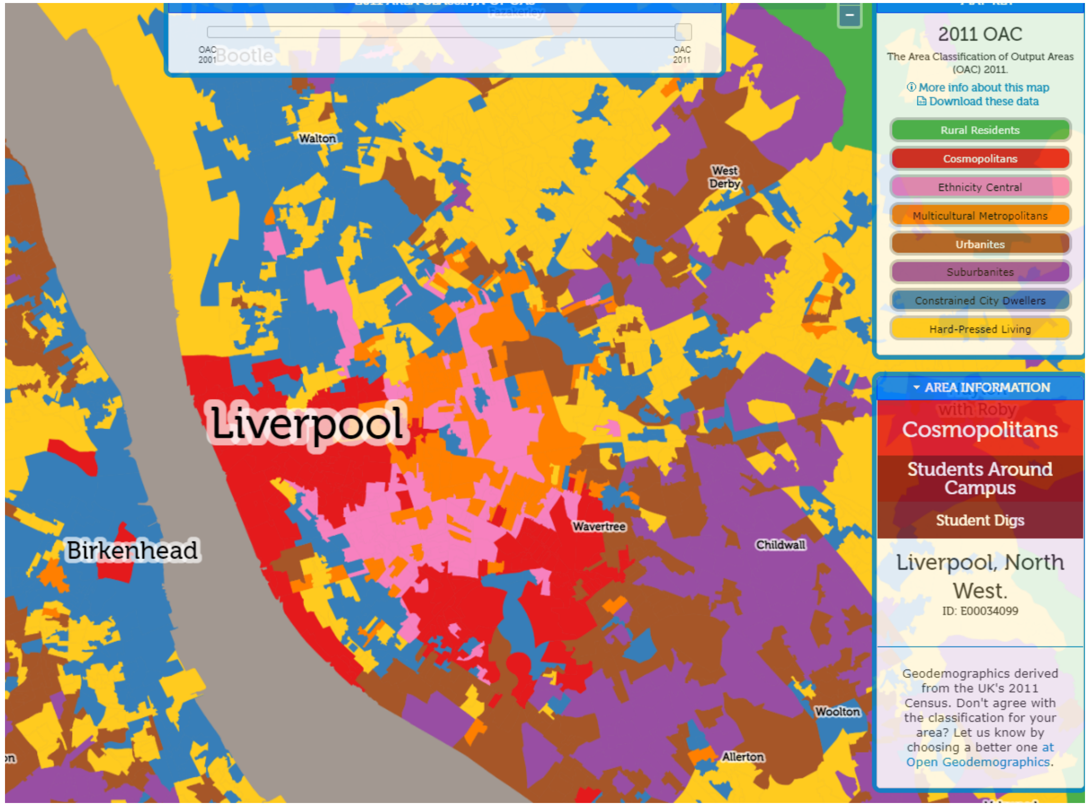
Assumptions of the approach

- Relationships are linear
- Explanatory variables are not multicollinear
- Errors are normally distributed
- Homoscedasticity of errors
- Independence of errors…

Where we have spatial data, independence of errors may be violated…

Linear regression – a recap

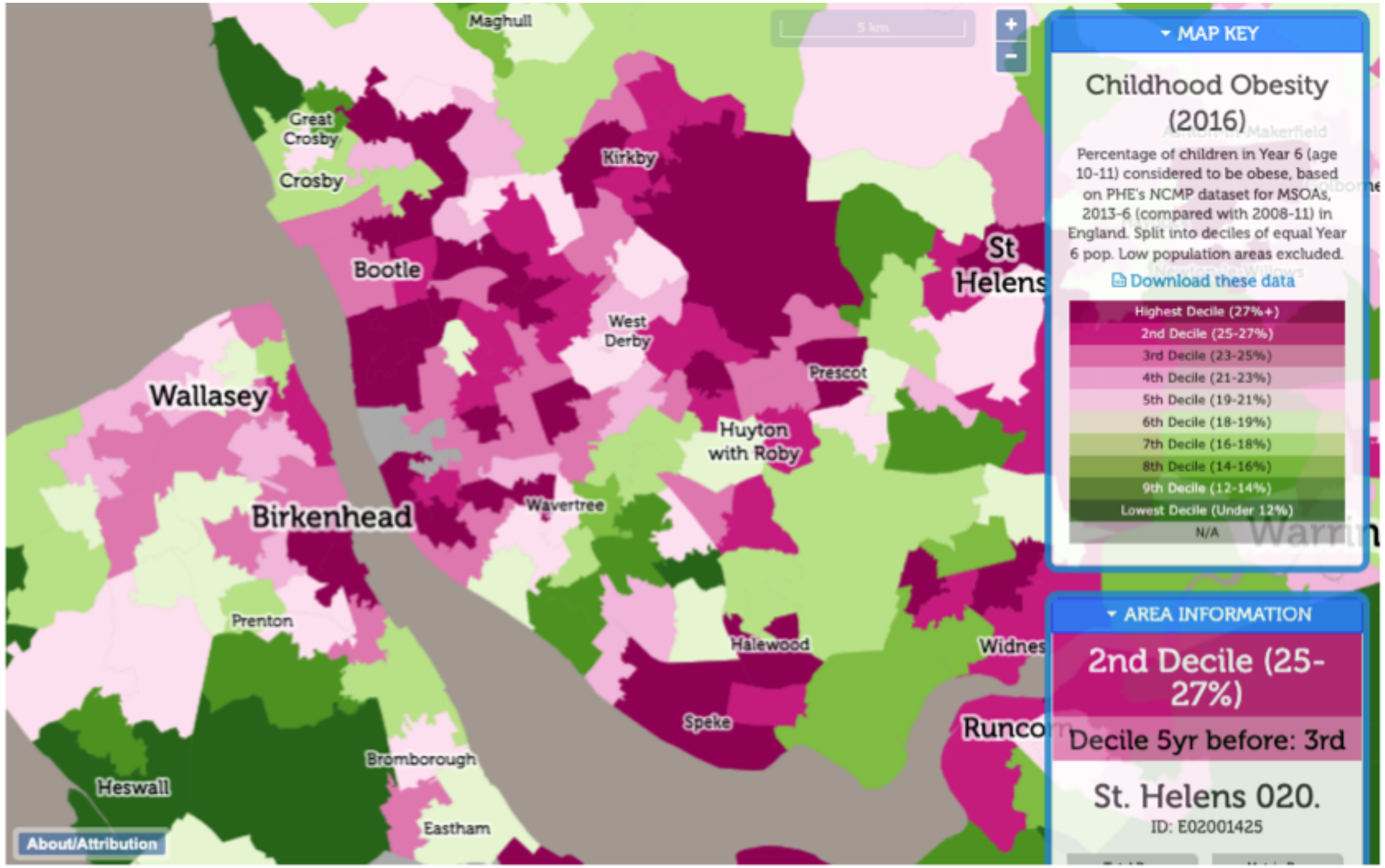
- Tobler's first law of geography "everything is related to everything else, but near things are more related than distant things"
- "Birds of a feather, flock together"
- If areas closer together are similar in characteristics, then we no-longer have independence of observations (and likely error terms)
- May reflect unobserved (or not measured) factors



Output Area Classification via maps.cdrc.ac.uk

Linear regression – a recap

Applies to health outcomes where spatial patterns are not random



Childhood obesity via maps.cdrc.ac.uk

Spatial regression

There are a lot of regression approaches that account for space, today we will focus on

- Spatial lag regression approaches
  - Spatially Lagged X-variables model
  - Spatial AutoRegressive model
- Spatial error regression approaches
  - Spatial Error Model
- Approaches that combine both spatial lag and spatial error models
  - Spatial Durbin Error

Spatial regression – spatial lag

SLX Model - Spatially Lagged X-variables:

$$y = \beta X + WX\theta + e$$

Here we extend the OLS equation by adding in $WX\theta$ where W is the spatial weight, X is our explanatory variable and θ is the spatially lagged coefficient applied to X.

It implies that changes in X in an area have a spillover effect on surrounding areas (depending on how W is defined).

Spatial regression – spatial lag

SAR Model - Spatial AutoRegressive model:

$$y = \rho W y + \beta X + e$$

Here we extend the OLS equation by adding in $\rho W y$ where *W* is the spatial weight and ρ is the spatially lagged correlation to our outcome variable *y.*

Since *y* is now on both sides of the equation, we violate the exogeneity assumption of OLS.

Spatial regression – spatial lag

Interpreting spatially lagged coefficients requires caution

- Direct effects – the immediate association (or effect) of an explanatory variable on an outcome
- Indirect effects represent how variables in surrounding areas influence an outcome
- Direct effect + indirect effect = total effect -> need to consider both effects to get the full picture of how processes act
- To get standard errors and p-values, need to use simulations rather than estimate directly (unlike SLX model)

Spatial regression – spatial error

SEM - Spatial Error Model:

$$y = \beta X + u$$

$$u = \lambda W u + e$$

We extend the OLS equation by changing our error term. Here, we define u as our spatial autocorrelation of the error termed $\lambda W u$ where λ is a coefficient controlling W our spatial weight and u *represents the* error terms. We still have e as our general error term in the model.

We violate our OLS assumption of independence of error terms.

Spatial regression – spatial lag and error

Spatial Durbin Model:

$$y = \beta X + WX\theta + u$$

$$u = \lambda Wu + e$$

An extension of SEM by adding in $WX\theta$ to explain y. Here, $WX\theta$ is the spatial lag of x-variables through applying the spatial weight W, adjusted by coefficient θ, to the independent variables X.

Can estimate direct and indirect effects.

Spatial regression

How do you select a model?

- Model fit statistics
- Test for spatial dependence in data and nature of this (always check for clustering of model residuals)
- Hausman test
- Conceptual reasons for spatial effects (e.g., lagged or spillover effects)

Criticisms

- All models are for linear regression, binary/count models harder to estimate using generalized versions of these approaches (Bayesian approaches more helpful here)
- Be careful in interpreting spillover effects – association based

Case study

Dearden et al. 2020. Exploring the histories of health and deprivation in Britain, 1971–2011. *Health & Place* **61**: 102255. https://doi.org/10.1016/j.healthplace.2019.102255

- Large geographical inequalities in health across Great Britain
- Neighourhood deprivation is a fundamental cause of inequalities
- Ecological analyses can help profile the changing nature of inequalities, including who and where is affected
- Uses Census data (Great Britain) converted to 1x1km^2 grids
- Outcome variable is self-reported Limiting Long-Term Illness (%)

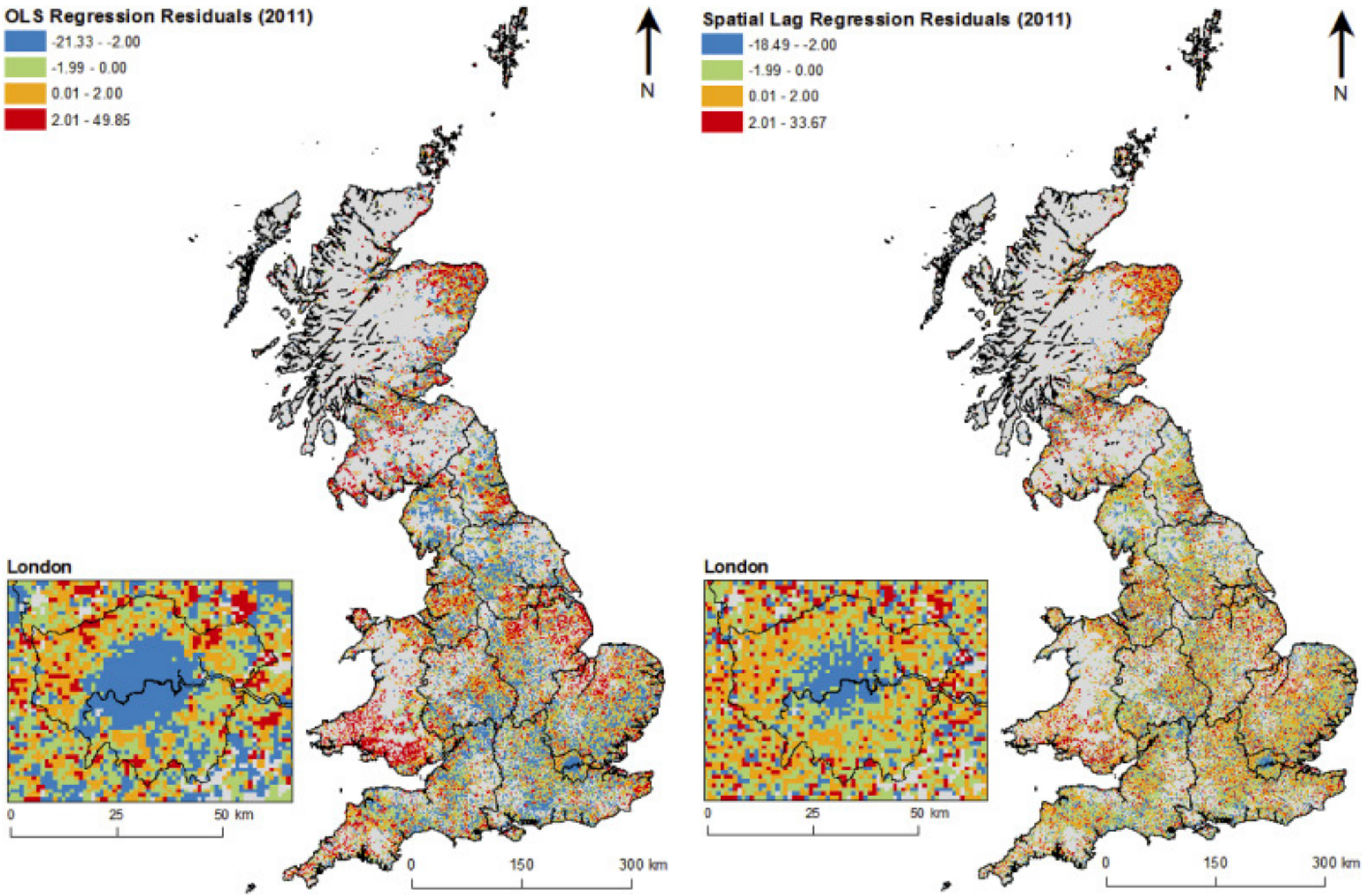
Case study

	1991		2001		2011	
	OLS	**SLQC**	**OLS**	**SLQC**	**OLS**	**SLQC**
Regression Coefficients						
Constant	9.83*	2.58*	15.74*	3.28*	15.78*	3.22*
Townsend Score	0.57*	0.30*	0.77*	0.39*	0.88*	0.47*
Aged 65 and over	0.25*	0.13*	0.45*	0.27*	0.43*	0.27*
Aged 0-14	−0.11*	−0.07*	−0.18*	−0.08*	−0.18*	−0.08*
Not UK Born	−1.57*	−0.46*	−2.44*	0.83*	−2.77*	−1.00*
Not White	0.68*	0.02*	0.54*	0.13*	0.25*	0.05*
ρ		0.73*		0.69*		0.66*
Model Fit						
R2	0.45	0.78	0.52	0.80	0.55	0.77
AIC	617,187	519,471	698,151	605,212	685,644	603,440
log-likelihood	−308,588	−259,728	−349,070	−302,599	−342,816	−301,713

Significant spatial patterns observed and attenuate the associations for coefficients observed in the OLS model

Case study

Comparing model residuals shows how the spatial model helps to account for the unexplained patterns in the OLS model



Further reading

- Anselin L. 2005. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics* **27**(3): 247-267. https://doi.org/10.1111/j.1574-0862.2002.tb00120.x
- Anselin L. 2008. Spatial Regression. pp 255-275. In Fotheringham and Rogerson (Eds.) The SAGE Handbook of Spatial Analysis. SAGE: London.
- Arbia G. 2014. A Primer for Spatial Econometrics. Palgrave: London.
- Rey S, et al. 2021. Spatial Regression. In *Geographic Data Science with Python*. https://geographicdata.science/book/notebooks/11_regression.html