Using geographically weighted regression to explore spatial variation in survey behaviour

Kaisa Lahtinen

Chris Brunsdon

Sarah Butt

Aims

- Discovering any evidence of spatial variation in response behaviour in social surveys
- Exploring whether geographically weighed regression can be used to discover these relationships further

Background to nonresponse research

- Decline in the response rates
- Nonresponse a joint outcome of individual and neighbourhood characteristics
- Nonresponse rates have been found to vary geographically
- But little existing research into whether the drivers of nonresponse would also vary spatially

Data

- Using European Social Survey (ESS)
    - Nationally representative
    - Response rate 53%
    - N = 4520

Data

- Using European Social Survey (ESS)
  - Nationally representative
  - Response rate 53%
  - N = 4520
- A set of theory driven control variables
  - ESS interviewer observations: living in a flat
  - Census 2011 data at LSOA level: population density, ethnic heterogeneity, rate of owner occupancy, proportion unemployed, proportion of 16 to 24 year olds
  - Recorded crime figures on violent crime at LA level

Methods

- Evidence for any spatial variation
  - Logistic regression
  - Introducing geographical constrains step by step
    - Model 1 – Baseline (global) model assuming no geographic effect
    - Model 2 – Global model including regional (NUTS1) dummies i.e. allowing for geographic variation but independent of other predictors.
    - Model 3 – Global model including interactions between region and other predictors.

Methods

- Evidence for any spatial variation
  - Logistic regression
  - Introducing geographical constrains step by step
    - Model 1 – Baseline (global) model assuming no geographic effect
    - Model 2 – Global model including regional (NUTS1) dummies i.e. allowing for geographic variation but independent of other predictors.
    - Model 3 – Global model including interactions between region and other predictors.
- Regional regression models

Methods

- Evidence for any spatial variation
  - Logistic regression
  - Introducing geographical constrains step by step
    - Model 1 – Baseline (global) model assuming no geographic effect
    - Model 2 – Global model including regional (NUTS1) dummies i.e. allowing for geographic variation but independent of other predictors.
    - Model 3 – Global model including interactions between region and other predictors.
- Regional regression models
- Geographically weighted regression (GWR)
  - Bivariate models
  - Adaptive bandwidth (bisquare kernel)

Methods: geographically weighted regression

- GWR is using ‘a moving window’ technique when calculating regression estimates

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 X_{1it} + \cdots \beta_k X_{kit}$$

- Adaptive bandwidth
- Bisquare weighting function

Results: controlling for regional variation

- Including regional dummies did not improve the model fit

- Including regional interaction terms made a very significant improvement to the fit of the model

Results: regional regression models

- Varying fit of regional regression models
- Not just difference in magnitudes but also on the direction of predictors
- Evidence for spatial variation in the predictors

Results: Geographically weighted regression

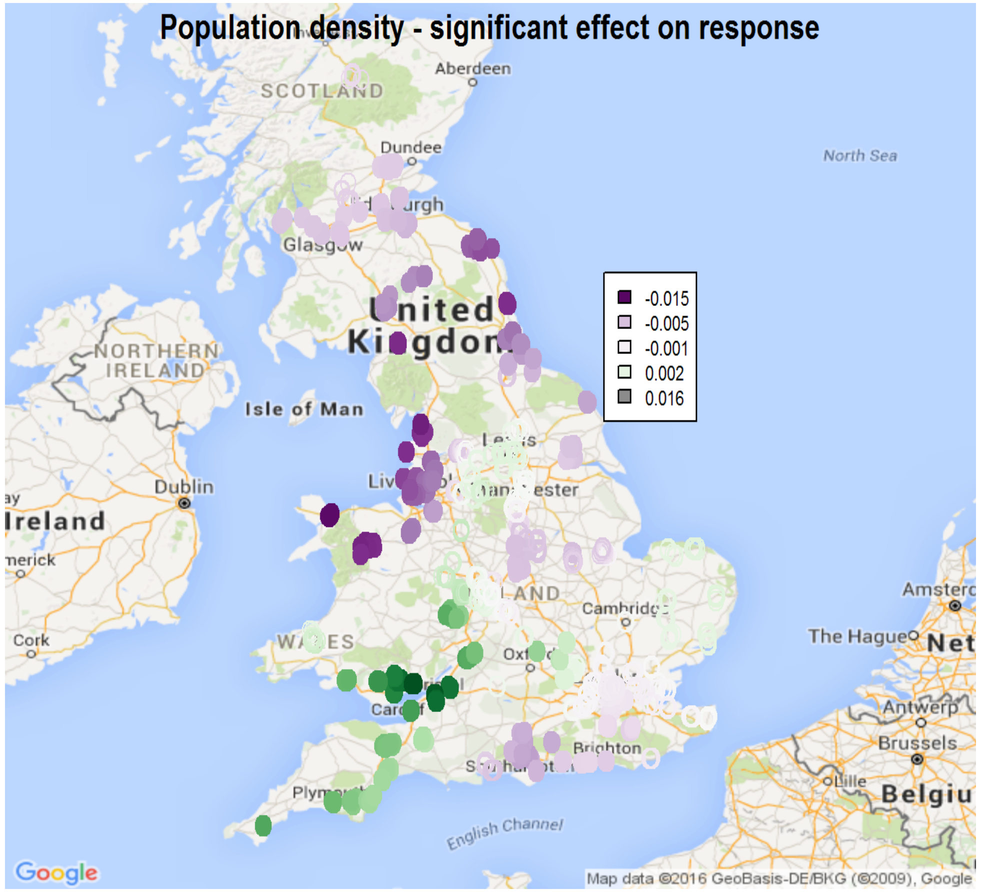
- Non uniform relationship between predictors and likelihood to participate in a survey

Table 1 Summary of GWR coefficients

Number of data points= 4146, adaptive bandwidth= 821, pseudo R square= 0.059

Variable	**Minimum**	**1st quartile**	**Median**	**3rd quartile**	**Maximum**
Population density	-0.0148	-0.0046	-0.0011	0.0018	0.0158
Fractionalisation index	-2.4460	-1.207	-0.4815	0.3716	2.454
% owner occupation	-3.932	-0.6548	-0.0704	0.3391	2.239
Living in flat	-1.159	-0.7448	-0.4928	-0.3697	0.3937
Violent crime	-0.0043	-0.001	-0.0001	0.0006	0.002
% unemployed	-16.28	-5.578	-0.906	4.424	16.180
16 to 24 year olds	-7.113	-2.213	-0.896	0.087	8.283

Results: GWR for population density



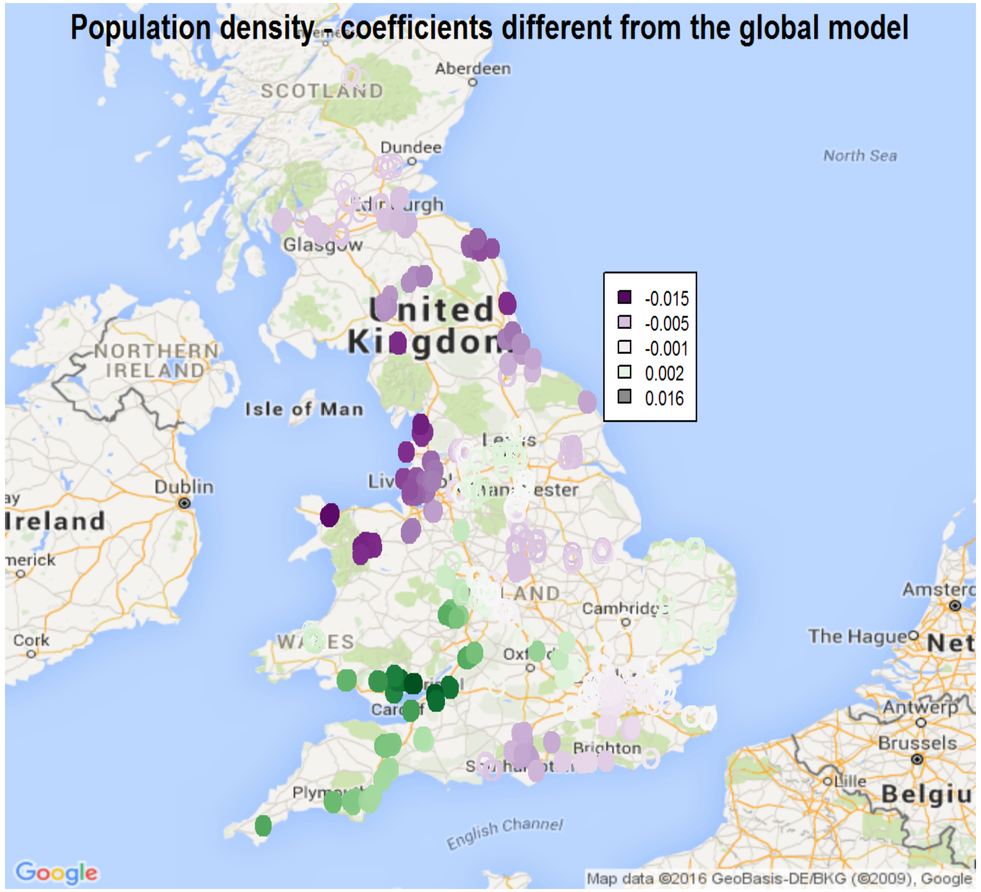
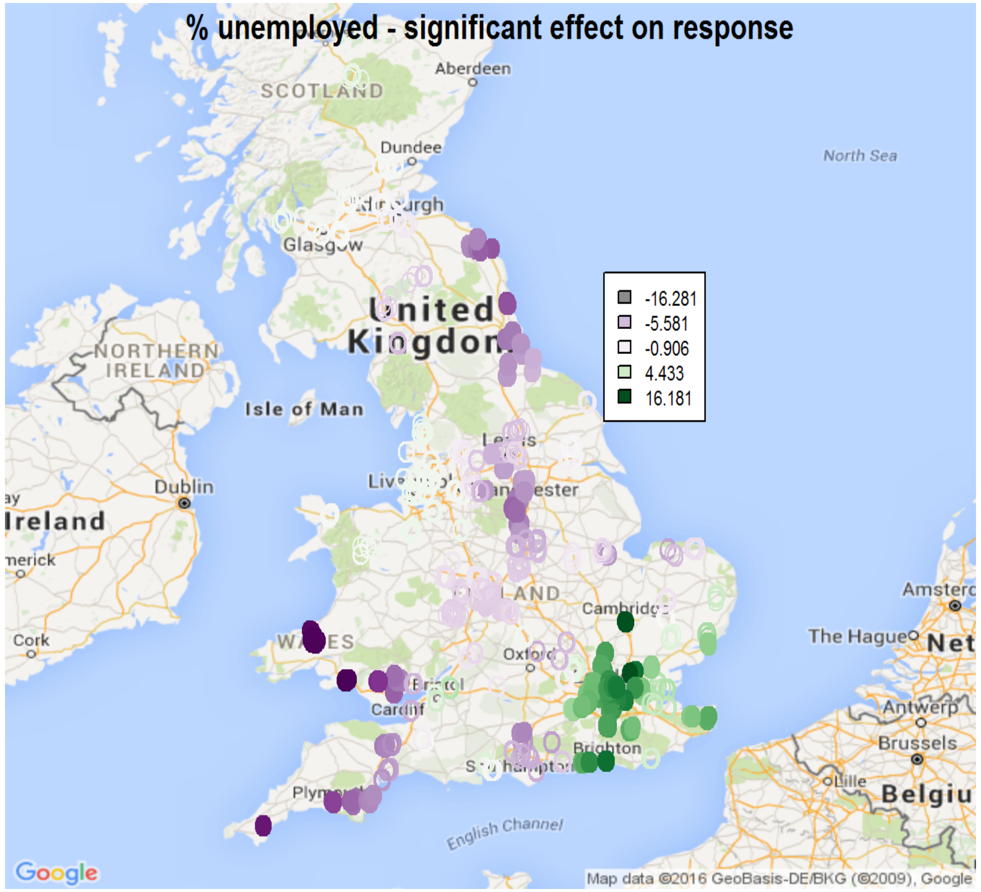


Figure 1 and 2 On the left the coefficients for population density estimated with GWR, on the right difference from global estimate

Results: GWR for unemployment



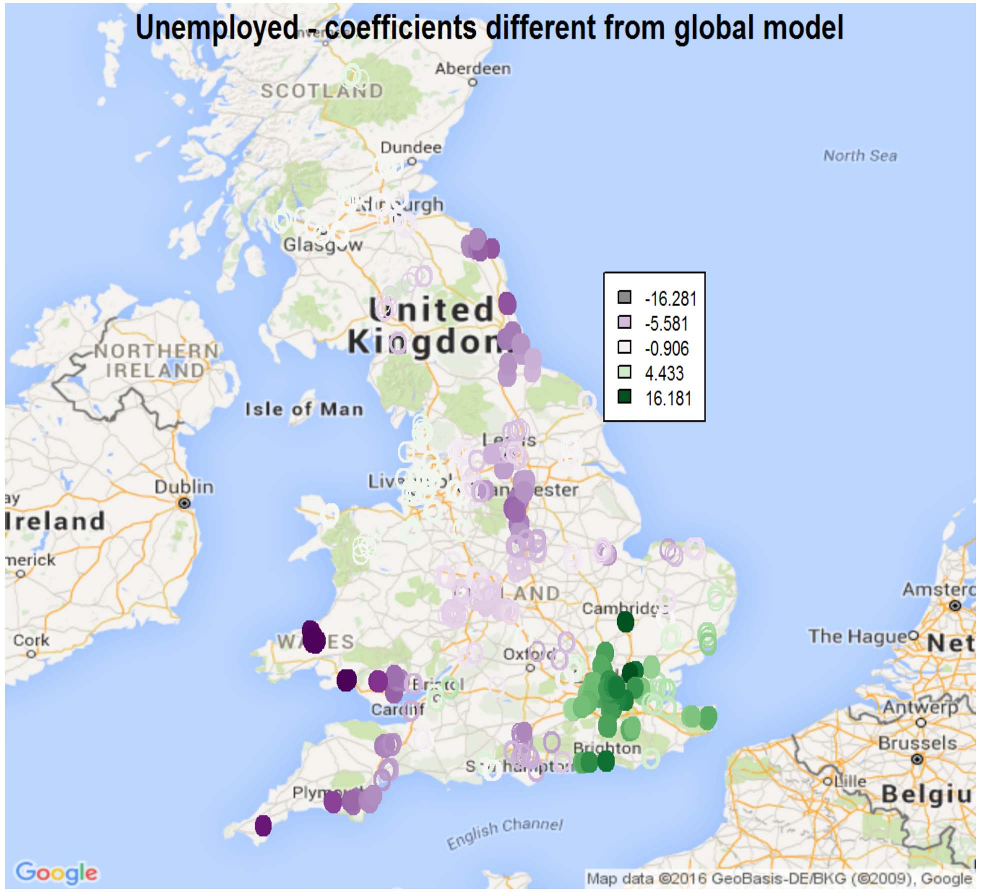
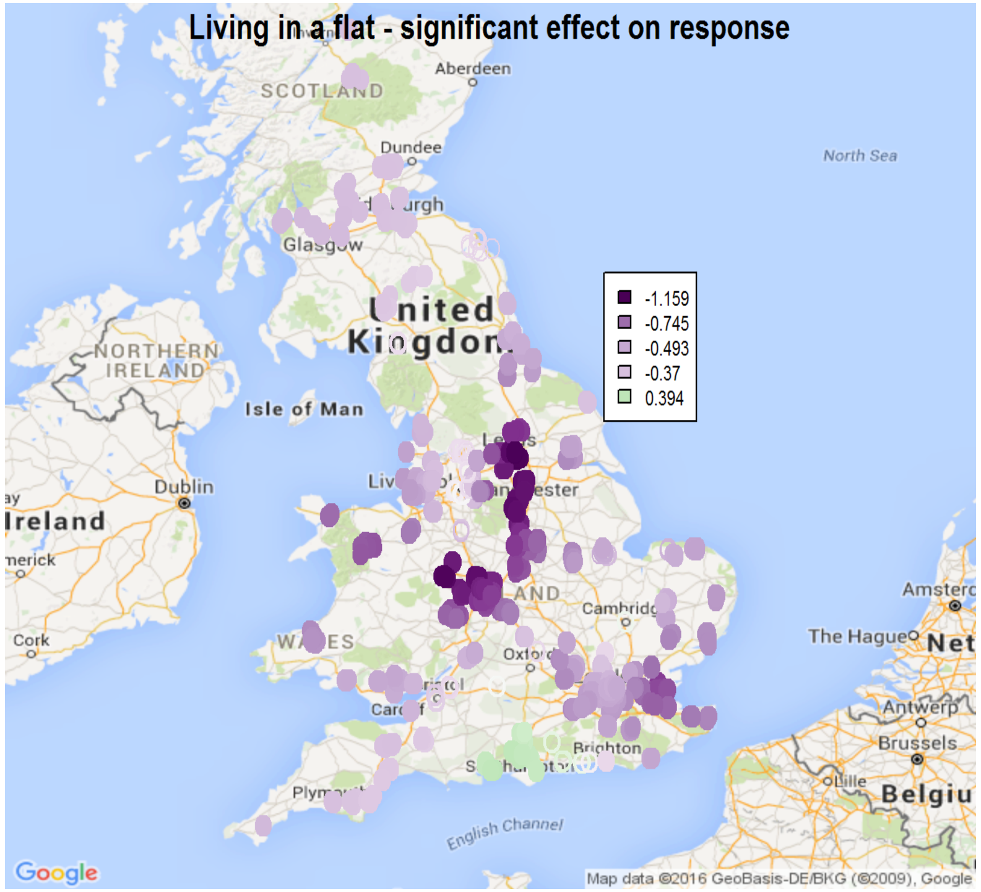


Figure 3 and 4 On the left the coefficients for unemployment estimated with GWR, on the right difference from global estimate

Results: GWR for Living in a flat



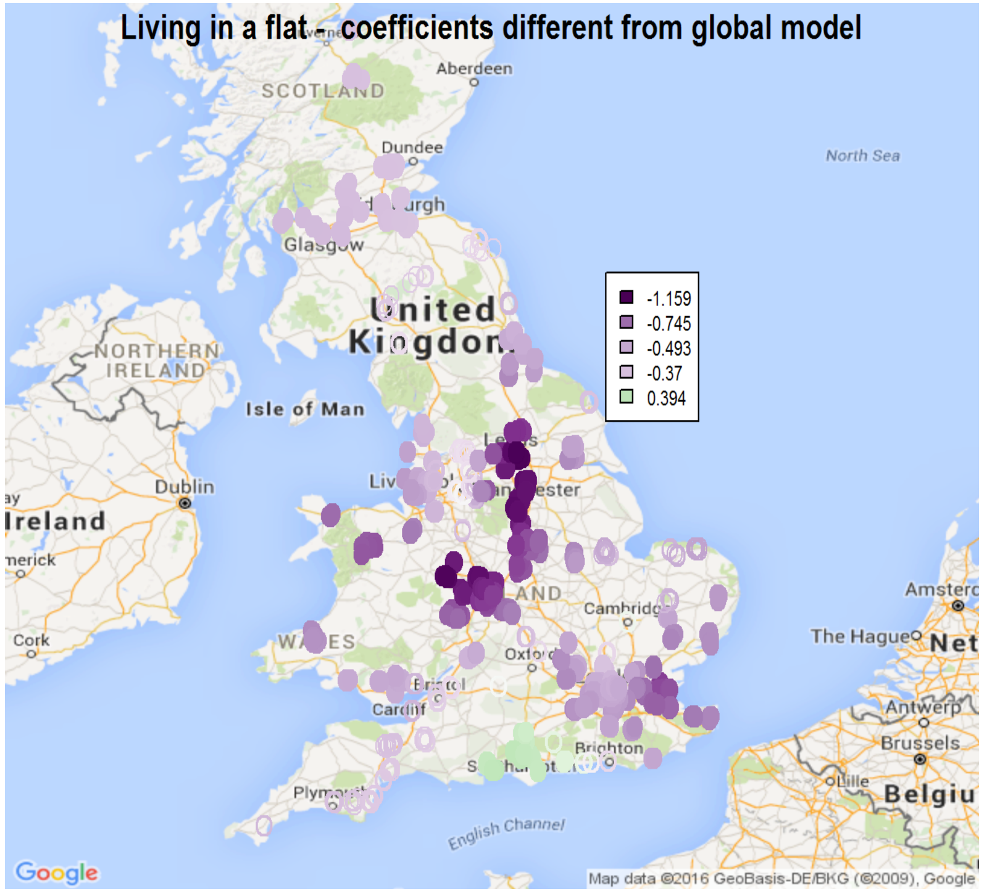


Figure 5 and 6 On the left the coefficients for living in a flat estimated with GWR, on the right difference from global estimate

Conclusion

- There is a spatial dimension in survey participation behaviour
- Using regional or other administrative boundaries does improve our model, however these still fail to capture the full picture
- GWR results help us to understand these dynamics further

Caveats

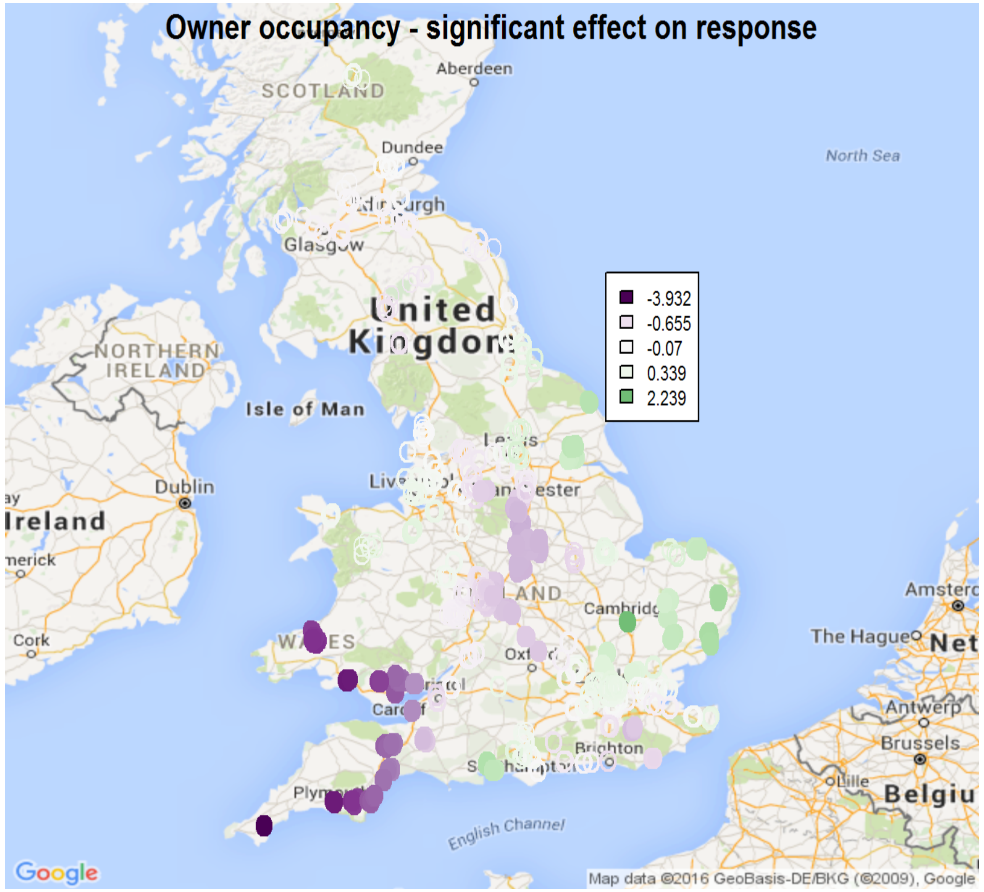
- Relatively small sample size
- Time of data collection
- Limited model of nonresponse

Thanks! Questions?

Kaisa.Lahtinen@Liverpool.ac.uk

Acknowledgements: This work is part of the 'Auxiliary Data Driven Nonresponse Bias Analysis (ADDReponse)' project funded by the UK Economic and Social Research Council (grant ES/L013118/1).

Results: GWR for owner occupancy



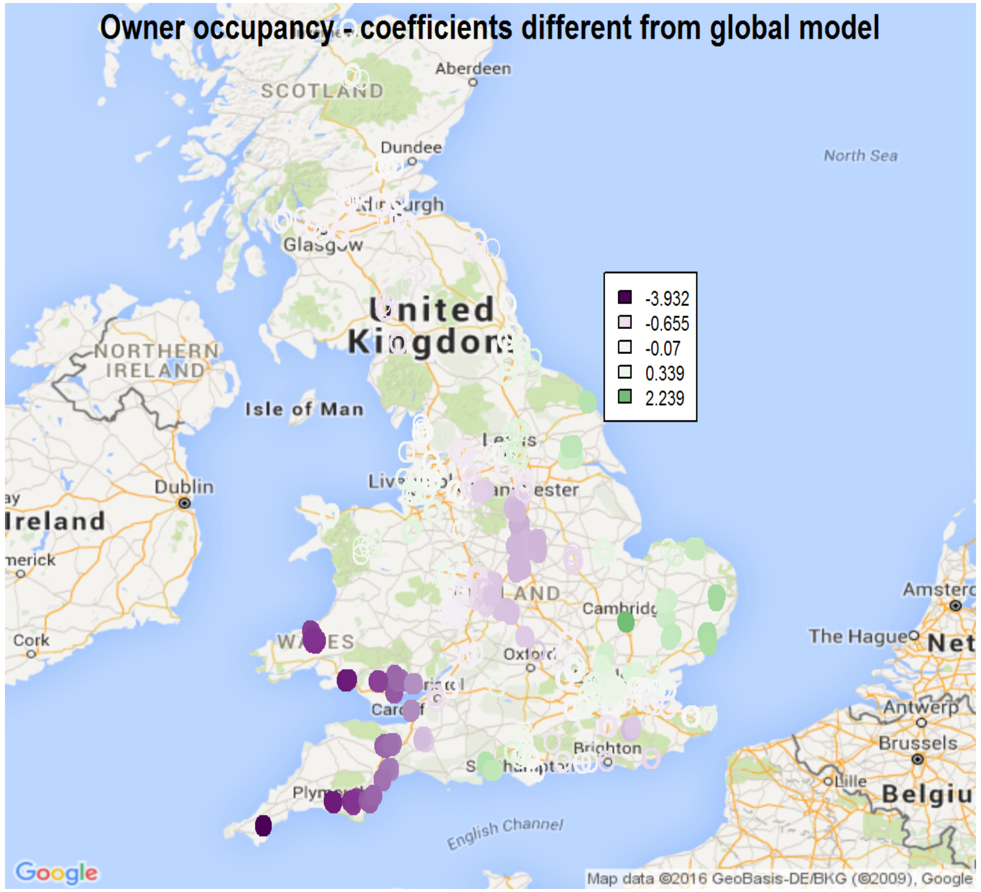


Figure 7 and 8 On the left the coefficients for owner occupancy estimated with GWR, on the right difference from global estimate

Results: GWR for fractionalisation index

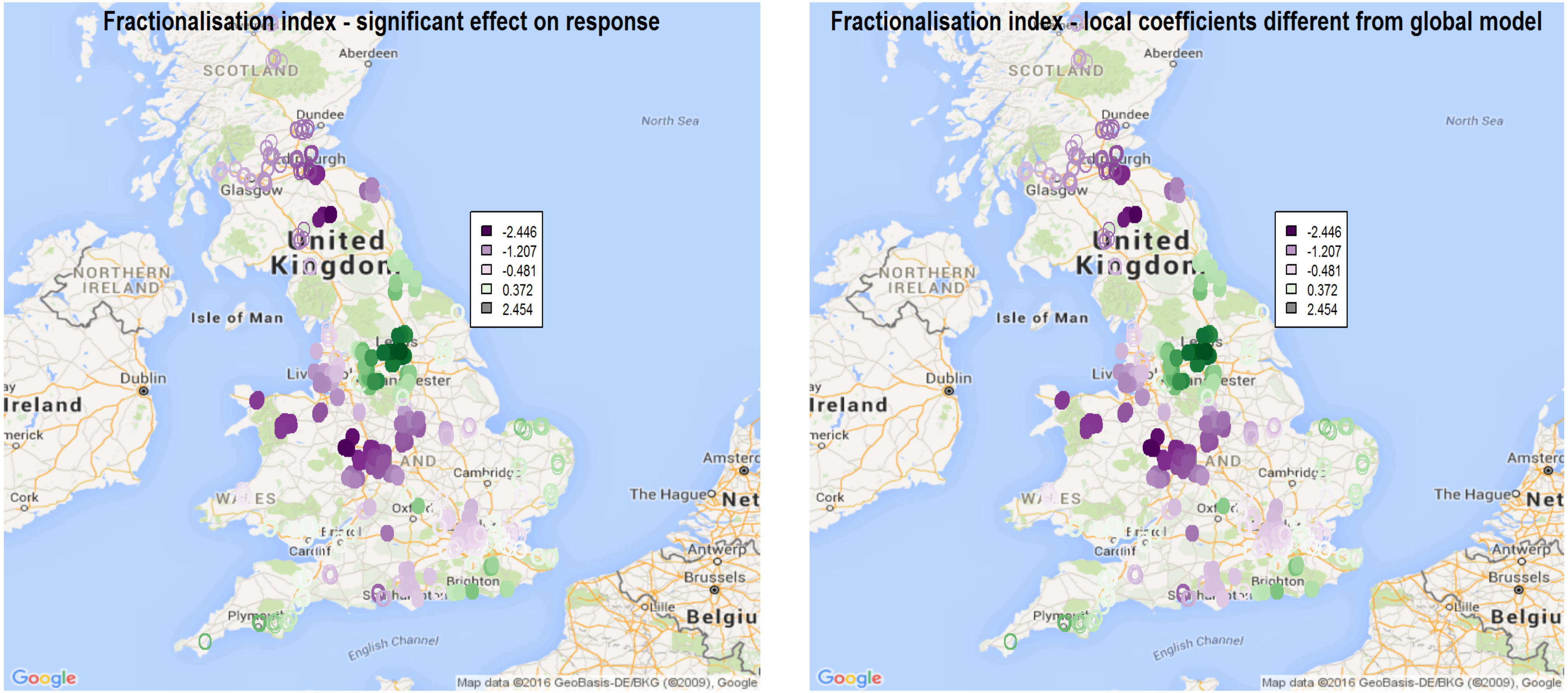
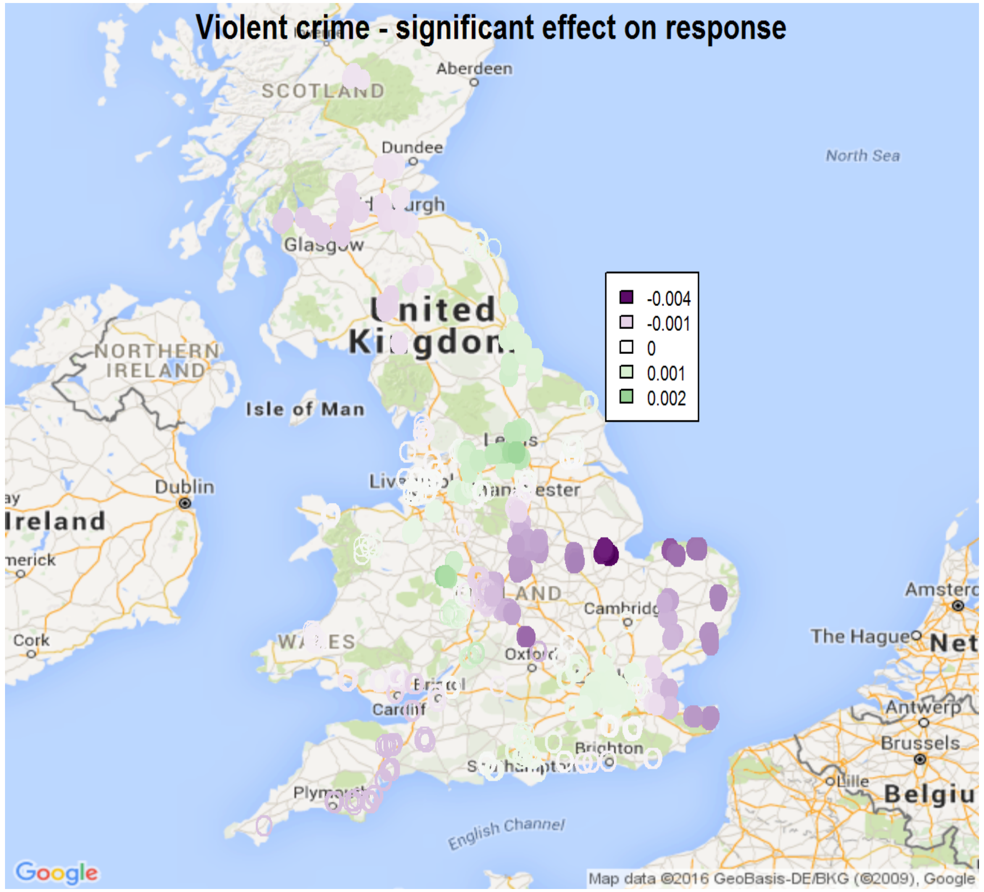


Figure 9 and 10 On the left the coefficients for ethnic fractionalisation index estimated with GWR, on the right difference from global estimate

Results: GWR for violent crime



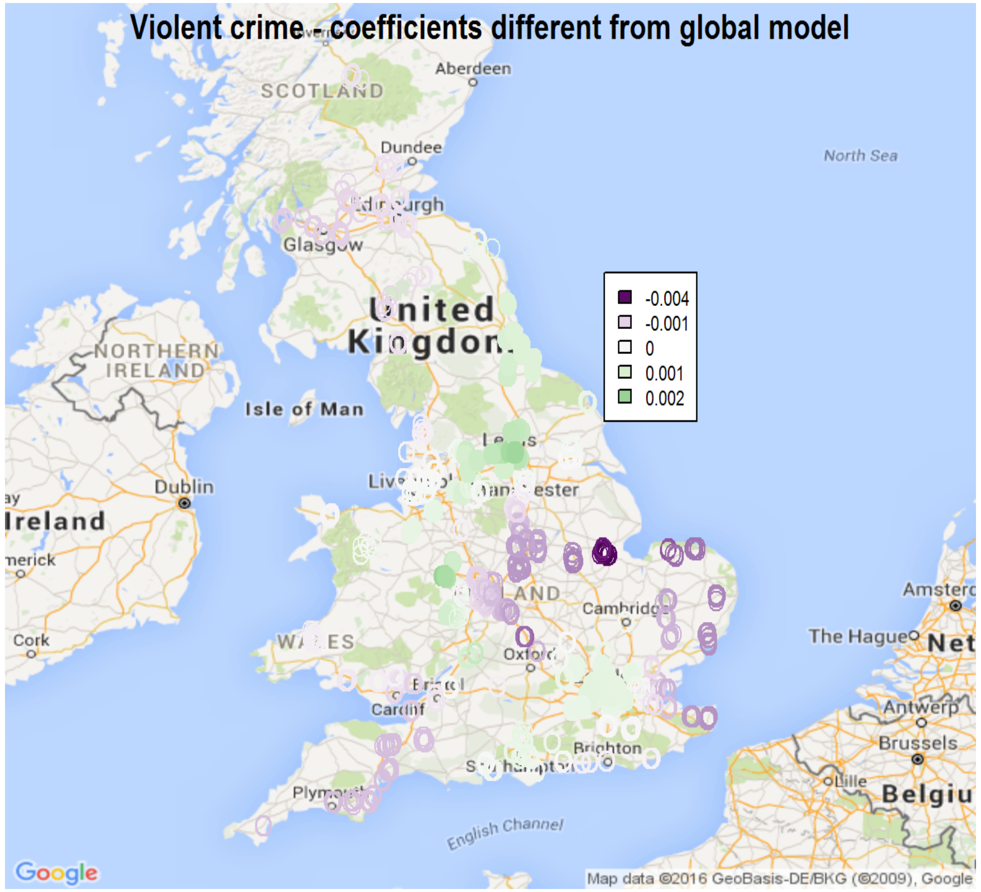
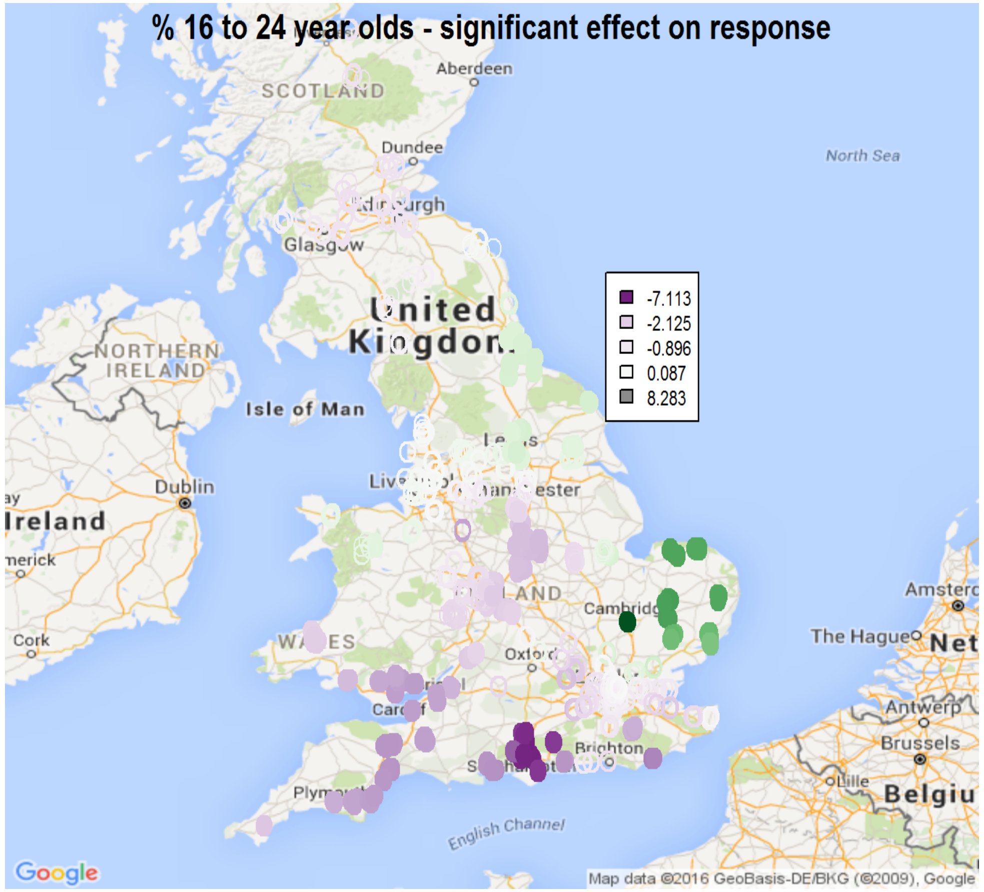


Figure 11 and 12 On the left the coefficients for violent crime estimated with GWR, on the right difference from global estimate

Results: GWR for 16 to 24 year olds



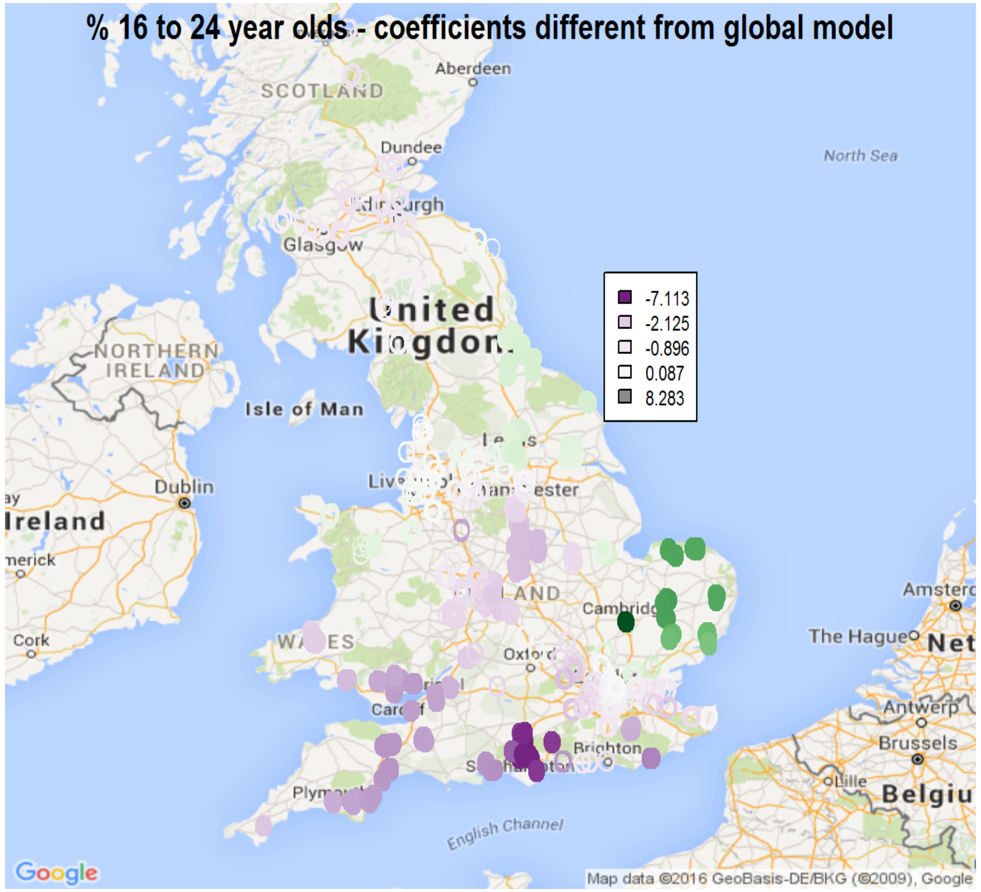


Figure 13 and 14 On the left the coefficients for 16 to 24 year olds estimated with GWR, on the right difference from global estimate