Homework on Spatial Data Analysis

Dina Yakovleva, Dmitry Donetskov, Ekaterina Korolkova

July 2, 2018

1 Assignment

The cautious analyst will use these results to inform future hypothesis testing methods; it is safe to test whether relationships vary across space using the Leung significance test, but other conclusions are best based in future analyses that are, in part, motivated by the suggestive results of the GWR analysis.

- What about the GWR analysis could inform the OLS and spatial regression analyses?
- Are the maps suggestive of spatial regimes?
- Do they suggest that the inclusion of specific interaction terms would be helpful?
- Based on these results, what other variables might you include in future analyses?

2 Answers

2.1 What about the GWR analysis could inform the OLS and spatial regression analyses?

We understand the question as "*What additional information the GWR analysis can bring on top of OLS and Spatial AutoRegression (SAR)?*".

The main advantage of the GWR analysis is the use of distance-weighted matrix to produce locally linear regression estimates for every observation in space. OLS does not take the spatial aspect into account at all and, thus, may frequently lead to invalid conclusions (due to the Simpson's paradox). SAR does include the spatial aspect in the form of autocorrelation coefficient but it is not free of errors if residuals are spatially non-stationary. Therefore, GWR is a more appropriate model to use when OLS and SAR have to be rejected as violating the underlying assumptions.

Results of GWR can probably be used to identify those regions where we can safely use OLS or SAR i.e. results of GWR can help to interpret if data conforms to assumptions like the homoscedasticity and the independence of residuals. Our intuition is that regions with homoscedastical looking data (on those graduated color maps) can be analyzed with OLS/SAR i.e. regions with similar variations can probably be analyzed with OLS/SAR, so it can be used as an additional diagnostic tools for that reason.

Also, we can try to compare AIC, SS of residuals and similar metrics for GWR and OLS/SAR to see if there is any benefit in using GWR as a local regression model or we are safe to stay with OLS/SAR as a global regression model (e.g. in case of AIC for them to be considerably less). Moving to a global regression model in a safe way based on the GWR analysis has the benefit of getting an explanatory model for the whole population (or a considerable part of it).

2.2 Are the maps suggestive of spatial regimes?

Yes, they are as there is considerable variation of the GWR statistics over different regions. For example, R2 are larger for the south which implies the GWR model can produce more significant results for the south.

2.3 Do they suggest that the inclusion of specific interaction terms would be helpful?

The multicollinearity effect is considered to be stronger in the GWR model than in global regression models [1]. Therefore it might be reasonable to use an interact term instead of separate independent variables.

Looking at the maps of R^2 and t-statistics for PFHH, one can notice discrepancy between them which suggests there is other variable contributing to high values of R^2 for some regions where t-statistics for PFHH are not that high.

Given [1], it would be helpful to identify that variable with additional diagnostic and, probably, replace them two with one interaction term. For this particular model, judging by the correlation pair plot, it is PFHH * PUNEM.

2.4 Based on these results, what other variables might you include in future analyses?

The domain of question suggests including PWKCO (Proportion work outside of county of residence) and PERPOV (Persistent poverty, 1970-2000) in future analyses. The first is an indicator of under-industrialized regions and the second is an indicator of "poverty trap".

References

1. *Wheeler D.*, *Tiefelsdorf M.* Multicollinearity and correlation among local regression coefficients in geographically weighted regression // Journal of Geographical Systems. — 2005. — June. — Vol. 7, no. 2. — P. 161–187.