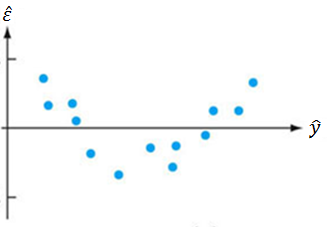
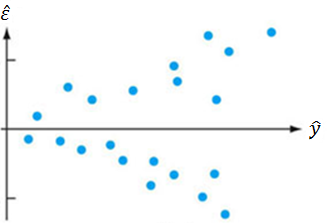
OLS Regression Problems

1. When residuals contain a systematic pattern, the assumption of randomness of residuals is not met.
   1. The first situation is where we may see a non-linear pattern in the residuals if we were to plot them against the predicted values of the dependent variable, . For example, in the instance below, we have a quadratic trend in the residuals, and we may be better fitting a polynomial (quadratic) model than a linear model.



* 1. The second situation is where we may see more obvious heteroscedasticity. Again, we can plot residuals against the predicted values of the dependent variable, .



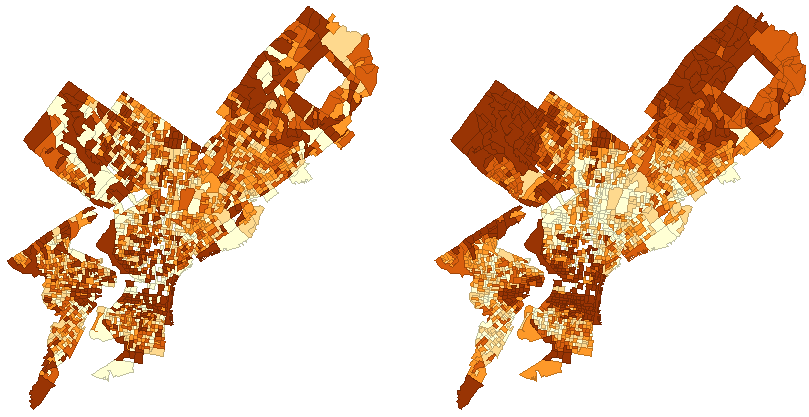
The figure above allows us to visually test for heteroscedasticity. Here, heteroscedasticity is obvious, because the variance (spread) of is smaller for low values of than for high values of .

We can also use the following tests in GeoDa to test the null hypothesis of homoscedasticity, against the alternative hypothesis of heteroscedasticity.

* + 1. Koenker-Bassett Test
    2. Breusch-Pagan Test
    3. White Test

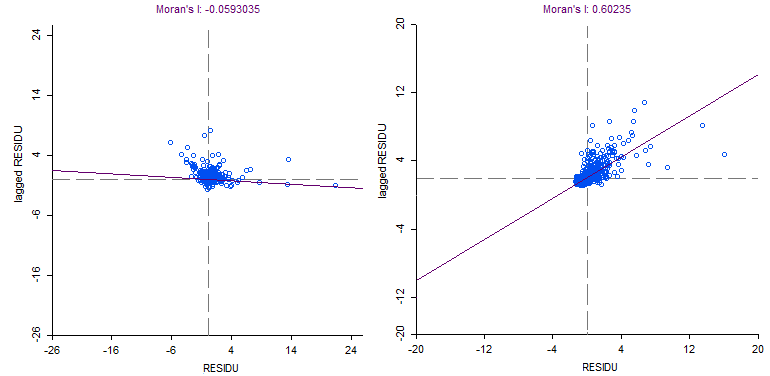
Ideally, we would want to fail to reject the null hypothesis for the alternative hypothesis (i.e., get a p-value > 0.05).

* 1. The third type of problem with residuals arises when we have dependent observations (and residuals). Specifically, if we have spatially autocorrelated OLS residuals, there is systematic under-prediction or over-prediction in certain parts of the study region; furthermore, the significance estimates for the *β* coefficients in OLS may be incorrect (inflated). There are a couple ways to check for spatial autocorrelation of the OLS residuals, though the first thing to do might be to map the residuals. We want the map of the residuals to look random, like the map on the left, and not spatially autocorrelated, like the map on the right.



After looking at the residual maps, we would want to

* + 1. Compute the Moran’s I of the residuals. Ideally, we will see a Moran’s I close to 0, like in the plot on the left below. In the plot on the right, there is an obvious problem with spatial autocorrelation of residuals.



* + 1. Regress residuals on spatially lagged residuals *W*. Ideally, we would see that there is no relationship between and *W* – that is, that the coefficient of *W* is not significantly different from 0.

1. When residuals are not normal, we may run into some problems for OLS (and spatial) regression. We can check for normality by:
   1. Examining the histogram of residuals
   2. Looking at the Jarque-Bera test in Geoda
      1. The null hypothesis is that residuals are normal, and the alternative hypothesis is that they are not normal. We want to not be able to reject the null hypothesis (i.e., get a p-value of 0.05 or higher).
2. When we have multicollinearity in multiple regression, we may run into serious issues with estimating regression parameters. We may check for multicollinearity by:
   1. Looking at the correlation matrix of predictors. If correlations are higher than 0.8 (or some will say 0.9) or lower than -0.8 (or -0.9), we have multicollinearity (severe multicollinearity).
   2. Looking at the Multicollinearity condition number in GeoDa. If it is more than 30, we have a problem with multicollinearity.
   3. Another (more rigorous) way is regressing each predictor *βi* on *all* the remaining predictors. If the *R2* in any of these regressions is greater than 0.8 (or some will say 0.9), we have problems with multicollinearity.
      1. We can also look at Variance Inflation Factors (VIFs) to determine issues with multicollinearity. If the VIF > 4, we have some evidence of multicollinearity, and if VIF > 10, we have a serious issue with multicollinearity.
   4. We may use ridge or lasso regression to address issues with multicollinearity.