

Regressions: Spatial Lagged Dependent Variable Models

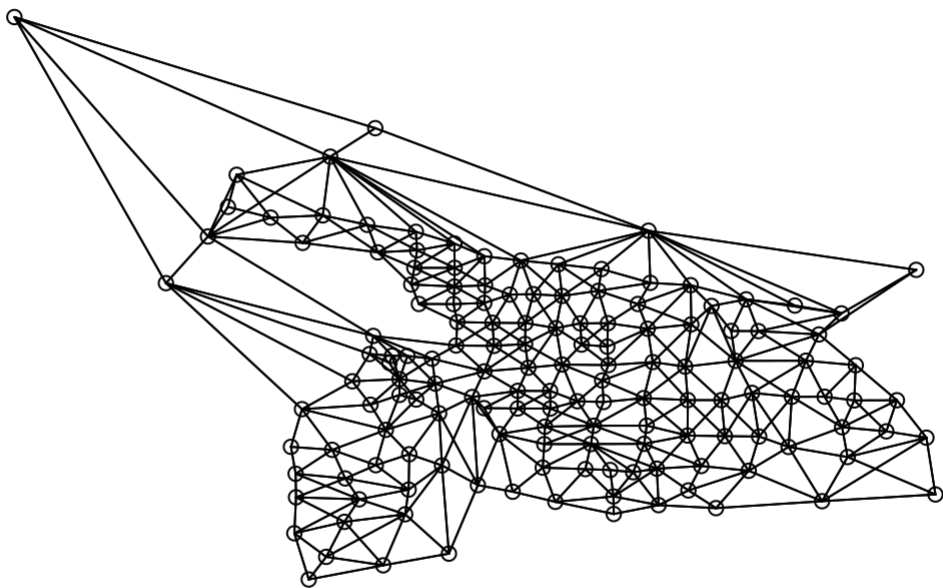
Model Overview

One concern of running an analysis on canopy coverage is that there will be spatial autocorrelation. This occurs when units of observations cluster (in our case Census Tracts), and similar dependent variables are seen. It seems logical that if a tract has high canopy coverage, regardless of any demographic statistics, its neighboring tracts are likely to have high canopy coverage as well.

Thus, we decided to use a spatial lagged dependent variable model to look at the relationship between different demographic data and canopy coverage. In its general form, the spatial lagged dependent variable is written as

$$Y_i = \rho WY_N + \beta X + e$$

where Y_i is tract i's canopy coverage, and ρ is a scalar of WY_N . WY_N is a matrix of the canopy coverage of tract i's neighbors, weighed by the matrix W , depending on distance. Below is an image of the distance of individual tracts from each other.



$\beta X + e$ is the matrix of the independent variables, X, multiplied by their coefficients, included in the matrix β , and e is the error term.

By choosing this type of model, we can account for the fact that neighboring tracts are going have similar canopy coverage irregardless of their demographic data. Thus we can isolate the relationship between demographics and canopy coverage in Portland.

Model Results

Our model included the independent variables for the median family income, percentage of the total population which is African American, median gross rent, gini index, and homeownership of a tract. The coefficients of those variables, and the intercept, are included below.

mlsl\$coefficients

```
##              (Intercept)
##              -2.225319e-01
##              `Median Family Income`
##              2.199398e-06
##  `% Total Population: Black or African American Alone`
##              -2.462425e-04
##              `Median Gross Rent`
##              -6.988772e-05
##              `Gini Index`
##              4.601479e-01
##              `% Occupied Housing Units: Owner Occupied`
##              1.334382e-03
```

The variable % Total Population: Black or African American Alone is not statistically significant. When running alternative regressions for the proportion of the population that is white, there was not statistical significance as well. However, it should be noted that the black coefficient was consistently negative in our alternative regressions, and white was consistently positive.

Median gross rent had a negative and statistically significant effect on canopy coverage. Since we control for income and income inequality with our other variables, this likely reflects the demand for rental properties in a tract. So, as more people demand rental properties, the expected canopy coverage decreases.

Median family income is statistically significant and has a positive effect on canopy coverage. This is unsurprising, and is by far the most statistically significant.

The Gini Index has a positive effect, which means as income distribution becomes more unequal, the expect canopy coverage increases. However, this could either mean more poor people or more concentrated wealth, which has different implications.

Homeownership also has a positive and stastically significant effect. This means that places where people are homeowners are more likely to have a higher canopy coverage.

Our choice of model proved useful, given the ρ coefficient's value below.

mlsl\$rho

```
##      rho
## 0.5671385
```

This is a sizeable ρ for scaling the effect of surrounding tracts' canopy coverage on a given tract's expected canopy coverage. This rho, against the null hypothesis that it should have no effect, is statistically significant at the 1% level.

These spatial dependent variable lags proved useful. Below is the output of a Lagrange Multiplier Test for spatial autocorrelation, with the null hypothesis being that there is not autocorrelation.

mlsl\$LMtest

```
##      [,1]
## [1,] 0.9442622
```

This value translates to a p-value of 0.33118. Thus, our spatial lagged dependent variables appears to resolve the issue of spatial autocorrelation, and our coefficients can be interpreted as robust, as we have effectively controlled for the fact that tracts close to each other will have similar canopy coverage, irregardless of demographics.

Complete Model Output

summary(mlsl)

```
##
## Call:
## lagsarlm(formula = f1, data = tract_stuff, listw = lw, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.245473 -0.043448 -0.010974  0.027142  0.282069
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##
##              (Intercept)              Estimate
## `Median Family Income`              -2.2253e-01
##  `% Total Population: Black or African American Alone` 2.1994e-06
## `Median Gross Rent`              -2.4624e-04
## `Median Gross Rent`              -6.9888e-05
## `Gini Index`              4.6015e-01
##  `% Occupied Housing Units: Owner Occupied` 1.3344e-03
##
##              Std. Error z value
## (Intercept) 7.4688e-02 -2.9795
## `Median Family Income` 4.8793e-07 4.5076
##  `% Total Population: Black or African American Alone` 1.2761e-03 -0.1930
## `Median Gross Rent` 3.8963e-05 -1.7937
## `Gini Index` 1.2367e-01 3.7207
##  `% Occupied Housing Units: Owner Occupied` 5.0377e-04 2.6488
##
##              Pr(>|z|)
## (Intercept) 0.0028871
## `Median Family Income` 6.556e-06
##  `% Total Population: Black or African American Alone` 0.8469883
## `Median Gross Rent` 0.0728639
## `Gini Index` 0.0001987
##  `% Occupied Housing Units: Owner Occupied` 0.0080779
##
## Rho: 0.56714, LR test value: 41.61, p-value: 1.1144e-10
## Asymptotic standard error: 0.074999
##      z-value: 7.5619, p-value: 3.9746e-14
## Wald statistic: 57.183, p-value: 3.9746e-14
##
## Log likelihood: 158.7975 for lag model
## ML residual variance (sigma squared): 0.0062245, (sigma: 0.078895)
## Number of observations: 146
## Number of parameters estimated: 8
## AIC: -301.59, (AIC for lm: -261.99)
## LM test for residual autocorrelation
## test value: 0.94426, p-value: 0.33118
```