Predicting Commuter Modal Choice in Chester County, Pennsylvania Using Spatial Regression

Dan Holton

SYS 502

May 5, 2002

Table of Contents

Background

Philadelphia, Pennsylvania, like most major cities, suffers from large external costs associated with heavy automobile use (e.g., pollution, opportunity cost of idle time on congested highways, accidents). The Southeastern Pennsylvania Transit Authority (SEPTA) is therefore very interested in increasing the number of workers who use public transit, rather than the automobile, to commute from homes in the suburbs to jobs in Philadelphia's central business district, known as Center City. One current proposal to accomplish this goal is the development of the Schuylkill Valley Metro (SVM), a new rail line linking Philadelphia and Reading, providing Center City rail access to growing suburban areas along the Schuylkill River in Chester, Montgomery and Berks Counties. To justify a project of this enormous size and cost, SEPTA must be reasonably certain that a sufficient number of commuters in these areas will choose this new rail line over the automobile.

This report certainly does not attempt to assess the projected increase in SEPTA rail ridership with the construction of the SVM. But, what it does attempt is to develop a spatial regression model to predict the percentage of Center City commuters who use rail in a given census tract using as explanatory variables those socioeconomic and location characteristics of the census tract that are considered to influence modal choice decisions.

Three census tract characteristics were chosen as explanatory variables for this study:

1. Median Income Level - Public transit has long battled the stereotype that riders are YAFU's, an acronym indicating that public transit riders are Young (students), Aged (senior citizens), Feeble (handicapped), or Unemployed (poor). The implication is that affluent commuters are less likely to use transit. Hence, we expect our model to indicate that the percentage of Center City commuters who choose rail decreases with an increase in median income level.
2. Percentage of Housing Units with Two or More Vehicles – Some commuters use public transit simply because they do not own an automobile, or because the family unit owns only one automobile and that vehicle is used by another member of the family for commuting purposes. Housing units with two or more automobiles have more choices for commuting, so we would expect our model to indicate that the percentage of Center City commuters who choose rail decreases with an increase in the percentage of housing units with two or more vehicles. Another way to consider this variable is that commuters who choose rail do not have the need to own a second automobile.

3. Percentage Increase in Travel Time Using Transit – Some commuters use public transit because of the improved reliability in travel time. While urban highways may be very fast at midnight or at 6am, many commuters do not have the luxury of choosing their working hours and therefore encounter congested highways during rush hour periods. This can lead to very high travel time variability for auto users. Public transit, which may have a longer mean travel time, does provide the advantage of very low variability in travel time. If the increase in travel time for transit users is relatively low compared to auto travel time, then transit becomes an attractive option. Therefore, we would expect our model to indicate that the percentage of Center City commuters who choose rail decreases with an increase in the percentage increase in travel time using transit.

Therefore, the specific research questions explored in this report can be summarized as follows:

1. Can the percentage of Center City commuters who use rail in a given census tract be predicted based on median income level, percentage of housing units which have two or more vehicles, and the percentage increase in travel time using transit in that census tract?
2. Is there spatial autocorrelation in the phenomena, and if so, what model most effectively accounts for this spatial autocorrelation?

Chester County, Pennsylvania was chosen as the study area for a number of reasons. First, it is one of the five counties in Pennsylvania (along with Delaware, Montgomery, Bucks, and Philadelphia) that are collectively known as the Philadelphia metropolitan area. Second, Chester County has only one major commuter rail line, SEPTA's Regional 5 (R5), also known as the "Main Line", and limited choices for other modes of public transit (e.g., buses). While other Pennsylvania counties, such as Delaware, have a higher percentage of workers that commute to Center City, these counties also have multiple public transit choices (e.g., multiple rail lines, light rail, subway, bus service), adding undue complexity to the model. Third, the author has a good deal of intuition and anecdotal evidence regarding commuting patterns in the county, having lived in Chester County and commuted to Center City for three years.

Figure 1 displays a map of the five-county metropolitan area in Pennsylvania, providing a visual picture of Chester County's proximity to Philadelphia.

Figure 1: 5-County Philadelphia Metropolitan Area in Pennsylvania

Data

Census Tract Shape Files

The 1990 decentennial census was chosen for the study, as journey-to-work data from the 2000 census will not be published until July, 2002. 1990 census tract shape files were obtained directly from the US Census Bureau (www.census.gov/geo/www/cob/tr1990.html). Centroid locations specified by longitude and latitude were calculated using a custom Visual Basic for Applications (VBA) script in ESRI's AcrMap 8.1 software package.

Census Tract Socioeconomic and Demographic Data

The 1990 census Summary Tape File 3 (STF3), which contains all socioeconomic and demographic variables, was also obtained directly from the US Census Bureau (http://homer.ssd.census.gov/cdrom/lookup). The following census tract variables were obtained from STF3 for all tracts in Chester County:

STF3 Category	STF3 Code	Data Element
Persons	P0010001	Total Unweighted Sample Count of Persons
Place of Work – MSA/PMSA Level	P0470001	Worked in MSA/PMSA of residence: Central city
Place of Work – MSA/PMSA Level	P0470002	Worked in MSA/PMSA of residence: Remainder of this MSA/PMSA
Place of Work – MSA/PMSA Level	P0470003	Worked outside MSA/PMSA of residence: Worked in a different MSA/PMSA: Central city
Place of Work – MSA/PMSA Level	P0470004	Worked outside MSA/PMSA of residence: Worked in a different MSA/PMSA: Remainder of different MSA/PMSA
Place of Work – MSA/PMSA Level	P0470005	Worked outside MSA/PMSA of residence: Worked outside any MSA/PMSA
Means of Transportation to Work	P0490001	Car, truck, or van: Drove alone
Means of Transportation to Work	P0490002	Car, truck, or van: Carpooled

STF3 Category	STF3 Code	Data Element
Means of Transportation to Work	P0490003	Public transportation: Bus or trolley bus
Means of Transportation to Work	P0490004	Public transportation: Streetcar or trolley car
Means of Transportation to Work	P0490005	Public transportation: Subway or elevated
Means of Transportation to Work	P0490006	Public transportation: Railroad
Means of Transportation to Work	P0490007	Public transportation: Ferryboat
Means of Transportation to Work	P0490008	Public transportation: Taxicab
Means of Transportation to Work	P0490009	Motorcycle
Means of Transportation to Work	P04900010	Bicycle
Means of Transportation to Work	P04900011	Walked
Means of Transportation to Work	P04900012	Other means
Means of Transportation to Work	P04900013	Worked at home
Median household income in 1989	P080A001	Aggregate Household Income in 1989
Occupancy Status	H0040001	Housing units: Occupied
Occupancy Status	H0040002	Housing units: Vacant
Tenure by Vehicles Available	H0370001	Owner occupied: None
Tenure by Vehicles Available	H0370002	Owner occupied: 1
Tenure by Vehicles Available	H0370003	Owner occupied: 2

STF3 Category	STF3 Code	Data Element
Tenure by Vehicles Available	H0370004	Owner occupied: 3
Tenure by Vehicles Available	H0370005	Owner occupied: 4
Tenure by Vehicles Available	H0370006	Owner occupied: 5 or more
Tenure by Vehicles Available	H0370007	Renter occupied: None
Tenure by Vehicles Available	H0370008	Renter occupied: 1
Tenure by Vehicles Available	H0370009	Renter occupied: 2
Tenure by Vehicles Available	H03700010	Renter occupied: 3
Tenure by Vehicles Available	H03700011	Renter occupied: 4
Tenure by Vehicles Available	H03700012	Renter occupied: 5 or more

Percentage of Center City Commuters Who Use Rail

This variable was calculated from the census tract socioeconomic and demographic data in STF3 as follows:

> Percentage of Center City commuters who use rail = (Public transportation: Railroad) / (Worked in MSA/PMSA of residence: Central city) x 100

Ten census tracts (3012.01,3012,3059,3104.98,3103,3105,3081,3005,3053,3041.01) had 0 Center City commuters and were therefore removed from the analysis.

The assumption inherent in this calculation is that all workers who use rail for journey-to-work are commuting to Center City. This may not always be the case. For example, there may be residents of, say, Downingtown, who use the R5 rail line to commute to a job in, say, Berwyn. However, it is believed that this number of commuters is extremely low, as there are few businesses within walking distance of most R5 rail stations in Chester County. This indicates that this category of commuters would need to find some other means of transportation to from the R5 station to the job site, and the most likely means, bus service, is infrequent in these areas.

A histogram of these values indicates that the distribution is slightly skewed to the right. Unfortunately, a log transformation of the data eliminates approximately 20% of the census tracts. Therefore, the spatial regression analysis will be performed with both raw and transformed data points.

Median Income Level

This variable is obtained directly from STF3 with no required assumptions or calculations.

Percentage of Housing Units with 2 or More Vehicles

This variable was calculated from the census tract socioeconomic and demographic data in STF3 as follows:

> Percentage of housing units with 2 or more vehicles = [(Owner occupied: 2) + (Owner occupied: 3) + (Owner occupied: 4) + (Owner occupied: 5 or more) + (Renter occupied: 2) + (Renter occupied: 3) + (Renter occupied: 4) + (Renter occupied: 5 or more)] / (Occupied Housing Units) x 100

Unfortunately, the census tracks vehicle ownership by housing unit, not by household. In other words, there may be multiple households living in one housing unit, and in general, we would expect members of the same household, not housing unit, to share a vehicle. However, most of the census tracts in the study area had a very small (1-2%) difference between the number of households and number of occupied housing units, and all census tracts had less than a 10% difference.

Percentage Increase in Travel Time Using Transit

A number of assumptions and estimates were made the calculation of this variable. The following procedure was used to estimate the percentage increase in travel time using transit:

- Driving times from each Chester County zip code to Center City (15^{th} and Market Streets) were calculated using Mapquest software (www.mapquest.com).
- Census tract centroids, in latitude and longitude, were mapped to associated zip codes using a custom program written in Matlab. The driving time from each census tract centroid to Center City was set to the travel time from its associated zip code to Center City.
- Street addresses for each R5 station were obtained from SEPTA (www.septa.org).

- The latitude and longitude of each R5 station address was calculated using ZipInfo software (www.zipinfo.com).
- The R5 station nearest to each census tract was calculated using latitude and longitude coordinates in Excel.
- Driving times from each census tract zip code to the nearest R5 station were calculated using Mapquest software.
- Transit times from each R5 station to Center City (Suburban Station at 16th Street and JFK Boulevard) were obtained from SEPTA.
- Total travel time for rail for each census tract was set equal to the driving time to the nearest R5 station plus the transit time from the station to Center City.
- The percentage increase in travel time using transit for each census tract was then calculated as follows:

  Percent increase in travel time using transit = (total travel time for rail – driving time) / (driving time) x 100

There are a number of assumptions made in this calculation, all of which may contribute to limitations with the model:

- <u>No congestion on roads</u> – All driving time estimates assume that there is no congestion on roads. This implies that commuters will choose a travel time such that congestion is minimal or nonexistent, which is not realistic. It is true that both auto users and rail users will be affected by congestion on roads. However, this effect is much greater for auto users because they spend more time on roads and they travel on more congested highways to reach Center City. The rail users may experience some congestion traveling to the R5 station, but it is not nearly as significant as the congestion experienced by auto commuters to Center City.
- <u>No R5 delays</u> - Transit times on the R5 assume no delays. While this is not perfectly realistic, it is much more realistic than the assumption of no congestion on roads, as SEPTA provides a money-back guarantee for delays greater than 15 minutes.
- <u>Identical egress time</u> – We assume egress time in Center City is the same for auto and rail users. That is, it takes the same amount of time for rail users to walk from Suburban Station to work as it does for auto users to park (presumably closer to work) and then walk from the parking facility to work.
- <u>Transit users will choose the closest R5 station</u> - This is not always the case, as users may drive farther to another station in order to save transit time. Consider the commuter in Pottstown, who may drive 20 minutes directly south on Route 100 to the Exton station in order to ride 50 minutes

on the train. Or, that same commuter may drive 30 minutes east on Route 422 to the Devon station in order to ride 35 minutes on the train, saving 5 minutes of overall travel time. Linked trips also come into play, as commuters may drive to a different station because it is close to, say, Home Depot, and they need to do some shopping after work. These complex modal choice scenarios are beyond the scope of this analysis and clearly present problems with the model.

In addition to these assumptions, perhaps the largest source of error in the travel time calculations lies in the discrepancies between travel times originating from zip codes and travel times originating from census tract centroids. This mapping of census tract centroids to zip codes was a “necessary evil”, due to the availability of driving time estimates based on zip codes only. However, this clearly leads to uncertainty in our driving time estimates.

Analysis

Visual Analysis

Figures 2-5 provide visual pictures of the variable under question (percentage of Center City commuters who use rail), as well as the three explanatory variables (median income level, percentage of housing units with two or more vehicles, and percentage increase in travel time using transit) in Chester County.

Figure 2: Percentage of Center City Commuters Who Use Rail

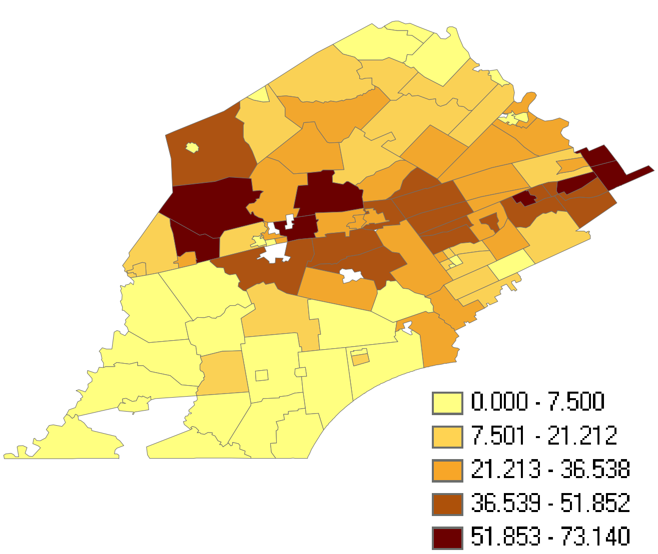


Figure 3: Median Income Level

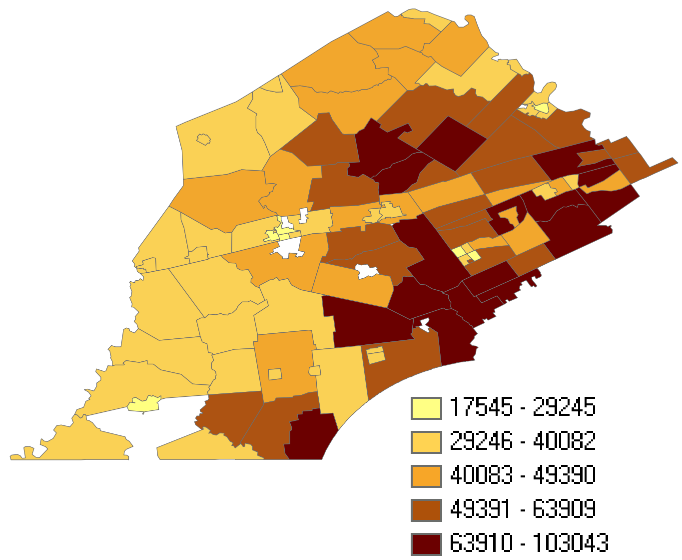


Figure 4: Percentage of Housing Units with Two or More Vehicles

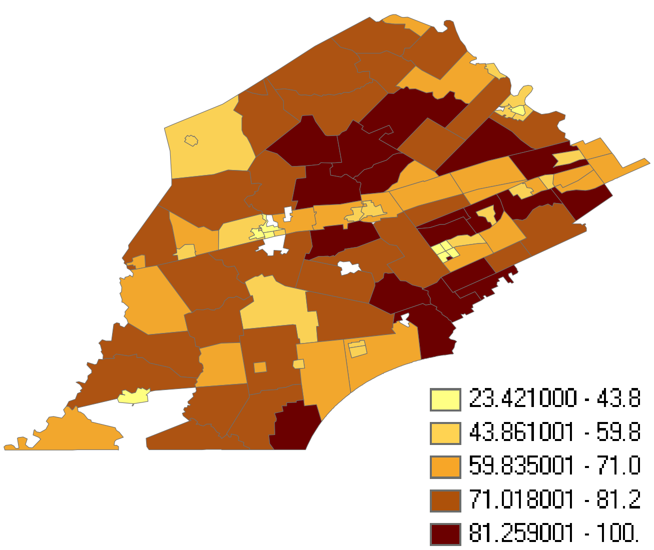
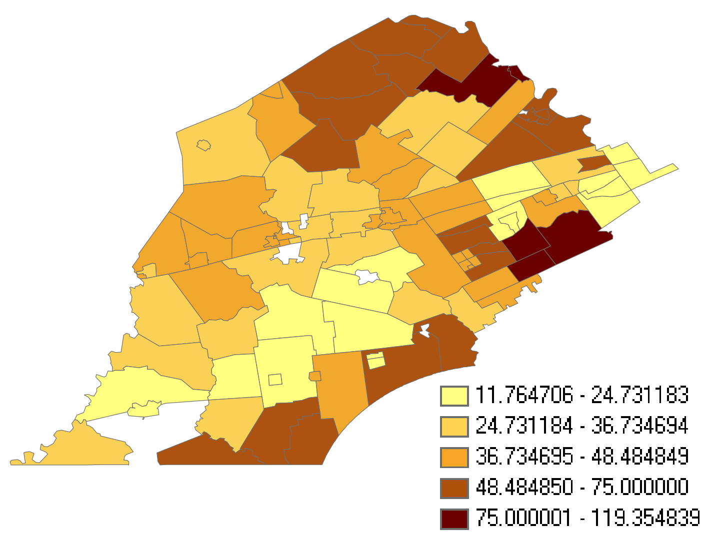


Figure 5: Percent Increase in Travel Time Using Transit



Figures 2-5 provide a number of visual insights. First, Figure 2 indicates that that there is band of high values running from east to west in the northern half of Chester County. This band roughly matches the geographic location of the R5 line. This image agrees with our intuition that commuters who live closer to an R5 station are more likely to choose rail over the automobile (i.e., the increase in travel time by using rail is relatively small).

Figure 3 indicates that there are higher levels of median income in census tracts in the northeast region of the county. This region is the point closest to Philadelphia, so this image is also intuitive. That is, tracts closest to Philadelphia have the highest median income levels, with income levels decreasing with distance from the city. But, it is difficult to assess visually if income level has any impact on the percentage of rail commuters.

Figure 4 indicates an overwhelmingly high percentage of housing units with two or more vehicles in census tracts throughout the county. This would suggest that this explanatory variable is not likely to be significant in the regression model. But, there does seem to be a band of low values running from east to west along the geographic location of the R5 line. This might suggest that commuters who reside close to the R5 line (i.e., within walking distance to a station) might be less likely to own two or more vehicles, and perhaps this will prove to be a significant explanatory variable after all.

Figure 5 indicates that a large percentage increase in travel time for transit in the northeastern and southeastern regions of Chester County. This phenomenon can be explained in light of the location of these regions with respect to major highways and to the R5 line. Commuters in the northeastern region of the county have access to Route 422, a major highway providing fairly quick access to Route 76 and therefore Center City. Likewise, commuters in the southeastern region have access to Route 1, providing relatively quick access to Interstate 95 and therefore Center City. These areas are not very accessible to the R5 line, so the commute via transit is relatively long with respect to the auto commute.

Ordinary Least Squares (OLS)

An OLS regression provided the following parameter estimates, with an R-Square value of only 0.14:

Parameter	**Estimate**	**Significance**
Y-Intercept	26.843679	0.0031
Median Income Level	0.0004687	0.0050
Percentage of Housing Units with Two or More Vehicles	-0.251504	0.1520
Percentage Increase in Travel Time with Transit	-0.210509	0.0118

The results of the OLS regression indicate the following:

- Only 14% of the deviation from the mean percentage of Center Commuters who use rail is explained by the three explanatory variables.
- The impact of the median income level is significant. However, the parameter is positive, which contradicts our original hypothesis that the percentage of Center City commuters who choose rail decreases with an increase in median income level.
- The impact of the percentage of housing units with two or more vehicles is not statistically significant.
- The impact of the percentage increase in travel time with transit is significant, and the negative parameter supports our original hypothesis that the percentage of Center City commuters who choose rail decreases with an increase in the percentage increase in travel time using transit.

While the OLS regression does indicate statistical significance in two of the three explanatory variables, further analysis indicates that there is significant spatial autocorrelation in the OLS residuals. Specifically, a regression of OLS residuals using the residual of a census tract's nearest neighbor as an explanatory variable yields the following results, with an R-Square value of 0.24:

Parameter	**Estimate**	**Significance**
Y-Intercept	-0.275616	0.8588
Nearest Neighbor Residual	0.5436421	<0.0001

In other words, an OLS residual in a census tract can be predicted by the OLS residual in its nearest neighbor census tract. This is evidence of spatial autocorrelation in the residuals. We can further test for this autocorrelation through random permutation tests. Essentially, we simulate the structure of the residuals under the null hypothesis of no spatial autocorrelation, and then test the significance of four sample indices (Moran's I-coefficient, rho statistic, correlation coefficient, and the Durbin and Watson, or D-W, measure) in light of this null hypothesis. Specifically, we complete the following procedure for each index:

- Compute the sample statistic using the OLS residuals.
- Construct 999 random permutations of the residuals, creating 999 new data sets.
- Using the 999 new data sets, compute 999 test statistics.
- Rank each sample index within the total set of 1000 indices. (It should be noted that the D-W measure is ranked from lowest to highest, with rank 1 for the lowest D-W measure and 1000 for the highest D-W value, since lower D-W values indicate higher positive autocorrelation. The other three indices are ranked from highest to lowest.)

This rank indicates that spatial autocorrelation is significantly positive at the level of the rank divided by 1000. For instance, if the rank of the sample index is 5 (out of the 1000 indices), then we can say that spatial autocorrelation among the OLS residuals is significantly positive at the 0.005-level. In other words, there is only a 0.5% probability of getting, by chance alone, this high of an index (in the case of Moran's I-coefficient, rho statistic, and correlation coefficient), or this low of an index (in the case of the D-W measure), given a true hypothesis of no spatial autocorrelation among residuals.

The following table summarizes the results of the random permutation test:

Index	**Moran (I)**	**Rho (p)**	**Correlation (r)**	**D-W (d)**
Max	0.2453	0.6941	0.4126	1.8737
Min	-0.2366	-0.6200	-0.3830	0.8628
Value	0.4594	0.8656	0.6306	0.6119
Significance	0.0010	0.0010	0.0010	0.0010

Each of the four indices indicates that there is extremely significant spatial autocorrelation in the OLS residuals. This suggests that either the Spatial Autoregression (SAR) or Spatial Lag (SL) models, which take into account this spatial autocorrelation among residuals, would be better suited for the problem at hand.

Spatial Autoregression (SAR) and Spatial Lag (SL)

Using a boundary-share weight matrix, the Spatial Autoregression and Spatial Lag models yielded the following results.

Final Regression Results:

Parameter	**Coefficient (SAR)**	**Coefficient (SL)**	**P-Value (SAR)**	**P-Value (SL)**
Constant (Y-intercept)	20.717906	6.391152	0.016411	0.316216
Median Income Level	0.000038	0.000180	0.821327	0.120052
Percentage of Housing Units with Two or More Vehicles	-0.016071	-0.045190	0.917684	0.710594
Percentage Increase in Travel Time with Transit	-0.07463940	-0.12134718	0.379139	0.034675

Autocorrelation Results:

Parameter	**Value**	**P-Value**
Rho (SAR)	0.727823	0.000000
Lambda (SL)	0.701116	0.000000

Goodness-of-Fit Results:

Measurement	**Value (SAR)**	**Value (SL)**
Pseudo R-Square	-0.022427	0.077066
Pseudo R-Square Adjusted	-0.053410	0.049099
Squared Correlation	0.099603	0.589510
Log-Likelihood	-421.153520	-416.026862

Tests of Model:

Test	**Value (SAR)**	**Value (SL)**	**P-Value (SAR)**	**P-Value (SL)**
Likelihood Ratio	43.565779	53.819095	0.000000	0.000000
Common Likelihood Ratio	18.481015	N/A	0.000350	N/A

Repeating the random permutation test for the SAR residuals yields the following results:

Index	Moran (I)	Rho (p)	Correlation (r)	D-W (d)
Max	0.2395	0.5986	0.3559	2.0411
Min	-0.2905	-0.7428	-0.4282	0.9303
Value	-0.0501	-0.1464	-0.0857	1.4425
Significance	0.7210	0.7210	0.7200	0.7330

Repeating the random permutation test for the SL residuals yields the following results:

Index	Moran (I)	Rho (p)	Correlation (r)	D-W (d)
Max	0.2651	0.6243	0.3890	1.9387
Min	-0.2487	-0.8819	-0.3913	0.9194
Value	-0.0856	-0.2325	-0.1411	1.5396
Significance	0.8420	0.8150	0.8230	0.8940

The SAR model hypothesizes that:

- The dependent variable (percentage of Center City commuters who use rail) can be described by a linear combination of explanatory variables (median income level, percentage of housing units with two or more vehicles, percentage increase in travel time with transit) plus some residual.
- This residual can be described as the sum of spatial dependencies with other residuals plus some internal effect, or final residual.
- These internal effects, or final residuals, are independent and identically (multinormally) distributed (IID).

Conversely, the SL model hypothesizes that:

- Spatial dependencies exist directly among the values of the dependent variable (percentage of Center City commuters who use rail).
- Residuals are independent and identically (multinormally) distributed (IID).

These results suggest the following:

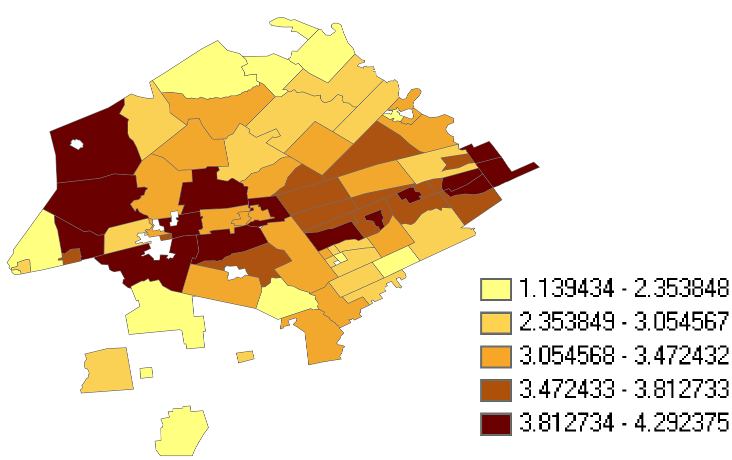
- While the SAR model suggests that none of the explanatory variables are significant, the SL model suggests that the percentage increase in travel time with transit is significant in predicting percentage of Center City commuters who use rail.

- While all four goodness-of-fit measures favor the SL model, the Pseudo R-Square and Pseudo R-Square Adjusted values for the SL model indicate that very little of the variation from the mean percentage of Center City commuters who use rail is accounted for by the SL model (8% and 5%, respectively).
- All four sample indexes indicate that spatial autocorrelation has been accounted for in both the SAR and SL models.
- The Common Likelihood Ratio is high, which suggests that it is likely that spatial dependencies are captured better by spatial lags than by autoregressive residuals.

Transformed Data Analysis

To eliminate possible errors due to a skewed distribution, the same SAR and SL analyses were conducted on transformed (natural logarithm) values of percentage of Center City commuters who use rail. The data transformation required the elimination of approximately 20% of the census tracts from the study area, since there were a good number of tracts in Chester County who had 0 rail commuters. Figure 6 provides a visual picture of the new study area with transformed data.

Figure 6: Natural Logarithm of Percentage of Center City Commuters Who Use Rail



Using a boundary-share weight matrix, the Spatial Autoregression and Spatial Lag models yielded the following results with respect to the transformed data.

Final Regression Results:

Parameter	**Coefficient (SAR)**	**Coefficient (SL)**	**P-Value (SAR)**	**P-Value (SL)**
Constant (Y-intercept)	3.517335	2.538589	0.000000	0.000000
Median Income Level	0.000008	0.000004	0.287870	0.478992
Percentage of Housing Units with Two or More Vehicles	-0.006951	-0.001948	0.403233	0.777309
Percentage Increase in Travel Time with Transit	-0.007197	-0.009375	0.041318	0.000935

Autocorrelation Results:

Parameter	**Value**	**P-Value**
Rho (SAR)	0.326584	0.013010
Lambda (SL)	0.317354	0.000017

Goodness-of-Fit Results:

Measurement	**Value (SAR)**	**Value (SL)**
Pseudo R-Square	-0.025218	0.132169
Pseudo R-Square Adjusted	-0.066226	0.097456
Squared Correlation	0.145665	0.332280
Log-Likelihood	-74.440432	-67.904351

Tests of Model:

Test	**Value (SAR)**	**Value (SL)**	**P-Value (SAR)**	**P-Value (SL)**
Likelihood Ratio	3.698743	16.770904	0.054453	0.000042
Common Likelihood Ratio	23.092928	N/A	0.000039	N/A

Repeating the random permutation test for the SAR residuals yields the following results:

Index	Moran (I)	Rho (p)	Correlation (r)	D-W (d)
Max	0.2733	0.7825	0.4364	1.9128
Min	-0.2547	-1.0669	-0.4792	0.8310
Value	-0.0208	-0.0586	-0.0349	1.3971
Significance	0.5370	0.5250	0.5330	0.6320

Repeating the random permutation test for the SL residuals yields the following results:

Index	Moran (I)	Rho (p)	Correlation (r)	D-W (d)
Max	0.2350	0.6697	0.3588	2.0004
Min	-0.2553	-0.9880	-0.4633	0.9026
Value	-0.0711	-0.1919	-0.1168	1.5127
Significance	0.7430	0.7050	0.7200	0.8230

The results of the analysis on the transformed data suggest the following:

- Both the SAR and SL models indicate that the percentage increase in travel time with transit is significant in predicting the natural logarithm of the percentage of Center City commuters who choose rail.
- All four goodness-of-fit indicators again favor the SL model, and with this transformed data the SL model is accounting for a much-improved 13% of the variation from the mean natural logarithm of the percentage of Center City commuters who choose rail.
- All four sample indexes again indicate that spatial autocorrelation has been accounted for in both the SAR and SL models.
- Again, the Common Likelihood Ratio is high, which suggests that it is likely that spatial dependencies are captured better by spatial lags than by autoregressive residuals.

Weight Matrices

A boundary-share weight matrix was used for each calculation, since it is generally the most reliable weight matrix for census tracts that do not follow a standard lattice shape, such as those in Chester County. However, to confirm consistency in results, a nearest neighbor and a symmetric nearest neighbor weight matrix were also used with both the SAR and SL analyses. While there were some slight differences in various values, the overall results (sign, relative magnitude, and significance) mostly confirmed those yielded using a boundary-share weight matrix. There were, however, a few exceptions:

- The Common Likelihood Ratio for the SAR model using a symmetric nearest neighbor weight matrix was smaller and less significant (0.135244) than previous results indicated.
- The nearest neighbor matrix indicated that there was less significant spatial autocorrelation in both the SAR (rho) and SL (lambda) models.

Conclusions

From the results of our spatial regression analyses, we can draw several tentative conclusions. First, it is clear that there are significant explanatory variables missing from the regression model, as the model as-is explains a very small amount of the variation from the mean percentage of Center City commuters who use rail. This is not surprising, given the tremendous amount of factors that go into modal choice. Second, neither of the two explanatory variables that can be characterized as socioeconomic data (i.e., median income level and percentage of housing units with two or more vehicles) are significant factors in predicting the percentage of Center City commuters who use rail. Third, it is likely that the percentage increase in travel time using transit is a significant factor in predicting the percentage of Center City commuters who use rail. Fourth, there is some evidence that the spatial dependencies in these phenomena are modeled more closely by spatial lags (i.e., spatial autocorrelation in the dependent variable) than by spatial autoregression (i.e., spatial autocorrelation in residuals).

The scope of this project was limited to Chester County, Pennsylvania for two main reasons. First, the data used in the analysis required intensive manual work to identify zip code locations of census tract centroids and associated driving times to Center City and to SEPTA train stations. Second, complexities in transit modal choice were eliminated from the model by having only one rail line in the study area. Further research should consider other counties in the Philadelphia metropolitan area, both in Pennsylvania, Delaware, and New Jersey. In addition, further research should consider additional explanatory variables, especially variability in travel time.