Ann Hartell

# Sprawl and Commuting: Exploring New Measures of United States Metro Regions

SRE-Discussion 2015/07 2015

**Discussion Paper**

# Sprawl and Commuting: Exploring New Measures of United States Metro Regions

Ann Hartell
Doctoral Candidate
Institute for Multi-Level Governance and Development
WU/Vienna University of Economics and Business
ahartell@gmail.com

Paper originally presented at the 94th Annual Meeting of the
Transportation Research Board
Washington, DC
January 2015

**ABSTRACT**
The degree of connectivity and proximity that results from the configuration of land uses and associated transport networks is an important concept in much of the transportation research agenda. A substantial body of work has developed around the idea that compact, mixed-use development with multimodal transport options will shape travel behavior, increasing the use of transit, walking, and cycling for routine travel. Yet empirical evidence is somewhat mixed. One of the reasons for this uncertainty is the difficulty of defining and measuring sprawl in a meaningful way for use in quantitative analyses, rather than using regionally idiosyncratic or mono-dimensional definitions of sprawl. A recently released national dataset measuring multiple dimensions of urban form offers an opportunity to explore the relationship between transportation and sprawl.

This study uses a series of spatial regressions to model effects on the share of a county's workers who commute by driving alone. The results for income are found to be robust across various model specifications, confirming the well-established, positive relationship between income and driving to work. The results for the Street Accessibility Factor suggest characteristics of the street network are related to the choice to commute by driving alone, with more compact street networks and greater connectivity associated with reduced driving alone. The Land Use Mixing Factor has little power in explaining travel behavior, despite its intuitive appeal as the land use component of the commute mode decision.

*Keywords*: Sprawl, Land use transportation interaction, Accessibility, Commute mode, Spatial modeling

## INTRODUCTION

There is a long-recognized connection between the patterns of urban development and patterns of human social and economic behavior across space. The degree of connectivity and proximity that results from the configuration of land uses and associated transport networks is an important concept in much of the transportation and urban studies research agenda, tracing back to the classic writings of von Thünen (*1*) and Christaller (*2*). Today, interest in addressing environmental impacts (water quality, air pollution, and land conservation), climate change impacts, and socio-economic impacts (physical health, quality of life, equity) of human settlement patterns has sustained an interest in evaluating the land use-transportation interaction, frequently focusing on the differences between compact versus sprawling places. Generally, compact places are defined as places with high densities, mixed land uses development, and multimodal transportation systems that deliver a high degree of accessibility. Sprawl, on the other hand, is the condition of low density, segregated land uses, and an autocentric (mono-modal) transportation system. These general concepts offer an intuitive framework for understanding the direct and indirect effects of particular configurations of land use and transportation systems, but are of little help in framing robust, quantified studies that can provide the kind of evidence needed for targeted projects, programs, or policies. Thus, to make meaningful contributions, research must continue to seek improvement in measuring and understanding the relationship between sprawl and outcomes that improve or erode desired conditions.

This study contributes to this ongoing effort by exploring recently released measures of sprawl and their relationship to commuting behavior. Using national indices of urban form, it models the relationship between urban form characteristics and the share of workers who commute by driving alone. Do greater street system connectivity and increased jobs accessibility reduce the share of workers who commute by driving alone? If so, is the effect different for workers who commute to a county that has a different level of compactness than their county of residence? Turning to the question of equitable effects, is the relationship between urban form and travel behavior different for low income workers compared with high income workers?

## LITERATURE REVIEW

Much of the relevant sprawl literature is structured around claims that sprawl promotes automobile dependency through a combination of low density development and a deficit of functional infrastructure for other modes. In contrast, compact, dense environments are theorized to discourage auto use and promote the use of alternative modes for routine travel.

In a meta-analysis of 50 quantitative studies of the relationship between urban form and travel published through 2009, Ewing and Cervero (*3*) sought generalizable patterns travel behavior that could be used by urban planning practitioners. Overall, they found that mode choice was inelastic to characteristics of urban form, although increased density, job accessibility, proximity to downtown, land use mixing, and street intersection density were found to have some effect on reducing driving, and increasing the use of active transport modes. Khattak and Rodriguez (*4*) found significant differences in mode shares in a study of a matched pair of neighborhoods: residents in a compact 'New Urbanist' neighborhood did substitute walking and bus modes for auto trips compared with residents in a low-density suburban neighborhood. Still, the evidence for the urban environment being an important determinate of travel behavior is mixed. In the Special Report *Does the Built Environment Influence Physical Activity?,* the Transportation Research Board summarized a wide-ranging examination of the evidence by noting that 'a supportive built environment alone is not sufficient . . . . nevertheless, it can play a facilitating role' in shaping mode choice in favor of a higher use of alternative modes. (*5*, p. 159).

As for the environmental and climate change impacts of sprawl, the high-profile report *Growing Cooler* (*6*) combined a review of literature with original calculations to argue for policies to mitigate a wide range of negative effects of sprawl, especially rising CO2 emissions. In addition to CO2, other studies have shown a link between sprawl and other air pollutants compared with compact metropolitan areas (*7*). Some of this effect, however, is being mitigated by improvements to vehicle technologies (see, e.g. *8*). Brabec et al (*9*) found high levels of impermeable surface, a feature of extensive development, negatively affects water quality by increasing rainwater runoff which carries more pollutants to water resources.

Recently, a revived interest in the relationship between public health and transport has fueled a flurry of studies on the effects of sprawl on health. Most of these studies find a negative effect on health, largely explained by lower levels of physical activity—theorized to be a result of automobile dependency—in low-density areas (see, e.g. *10*).

Other scholars have been critical of claims of the effects of sprawl. Feng et al. (*11*) noted considerable heterogeneity in methods, especially the definitions of sprawl employed, the selection of geographic scale for measuring sprawl, and cross-jurisdictional differences affecting the quality and comparability of the underlying data. These authors assert that the sensitivity of models to specification changes in urban form metrics makes it difficult to draw strong conclusions that can inform policy or practice. Eid et al. (*12*) join this criticism of methods and lack of robustness in their longitudinal study of the relationship between sprawl and obesity. Using panel data from the National Longitudinal Youth Survey (NLYS) for residential location and Body Mass Index and controlling for changes in the urban environment after a household move, they found no significant relationship between weight gain and urban environment. These findings strongly suggest self-selection in residential location decisions explains the superficial association between obesity and sprawl, adding another complication for the design of sprawl studies. Notably, Eid et al. (*12*) used very limited measures of urban form coupled with a geographic buffer approach that likely introduces spatial aggregation issues. Problematic and dramatically varied definitions of sprawl—either too simplistic or too aggregated—no doubt contribute to the lack of clear evidence on sprawl impacts across studies.

Cutsinger et al. (*13*) make a pointed argument for improved rigor and provided a factor analysis of urban form components as an example of how to define a quantified definition of sprawl (see also *14*). This work is updated by Sarzynski et al. (*15*) who suggest that effective policies to improve urban form-related quality of life must be informed by studies that account for the diversity of US cities and the complexity of interrelated land use, transport, demographic, and historical factors. Further, land development, housing, transport, and other socio-economic processes are addressed through different regulatory mechanisms with different policy frameworks. One-dimensional definitions or measures of urban form are unlikely to account for the complexity of these interrelationships and their effects. Thus, from a policy-making perspective, while aggregating various dimensions of urban form into a single metric may appeal to those seeking easy-to-report rankings, this approach does not provide adequate clarity to develop targeted, effective and implementable policies.

Addressing some of the definitional concerns, Kelly-Schwartz et al. (*16*) used data from 29 Metropolitan Statistical Areas (MSAs) to model sprawl effects on measures of physical health. They found no significant relationship between composite measures of sprawl. However, when urban form was measured using narrowly focused metrics, differential effects were seen: higher street network connectivity was associated with better health (as reported by survey respondents and physicians), which they attribute to a transport effect through higher rates of walking and transit use. Density alone had a negative effect on health. Thus the effects of various dimensions of sprawl might have different effects from one another or from any composite sprawl measure.

A few studies have considered social effects of sprawl. For example, Brueckner and Largey (*17*) tested Robert Putnam's claim that sprawl reduces social capital and weakens social networks by decreasing informal interactions. They found, however, that increased density was associated with less interaction among neighbors, having fewer close friends, and less frequent informal socializing. Using a more complex approach to defining the built environment than simple density, Kamrazzman et al (*18*) found higher levels of trust and social capital among residents or compact, transit-oriented developments in Australia. Whether such studies can be taken as clear evidence is unclear, given the disparate definitions of sprawl used.

There is a vast literature on the determinants of commute mode choice, as well as many papers on the health, social, and economic impacts of sprawl; this section presents only a summary view. The lack of consensus across studies suggests continued work is needed to disentangle the complex factors involved. Such work must highlight convergence on a robust and policy-relevant definition of the sprawl phenomenon, an area ripe for improvement. The focus of this study is to employ a recently released dataset of urban form characteristics in a transportation-related analysis. It offers insights into the complex relationship between land use and transportation, as well as observations on the suitability of this new data resource for transportation-related analyses.

## DATA

The measures of urban form are variables developed by a research team funded by the National Institutes of Health, Smart Growth America, and the Ford Foundation (*19*). Using data dating largely to 2010, the researchers calculated sprawl measures for MSAs with populations of 200,000 or more: 221 metro regions and almost 1,000 counties in the US (a few places in or partially in Massachusetts are not included or only have partial data because that state does not participate in the Local Employment Dynamics (LED) data program, source of several important components of the index). The result is a set of urban form variables that can be used for analysis at the national scale (*20*).

The developers of the sprawl index drew extensively from existing research literature to identify elements to include in their index, and organized a wide range of measures into subcategories that measure different dimensions of urban form: development density, centering of population and jobs, land use mixing, and street accessibility. This analysis uses the latter two dimensions, which measure jobs accessibility and the connectivity and density of the street network—the measures directly related to the commute trip and to the areas of potential influence by transportation agencies. The documentation of the construction of the components notes the relationship of these two factors to travel behavior. Each dimension is constructed from multiple metrics:

1. Land Use Mixing Factor (MIX):
    - job-population ratio at the block group level, weighted by the sum of jobs and residents as a percent of the county total; averaged for the county
    - an entropy measure of the variation in jobs by employment sector (retail, entertainment, health, education, personal services), weighted by the block group share of the county total population and employment
    - Walk Score values for Census tracts, weighted by the tract share of the county total population and employment

2. Street Accessibility Factor (STRTACC):
    - average street block size; excluding rural blocks defined as larger than one square mile
    - percentage street blocks that are less than one-hundredth of a square mile in size
    - intersection density for urban and suburban census tracts; excluding rural tracts with population densities of less than 100 persons per square mile
    - percentage of intersections that are 4-or-more-way; excluding rural tracts

For each Factor, principal components were extracted from that Factor's set of metrics. These principal components thus represent the vector that captures the greatest amount of variation in the dataset. The first principal component for each Factor is then used to represent that Factor. For the Street Accessibility Factor, the first component has an eigenvalue of 2.39 and explains almost 60% of the variation in the data. For the Land Use Mixing Factor, the first component has an eigenvalue 2.30 and explains just over 75% of the variation in the data. Each first component is then standardized with a mean of 100 and a standard deviation of 25. The standardized values are reported as the Factor scores for counties and for MSAs. Similar pre-processing was carried out for the other two Factors. Then, to create a composite sprawl index, the four Factors were summed and the summed values again standardized with a mean of 100 and a standard deviation of 25. Full details of the construction of the sprawl index are provided in Ewing and Hamidi (*21*).

The substantial amount of preprocessing involved in developing these measures has advantages and disadvantages. Using a principal components analysis approach is a strategy to reduce the number of variables involved while retaining the variation in the original data. However, this step can be affected by outliers in the data. The construction of the Factors also assumes that the metrics used are valid measures of the respective construct; unexplained variation is not a critical element that is being left unaccounted for and metrics are not imperfect proxies of the true environmental factor. The process of standardization may yield a distribution that is easier to work with in quantitative settings; however, it may also unduly smooth the data and mask important characteristics of the data.

The data source for commuting and demographic variables is the American Community Survey (ACS) 5-year estimates for 2006-2010 for US counties (*22*). This introduces potential temporal mismatch with the sprawl Factors, however it is unlikely that there were dramatic changes in urban structure at the county level over this brief period. Shorter time-frame estimates are provided by the ACS (1- and 3-year) which would allow a closer temporal match with the 2010 urban form data; however, those data rely on much smaller samples and in many cases have very large margins of error. This is particularly a problem for modeling mode choice because in many places, cycling and walking are 'rare behaviors' and the small-sample estimates cannot be considered reliable (see *23*). For this analysis, the need for precise temporal congruency was balanced against the interest in reliability across a wide range of geographies.

The ACS is also the source for data on the number of workers and whether their commute is to a job site outside their county of residence. It is important to note that income is included by using the number of workers in the county, categorized by income level compared with the federal poverty threshold (*24*). Rather than a raw figure for dollar income, this takes into account the household size and composition, adjusted to the rate of inflation. This measure of income is perhaps more meaningful than a simple dollar figure in that it relates income to a measure of household need and 'basement level' expenditures for household needs.

For the analysis, the full set of 994 counties for which sprawl data were available was reduced. First, counties where data limitations did not allow the calculation of all sprawl index components (mostly counties in Massachusetts) were removed. Second, because a valid representation of mode split is an important part of the analysis, counties that had unacceptable margins of error for ACS variables were also removed. Specifically, those counties where the ACS estimates for walking or cycling had margins of error that crossed zero were removed from the analysis set (a negative mode share is not possible). Because the analysis is inherently spatial, and spatial autocorrelation effects were anticipated, MSAs comprised of a single county were also removed. These spatial 'islands' can bias coefficients in spatial models. Rather than mixing these single-county cases in with multi-county MSAs where interaction effects among counties are likely both in terms of land use and commuting behavior, it would be more appropriate to treat them as a different class of MSAs and model these cases separately. This left 699 counties in the continental US for the analysis with good representation nationwide.

## ANALYSIS

This study uses a series of regressions to investigate the relationship between demographic and commuting variables and two index variables of urban form. The sequential models also allow for evaluation of the robustness of estimates across specifications. Models are estimated using GeoDa 1.5.37 (2014 beta release) and GeoDa Space 0.8.7, both available for download from the Arizona State University's GeoDa Center for Geospatial Analysis and Computation (*25*).

A model is specified to estimate the contribution of increasing the number of workers, classified by income, to the overall share of workers who commute by driving alone. Two sprawl factors are used, along with a Gini index to measure the difference in urban form within each MSA. The Gini index is constructed to account for possible differences between the location of residence and the location of work (as well as the counties through which a worker might commute). A better measure would be to include the Sprawl Index values for both county of residence and county of employment, however ACS data are limited to providing county of residence along with the number of workers who commute to a county outside their county of residence. Commonly used to measure income inequality, the Gini index is the share of total income earned by each cumulative segment of the population (see *26*). In this application, it measures the variation of the Overall Sprawl index within the group of counties in each MSA. The Gini index is based on the composite sprawl index and is included to capture a wide range of spatial characteristics of the MSA, not only of the county of residence.

The regression takes the form:

$$\text{PCTWRKRSOLO} = \alpha + \beta(\text{MIX}) + \beta(\text{STRTACC}) + \beta(\text{GINI}) + \beta(\text{PctWrkOutC}) +$$
$$\beta(\text{Tot\_100Pov}) + \beta(\text{Tot100\_149Pov}) + \beta(\text{Tot150Plus}) + \varepsilon$$

Variable definitions and descriptive statistics are provided in Table 1.

**TABLE 1 Variable definitions and descriptive statistics (N = 669)**

| Variable Name and Description | Min | Max | Mean | Median | St. Dev. |
|---|---|---|---|---|---|
| PCTWRKRSOLO (% of workers for whom poverty status is calculated who commute by driving alone) | 6.93 | 89.04 | 79.56 | 81.12 | 7.53 |
| MIX (Land Use Mixing Factor) | 35.65 | 177.53 | 107.91 | 110.29 | 20.16 |
| STRTACC (Street Accessibility Factor) | 51.82 | 230.33 | 103.81 | 100.50 | 24.99 |
| GINI (Gini index calculated using composite sprawl index values for each MSA where 100 = perfect inequality and 0 = perfect equality) | 0.00 | 22.57 | 7.53 | 8.10 | 5.45 |
| Tot_100Pov (# of workers with income below 100% of poverty threshold, in 100s) | 0.45 | 3210.51 | 93.43 | 44.00 | 1187.52 |
| Tot100_149 (# of workers with income from 100 – 149% of poverty threshold, in 100s) | 0.75 | 3752.92 | 93.08 | 41.79 | 206.57 |
| Tot150Plus (# of workers with income above 150% of poverty threshold, in 100s) | 35.37 | 36860.40 | 1355.84 | 652.47 | 2369.35 |
| PctWrkOutC (% of workers who commute outside their county of residence) | 1.50 | 81.80 | 35.33 | 36.90 | 19.06 |

## RESULTS

### OLS Model

The first model estimated is an OLS model, which has an adjusted R-square of 0.374. Signs are generally as expected, with higher STRTACC values having a negative effect on the percent of workers commuting by driving alone (see Table 2). The MIX variable, however, is only very marginally significant ($p = 0.105$). By income group, adding workers earning below 100% of the poverty threshold reduced the overall share of workers who commute by driving alone, while those earning 100-149% of the poverty threshold increased PCTWRKRSOLO. This is in keeping with established research on the higher share of low income workers who use transit or other non-auto modes to commute, and lower car ownership rates for these households. The effect of GINI is also negative and significant. Because GINI increases as diversity among counties in an MSA increases, the results indicate that metros with greater diversity in sprawl scores have reduced shares of commuters who drive alone to work. This is likely capturing the effect of cities with very dense cores which are very different from the balance of the metro. However, the share of workers who commute outside their county of residence is insignificant, which may simply reflect the irrelevance of county boundaries, which are political and governance units, for meaningfully defining commuting zones in most MSAs. It is important to note that because of the wide range of size of counties this variable cannot be interpreted as a distance effect.

**TABLE 2 OLS Model Results**

| Variable | Coefficient | Std. error | t-Statistic | Probability |
|---|---|---|---|---|
| Constant | 98.809 | 1.678 | 58.894 | 0.000 |
| MIX | -0.0271 | 0.0167 | -1.621 | 0.106 |
| STRTACC | -0.125 | 0.013 | -9.353 | 0.000 |
| GINI | -0.168 | 0.055 | -3.035 | 0.003 |
| Tot_100Pov | -0.062 | 0.010 | -6.220 | 0.000 |
| Tot100_149 | 0.047 | 0.017 | -2.218 | 0.034 |
| Tot150Plus | 0.0004 | 0.0003 | 1.092 | 0.275 |
| PctWrkOutC | -0.036 | 0.017 | -2.218 | 0.034 |

Dep. variable: PCTWRKSOLO N = 669 Adj. R-Square = 0.374 F-statistic: 58.092

Model diagnostics for the OLS model are highly significant for heteroskedasticity of errors (Jacque-Bera) and for spatial effects (Lagrange Mulitplier). The latter indicate the spatial lag method rather than a spatial error method is appropriate for addressing the spatial structure in the data. Shifting to a spatial lag model involves creating a spatial weight matrix and adding a variable to the regression (WY) to estimate the spatial dependence present in the dependent variable. For this model, weights for the spatial weight matrix are based on a first-order contiguity, with each county assigned those counties with which it shares a border as its neighbor(s). These weights are then normalized for each county (*27*; note that a second-order contiguity criterion made the geographic extent of the contiguous counties implausibly large for commuting interaction effects in Western states with large counties).

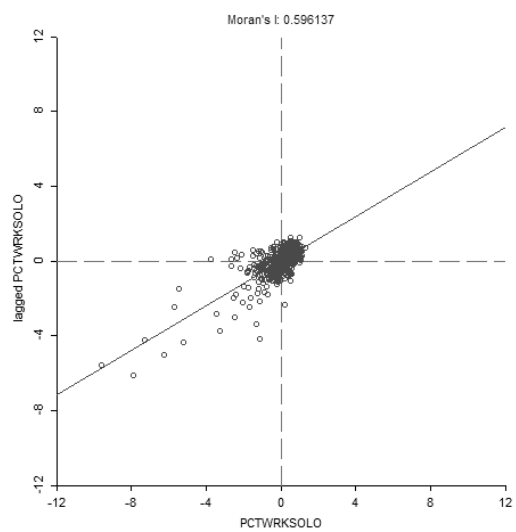


**FIGURE 1 Moran's I scatter plot (X-axis = dependent variable; Y-axis = lagged dependent variable).**

Further information on the spatial structure in the data is seen in the Moran's I scatter plot, which reveals a series of counties with 'low-low' values; that is, counties with low values for the dependent variable are located near other counties with low values (lower left quadrant, Figure 1). These observations are counties located in several of the most urbanized MSAs in the country including New York, Boston, San Francisco, and Washington DC. Treating these cases as true outliers and removing them from the dataset was explored; however, this step did not fully correct the pattern as another set of lower sprawl counties, frequently from the same MSAs as those removed, emerged as extreme low-low observations. This suggests the data may contain subclasses of metro regions with meaningful differences for commute mode choice, and that perhaps model fit would be improved using a classification of metros rather than a pooled regression.

**Spatial Lag Model**

The spatial lag model, shows improved overall fit compared with an OLS model using the same variables (with the addition of *WY*). Log likelihood improves from -2138.78 to -1987.08; the Akaike Information Criterion (AIC) and Schwarz criterion drop by similar magnitudes. (Spatial lag models yield a pseudo R-square value which cannot be compared with OLS R-square values.) Tests for heteroskedasticity of errors remain significant, indicating persistent issues of non-normality of errors. The error issue is not, however, attributable to spatial effects. The Moran's I for the spatially lagged residuals for this second model are essentially zero (z-value of 0.00104); as expected the spatially lagged dependent variable eliminated the spatial autocorrelation (*28*).

Also as expected the spatial lag term is highly significant with a coefficient value of 0.579, indicating that the spatial structure of the data is picked up by this variable (see Table 3 for full model results). At the same time the GINI variable becomes insignificant; the contribution of this variable is captured in the spatial lag term, as it should be. The other geographic variable, PctWrkOutC, also becomes insignificant, suggesting the effect of inter-county commuting is accounted for in the spatial lag term. This may be an effect of small counties in large metro regions, where inter-county commutes are common and the urban form characteristics in these core counties and the counties immediately adjacent are quite similar in terms of the effect on the choice to commute by driving alone. The Street Accessibility Factor remains highly significant, although its effect is diminished; the effect of the street system is partially accounted for in the spatial lag term. Still, every increment increase in this Factor is associated with a marginal decrease in the share of workers who commute to work by driving alone by nearly 0.10%. On the other hand, the Land Use Mixing Factor proves unstable; it becomes insignificant (p-value = 0.241).

**TABLE 3 Spatial Lag Model Results (Maximum Likelihood Estimation)**

| Variable | Coefficient | Std. error | z-Value | Probability |
|---|---|---|---|---|
| Lag_PCTWRKSOLO | 0.579 | 0.023 | 25.009 | 0.000 |
| Constant | 43.416 | 2.435 | 17.831 | 0.000 |
| MIX | 0.014 | 0.012 | 1.172 | 0.241 |
| STRTACC | -0.953 | 0.0099 | -9.617 | 0.000 |
| GINI | -0.061 | 0.040 | -1.523 | 0.128 |
| Tot_100Pov | -0.066 | 0.007 | -9.079 | 0.000 |
| Tot100_149 | 0.047 | 0.007 | 7.056 | 0.000 |
| Tot150Plus | 0.0008 | 0.0003 | 3.239 | 0.0012 |
| PctWrkOutC | -0.014 | 0.012 | -1.139 | 0.255 |

Dep. variable: PCTWRKSOLO N = 669 Log likelihood: -1987.08

This instability is in contrast to the variables for workers by income group. The significance level and coefficient for the poorest workers (Tot_100Pov), and workers in the middle income group (Tot100_149) are consistent across the OLS and spatial lag specifications. The highest income group, those at 150% or more of the poverty threshold, become significant in the lag model with a small but positive effect on the share of workers driving alone to work. It is likely there is a spatial component to the distribution of higher income workers, which is better specified in the spatial lag model, allowing the marginal effect of increasing the number of workers in this income class to become apparent. Interestingly, the small coefficient signals these workers do not overwhelmingly choose to drive alone. Whether this is because they have more alternatives from which to choose as a result of the choice of residential or work locations, or because of some preference structure of this population, cannot be determined from these results.

In order to identify potential improvements to the model specification, the error plots were explored. Many of the large and negative residual values are for highly urbanized counties, the same counties identified as having extreme low-low values in the Moran's I plot (Figure 1). This persistent pattern suggests there may be meaningful subgroups in the data. For models of mode choice, these groups could perhaps be defined using transport system variables that more explicitly account for differences in the mode choice set for commuters in different counties and MSAs. Thus the third model takes a spatial regime approach to estimate a two-stage least squares model with the counties grouped.

**Spatial Regimes Model**

While the Street Accessibility Factor focuses on characteristics associated with walkable access or perhaps bikability, it does not include any measures of transit. Naturally, transit availability will be lower in places at the low end of the scale and high only in places at the high end, where high density, high street connectivity, high centeredness, and high land use mixing provide a context more conducive to providing robust transit service that will attract commuters. Still, if transit is a defining characteristic of types of urban places and environments that shape commute mode choice, adding a categorical variable for transit may provide a meaningful separation of counties. This may lead to model results that better tease out the effects of the Street Accessibility and Land Use Mixing Factors.

To implement this, a third model is developed with a variable indicating the presence or absence of heavy rail. Heavy rail transit is defined as public transit service operated over fixed tracking, in a dedicated right-of-way, operating equipment that can carry heavy passenger volumes (*29*). Heavy rail transit is associated with the availability of a suite of complimentary transit services including bus and streetcar. The availability of high-volume transit in at least part of the metro, along with a range of complementary/feeder services supports the idea that places with heavy rail transit are in some ways categorically different from non-heavy rail metros. For this reason, all the counties in a metro where heavy rail was operating in <u>any</u> county in that metro was included in the heavy rail group. In all, 8 metro regions had heavy rail transit in 2010; 94 counties were located in metros where heavy rail transit operated (*30*).

The advantage of the spatial regime approach is that it allows direct comparison of coefficient estimates across two categories of observations, which is not the case when two groups of observations are modeled separately. Thus it provides additional information about the robustness and stability of the variables used. In addition to different coefficient estimates for the two groups, Chow tests indicate whether estimates are significantly different, either in their significance level or their coefficients.

Table 4 reports results of the spatial regime model. The global Chow test is highly significant, indicating a significant difference between the models for heavy-rail metros versus metros without heavy rail. The effect of the Street Accessibility Factor is significantly different for the two categories of metros: for heavy rail metros, -0.193, for non-heavy rail metros, -0.029. The effect is somewhat less significant for the latter category (p-value = 0.0204).

The results for the Land Use Mixing Factor are different for the two categories at the 0.01 significance level. This variable is insignificant for non-heavy rail metros. For heavy rail metros, it is highly significant, with a positive effect on the share of workers commuting by driving alone. This counterintuitive result may stem from the ability of this variable to only measure county-level employment opportunity, without providing information to explain commuters' choices to drive alone. The GINI variable remains insignificant for both categories; its effect is likely still captured in the sizeable coefficient for the spatial weight matrix (Lag_PCTWRKSOLO).

Notably estimates for changes in the number of workers by income are largely consistent with the previous specifications. They are significant for both county subgroups and for all income categories. Comparing groups of counties, the Chow test finds a significant difference for workers below 100% of the poverty threshold. This difference is attributable to the difference in the size of the coefficients: the marginal reduction in the share of commuters who drive alone to work is nearly twice as large in heavy rail metros compared with the non-heavy rail group (-0.088 and -0.047, respectively), reflecting the greater share of the lowest-income workers who use less costly modes for commuting. Thus, being in an MSA where heavy rail transit operates clearly provides more commuting options for this group.

The differences for the other two income levels do not show a strong difference across groups of counties, although the coefficients do vary somewhat. Adding workers with incomes at 100% to 149% of the poverty threshold results in a greater marginal increase in the share of all workers who commute by driving alone. The effect is similar for those earning 150% or more of the poverty threshold. Overall, the results for the income category variables are robust across the models.

Despite these differences visible across the categories of counties, the value of the coefficient for the spatial lag term is quite high, indicating that there is a substantial increase in the size of the coefficients when the spatial effect is combined with the marginal effect (coefficient value). This outcome, coupled with the persistent heteroskedasticity of errors, suggest remaining specification problems, which could flow from nonnormal distributions or nonlinear relationships in the underlying data, error in the Census estimates or the sprawl Factors, or from any of the various issues that commonly plague models using spatial data (*31*).

**TABLE 4 Results of Spatial Two-Stage Least Squares with Regimes**

| Variable | Coefficient | Std. error | z-Statistic | Probability |
|---|---|---|---|---|
| **Counties in Non-Heavy Rail Metros (N = 575)** | | | | |
| Constant | 22.963 | 8.258 | 2.7811 | 0.0054 |
| MIX | 0.0073 | 0.014 | 0.527 | 0.598 |
| STRTACC | -0.029 | 0.012 | -2.319 | 0.0204 |
| GINI | -0.010 | 0.041 | -0.245 | 0.806 |
| Tot_100Pov | -0.047 | 0.012 | -3.923 | 0.00008 |
| Tot100_149 | 0.033 | 0.0104 | 3.201 | 0.0014 |
| Tot150Plus | 0.0008 | 0.0004 | 2.199 | 0.0279 |
| PctWrkOutC | 0.0124 | 0.01199 | 1.036 | 0.3004 |
| **Counties in Heavy Rail Metros (N = 94)** | | | | |
| Constant | 25.952 | 10.419 | 2.491 | 0.0128 |
| MIX | 0.1043 | 0.035 | 2.970 | 0.00297 |
| STRTACC | -0.1933 | 0.0349 | -5.527 | 0.000 |
| GINI | -0.044 | 0.108 | -0.409 | 0.682 |
| Tot_100Pov | -0.0879 | 0.0148 | -5.951 | 0.000 |
| Tot100_149 | 0.060 | 0.0105 | 5.738 | 0.000 |
| Tot150Plus | 0.00158 | 0.0004 | 3.5456 | 0.00039 |
| PctWrkOutC | 0.0835 | 0.0439 | 1.9027 | 0.0571 |
| **Global** | | | | |
| Lag_PCTWRKSOLO | 0.7399 | 0.0959 | 7.719 | 0.000 |

Dep. variable: PCTWRKSOLO

**Future Extensions**

There are extensions of these models that could provide further insights. First, to investigate possible scale mismatch issues, the analysis could be repeated at the metro scale. It may be that the characteristics of the metropolitan region as a whole are a better predictor of travel behavior choices because it better represents the spatial processes at work in commuting decisions (*32*). Second, alternatives to incorporating measures of transit service or accessibility into the analysis should be explored, perhaps as an additional independent variable. This appears to have important implications the performance of the sprawl index Factors and certainly has important implications for the available choice set for commuters. Further, without explicitly including some measure of transit, it is impossible to know if the effects of the sprawl indices are partial proxies for a multi-modal transportation system. Other characteristics of interest could include more demographic variables such as household characteristics, vehicle ownership, and a fuller accounting of mode split, although such variables may introduce complex collinear relationships. Finally, a different spatial weight matrix could improve the model, perhaps one defining the extent of spatial interaction as occurring among all the units in each metro but not among metros (*27*).

## INTERPRETATION AND IMPLICATIONS

The results for income are found to be robust across various model specifications. This confirms a general relationship between income and commuting behavior that holds across metro areas. The results for the Street Accessibility Factor suggest the configuration and density of the street network is an important environmental element in the choice to commute by driving alone, with more compact street networks associated with reduced driving alone. This Factor may be of value in studies of the determinants of travel behavior, although in metros without heavy rail transit systems, it is somewhat less able to account for variation in the level of commuting by driving alone. It is likely that in analyzing the relationship to travel behavior, this Factor should be augmented with additional measures such as transit service or transit accessibility. The results for the Land Use Mixing Factor indicate this variable has limited power to explaining commuting behavior, despite its intuitive appeal as the land use component of the commute mode decision process. Perhaps it is best employed in combination with the other Factors from the sprawl index.

Without question, the sprawl index provides an intriguing dataset for researchers and practitioners seeking insights into the relationship between transportation and urban form. The data open the opportunity for enriched exploratory work. However, the variation in land use patterns and transportation systems, coupled with the highly pre-processed nature of the data, suggest there are important caveats to using the sprawl indices and component Factors.

Despite the massive effort of data compilation and processing represented in the sprawl index, it may or may not represent a definitive method for measuring the phenomenon of sprawl. It may be the case that sprawl represents a deeply complex phenomenon, perhaps with synergistic, nonlinear, and/or threshold effects. Such a view suggests that quantitative approaches that seek to isolate components so as to develop targeted interventions or policies face serious methodological challenges which could leave practitioners and decision makers with great uncertainty when confronted with sweeping arguments of advocates, whether pro- or anti-sprawl.

## ACKNOWLEDGMENTS

The author thanks Julia Koschinsky, Associate Research Professor at Arizona State University, for her most helpful comments on previous work from which this paper was developed. Thanks are also given to Reid Ewing, Shima Hamidi, their colleagues, and their funders for developing the sprawl indices, with particular appreciation for their making their data available to other researchers.

## REFERENCES

1. von Thünen, J. H. *Isolated state: an English edition of Der isolierte Staat*. Pergamon Press, Chicago, 1966.
2. Christaller, W. (1933). Die zentralen Orte in Süddeutschland: Eine ökonomischgeographische Untersuchung uber die Gesetzmässigkeit der Verbreitung und Entwicklung der siedlungen mit städtischen Funktionen. Presented in B. J. L. Berry and F.E. Horton. *Geographic perspectives on urban systems with integrated readings*. Prentice-Hall: Englewood Cliffs, NJ, Geographic, 1970, pp 171-173.
3. Ewing, R. and R. Cervero. Travel and the Built Environment. A Meta-analysis. *Journal of the American Planning Association*, Vol. 76, No. 2, 2010, pp 265-294.
4. Khattak, A. J. and D. Rodriguez. Travel behavior in neo-traditional neighborhood developments: A case study in USA. *Transportation Research Part A*, Vol. 39, 2005, pp 481-500.
5. Transportation Research Board. *Does the Built Environment Influence Physical Activity?* Special Report 282. Transportation Research Board and Institute of Medicine of the National Academies, Washington DC, 2005.
6. Ewing, R., K. Bartholomew, S. Winkelman, J. Walters and D. Chen. *Growing Cooler: The Evidence on Urban Development and Climate Change*. Urban Land Institute, Washington DC, 2007.

7. Bereitschaft, B. and K. Debbage. Urban Form, Air Pollution, and CO2 Emissions in Large U.S. Metropolitan Areas. *The Professional Geographer*, Vol. 65, No. 4, 2013, pp 612-635.
8. Kahn, M. E. and J. Schwartz. Urban air pollution progress despite sprawl: The “greening” of the vehicle fleet. *Journal of Urban Economics*¸ Vol. 63, Issue 3, 2008, pp 775-787.
9. Brabec, E., S. Schulte, and P.L. Richards. Impervious Surfaces and Water Quality: A Review of Current Literature and Its Implications for Watershed Planning. *Journal of Planning Literature*, Vol. 16, No. 4, 2002, pp 499-514.
10. Ewing, R., T. Schmid, R. Killingsworth, A. Zlot, and S. Raudenbush. Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity. *American Journal of Health Promotion*, Vol. 18, No. 1, 2003, pp. 47-57.
11. Feng, J., T.A. Glass, F.C. Curriero, W.F. Stewart, and B.S. Schwartz. The built environment and obesity: A systematic review of the epidemiologic evidence. *Health & Place*, Vol. 16, 2010, pp 175-190.
12. Eid, J., H.G. Overman, D. Puga, and M.A. Turner. Fat city: Questioning the relationship between urban sprawl and obesity. *Journal of Urban Economics*, Vol. 63, 2008, pp 385-404.
13. Cutsinger, J., G. Galster, H. Wolman, R. Hanson, D. Towns. Verifying the Mulit-Dimensional Nature of Metropolitan Land Use: Advancing the Understanding and Measurement of Sprawl. *Journal of Urban Affairs*, Vol 27, No. 3, 2005, pp 235-259.
14. Galster, G., R. Hanson, M. R. Ratcliffe, H. Wolman, S. Coleman , and J. Freihage. Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept. *Housing Policy Debate*, Vol 12, Issue 4, 2001, pp 681-717.
15. Sarzynski, A., G. Galster, and L.Stack. Typologies of sprawl: investigating United States metropolitan land use patterns. *Urban Geography*, Vol. 35, No 1, 2014, pp 48-70.
16. Kelly-Schwartz, A.C., J. Stockard, S. Doyle, and M.Schlossberg. Is Sprawl Unhealthy? A Multi-level Analysis of the Relationship of Metropolitan Sprawl to the Health of Individuals. *Journal of Planning Education and Research*, Vol. 24, No. 2, 2004, pp 184-196.
17. Brueckner, J. and A.G. Largey. Social interaction and urban sprawl. *Journal of Urban Economics*, Vol. 64, Iss. 1, 2008, pp 18-34.
18. Kamrazzman, M., L. Wood, J. Hine, G. Currie, B. Giles-Corti, and G. Turrell. Patterns of social capital associated with transit oriented development. *Journal of Transport Geography*, Vol. 35, 2014, pp. 144-155.
19. Ewing, R. and S. Hamidi. 2014. Measuring Sprawl 2014. Smart Growth America: Washington DC. 51 pp. Available at: http://www.smartgrowthamerica.org/documents/measuring-sprawl-2014.pdf Accessed 10 May 2014.
20. National Cancer Institute. *Updated Urban Sprawl Data for the United States*. Available at: http://gis.cancer.gov/tools/urban-sprawl Accessed 10 June 2014.
21. Ewing, R. and S. Hamidi. *Measuring Urban Sprawl and Validating Sprawl Measures.* Metropolitan Research Center. Available at: http://gis.cancer.gov/tools/urban-sprawl/ Accessed 10 May 2014.
22. US Census Bureau. 2014. American Community Survey Main. Website. Available at: http://www.census.gov/acs/www/ Accessed 6 June 2014.
23. Folch, D.C., D. Arribas-Bel, J. Koschinsky, and S.E. Spielman. *Uncertain Uncertainty: Spatial Variation in the Quality of American Community Survey Estimates*. Working

Paper Number 01. ASU GeoDa Center for Geospatial Analysis and Computation. Available at: https://geodacenter.asu.edu/drupal_files/2014-01.pdf Accessed 10 June 2014.

24. US Census Bureau. *How Poverty is Calculated in the ACS*. Available at: http://www.census.gov/hhes/www/poverty/poverty-cal-in-acs.pdf Accessed 6 June 2014.
25. Arizona State University. *GeoDa Center for Geospatial Analysis and Computation*. Available at: http://geodacenter.asu.edu/ Accessed 14 June 2014.
26. Sen, A. Informational Bases of Alternative Welfare Approaches. *Journal of Public Economics*, Vol. 3l, 1974, pp 387-403.
27. Getis, A. Spatial Weights Matrices. *Geographical Analysis*, Vol. 41, 2009, pp 404-410.
28. Anselin, L. *Exploring Spatial Data with GeoDa: A Workbook*. Center for Spatially Integrated Social Science, University of California, Santa Barbara, 2005.
29. American Public Transit Association. *Mode of Service Definitions*. Available at: http://www.apta.com/resources/statistics/Pages/glossary.aspx#8 Accessed 6 June 2014.
30. American Public Transit Association. *Public Transportation Ridership Report, Fourth Quarter 2010*. Available at: http://www.apta.com/resources/statistics/Documents/Ridership/2010_q4_ridership_APTA.pdf Accessed 30 May 2014.
31. O'Sullivan, D. and D. J. Unwin. *Geographic Information Analysis, Second Edition.* John Wiley & Sons, Hoboken, 2010.
32. Getis, A. Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, Vol. 37, 2007, pp 491-496.

Multilevel Governance and Development
Wirtschaftsuniversität Wien
Institutsvorstand : ao.Univ.Prof. Dr. Gunther Maier
Welthandelsplatz 1
A-1020 Wien, Austria
Tel.: +43-1-31336/4777 Fax: +43-1-31336/705 E-Mail: mlgd@wu.ac.at
http://www.wu.ac.at/mlgd