TRB Annual Meeting

Index of Employment-Worker Balance by Transit Station Mode

--Manuscript Draft--

Full Title:	Index of Employment-Worker Balance by Transit Station Mode
Abstract:	Employment-Worker Balance (EWB), a key to economic growth through agglomeration economies is also a key to social equity. This is due to its ability to both increase workers' access to employment and firms' access to a strong, diverse, and resilient workforce. Smart Growth advocates frequently identify Employment-Worker Balance as a key metric in compact urban design. Because of its potential synergistic effects with EWB, another key element of Smart Growth, Fixed-Rail Transit systems (FRT), needs to be studied for its effects on EWB: is the latter improved by the former, and for which job sectors and which workers? Principle Component Analysis will be used to produce a EWB Index that is able to map EWB across multifarious spatial contexts across the U.S., taking into its scope the multiple types of transit system modes, real estate types, and the many sectors of the economy that surround FRT stations. The EWB Index will provide a tool for practitioners and researchers to utilize in regression analysis, and policy and decision support. The paper will follow up on this significant increase of available evidence to work towards further theoretical refinement of EWB.
Manuscript Classifications:	Sustainability and Resilience; Transportation and Sustainability; Transportation and Economic Development AMS50; Economic Development; Land Use and Multiuse; Transit Oriented Development
Manuscript Number:	TRBAM-21-01296
Article Type:	Presentation and Publication
Order of Authors:	Robert E. Hibberd
	Arthur C. Nelson, PhD

Index of Employment-Worker Balance by Transit Station Mode

Robert E. Hibberd
School of Geography, Development & Environment
University of Arizona, Tucson, AZ, 85721
Email: rhibberd@email.arizona.edu

Arthur C. Nelson
Professor of Planning & Real Estate Development
College of Architecture, Planning + Landscape Architecture
University of Arizona, Tucson, AZ, 85721
Email: acnelson@email.arizona.edu

Word Count: 5,977 words + 6 tables & figures (250 words per table) = 7,477 words

Submitted July 31, 2020

ABSTRACT

Employment-Worker Balance (EWB), a key to economic growth through agglomeration economies is also a key to social equity. This is due to its ability to both increase workers' access to employment and firms' access to a strong, diverse, and resilient workforce. Smart Growth advocates frequently identify Employment-Worker Balance as a key metric in compact urban design. Because of its potential synergistic effects with EWB, another key element of Smart Growth, Fixed-Rail Transit systems (FRT), needs to be studied for its effects on EWB: is the latter improved by the former, and for which job sectors and which workers? Principle Component Analysis will be used to produce a EWB Index that is able to map EWB across multifarious spatial contexts across the U.S., taking into its scope the multiple types of transit system modes, real estate types, and the many sectors of the economy that surround FRT stations. The EWB Index will provide a tool for practitioners and researchers to utilize in regression analysis, and policy and decision support. The paper will follow up on this significant increase of available evidence to work towards further theoretical refinement of EWB.

Keywords: Transit-Oriented Development, income match, employment-worker balance index, principle component analysis

INTRODUCTION

Debate is long-standing and continuous regarding the relative merits of accessibility and mobility as separate paradigms of human flows. The EWB inherently is built upon a measure of access; however, the mode of travel between places changes the salient components of access. The built environment must support whatever transportation technology is chosen to reach one's destination. Some researchers assert the lack of a strong connection between accessibility and the built environment, citing instead the importance of individual behavior and constraints placed upon individuals by social structures (1,2). However, mobility is one way to increase accessibility. It is most available and enjoyed by those who can pay to use mobilizing infrastructures and technologies. These, moreover, are ever-costlier to jurisdictions as they are required to support the transport of ever-larger magnitudes of people across ever-longer distances. Sprawling development necessitates the near-complete human reliance upon the automobile-as-prosthetic-device, and thus the infrastructure to facilitate its use. The continuance of sprawling development will exacerbate financial excesses in real estate development and public infrastructure spending (3). Others emphasize the need to balance mobility and accessibility. The former is the ability to move about the region in order to access needed land uses; the latter is "the relative connectedness of an area" (4). Paez et al. (5) provide clarification in their definition, describing accessibility as "the joint result of a transportation network and the geographical distribution of activities." Regianni et al. (6) describe accessibility as "the potential of opportunity for interaction," which aids in economic growth. Mode of transportation is also critical. The empirical data measured by studies such as Ewing & Hamidi (7) and Levin (8) demonstrate the importance of the built environment to accessibility. Relative accessibility is measured by one's ability to utilize needed land uses.

QUESTIONS & HYPOTHESES

Can we directly measure the effect on EWB from transit stations via an index that is sensitive to different kinds and levels of transit across metros?

- Hypothesis 1: Different combinations of economic sectors and transit modes will load on different PCs
- Hypothesis 2: Different modes of travel will load differently on the Employment-Worker Balance Index (EWBI).
- Hypothesis 3: These differential loadings will produce a variety of EWB regimes or clusters.

LITERATURE REVIEW

Graaskamp (9) emphasized the value (situs) of a site as "related to the functional needs of the activity and not the site." Linkages between a site and the surrounding region facilitate accessibility, and the "costs of friction" are those of the stress, time, and costs accruing to each of the needed linkages for a specific activity (9). Employment-worker balance enhances those linkages between sites, both for the workforce and for the workplace, easing the costs of friction through greater accessibility. TOD enhances the EWB inasmuch as it is relevant, through built environment characteristics and transit node interconnectivity that draws people to utilize the site for both land uses and access to the regional transit network. Levine (8) argued that while "commute time remains a strong determinate of residential location at the regional scale," the salient improvement in accessibility accruing from an employment-worker balance derives from the increased match between workplace and residential location due to a greater range of options from which to choose in both housing and transportation. This principle is consistent with the theory of the gravity model, as a multitude of land uses in close proximity will increase the pull effect of a location. Relaxing zoning regulations that promote and subsidize spatially separated, single-use and lower-density development may promote better EWB, as demonstrated by Levine's (8) discrete choice model.

Worker accessibility is of significant value to both worker and employer. The "drive till you qualify" crowd living out in the suburbs or exurbs pay thousands of dollars more annually for transportation costs, when considering both monetary cost and time spent, and many choose to live nearer to work when given the option. The positive market response to development of residences nearer employment clusters negates the argument that a regulatory promotion of EWB is an interference with the market (8). Moreover, firms regularly demonstrate the importance of workforce accessibility to the health of the firm. Employers regularly place firms on the basis of an analysis of local workforce educational attainment and spatial concentration (10).

DATA & METHODS

The quantitative analysis of multivariate processes and phenomena in the social sciences requires the combination of many indicators of these phenomena, which further requires a structured paradigm or theory to both formulate and to interpret the analysis. A common definition of an index is informative: it is a measure of an abstract theory that combines multiple indicators.

Method examples for creation of indices in the literature include both Factor or Principle Component Analysis (PCA) and multiple regression. In the PCA realm, Ewing & Hamidi (7) use the Census Tract to provide local "sprawl-like" measures, applying their PCA methodology, which was originally used at the metropolitan area scale. The PCA combines many correlated covariates in vector space to reveal the latent processes jointly explained by the correlated variables. In the realm of regression, the U.S. Department of Housing and Urban Development's (HUD) Location Affordability Index (LAI) relies upon a complex Structural Equation model (SEM), which maps out direct and indirect causal pathways between endogenous variables.

The choice of either method requires weighing the tradeoffs of positive and negative aspects of each, given the unique requirements of each study undertaken. PCA is non-parametric, a probable source of advantage over regression. Regression modeling with fixed effects may provide some advantages over PCA, as it can control for noisy differences between unique places using various fixed effects (11). The EWBI will utilize the strengths of the PCA approach, and a further study in chapter 5 will follow up with a fixed effects spatial regression to evaluate the EWBI for applicability to various demographic groups across the US.

Transit systems for this study were derived from General Transit Feed Specification (GTFS) static files, which most transit authorities across the United States provide in accordance with the Google GTFS data standard. Transit authorities prepare their data about stops and routes along the various modes of public transportation available in their communities, including local, express, and rapid bus routes, commuter rail transit, light rail, streetcar rail, and heavy rail subway-metro systems. The GTFS standard tables were processed through ArcGIS Model Builder.

The study will review transit systems for the year 2016 in the cities of Atlanta, Cleveland, Eugene, and Minneapolis. These cases represent a wide variety in terms of region, population size, economy, transit modes (e.g., streetcar or bus rapid transit), and urban form. The study area is restricted to the U.S. Census Urbanized Area of the counties of the metropolitan area that are served by transit systems. The transit system modes for each city are as follows:

- Atlanta: streetcar (SCT), heavy rail transit (HRT)
- Cleveland: light rail (LRT), bus rapid transit (BRT)
- Eugene: bus rapid transit (BRT)
- Minneapolis: LRT, BRT & commuter rail (CRT)

Commutesheds from LEHD Origin-Destination Tables

The data tables for jobs and workers were gathered from the U.S. Census Bureau's Longitudinal Employment-Housing Database (LEHD) job data tables for census blocks were downloaded from the U.S. Census Bureau's On the Map website in shapefile format. The LEHD Origin- Destination Employment Statistics (LODES) tables provide full counts, rather than samples, of wage and salary jobs covered by unemployment insurance, with strict enforcement of privacy for individual respondents. These tables provided the variables for study about the location of jobs and their pay level, as well as workers and their pay scale. The former are found in the Work Area Characteristics (WAC) files, detailing the workplace location and other data for the employees that are enumerated in the file. Jobs totals are provided, along with a breakout of jobs by age of employee, by pay ranges, and by jobs according to the North American Industry Classification System (NAICS) job sector categorization. The Residence Area Characteristics (RAC) file provides data on the residence location of workers, including the same variables as the WAC file, but from the basis of the residence location of the enumerated workers, which may or may not include the residence census block. Benner & Karner (12) point out the limitations of the LEHD earnings classification, including the lack of an index to inflation and the significant variation in the number of workers who fall into each category as one controls for metropolitan statistical area. This study will utilize a classification of income based on NAICS job sectors, following Nelson and Ganning (13). Street and intersection data will come from the U.S. Census Bureau's TIGER Line data set, with post-processing done in GIS.

Commuteshed sums of workers for a dissimilarity index and internal capture ("residence ratio" in Kain (14)) is measured using an origin-destination cost matrix, which maps the Euclidean distance from each origin to each destination to which it is connected. The distance method is a 3-mile cutoff.

The commuteshed is derived in GIS by a search from each origin census block group to all CBGs listed as destinations. An origin-destination cost matrix selects all destinations within the 3-mile Euclidean distance threshold. A one-to-many relationship exists between origins and destination. Therefore, the cost matrix provided the required lookup table between origins and destinations. Summing the workers in the commuteshed of each CBG was calculated as follows:

$$workers\ in\ CBG\ commuteshed \;=\; \sum_i^n \sum_j^m C_{ij} \qquad (1)$$

Where the total number of workers commuting from an origin i, to a destination, j, or C_{ij}, within the 3-mile range is considered part of the commuteshed. The number of origins, i, is denoted by n, and m is the number of destinations j per origin, i. This calculation is done by summing each origin-destination CBG pair, and then again for the origin CBG. The origin census block group ID provided the basis for summing those workers working at job sites within about 3 miles from home. This method was used for both the numerator and the denominator in the internal capture equation. Each census block group gets evaluated for the number of workers at each destination, and a sum is made of workers who live or work and who both live and work in the commute shed. This enables use of the equation for internal capture in each cluster:

$$internal\ capture = \frac{2 * workers\ living\ \&\ working}{workers\ living + workers\ working} \qquad (2)$$

Commuteshed sums for the dissimilarity index were calculated with the same method. The dissimilarity index gives a measure of distribution or concentration of subsets of a data population. This study applies it to the level of "income match" in a CBG commuteshed. Income match (15) determines the relative balance in a location of workers and relevant housing by income. The dissimilarity index is computed thusly:

$$DI = 0.5 \sum_{i=3}^{N} \left| \frac{r_i}{R} - \frac{w_i}{W} \right| \qquad (3)$$

Where r_i is the number of workers of a given income level subset residing in the commuteshed, R is the number of all workers residing in the commuteshed, w_i is the number of workers of a given income level subset working in the commuteshed, and W is the number of all workers working in the commuteshed.

Spatial Cluster Analysis – Identifying Centers & Sub-Centers

Centers of employment and residential land use will be identified through spatial cluster analysis, which relies upon spatial dependency or autocorrelation between objects in terms of one measured variable. Moran's I, a global measure of spatial autocorrelation, a spatially-weighted version of the Pearson correlation coefficient (16), is the most appropriate analysis to begin with, as it determines overall levels of spatial clustering in a given region or total study area. Then, if it identifies statistically significant clustering, this finding indicates that more neighborhood-level measures can be used (and at what distance band), such as the Getis & Ord Gi* statistic, which identifies neighborhood-level hot or cold spots of a given variable, assigning z scores and p values for quantification. The most intense employment cluster in the region is the CBD (17).

Moran's *I* is defined as

$$I = \left(\frac{1}{s_y^2}\right) \frac{\sum_i^N \sum_{\{j:i \neq j\}}^N w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i^N \sum_{\{j:i \neq j\}}^N w_{ij}} \tag{4}$$

Where $\bar{y} = \sum_i {y_i}/{N}$, $s_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$. y_i are counts, although alternative versions of Moran's *I* utilize continuous values (16). The metric provides a cross-product, as it sums the covariance between each point and each of its neighbors, providing the sum of covariance (deviation from the mean at y_i multiplied by the deviation from the mean at y_j) for all sets of adjacent neighbors, and then it divides it by the global variance, s_y^2. The resulting index ranges between -1 and 1, from a spatially dispersed pattern, to a spatially clustered one. This metric can be used at various distance bands, defined in the equation by assigning all features within the desired distance band a value of 1 in the matrix, w_{ij}. The various peaks in the score can represent neighborhoods in which the underlying spatial associations are strongest, and it is not necessarily true that each phenomenon has only one peak.[1] The researcher may then choose the peak distance band at which the phenomenon being studied is operative (**figure 1**).

The Getis & Ord Gi* metric measures the degree of association resulting from the concentration of weighted points or areas and the other weighted points or areas within a given neighborhood, which is defined by distance *d* from the origin *i*. The Gi* metric is defined as follows,

$$Gi^*(d) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j} \tag{5}$$

Where w_{ij} is the matrix of weighted points within each neighborhood, $w_{ij}(d)$. The matrix is a set of binary values designating whether each location *j* is within distance *d* of the origin location *i*. Each weighted point has the attribute value, x_i or x_j. The metric has a null hypothesis of spatial independence (18). Moran's *I* is a useful starting point for using local-scaled metrics of spatial association, by defining distance bands at which association may be strongest. This distance then becomes the definition for the neighborhoods in the Gi* statistic (distance *d* in equation 2 above).

Centering is evaluated using non-parametric global and local spatial autocorrelation metrics, and sub-centering is evaluated using non-parametric geographically weighted regression (GWR). It fits a separate regression model to each observation according to a sample of observations taken from a neighborhood kernel centered on the observation. The kernel can be fixed in size or adjusted at each observation for size to capture *k* observations to make the sample sufficiently large. The result is a set of unique coefficients and error terms associated with each

[1] ESRI ArcGIS Desktop Help. "Incremental Spatial Autocorrelation." Accessed 7-25-2017. http://desktop.arcgis.com/en/arcmap/10.4/tools/spatial-statistics-toolbox/incremental-spatial-autocorrelation.htm.

observation in the study sample. This produces a local statistic that answers for a lack of structural integrity in some explanatory variables that vary significantly across space. It is specified thusly:

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)\, X_{ik} + \varepsilon_i \quad i = 1, \ldots, n \tag{6}$$

It answers for spatial non-stationarity by fitting a regression to each observation, *i,* estimating the dependent variable by estimating a constant, β_0, and a vector of parameters, β_k, at each spatial location designated by the coordinates of i, (u_i, v_i). The error term is also determined for each location, i. The parameters $\beta_k(u_i, v_i)$ are estimated by adding a spatial weights matrix, $W(u_i, v_i)$ to the traditional OLS parameter estimation equation. The **W** matrix models spatial relationships using either a fixed or adaptive continuous kernel, with a distance decay function estimating the spatial relationship. Common forms are a Gaussian distance decay or bisquare weighting function. The adaptive kernel chooses a varying distance bandwidth in order to capture the same number of nonzero weights per observation i (19).

To evaluate centering, one identifies and culls those centers with employment that consist of more than 75% of all jobs in one single economic sector (20). This eliminates large single-sector land uses. Standard deviations of the Gi* statistic will be used to evaluate the magnitude of centering. Positive residuals from GWR regression of employment density on distance from CBD represent sub-centers. One threshold employment density for sub-centers is 20 jobs per acre. Alternatively, the magnitude of the positive residuals, proxies for the intensity of centering, with 2.5 standard errors being used in the literature as a cutoff (17,20). The study will use the latter approach to generalize the results to multiple levels of density across the urban hierarchy.

TABLE 1. Place-Based Job Sectors in the Study Allocated by Wage Category

NAICS	Description	Mean Annual Wages, 2013	Wage Category	Share of Jobs
44	Retail Trade	$25,779	Lower	
56	Administrative, Support, Waste Mgmt., Remediation	$35,931	Lower	
61	Educational Services	$35,427	Lower	
71	Arts, Entertainment and Recreation	$32,188	Lower	
72	Accommodation and Food Services	$17,453	Lower	
81	Other Services (except Public Administration)	$29,021	Lower	
	Weighted Mean Wages and National Share of Jobs		***~$30,000***	***~33%***
48	Transportation and Warehousing	$45,171	Middle	
53	Real Estate and Rental and Leasing	$46,813	Middle	
62	Health Care and Social Assistance	$44,751	Middle	
92	Public Administration	$51,340	Middle	
	Weighted Mean Wages and National Share of Jobs		***~$50,000***	***~33%***
22	Utilities	$94,239	Upper	
31	Manufacturing	$54,258	Upper	
42	Wholesale Trade	$65,385	Upper	
51	Information	$83,677	Upper	
52	Finance and Insurance	$88,677	Upper	
54	Professional, Scientific and Technical Services	$75,890	Upper	
55	Management of Companies and Enterprises	$105,138	Upper	
	Weighted Mean Wages and National Share of Jobs		***~$70,000***	***34%***

Source: Adapted from (13).

Principle Component Analysis

Factor analysis reveals latent variables from a series of highly correlated observed variables. One major variant, Principle Component Analysis (PCA), reduces the noise in a set of correlated variables, revealing with greater clarity the underlying signal of the phenomena being explained by the variables by highlighting their shared variance and removing white noise. Each component is a group of correlated variables that load highly on the component, meaning they are closely related to it. It is also known as Empirical Orthogonal Functions (EOF) due to non-parametric fitting of eigenvectors to highly correlated covariates. Each EOF is orthogonal to the others.

The two most common applications of PCA involve explaining the variance across 1) a time series (21), or 2) a set of variables (22). The first approach reduces the noise in a single variable across many time periods, while the second approach reduces noise for many variables for a single-year cross-sectional data set. The study relies upon PCA to reveal job growth dynamics in redlined zones and nearby transit stations, hypothesizing that transit proximity increased job growth over the years of the study, but did not provide this benefit to all segments of the population.

Table 2 lists the variables to be used in the PCA, with justifications from the literature. The variables' data sources include the US Census Bureau's ACS and LEHD data sets, the GTFS transit data format, and various GIS and R-derived spatial processes.

The method consists of taking four matrices to translate from variables to component scores for each observation in space. The process begins (**figure 3**) with a) the base matrix of N observations by M variables translated into standardized z-score format, also known as a centered matrix, and next, b) the correlation coefficient matrix is taken, producing a square matrix of M dimensions. Then, c) the singular value decomposition method (SVD) creates matrices of eigenvectors (as wells as eigenvalues in a separate matrix). The eigenvalues are presented as the diagonal of the **S** matrix in SVD, and the proportion between each value of the diagonal and the trace (the sum of the diagonal) provides the explained variance of each eigenvector. Finally, d) the eigenvectors, found in the **U** matrix of the SVD output, are used as weights in linear combination with the original data values to produce the components (i.e., signals or scores). The individual elements of the components in (d) are also known as score coefficients.

The literature cites four reasons to rotate the vector space (a linear transformation process). Rotation provides such advantages as insensitivity to the shape of the spatial domain and subdomains underlying the data, no trouble with sampling error, and an accurate picture of physical relationships within the original data (23). Also, rotated space typically increases the explained variance and the loadings of each component on the relevant original variables, by better fitting the components to the data (24). Further, rotating and gaining a better fit of the components to the variable vectors also reduces secondary loadings of the variable vectors on the components, in effect causing each factor to represent one process captured by the clustering (i.e., correlated) variable vectors (22,23).

TABLE 2. Variables Included in EWBI, Justification & Literature Sources

Variables	Description & Justification	Selected Literature
Distance to nearest FRT stations by mode and station's centering intensity (GWR score)	Vector of measures of node and place attributes of neighborhood transit stations.	(25), (17,20,26)
Place-Based Employment by sector group, age, and income	Categorized vectors of employment, they are a necessary input to capture demographic interactions. Firms compete for location.	(13,14,27)
Place-Based Workers by sector group, age, and income	Vector of workers. Worker demographics greatly affects commute.	(13,14)
Housing by tenure and quality	Vector of housing. Renters can move residences easier than owners. Housing quality increases in newer development, with fewer vacancies.	(15)
Commuting mode, time	EWB highly dependent upon commute mode and shed, a vector measured in time or distance. Link between station proximity and mode choice to work.	(15,25)
Vehicles per Household	Vector proxy for automobile dependency	(25,28)
Intersection density	A measure of urban compactness and walkability.	(7,15)
Road network density	A measure of urban compactness and walkability.	(15,27)
Strength of employment density; centering/subcentering	Higher EWB results in lower VMT and VHT and facilitates substitution of other travel modes for the automobile.	(7,17,20,28)
Strength of housing density	Higher EWB results in lower VMT& VHT and facilitates substitution of other travel modes for the automobile. Clustering of housing should increase EWB.	(7,17,20)
Distance to CBD	Regional context of the neighborhood	(20,27)
Dissimilarity Index of income match for place-based jobs	Degree to which the neighborhood employment sectors is matched with workers' job sectors	(15)
Internal capture	Workers living and working in the same commute shed as % of total	(14,15,20)

Diagnostic tests for the PCA include Kaiser's Criterion, which calls for keeping only those components that have an eigenvalue of 1 or higher, due to lower eigenvalues providing insufficient information to retain. The Broken Stick test employs a line above which the components are considered significant, by applying a random component distribution, above which the eigenvalues of the components should fall to be retained.

Mapping the EWBI to the underlying census enumeration units will be done following the method in Plane & Rogerson (22). Each component of interest to the study will be mapped onto the underlying block groups. This is done by a linear combination of the vector of weighted normalized variables used in each component. Each component produces a "component score coefficient" for each variable. This coefficient is used to weight a vector for each variable. Then the weighted vectors are linearly combined and the resulting component scores can be mapped in GIS to visualize the spatial distribution of each component, classified by component score ranges (less than -1, -1 to 0, 0 to 1, and greater than 1). These maps will reveal spatial concentrations of high or low values of EWB, with multiple compound characteristic regimes, denoting various types or varieties of what may be considered EWB. These types may be delineated by various demographics groups or employment sectors, or other heretofore unconsidered subgroups.

Analysis outputs will include:

- Global Moran's *I* plots
- tables of loadings and explained variance
- scree plots showing variance for each PC
- component significance tests
- thematic maps of component scores

RESULTS & DISCUSSION

The Moran's *I* results show significant variation in spatial autocorrelation intensity and scale across the four metropolitan areas of the study. Figure 3 shows z-scores at 2,000-meter intervals. Atlanta shows a major peak at approximately four kilometers, with a precipitous drop in intensity thereafter. Cleveland's intensity is high at two kilometers, also dropping precipitously thereafter. Eugene shows a markedly different pattern, with a peak at approximately three kilometers, and two thereafter. Its highest z-score is about 2, considerably lower than the other metropolitan areas. Minneapolis demonstrates larger-scale land use concentrations than the other metropolitan areas in the study. It has relatively low intensity at the local scale and begins a sharp climb to its peak at five kilometers, drops until ten kilometers, and then climbs considerably to sixteen kilometers. Its intensity of concentration is high.

The results for the varimax rotation were unsuitable, as most of the loadings were near zero. Therefore, the rotated components were dropped from the study. The original PCA results are utilized for the study. The Broken Stick test indicated that the following PCs were retainable, as they were above the expected component value in the case of a random component distribution. For Atlanta, the first 3; for Cleveland, 3; for Eugene, 4; and, for Minneapolis, 3.

Only significant components with 9% or more variance explained were retained, the first two for Atlanta and Cleveland, and the first 3 for Eugene and Minneapolis. The strength of the loadings are modest for all components, but the expected patterns emerge, with multiple regimes displayed across the different components. All loadings will be evaluted in this context. Each city shows varying intensity of response to transit by mode, and to jobs by sector and income.

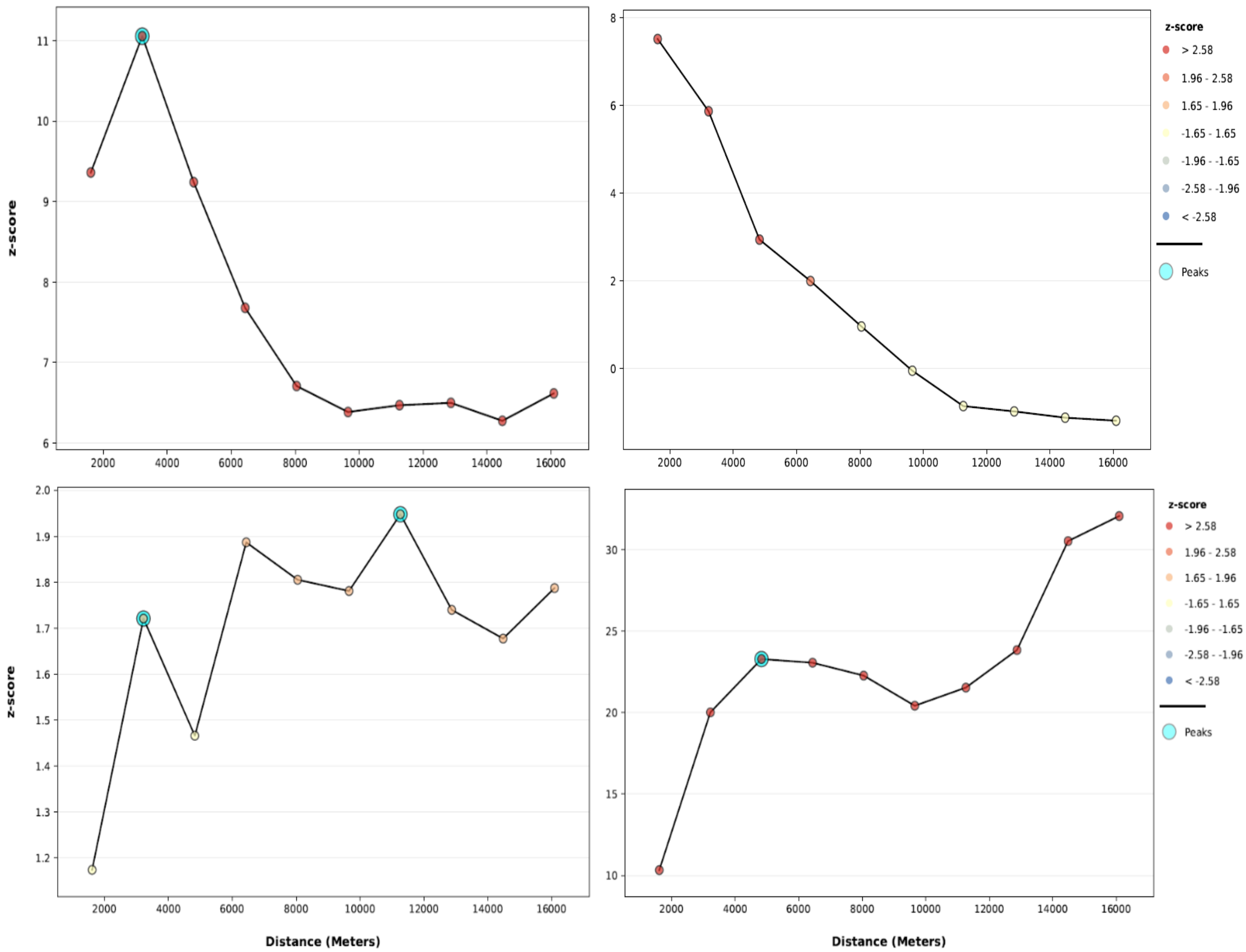

Figure 1. Results of Global Moran's I at various scales.

TABLE 3. Loadings & Explained Variance

	Atlanta		**Cleveland**		**Eugene**			**Minneapolis**		
	PC1	PC2	PC1	PC2	PC1	PC2	PC3	PC1	PC2	PC3
Total Jobs	0.07	**0.31**	0.06	**0.31**	-0.03	**-0.33**	-0.13	0.04	**-0.30**	0.18
Upper Income Jobs	0.07	**0.27**	0.05	**0.28**	-0.02	**-0.26**	**-0.19**	0.04	**-0.26**	0.19
Mid Income Jobs	0.03	**0.21**	0.02	0.18	-0.05	**-0.24**	-0.03	0.02	**-0.22**	0.08
Lower Income Jobs	0.06	**0.26**	0.06	**0.29**	-0.02	**-0.27**	-0.07	0.04	**-0.23**	0.13
Manufacturing Jobs	0.03	0.10	0.02	0.10	0.00	-0.13	**-0.19**	0.04	-0.08	0.11
Light Industry Jobs	0.05	0.14	0.05	0.18	0.00	-0.15	**-0.20**	0.04	-0.15	0.14
Retail-Lodging-Food Jobs	0.08	**0.22**	0.07	**0.22**	-0.01	**-0.26**	-0.09	0.04	**-0.22**	0.13
Knowledge Jobs	0.06	**0.27**	0.05	**0.29**	-0.03	**-0.25**	-0.06	0.03	**-0.25**	0.16
Office Jobs	0.05	**0.26**	0.06	**0.28**	-0.02	**-0.30**	-0.08	0.03	**-0.26**	0.18
Education Jobs	0.01	0.10	0.03	**0.24**	-0.02	-0.07	0.00	0.01	-0.09	0.03
Health Jobs	0.02	0.16	0.00	0.14	-0.05	**-0.19**	-0.01	0.02	-0.14	-0.01
Arts-Ent-Rec Jobs	0.03	0.17	0.05	**0.26**	-0.02	**-0.20**	-0.02	0.03	**-0.20**	0.09
Total Workers	**0.23**	-0.04	**0.22**	-0.03	**0.25**	-0.03	0.00	**0.23**	0.02	-0.03
Upper Income Workers	0.20	0.03	**0.21**	-0.02	**0.25**	-0.02	-0.02	**0.23**	0.03	0.02
Mid Income Workers	**0.20**	-0.08	**0.20**	-0.03	**0.25**	-0.01	-0.02	**0.22**	0.02	-0.08
Lower Income Workers	**0.22**	-0.06	**0.21**	-0.03	**0.24**	-0.05	0.04	**0.22**	0.00	-0.09
Manufacturing Workers	**0.20**	-0.06	**0.19**	-0.04	**0.21**	-0.01	-0.05	**0.20**	0.08	0.06
Light Indus Workers	**0.21**	-0.07	**0.20**	-0.05	**0.23**	-0.01	-0.06	**0.22**	0.06	0.05
Retail-Lodging-Food Wrkrs	**0.21**	-0.07	**0.20**	-0.03	**0.23**	-0.06	-0.01	**0.22**	0.01	-0.07
Knowledge Workers	0.16	0.07	**0.20**	0.02	**0.22**	-0.03	0.05	**0.20**	-0.02	-0.03
Office Workers	**0.23**	-0.03	**0.21**	-0.01	**0.24**	-0.05	0.04	**0.22**	-0.01	-0.06
Education Workers	0.19	-0.05	**0.19**	-0.05	0.16	0.00	0.10	**0.19**	0.01	-0.07
Health Workers	**0.21**	-0.07	**0.19**	-0.02	**0.25**	-0.02	0.00	**0.20**	0.00	-0.11
Arts-Ent-Rec Workers	**0.19**	-0.01	0.19	0.02	**0.19**	-0.03	0.04	**0.21**	-0.03	-0.10
Standard deviation	4.21	2.90	4.52	2.97	3.88	2.78	2.23	4.19	2.95	2.29
Proportion of Variance	33%	16%	38%	16%	28%	15%	9%	32%	16%	10%
Cumulative Proportion	33%	48%	38%	54%	28%	43%	52%	32%	48%	57%
Broken Stick Sig. Test	ATL: 3 PCs		CLV: 3 PCs		EUG: 4 PCs			MINN: 3 PCs		

TABLE 3. (Continued) Loadings & Explained Variance

	Atlanta		Cleveland		Eugene			Minneapolis		
	PC1	PC2	PC1	PC2	PC1	PC2	PC3	PC1	PC2	PC3
Drove Alone to work	**0.22**	-0.02	**0.21**	-0.05	**0.24**	-0.01	0.00	**0.22**	0.03	-0.02
Carpooled to work	0.11	-0.03	0.10	-0.01	0.13	-0.01	-0.03	0.11	-0.01	-0.10
Transit to work	0.03	0.04	0.01	0.07	0.05	-0.07	0.18	0.03	-0.12	**-0.25**
Walked to work	0.03	0.15	0.04	0.15	-0.05	-0.11	**0.22**	0.02	**-0.20**	-0.11
Work at Home	0.13	0.03	0.11	-0.02	0.07	0.02	0.03	0.11	0.00	-0.01
Distance to LRT			0.12	**-0.09**				0.11	0.13	**0.25**
Distance to HRT	0.11	**-0.13**								
Distance to BRT			0.14	**-0.10**	-0.02	0.08	**-0.21**	0.10	0.11	**0.23**
Distance to SCT	0.11	**-0.12**								
Distance to CRT								0.02	0.07	0.14
Households - No Vehicle	0.03	0.07	0.01	0.09	0.00	-0.17	0.18	0.02	-0.17	**-0.21**
Households - 1 Vehicle	0.17	0.08	0.13	0.05	0.13	-0.11	0.13	0.11	-0.14	-0.18
Households - 2+ Vehicles	**0.20**	-0.08	**0.20**	-0.08	**0.22**	0.04	-0.07	**0.21**	0.08	0.04
Commute Time Fwr 5 Min	0.07	0.06	0.08	0.04	0.02	-0.06	-0.03	0.07	-0.07	-0.02
Commute Time 5 to 14	0.15	0.11	0.16	0.03	0.13	-0.11	0.14	0.16	-0.10	-0.10
Commute Time 15 to 29	0.18	0.04	0.18	-0.02	**0.21**	0.00	0.09	0.18	-0.04	-0.14
Commute Time 30 to 44	0.17	-0.06	0.19	-0.05	0.13	-0.01	-0.01	**0.19**	0.03	-0.03
Commute Time Grtr 45	0.17	-0.10	0.15	-0.04	0.11	0.02	-0.01	0.17	0.05	0.00
Distance to CBD	0.11	-0.12	0.12	-0.10	-0.03	0.08	**-0.23**	0.10	0.14	**0.25**
Population Density / mile	0.00	0.08	-0.06	-0.01	-0.03	-0.04	**0.33**	-0.04	-0.09	**-0.28**
Worker Density / mile	0.03	**0.27**	0.01	**0.25**	-0.07	**-0.22**	0.02	0.01	-0.29	0.09
Total Occupied Housing	**0.22**	0.01	**0.19**	0.01	**0.21**	-0.08	0.07	**0.21**	-0.06	-0.12
Owner Occupied	0.17	-0.08	0.18	-0.09	0.19	0.08	-0.17	**0.19**	0.09	0.07
Renter Occupied	0.13	0.09	0.07	0.10	0.07	-0.17	**0.24**	0.06	-0.17	-0.23
Vacancy Rate	-0.07	0.05	-0.11	0.05	-0.08	-0.10	-0.05	-0.03	-0.06	-0.03
Median Year Built	0.14	-0.01	0.14	-0.04	0.10	-0.04	-0.16	0.13	0.02	0.16
Dissimilarity Index - Income Match	-0.04	-0.02	-0.05	-0.02	-0.01	0.03	-0.03	-0.04	0.03	-0.02
Internal Capture	0.16	-0.05	0.18	-0.05	-0.01	0.06	-0.19	0.15	0.03	0.03
GWR subcenter Std. Dev.	0.07	**0.29**	0.06	**0.31**	-0.03	**-0.32**	-0.18	0.05	**-0.29**	0.19
Gi* Centering Z Score	0.03	**0.20**	0.06	0.17	-0.07	-0.16	**0.22**	0.02	-0.12	-0.08
Road Network Density	-0.05	0.16	-0.12	0.07	-0.07	-0.05	**0.31**	-0.09	-0.11	**-0.24**
Intersection Density	-0.06	0.16	-0.13	0.07	-0.08	-0.07	**0.33**	-0.10	-0.14	**-0.25**
Standard deviation	4.21	2.90	4.52	2.97	3.88	2.78	2.23	4.19	2.95	2.29
Proportion of Variance	33%	16%	38%	16%	28%	15%	9%	32%	16%	10%
Cumulative Proportion	33%	48%	38%	54%	28%	43%	52%	32%	48%	57%
Broken Stick Sig. Test	ATL: 3 PCs		CLV: 3 PCs		EUG: 4 PCSs			MINN: 3 PCs		

Note: Negative values for loadings on transit modes denote a proximity-based positive influence.

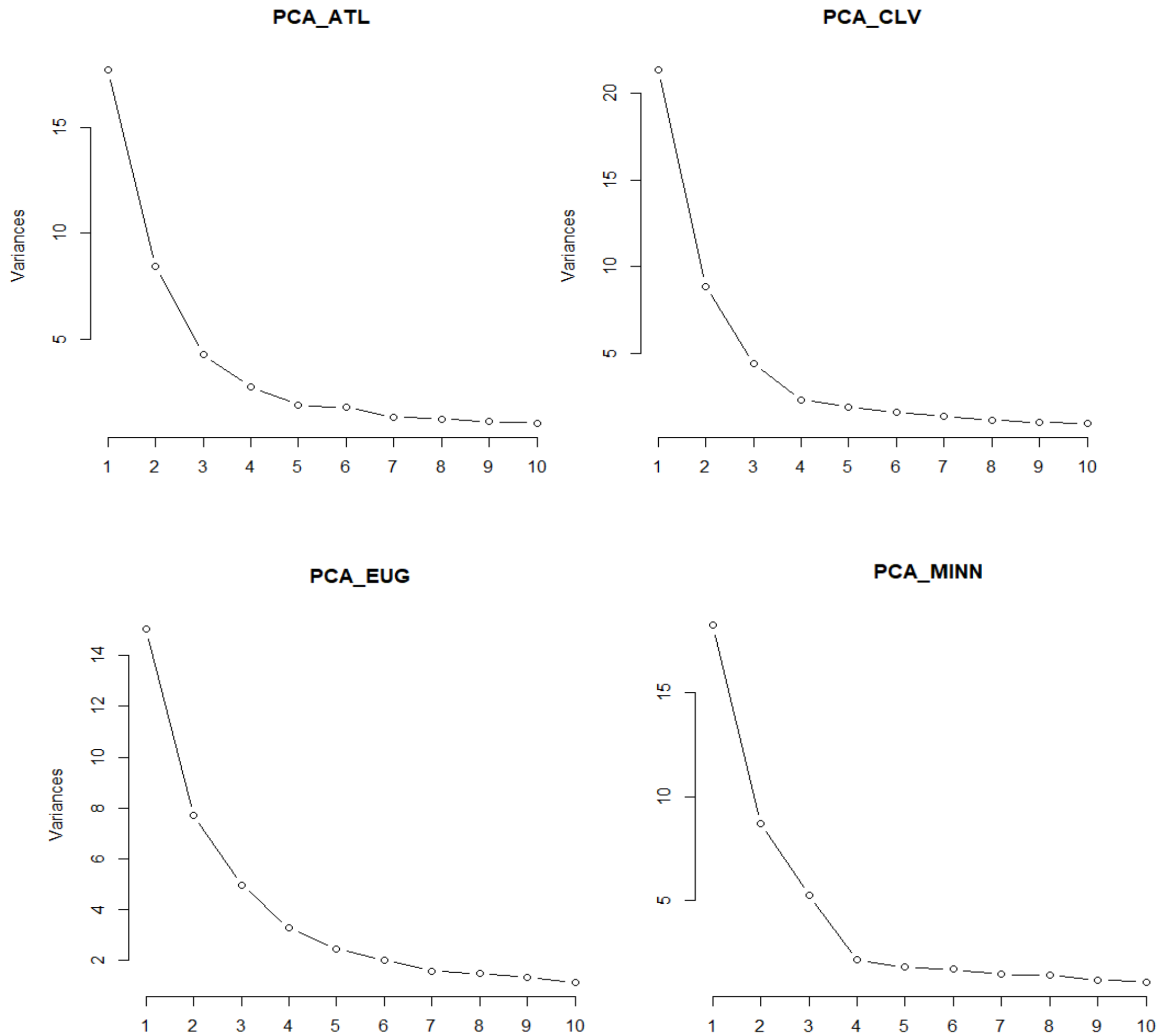

Figure 2. PCA Scree plots, showing PCs by variance

The loadings (**table 3**) show, most saliently, that in the study cities the location of jobs and residences are not loading on the same components. In Atlanta, the places with jobs also have the strongest access to transit, but the spatial relationship between workers' residences and transit stations is weak. Atlanta's jobs load on the same PC across various income levels and sector groups. Retail, Lodging and Food jobs load on the same PC as Office and Knowledge jobs. The same is the case in Cleveland. For all of the study cities, the jobs and worker residences load on different PCs, indicating that their locations are not highly correlated. In Minneapolis, jobs and residences do not load highly on any PCs, which may indicate a lack of strong spatial concentration of these variables. Eugene's loadings suggest that it is mainly a suburb for workers in Portland, as residences load highly but jobs load negatively. Atlanta and Cleveland show residents loading highly on PC1 and jobs loading highly on PC2. Minneapolis residents load highly on PC1 and jobs load negatively on PC2 and somewhat highly on PC3.

This suggests that Minneapolis consists of more complex spatial regimes than the other cities. As Atlanta is also a large and complex city, the results also demonstrate the explanatory limitations of the study methods, and the need to further the scale effects of these cities' spatial regimes.

The commuting modes show the ongoing dominance of the automobile. All cities show a dominant pattern of commuters driving alone to work, followed by carpooling. Work from home loaded more highly for all cities than did transit use. For Eugene, walking to work loaded highly on the same PC as proximity to BRT. Transit proximity loaded in varying ways across the cities, and all but Minneapolis responded positively to proximity to transit stations, with modestly negative loadings on the distance to the transit station. As stated above, a negative loading on distance indicates a positive loading on transit station proximity, which may be interpreted as a positive response to station proximity by the surrounding land uses. In Atlanta and Cleveland, job locations and transit station proximity loaded highly on the same PC, suggesting that job locations are more highly served than residential neighborhoods. Eugene responded well to proximity to BRT stations, but its job variables loaded negatively on all components, which further supports the interpretation of Eugene as a suburb of Portland.

Households with no vehicle loaded weakly on the components all cities, and households with 2 or more loaded most highly. Longer commute times loaded most highly across all the cities. Density and distance to CBD varied in relevance across the cities. Job density loaded positively in Atlanta and Cleveland, but negatively in Eugene and Minneapolis. Population density loaded positively in Eugene, but elsewhere was weak or negative. Total occupied housing was most relevant in all cities, with owners being more relevant than renters in most contexts, but renters loaded more highly than owners in Eugene. Centering and subcentering (polycentric development) loaded positively in Atlanta, Cleveland, and Eugene, but loaded negatively in Minneapolis. This suggests a more polycentric development in the first 3 cities, but a more dispersed development in Minneapolis. The built environment appears to be less walkable in all of these cities but Eugene, as the road network and intersection densities loaded weakly or negatively in all but Eugene, which had positive loadings on PC3, which also has high loadings from proximity to BRT.

All study cities demonstrated weak loadings from income match and internal capture measures. This is an expected outcome, with a manifest need in most US cities for a greater emphasis on employment-worker balance.

Maps of PC1 scores (**figure 5**) reveal a wide variation in the index across the study cities. Each city shows some degree of high intensity along road networks. This is quite pronounced in Cleveland and seen in a radial pattern of low intensity in Atlanta, and in the north region of Minneapolis.

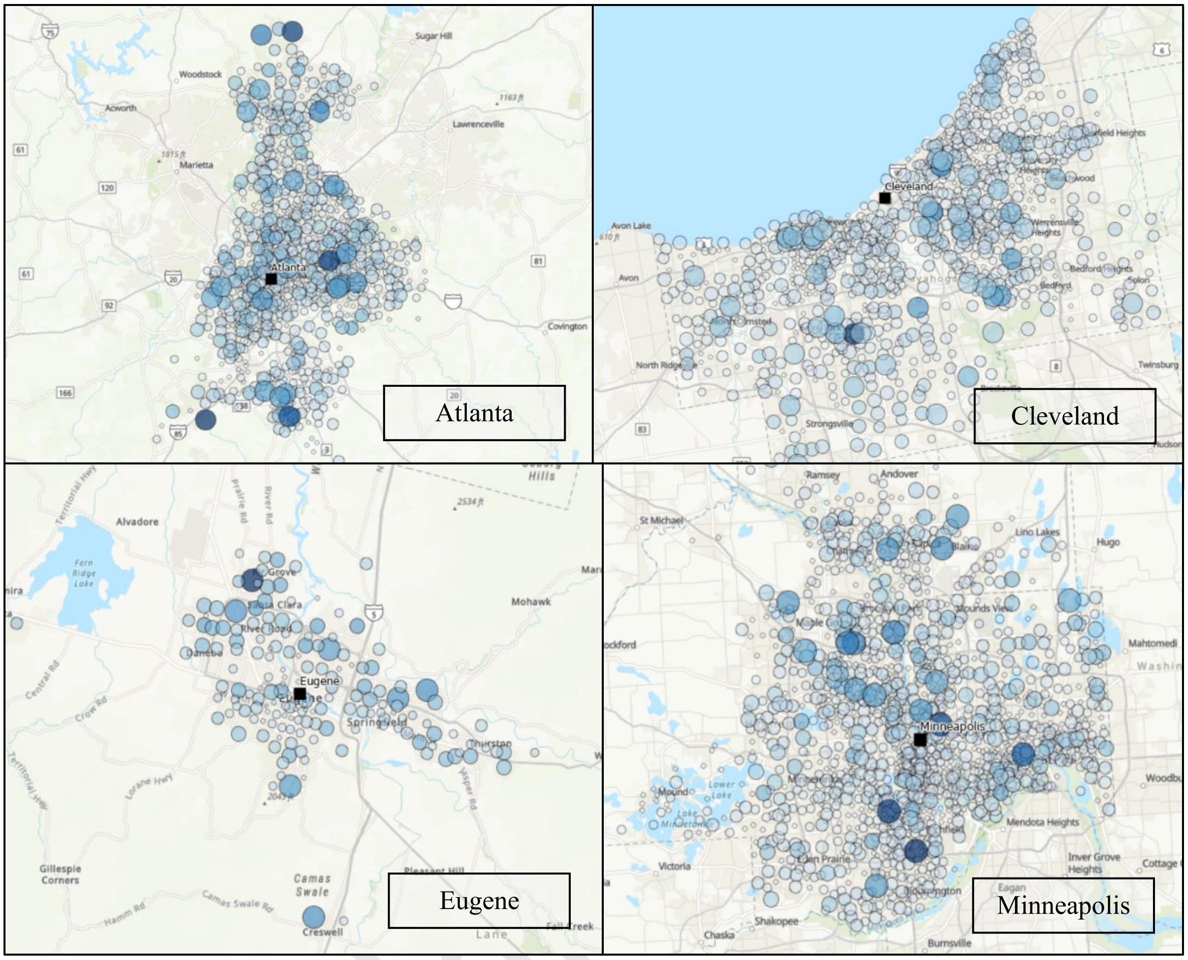

Figure 3. PC1 Scores for Atlanta, Cleveland, Eugene, And Minneapolis. Graduated symbols range from small to large circles in light to dark blue, and high to low transparency to denote score magnitude. Atlanta's scores range from -2.3 to 27. Cleveland's scores range from -3.2 to 35.6. Eugene's scores range from -3.3 to 19.8. Scores range from -1.7 to 31.4 for Minneapolis. *Base Map Sources:* Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors and the GIS User Community

CONCLUSION

Theoretical implications of the employment-worker balance phenomenon are drawn from the spatial and attribute clusters revealed by the EWBI. A more accessible workplace translates to a more productive and resilient workforce through potential improvements in work-life balance and overall cost of living, which in turn benefits the firm through higher output. Additionally, existing discrepancies in EWB near transit stations reveal low-hanging fruit for planners who wish to increase economic and housing resiliency. The employment-worker spatial regimes identified in this study through PCA may require targeted solutions to increase EWB. This may reveal some significant patterns of outcomes to transit development. One main implication is that there is a great deal of potential to develop spatial balance between employment and worker residence. The built environment in Eugene far better supports walkability than in the other larger cities of the study. The built environment also plays a role in a positive response to transit proximity. Road and intersection densities seem to correlate well with a positive response to transit.

Workers remains separated from their workplaces. This may be seen by a portion of the population as a significant benefit, but many are paying excessive transportation costs, spending excessive time to commute, and high municipal taxes to support this separation of land uses. These results have significant workforce as well as workplace implications, as accessibility outcomes provide agglomeration economies. The regions in which workers have greater TOD-driven access to firms also provide a more business-friendly environment with increased *situs* via a more accessible, active workforce. When appropriate housing is provided for workers of all sectors of the economy, greater economic diversification is possible.

The results indicate a modestly positive response to TOD. The political implications of increasing employment-worker balance depend upon the local typology of imbalance needing correction. In neighborhoods that are job-rich and housing poor for a lower- to moderate-income worker, challenges may include potential for local opposition from businesses that benefit from larger numbers of workers than residents, businesses seeking to protect their market share from newcomer firms, or from residents who fear negative externalities of lower or moderate-income housing development in their neighborhoods. Neighborhoods with upper-income jobs that seek to improve EWB may face gentrification pressures. Bedroom communities for blue-collar workers needing more jobs may face challenges from industrial externalities (10).

ACKNOWLEDGMENTS

This manuscript was prepared as part of the project sponsored by the National Institute of Transportation and Communities (under Project Number 1253 and a Dissertation Fellowship), with match provided by the University of Arizona.

DISCLOSURES

Research reported herein was supported by the National Institute of Transportation and Communities through funding from the U.S. Department of Transportation's University

Transportation Center Program as well as the University of Arizona. Views expressed herein do not necessarily represent those of the sponsors.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: Hibberd, Nelson; data collection: Hibberd; analysis and interpretation of results: Hibberd, Nelson; draft manuscript preparation: Hibberd, Nelson. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Weber J, Kwan M-P. Evaluating the Effects of Geographic Contexts on Individual Accessibility: A Multilevel Approach [1]. Urban Geogr [Internet]. 2003;24(8):647–71. Available from: http://www.tandfonline.com/doi/abs/10.2747/0272-3638.24.8.647
2. Schwanen T, Dijst M. Travel-time ratios for visits to the workplace: The relationship between commuting time and work duration. Transp Res Part A Policy Pract. 2002;36(7):573–92.
3. Nelson AC, Marshall JT, Juergensmeyer JC, Nicholas JC. Market Demand-based Planning and Permitting. Chicago: American Bar Association Press; 2017.
4. Sanchez TW. The Connection Between Public Transit and Employment. 2007;4363.
5. Paez, Antonio., Scott, Darren., Morency C. Measuring Accessibility: Positive and Normative Implementation of Various Accessibility Indicators. J Transp Geogr [Internet]. 2012;25:141–53. Available from: http://journals1.scholarsportal.info/details.xqy?uri=/09666923/v25inone_c/141_mapaniovai.xml
6. Reggiani A, Bucci P, Russo G, Haas A, Nijkamp P. Regional labour markets and job accessibility in City Network systems in Germany. J Transp Geogr [Internet]. 2011;19(4):528–36. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2010.05.008
7. Ewing, Reid, and Hamidi S. Measuring Urban Sprawl and Validating Sprawl Measures [Internet]. Washington, D.C.; 2014. Available from: https://gis.cancer.gov/tools/urban-sprawl/
8. Levine J. Rethinking Accessibility and Jobs-Housing Balance. J Am Plan Assoc. 1998;64(2):133–49.
9. Graaskamp JA. Fundamentals of Real Estate Development. ULI-The Urban Land Institute. Washington, DC; 1981.
10. Weitz J. Jobs-Housing Balance. Chicago, Ill.; 2003.
11. Sommervoll Å, Sommervoll DE. Learning from man or machine: Spatial fixed effects in urban econometrics. Reg Sci Urban Econ. 2019;

12. Benner C, Karner A. Low-wage jobs-housing fit : identifying locations of affordable housing shortages. Urban Geogr [Internet]. 2016;37(6):883–903. Available from: http://dx.doi.org/10.1080/02723638.2015.1112565

13. Nelson AC, Ganning J. National Study of BRT Development Outcomes [Internet]. Portland, OR; 2015. Available from: https://nitc.trec.pdx.edu/research/project/650

14. Kain JF. Housing Segregation, Negro Employment, and Metropolitan Decentralization. Q J Econ [Internet]. 1968;82(2):175–97. Available from: http://www.jstor.org/stable/1885893

15. Stoker P, Ewing R. Job – Worker Balance and Income Match in the United States. Hous Policy Debate [Internet]. 2014;24(2). Available from: http://dx.doi.org/10.1080/10511482.2013.852604

16. Jackson MC, Huang L, Xie Q, Tiwari RC. A modified version of Moran's I. Int J Health Geogr [Internet]. 2010;9:33. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2903534&tool=pmcentrez&rendertype=abstract

17. Hajrasouliha AH, Hamidi S. The typology of the American metropolis: monocentricity, polycentricity, or generalized dispersion? Urban Geogr [Internet]. 2016;3638(April):1–25. Available from: http://www.tandfonline.com/doi/full/10.1080/02723638.2016.1165386

18. Getis A, Ord JK. The Analysis of Spatial Association. Geogr Anal. 1992;24(3):189–206.

19. Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. Int J Geogr Inf Sci. 2010;24(3):383–401.

20. Park K, Ewing R, Sabouri S, Choi D, Hamidi S, Tian G. Guidelines for a Polycentric Region to Reduce Vehicle Use and Increase Walking and Transit Use. J Am Plan Assoc [Internet]. 2020;0(0):1–14. Available from: https://doi.org/10.1080/01944363.2019.1692690

21. Behera SK, Rao SA, Saji HN, Yamagata T. Comments on "A cautionary note on the interpretation of EOFs." J Clim. 2003;16(2):1087–93.

22. Plane DA, Rogerson PA. The Geographical Analysis of Population with Applications to Planning and Business. New York: John Wiley & Sons; 1994.

23. Richman MB. Rotation of Principle Components. J Climatol. 1986;6:292–335.

24. Cadwallader MT. Migration and residential mobility: Macro and micro approaches. Madison, Wis.: Univ of Wisconsin Press; 1992.

25. Nelson AC, Hibberd R. The Link between Transit Station Proximity and Real Estate Rents , Jobs , People and Housing with Transit and Land Use Planning Implications The Link between Transit Station Proximity and Real Estate Rents , Jobs , People and Housing with Transit and Land. Portland, OR; 2019.

26. Bertolini L. Nodes and places: Complexities of railway station redevelopment. Eur Plan

Stud. 1996;4(3):331–45.

27. Ewing R, Cervero R. Travel and the built environment: a synthesis. Transp Res Rec [Internet]. 2001;1780(Paper No. 01-3515):87–114. Available from: http://mrc.cap.utah.edu/wp-content/uploads/sites/8/2015/04/fulltext.pdf

28. Cervero R, Duncan M. Jobs-Housing Balauce or Retail-Housing Mixing ? J Am Plan Assoc. 2008;74(2):475–90.