Modeling the Economic Impact of Tennessee's FastTrack Program through Spatial Regression Analysis

Benjamin Browning
GEOG870: Capstone in GIS
Dr. Fritz Kessler
23 April, 2024

Introduction

Corporate incentive programs are one of the predominant tools economic development organizations (EDOs) use in the United States to stimulate local economies. Through incentive programs, EDOs offer financial assistance to encourage firms to relocate or expand within a jurisdiction. The financial assistance offered could consist of grants, subsidies, tax cuts, and occasionally favorable loans, and the amount of assistance provided is often proportional to the amount invested and new jobs promised by the company. In Tennessee, the Department of Economic and Community Development's (TNECD) FastTrack grant program is the main incentive source for firms relocating or expanding within the state. Companies that receive FastTrack grants can use the money for job training assistance, infrastructure development, or general expenditures (TNECD, n.d.).

Moreover, EDOs provide financial assistance assuming that the investments made by the company will boost the local economy, especially employment and income levels. However, there is no consensus on whether this assumption is true or universally applicable. Therefore, this capstone paper will address the assumption of economic benefits by modeling the relationship between FastTrack grants and changes in the civilian labor force size, employment, household income, and home values at the census tract level in Tennessee. Specifically, four spatial regression models, i.e., one for the labor force size, employment, income, and home values, that conform to all necessary assumptions will be developed through this research.

Literature Review

Incentive Spending

Corporate incentive programs designed to boost local economies comprise a substantial portion of state and local budgets. In 2014, on average, states spent 38 percent of their corporate tax revenue on corporate tax incentives (Slattery & Zidar, 2020). Annually, EDOs nationwide spend an estimated $45 to $90 billion total on incentives (Parilla & Liu, 2018). Furthermore, incentive programs are implemented expecting that incentivized corporate activity will produce a measurable economic benefit for the jurisdiction where it occurs. EDOs frequently measure the success of their programs by tying them to economic metrics, such as employment, population, income, or local GDP (Gonzales et al., 2019), and this research follows suit. However, in most cases, job creation is the primary motivation behind incentive programs. Therefore, program efficiency is frequently measured as costs per job. The national average cost per job created between 2002 and 2017 by incentive programs was $45,785, and in Tennessee, where this study takes place, it was $11,805 (Slattery & Zidar, 2020). Inevitably, the high price tag of economic incentive programs has led to much debate.

Spillover and Agglomeration Effects

Policymakers justify the expense of incentive programs by citing the economic phenomenon known as "spillover." Spillover refers to the effects that occur as an indirect result of other economic activity. In the context of economic development, existing research into manufacturing agglomeration as a determinant of industrial location and Foreign Direct Investment (FDI) provides a potential but not irrefutable mechanism for the occurrence of spillover. To elaborate, it is asserted that new companies opening in an area attract other companies in their supply chain to the same area, thus creating additional benefits. This reasoning is supported by Esiyok and Ugur (2017) who found that increases in FDI within a province of Vietnam resulted in increases in nearby provinces' FDI. Furthermore, Kline and Moretti's (2014) research indicated that substantial federal investment in the Tennessee Valley Authority (TVA) in 1933 accelerated the regional economy's employment, manufacturing, and agricultural production, largely due to manufacturing agglomeration in the following decades.

While support for the industrial agglomeration argument is quite prominent, other research conflicts with claims of industrial agglomeration and subsequent economic spillover. Takano et al. (2018) found that agglomeration effects are only statistically significant for heavy but not light industries. This finding suggests that spillover effects may also only apply to heavy industry investments though incentives are typically provided to all industry types. Also, Edmiston (2004, p. 316) indicated that "dispersive forces outweigh agglomeration forces with new employment" in manufacturing. In other words, while new plants create new jobs, they may cost some existing jobs at nearby plants that compete for resources. In those instances, the employment growth is positive, but the spillover is negative, meaning the initial costs per job created were underestimated, as net job growth did not meet its target (Edmiston, 2004). Moreover, linking industrial agglomeration to growth following investments in the TVA could be a misinterpretation. The TVA investment was ultimately an investment in publicly owned energy infrastructure, and as Greenstone et al. (2010) and Villaverde and Maza (2012) suggested, access to infrastructure — which multiple companies likely need to share — is a potentially greater determinant of plant location and FDI than intentional proximity to other plants. While industry agglomeration is not the primary concern of this paper, understanding agglomeration is important to understanding the hypothetical mechanism behind economic development spillover. Therefore, as no consensus exists on agglomeration effects nor is there cross-sector uniformity, the methodology implemented must ensure that spillover effects on a location from all nearby firms are captured, either explicitly or as a net value.

Because the evidence on agglomeration effects is unclear, the evidence on spillover effects is also unclear. Many researchers have analyzed spillover from incentive programs but achieved mixed results depending on the variables and methods employed. Slattery and Zidar (2020) found no evidence of employment spillover from subsidies. Moreover, Bundrick and Snyder (2018) and Bundrick and Yuan (2019) assessed Arkansas's Quick Action Closing Fund (QACF) program using ordinary least squares (OLS) and synthetic control models, respectively. Bundrick and Snyder (2018) found no statistically significant relationship between QACF subsidies and growth in employment and other private establishments within the county where an investment is made, but they did find a slightly negative effect on the growth of establishments in neighboring counties through a fixed effect approach. This is similar to Edmiston's (2004) findings that new manufacturing plants create new jobs at the expense of some existing jobs. Furthermore, Bundrick and Yuan (2019) reported improvements in income and poverty levels at the county level within the first year following a subsidized investment, but no long-term increases were evident. In contrast, Greenstone et al. (2010) found that new plants substantially boost the productivity of nearby existing plants from the same industry and increase land and labor costs. Within this paper's context, land and labor cost increases can be understood as beneficial income and home value increases. Also, while Edmiston (2004) identified zero or negative spillover on employment from new manufacturing plants, they also found that expansions of existing plants have substantial positive spillover.

Spatial Approaches

This capstone paper will refine the existing body of evidence on subsidized investments' impact by focusing on a few highly relevant, frequently used econometrics, specifically the labor force size, employment, income, and home values. However, it deviates from existing research by implementing spatial regression models. Spatial regression can be used to account for spatial heterogeneity, externalities from neighbors, omitted variable bias, and model uncertainty (LeSage & Pace, 2009), which are all a concern in the research at hand. Also, Chi and Zhu (2020) recommend implementing spatial regression models whenever spatial dependencies are identified in variables and model residuals. When dealing with explanatory variables, spatial dependency could even be conceptual (Chi & Zhu, 2020), as is the case with firm relocation externalities. Therefore, this capstone paper assumes and models spatial dependencies within the data using spatially lagged variables, which are weighted neighborhood averages.

Furthermore, this paper expands on existing research by analyzing the topic at a more granular scale. Analyzing at the census tract level instead of the county level allows impact distribution to be identified within counties and across county borders. This is important as the location of a firm within a county is likely to affect how its externalities are dispersed. For example, a firm established near the county's edge could benefit the neighboring county just as much as the county it is in due to its proximity. However, neighboring counties are not as likely to experience those benefits when a firm is in the center of the county.

Furthermore, while spatial regression has not been applied to subsidized investment research, existing research on FDI determinants provides several potential templates for a spatially aware approach to economics. For the most common forms of spatial regression, the spatial weighting matrix — abbreviated as the W matrix — is critical, though how it is implemented varies (Corrado & Fingleton, 2012). In this capstone paper, the W matrix determines how far and strongly a firm's investment is represented by the spatially lagged variable. Esiyok and Ugur

(2017) used two nearest neighbor-based and distance-based W matrices when lagging their dependent variable and found that FDI agglomeration was less impactful at greater distances. Villaverde and Maza (2012) also lagged their dependent variables while testing for interregional competition for FDI in Spain using Generalized Least Squares (GLS), but they are unique in that they implemented lagged independent variables as instrumental variables. As Corrado and Fingleton (2012) note, testing models with spatially lagged independent variables before adding lagged dependent variables can help mitigate model bias and misspecification. In their research, Takano et al. (2018) followed this advice when constructing a negative binomial regression model with Eigenvector Spatial Filtering (ESF). Notably, they did not report any major differences in their study between the lagged independent variables constructed with a five nearest-neighbors matrix and an inverse distance squared matrix with a 45-kilometer threshold. Also, the models with ESF produced the most statistically significant variables of the models tested (Takano et al., 2018). This is consistent with Griffith and Chun's (2016) findings that ESF greatly reduces omitted variable bias as measured by Ramsey's regression equation specification error test (RESET).

Methodology

Data Selection

Data for this research will come from two sources: the TNECD and the U.S. Census Bureau. The TNECD's *Projects with Contracted FastTrack Grants since 2011* dataset provides a list of investments made within Tennessee from January 2011 to November 2023 (Tennessee Department of Economic and Community Development, 2023). Relevant variables from this dataset include grant totals and address fields. The grant totals serve as the common independent variable for all regression models, and the address field can be geocoded into point data and aggregated.

The economic data comes from the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates. The TIGER/Line 2011 Census Tract boundary dataset with 2007-2011 ACS estimates contains the three dependent variables (U.S. Census Bureau, 2023c). The 2011 dataset provides the starting year for measuring change in the variables over the study period. Currently, the equivalent 2022 dataset is not available. However, a comparable dataset can be created by joining the TIGER/Line 2022 tract boundary dataset for Tennessee (FIPS Code = 47) with relevant tables using tract GEOIDs (U.S. Census Bureau, 2022). The DP03: Selected Economic Characteristics table provides employment counts and median household income per tract (U.S. Census Bureau, 2023a), and the DP04: Selected Housing Characteristics table provides median home values (U.S. Census Bureau, 2023b). Once these two datasets are joined to the associated tract boundaries, they will serve as the endpoint for measures of change. Ultimately, the percent change from 2011 to 2022 will be calculated using these datasets.

Data Preparation

Before further analysis, data from separate files must be joined into a single tract boundary file. First, the DP03 and DP04 tables must be joined to the 2022 Census Tract boundaries in ArcGIS Pro using the GEOID field. Next, the 2022 data is reaggregated to match the 2011 tract borders. Data is aggregated to the 2011 data for two reasons. First, because tract boundaries changed during the 2020 Census, most years included in the study period share the same boundaries as the 2011 tract data. Second, converting the 1,701 tracts from the 2022 set into the 1,497 tracts from the 2011 set should minimize aggregation errors. The data can be aggregated using the Summarize Within tool found in the GeoAnalytics Desktop toolbox. The new employment and labor force fields are calculated using a sum with the "count" quantity type. Median home values and median household income — because they represent measures of central tendency — are better calculated as weighted averages, using the "rate" quantity type. Once these three fields have been aggregated and merged with the 2011 dataset, the percent change from 2011 to 2022 can be calculated. To prepare the FastTrack dataset, the project addresses need to be geocoded, aggregated to the 2011 tract boundaries, and merged with the economic data. Project aggregation also uses the Summarize Within tool with grant totals for each tract calculated as a sum with the "count" quantity type.

Exploratory Analysis

Once the data is consolidated, exploratory tests will be implemented on each variable. To understand non-spatial distributions for each variable, measures of central tendency, i.e., mean and median, and measures of distribution, i.e., quartile range, minimum, maximum, standard deviations, skewness, and kurtosis, will be calculated. Understanding the non-spatial distribution will help identify extreme outliers that may have resulted from aggregation.

Exploratory spatial tests will help identify spatial autocorrelation and indicate the variables' spatial distribution. Most importantly, spatial tests will help identify the appropriate parameters for weighting matrices implemented during the regression analysis. Weighting matrices define the neighborhood and weighting scheme applied when calculating lagged variables. Using functions from the *spdep* package in R Studio, Moran's *I* will be used to test each variable for spatial autocorrelation with varying weighting matrix parameters. Following the advice of Chi and Zhu (2020), the weighting matrices that show the greatest spatial autocorrelation while being highly significant will be chosen to represent spatially dependent variables in the regression models. Before modeling, local Moran's *I* tests using the chosen matrices will be implemented to identify erroneous spatial outliers, and lagged grant total variables will be calculated for each matrix.

Regression Analysis

After completing exploratory analysis, three series of regression models with various combinations of spatial regression terms will be tested using functions from the *spatialreg* package within R Studio. The first series will consist of spatially lagged X (SLX) models — also known as spatial cross-regressive models — which expand standard OLS models by including a lagged independent variable. The lagged independent variable is necessary for this research because most census tracts do not contain FastTrack-incentivized projects, strongly skewing the data negatively. Additionally, research suggests that project effects are spatially distributed over an unknown distance, and when the conceptual basis for a relationship between the lagged independent variable and the dependent variable exists, the SLX model is useful (Chi & Zhu, 2020). The general form of an SLX model is

$$y = X\beta + WX\theta + \varepsilon \quad (1)$$

where WX is the spatially lagged form of X created using matrix W, and θ is the coefficient (Chi & Zhu, 2020). In R Studio, the SLX model will be implemented via the *lmSLX* function.

Following the example of Takano et al. (2018), the second series will introduce eigenvector spatial filtering (ESF) to the SLX model. ESF controls for omitted spatial variables by extracting eigenvectors from residuals using the spatial weight matrix C to produce "synthetic proxy variables" for the right-hand side of the model (Griffith & Chun, 2016). Because labor force size, employment, income, and home values are simple variables that could be impacted by an indeterminable number of factors, control variables are necessary to isolate the impact of grant totals on the dependent variables, minimize omitted variable bias, and remove spatial autocorrelation. ESF accomplishes this using selected eigenvectors without substantially increasing model complexity. The standard ESF equation is

$$y = X\beta + E_k\beta_k + \varepsilon \quad (2)$$

where E_k is the selected set of eigenvectors used to create the spatial filter (Griffith & Chun, 2016). In R studio, eigenvectors can be extracted from residuals and selected with step-wise regression using the *mem.select* function from the *adespatial* package. ESF has also been known to improve a model's normality and homoscedasticity (Thayn & Simanis, 2013).

The third series of models will implement a Spatial Durbin Model (SDM), which expands the SLX model by adding a lagged dependent variable as an explanatory variable (LeSage & Pace, 2009). While the SLX model with ESF certainly controls for spatial autocorrelation among the dependent variable, it does not explicitly indicate its presence. Therefore, including the lagged dependent variable will allow the model to identify endogenous spatial dependence within the dependent variables. The equation of an SDM model is

$$y = \rho Wy + X\beta + WX\theta + \varepsilon \quad (3)$$

where ρ is the coefficient of the spatially lagged dependent variable, Wy (Lesage & Pace, 2009). The inclusion of the Wy term has a similar effect in reducing residual spatial autocorrelation as the inclusion of an eigenvector spatial filter. The SDM model can be implemented in R Studio using the *lagsarlm* function.

Model fit will be assessed using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which penalizes too many variables (Chi & Zhu, 2020). Furthermore, because OLS models must meet the assumptions of homoscedasticity, normality, and independence within model residuals, additional statistical tests will be implemented to test these assumptions. Following the recommendations of Chi and Zhu (2020), autocorrelation will be tested using Moran's I, and heteroscedasticity will be tested with the Breusch-Pagan test. Normality can be tested using the Kolmogorov-Smirnov test. Should the normality test fail, Box-Cox transformations will be implemented as specified by Griffith (2013). The better model should fail to reject the null on all these tests and have the lowest AIC and BIC scores.

Results

Exploratory Analysis

Once data preparation was completed in ArcGIS Pro, the data was loaded into R Studio. **Figure 1** shows a map of the grant totals for each district. Then, null values were removed, and summary statistics were calculated for each variable. Upon observing high levels of skewness and kurtosis in all the variables, outliers were removed from the dependent variables using Tukey's Method with two interquartile ranges. Summary statistics were then calculated again. The results can be found in **Table 1**. Even after outlier removal, positive skewness and kurtosis were still found in the variables. Maps created for each dependent variable after outlier removal can be found in **Figure 2**. Following this step, the grant totals were also divided by one million to make coefficient reporting and eigenvector extraction in later steps more understandable and computationally feasible.

Then, following the advice of Chi and Zhu (2020), 18 inverse distance weighted spatial weighting matrices with varying distance and power parameters were generated for each dependent variable using the *spdep* package. The Moran's *I* score and significance level for each were then compared. For all variables, the spatial weighting matrix that used the minimum distance threshold, which was 64,299 feet or 12.18 miles, with a power of two resulted in the greatest statistically significant Moran's *I* values. **Table 2** shows the Moran's *I* values and their associated p-values calculated from the selected matrix for each variable. Using the same matrices, Local Indicators of Spatial Association (LISA) cluster maps, shown in **Figure 2**, were created. All variables experienced a clustering of high values around the Greater Nashville Area and clusters of low values in the eastern and western portions of the state. This is particularly noticeable for the home values variable. Also using the same spatial weighting matrices, a lagged version of the grants variable was calculated for use in each regression analysis. **Figure 4** shows

Figure 1

Grant Totals for Each Census Tract

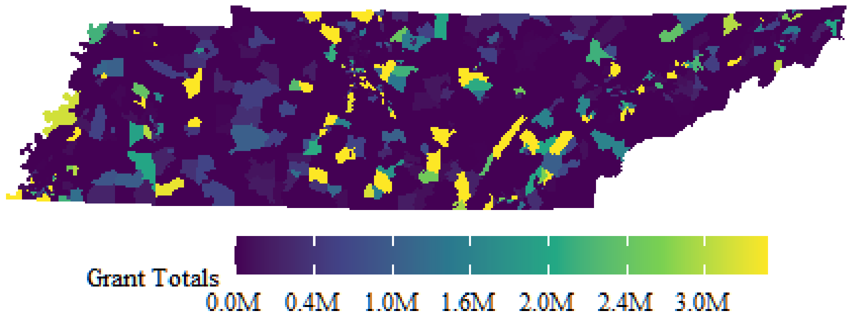


Figure 1: The map above shows the grant total for each census tract in millions. The scale has been truncated to focus on values below the 95th percentile. The maximum grant total is 78.6 million dollars.

the mapped lagged variable associated with each dependent variable's matrix. The variable is generally the same for each matrix, though some slight differences exist based on the outliers that were removed before creating the matrix. These differences are most noticeable around Nashville in the income and home value maps.

Table 1

Variable Summary Statistics

variables	mean	median	IQR	minimum	maximum	skewness	kurtosis
civ_change	7.79	4.99	27.37	-58.50	90.78	0.62	3.65
emp_change	13.03	10.01	28.50	-60.16	98.85	0.66	3.65
inc_change	45.34	42.74	37.32	-33.97	155.75	0.62	3.74
hval_change	61.32	57.08	41.87	-35.05	189.70	0.69	3.81
grants_total	659,312.37	0.00	45,500.00	0.00	78,624,180.00	14.56	269.09

Table 2

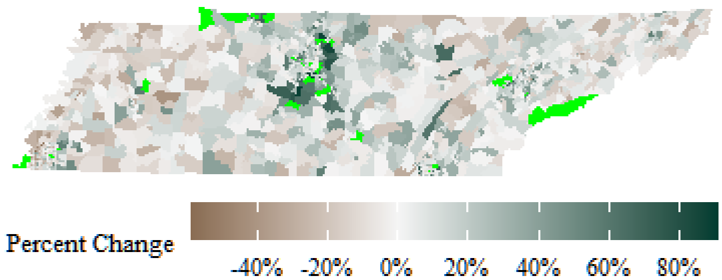
Spatial Autocorrelation in the Dependent Variables

Statistic	Labor Force	Employment	Income	Home Value
Moran's I	0.2350	0.2121	0.1391	0.3890
p-value	1.36E-72	2.57E-59	5.04E-27	8.46E-191

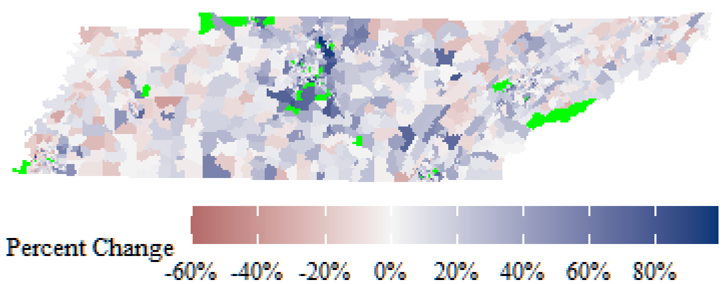
Figure 2

Maps of the Dependent Variables

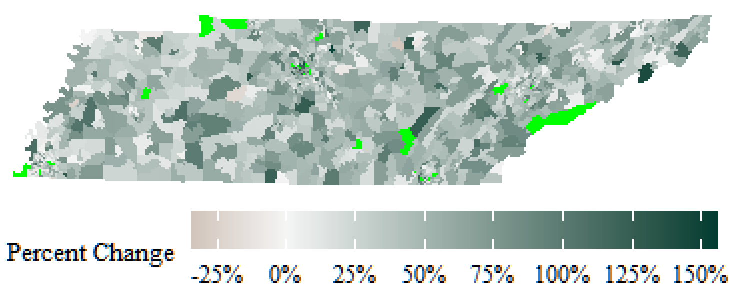
A



B



C



D

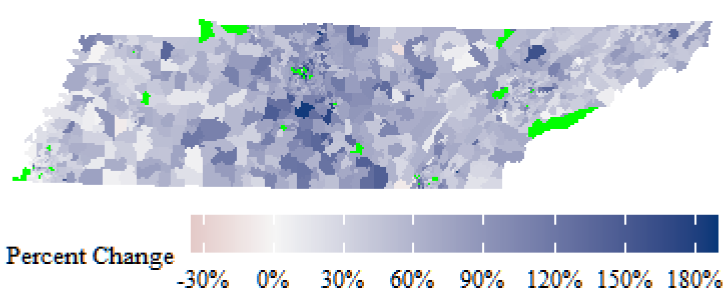


Figure 2: The percent change in each variable from 2011 to 2022 was calculated for each census tract and mapped. (A) shows change in the civilian labor force size. (B) shows how employment counts changed. (C) displays change in median household incomes, and (D) shows change in the median home values. Outliers and null values are represented in bright green.

Figure 3

LISA cluster maps for the Dependent Variables

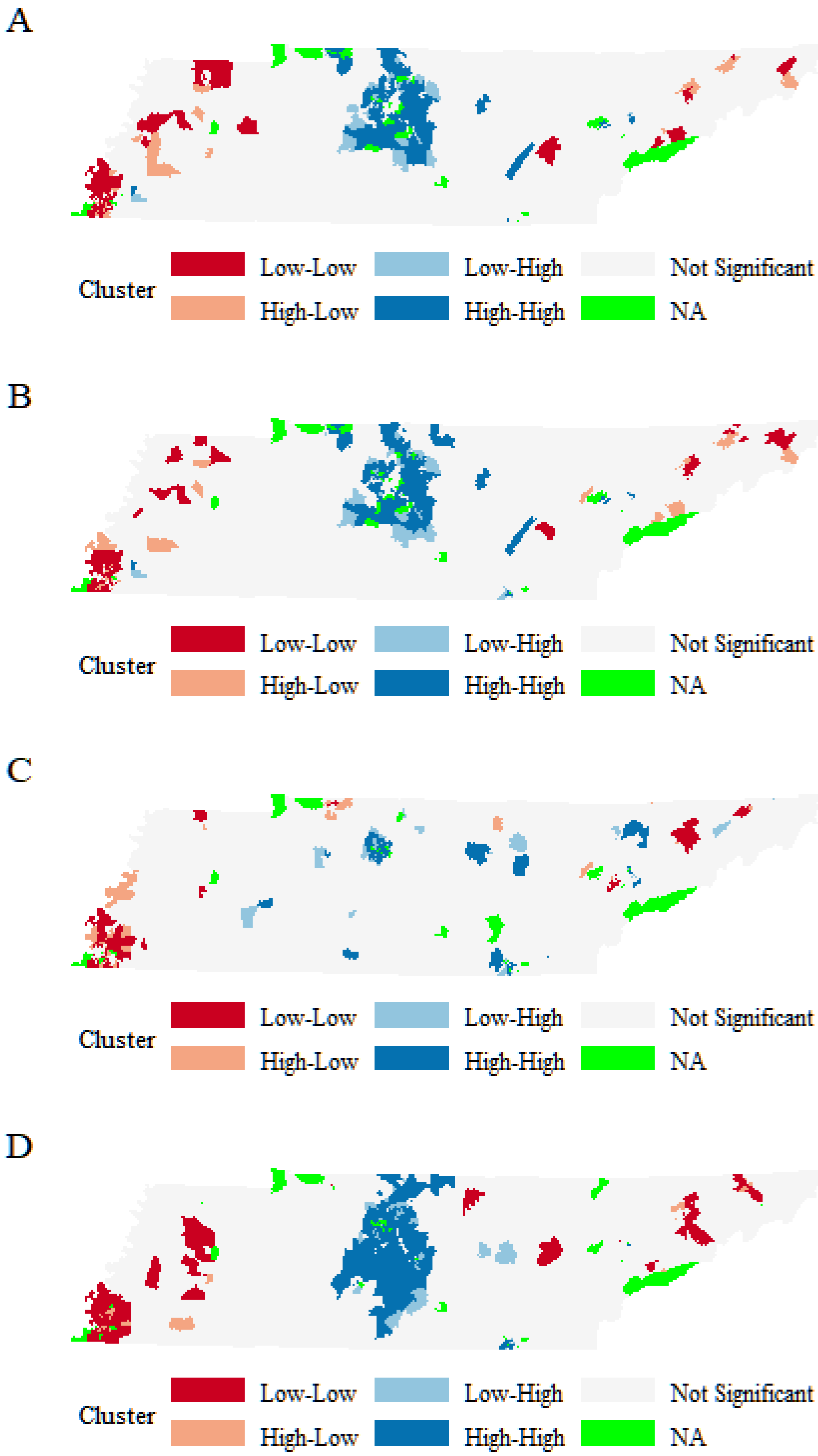


Figure 3: LISA cluster maps were generated for each dependent variable: (A) civilian labor force change, (B) employment count change, (C) household income change, and (D) home value change. High value clusters tend to appear around the Nashville area, while low value clusters tend to appear around the Memphis area.

Figure 4

Maps of Lagged Grant Totals for Each Variable Matrix

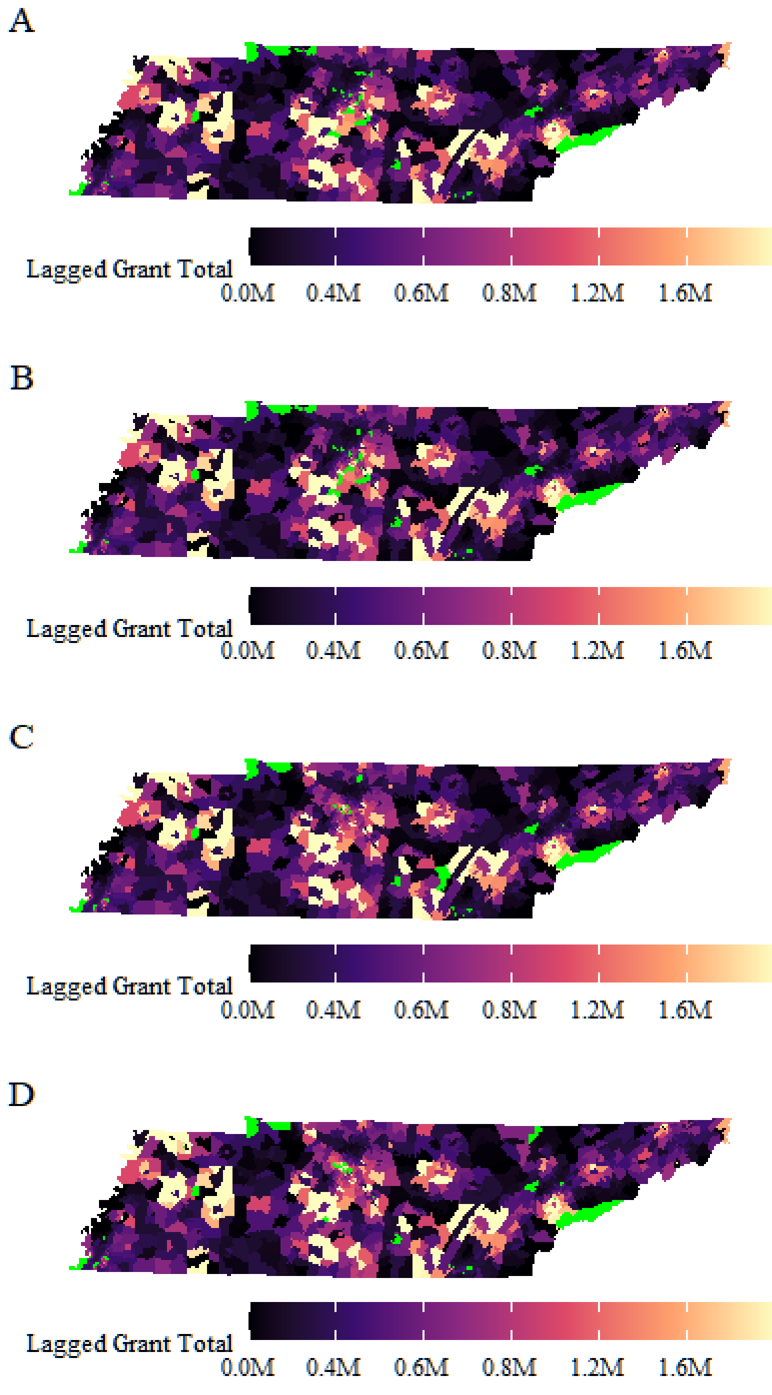


Figure 4: A lagged version of the grant totals was calculated using a spatial weighting matrix corresponding to each dependent variable. The scale bar has been truncated to better show the data below the 95th percentile. The maximum lagged value is 12.6 million. (A) shows the lagged grants calculated from the civilian labor force weighting matrix. (B) shows the lagged grants from the employment matrix. (C) shows the lagged grants for income, and (D) shows the lagged grants for home values.

Regression Analysis

Following the creation of the lagged variables, each variable was fitted with five spatial regression models: the SLX, the SDM, the SDM with a Box-Cox transformation, the SLX with ESF, and the SLX with ESF and a Box-Cox transformation. The spatial filters for the ESF models were constructed using step-wise regression and a queen-contiguity weighting matrix. In all instances where the Box-Cox transformation occurred, the absolute value of the minimum plus one was added to the dependent variable to eliminate negative and zero values before the square root was taken. The fit of the models was then compared by the models' AIC and BIC scores, as well as how well they met the linear modeling assumptions based on the Moran's *I* (MI), Kolmogorov-Smirnov (KS), and Breusch-Pagan (BP) tests.

Table 3 shows the test result statistics of the models created for the civilian labor force change variable. The Box-Cox transformed SLX model with ESF had the lowest AIC and BIC scores, but the BP test was highly significant, indicating the model did not meet the assumption of homoscedasticity. The Box-Cox transformed SDM model had the second lowest AIC and BIC scores, but the KS test statistic was significant, meaning the normality assumption was not met. The other three models had much higher AIC and BIC scores and did not meet the normality or homoscedasticity assumptions. Also, while the SLX model had substantial residual spatial autocorrelation, the SDM and ESF models were successful at significantly reducing the residual spatial autocorrelation as shown by the reduction of the MI score. This successful spatial autocorrelation reduction was consistent across all four variables' models.

The model statistics for the employment change variable can be found in **Table 4**. Much like the labor force change models, the Box-Cox transformed SLX with ESF and SDM models presented the lowest AIC and BIC scores. Once again, the SLX with ESF model residuals were not homoscedastic, and the SDM model residuals were not normally distributed. The other three models' residuals were also not normally distributed, but the standard SLX and SDM models did have homoscedastic residuals.

Table 3

Civilian Labor Force Change Model Statistics

Model	AIC	BIC	Moran	KS	BP
SLX	13140	13161	0.2362 ***	0.0549 ***	6.8443 *
SDM	12932	12958	-0.0091	0.0533 ***	6.4768 *
SDM (Box-Cox)	4857	4883	-0.0127	0.0252 *	2.8927
SLX+ESF	12662	12863	0.0089	0.0418 ***	85.042 ***
SLX+ESF (Box-Cox)	4601	4797	0.0030	0.0210	103.62 ***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Table 4

Employment Change Model Statistics

Model	AIC	BIC	Moran	KS	BP
SLX	13285	13307	0.2138 ***	0.0637 ***	4.9245
SDM	13112	13138	-0.0068	0.0564 ***	5.8616
SDM (Box-Cox)	4895	4921	-0.0090	0.0263 *	3.2441
SLX+ESF	12868	13090	0.0034	0.0428 ***	76.373 ***
SLX+ESF (Box-Cox)	4661	4883	-0.0011	0.0194	72.926 **

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

The model statistics for the income change variable, found in **Table 5**, follow a similar pattern to the AIC and BIC scores. However, neither Box-Cox transformed model passed the normality test, and both passed the homoscedasticity test. Once again, the other three non-transformed models did not meet the normality assumption, but the non-transformed SDM model met the homoscedasticity assumption while the others did not. The model statistics for the home value change variable are in **Table 6**. The KS test statistics for all models indicate highly significant non-normality. Also, only the SDM models have homoscedastic residuals. As with the other variables, the Box-Cox transformed models had the lowest AIC and BIC scores.

Table 5

Income Change Model Statistics

Model	AIC	BIC	Moran	KS	BP
SLX	13888	13909	0.1297 ***	0.0474 ***	6.7924 *
SDM	13818	13844	-0.0125	0.0509 ***	3.8476
SDM (Box-Cox)	5496	5522	-0.0131	0.0261 *	2.1675
SLX+ESF	13620	13841	-0.0030	0.0410 ***	69.464 **
SLX+ESF (Box-Cox)	5292	5513	-0.0157	0.0239 *	39.301

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

Ultimately, none of the models for any of the variables met the assumptions of both normality and homoscedasticity as required for a valid model fit. However, the Box-Cox transformed SDM models produced the nearest-to-normal models that were also homoscedastic and minimized spatial autocorrelation. While the models are not statistically valid, **Table 7** shows the coefficients and statistical significance of the independent variables produced by the Box-Cox transformed SDM models for all the dependent variables. The grant total within a tract has a positive, statistically significant impact on the changes in the civilian labor force and employment, while the lagged grant total is not statistically significant. Both the grant total and

Table 6

Home Value Change Model Statistics

Model	AIC	BIC	Moran	KS	BP
SLX	14200	14221	0.3731 ***	0.0542 ***	9.3993 **
SDM	13766	13793	-0.0409	0.0699 ***	5.6151
SDM (Box-Cox)	5240	5267	-0.0431	0.0417 ***	1.0921
SLX+ESF	13449	13734	-0.0062	0.0582 ***	123.74 ***
SLX+ESF (Box-Cox)	4925	5209	-0.0018	0.0304 **	91.292 ***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

lagged grant totals have a positive, statistically significant impact on changes in income, and only the lagged grant total has a significant relationship with home value changes. For all the dependent variables, the lagged response variable was positive and statistically significant.

Reflection

This research did not accomplish its objective of creating a statistically valid regression model for each dependent variable. Attempts to correct the data's skewness through outlier removal and Box-Cox transformations were unsuccessful at producing normally distributed and homoscedastic model residuals. Even when models were nearly normal and homoscedastic, the AIC and BIC scores were relatively high, indicating that the models were not a good fit. This research also attempted to identify the relationship between FastTrack grants and the four dependent variables. While the coefficients in **Table 7** provide some indication of that relationship, the results cannot be treated as accurate representations of the relationship due to the poor model fit and failure to meet the normality assumption. Therefore, they should not be used to make any conclusions about the FastTrack program's impact.

Table 7

SDM (Box-Cox) Model Coefficients for Each Dependent Variable

Model	Labor Force	Employment	Income	Home Value
(Intercept)	3.6447 ***	4.1031 ***	5.8097 ***	3.0999 ***
grants_total_mil	0.0336 **	0.0330 **	0.0321 **	0.0092
lag.grants_total_mil	0.0633	0.0761	0.1449 **	0.1422 ***
lag.Y	0.5429 ***	0.5097 ***	0.3278 ***	0.6694 ***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

While the exact source of the data's skewness and the models' poor fit will likely remain unknown, several potential problem points exist, including the choice in the years for percent-change calculations, the chosen geographic scale, and the necessary inclusion of the WX term. The choice of start and end points of the study period matters because the change in the variables and amount of grants provided was likely not consistent year-over-year. Picking different start and end years could potentially lead to different calculations of the percent change variables, meaning a different data distribution and levels of correlation. When simply picking a start and end date, this problem cannot be avoided. However, future analysis could implement temporal analysis techniques to consider the year-over-year impact, so that yearly changes and grant totals are identified explicitly.

Furthermore, this research deviated from most existing research on economic development by using census tracts rather than counties for its analysis. While the choice to use census tracts allowed for a more granular study, it introduced several challenges. First, because of the modifiable areal unit problem, the county-level estimates, data distribution, and summary statistics are likely quite different than the tract-level estimates, distribution, and summary statistics. While a skewed distribution in the tract-level data inhibited the fitting of valid models, a county-level analysis may not have the same problem. Second, because census tracts change boundaries following every census, the data from 2022 had to be aggregated to the 2011 boundaries before it could be used, which introduced the potential for inaccuracies. In contrast, county borders rarely change, so aggregating the 2022 estimates to the 2011 borders would not be necessary.

Third, after aggregating the project data into the census tracts, most tracts did not have a project within them, meaning the grant totals for those tracts were zero. To prevent the high count of zero values from skewing the model, the lagged grant total variable, the WX term, was introduced as a proxy variable that represented the impact of nearby projects. However, the introduction of the WX term explicitly introduced spatial autocorrelation into the model. According to D. A. Griffith (personal communication, April 17, 2024), this could have caused the ESF models to double-adjust for spatial autocorrelation, resulting in unexpected increases in the residual heteroscedasticity. If the analysis were conducted at the county level, using the WX term as a proxy variable would not be necessary, as most counties have received a FastTrack project during the study period. Then, the merit of the WX term in improving model fit could be tested. With these points in mind, any future research should ensure that the chosen geographic scale does not require the use of spatial regression terms without first proving that they improve model fit. Doing so should minimize the risk of introducing spurious variables that negatively impact the residual homoscedasticity or normality.

In future research, once the points of concern mentioned in this paper have been resolved, additional FastTrack project variables, such as the number of promised jobs, could be valuable in improving model fit. It is possible that the grant totals alone do not provide a clear enough indication of program accomplishments. Introducing additional variables could help determine the circumstances under which the FastTrack program performs well. Also, rather than using the grant totals, considering the three FastTrack grant types separately could provide the TNECD valuable insight into what program components are most effective.

Conclusion

After this study's conclusion, the analysis was still unable to provide conclusive evidence regarding the Tennessee FastTrack grant program's efficacy. While the results indicated that there may be a positive relationship between the grants and changes in the civilian labor force, employment, household income, and home values when examined at the census tract level, an approximation of the true relationship cannot be ascertained with statistical certainty. Therefore, additional research into the program is necessary to confidently determine the impact of the FastTrack program on local economies. This research could not overcome the challenges presented by the data's underlying distribution. To avoid this problem, future research could work with different data distributions by analyzing metrics at the county level rather than the census tract level, exploring year-over-year changes, or introducing additional explanatory variables. Of these three options, county-level analysis has the clearest benefits, though a combination of all three would provide additional insight into the FastTrack program.

References

Bundrick, J., & Snyder, T. (2018). Do business subsidies lead to increased economic activity? Evidence from Arkansas's Quick Action Closing Fund. *The Review of Regional Studies*, *48*(1), 29–53. https://doi.org/10.52324/001c.8005

Bundrick, J., & Yuan, W. (2019). Do targeted business subsidies improve income and reduce poverty? A synthetic control approach. *Economic Development Quarterly*, *33*(4), 351–375. https://doi.org/10.1177/0891242419875502

Chi, G., & Zhu, J. (2020). *Spatial Regression Models for the Social Sciences*. SAGE Publications.

Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science*, *52*(2), 210–239. https://doi.org/10.1111/j.1467-9787.2011.00726.x

Edmiston, K. D. (2004). The net effects of large plant locations and expansions on county employment. *Journal of Regional Science*, *44*(2), 289–319. https://doi.org/10.1111/j.0022-4146.2004.00338.x

Esiyok, B., & Ugur, M. (2017). A Spatial Regression Approach to FDI in Vietnam: Province-level Evidence. *The Singapore Economic Review*, *62*(02), 459–481. https://doi.org/10.1142/s0217590815501155

Gonzales, C., Kerlin, M., Schaff, R., & Tucker-Ray, S. (2019, September 13). *How state and local governments win at attracting companies*. McKinsey & Company. https://www.mckinsey.com/industries/public-sector/our-insights/how-state-and-local-governments-win-at-attracting-companies#/

Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy*, *118*(3), 536–598. https://doi.org/10.1086/653714

Griffith, D. A. (2013). Better articulating normal curve theory for introductory mathematical statistics students: Power transformations and their back-transformations. *The American Statistician*, *67*(3), 157–169. https://doi.org/10.1080/00031305.2013.801782

Griffith, D. A., & Chun, Y. (2016). Evaluating Eigenvector Spatial Filter Corrections for Omitted Georeferenced Variables. *Econometrics*, *4*(2), 29. https://doi.org/10.3390/econometrics4020029

Kline, P., & Moretti, E. (2014). Local Economic Development, agglomeration economies, and the big push: 100 Years of Evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics*, *129*(1), 275–332. https://doi.org/10.1093/qje/qjt034

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

Parilla, J., & Liu, S. (2018, March). *Examining the Local Value of Economic Development Incentives*. Brookings Institution. https://www.brookings.edu/wp-content/uploads/2018/02/report_examining-the-local-value-of-economic-development-incentives_brookings-metro_march-2018.pdf

Slattery, C., & Zidar, O. (2020). Evaluating state and local business incentives. *Journal of Economic Perspectives*, *34*(2), 90–118. https://doi.org/10.1257/jep.34.2.90

Takano, K., Tsutsumi, M., & Kikukawa, Y. (2018). Spatial modeling of industrial location determinants in Japan: Empirical analysis using spatial econometric approaches. *Review of Urban and Regional Development Studies*, *30*(1), 26–43. https://doi.org/10.1111/rurd.12073

Tennessee Department of Economic and Community Development. (n.d.). *Incentives & Grants*. Tennessee Department of Economic and Community Development. https://tnecd.com/advantages/incentives-grants/

Tennessee Department of Economic and Community Development. (November 8, 2023). *Projects with Contracted FastTrack Grants since 2011* [Unpublished raw data]

Thayn, J. B., & Simanis, J. M. (2013). Accounting for spatial autocorrelation in linear regression models using spatial filtering with eigenvectors. *Annals of the Association of American Geographers*, *103*(1), 47–66. https://doi.org/10.1080/00045608.2012.685048

U.S. Census Bureau. (2022). TIGER/Line 2022 Tennessee Census Tract Boundaries [Dataset]. https://www2.census.gov/geo/tiger/TIGER2022/TRACT/tl_2022_47_tract.zip

U.S. Census Bureau. (2023a). *DP03: Selected Economic Characteristics*: 2022 ACS 5-year Estimates for Tennessee [Dataset]. https://data.census.gov/table/ACSDP5Y2022.DP03?g=040XX00US47$1400000

U.S. Census Bureau. (2023b). *DP04: Selected Housing Characteristics*: 2022 ACS 5-year Estimates for Tennessee [Dataset]. https://data.census.gov/table/ACSDP5Y2022.DP04?g=040XX00US47$1400000

U.S. Census Bureau. (2023c). TIGER/Line with 2007-2011 Detailed Tables for Tennessee Census Tracts [Dataset]. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.2011.html#list-tab-1656998034

Villaverde, J., & Maza, A. (2012). Foreign direct investment in Spain: Regional distribution and determinants. *International Business Review*, *21*(4), 722–733. https://doi.org/10.1016/j.ibusrev.2011.08.004