

Assessment of Spatial Correlation Patterns of Unit Price Bids and External Factors

Minsoo Baek and Baabak Ashuri, Ph.D.
Georgia Institute of Technology
Atlanta, Georgia

The generalized linear modeling (GLM) approach overlooks the spatial correlation between the unit price bids and the geographical location of a project, leading to over or underestimating the significance of explanatory variables. State departments of transportation (DOTs) often encounter over or underestimation of their construction costs, which can cause cost overruns, inefficient budget allocation, and project delays or cancellation. The overarching objective of this study is to explain variation in the submitted unit price bids for asphalt line items used in highway projects. The study applies geographically weighted regression (GWR) analysis to explore spatial variations of relationships between submitted unit price bids and explanatory variables and to create a set of local models with the capability to describe variations in submitted unit price bids. The regression analysis of submitted unit prices for asphalt line items, drawn from in highway pavement projects in the state of Georgia between 2008 and 2015, examines key variables, including the quantity of the bid item, the total contract price, the number of bids, the dollar value of a project, and the price index of asphalt cement. The results indicate significant spatial variation in the relationship between submitted unit price bid and the identified explanatory variables. Not only is GWR analysis capable of more accurately describing variation in submitted unit price bids, but it also provides a means of exploring the geographical variation of construction costs in a graphical manner.

Key Words: Asphalt Line Items, External Factors, Spatial Correlation, Unit Price Bids

Introduction

State departments of transportation (DOTs) experience significant cost differentials resulting from variability in construction market and economic conditions. When determining construction costs, DOTs must deal with uncertainty caused by significant cost variation. Such uncertainty is exacerbated by significant variability in the construction market and economic conditions. A high degree of uncertainty leads to significant variations in highway construction costs. The major consequences of such variations are poor investment decisions and inaccurate estimations of construction costs, that is, over or underestimation of construction costs, which can cause cost overruns, inefficient budget allocation, and project delays or cancellation. Thus, state DOTs must be capable of analyzing uncertainty related to the impact of geographical variability of the construction market and economic conditions on construction costs.

Several studies have emphasized spatial correlation in estimations of construction costs. For instance, Zhang et al. (2014) used surface interpolation methods to analyze cost indexes for studying spatial correlation with geographical locations and found a significant spatial correlation between cost indexes and specific geographical locations. They concluded that an adjustment in cost data increased the accuracy of cost estimates for construction projects. Another study carried out by Zhang et al. (2016) analyzed RSMean's city cost index (CCI) to develop location adjustment factors for realistic cost estimates and concluded that improvement in the geographical interpretation of CCI by incorporating local economic conditions increases the accuracy of cost estimates for a construction project. Migliaccio et al. (2009) also used the RSMeans' CCI national reference data to conduct a spatial analysis and showed a strong spatial correlation between proximity and CCI value. Migliaccio et al. (2012) also used spatial analysis to explain the spatial patterns of construction costs with socioeconomic variables such as the population, the population growth percentile, and he household growth rate. The authors concluded that the impact of each covariate differed from state to state. Although these studies indicate a significant spatial correlation between the cost and geographical location of a construction project, they do not analyze actual construction projects or the impact of covariates such as project characteristics, construction market, and economic conditions on construction costs.

Numerous studies have identified factors that affect the costs of construction. For example, Herbsman (1986) applied regression analysis to develop a forecasting model for highway construction projects and concluded that construction market factors such as the costs of material, labor, and equipment and the annual bid volume significantly impacted the total cost of a construction project. Wilmot and Cheng (2003), conducting a regression analysis to identify factors and develop an estimation model for construction costs, determined the impact of several key factors such as location, bid volume, and prices of material. In another study, Akintoye and Skitmore (1993) used a regression approach to identify factors affecting the tender price index and claimed that economic factors such as the unemployment rate and manufacturing profitability significantly impacted the tender price index. To examine variations in the bid price, Wang and Liu (2012) used construction market factors such as number of bidders, asphalt price index, and diesel price index to conduct regression analysis. Shrestha and Pradhananga (2010) also studied the effect of competitive bidding on variations in bid prices using 113 public street projects in Clark County, Nevada. Through regression analysis, the authors concluded that the number of bidders significantly impacted bid prices. Thus, although many studies have examined the impact of the construction market and economic conditions on construction costs using regression analysis, few have focused on either assessing spatial variation in the construction costs of highway construction projects or analyzing spatial relationships between construction costs and external factors that impact construction costs.

Research Objectives

The main objective of this study is to explain variations in submitted unit price bids for asphalt line items used in highway construction projects by incorporating external factors: construction market and economic condition factors. To achieve this main objective, this study also has the following sub-objectives:

1) To identify potential factors that affect submitted unit price bids
2) To develop explanatory models for describing variations in submitted unit price bids
3) To identify spatial correlation between submitted unit price bids and external factors

Research Methodology

This study shows a significant spatial correlation between the cost and geographical location of a construction project—a correlation that a generalized linear modeling (GLM) approach may overlook. Using geographically weighted regression (GWR) analysis, we develop explanatory models for describing variations in submitted unit price bids. GWR enables the identification of parameters for each location in space and complex spatial variations in parameter estimates (Brunsdon et al. 1996). In addition, hot spot analysis to measure spatial variation in submitted unit price bids and geographical location of a project. A hot spot analysis is a statistical method for assessing geographical clustering, which identifies the locations of statistically significant high- and low-value clusters of construction costs by evaluating each feature within the context of neighboring features and against all features in the dataset.

Hypothesis Test of the Hot Spot Analysis:

Null hypothesis (H_0): Submitted unit price bids do not significantly differ across various geographical locations of projects.

Alternative hypothesis (H_1): Submitted unit price bids significantly differ across various geographical locations of projects.

In addition, this study applies a Moran's I test on the regression residuals of both GLM and GWR analysis to determine whether they are spatially auto-correlated. Since the spatial autocorrelation of regression residuals implies significant under or overestimates of the variances of regression coefficients, it should be diagnosed. Spatial autocorrelation is determined by Moran's index, whose value is in a range between – 1 and +1. A Moran's index close to 1 indicates clustered residuals while a value close to -1 indicates dispersed residuals. If a Moran's index is close to 0, it indicates that residuals are randomly distributed over all geographical locations (no autocorrelation) (Li et al. 2007).

Hypothesis Test of Moran's Index:

Null hypothesis (H_0): Regression residuals are randomly distributed with the geographical locations of projects.

Alternative hypothesis (H_1): Regression residuals are not randomly distributed with the geographical locations of projects.

A spatial interpolation tool, the natural neighbor, is used to visually assess the spatial variation of the relationship between submitted unit price bids and explanatory variables. The degree of the relationship between submitted unit price bids and explanatory variables are explored using three classifications: (1) high correlation; (2) medium correlation; and (3) low correlation. Lastly, to explain spatial variation, this study uses seven districts of the Georgia Department of Transportation (GDOT). Although spatial variation may not be exactly explained by the seven districts, it will serve as a guide to accurately identifying the geographical locations for this study.

Dataset

This work applied regression modeling to explain variations in the unit prices for bid line items by explanatory variables. It used the submitted unit price bid for one asphalt line item, hot-mix recycled asphaltic concrete, the most common pavement material used by state DOTs in the United States (Kandhal et al. 1995). This paper collected 1,391 historical sample data from the GDOT between 2008 and 2015. This study focuses on pavement projects, including widening and resurfacing projects. Descriptive statistics appear in Table 1.

Table 1.

Descriptive statistics of unit price bids for hot-mix recycled asphaltic concrete

	Total	Resurfacing	Widening
Number of Projects	1,391	1,034	357
Mean	68.037	66.496	72.498
Min	44.47	44.47	51.970
Max	102.03	95.42	102.030

We classified the explanatory variables for regression modeling into two groups: project-specific and external factors. Twelve candidate explanatory variables were displayed for regression analysis. We used detailed project-specific information collected in the data compilation to develop project-specific factors, including the number of bid line items, the total bid price, the duration and size of a project, and the number of bidders of the project. Depending on the project-specific conditions, the level of uncertainty varies. Thus, project-specific conditions should be taken into account in explanations of variations in submitted unit price bids. As rapidly changing market conditions also have a significant impact on estimating construction costs, this study includes the external variables of construction market and economic conditions to measure their impact on submitted unit price bids. The following lists the variables examined in this study:

Project-specific Variables:

- Project duration: The period between notice to proceed and date of completion (Retrieved from the GDOP GeoPi system) (in days)
- Quantity of the bid item: The volume of the asphalt line item in the submitted bid (retrieved from the BidTabs database) (in tons)

- Total contract price: The lowest total bid price submitted by highway contractors that bid on a project (retrieved from the BidTabs database) (in dollars)
- Number of bidders: The number of highway contractors that submitted bids for the project (retrieved from the BidTabs database)

External Variables:

- Total number of projects at the state level: The number of projects awarded in the same month that a project is awarded in the state of Georgia (retrieved from the Bid Express online bidding system)
- Total dollar value of projects at the state level: The total dollar value of projects awarded in the same month that a project is awarded in the State of Georgia (retrieved from the Bid Express online bidding system) (in dollars)
- Georgia fuel price index: The average statewide selling price of unleaded regular gasoline and diesel fuel in Georgia (retrieved from the GDOT website) (in dollars/gallon)
- Georgia asphalt cement price index: The average selling price of asphalt cement, collected from approved local asphalt cement suppliers as reported in the GDOT monthly survey (retrieved from the GDOT Office of Materials) (in dollars/ton)
- Number of hires: The total number of additions to the payroll in the U.S. construction industry during the month that a project is awarded, provided by the Job Openings and Labor Turnover Survey (JOLTS) (retrieved from the U.S. Bureau of Labor Statistics)
- Unemployment: The number of people eligible to work but unable to find a job, measured at the county level in the State of Georgia (retrieved from the U.S. Bureau of Labor Statistics) (in the number)
- Value of construction put in place (i.e., pavement): The monthly estimate of the total dollar value of pavement construction work completed in the southern region of the United States (retrieved from the U.S. Census Bureau) (in millions of dollars)
- Population: The number of individuals who reside in the State of Georgia, measured at the county level. (retrieved from the U.S. Census Bureau)

Results and Discussion

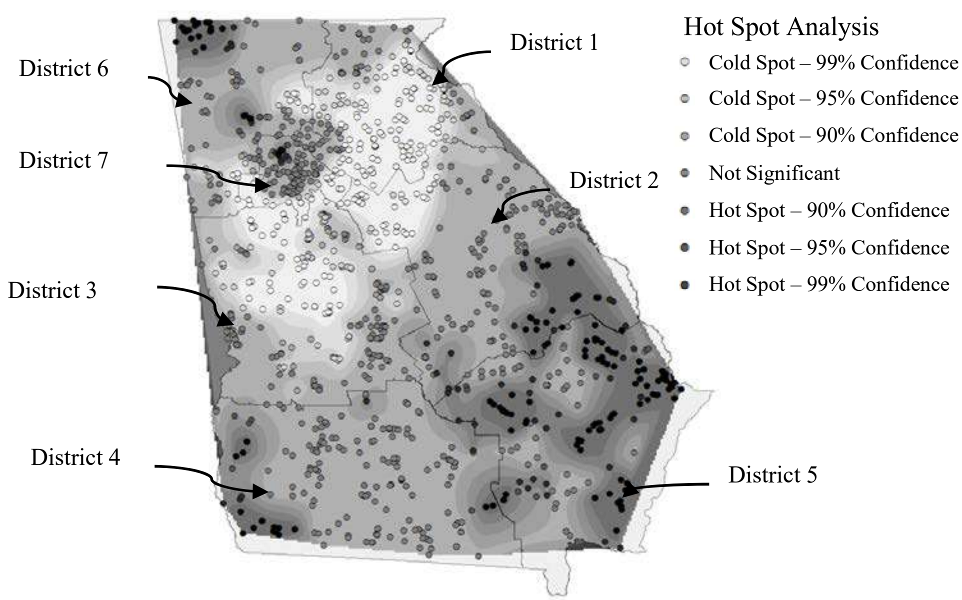

Figure 1: Hot spot analysis for submitted unit price bids

Hotspot Analysis:

To diagnose the presence of significant spatial variation between submitted unit price bids and geographical location of the projects, this study entails a hotspot analysis. As shown in Figure 1, submitted unit price bids in hot spot areas, mostly Districts 4, 5, and 6, are significantly higher than the overall mean of submitted unit price bids at a 99% confidence level. One explanation for this finding is that Districts 4 and 5 have difficulty procuring important resources such as labor, materials, and equipment. Another explanation is that very few asphalt plants are located in southern Georgia. In addition, District 6 is a mountainous terrain, which might decrease the level of productivity on projects and increase the unit prices on submitted bids for asphalt line items. Conversely, submitted unit price bids in cold spot areas, mostly Districts 1 and 3, are significantly lower than the overall mean of unit prices. One explanation for this finding is that Districts 1 and 3 have better accessibility to procuring asphalt cement and aggregates from North Georgia. Thus, Districts 1 and 3 receive relatively lower submitted unit price bids than other districts. One conclusion from these findings is that significant spatial variation exists between unit price bids and geographical locations, which rejects the null hypothesis of this test.

Generalized Linear Model (GLM):

This study entails a stepwise selection procedure for identifying key variables for regression modeling. To select the optimal subset model, Akaike's information criterion (AICc) and the adjusted R squared are used. Through a stepwise procedure, we identify five significant variables—the quantity of the bid item, the total contract price, the number of bidders, the total dollar value of a project at the state level, and the asphalt cement price index—for regression modeling. We apply ordinary least squares (OLS) to estimate the parameters of significant variables in the regression model. Table 2 presents the OLS regression model with identified variables. Since the variance inflation factors (VIFs) for identified variables are less than 10, a multicollinearity issue does not exist in the model. In addition, the identified variables have the power to explain variation in the submitted unit prices at a 95% confidence level (P-value < 0.05). Both the quantity of the bid item and the number of bidders show a negative relationship with submitted unit prices while the total contract price, the total dollar value of projects, and the asphalt cement price exhibit a positive relationship with submitted unit price bids. Overall, the developed OLS regression model explains 32% of the variation in submitted unit prices for asphalt line items.

Table 2

Results of ordinary least squares (OLS) regression

Model	Coefficient	Standard Error	t value	P-value	VIF
Intercept	66.950	1.780	37.615	0.000	----
Quantity of the bid item (in transformed natural logarithmic form)	-2.071	0.147	-14.071	0.000	1.100
Total contract price	2.840×10^{-7}	3.226×10^{-8}	8.803	0.000	1.161
Number of bidders	-1.070	0.119	-8.966	0.000	1.024
Total dollar value of projects at the state level	2.050×10^{-8}	5.103×10^{-9}	4.016	0.000	1.108
Asphalt cement price index	0.041	0.002	18.014	0.000	1.066
Akaike's information criterion (AICc)	9385.030				
Adjusted R-squared	0.312	32%			

Table 3

Global Moran's I summary for OLS regression

Types	Values
Moran's index	0.372
z-score	17.795
p-value	0.000

Table 3 provides the results of Moran's I test. Since the Moran's index is not close to 1 and the z-score is greater than 2.58 (99% confidence level), the residuals of the OLS regression model spatially auto-correlated (or clustered). The results of this test indicate a rejection of the null hypothesis—that the regression residuals are randomly distributed with the geographical locations of projects. Thus, one should remedy autocorrelation issues to prevent over or underestimations of the significance of explanatory variables.

Geographically Weighted Regression (GWR):

With the identified variables from the stepwise select process, this work applies geographically weighted regression analysis to estimate local variable coefficients for each location. Table 4 provides the results of GWR analysis including the mean, minimum, median, maximum, and standard deviation of the coefficients of the local models. The coefficients of the identified variables show significant variability with regard to the impact of the identified variables on submitted unit price bids. In addition, the results of the analysis indicate that the GWR model can explain 51.2 % of variation in submitted unit prices for asphalt line items.

Table 4

Results of Geographically Weighted Regression (GWR)

Model	**Mean**	**Minimum**	**Median**	**Maximum**	**Standard Deviation**
Intercept	65.519	52.342	64.443	91.671	8.701
Quantity of the bid item (in transformed natural logarithmic form)	-2.015	-3.044	-1.976	-0.923	0.353
Total contract price	3.000×10^{-7}	1.085×10^{-8}	3.176×10^{-7}	4.452×10^{-7}	1.013×10^{-7}
Number of bidders	-0.715	-3.868	-0.583	0.113	0.624
Total dollar value of projects awarded in the same month at the state level	2.248×10^{-8}	-1.423×10^{-8}	2.305×10^{-8}	4.293	1.042×10^{-8}
Asphalt cement price index	0.039	0.015	0.040	0.055	0.007
Akaike's information criterion (AICc)	8994.900				
Adjusted R-squared	0.512	51.2%			

To visually examine the spatial variations of relationships between submitted unit price bids and key identified variables, we employ a spatial interpolation tool, the natural neighbor, on the coefficients of the variables. Figure 2 depicts spatial variation in the relationship between submitted unit price bids and the identified variables. The quantity of the bid item exhibits a significant negative relationship with the submitted unit price bids in Districts 3 and 5, and the total contract price shows a significant positive relationship with the submitted unit price bids in the western regions of Georgia, particularly District 3. In addition, although the number of bidders and the number of submitted unit price bids have a significant positive relationship in some parts of Districts 2 and 4, they do not show this relationship in the northern, middle, and southern regions of Georgia. The total dollar value of a project at the state level is more strong correlated in the northern and eastern regions of Georgia, Districts 3, 6, and 7, than in other regions. In addition, the asphalt cement price index is more strongly correlated in the northern regions of Georgia, Districts 1, 3, 6, and 7, than in other regions.

A summary of Moran's I test for the GWR model appears in Table 5, which shows that Moran's index significantly decreases, and its values are close to 0. These results indicate that the GWR model minimizes uncertainty in the significance of the identified variables. However, the z-score is still greater than 2.58 at a 99% confidence level, which indicates that the residuals of the GWR model are spatially clustered. Thus, to improve the GWR model, one must identify other potential variables.

Table 5

Global Moran's I summary for GWR

Types	**Values**
Moran's index	0.192
z-score	9.199
p-value	0.000

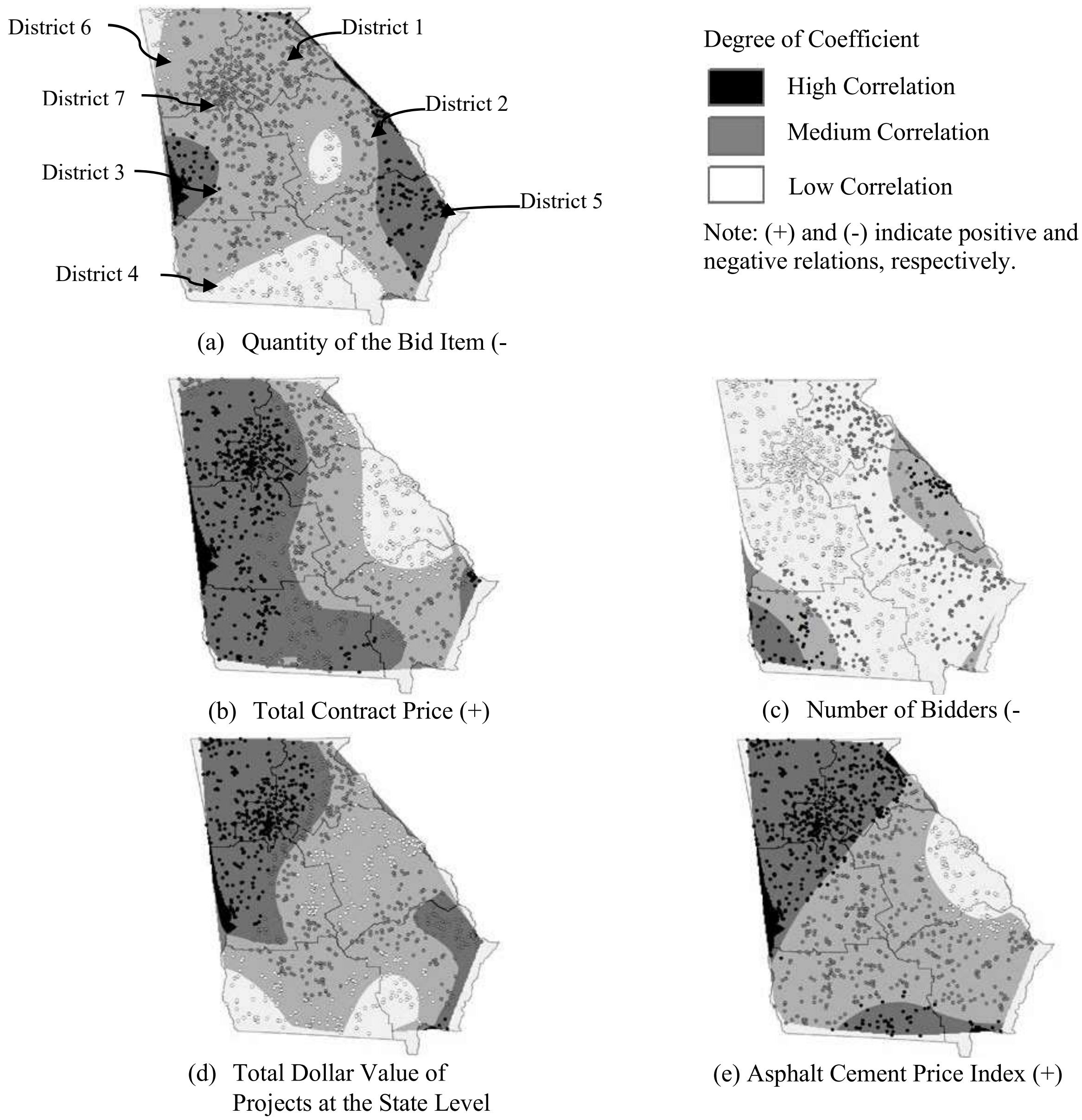

Figure 2: Spatial interpolation for identified variables: (a) the quantity of the bi item (-); (b) total contract price (+); (c) number of bidders (-); (d) total dollar value of projects at the state level (+); and (e) asphalt cement price index (+)

Conclusion and Recommendations

The overall contribution of this study to the body of knowledge is a preliminary understanding of the relationship between construction costs and the geographical location of a project and spatial relationships between construction costs and external covariates, including project characteristics, construction markets, and economic conditions. The study used GWR (geographically weighted regression) to spatially examine variations in submitted unit prices for bid line items. The identified variables—the quantity of the bid item, the total contract price, the number of bidders, the total dollar value of a project at the state level, and the asphalt cement price index—showed power to explain variation in submitted unit prices. Using a spatial interpolation tool, the author of this study examined the significance of spatial correlations between submitted unit price bids and the identified variables and concluded that when cost estimators use the identified variables for estimating construction costs, they should assign various weights to the identified variables with the spatial relationships. This study proved that the use of GWR analysis provides greater capability of describing variations in submitted unit price bids for highway construction projects. In addition, the proposed approach provides insight into the exploration of geographical variation in a graphical manner. However, as Moran's I test for the GWR model shows autocorrelation, indicating clustered residuals with geographical locations, the development of a more accurate GWR model requires further investigation of other potential variables, which should minimize the over or underestimation of the significance of explanatory variables. The findings of this study should help state DOTs determine more accurate construction costs by considering the geographical locations of the projects. For instance, to adjust the preliminary cost projections for a project, they could use a cost estimator that takes the spatial patterns of construction costs into account. By employing a cost estimator that uses spatial relationships between construction costs and important variables, a state DOT is able to adjust and prepare bids prior to letting a project.

References

Akintoye, A., and Skitmore, M. (1993). "Macro models of UK construction contract prices." Civil Engineering Systems, 10(4), 279 – 299.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996). "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." Geographical Analysis, 28(4), 281 – 298.

Herbsman, Z. (1986). "Model for Forecasting Highway Construction Cost." Transportation Research Record.

Kandhal, P. S., Rao, S. S., Watson, D. E., and Young, B. (1995). "Performance of Recycled Hot Mix Asphalt Mixtures." NCAT Report 95-01. Auburn, AL.

Li, H., Calder, C. A., and Cressie, N. (2007). "Beyond Moran's I: Ttesting for Spatial Dependence Based on the Spatial Autoregressive Model." Geographical Analysis, 39(4), 357-375.

Migliaccio, G. C., Guindani, M., D'Incognito, M., and Zhang, L. (2012). "Empirical Assessment of Spatial Prediction Methods for Location Cost-Adjustment Factors." Journal of Construction Engineering and Management, 139(7), 858 – 869.

Migliaccio, G. C., Zandbergen, P., and Martinez, A. A. (2009). "Assessment of Methods for Adjusting Construction Cost Estimates by Geographical Location." In Construction Research Congress 2009: Building a Sustainable Future, 886 – 895.

Shrestha, P. P., and Pradhananga, N. (2010). "Correlating Bid Price with the Number of Bidders and Final Construction Cost of Public Street Projects." Transportation Research Record: Journal of the Transportation Research Board, 2151, 3 – 10.

Wang, Y., and Liu, M. (2012). "Prices of Highway Resurfacing Projects in Economic Downturn: Lessons Learned and Strategies Forward." Journal of Management in Engineering, 28(4), 391 – 397.

Wilmot, C. G., and Cheng, G. (2003). "Estimating Future Highway Construction Costs." Journal of Construction Engineering and Management, 129(3), 272 – 279.

Zhang, S., Migliaccio, G. C., Zandbergen, P. A., and Guindani, M. (2014). "Empirical Assessment of Geographically Based Surface Interpolation Methods for Adjusting Construction Cost Estimates by Project Location." Journal of Construction Engineering and Management, 140(6), 04014015.

Zhang, S., Bogus, S. M., Lippitt, C. D., and Migliaccio, G. C. (2016). "Estimating Location-Adjustment Factors for Conceptual Cost Estimating Based on Nighttime Light Satellite Imagery." Journal of Construction Engineering and Management, 143(1), 04016087.