Published in final edited form as:

J Constr Eng Manag. 2013 July 1; 139(7): 858–869. doi:10.1061/(ASCE)CO.1943-7862.0000654.

Empirical Assessment of Spatial Prediction Methods for Location Cost Adjustment Factors

Giovanni C. Migliaccio¹ [Assistant Professor], Michele Guindani² [Assistant Professor], Maria D'Incognito³ [PhD Candidate], and Linlin Zhang⁴ [PhD Candidate]

Giovanni C. Migliaccio: gianciro@uw.edu

¹University of Washington, Department of Construction Management, Box 351610, Seattle, WA 98195-1610, phone: +1 (206) 685-1676, fax: +1 (206) 685-1976

²The University of Texas MD Anderson Cancer Center, Department of Biostatistics - Unit 1411, P. O. Box 301402, Houston, TX 77230-1402, phone: +1 (713) 563-4285

³Politecnico di Bari, DICATECh - Department of Civil, Environmental, Building Engineering, and Chemistry, Bari, Italy

⁴Rice University, Department of Statistics, Houston, TX

Abstract

In the feasibility stage, the correct prediction of construction costs ensures that budget requirements are met from the start of a project's lifecycle. A very common approach for performing quick-order-of-magnitude estimates is based on using Location Cost Adjustment Factors (LCAFs) that compute historically based costs by project location. Nowadays, numerous LCAF datasets are commercially available in North America, but, obviously, they do not include all locations. Hence, LCAFs for un-sampled locations need to be inferred through spatial interpolation or prediction methods. Currently, practitioners tend to select the value for a location using only one variable, namely the nearest linear-distance between two sites. However, construction costs could be affected by socio-economic variables as suggested by macroeconomic theories.

Using a commonly used set of LCAFs, the City Cost Indexes (CCI) by RSMeans, and the socio-economic variables included in the ESRI Community Sourcebook, this article provides several contributions to the body of knowledge. First, the accuracy of various spatial prediction methods in estimating LCAF values for un-sampled locations was evaluated and assessed in respect to spatial interpolation methods. Two Regression-based prediction models were selected, a Global Regression Analysis and a Geographically-weighted regression analysis (GWR). Once these models were compared against interpolation methods, the results showed that GWR is the most appropriate way to model CCI as a function of multiple covariates. The outcome of GWR, for each covariate, was studied for all the 48 states in the contiguous US. As a direct consequence of spatial non-stationarity, it was possible to discuss the influence of each single covariate differently from state to state. In addition, the article includes a first attempt to determine if the observed variability in cost index values could be, at least partially explained by independent socio-economic variables.

CE Database subject headings

Estimation; Geostatistics; Construction costs; Planning; Budgeting

Introduction

In projects, success is achieved when, under a specified budget within a scheduled duration, predefined quality requirements are met (Duncan, 1996). In construction, cost estimating is the determination of quantities and the prediction of the costs required to construct and commission a facility within a defined scope. The accuracy of cost estimates in the early phases varies within an expected range that spans from -100% to +200% (Christensen and Dysert, 2011). However, these preliminary cost estimates are crucial to decision-making during early phases of a project with accuracy of preliminary cost estimation methods being a primary determinant to a project's success.

In practice, the most typical preliminary estimating procedure is the stochastic method. This method mainly utilizes adjustment factors to derive the cost estimating relationships among the independent variables. The use of these adjustment factors allows obtain a quick order-of-magnitude estimate of the facility cost by simply adjusting historically based estimates by time, project location and size (Remer et al., 1998). When adjusting for project location, various datasets provide location cost adjustment factors (LCAFs) for sample locations in North America, but, obviously, not all locations are included. To adjust estimates for un-sampled locations, the construction industry has adopted a simple, proximity-based interpolation method, the Nearest Neighbor (NN) method, which considers only the shortest linear-distance between two sites. Previous research has compared the accuracy of the NN method against other interpolation methods and found that the Conditional Nearest Neighbor (CNN) method outperforms the NN method (Martinez et al., 2010; Zhang, 2010). The CNN method revises the NN method by taking into account state jurisdictional boundaries. In the CNN method, the values of un-sampled locations are assumed equal to the LCAF value of the closest location within the same state as the project location. Spatial interpolation methods, such as the NN and CNN methods, utilize a set algorithm that estimates values for un-sampled locations by interpolation of sampled LCAF values based on historical construction costs in surrounding locations. However, sampled cost adjustment factor values can be affected by many other site factors, such as population, households, and income at the sampled location. Simple interpolation would ignore the effect of these values. On the contrary, spatial prediction methods use statistical analysis to examine or quantify the relationships among a set of independent variables (covariates) and the available LCAF values. The information obtained from that analysis can then be used to estimate the LCAF values for un-sampled locations. In this way, values are “predicted” out of a set of variables instead of being “interpolated” out of a set of sampled values.

The objective of this paper is to evaluate and assess the accuracy of various spatial prediction methods in estimating the value of LCAFs for un-sampled locations with respect to spatial interpolation methods. In addition, we attempt to determine if the observed variability in LCAFs could be partially explained by independent socio-economic variables.

The presence of local spatial variability is commonly ascribed in the geo-statistical literature to a pattern of factors. Besides the mere effect of spatial proximity (so that the local costs may be influenced by the general costs in a given area), the spatial patterns of the observed construction costs may be explained by a set of unknown (and sometimes unobserved) local covariates. The rationale is that, once the available variables are taken into account in the analysis, the performance of the regression-based methods should be equivalent or better than that of the existing proximity-based methods. Instead, if the regression methods are not able to capture the observed variability appropriately, we could be inclined to conclude that either mere proximity plays an important role in costs' determination or there are other variables (yet not considered in the analysis) that may be important to consider. In particular, this research investigated the relationship between a commonly used set of LCAFs, the Weighted Average City Cost Indexes (CCI) by RSMeans and the socio-economic variables included in the ESRI Community Sourcebook.

Literature Review

Location cost adjustment factors

Price variability of equipment, plants, or buildings has a temporal component depending on the competitive market conditions and the general state of inflation or deflation of a country's currency, and a geographical component within and between countries. These differences in price from one time to another and from one location to another can be measured by cost indexes (Humphreys, 2005), also called cost adjustment factors. In literature, the words “index” and “factor” are often used interchangeably. In cost engineering literature, a cost index is commonly defined as the ratio between the cost of a given facility at a given time and place, and the cost of a similar facility at a base time and place. Taking this definition into consideration, a cost index has to be intended as the factor that allows adjusting a base estimate to obtain a new estimate of a similar facility. Hence, cost indexes and adjustment factors are technically synonymous. In this paper, the term cost adjustment factor is used to indicate, in a general sense, every factor that “may be reasonably expected to influence project cost through the expected construction period” (Headquarters, Department of the Army - HQDA, 1994:1). According to the HQDA (1994), the budgetary cost at a specific location and under specific conditions can be estimated for repetitive type facilities by adjusting sample unit costs with appropriate factors that are considered relevant. In particular, five cost factors, derived from accurate data forms, are defined and constantly updated by HQDA (1994): (a) *Size adjustment factors*, (also called cost capacity factors or scale-up factors; Remer and Mattos, 2003), used to adjust the cost of similar projects of different sizes or capacities; (b) *Location adjustment factors* (also called simply location factors), used to adjust project costs for different geographical locations; (c) *Cost escalation factors* (also named inflation indexes, historical cost indexes or simply escalation factors), that track the change in costs for similar projects in different time periods, taking into consideration the inflationary factors that apply to construction costs; (d) *Technological updating adjustment factor*, accounting for logical advances in equipment and operational techniques and obsolescence issues; (e) *Design Contingency adjustment factors*, used to cover component items that cannot be analyzed or evaluated at the time the facility cost estimate is prepared, such as engineering and design progresses.

The geographical adjustment of a project's construction costs is commonly done through location factoring. The process of “factoring” implies that the cost of an item is estimated by multiplying the known cost of a related item by a calculated factor (Hollmann and Smith, 2007). According to Pietlock (2006), a LCAF is an instantaneous overall total project factor for translating the total cost of the project cost elements of a defined construction project scope of work from one geographic location to another. LCAFs recognize differences in productivity and costs for labor, engineered equipment, commodities, freight, duties, taxes, procurement, engineering, design, and project administration. Usually, the cost of land, scope/design differences for local conditions and codes are not included in a LCAF (Humphreys, 2005). LCAFs also do not reflect abnormal differences due to unique site consideration, such as historical preservation (HQDA, 1994).

The use of LCAFs is considered a valuable tool for the development of the earliest screening decisions in construction, in particular when the scope of the facility is highly uncertain. In fact, the use of LCAFs is only recommended during preliminary project evaluations. Further, LCAFs are not intended to be used when preparing appropriation-quality estimates (Pietlock, 2006). The utilization of LCAFs has the advantage of being easy, repeatable, and updateable in time. On the other hand, LCAFs are limited in accuracy since they generally reflect only the relative cost to exactly replicate a facility in another location without considering site-specific cost effects, such as climate and geological features (Humphreys, 2005). Thus, if the design is not identical, the cost differences cannot be taken into account through the application of location factoring. Moreover, most published LCAFs only include construction costs and not the engineering and project management costs incurred for the project (Russell, 2002).

Despite its potential weaknesses, estimation by adjustment factors is a very common approach for all types of construction. Due to the popularity of this approach, numerous cost adjustment factors have been developed and are now commercially available. Some examples of factors available in North America for the construction sector have been described in literature (McCabe et al., 2002; Remer and Mattos, 2003; Humphreys, 2005; OCS, 2001). In North America, the RSMeans CCIs are among the most commonly used LCAFs. RSMeans CCIs have been defined through a research on the building types most often constructed in the US and Canada. RSMeans utilizes a basket analysis approach that results in the determination of CCIs through a bottom up approach (Sullivan et al., 2010), where prices of individual basket elements are compiled into a composite model building. Nine different types of buildings were identified and combined to create a composite model used as reference building. The indexes for the composite model have been calculated through a combination that takes into consideration the specific quantities of 66 commonly used construction materials, specific labor hours for 21 building construction trades, and specific days of equipment rentals for 6 types of construction equipment used for the installation of the 66 materials by the 21 building trades. In addition, RSMeans computes the CCI 30-City National Average Index, which it defines on its website as “a quarterly 30-city market basket analysis of the most commonly used construction materials, labor trades and construction equipment.” The CCI for each city is then calculated as a percentage ratio of a specific city's cost to the National Average cost of the same item at a defined time period. The CCI is broken down by trade division numbers that refer to the different cost

components (e.g., “Contractor equipment”, “Site, infrastructure and demolition”, “Concrete”, etc.). For each city, a weighted average CCI is also reported. The average is weighted by giving more value to the more expensive components of a construction project and less influence to those trades that are usually the least expensive. To estimate the construction cost in a location A, knowing the construction cost of a similar facility in the location B, it is necessary to multiply the known cost for the ratio between the CCI for the location A and the CCI of the location B (RSMMeans, 2008) [1].

$$\text{City A Cost(unknown)} = \text{City B Cost(known)} \cdot \frac{\text{City A Index}}{\text{City B Index}} \quad [1]$$

Location cost adjustment factors are also used in early cost estimating outside North America. Building Cost Information Service (BCIS) location factors are commonly used in the United Kingdom. BCIS Location factors are regional pricing factors that indicate the general variability of pricing levels around UK. Suggesting a bottom-up approach, information from sampled price bills of quantities are compiled and statistically processed to determine BCIS values (Johnson, 2007). The default BCIS location factor is UK mean location (with a value of 100). Regional, County, and/or District Location Factor are used to adjust to the “UK Mean” level at the effective date and to make broad adjustments for differences in location (BCIS, 2002). The German BKI *Baukosten-Regionalfaktoren* (i.e., Construction costs regional factors) are developed by the *Baukosteninformationszentrum Deutscher Architektenkammern* (German chamber of architects). These 400 construction cost adjustment factors are available for all urban and rural districts in Germany to illustrate how high the average building construction cost differences are (Stoy et al. 2008). These factors are updated yearly in a book that also reports on country-level construction cost adjustment factors for all countries within the European Union (BKI, 2011).

Variables affecting construction costs

The construction cost of a facility is a unique combination of different localized variables (Johnson, 2007). Several studies have focused on the variables that may impact cost, schedule, and/or price of a construction project. At international level, the cost of construction is likely to vary enormously depending on a variety of factors, including the economic, political, and legal conditions of the particular country or the availability of resources, levels of skills, technology, degree of risk, infrastructure, financing, and communications (Jaggar et al., 2002). The physical characteristics of a particular site, its size, accessibility, and topography are also relevant (Johnson, 2007). Humphreys (2005) has identified a list of variables that can impact the cost of international construction projects. His list includes factors such as local material quality and availability, labor availability, equipment availability, labor productivity, import duties, import licenses, customs, local taxes, language, length of workweek, holidays, inflation, fluctuating exchange rates, religious customs, buy-local laws, shipping cycles, weather and climatic impacts, workforce level of education, communications, language, logistics, workforce housing, and training. Similarly, Walewski (2005) has found that estimate uncertainty is one of the risks that most extremely affect success of international projects with estimate uncertainty being majorly affected by location-specific factors.

One important point stated by the aforementioned studies is that geographical location has certainly a considerable influence on the likely financial implications of construction projects at macro (country-specific), meso (region-specific), and micro (site-specific) levels. At the macro level, lack of experience in a particular international jurisdiction is perceived as having a great impact on the occurrence of overestimates and underestimates (Walewski, 2005). At the meso level, remoteness and distance from major cities or supply centers can often aggravate the above indicated issues (Humphreys, 2005). This is particularly evident in the case of large conurbations (Johnson, 2007). Dursun and Stoy (2011) have demonstrated that the interaction between the cost and the location of a building at the country level is fundamental for the determination of construction duration in Bromilow's Time-Cost Model. At the micro level, geographical location has been acknowledged as a key cost driver by several authors, including Martin (2004) and Stoy et al. (2008).

Irrefutably, scholars and practitioners seem to have widely accepted that construction costs result from a combination of different localized variables and, therefore, they are extremely sensitive to geographical variations at different scales. Hence, the choice of spatial interpolation methods for the definition of unknown location cost adjustment factors appears justifiable. As the literature highlights, construction costs may vary enormously depending on a variety of factors related to political geography (such as the economic, political, and legal conditions of a state or region) more than the mere physical geography.

Macroeconomic aspects of the construction industry

Construction is a “principal component of the real sector of [any] national economy” (Olatunji, 2010:747), but how that role manifests itself may vary greatly from one nation to another (Ruddock, 2007). If the whole value chain of the industry is taken into consideration, the contribution of the construction industry to gross domestic product (GDP) in highly developed countries is approximately 20% (Ruddock, 2007). Moreover, construction represents a stimulus to macroeconomic growth through its contribution not only to GDP, but also to gross national income (GNI), fixed capital formation (FCF), and resource employment (RE) and by fostering other industries to produce (Hillebrandt, 2000, cited in Olatunji, 2010), as can be explained by the multiplier effect theory. For these reasons, investing in the construction industry represents an important public policy tool often used by central and local government to accelerate development and create employment, especially in recession periods (Wibowo, 2009). The multiplier effect suggests that the relation among the sectors within an economy is biunivocal. This means that a change in the national and local economy may produce a certain impact on the construction sector, and viceversa. The time and geographical effects are undoubtedly to be considered. First of all, the size of multiplier is negatively affected by leakages in the form of taxes and spending on imports. Secondly, the effect of the multiplier is not immediately visible because the additional spending brings idle resources into production and leads to additional real output rather than to increased prices (Gwartney et al., 2006).

In order to understand how the construction industry is affected by the general and local status of the economy, the use of socioeconomic indicators seems the most appropriate approach. Socioeconomic indicators provide a snapshot of the status of a national economy

at different scales. Since they represent the local differences in the economy of a nation, socioeconomic indicators can be used to understand the interaction between the construction sector and the economy at a local level.

Research Methodology

The objective of this research is to evaluate and assess various spatial prediction methods and compare their accuracy in estimating the value of LCAF for un-sampled locations against spatial interpolation methods. A subset of the 2008 City Cost Indexes (CCIs) dataset that included only U.S. cities was used. The 2008 CCIs dataset was obtained from the RSMMeans reference book. Prediction models relied on other socio-economic variables (e.g., population, households, number of families, races, income, and spending potential) that were acquired from the 2008 ESRI Community Sourcebook (ESRI 2008a).

A traditional correlation analysis and stepwise regression analysis were used to: (a) determine which ESRI covariates (attributes) are more associated with CCI values; (b) remove redundant attributes from the data; and thus (c) improve the performance of the prediction model. Pearson's correlation matrix and variance inflation factor were calculated to assess the multicollinearity of covariates. Once the relevant variables have been selected, two prediction models, a global, constant-coefficients regression model and a geographically weighted regression (GWR) model were used and compared against three spatial interpolation models: Nearest Neighbor (NN), Conditional Nearest Neighbor (CNN), and Inverse Distance Weighted (IDW). NN and CNN methods were described in the introduction of this paper. IDW interpolation method is one of the most frequently used deterministic models in spatial interpolation. Under IDW, values at unknown points are estimated using sampled values and distance to nearby known points. The farther away the point, the less weight the point has in helping define the value at an unsampled location.

In this study, the statistical analysis was based on a data set consisting of: city names, state names, ZIP codes, longitudes, latitudes, CCI values, and 65 ESRI covariates. The 2008 CCI dataset includes data for cities in U.S. and Canada. For this study, only data on 649 cities in all the 48 states in the contiguous US were initially extracted. The data set initially included 602 cities that were later reduced to 564, due to the missing values in response variables. These 564 cities corresponded to 2990 different ZIP codes. The 2008 ESRI Community Sourcebook contains 78 types of socio-economic variables such as population, housing, and income, based on ZIP codes. The data from Census 2000 are the main basis for ESRI's yearly projections, as described in the Update Methodology published every year by ESRI. For this study, it was chosen to focus only on 2008 projections. Hence, 13 out of 78 covariates were excluded from the analysis, since those 13 variables were data from the Census 2000 and ESRI Projections for 2013.

Data Analysis and Results

Correlation and Stepwise Regression Analysis

First, a Pearson's correlation test for each pair of covariate and CCI values was performed, and p-values were adjusted with Bonferroni correction. Variables with the highest

correlation coefficients and adjusted p-values less than 0.05 were selected. Then a Pearson's correlation matrix of the selected variables was calculated, and only one of two collinear variables was selected. As a result, 17 covariates were found to correlate with CCI values. Finally, stepwise regression analysis, with a combination of forward selection and backward elimination, was used to choose the final predictors. Eventually, 13 out of 17 covariates were selected. After a standard regression analysis was used to study the association between CCI values and these covariates, all 13 variables were found to be significant (p-value < 0.05). Significant here means that the null hypothesis of a zero regression coefficient for that variable would be rejected at the level of significance $\alpha=0.05$ in that location. The variance inflation factor (VIF) was also computed from the regression analysis to reveal the multicollinearity of the covariates. All VIF values of covariates were less than 5, indicating that each of the selected covariates was uncorrelated with the other covariates.

The 13 variables finally included in the analysis were:

- V1: Median Owner Occupied Housing Units Value 2008
- V2: Percentage of Asian Population 2008
- V3: Percentage of Two or More Races Population 2008
- V4: Household Income Base 2008
- V5: Population 2008
- V6: Percentage of Owner Occupied Housing Unit Value between \$175000 and \$400000 2008
- V7: Percentage of Other Population 2008
- V8: Median Household Income 2008
- V9: Percentage of Population 5-9 Years Old 2008
- V10: Growth Rate of Household (%) from 2000 to 2008
- V11: Population Growth Percentil
- V12: Percentage of the number of households having an income between \$25000 and \$50000 2008
- V13: Percentage of White Population 2008

Global Regression Analysis

In the standard regression analysis, the regression parameters are constant across space, hence we also call it global. The regression model can be written as [2]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad [2]$$

where Y is the response variable (dependent variable), $X_1, X_2, \dots, X_p$ are explanatory variables (covariates/predictors), $\beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients (parameters) to be estimated and assumed constant across space, and ε is the error term that is assumed to be

normally distributed. The regression analysis was carried out with the R package “stats”. R is a free software environment for statistical computing and graphics, and “stats” is one of the basic packages designed for regression analysis in R. The regression analysis provided evidence that all 13 variables are significant across the space, that is, each of these 13 attributes has a significant effect on the study areas. Here, the regression coefficients (the effect of the attributes) are assumed spatially constant. However, this is not necessarily the case. For example, population growth may be an important factor to explain a rising construction cost in an area with an already dense population, but it may be less relevant in an area where land is available and there are no construction impediments. In order to explore the complex relationships that vary over space, a spatially varying regression model was chosen.

Geographically-weighted Regression Analysis

Geographically weighted regression (GWR) provides an elegant way to explore the spatial variation in the model. Unlike global regression, it takes spatial non-stationarity explicitly into account in the regression model. The GWR model can be written as [3]:

$$Y(u, v) = \beta_0(u, v) + \beta_1(u, v)X_1 + \beta_2(u, v)X_2 + \dots + \beta_p(u, v)X_p + \varepsilon(u, v) \quad [3]$$

where (u,v) are the coordinates of the location, that is, the longitude and latitude in the data set. An estimate of the regression coefficients and a predicted value at the location (u,v) can be obtained by least squares fit. GWR analysis was implemented by using the R package “spgwr”.

In the GWR results, the response variables are the 2008 CCI values, whereas the explanatory variables are the selected 13 covariates. The significance of regression coefficients was assessed by computing the 95% confidence intervals of the coefficients

$$[\hat{\beta}_i(u, v) - 1.96 \times SE_{\hat{\beta}_i(u, v)}, \hat{\beta}_i(u, v) + 1.96 \times SE_{\hat{\beta}_i(u, v)}]$$

where $\hat{\beta}_i(u, v)$ is the estimate of coefficient $\beta_i(u, v)$ at location (u, v), $SE_{\hat{\beta}_i(u, v)}$ is the standard error for the coefficient estimate $\hat{\beta}_i(u, v)$, $i = 1, 2, \dots, 13$. If the confidence interval of the regression coefficient contains 0, the corresponding covariate has a non-significant effect on CCI values; if the confidence interval does not contain 0, the covariate has a significant effect on CCI values.

GWR Performance Measurement

In order to measure the performance of GWR and compare it with global regression and the selected interpolation methods, the residual was defined as the difference between the true value and the predicted value, that is: $\varepsilon = Y - \hat{Y}$. Where ε is the residual, Y is the actual value, and $\hat{Y}$ is the predicted value. Therefore, a positive residual (Figure 1) means the value is underestimated. On the contrary, a negative residual (Figure 1) represents an overestimation of the value. The mean, median, and standard deviation of the residuals in

absolute value for GWR analysis and global regression analysis were calculated and compared with the residuals of the three spatial interpolation methods mentioned above. The comparison results are showed in Table 1.

The maps in Figure 1 and Figure 2 show the distribution of the error estimates across the continental US. As indicated in Table 1, GWR provided a better prediction of the CCI values since the predictive errors were much smaller than those of global regression, and the three surface interpolation methods NN, CNN, and IDW. The prediction errors in global regression are greater than in GWR analysis, and the residuals are less spatially correlated in GWR analysis. Moreover, a detailed inspection of the nature of the error estimates shows that global regression was significantly affected by overestimates and underestimates with a prevalence of overestimates. On the other hand, GWR was more precise as shown by the large number of white dots in Figure 2. The distribution of the frequency of the GWR residuals in the ZIP codes included in the model follows a Gaussian curve, with the mean, median, and mode falling in the interval between -1 and 1, which has been considered as the closer interval to the exactness of the forecast (Figure 3).

Successively, the distribution of GWR Residuals was studied with respect to the population density taking into account the 250, 500, and 1,000 ZIP codes with the highest and lowest density, expressed in people/sq km. The graph in Figure 4 illustrates that the three samples of the ZIP codes with the highest population density generally follow the bell curve. On the contrary, the low density curves are characterized by a local maximum in the high overestimation interval ($\hat{\varepsilon} < -5$). A possible explanation of this occurrence is that the socio-economic variables chosen for the model are not representative of the actual situation of large land areas where few people live. While the CCI values are found through local quarterly-updated price surveys of specific materials and labor, the forecast CCI derive from the modelization of socio-economic covariates. This may mean that when the population is more disperse in an area, or when the population is scarce, the socio economic variables are less capable of describing the fluctuation of construction costs, thus the model is less accurate. Further studies may help to refine the model taking into account the peculiarity of the areas with low population density.

Discussion

In this section, the most relevant results obtained from the GWR analysis are presented and discussed. Data analysis has been conducted at the state level, taking into account the significance of each variable in the 48 states included in the model. Since it is unlikely that every covariate has a homogeneous impact on the dependent variable (i.e., CCI) in the entire state, an evaluation of the significance in term of predominance has been performed. Therefore, a covariate has been considered significant (S) in a state if it is significant in more than 66% of the ZIP Codes across the state. Likewise, a covariate has been considered Non Significant (N/S) if the significance occurs in less than 33% of the ZIP Codes across the state. If the significance falls in the interval between 33% and 66%, the state has been marked with the symbol (!).

Later, states have been divided into three groups with respect to the number of significant ESRI covariates in the regression model. The results of this analysis for each state are displayed in the map in Figure 5. In the map, the first group contains the states with high relevance of the variables (9-13 significant variables out of 13); the second group includes the states characterized by medium relevance (5-8 out of 13), while the last group contains the states with low relevance (0-4 significant variables). The map in Figure 5 shows that only 8 states were marked by a high number of significant covariates (ME, VT, NH, MA, RI, NJ, ND, and UT), while in 23 out of 48 states the significance of the covariates was low. The map does not show a distinctive geographical pattern in terms of significance of the variables. In northeastern states, a large number of the 13 covariates are in some way correlated to the CCI. At the same time, all the Pacific states (WA, OR, and CA), along with most of the southern states, show a very low significance of the covariates. This occurrence does not necessarily mean that the covariates selected for this study are not capable of explaining the relation between construction costs and socio-demographic indicators. On the contrary, this analysis may suggest predictors for further regression analysis at the state (or region) level, together with other socio-economic and demographic variables that may be available in local records but are not included in the ESRI sourcebook.

In Table 2, the sign of the parameters referring to the covariates is studied. The positivity of the parameter means that an increase in the value of the variable to which the parameter refers is correlated to an increase of the dependent variable (CCI), and vice versa. On the contrary, if the sign is negative, an increase of the former will induce a decrease of the latter, and vice versa. The sign of the correlation coefficient was studied taking into account not only the significant (S) cases, but also the uncertain cases (!), in which the covariate was deemed significant between 33% and 66% of the ZIP Codes across the state.

For instance, the Median Owner Occupied Housing Units Value 2008 (V1) is significant and positive in Connecticut (CT). This means that an increase in the CCI (and thus, the construction cost) is correlated with an increase in the Median Owner Occupied Housing Units Value with respect to the year 2000. On the other hand, the Median Household Income 2008 (V8) is significant and negatively correlated with the CCI: this means that an increase in the CCI occurs when a decrease of the Median Household Income happens from 2000 to 2008, or vice versa.

Data analysis illustrates that a negative correlation occurs in 55% of the cases. Emblematic are the outputs for the covariates Percentage of Population 5-9 Years Old 2008 (V9), Growth Rate of Household (%) from 2000 to 2008 (V10), and Percentage of White Population 2008 (V13), where the negative correlation occurs in all the states where they were deemed significant (Table 2). On the contrary, the positive correlation is particularly evident in the following covariates: Median Owner Occupied Housing Units Value 2008 (V1), Percentage of Two or More Races Population 2008 (V3), and Household Income Base 2008 (V4), where the positivity is above 90%.

In a following phase, the analysis has been focused on the covariates, in order to highlight the ones with the highest percentage of significance and non significance across the states. The results are summarized in Table 3, that reports the ranking of the covariates in terms of

significance (first column). The results show that only 1 out of 13 variables is significant in more than 50% of the states (V10 - Growth Rate of Household from 2000 to 2008), and half of them are significant in less than 40% of the states. As we could expect from the supply-demand rule behind the construction market, 4 of the 6 most significant variables are related to Family and Households, while the other 2 are related to Housing and Income. On the other hand, 4 variables report a percentage of non-significance above 50%. Such occurrence does not necessarily mean that the covariates are not appropriate to define the GWR model, and therefore the correlation between socio-economic descriptors and construction costs. This is the essence of spatial heterogeneity: the structure of the model changes from place to place across the study area as the parameter estimates change in relation to each other in the model (Charlton and Fotheringham, 2009; Brunsdon et al., 1998).

Figures 6 through 8 show the coefficient distribution for 3 selected covariates that have shown the most interesting features in terms of significance effect of the covariates. Figure 6 shows the effect of Household Growth Rate from 2000 to 2008 on CCI. As the map displays, this covariate is not significant in none of the ZIP codes of the western states (WA, OR, ID, CA, NV, MT, and CO), and in OH, and it is scarcely significant (only few ZIP codes) in part of the Great Lakes Area (IN, MN, and WI), almost in all the South-western states (AZ, UT, NM, and TX), and in FL. In the areas where the covariate appears significant, a negative correlation is reported. Upstate NY is an area where this negative correlation is particularly evident. Moreover, the magnitude of the regression coefficient suggests that the influence of the covariate in that area is somehow stronger than in other areas. Thus, it is worth trying to infer some potential reasons, without pretending to fully explain the event, since it would require an additional effort that is beyond the scope of this work.

$$HH \text{ growth rate} = \left[\left(\frac{HH_{08}}{HH_{00}} \right)^{1/8.25} - 1 \right] \cdot 100 \quad [4]$$

ESRI household growth rate 2008 is calculated as an annual compound rate of change from 2000 to 2008, represented by the ratio between the number of households in the reference year 2000 (HH_{00}) (Census data) and the forecast for the current year (2008) (HH_{08}), as explained in Eq. [4]. According to the State of the Nation's Housing 2010 report (JCHSHU, 2010), the US household growth from 2005 to 2009 fell from about 1.2–1.4 to less than 1.0 million per year. Most probably, in 2008 the household growth rate decreased even in upstate NY, with respect to 2000 data. Regarding the dependent variable (the CCI), it is acknowledged that New York City (NYC) is one of the most expensive construction markets in the US. In 2008, first year of the recession in the US, construction costs in NYC soared at an annual rate of around 12%, falling only in 2009 due to the combination between the national recession and a steep drop in residential and commercial construction (NYBC, 2011). For all the reasons cited above, it is possible to deduce that the correlation in this particular area is defined by a fall of the household growth and a spike in the construction costs. Anyway, this does not automatically mean that the two variables are causally related: correlations can also reflect non-causal relationships. As a matter of fact, co-variation reflects a causal relationship only if it is coherent in terms of time order (the causal factor

must come before the supposed effect), effectiveness of change (the dependent variable must be capable of being changed), and theoretical plausibility (a causal relationship can actually be inferred) (De Vaus, 2001).

The map depicted in Figure 7 presents the regression coefficients for the covariate Median Owner Occupied Housing Units Value 2008. The effect of the covariate is significant in 24 states, and highly significant in 17 states, where the significance rate is 100% in all the ZIP codes included in the model. The states where the effect is not significant are in the West (WA, OR, CA, CO, and AZ), in the South East (AR, FL, MS, KY, LA, OK, and TN), and OH, MN, and PA. The sign of the coefficient is mainly positive, with higher magnitude in the Great Lakes region. In this specific case, the explanation of the positivity appears less fuzzy, since that the housing values (and therefore, the housing market) and the construction costs are directly correlated, in ordinary contexts. It is anticipated that, if the analysis had been performed with the 2009 dataset, the output would have been slightly different due to the housing bust that affected the country in 2008. The 2008 dataset, in fact, refers to projections made in 2007, while the recession officially started in December 2007 (NBER, 2012).

The map in Figure 8 shows the effect of the Percentage of the population between 5-9 years old on CCI. Even in this case, the covariate is not significant in numerous areas, such as almost all the Southeastern states, all the South-western states, part of the western states (WA, OR, CA, and CO) and part of the Midwest (IN, KS, MN, MO, and OH). Where the effect is significant, the sign of the coefficient is mainly negative. The map highlights a strong negative correlation in metropolitan areas around important cities, such as Boston, Chicago, and Washington DC. The correlation can be explained in terms of a contraction of young population in areas where the construction market is characterized by a different consumer's profile (more professionals than families) and more focused on high-quality housing. Similarly, other potential explanations include that families with kids are induced to leave metropolitan areas because of lack and/or high cost of large size housing units, or also that families with kids are induced to leave metropolitan areas to find suburban neighborhoods that are more kid-friendly. Whereas this phenomenon is well known, these “migrant” families are usually absorbed within a city limits. However, cost of construction in large conurbations, such as Boston, Chicago, and Washington DC, seems to be forcing families with kids away into more rural areas.

Finally, an analysis conducted on the significance of the covariates at the ZIP codes level shows that in some of the most important cities of Texas, Florida, Minnesota, Colorado, Oregon, and Washington, none of the covariates in the study has a significant effect on construction cost. This happens, for instance, in Houston and San Antonio (TX), Miami and Orlando (FL), Minneapolis (MN), Cincinnati (OH), Denver (CO), Portland (OR), and Seattle (WA). Whereas ESRI variables were not significant in these cases, probably, some other factors not considered in the study could affect the CCI values in those areas.

Conclusions and Future Work

The use of adjustment factors to obtain preliminary cost estimates has been widely adopted although the literature has also highlighted the limits of this estimating methodology. Due to its easiness and to the availability of commercial and governmental datasets, estimating procedures through adjustment factors have progressively become common practice. The main problem in the use of location cost adjustment factors is that not every location is sampled. Therefore, it is necessary to estimate them for the locations where they are not readily available. A common approach is to use interpolation. Industry practitioners have largely relied on a simple interpolation method, which considers only the shortest linear-distance between two sites. However, construction costs may be influenced by numerous factors other than the geographical location, as suggested by macroeconomic theories.

The contribution to the body of knowledge of this study is an improved understanding of the role of local socio-economic covariates to explain geographical variability in construction costs. Two spatial prediction models were created to include the socio-economic variables in the prediction of the CCI values: a Global Regression Model, in which the regression parameters are constant across space, and a Geographically Weighted Regression Model (GWR), that considers the spatial non-stationarity. The two regression-based prediction models were compared with three surface interpolation methods. The results reveal that GWR outperformed the other models considered in the research. The data analysis, conducted for each covariate at a state level, illustrates that the effect of the covariates is not significant in almost 50% of the states. In most cases, the state territory is not interested by a uniform pattern, showing areas where each variable is significant, and areas where it is not. In some cities, such as Houston, Orlando, Minneapolis, or Seattle, none of the variables is significant. Presumably, the fact that these cities are cores of important metropolitan statistical areas (US Census Bureau, 2011) may be a reasonable explanation of this peculiarity.

Causality and co-variability are not synonymous. The construction of a plausible causal model that could link the CCI and the socio-economic figures was beyond the scope of the present work. Hence, this article focuses on the inference of possible causal relations, which will be investigated in the future by taking into account more detailed geographical features other than the state boundaries. The division into metropolitan and micropolitan statistical areas may in fact clarify why, in different areas within a state, the construction costs are not affected by the same variables. Moreover, a thorough study on the causal relationships between the CCI and the socio-economic variables may help the industry to obtain more comprehensive and realistic early cost estimations.

This paper also contributes to practice by providing theoretical ground that would justify the creation of practical tools to support cost engineers. In an industry where computing power is used to support a large variety of decisions, the authors have identified a knowledge gap that has limited the creation of statistically-based software applications in support to cost engineers. Using the results of this study, various commercial entities (e.g., RS Means) could enhance their online tools by uploading publicly available socio-economic variables and allowing users to perform geostatistical analysis. As a result, a cost engineer could input

the location of a project and obtain the most accurate location adjustment factor through a mix of interpolation and geostatistical prediction techniques.

References

- BCIS. BCIS Online User Guide - September 2002. 2002. Available on www.library.ulster.ac.uk/electronic/guides/bcis.pdf. Retrieved 9/5/2011
- BKI Baukosteninformationszentrum Deutscher Architektenkammern. BKI Baukosten-Regionalfaktoren 2011 für Deutschland und Europa. Stuttgart, Germany: 2011.
- Brunsdon, Chris; Fotheringham, AStewart; Charlton, MartinE. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1998; 47(3):431–443.
- Charlton, MartinE; Fotheringham, AStewart. Geographically Weighted Regression - White Paper. 2009. Available on http://ncg.nuim.ie/ncg/GWR/GWR_WhitePaper.pdf. Retrieved 01/06/2012
- Christensen, Peter; Dysert, LarryR. Cost Estimate Classification System AACE International Recommended Practice No 17R-97. AACE Inc; 2011.
- De Vaus, DavidA. *Research Design in Social Sciences*. SAGE Publications Ltd; London, UK: 2001.
- Duncan, WilliamR. *A Guide to the Project Management Body of Knowledge*. Project Management Institute Publishing Division; Sylva, NC: 1996.
- Dursun, Onur; Stoy, Christian. Time-cost relationship of building projects: Statistical adequacy of categorization with respect to project location. *Construction Management and Economics*. 2011; 29(1):97–106.
- ESRI. *Community Sourcebook America*. 2008. ESRI; Vienna, VA: 2008a.
- ESRI. [Accessed 11/2/2011] ESRI Demographic Update Methodology: 2008/2013. 2008b. Available on www.esri.com/library/whitepapers/pdfs/demographic-update-methodology.pdf
- Gwartney, JamesD; Stroup, RichardL; Sobel, RussellS; Macpherson, DavidA. *Economics: private & public choice*. Eleventh. Thomson South-Western; Mason, OH: 2006.
- Headquarters, Department of the Army – HQDA. TM 5-800-4, Programming Cost Estimates for Military Construction. Washington, DC: 1994.
- Hillebrandt, PatriciaM. *Economic Theory and the Construction Industry*. Macmillan; London, UK: 2000.
- Hollmann, JohnK; Smith, Peter. ICEC International Location Factor Cost Model. Proceedings of the 11th PAQS Congress; Auckland, New Zealand. 2007.
- Humphreys, KennethK. *Project and Cost Engineers' Handbook*. Fourth. Marcel Dekker; New York: 2005. Cost Indexes, Escalation, and Location Factors.
- Jaggar, David; Ross, Andy; Smith, Jim; Love, Peter. *Building Design Cost Management*. Blackwell Science Ltd; Oxford: 2002. Design Cost Management: the feasibility stage; p. 62-75.
- Johnson, VB. *Laxton's Building Price Book - Major & Small Works*. 179th. Melbourne: Butterworth-Heinemann; 2007. Regional variations; p. XXIX
- Joint Center for Housing Studies of Harvard University - JCHSHU. *The State of the Nation's Housing 2010*. 2010. Available on www.jchs.harvard.edu/research/publications/state-nations-housing-2010. Retrieved 11/15/2011
- Martin, Joe. Database for elemental cost planning. AACE International Transactions - 48th AACE International Annual Meeting. 2004
- Martinez, AdamA; Migliaccio, GiovanniC; Zandbergen, Paul; Zhang, Su. Assessment of Methods for Adjusting Construction Cost Estimates by Geographical Location. AACE International Transactions - 53th AACE International Annual Meeting; Seattle, WA. 2009.
- McCabe, BY.; O'Grady, J.; Walker, F. *A Study of Construction Cost Sources*. Proceedings of the Annual Conference of the Canadian Society for Civil Engineering; Montréal, Canada. 2002.
- National Bureau of Economic Research - NBER. *US Business Cycle Expansions and Contractions*. 2012. Available on www.nber.org/cycles.html. Retrieved 7/12/2012

- New York Building Congress - NYBC. New York City Construction Costs on the Rise. New York City Construction Outlook Update. 2011. Available on www.buildingcongress.com/outlook/032511.html. Retrieved 11/15/2011
- Olatunji OA. The impact of oil price regimes on construction cost in Nigeria. *Construction Management and Economics*. 2010; 28(7):747–759.
- McCabe, Brenda; O'Grady, John. Ontario Construction Secretariat – OCS. A Guide to Construction Cost Sources. 2001. prepared by Available on www.ogradys.on.ca/Downloads/Construction%20Industry.html. Retrieved 09/01/2011
- Pietlock, BernardA. Developing Location Factors By Factoring - As Applied In Architecture, Engineering, Procurement, and Construction. AACE International Recommended Practice No 28R- 03. 2006
- Remer, DonaldS; Mattos, FernandoB. Cost and scale-up factors, international inflation indexes and location factors. *International Journal of Production Economics*. 2003; 84(1):1–16.
- Remer, DonaldS; Huynh, Leon; Agarwal, Kanishka; Auchard, BrianJ; Heaps-Nelson, Tom. A compilation of inflation and location indexes. *International Journal of Production Economics*. 1998; 54(1):41–55.
- RSMeans. Building Construction Cost Data – 66th Annual Edition. RSMeans Construction Publishers & Consultants; Kingston, MA: 2008.
- Ruddock, Les. The Economic Value of Construction: Achieving a Better Understanding. Proceedings of COBRA 2007; Atlanta, GA. 2007.
- Russell, RodneyL. Estimating using international factors. AACE International Transactions -46th AACE International Annual Meeting. 2002
- Sullivan, WilliamG; Wicks, ElinM; Koelling, CPatrick. *Engineering Economy*. 15th. Prentice Hall; New York: 2010. Cost-Estimation Techniques.
- Stoy, Christian; Pollalis, Spiro; Schalcher, Hans-Rudolf. Drivers for cost estimating in early design: Case study of residential construction. *Journal of Construction Engineering and Management*. 2008; 134(1):32–39.
- US Census Bureau. 2010 Census Demographic Profile Summary File. 2011. Available on www.census.gov/prod/cen2010/doc/dpsf.pdf. Retrieved 11/2/2011
- Walewski, John. Unpublished Doctoral Dissertation. University of Texas; 2005. International Project Risk Assessment.
- Wibowo, Agung. Discussion Paper. Construction Management, Civil Engineering, Department, Diponegoro University; Indonesia: 2009. The Contribution of the Construction Industry to the Economy of Indonesia: A Systemic Approach.
- Zhang, Su. Unpublished Master Thesis. University of New Mexico; 2010. Validation of Geographically Based Surface Interpolation Methods for Adjusting Construction Cost Estimates by Project Location.

Figure 1. Global Regression Analysis – Error

Figure 2. Geographically Weighted Regression Analysis – Errors

Figure 3. Frequency distribution of GWR residuals for all the ZIP Codes

Figure 4. Comparison of frequency distributions of GWR errors

Figure 5. Analysis of states in terms of significance of the covariates

Figure 6. GWR Analysis – Coefficients Distribution of Household Growth Rate from 2000 to 2008

Figure 7. GWR Analysis – Coefficients Distribution of Median Owner Occupied Housing Units Value 2008

Figure 8. GWR Analysis – Coefficients Distribution of Percentage of Population between 5-9 years old 2008

Table 1

Comparison Summary

Descriptive Statistics	Surface Interpolation Method (adapted from Zhang 2010:84)			Regression-Based Prediction Model		
	NN	CNN	IDW	GWR	Global Regression	
Mean	3.55	2.89	2.94	2.22	7.49	
Median	2.40	1.90	2.20	1.44	6.65	
Std. Dev.	3.81	2.93	2.80	2.38	5.14	

Note: Statistics for the three surface interpolation methods were based on city-level analyses whereas statistics for regression-based prediction models were computed at ZIP-code level.

Table 2
Analysis of the positivity of the correlation for significant covariates across states

	COVARIATES												
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
AL	S (+)	! (+)					! (+)	! (+)		S (-)			S (-)
AR		S (+)	! (-)		S (-)	S (+)		S (+)	S (-)	S (-)	S (-)		
AZ													
CA						! (-)	! (-)						! (-)
CO													
CT	S (+)			S (+)	S (+)			S (-)	S (-)	S (-)	S (+)	S (-)	! (-)
DC						S (-)			S (-)	S (-)			S (-)
DE	S (+)	S (+)		! (+)	! (+)			S (-)		S (-)		S (-)	S (-)
FL											! (-)		
GA	S (+)	! (+)					! (+)		S (-)	S (-)			S (-)
IA	S (+)	! (+)	S (+)	S (+)		S (+)		! (+)	S (-)	S (-)	S (-)		S (-)
ID	S (+)	S (+)	S (+)		S (+)	S (-)	! (-)		S (-)		S (-)		! (-)
IL	S (+)	S (+)		S (+)	S (+)	S (+)		S (-)	S (-)		!		S (-)
IN	S (+)	S (+)		S (+)	S (+)	! (+)	S (+)	! (-)			S (-)		S (-)
KS	! (+)		S (-)		S (-)	S (+)	! (-)			S (-)	S (-)		S (-)
KY		S (+)	! (+)				! (-)	S (+)		! (-)	S (-)	! (-)	
LA		! (+)								S (-)			! (-)
MA	S (+)	! (+)	S (+)	S (+)	! (+)	S (+)		S (-)	S (-)	S (-)	S (+)	S (-)	S (-)
MD	S (-)		S (-)			! (+)			S (-)	S (-)			S (-)
ME	S (+)	S (+)	S (+)	S (+)	S (+)	! (-)		S (-)	S (-)	S (-)	S (-)	S (-)	S (-)
MI	! (+)		! (-)	! (+)	S (+)	! (+)		! (-)	S (-)	! (-)	S (-)	S (-)	S (-)
MN													
MO	! (+)	! (+)			! (+)	! (+)	! (-)	S (+)		! (-)	S (-)	! (+)	S (-)
MS		! (+)						S (+)	! (-)	S (-)	! (-)		
MT	S (+)		S (+)		S (+)	S (+)	! (-)		S (-)				
NC	! (+)	! (+)	! (+)		S (-)		! (-)	! (-)		! (-)			

	COVARIATES												
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
ND	S(+)	S(+)	! (+)		S(+)	S(+)	S(-)	S(+)	S(-)	S(-)	S(-)	! (-)	S(-)
NE	S(+)					! (+)		! (+)		S(-)	S(-)		S(-)
NH	S(+)	S(+)	S(+)	S(+)		S(-)		S(-)	S(-)	S(-)	S(+)	S(-)	S(-)
NJ	S(-)	! (-)		S(+)	S(+)	S(-)		S(-)	S(-)	S(-)	S(+)	S(-)	
NM	S(+)	S(+)	S(+)	! (-)		S(+)	S(-)				S(-)	S(+)	
NV	S(+)	S(+)	S(+)			S(-)			S(-)		S(-)		S(-)
NY	! (+)	! (-)		S(+)	S(+)	S(-)	! (+)	S(-)	S(-)	S(-)	S(-)	! (-)	! (-)
OH											! (-)	! (-)	! (-)
OK		S(+)	! (-)			! (+)					! (-)		
OR													
PA		! (+)	! (-)	! (+)				! (-)		S(-)		! (-)	! (-)
RI	S(+)		S(+)	S(+)	S(+)	S(-)		S(-)	S(-)	S(-)	S(+)	S(-)	S(-)
SC	! (+)					! (-)				! (-)			
SD	S(+)	S(+)	! (-)			S(+)	! (-)	S(+)	S(-)	S(-)	S(-)		S(-)
TN		S(+)	S(+)	! (+)		! (-)		S(+)	! (-)	S(-)	! (-)	! (-)	
TX		! (+)											! (-)
UT	S(+)	S(+)	S(+)		S(+)	S(+)			S(-)		S(-)	S(+)	S(-)
VA	! (+)		! (+)			! (-)	! (-)	! (-)		S(-)		S(-)	! (-)
VT	S(+)	S(+)	S(+)	S(+)	S(+)	! (-)	! (-)	S(-)	S(-)	S(-)	S(+)	S(-)	S(-)
WA													
WI	S(+)	! (+)	! (+)	S(+)	S(+)	S(+)		S(-)	S(-)			! (-)	S(-)
WV	! (-)	S(+)	S(-)			S(-)	S(-)	S(+)		S(-)	S(-)	S(-)	
WY	S(+)	S(+)	! (-)					S(+)	S(-)		! (-)	! (+)	

Legend:
S (+): The significance is positive
! (+): The significance, where applicable, is positive
S (-): The significance is negative
! (-): The significance, where applicable, is negative

Table 3

Analysis and ranking of the significance of the covariates across states

Variables	S	N/S	!	S [%]	N/S [%]	! [%]
V 10) Growth Rate of Household (%) from 2000 to 2008	25	19	5	51,0%	38,8%	10,2%
V 1) Median Owner Occupied Housing Units Value 2008	24	17	8	49,0%	34,7%	16,3%
V 9) Percentage of Population 5-9 Years Old 2008	23	24	2	46,9%	49,0%	4,1%
V 11) Population Growth Percentile	23	19	7	46,9%	38,8%	14,3%
V 13) Percentage of White Population 2008	22	18	9	44,9%	36,7%	18,4%
V 8) Median Household Income 2008	20	21	8	40,8%	42,9%	16,3%
V 2) Percentage of Asian Population 2008	19	17	13	38,8%	34,7%	26,5%
V 6) Percentage of Owner Occupied Housing Unit Value between \$175000 and \$400000 2008	19	18	12	38,8%	36,7%	24,5%
V 5) Population 2008	17	29	3	34,7%	59,2%	6,1%
V 3) Percentage of Two or More Races Population 2008	15	23	11	30,6%	46,9%	22,4%
V 12) Percentage of the number of households having an income between \$25000 and \$50000 2008	13	27	9	26,5%	55,1%	18,4%
V 4) Household Income Base 2008	12	32	5	24,5%	65,3%	10,2%
V 7) Percentage of Other Population 2008	4	32	13	8,2%	65,3%	26,5%

Legend:

S: Significant

!: Marginally Significant

N/S: Non Significant