Empirical Comparison of Methods for Estimating Location Cost Adjustments Factors

G.C. Migliaccio¹, Paul Zandbergen², and A.A. Martinez³

Abstract

Location factors are used to adjust conceptual cost estimates by project location. Presently, the construction industry has adopted a simple, proximity-based interpolation method to estimate location factors for missing locations. Although this approach is widely accepted, its validity has not been statistically substantiated. This study assessed the current method of adjusting conceptual cost estimates by project location and compared its performance against two alternative spatial interpolation methods. A Moran’s I test was used to confirm the presence of strong spatial autocorrelation, which supports the use of proximity-based methods. Additional statistical evaluations of current and alternative methods were also conducted. Results provided statistical justification for the current method. However, an alternative method was proven to outperform the current method. Moreover, several opportunities for future research were identified as a result of this exploratory study.

CE Database subject headings

Estimation; Geographic information systems; Construction costs; Planning

¹ Assistant Professor, University of Washington, Department of Construction Management, Box 351610, Seattle, WA 98195-1610, email: gianciro@uw.edu, phone: +1 (206) 685-1676, fax: +1 (206) 685-1976
² Associate Professor, University of New Mexico, Department of Geography, MSC01 1110, Albuquerque, NM 87131-0001. Email: zandberg@unm.edu, phone: +1 (505) 277-3105, fax: +1 (505) 277-3614
³ Project Manager, Intel Corporation, Rio Rancho, NM, email: adam.a.martinez@intel.com
INTRODUCTION

Accuracy of cost estimates is fundamental to the success of construction projects. Cost estimates are developed regularly throughout the project lifecycle. In preliminary stages, such as planning and conceptual design, estimates help owners determine general financial feasibility for budgeting and programming purposes and form the basis of project scoping.

As there are various project stages in which cost estimates are produced, a recommended practice for their classification was established by the Association for the Advancement of Cost Engineering International (AACEI) (Christensen & Dysert 2003). AACEI identified five cost estimate classes throughout the entire lifecycle of a project using project scope definition as the most significant factor to categorize each cost estimate class together with its end usage and expected accuracy range. Lower accuracy of early estimates is mostly due to undefined project scope, which cannot be avoided because the owner usually has only a rough understanding of the project scope at the conceptual level, and this understanding evolves throughout design. This time-specific asymmetry of information affects the magnitude of early inaccuracies because an estimator cannot accurately account for changes and consequent risks in an idea that has not been fully formulated.

Estimate accuracy is also limited by the adjustment methods used to develop the estimate. While there are multiple adjustments that are considered in conceptual cost estimation (i.e., time, location, size, complexity) (Gould 2002), this study aims at assessing and improving current practice for adjusting cost estimates by geographic location. While study results may be applied to all classes of cost estimates, the authors expect that they are more beneficial to owner-developed early-stage cost estimates.

Location cost adjustment factors (LCAFs) are used to distinguish geographical locations of interest, usually represented as cities, and are used to predict cost implications. Although various organizations have sampled data and developed LCAF datasets for many cities throughout the United States, not all cities are included in these datasets. Assuming that these datasets accurately describe construction costs for specific categories of projects, the main issue is that these organizations sample information only from a limited number of locations. For instance, the RS Means dataset includes values for only about 650 out of the more than 30,000 cities within the continental U.S.

With this in mind, the following primary research question for this study was established: How should a cost estimate be adjusted for a location that is not included in a LCAF dataset? This question has already been raised by practitioners, and a response has been provided by at least one of the organizations developing LCAF datasets. According to Waier (2006), “For a city not listed [in the RS Means dataset], use the factor for a nearby city with similar economic characteristics” (p. 586). However, this suggestion is somewhat vague and left up to interpretation. The suggested relevancy of a “nearby city” indicates that proximity should be a key element in performing location factor interpolation. Whereas identifying a
nearby city with “similar economic characteristics” was also suggested, guidance on what characteristics may affect construction costs is lacking. Similarly, scholars have suggested a different hybrid approach utilizing a combination of proximity selection and targeted local data collections (Johannes et al 1985). This approach, however, requires the identification of local input prices for components “that are believed to differ considerably from the respective input prices in the closest city available [in the dataset]” (pp.77).

To date, both practitioners and scholars have adopted a simple solution to this problem by using a simple, proximity-based interpolation method that adopts the LCAF value of the closest sampled location to estimate the factor for an unsampled location. For instance, RS Means at the question “My city or town is not listed in the CCI. What do I do?” provides the following answer:

Use the factors for the nearest city that you can find. If a city or town is too small to have its own union hall workers, who wish to belong to the union, will travel to the nearest larger city that does have a union hall (RS Means website).

In this article, this method is described as “nearest neighbor” (NN) interpolation and was considered the “current” method. Although this method is commonly used, its validity has not been statistically substantiated. Assessing this method is the main contribution of this article that aims at improving accuracy of conceptual cost estimates. The article also provides initial answers to the primary research question together with the following detailed research questions:

- Can statistical analysis provide justification for the current NN method?
- What are possible alternatives to the current NN method that may potentially increase accuracy of location adjustments?
- Can these alternate methods be statistically proven to produce a more accurate estimate?

BACKGROUND

Use of Location Factors in Preliminary Estimating

Historical cost data are commonly used to prepare preliminary estimates. However, certain variables are known to affect construction cost, including quantity, geographic location, and project type and characteristics (Ashur and Crockett 1997). To take them into account, adjustments are made for project specific characteristics such as project date, size, location, and complexity. Project costs can be adjusted based on the unit area or volume of a building, or occupancy units (e.g., number of hospital beds). Popescu et al. (2003) described a common procedure of applying cost estimate adjustments:

- Determine the usable area of the building, volume, or number of occupant units.
- Select from the most recently published standards for the type of building that most closely matches the project, the unit area, unit volume, or occupancy unit standard cost.
- Adjust […] costs to a projects location using regional adjustment factors (p. 59).
Included in this procedure is the location adjustment, which is the focus of this research. Location adjustments are performed using LCAFs. Pietlock (2006) describes location factors as follows:

A location factor is an instantaneous (i.e., current—has no escalation or currency exchange projection), 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. This factor recognizes differences in productivity and costs for labor, engineered equipment, commodities, freight, duties, taxes, procurement, engineering, design, and project administration. The cost of land, scope/design differences for local conditions and codes, and differences in operating philosophies are not included in a location factor (p. 1).

Location factors are primarily used by owners in developing estimates based on a not completely defined project scope and are not intended to be used for higher quality estimates that are developed by contractors using data based on actual construction material costs and productivity rates. Despite their widespread use, using location factors when creating a conceptual cost estimate offers several challenges:

- Published cost standards seldom represent 100% of the project under consideration.
- The location factor of adjusting a city or community is not accounted for in the published standard.
- The time factor involved in extrapolating future construction cost variations may differ (Popescu et al. 2003; p. 59).

RS Means city cost indexes (CCI) and U.S. Department of Defense (DoD) area cost factors (ACF) are two annually published LCAF datasets. RS Means values are primarily used for commercial and light industrial construction projects whereas ACF values are primarily used for military projects. The United States of America Department of Defense created a unified facilities criteria design guide for location adjustment using location factors. This design guide refers to the ACF indexes:

The ACF index is used in adjusting estimated costs to a specific geographical area. The factors reflect the average surveyed difference for each location in direct costs between that location and the national average location (U.S. DoD 2008; pp. C-6).

In addition, the Air Force Civil Engineering Support Agency (2005) describes ACF in the Historical Air Force Construction Cost Handbook:

Location Factors or Area Cost Factors (ACF) are used by all DoD services to adjust average historical facility cost to a specific project location. This allows increased accuracy in identifying project costs during initial project submissions or when specific design information is not available. The area cost factor index takes into consideration the cost of construction material, labor and equipment, and other factors such as weather, climate, seismic conditions, mobilization, overhead and profit, labor availability, and labor productivity for each area (p. 73).

Factors such as weather, climate, and labor productivity are clearly reflected in the ACF index. On the other hand, the RS Means CCI dataset, which was used in this study, does not reflect productivity, and it is not clear if it reflects the effect of weather and climate:
the CCI [dataset] does not take into consideration factors such as: managerial efficiency, competitive conditions, automation, restrictive union practices, unique local requirements and regional variations due to specific building codes. Moreover, it does not include the productivity variations between trades or cities (Waier 2006; pp.644).

Seminal Work on Development of Location Factors

Besides being recurrently mentioned in handbooks and used in practice, little research has been conducted on location adjustments. Johannes et al. (1985) introduced the concept of an “area cost factor” as an input decision for construction expansion. This cost factor can be described by the construction cost in new areas relative to the cost in another area. The primary purpose of Johannes was to explore how the economic theory of cost functions can be used to construct theoretically sound area cost factors. Johannes et al claimed that by knowing the prices of inputs and the level of output, it was possible to derive the minimum cost of producing any amount of output, also known as a “cost function.” Once a cost function has been developed, different regions can be compared using a cost factoring method to determine exactly what the regional cost differences are. Sponsored by the U.S. Department of Commerce, Johannes et al used the Cobb-Douglass production function to determine the average cost of construction in a generic region against the Washington DC’s baseline. Johannes’s regional cost mostly depends on two factors: the area cost factor and the size cost factor. The former is dependent on differences in labor and material prices, in the amount of construction activity, and the production technology. These factors were used to develop a datasets of LCAFs for military construction projects. Johannes et al suggested an approach to LCAF interpolation that was based on proximity and local data collections, but it did not validate it. Whereas their research is seminal to LCAF development, it does not provide an approach that would be feasible at the conceptual stage because of the lack of details on production technologies and specific materials to be employed in a project.

Geographic Information Systems

In all sciences, there is an underlying aspiration to understand how the physical world works. Geographic information science is a discipline in which people try to understand how the world works by evaluating and describing human relationships with the Earth (Poku & Arditi, 2006). To this end, geographic information systems (GIS) were created as tools to visualize and analyze these spatial relationships. GIS tools incorporate database files with geographically referenced thematic data, meaning files can contain a geographic location as well as specific themes or attributes that pertain to the location. GIS tools have been around since the early 1980’s and were one of the fastest growing computer-based technologies of the 1990’s (Bolstad, 2005). GIS have advanced technologically, and have been used in a multitude of industries as analytical, managerial, and visualization tools.

GIS have been successfully implemented in various fields, including construction-related fields with contributions in several aspects of construction project controls including scheduling, planning, and even
material procurement. However, GIS contribution to cost estimation, especially at a conceptual level, has been minimal. Cheng & O’Connor (1996) evaluated the use of GIS for enhanced construction site layout. Similarly, Cheng & Yang (2001) studied GIS-based cost estimates integrated with material layout planning. Zhong et al. (2004) studied GIS-based visual simulation methodologies and their applications in concrete dam construction processes. Oloufa et al. (1994) integrated GIS for construction site investigation. Li et al. (2003) proposed an internet-based geographical information system for E-commerce applications in construction material procurement. Even with all these examples, the full potential of GIS in the construction industry has not been realized (Jeljeli et al., 1993). In addition, researchers have indicated that despite widespread application of GIS in the construction industry, project visualization involving GIS has not yet been used to its full potential (Bansal & Pal, 2007).

Few contributions on using GIS for construction cost data handling and visualization, however, exist. Ashur and Crockett (1997) demonstrated that GIS can be used to analyze cost data and facilitate cost estimation through the power of geographic management. A fundamental concept of GIS is the ability to integrate geographic systems and database spreadsheet information systems. Using GIS, Ashur and Crockett retrieved and displayed unit price data for road projects. Typically, transportation agencies estimate construction project costs based on historical bid data. Ashur and Crockett showed that with the aid of GIS, a systematic information collection, organization, and storage process can be used so that relevant historical cost data can be retrieved. They concluded that using such technology would “assist in easing the ever-increasing demand to analyze information to support more effective decision making” (pp. 112). Later, Bansal & Pal (2007) researched the effect of using GIS for building cost estimation and visualization. They proposed a 5-step procedure for quantity takeoff that integrate the use of computer aided design (CAD) tools and GIS to generate location-specific bills of quantity (BOQ).

For the purpose of this research, GIS tools were utilized to visualize the spatial relationships between US cities with RS Means CCI location factors. In addition, statistical testing within GIS was used to test autocorrelation between proximity of cities and CCI values. GIS was utilized in this study mainly through its functionality of spatial estimation methods. Spatial estimation incorporates interpolation and prediction techniques. Interpolation and prediction techniques allow us to estimate variables at locations where they have not been measured. According to Bolstad (2005), spatial prediction differs from spatial interpolation because it uses a statistical fitting process. Spatial prediction uses rules and equations whereas interpolation only uses a set algorithm. Bolstad (2005) admits, “Our distinction between spatial prediction and interpolation is artificial, but it is useful in organizing our discussion, and highlights an important distinction between our data-driven models and our fixed interpolation methods” (p. 409). Due to the ambiguous distinctions between the two techniques, in our study there was no distinction between interpolation and prediction. Instead, the two terms are used interchangeably, both referring to spatial
estimation. Bolstad reveals that there are many spatial estimation methods, but the following are the most common: (a) Thiessen (Nearest Neighbor) Polygon, (b) Local Averaging (Fixed Radius), (c) Inverse Distance Weighted, (d) Trend Surface, and (e) Kriging (p. 428). Each respective method has inherent advantages and disadvantages and no method has been proven to continually outperform all others.

In this study, the authors utilized the nearest neighbor spatial estimation method as well as a similar version of the local averaging method. Bolstad conceptually defines nearest neighbor as the simplest method, in the sense that the mathematical function used is simply equality function and the nearest point is used to assign a value to an unknown location. In local averaging, cell values are defined based on the average of nearby samples. The number of samples depends on what search radius value is defined. In this study, a traditional search radius value was not defined. Instead, the state boundary was used to define the spatial extents of the search. To demonstrate this concept, all values within a state were averaged to estimate a collective value used for every potential project location within the state. In addition, this study incorporated spatial auto-correlation measured within GIS. Bolstad (2005) defines spatial auto-correlation as the tendency of nearby objects to vary in concert, meaning high values are found near high values, and low values are found near low values. If auto-correlation between variables that affect location adjustment accuracy is studied, this knowledge can be incorporated into the estimation process.

RESEARCH METHODOLOGY

Research Objectives and Process

The specific focus of this research was to assess spatial interpolation LCAF methods and compare their performance. LCAF methods based on Thiessen polygon and local averaging approaches were identified and compared. All methods aimed at the identification of a “twin location” among the dataset of sampled locations that would estimate the location factor for an unsampled location. It is important to understand what is meant by twin location. This twin location is what each location adjustment method found as the ideal alternative to an actual city that lacks a location factor. It varies depending on which interpolation method was considered. The location factor value for the twin location is what would be used if the original city did not have a value. The approach for selecting the twin location is what differentiates each interpolation method.

The initial research goal was to identify criteria that could be used to identify twin locations. Whereas the literature identified several criteria that are expected to affect construction costs, many criteria cannot be easily adopted for construction estimates. The authors have selected two criteria that can be easily transferred into interpolation methods. First, the authors assessed the value of Proximity (Criterion 1). This assessment was performed through the computation of the Moran’s I test statistic.
Whereas proximity in term of travel time could be seen as a more meaningful criterion for construction cost, this study adopted linear distance for simplicity. Another criterion, State Jurisdiction (Criterion 2) was analyzed and methods based on these initial two criteria were compared, including:

- **Nearest Neighbor (NN):** an estimator would select the value of the closest sampled location to estimate the LCAF for an unsampled location. To use this approach, Thiessen polygons would be drawn as shown in Figure 1a and 1b. This is the industry’s current method.

- **Conditional Nearest Neighbor (CNN):** this method evolves upon the NN method by selecting the value of the closest sampled location within the same state as the unsampled location. This approach considers the impact of state regulations and policies on construction costs and relies on Thiessen polygons drawn at the state-level as shown in Figure 1c.

- **State Average (ST AVG):** this method partially mitigates the effect of proximity by assigning to all locations within a state the average LCAF value among the sampled locations. In this case, the impact of state jurisdiction is spread out over all cities within the state. In this case, the twin location for all cities within the state is virtually assumed as an imaginary location located in the state capital with a location factor equal to the average value of the within-state LCAFs. This is a non-spatial average whereas the value would result from the local averaging approach applied at the state-level as shown in Figure 1d.

The location factor of the twin location was used as an estimated location factor pertaining to each city in the RS Means CCI dataset under the assumption that such city wouldn’t have a LCAF value. The comparison between the actual and estimated values was used to assess the performance of each interpolation method. The performance of each LCAF method was compared with descriptive and inferential statistical evaluations, including (a) mean, median, standard deviation, and variance of error, (b) histograms, (c) box plots, (d) Levene’s tests, and (e) Mann-Whitney tests.

**Data Collection and Sample Size**

The authors used the LCAF dataset published by RS Means in the 2006 Building Construction Cost Data book. These factors are named City Cost Indexes (CCIs). CCIs are the result of a research on the nine building types that are most often constructed in the United States and Canada. Results of this research are combined to create a composite model, used as reference building. Taking 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, RS Means calculates the indexes for the composite model. By averaging index values for the 30 major US cities, RS Means obtains the 30-City National Average Index that is assumed equal to the National Average. Therefore, the CCI for each city is...
calculated as a percentage ratio of a specific city’s cost to the national average cost of the same item. 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 and less influence to those trades that are usually the least expensive. This weighted average provides the LCAF for a given location. To estimate the construction cost in a location A, known 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 (RS Means, 2008).

In addition, RS Means limits the scope of application of its dataset. For instance, projects are limited to new commercial or industrial projects that cost $1,000,000.00 or more. The 2006 dataset included 649 sampled observations for the continental U.S. It was assumed that the statistical observations made using this national-level sample size was an accurate estimation of what should happen for the entire population. The entire population in this study included all cities within the contiguous United States, meaning all the 25,515 locations that are defined as “populated places” by the U.S. Census. According to proven statistical rules, such as the law of large numbers and the central limit theorem, it can be inferred that a larger sample size leads to increased precision in hypothesis tests.

According to Lenth (2001), sample size should be “large enough” that an effect of such magnitude as to be scientifically significant will also be statistically significant. Lenth continues, “Sample size is important for economic reasons: An under-sized study can be a waste of resources for not having the capability to produce useful results, while an over-sized one uses more resources than are necessary” (p. 2). While sample size determination is a common statistical problem, there are many limitations in the research design itself affecting sample size outcomes. The following are examples of these limitations: (a) Cost considerations, (b) Complexity of the design, (c) Research Deadlines, and (d) Minimum acceptable level of precision. These limitations make sample size determination a somehow subjective process. One thing that does hold true is that as the sample size increases, so does the precision of hypothesis test outcomes. In this research, we used two datasets to explore the influence of sample size on research results: (a) the complete CCI dataset that included 649 values out of the around 25,515 cities as represented in Figure 2a, and (b) a smaller sub-sample that included all values from a region that was randomly chosen to represent the national-level sample, as represented in Figure 2b. This “regional-level” sample included the 82 available CCI observations out of the 2,279 cities in New Mexico, Colorado, Arizona, Utah, Nevada, and California.

<Figure 2 here>
**Research Methods**

Since proximity-based interpolation is the preferred choice for LCAF estimation, the authors decided to substantiate the validity of this approach by conducting some specific analysis. To this end, the global Moran’s I test statistic, within ArcMap GIS software, was used to evaluate the degree of spatial autocorrelation between RS Means values and proximity. If evidence of significant auto-correlation results from the Moran’s I tests, it will substantiate the validity of proximity-based spatial-interpolation, and ultimately provide statistical justification of the NN method. According to Bolstad (2005):

- Moran’s I values approach a value of +1 in areas of positive spatial correlation, meaning large values tend to be clumped together, and small values clumped together.
- Values near zero occur in areas of low spatial correlation (pg. 412).

A negative correlation is shown as the Moran’s I values approach a value of –1 and a positive correlation is shown as the Moran’s I values approach +1. Correspondingly, a statistical Z score was calculated as part of the Moran’s I test. The Z-score evaluated if the null hypothesis should be rejected. The null hypothesis in this research essentially stated that "there was no spatial clustering of cities with similar CCI values." Using a 95% confidence interval and a 0.05 significance level, the Z-score must be less than –1.96 or greater than 1.96 to reject the null hypothesis with statistically significant confidence.

The evaluation of the homogeneity of variance between groups was necessary to determine what types of statistical tests were appropriate in this study. To this end, the authors utilized the Levene’s test because it was considered a common approach to determine homogeneity of variance between groups. According to Kault (2003):

- The common test for equal variance is called the Levene test. This test uses the principle that equal variances within each group by definition means equal values for the average of the square of the differences between each value and the group average… The Levene test then does a preliminary ANOVA to see if there is evidence against the assumption that the size of the difference between a value and its group average is on average the same in every group (p. 202).

This test did not determine which method outperformed other methods, but was useful to identify what types of statistical tests were appropriate to this purpose. If the Levene’s tests showed results that provided evidence of different variances, non-parametric tests were appropriate.

In this study, non-parametric test were used in lieu of more traditional statistical tests (t-tests). One of the common t-test assumptions is that the data have the same variances. Since the Levene’s test determined that the initial interpolation methods had significantly different variances, non-parametric tests such as Mann-Whitney were deemed appropriate. Mann-Whitney is a non-parametric test, which was used to test whether two independent samples of observations have statistically differing medians. According to Kinnear (2004):
When there are serious violations of the assumptions of the t-test, nonparametric tests can be used instead…the comparable nonparametric test may lack the power to reject the null hypothesis…The Mann-Whitney test is an alternative to the independent samples t-test (p. 179).

**DATA ANALYSIS AND RESULTS**

**Moran’s I Test**

Moran’s I tests were conducted nationally and for each individual state within the U.S. This created national and state level results. At the state-level, only 46 states contained enough data points to measure spatial autocorrelation. Results indicated that 24 of these 46 states returned positive Moran’s Index values. Furthermore, 19 of these 24 states returned positive Moran’s Index values, and Z-scores greater than 1.96. Consequently, there was evidence of positive, statistically significant auto-correlation between proximity and RS Means CCI values at the state-level.

Results of the national-level Moran’s I test statistic returned a positive Moran’s index value of 1.59 and a significant Z-score of 82.18. The Z-score of 82.18 indicated results were highly significant, and the Moran’s index value of 1.59 indicated spatial clustering of RS Means values across the entire nation. The tests results at the national-level, act as an internal validation supporting proximity-based interpolation. Being the current NN interpolation method proximity-based, this test has proven the statistical validity of the NN method responding to our first research question.

**Performance Comparisons**

As mentioned earlier, the 2006 RS Means CCI dataset was used in this research. From this dataset, a total of 649 cities were referenced as points on a map using ArcMap GIS software. The actual CCI location factor values pertaining to each city from the RS Means dataset were added as attributes and spatially associated with each corresponding city. The cities were then exported as a new data layer and a map layout was created which displayed the United States and the cities with an RS Means CCI value as shown in Figure 1a.

For each location adjustment method, a “twin location” was selected to represent each city in the sample. Each city was associated to a vector of four values: (a) LCAF<sub>Act</sub>, (b) LCAF<sub>NN</sub>, (c) LCAF<sub>StAvg</sub>, (4) LCAF<sub>CNN</sub>. The following sequence describes this process (1) we assumed that a location in the CCI dataset was missing its CCI value (LCAF<sub>Act</sub>), (2) we utilized the three interpolation methods (NN, CNN, St AVG) to identify a twin location, (3) we associated to this city the values of the CCI from the up to three twin locations (LCAF<sub>NN</sub>, LCAF<sub>StAvg</sub>, LCAF<sub>CNN</sub>), (4) to determine the error of each interpolation method, we compared the CCI value of this twin location against the actual CCI value that was initially ignored. Therefore, the CCI value of the twin location was used as an estimated CCI value pertaining to each RS Means city. Since each city in the CCI dataset has an actual known value (LCAF<sub>Act</sub>), the difference between estimated and actual values is what distinguishes the performance of the methods.
This calculation produces an “error” between estimated and actual values. This approach to calculating error is a common practice in many scientific research studies. In this study, error took the form of an overestimate or an underestimate. When the difference between estimated and actual data was positive, it was interpreted that the value of the method-specific twin location was overestimating the cost of construction. Similarly, if the difference was negative this corresponded to an underestimate. This approach relies on the assumption that CCI values (or any other LCAF) are good predictors of local construction costs. Since the goal of the performance comparison was to assess performance of various interpolation methods, this assumption is reasonable.

Initially, the authors tested if the NN method became less reliable as the distance between RS Means CCI locations and their respective twin locations increased. Error for the NN method was measured as a function of distance to determine if the method became less reliable as the distance between CCI locations increased. The idea was to test observations of Waldo Tobler (1970). In Tobler’s first law of geography he stated that “…everything is related to everything else, but near things are more related than distant things.” To test this theory, error was measured as a function of distance as shown in Figure 3. Results from this figure indicated that many small errors occurred even at greater distances and many large errors occurred even at shorter distances. This suggested that the NN method did not become less reliable when proximity between actual and estimated CCI values increased. Unexpectedly, some odd patterns were found, such as greater errors (10 or greater) occurring at very short distances and smaller errors (5 or less) at greater distances. As a result, it was determined that Tobler’s first law did not entirely hold true for the NN interpolation method evaluated in this study. It was inferred that as the distance between a city and its respective twin location increased or decreased, there was no substantial evidence that the degree of error was proportional to the change in distance. In other words, error was not directly related to the distances between CCI locations and twin locations. Error did not become greater simply because of greater distances between a city and the twin location. Correspondingly, error did not become less due to shorter distances.

Descriptive statistics were considered for better analyzing the performance of the three methods. Absolute error values were considered in all calculations. Descriptive statistic comparisons were used to determine if an alternative method could be statistically proven to outperform the current method. Five levels of error were defined with the goal of easily visualizing the results: very low, low, medium, high and very high. The actual count and percentages of how many cities were included in these levels are shown in Table 1.

Absolute error was used in this evaluation. Results indicated that the CNN method had the least count and lowest percentage of the three methods at medium, high and very high error levels.
Correspondingly, it also has the highest count and percentage at very low and low error amounts. Therefore, CNN outperformed NN and ST AVG.

In addition, mean, median, and standard deviation of absolute error values for all three methods were calculated and summarized in Table 2.

The median and mean error for the NN method was less than ST AVG method, but greater than the CNN. In addition, standard deviation was highest in the NN method. Overall, results indicated that the CNN method had the least median, mean, and standard deviation of error. Therefore, CNN outperformed NN and ST AVG. Table 1 also shows a comparison of the number of overestimates, underestimates, and accurate estimates for each of the three interpolation methods studied at this time. Relative error (positive and negative error calculations) was used in the comparison. This analysis was conducted to determine if a set pattern could be observed and revealed a slight increase in overestimates for all methods. However, there were no relatively significant or extreme differences between the number of overestimates and underestimates for each method except that all three methods showed a slight tendency to overestimate. A comparison of the number of accurate estimates found that CNN outperformed ST AVG but did not outperform NN. However, these methods provided perfect estimates only for a few locations.

**Histograms**

Histograms of error from the methods evaluated in Phase 1 of this study were compared and analyzed. The error considered in these histograms was relative, meaning positive and negative values are shown. Histograms were incorporated because they can visually demonstrate statistical comparisons including the distribution of error and outliers. Figure 4 shows the national-level histograms comparison of NN, CNN and ST AVG.

Results from the comparison of NN and ST AVG histograms indicated that there were more outliers in the NN method. There were also higher frequencies of low error amounts in the NN method than the ST AVG method. Comparing NN and CNN, histograms indicated that CNN had higher frequencies of lower error. In addition, CNN had lower outlier values. Similarly, comparing CNN to ST AVG, histograms indicated that CNN had higher frequencies of low error and lower outlier values again. These histograms suggest the CNN method produces more accurate estimates.

As part of the empirical comparison, descriptive statistics for all methods were produced for both regional and national-levels. This was calculated in order to determine which method produced the lowest statistical error results. Table 3 shows the results at the regional and national levels using absolute error values.
National-level results indicated that CNN produced the lowest mean, median, standard deviation, and variance. ST AVG method produced some of the highest central tendency values whereas NN method offered larger variances and standard deviations. Similar results were found at regional-level.

Table 4 shows the results at the regional and national levels using relative error values.

Table 4

<table>
<thead>
<tr>
<th>Method</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>0.01</td>
<td>0.005</td>
<td>0.00025</td>
</tr>
<tr>
<td>ST AVG</td>
<td>0.02</td>
<td>0.015</td>
<td>0.00225</td>
</tr>
<tr>
<td>NN</td>
<td>0.04</td>
<td>0.030</td>
<td>0.009</td>
</tr>
</tbody>
</table>

At the national-level, CNN still produced the lowest Median, standard deviation and variance with ST AVG having the lowest mean. At the regional-level, CNN still produced the lowest standard deviation and variance. However, CNN produced the highest mean and median. In addition, CNN had one of the lowest median, standard deviation, and variance of error at the regional-level.

Box Plots

Continuing with the statistical analysis, box plots showing relative error for various methods at both the national and the regional levels were evaluated and are shown in Figure 5.

Results indicated that all methods showed evidence of outliers. These are the extreme values that deviate significantly from the rest of the data. Essentially, these are the circles or asterisks found above and below the whiskers. Box plots did not show a large difference in the medians among these methods. At the national level, CNN had the least spread between whiskers, and also had the lowest outlier values. There did not appear to be a large difference in the box sizes. At the regional level, CNN still outperformed the other two methods, but had higher outliers than ST AVG.

Levene’s Test

The Levene’s Test for equality of variance was conducted at the national level and its results are shown in Figure 6.

Results indicated that the P value (0.002) was less than the significance level (0.05) therefore there is evidence to reject null hypothesis that the variances between methods are the same. As part of the test, Bonferroni confidence intervals (CI) were shown. The CI for CNN was well separated from the other two methods, also showing evidence to reject the null hypothesis. Table 5 shows bi-variable Levene’s tests for homogeneity of variances at the national and regional levels.

Table 5

<table>
<thead>
<tr>
<th>Comparison</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN vs NN</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>CNN vs ST AVG</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>ABC vs ST AVG</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

National level test results indicated that there is significance less than .05 between CNN versus NN and respectively between CNN versus ST AVG. There is a statistically significant difference between the variances of these methods. Regional level test results had two comparisons that resulted in significance less than .05. These comparisons included CNN versus ST AVG and ABC versus ST AVG. There was a statistically significant difference between the variances of these comparisons.
**Mann Whitney Tests**

The Mann-Whitney test was evaluated at national and regional-levels. Table 6 shows the national level test results.

<Table 6 here>

For all methods, the null hypothesis was retained meaning there was no significant difference between the medians. Similar results were obtained at the regional level.

**CONCLUSIONS**

A common problem in the construction industry today involves cost estimate location adjustment for locations that do not have location adjustment factors. This study evaluated the current interpolation method used for estimating these unknown location adjustment factors. The current method referred to “nearest neighbor” interpolation, which was a spatial estimation technique based on linear distance and proximity. This technique basically estimates a variable for a city solely based on the same variable of the closest proximate city. This study used the 2006 RS Means CCI dataset to assess the statistical validity of this approach and compare its performance against two alternative methods. This study was limited to an internal validation of location adjustment methods and relied on the assumption that 2006 CCI values were accurate predictors of construction costs.

Strong spatial autocorrelation of the CCI values was confirmed using a Moran’s I test. This autocorrelation provides strong support for the use of proximity-based interpolation methods to estimate CCI values. This suggests the current location adjustment is statistically sound. Statistical testing analysis provided evidence that the CNN method outperformed all other methods at the national level. At a regional level, CNN performed well but results were inconsistent showing the importance of sample size to accurate LCAF estimates. The exploratory nature of this study also allowed the authors to identify several potential areas for future research, including the use of geographical cost functions that would overcome the limitation of selecting twin locations, and the identification of socio-economic variables that can are highly correlated with construction costs and can be used to develop geographically-based regression models (Migliaccio et al 2013). Similarly, a different approach to local averaging could be pursued that does not use state boundaries. Instead, various radiiuses could be used and their performance compared against socio-economics to understand how the area of influence of twin locations could vary depending on differing socioeconomics in a certain region.

**AKNOWLEDGMENTS**

The authors acknowledge Dr. Michele Guindani for his valuable insight on the statistical analysis approach adopted in this study. Mr. Su Zhang and Ms. Maria D’Incognito’s contribution to the literature review are acknowledged with grateful thanks. In addition, the authors want to thank the reviewers of this technical paper who have provided precious suggestion to improve this manuscript.
REFERENCES


FIGURE CAPTIONS LIST

Table 1 — Comparison by Magnitude of Error

Table 2 — Error Descriptive Statistics

Table 3 — Absolute Error Values Statistics

Table 4 — Relative Error Values Statistics

Table 5 — Levene’s Test Results

Table 6 — Mann-Whitney Test National-Level Results

Figure 1 — Interpolation Methods

Figure 1a - Thiessen Polygons Drawn using NN Method: Continental U.S.
Figure 1b - Thiessen Polygons Drawn using NN Method: New Mexico
Figure 1c - Thiessen Polygons Drawn using CNN Method: New Mexico
Figure 1d – State Average Method

Figure 2 — National-level and regional-level samples

Figure 2a – RS Means Dataset for Continental U.S.
Figure 2b – Subsample of RS Means Dataset for Southwestern Region

Figure 3 — NN Error as a Function of Distance

Figure 4 — National-Level Histogram Comparison of NN, CNN and ST AVG

Figure 5 — National-Level and Regional-Level Box Plots.

Figure 6 — Levene’s Test Results for CNN, NN, and ST AVG
Accepted Manuscript
Not Copyedited
Accepted Manuscript
Not Copyedited
<table>
<thead>
<tr>
<th></th>
<th>National-Level Results</th>
<th></th>
<th>Regional-Level Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NN</td>
<td>St Avg</td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Test for Equal Variances for values

95% Bonferroni Confidence Intervals for StDevs

Bartlett’s Test
Test Statistic: 20.00
P-Value: 0.000

Levene’s Test
Test Statistic: 6.46
P-Value: 0.002

Accepted Manuscript
Not Copyedited
### Table 1 — Comparison by Magnitude of Error

<table>
<thead>
<tr>
<th>Interpolation Methods</th>
<th>Comparison</th>
<th>VL (0-1%)</th>
<th>L (1-3%)</th>
<th>M (3-5%)</th>
<th>H (5-10%)</th>
<th>VH (&gt;10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST AVG</td>
<td>count</td>
<td>118</td>
<td>202</td>
<td>119</td>
<td>131</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>percentage</td>
<td>18%</td>
<td>32%</td>
<td>18%</td>
<td>20%</td>
<td>12%</td>
</tr>
<tr>
<td>NN</td>
<td>count</td>
<td>156</td>
<td>186</td>
<td>104</td>
<td>137</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>percentage</td>
<td>24%</td>
<td>29%</td>
<td>16%</td>
<td>21%</td>
<td>10%</td>
</tr>
<tr>
<td>CNN</td>
<td>count</td>
<td>178</td>
<td>218</td>
<td>93</td>
<td>110</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>percentage</td>
<td>27%</td>
<td>34%</td>
<td>14%</td>
<td>17%</td>
<td>8%</td>
</tr>
</tbody>
</table>

VL – Very Low; L – Low; M – Medium; H – High; VH – Very High

### Table 2 — Error Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>CNN</th>
<th>NN</th>
<th>ST AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Error</td>
<td>1.95</td>
<td>2.30</td>
<td>2.56</td>
</tr>
<tr>
<td>Mean Error</td>
<td>3.07</td>
<td>3.78</td>
<td>3.80</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.09</td>
<td>4.08</td>
<td>3.77</td>
</tr>
</tbody>
</table>

**Error Classification**

<table>
<thead>
<tr>
<th>Classification</th>
<th>CNN</th>
<th>NN</th>
<th>ST AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underestimates</td>
<td>332</td>
<td>333</td>
<td>327</td>
</tr>
<tr>
<td>Overestimates</td>
<td>301</td>
<td>303</td>
<td>314</td>
</tr>
<tr>
<td>Perfect estimates</td>
<td>15</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>649</td>
<td>649</td>
<td>649</td>
</tr>
</tbody>
</table>

CNN - Conditional Nearest Neighbor
NN - Nearest Neighbor
ST AVG - State Average

### Table 3 — Absolute Error Values Statistics

<table>
<thead>
<tr>
<th>METHOD</th>
<th>National Sample (Absolute Error Values)</th>
<th>Regional Subsample (Absolute Error Values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>NN</td>
<td>3.78</td>
<td>2.30</td>
</tr>
<tr>
<td>St Avg</td>
<td>3.80</td>
<td>2.56</td>
</tr>
<tr>
<td>CNN</td>
<td>3.07</td>
<td>1.95</td>
</tr>
<tr>
<td>MIN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td>MAX</td>
<td>St Avg</td>
<td>St Avg</td>
</tr>
</tbody>
</table>
Table 4 — Relative Error Values Statistics

<table>
<thead>
<tr>
<th>METHOD</th>
<th>National Sample (Relative Error Values)</th>
<th>Regional Subsample (Relative Error Values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>NN</td>
<td>0.23</td>
<td>0.10</td>
</tr>
<tr>
<td>St Avg</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>CNN</td>
<td>0.16</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 5 — Levene’s Test Results

<table>
<thead>
<tr>
<th>National Level</th>
<th>CNN vs. NN</th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN vs. St Avg</td>
<td>14.727</td>
<td>1</td>
<td>1294</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Regional Level</td>
<td>CNN vs. NN</td>
<td>2.316</td>
<td>1</td>
<td>162</td>
<td>.933</td>
</tr>
<tr>
<td>CNN vs. St Avg</td>
<td>6.372</td>
<td>1</td>
<td>162</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>St Avg vs. NN</td>
<td>0.437</td>
<td>1</td>
<td>162</td>
<td>.509</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 — Mann-Whitney Test National-Level Results

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Test</th>
<th>Significance</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 The distribution of national relative error for CNN vs. NN is the same across</td>
<td>Independent-Samples</td>
<td>0.742</td>
<td>Retain the null</td>
</tr>
<tr>
<td>categories of national level sample size.</td>
<td>Mann-Whitney U Test</td>
<td></td>
<td>hypothesis.</td>
</tr>
<tr>
<td>2 The medians of national relative error for CNN vs. NN are the same across</td>
<td>Independent-Samples</td>
<td>0.978</td>
<td>Retain the null</td>
</tr>
<tr>
<td>categories of national level sample size.</td>
<td>Median Test</td>
<td></td>
<td>hypothesis.</td>
</tr>
<tr>
<td>3 The distribution of national relative error for CNN vs. ST AVG is the same</td>
<td>Independent-Samples</td>
<td>0.893</td>
<td>Retain the null</td>
</tr>
<tr>
<td>categories of national level sample size.</td>
<td>Mann-Whitney U Test</td>
<td></td>
<td>hypothesis.</td>
</tr>
<tr>
<td>4 The medians of national relative error for CNN vs. ST AVG are the same across</td>
<td>Independent-Samples</td>
<td>0.912</td>
<td>Retain the null</td>
</tr>
<tr>
<td>categories of national level sample size.</td>
<td>Median Test</td>
<td></td>
<td>hypothesis.</td>
</tr>
</tbody>
</table>

Asymptotic significances are displayed. The significance level is .05.

Accepted Manuscript
Not Copyedited