A Modeling Approach to Indexing the Cost of Highway Construction and Maintenance Projects to A Responsive Highway User Fee Using Bid Items Data

By

Said Fathy El Said

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Civil Engineering

January 31, 2022

Nashville, Tennessee

Approved:

Robert Stammer, Ph.D., P.E.

Sanjiv Gokhale, Ph.D., P.E

Mark Abkowitz, Ph.D.

Kenneth Pence, Ph.D.

Paul Degges, P.E.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/NOMENCLATURE/SYMBOLS

ACF	Autocorrelation Function
ADF	Augmented Dicky Fuller
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AM100K	Contract Amount $100,000
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ASCE	American Society of Civil Engineers
ASPH	Asphalt
AWG	American Wire Gauge
BPI	Bid Price Index
BRCN	Bridge Construction
BRRP	Bridge Repair
CONC	Concrete
COVID-19	Coronavirus Disease 2019
CPI	Consumer Price Index
DF	Dicky Fuller
DOT	Department of Transportation
DUR	Contract Duration
EMER	Emergency
FF	Federally Funded
FHWA	Federal Highway Administration
GDPDEF	Gross Domestic Product Implicit Price Deflator
GEN	General Construction
GRDG	Grading
HTF	Highway Trust Fund
IMPS	Improvements
INCI	Incidental

ITS	Intelligent Transportation Systems
MA	Moving Average
MLR	Multiple Linear Regression
MTNC	Maintenance
NHCCI	National Highway Construction Cost Index
OH	Overhead
PACF	Partial Autocorrelation Function
PC	Portland Cement
PPI	Producer Price Index
PPIC	Producer Price Index Concrete
PPIS	Producer Price Index Steel
PVMK	Pavement Marking
R^2	R-Squared Coefficient of determination
SRFT	Surfacing
TC	Tack Coat
TDOT	Tennessee Department of Transportation
TNHCCI	Tennessee Highway Construction Cost Index
TRB	Transportation Research Board
VMT	Vehicle Miles Traveled

INTRODUCTION

The transportation system in the United States is primarily funded through fuel taxes (i.e., gasoline, diesel, and special fuels), with the gasoline tax as the primary source of funding. Fuel taxes, along with other direct and indirect excise taxes, were intended to provide the funds needed to finance construction and maintenance for the highway system. Unfortunately, the current collected taxes are not equivalent to the actual full costs of highway construction and maintenance projects. In fact, "The U.S. has been underfunding its highway system for years, resulting in an $836 billion backlog of highway and bridge capital needs" (ASCE 2017). Taxes on gasoline and diesel fuel were created to form an excise user fee. This user fee has gradually decreased in percent value over time and currently is not meeting the nation's financial demands for constructing and maintaining the nation's highway system.

The funding for highway construction and maintenance projects should ideally be based on system needs and current system usage. Several earlier research studies tested solutions focused on charging a fixed fee based on vehicle miles traveled and concluded that such a change does not fully address the loss of value due to the overall increase in vehicle fuel efficiency (Small and Van Dender 2007; Parry et al. 2007). While some suggested solutions create a mechanism that accounts for taxing all vehicles according to miles driven, the presented solutions still promote a fixed fee per mile (Atkinson et al. 2009). Replacing the excise tax on gallons of fuel sold with a fixed fee per miles driven is often suggested as a solution to tax electric vehicles. This is to account for the highway user fee as no taxes are collected similar to other vehicles using fossil fuels. Such solutions still do not account for inflation and do not adapt to changes that erode the value of the fixed fee (i.e., changes in material costs). Thus a per miles traveled tax is a similar fixed fee tax like the current per gallon consumed tax. Other solutions promote an adjustable fee per gallons of fuel based on national indices without consideration of local factors.

A highway financial funding system relying heavily on a fixed excise tax rate or adjusted based on global factors is not sensitive to local factors affecting the costs of projects. Such a funding system lacks the intended direct link between highway project costs and the collected user fees.

This research provides the tool to adjust the fuel tax based on current costs of highway and construction projects relative to project, local, and global factors. This proposed adjustment tool can be used to adjust the tax rates using the developed Highway Construction Cost Index. Such adjustments can be applied, regardless of the method by which the taxes are being charged. Thus the change to a fixed fee tax will change according to fiscal needs (i.e., based on miles traveled or gallons of fuel sold). However, this research does not address the need to identify and prioritize highway projects. This decision process is often performed during the planning stage using computer programming as is the case with the Tennessee Department of Transportation (TDOT) and many other state DOTs. Future research could combine the later mentioned planning decision process with the presented estimation and indexing process described in this research.

The first step in the methodology presented in this research starts by identifying the bid items that can be used in estimating project costs. The second step studies and analyzes the factors affecting the costs of the bid items and therefore the cost of highway projects. Then, a bid item price estimation model is developed based on the quantified change of highway construction and maintenance project costs. This bid item price estimation model should reasonably be able to estimate the future changes in the cost of contract bid items and therefore the cost of highway construction and maintenance projects. Using the future estimated cost of items and projects, an index can be developed to adjust the user fee accordingly. Applying an indexing mechanism to the existing fuel tax rate and quantifying the relation between costs and taxes will result in an equitable user fee to be charged per gallon of fuel. Such a mechanism could provide more stability and consistency for fiscal planning purposes.

The current project programming process in State Departments of Transportation (DOTs) is dependent on the estimated funds based on the projected gallons of fuel sold within the states. The gallon sale price includes an excise fuel tax amount. The gallons of fuel sold and consumed within a state is linked to the Vehicle Miles Traveled (VMT) and is influenced by the economy status on the state and the national level. Most recently, the VMT as seen in the following graph were severely reduced. The COVID-19 pandemic has influenced the VMT in the state of Tennessee and therefore affected the funds collected for highway construction and maintenance projects. As a result, the total VMT in the state of Tennessee in 2020 is less than 2019 and the total funds collected are less as well. The graph in Figure 1 shows a sudden drop of approximately 40% of traffic volumes (Comparing 2020 to 2019 volumes) following the pandemic quarantine measures applied in March of 2019. Then, a gradual recovery is noticed throughout 2020 to reach approximately 10% reduction of traffic volumes when compared to 2019.

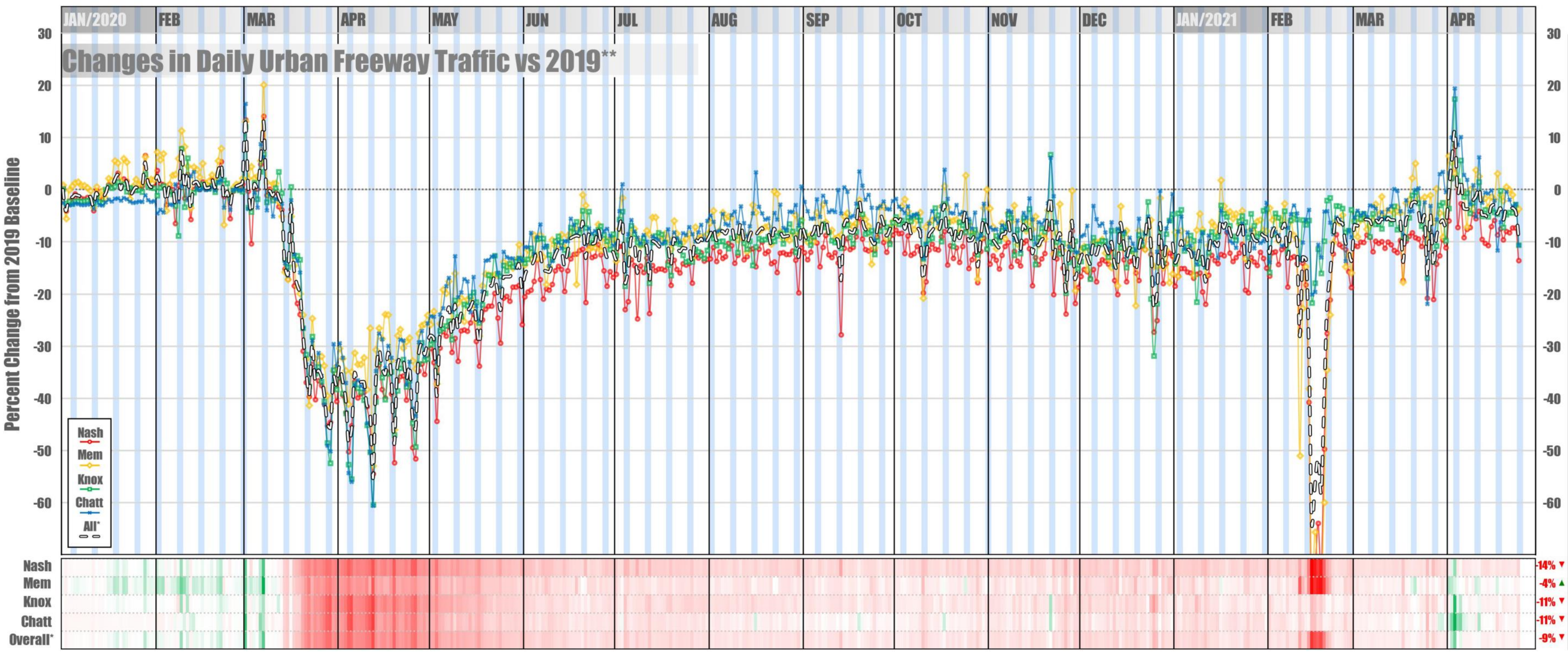

Figure 1 Changes in the Tennessee Daily Urban Freeway Traffic in 2020 vs 2019 (TDOT)

Thus, using the methodology developed by this research, DOTs can have more control on project funding by linking the excise tax rate to the highway construction and maintenance projects. In addition, quantifying the factors influencing the change in the cost of highway construction and maintenance projects allows the DOTs to influence bid prices and therefore project costs. Also, using the cost estimating model, the DOTs can estimate future bid item prices and therefore better estimate the cost of highway construction and maintenance projects. In addition, a State Highway Construction Cost Index can be used as a statewide local indicator and a reference to the highway projects spending and construction material price estimating.

While all bid line items affect the total cost of highway projects, a limited number of items can be used in the project costs estimation process. Most of the bid line items data include variations in observations, inconsistencies of non-standard pay items, and changes in the reporting methodology (Jeong et al. 2017). Selection criteria are used to edit the data and filter bid line items to select representative items according to the National Highway Construction Cost Index (NHCCI) by the Federal Highway Administration (FHWA) (NHCCI 2020). Some of the reviewed research presents a bid item selection method based on the weighted bid item costs versus the total project costs (Chou et al. 2006). The later mentioned research selected bid items that represented 80% of the total project costs and included the remaining 20% as a "roll-up" cost applied to the overall project cost estimate as a percentage cost markup. While the later mentioned selection method presents a simplified way of selection when compared to the NHCCI, this later method is not sensitive to the quality of the bid item price data. Future research can study the differences between the two selection processes.

In this research, Multiple Linear Regression (MLR) modeling techniques are used to develop an estimation model. The model is tested by regressing estimated data against actual costs using historical data that was not included in the model development. Later in the "Data" section, bid data are presented according to the number of observations, types of dependent variables, and limitations influencing the use of the data. The section then discusses the history of the current database and the need for developed data fields. The "Methodology" section includes Data Selection and Edits, Modeling Techniques, and

Indexing subsections. The Data Selection and Edits subsection:

1. Examines the bid line items of highway construction and maintenance projects,
2. Identifies significant bid items,
3. Applies data improvement techniques, and then
4. Calculates relative bidding data.

The research proposes a modeling and indexing approach using the 95% confidence level that accurately identifies the local and national factors controlling the change in bid item prices and indexes bid item prices. The developed index is used to calculate the projected change in the highway user fee that can be used to establish the desired link to the change in bid item prices.

The rest of the paper is organized as follows:

1. Quantifying the factors affecting the change in bid item prices and developing an estimation model using MLR are detailed in the Modeling Techniques subsection.
2. The Indexing subsection details indexing techniques applied to the user fee.
3. The “Results” section explains how the bid item price estimation model was tested and evaluated for the ability to produce useful data.
4. The “Discussion” and “Conclusion” sections summarize key research findings and conclusions.

LITERATURE REVIEW

Multiple earlier research efforts and reports were reviewed to identify:

1. Recommendations to change existing highway funding structure,
2. Indexing mechanisms addressing the highway system user fee, and
3. Modeling techniques.

Recommendations to Change Existing Structure

Some research performed national level studies and recommended adjusting fuel taxes on the federal level. These national research efforts and the state concepts share similarities between the funding source using a fixed excise tax rate, and addressing the rising costs of construction and maintenance of highways. The National Surface Transportation Infrastructure Financing Commission Report (Atkinson et al. 2009) identified the need to index fuel taxes and provided the following endorsement: “Congress should index all federal motor fuel taxes to inflation on a going forward basis.” The same report (Atkinson et al. 2009) recommended maintaining the security and sustainability of the Highway Trust Fund (HTF), applying immediate increases to the fixed fuel tax for the short and long term, and indexing the taxes to indices identified in the report.

The Transportation Research Board (TRB) formed a committee to research the long-term viability of the fuel taxes and alternatives for transportation financing. Reducing evasion and limiting exemptions, indexing tax rates, and reforming the use of debt finance were recommendations of the committee report titled, “The Fuel Tax and Alternatives for Transportation Funding.” (Board 2006)

Indexing Mechanisms Addressing the Highway System User Fee

Multiple indices were researched including the Consumer Price Index (CPI), Producer Price Index (PPI) for Highway and Street Construction, Gross Domestic Product Implicit Price Deflator (GDPDEF), and other indices created by industry organizations (Atkinson et al. 2009). Other research recommended to maintain the integrity of the current system by indexing tax rates using the CPI and other project costs indices compiled by the FHWA (Board 2006).

The FHWA produces the NHCCI that can be used to track price changes associated with highway construction costs on a quarterly basis (NHCCI 2020). The NHCCI is intended to replace the Bid Price Index (BPI) and is a way to enhance the FHWA process of monitoring the cost of highway projects by reviewing all construction projects, especially those costing more than $500,000. The NHCCI uses the Fisher Index in its mathematical calculations applied to the bid tabs data (collected by Oman Systems, Inc.) from all states and selects the data in a way to guarantee the consistency of the quality of the items.

The FHWA online published report (Atkinson et al. 2009) used several indexing and data editing techniques, including the NHCCI, and concluded that improvements are desirable and that additional research should be conducted. Additional research should focus on further developing computer programs to provide indices by individual states (NHCCI 2020).

Modeling Techniques

With regards to various modeling techniques, linear modeling techniques have previously been studied and used in estimating construction costs (Chou et al. 2006; Kyte et al. 2004; Lowe et al. 2006; Williams 2003; Williams 2002; Wright and Williams 2001). These studies have concluded that linear regression models have significant advantages with respect to accuracy, model creation, and model examination over other models. Simple linear regression has been used to identify predictors for final project costs (Williams 2003; Williams 2002). MLR was best used when multiple predictors were studied for final project costs (Chou et al. 2006; Kyte et al. 2004; Lowe et al. 2006; Wright and Williams 2001).

Independent variables used in modeling project costs include project characteristics related to project length, width, total cost, quality, duration, type, and location (Chou et al. 2006; Lowe et al. 2006;

Williams 2002). Other research studied bid data to include number of bids, low bid, mean and median bid, standard deviation, and bid item price spread (Williams 2003; Williams 2002; Wright and Williams 2001). Data transformation of cost and predictors has been introduced to reduce heteroscedasticity or varying standard deviations between predicted construction costs and independent model variables (Chou et al. 2006; Lowe et al. 2006; Williams 2003; Williams 2002; Wright and Williams 2001).

While linear regression assumes consistency in data variance or homoscedasticity, untransformed data may have random variation or heteroscedasticity that can be reduced by data transformation. Transformations using logarithm and the natural logarithm of the cost and the independent variables has shown varying results in improving the relationship. Earlier studies have investigated using neural network models and compared results to linear regression (Lowe et al. 2006; Williams 2002; Wright and Williams 2001). Some research studies found that neural networks can produce reasonable predictions, but linear regression was preferred over neural networks (Williams 2002; Wright and Williams 2001). Only one of the reviewed studies found that the neural network model performed slightly better than the regression model (Lowe et al. 2006).

One study used bid item quantities as the dependent variable for estimation instead of bid item prices (Chou et al. 2006). The same research used the historical local bid prices and escalation rates to achieve estimated project costs. This same study also used a general parametric function as a transformation that produced a general MLR. Then MLR was used to develop an estimation model (Chou et al. 2006).

DATA

This research uses the extensive Tennessee Department of Transportation (TDOT) database of highway contracts let to date. The database is used as a state DOT sample for bid line item prices. The database contains all contract line bid items included in highway contracts let between 2002 and 2016 with about 936,000 line bid items. The line items represent all bids on state highway projects for highway construction and maintenance contracts. Using the current data system presents an abundant source of reliable data with minimal flaws and inconsistencies compared to earlier available data. The database has 28,446 different bid items using 40 different units of measure. Line bid items are proposed on 6,396 highway construction and maintenance projects with a total award amount of $12.4 billion.

The line item prices data include the following data fields: letting date, contract award amount, bid quantity, number of bidders, project type, project location, estimated work duration, and number of projects in the letting. Using the letting dates, more variables are derived to include the change in fuel price, cost of construction materials, and CPI. The data has some limitations that are addressed by this research for including: non-standard bid items and ambiguous units. A full detailing of how these limitations are vetted is included in the next section entitled “Methodology”, and specifically in the accompanying Data Selection and Edits subsection. The data does not include project specific information from project plans and letting documents. Therefore, some limited factors may affect the change of bid item prices, but cannot be differentiated by the available data because project specific information is not included. While such information is included in the project specifications, contracting documents and plans, such information is not available through the bid item prices data used in the research. Therefore, some limited factors affecting the change of bid item prices are not included in the available data. For example, bid item prices will change due to the construction and maintenance tasks included within each bid item and differing site conditions between projects. To further explain, rural projects typically requiring long haul of materials will include an increase to the price that is not indicated in the data. Another example is information about geological conditions similar to rock excavation that is also not included in the data. Such information may contribute to an increase in the price of the bid item but cannot be differentiated by the available data.

Some bid line items are non-standard having the same item number but different descriptions or units of measure. For example, highway utility projects can have "per each" or "per pound" unit of measure for the same bid item, and thus introduces the possibility of a major inconsistency in bid prices between projects. For some other pay items, the unit of measure can make tracking price changes difficult (e.g., lump sum items as in construction stakes, lines, and grades). Similar to the project specific data, lump sum items are detailed in project letting documents. However, the data set does not include such details significant to the change in the lump sum bid item prices. In addition, some bid item categories are considered ambiguous or "suspect categories" as described by the NHCCI (NHCCI 2020). These categories relate to aspects that are specific to individual contracts such as start-up costs and incentives (e.g., Mobilization, Alternates, Bonuses, Time, and other categories).

The data set includes bid items that have no continuity. For the purpose of calculating an indexing tool, continuity of the bid items is necessary for the index and to link lagged values to available historical data. Some bid item observations have no lagged values and therefore cannot support the continuity needed for index development. In addition, some bid items have a different description for the same bid item number. Changes in the description are similar to introducing a different bid item under the same number and therefore disrupting the continuity needed for index development.

Some data items have observations that are considered outliers (NHCCI 2020). An example of an outlier is a project item with a small quantity that includes high overhead and fixed cost charges. Emergency contracts introduce a similar problem with outlier observations. Such contracts are known to have higher prices due to the high mobilization cost in allocating construction resources to small bid item quantities. Such mobilization is also required in a fast deployment process that is also known to increase the cost. The latter mentioned data issues and limitations are addressed by this research. A full detailing of how these issues are vetted is included in the next section entitled "Methodology" and the accompanying Data Selection and Edits subsection.

METHODOLOGY

Data Selection and Edits

During Data Selection and Edits, editing techniques are applied to eliminate problematic data as suggested by FHWA (NHCCI 2020). Non-standard pay items having the same item number but different descriptions or units of measure are excluded from the data set. In addition, ambiguous or "suspect categories" that cannot be used as a representative sample for all projects are removed. By applying the selection criteria, 34 bid items out of a possible 28,446 satisfied all selection criteria and are included in the data set. While the selected bid items represent 0.12% of all bid items in numbers, the cost of selected items represents approximately 15.6% of total cost of highway projects. Table 1 provides the list of the selected bid items by description, item numbers and unit of measure.

Table 1 Included Bid Items Using the NHCCI Techniques

	Item Description	Item Number	Unit of Measure
1	UNDERCUTTING	203-05	C.Y.
2	8-18" TEMPORARY SLOPE DRAIN	209-02	L.F.
3	CATCH BASIN PROTECTION (TYPE D)	209-40	EACH
4	BITUMINOUS MATERIAL FOR PRIME COAT (PC)	402-01	TON
5	BITUMINOUS MATERIAL FOR TACK COAT (TC)	403-01	TON
6	LOAD TRANSFER DOWELS	502-04	L.F.
7	PRECAST CONCRETE BOX CULVERT (8' X 4')	607-50	L.F.
8	CATCH BASINS, TYPE 12, > 4' - 8' DEPTH	611-12	EACH
9	CATCH BASINS, TYPE 14, > 4' - 8' DEPTH	611-14	EACH
10	CATCH BASINS, TYPE 38, 0' - 4' DEPTH	611-38	EACH
11	CATCH BASINS, TYPE 51, > 4' - 8' DEPTH	611-51	EACH

12	PRESTRESSED CONCRETE I-BEAM (TYPE II)	615-01	L.F.
13	PRESTRESSED CONCRETE BOX BEAM (39"X 36	615-02	L.F.
14	BRIDGE DECK SEALANT	617-01	S.Y.
15	CONCRETE COMBINED CURB & GUTTER	702-03	C.Y.
16	CEMENT CONCRETE DITCH PAVING	703-01	C.Y.
17	REINFORCED CONCRETE SLOPE PAVEMENT	709-04	C.Y.
18	AGGREGATE UNDERDRAINS (WITH PIPE)	710-02	L.F.
19	LATERAL UNDERDRAIN ENDWALL (3:1)	710-06	EACH
20	STEEL OVERHEAD SIGN STRUCTURE	713-09	EACH
21	FLAT SHEET ALUMINUM SIGNS (0.100" THICK)	713-13	S.F.
22	EXTRUDED ALUMINUM PANEL SIGNS	713-14	S.F.
23	CABLE (1/C # 10 AWG)	714-06	L.F.
24	REMOVAL OF LIGHT STANDARD & FOUNDATION	714-08	EACH
25	SIGNAL HEAD ASSEMBLY (220)	730-02	EACH
26	CANTILEVER SIGNAL SUPPORT (1 ARM @ 30')	730-23	EACH
27	GEOTEXTILE (TYPE IV) (STABILIZATION)	730-40	EACH
28	GEOTEXTILE (TYPE III)(EROSION CONTROL)	740-10	S.Y.
29	TEMPORARY SEDIMENT TUBE 8IN (DESCRIPT)	740-11	L.F.
30	OH CONDUIT VARIOUS	790-40	L.F.
31	CONNECT TO 2IN EX. STL MAIN W/ STOPPLE	791-06	EACH
32	10IN X 6IN TAPPING SLEEVE AND VALVE	795-07	EACH
33	8IN GATE VALVE ASSEMBLY	795-08	EACH
34	SODDING (NEW SOD)	803-01	S.Y.

The selected bid line items pass the elimination process being “Standard" by:

1. maintaining the same description and unit of measure,

2. having a non-problematic unit of measure that is specific and measurable, and
3. belonging to a non-suspect category with consistent standard specifications that are not contract specific.

The selected bid items have all the qualities required by the NHCCI. The selected bid items observations have lagged values, have not changed in description or unit of measure, and have more than 8 quarters worth of data. The duration of 8 quarters (2 years) is used by the NHCCI to identify the length of time required for the bid item observations to show continuity and therefore to be included in the analysis. Therefore, the selected bid items observations show the continuity and statistical validity required to be included in the analysis.

After satisfying the previous qualifiers, the bid items observations are edited for outliers. Statistical edits recommended by the NHCCI are applied sequentially to the selected bid items. The outlier observations are set to the average change in logged price for non-outliers in the same period. This particular threshold represents the 95th percentile of pay-items. Thus, only 5 percent of all pay-items had observations outside of the selection threshold.

The FHWA, in the most recent revision of the NHCCI (Index 2.0), included a change to the percentile used to determine the outlier observations being outside of the 99.7th percentile range (changed from the 95th percentile). This research is using the 95th percentile range as the limit to determine outliers. By using the same state bidding system (e.g., TDOT), the data is expected to have less outliers outside the 95th percentile. Therefore, using the 95th percentile appears more adequate for this research on the state level. The NHCCI use of the 97th percentile, can be more appropriate on the national level as such percentile is less limiting to the amount of data included in the percentile envelope. In other words, more variations exist on the national level due to the use of multiple databases from multiple states. The use of the 95th percentile on the national level excludes more data as outliers from the NHCCI and causes the index to become non-representative of the national highway construction cost. However, such variations do not exist on the state level due to the consistency of using a single database. Thus, the 95th percentile is more appropriate to be used by this research on the state level.

Additional qualifiers to bid items data suggested by the NHCCI are applied (NHCCI 2020). Bid item observations having an R-squared value greater than 0.60 from a regression of the “log change in price” on the “log change in quantity” are eliminated. Bid items meeting this criterion represent a break in the price-quantity relationship required by an index.

More relevant data to the bidding process and the highway program are integrated into the bid item price estimation model. Such data include the number of contractors bidding and the number of highway projects being let. This data is used to analyze the effects of letting variables on bid item prices due to the competition between contractors at the time of the project letting. In addition, some construction material prices including steel and Portland cement concrete prices at the time of the letting are added to the database to measure their effect on the change in the Bid Item Prices. These prices of materials (e.g., steel and Portland cement concrete prices) are an indication of the current price of construction materials at the bid time and are necessary for the process when compared to other selected bid items.

Some initial analyses revealed that linear equation fitting has an upward trend. This is due to the long period of increase in fuel prices (increase in construction material prices from 2002 to 2011) followed by the decrease of such prices noticed over the next five years. Therefore, a shorter duration of five years is used for the analysis to provide a more realistic outcome and a better fit of the model to the data available.

After applying the recommended data edits and elimination criteria to all observations in the database, the following parameters describe the satisfactory and edited observations. The letting periods used are between 2/11/2011 and 4/1/2016 (i.e., 5 years), and all emergency lettings are eliminated. The selected 34 bid items have 6 units of measure from the initially available 40 units of measure and contain approximately 31,000 bid lines that remain from 936,000 bid items originally.

According to most common trading and economic practices in data transformation, the “change in log price” is used rather than the “change in price”. This practice equalizes the “percent change in price” rather than the “arithmetic change in price”. For example, a “change in price” from \$5 to \$10 has the same percent change as a \$10 to \$20 change even though the arithmetic change is doubled. This application is used in controlling outlier observations to the 95% confidence envelope and to neutralize the effect of inflation (Atkinson et al. 2009; Darby 1982). The year 2002 is used initially as the base year to adjust price observations and contract award amounts. Also, the Bureau of Labor Statistics CPI is used as the cost inflation rate. However, a different base year is used in indexing, and will be further explained later in the research.

The letting dates are used as the unit time period. This is necessary to document "observed bid item prices" as the "letting average bid item price." Also, the oil price used in the regression is "the market price at the date of the letting day." Time period consistency is maintained by transforming the frequency of price changes to the corresponding letting date. TDOT typically has a letting every 6 weeks for a total of 8 scheduled lettings per year, not counting emergency and unscheduled lettings. Unscheduled lettings can increase the number of lettings similar to the year 2011, which had a total of 14 lettings compared to the typically 8 scheduled lettings per year.

The weighted average of the bid item price for the letting is calculated based on quantities of the line bid items. The sums of the product of bid item prices and quantities are divided by the sums of quantities for the same letting to calculate the letting weighted average observed price for each of the 34 selected bid line items. For example, the weighted average price of the Bituminous Material for Tack Coat (bid item 403-01) in the February 11, 2002 letting is $68.34. Contract award amounts range from $50,000 to $126,000,000. The award amounts are scaled using $100,000 as the scaling unit. Contract duration is divided by 30 days to calculate projects duration in months.

Modeling Techniques

Following the data selection and edits, the research focuses on quantifying the factors affecting the change in the bid item prices. In other words, studying how the change in a set of variables affects the change of bid item prices. Using R Studio Programming software, MLR modeling is applied to the available data. The relationships between the prices of selected bid items and a set of independent variables are regressed for their effect on changing item prices in Tennessee highway projects. These independent variables are: letting date, project duration, project cost, number of projects in the letting, number of contractors bidding, item quantities, construction material prices, fuel prices, project type, material prices, funding source, and project location (urban versus rural).

The available bid line items data is unique and affects the selection of the modeling technique of choice. The number of observations is large enough to eliminate experimenting with modeling techniques similar to the studentized residual using smaller amounts of data. Also, the majority of the explanatory variables are scalable to a numerical range (e.g., project duration, project cost, number of projects let, number of bidders, item quantities, material prices, and fuel prices). Such scalability allows for continuity of the explanatory variables over the observed numerical range, and this differs from non-continuous relationships (i.e., step functions). Also, the inherent competition among contractors through the lowest bid letting method leads to a common consistency in item prices among different projects being let during the same time window. Unlike one-of-a-kind projects, the high number of projects let per year and a common specialty (e.g., Highway Construction and Maintenance) also contribute to consistency of item prices versus having multiple project types of different specialties such as residential, medical, water treatment, or others.

Scatter plots are used to identify common trends between bid item prices as dependent variables (observations) and the independent variables. To present the process, Figure 2 shows scatter plots of one of the selected bid items (Bituminous Material for Tack Coat) prices versus Oil Price, Contract Letting Date (Time), Project Award Amount in $100,000, Bid Item Quantity, Number of Projects per Letting, the Steel Producer Price Index, and the Cement Producer Price Index.

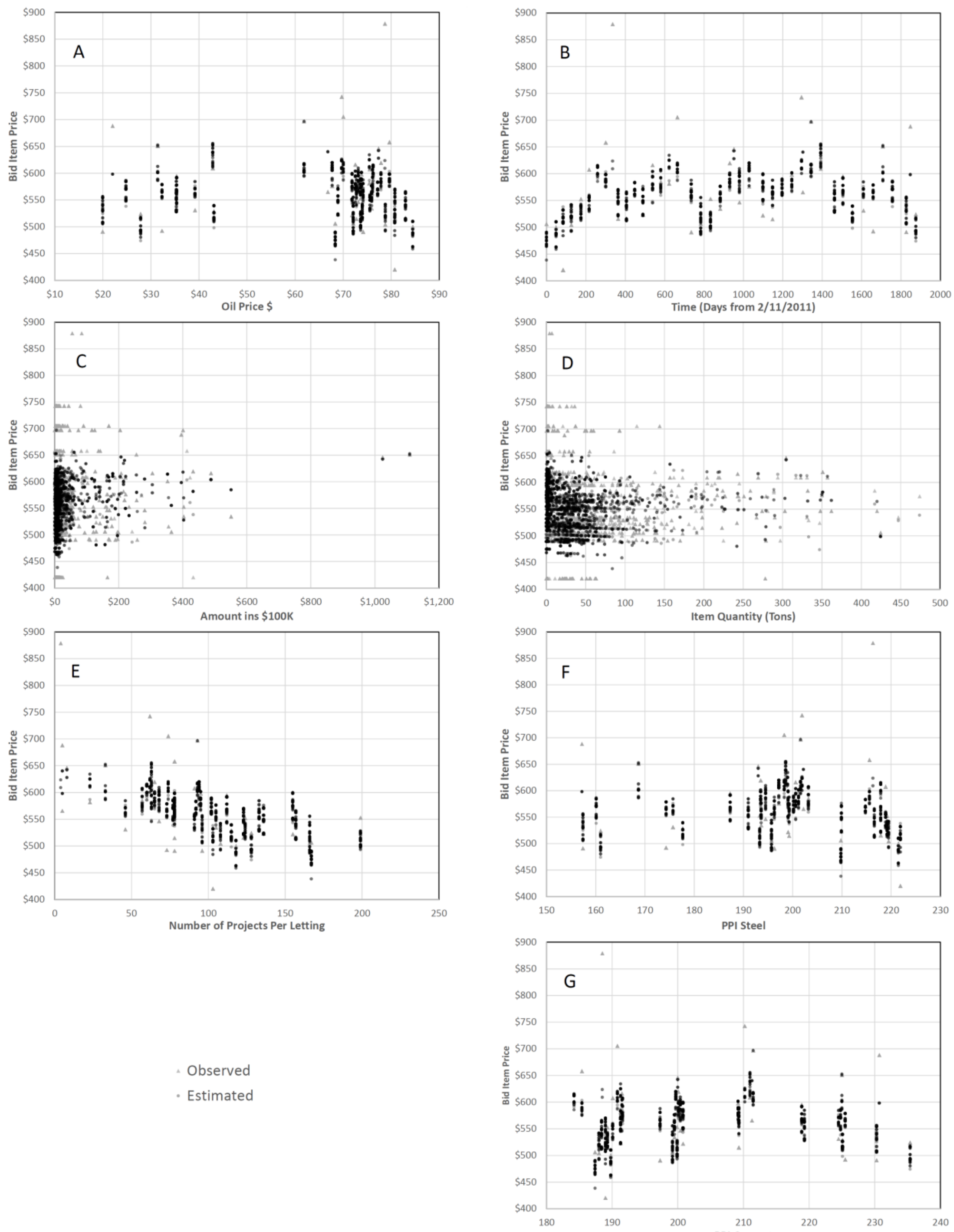

Figure 2 Scatter Plots (A-G) of Bid Item Prices versus Select Independent Variables

The scatter plots show some inconsistent spread with gaps of the data over the available data range similar to as in plots A, C, D, and F. This can be explained for plots A and F that have short time durations with low oil prices and low steel prices relative to the studied duration. This variation in prices is introducing gaps between low and high prices. However, plots C and D show more concentration of the data at the lower range of the independent variables. In other words, the majority of the projects have total award amounts of less than $1 million and total quantities of 100 tons or less of the sample bid item. Plot B shows some seasonal change of the price over time. Time series analysis is conducted and the results are discussed next.

Plots A, E, F, and G present the relationship between bid item prices and discrete variables associated with the letting date. Therefore, concentration of the price data around specific "discrete" values of the independent variables can be seen. As an example, plot F presents the bid item price relationship to the Producer Price Index (PPI) for cement. Note that the PPI Cement will have monthly discrete values produced by the Bureau of Labor Statistics. Therefore, one monthly value of the PPI Cement applies to all the bid item prices included in the same letting date, and causes the same PPI value to apply over a range of bid item prices. The same is correct for the Oil Price, Number of Projects in the letting, PPI Steel, and PPI Cement.

Plots E and G show better distribution of the bid item price over the range for independent variables. The relationships between the bid item price, the Number of Projects, and the Cement Producer Price Index (plots E and G) are indicating some nonlinear trends. However, the modeled linear relationships may provide significant representation of the relationship between the bid item price and the independent variables. Although linear relationships may provide a representation of trends, the correlations are not exactly linear in nature. Finally, there is a clear inverse relationship in plot E showing a lower bid item price relative to the increase in numbers of projects let.

In order to test bid items prices observations as time series, an Autoregressive Integrated Moving Average (ARIMA) model is used to detect patterns or seasonality in the observations (Hillmer and Tiao 1982). The steps of the Box Jenkins Method are used for analysis and selection of the ARIMA model through the identification, estimation, and the diagnostic checking steps. The 34 selected bid items series of quarterly prices are plotted against time as shown in Figure 3. Plotting the time series helps identifying characteristics of the data series to include: presence of data outliers, missing data values, series continuity, homoscedasticity, stationarity, and the presence of pronounced trends.

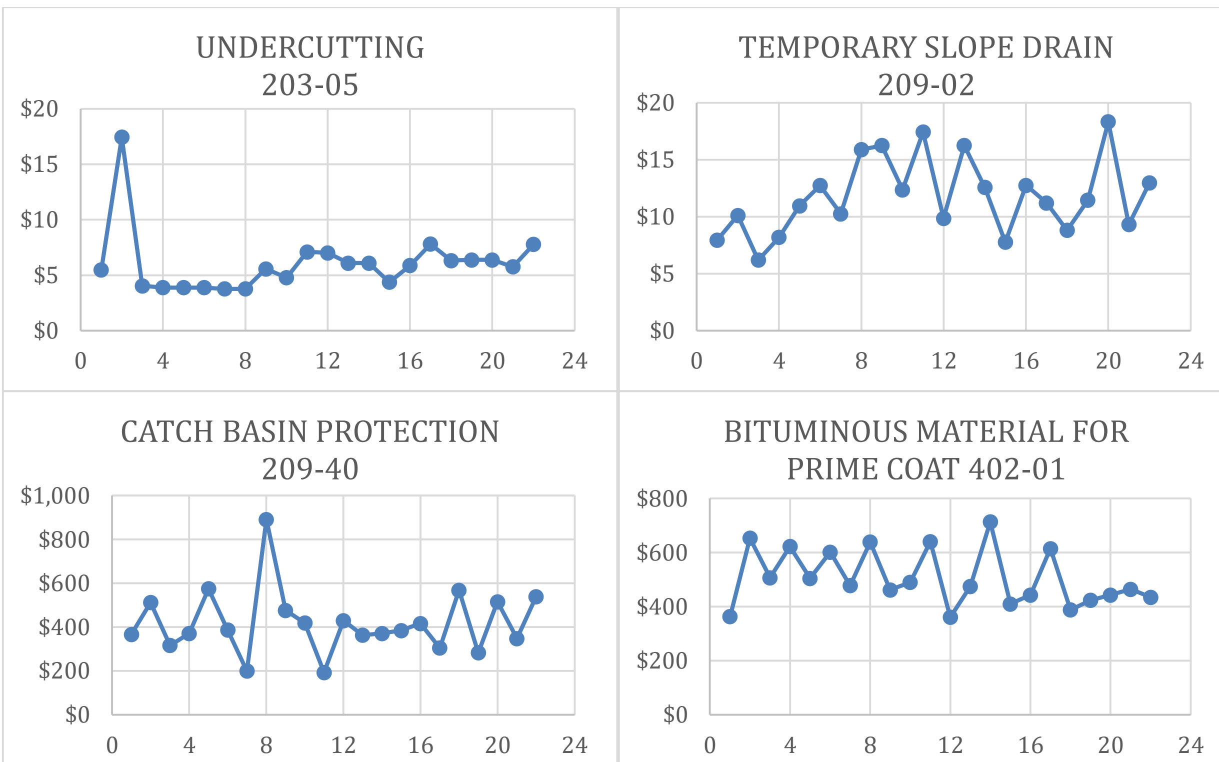

Figure 3 Bid Item Prices Quarterly Time Series

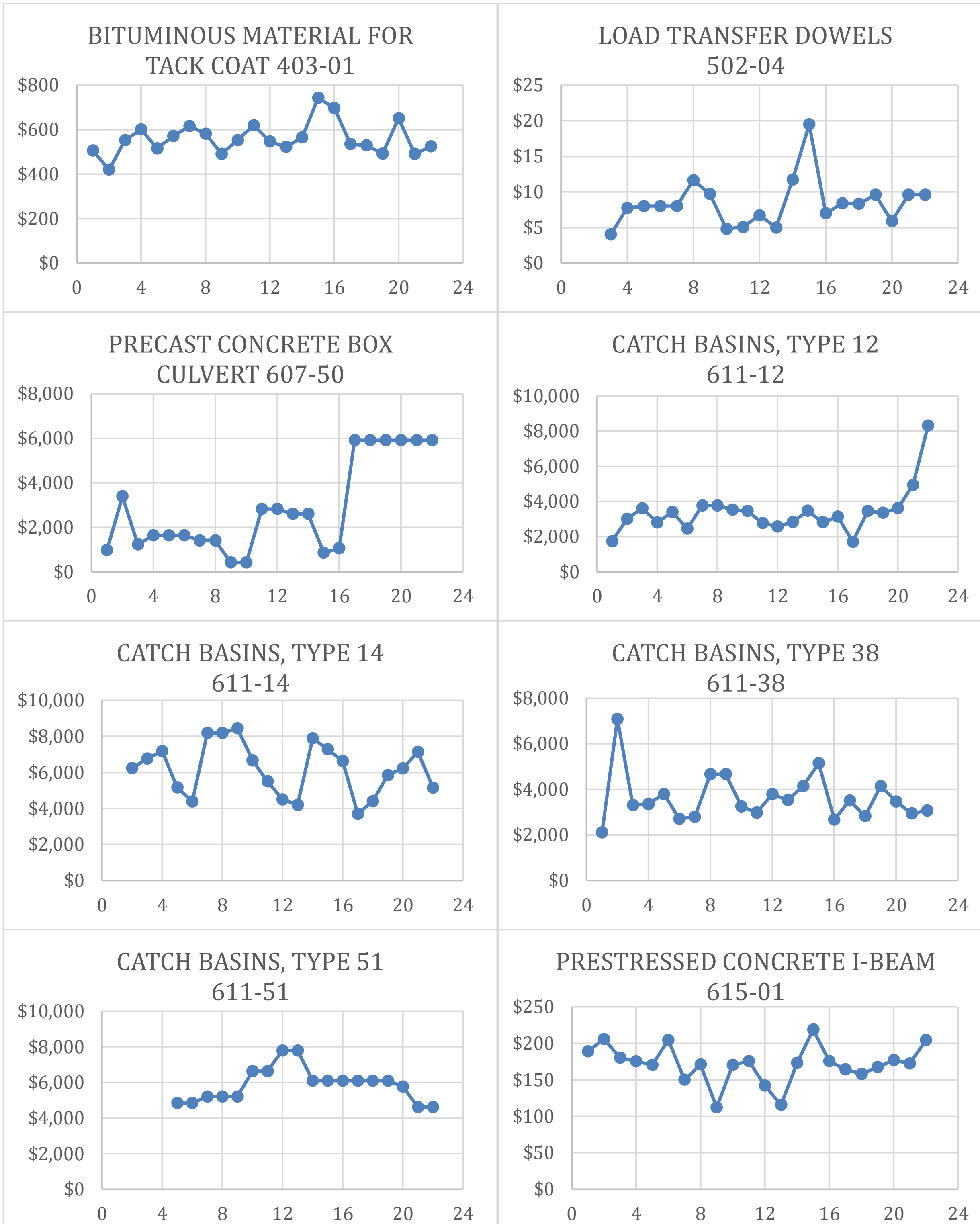

Figure 3 Bid Item Prices Quarterly Time Series (continued)

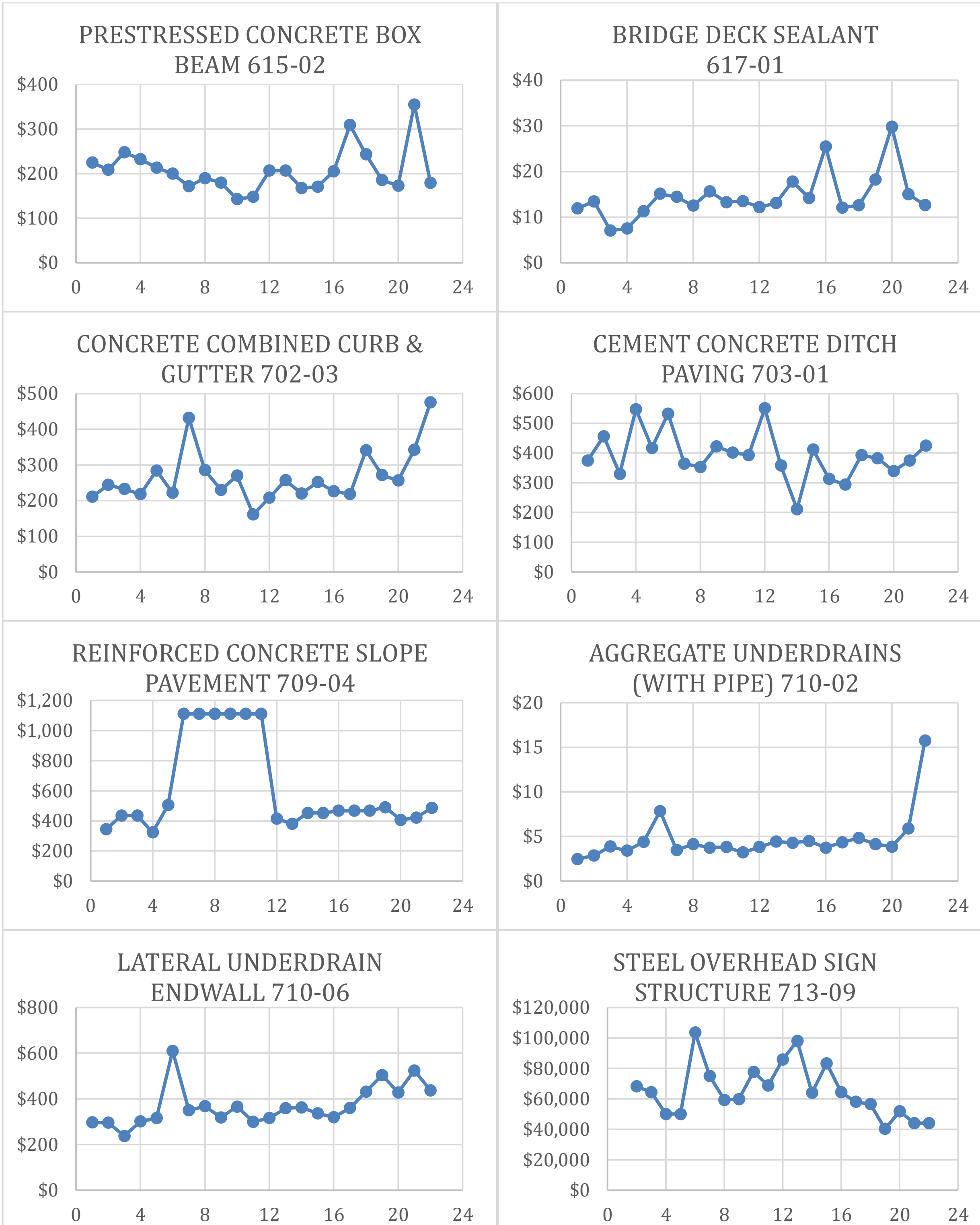

Figure 3 Bid Item Prices Quarterly Time Series (continued)

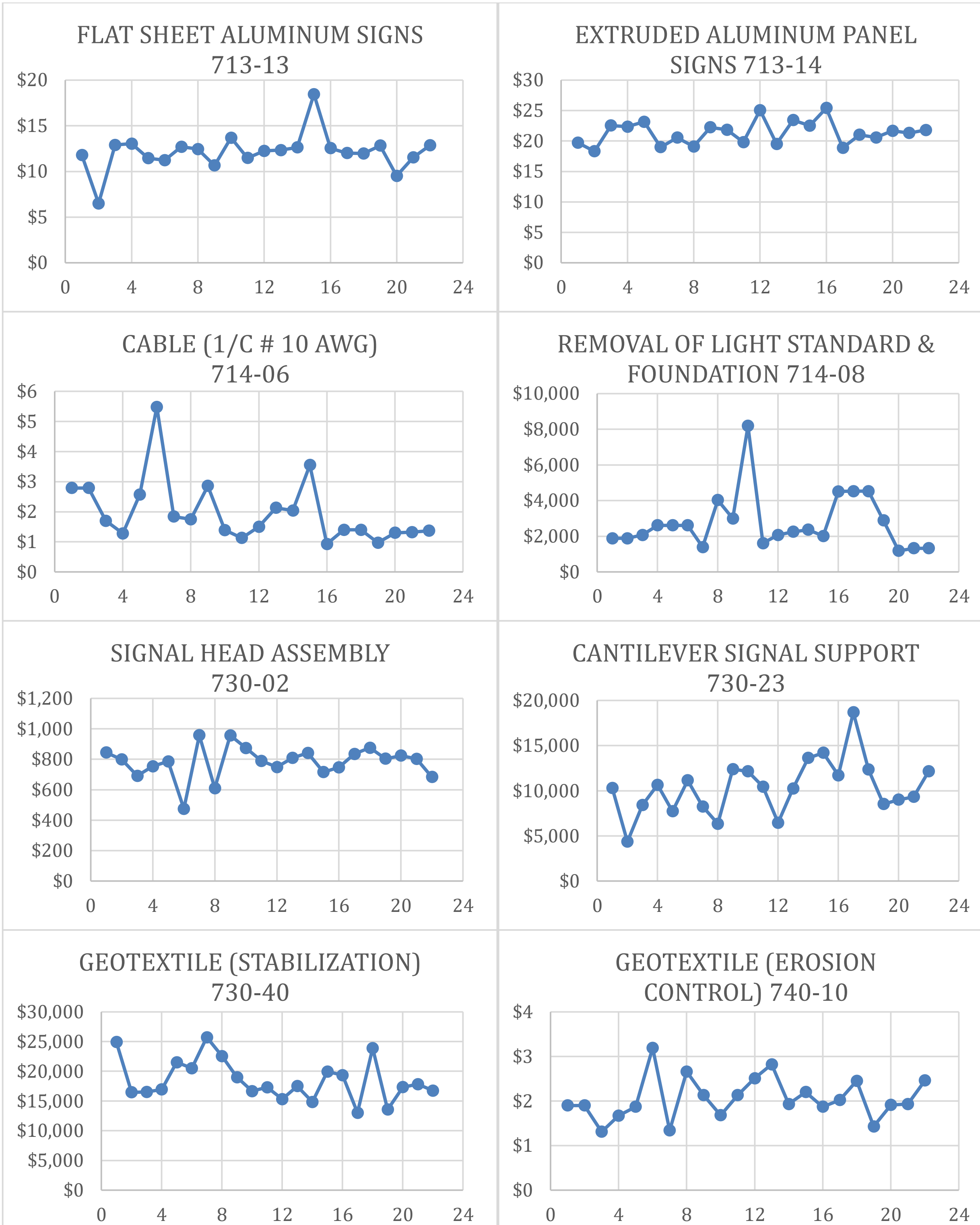

Figure 3 Bid Item Prices Quarterly Time Series (continued)

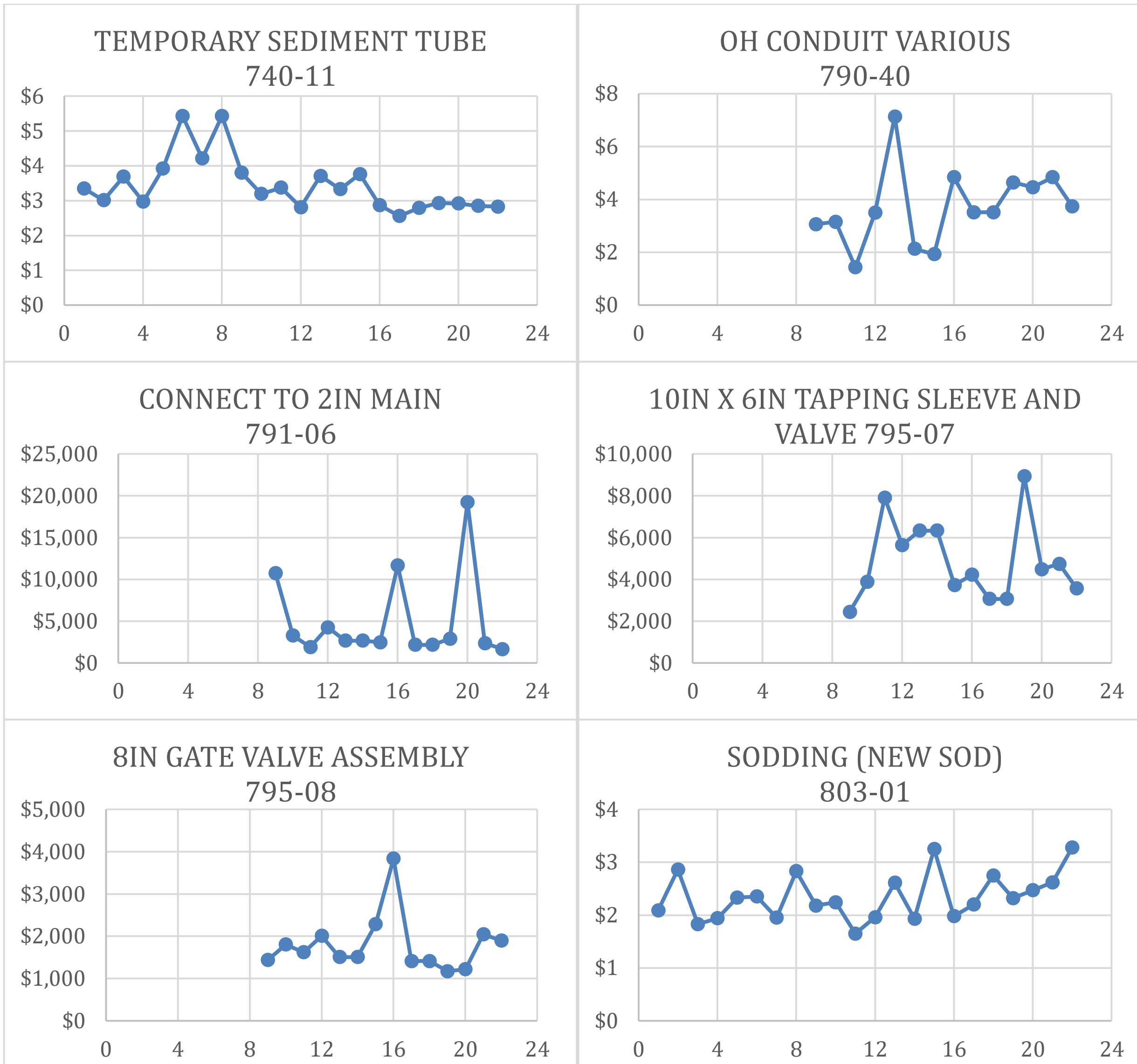

Figure 3 Bid Item Prices Quarterly Time Series (continued)

10 items (~30%) of the 34 selected bid item prices time series plots show outliers in the data similar to:

- UNDERCUTTING 203-05
- CATCH BASIN PROTECTION (TYPE D) 209-40
- LOAD TRANSFER DOWELS 502-04
- PRECAST CONCRETE BOX CULVERT (8' X 4') 607-50
- CATCH BASINS, TYPE 12, > 4' - 8' DEPTH 611-12
- CATCH BASINS, TYPE 38, 0' - 4' DEPTH 611-38

- REINFORCED CONCRETE SLOPE PAVEMENT 709-04
- AGGREGATE UNDERDRAINS (WITH PIPE) 710-02
- CABLE (1/C # 10 AWG) 714-06
- REMOVAL OF LIGHT STANDARD & FOUNDATION 714-08

Outliers reduce the stationarity of the items data, and are likely to result in lower test results. Also, 4 items (~12%) of the 34 selected bid item prices time series plots show missing values for the first 8 periods. This is an indication that the bid item numbers with the missing values were not used before the indicated time period. However, continuity of the available data may qualify the time series to be used with the ARIMA model. All series are continuous without missing intermediary or breaks in the data. Almost all the series can be described as stationary and have no pronounced trends or changing variance. A time series is considered stationary if the series has statistical equilibrium by changing uniformly about a fixed level.

The Dicky Fuller (DF) and the Augmented Dicky Fuller (ADF) tests examine the stationarity of the time series by testing the change of the mean and the standard deviation. Thus, the two tests can be used to compare the time series stationarity between the original data and the transformed data. By applying the Dicky Fuller tests, all bid item prices time series show better stationarity using transformed data by differencing than the original time series. The differenced data show better (i.e., higher) DF test results for all the 34 bid items. The differenced data show the same or better (i.e., lower) P-value for all 34 bid items. The time series for 12 bid item prices have P-values of 0.01 for both the original and the differenced data. A P-value of 0.01 is the minimum limit for the test being applied using the R Programming Language. Thus the results may actually be better than the minimum limit.

By applying the ADF test, nearly all of the bid item prices time series data show better stationarity through transformation by differencing than the original time series. The differenced data show better (i.e., higher) ADF test results for 32 out of total 34 bid items. The differenced data show better (i.e., lower) P-value for 31 out of the total 34 bid items indicating better stationarity using the differenced data. Therefore, the initial testing using the Dicky Fuller and the Augmented Dicky Fuller tests indicate that the ARIMA Analysis is likely to produce a better model using the transformed data by differencing.

The Autocorrelation Function (ACF) examines the correlation of a time series against itself but with lagged time periods. A positive value of the ACF is an indication of a direct relationship between the original and the lagged time series and the opposite is also true. A negative value of the ACF is an indication of an inverse relationship between for the original and the lagged time series. By examining the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF), the differenced data show better results than the original data for most of the selected bid items data at the first lagged value of the time series. By examining the Bituminous Material for Tack Coat plotted results, as an example, both the original and the transformed series show rapid decay to within the 95% confidence level as shown in Figure 4. This result of the ACF cutting off right after a very strong first lag is an indication of the likelihood to have an ARIMA model with better results using the Autoregressive part of the model to the first order.

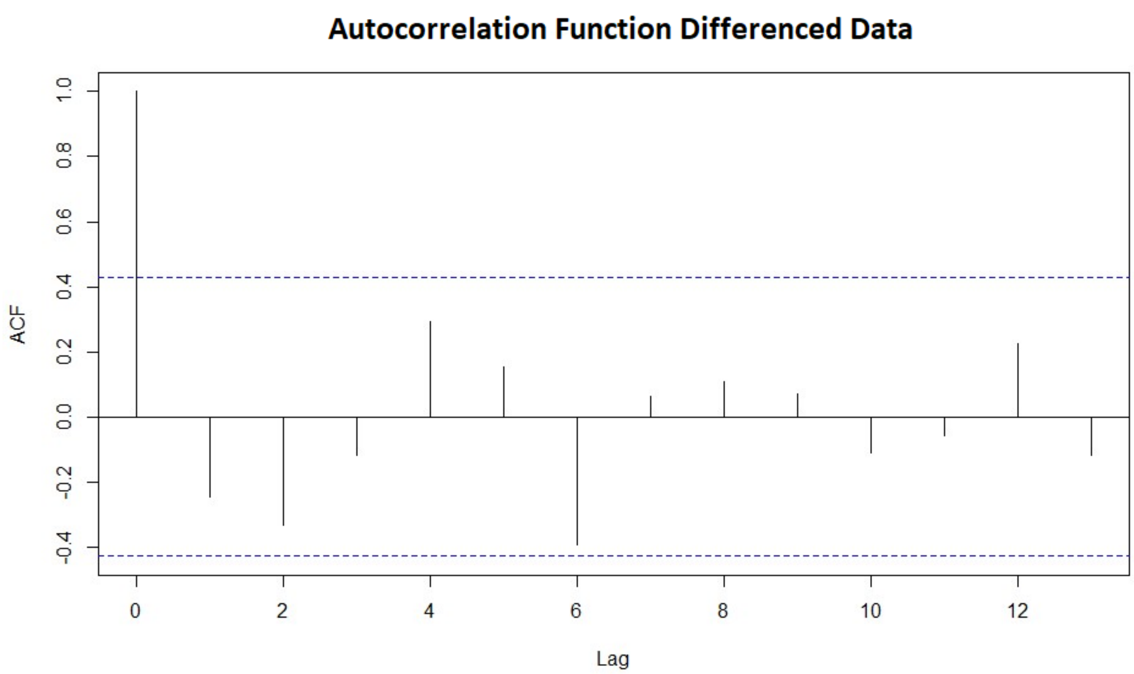

Figure 4 Bituminous Material for Tack Coat Autocorrelation Function (Differenced Data)

Using the R programming language and the ARIMA model, multiple models are tested by examining various coefficients for significance to select good fitting models. The best ARIMA models are selected by examining the goodness of fit of multiple models using the Akaike Information Criterion (AIC) as a selection parameter. Results are discussed more fully in the forthcoming "Results" section later.

Indexing

The third step in the research methodology is to index the item prices to calculate a local index for Tennessee highway projects. The research produces an index using the selected 34 bid items identified in the "Methodology" section.

At the time of this research, multiple states are using indexing techniques to adjust their respective excise fuel taxes. These fuel tax indices are either based on adjusting for inflation or global indices similar to the NHCCI. Nebraska is the only state with an index accounting for the change in gas prices and the transportation needs based on legislative decisions. Nebraska has a biannual adjustment using a variable tax that is set by the Department of Transportation Director to meet the highway funding appropriations made by the Nebraska Legislature (Nebraska 2018).

When using the NHCCI as the only index for the state, the process neglects local factors that may affect the change in highway construction and maintenance costs of projects. By using the previously mentioned independent variables in the estimation model development process, the resulting model and index for Tennessee would be more sensitive to local factors and global ones. More details on this will be explained in the forthcoming "Results" section.

This research uses the Fisher Ideal Index formula (Equation 1) for indexing the sample bid line item price. The Fisher index is generally believed to be a better approximation, superlative, simple, and easy to use as described in the FHWA report (NHCCI 2020). The Fisher Ideal Index uses a geometric average of Laspeyres Index and Paasche Index to dilute certain disadvantages or biases of these two techniques (NHCCI 2020; Wallace 1996). The Fisher Ideal Index takes the weights of the base period ($q_{j,0}$) and the current period ($q_{j,t}$) into account with calculations to accommodate for the substitution effect which differs from the Laspeyres and the Paasche techniques (NHCCI 2020; Wallace 1996). Laspeyres and Paasche techniques adjust to either the weights of the base period or the current periods. The Fisher Price Index equation is used to calculate the Tennessee Highway Construction Cost Index (TNHCCI). Values will be presented and compared to the NHCCI later in the "Results" section.

$$F(p) = \sqrt{\frac{\sum_{j=1}^{N} p_{j,t} q_{j,0}}{\sum_{j=1}^{N} p_{j,0} q_{j,0}} \times \frac{\sum_{j=1}^{N} p_{j,t} q_{j,t}}{\sum_{j=1}^{N} p_{j,0} q_{j,t}}} = \sqrt{Laspeyers\ Index \times Paasche\ Index} \qquad (1)$$

$F(p)$	=	Fisher Index,
j	=	The bid item number ranging from 1 to N,
N	=	Total number of bid items to be indexed (i.e., 34),
$p_{j,t}$	=	The price of the item j in the period t,
$p_{j,0}$	=	The price of the item j in reference to the base period 0,
q_j	=	Represents the quantity of the item j in period t or base period 0.

The base period is then considered at letting period 0 that is the letting of 2/11/2011. The NHCCI is then transformed using the same base period equating the base period to an index of 1 and calculating all consecutive periods respectively.

$$(Transformed)NHCCI_t = \frac{NHCCI_t}{NHCCI_0}$$

$(Transformed)\ NHCCI_t$	= The transformed NHCCI value for the period t,
$NHCCI_t$	= The NHCCI value for period t,
t	= The letting period ranging from 0 to 42 (2/11/2011 to 4/1/2016),
$NHCCI_0$	= The NHCCI value for period 0,

The same index transformation equation is used to calculate the annual values for the Tennessee Highway Cost Index. These annual values of the index are used to calculate the change in the gas tax to present an annual value instead of calculating the change in the tax based on the letting period. The annual index is subtler with less volatility when compared to the letting period index. Calculated values are presented in the "Results" section.

Figure 5 shows the Tennessee Highway Construction Cost Index (TNHCCI) relative to the NHCCI and is calculated using estimated bid item prices. After applying the developed model to calculate the bid item prices, the bid item data for the study period (2011-2016) are indexed and plotted to show the "TNHCCI" change over time relative to the NHCCI. The left vertical (y) axis has the index values using the base period of 1 at 2/11/2011, as previously discussed. The TNHCI 4 periods moving average shows the index with a smoothed out trend line. When the indices are applied to the gas tax values starting at the base period, the resulting gas tax values follow the indices change and values are shown on the right vertical axis starting with 20 cents at 2/11/2011. A gas tax of 20 cents was the gas tax at the time before the recent gas tax increases starting in 2017. Results are presented with details in the forthcoming "Results" section.

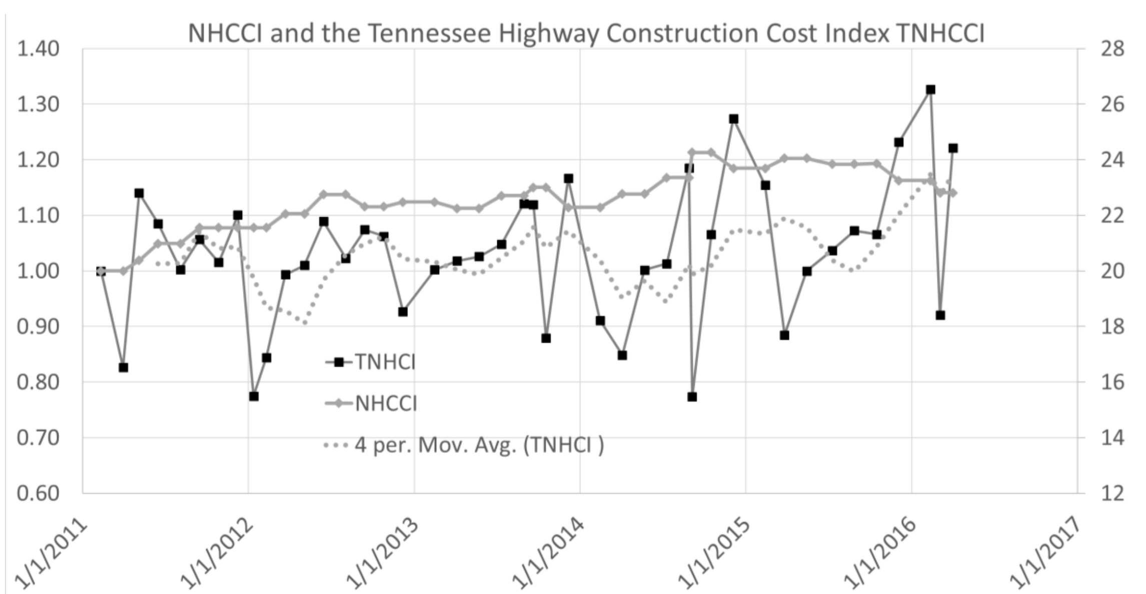

Figure 5 NHCCI, TNHCCI, and 4 Periods Moving Average (Jan 2011 - April 2016)

RESULTS

Significant Variables

The data for the 34 selected bid items were evaluated using a correlation matrix. A total of 26 independent variables were then tested for multi-collinearity. Some variables showed consistent multi-collinearity with most of the bid items (e.g., project cost and duration). Such variables are noted and then were examined for their multi-collinearity effect on other variables during the modeling process. Variables with high multi-collinearity were removed using a stepwise process with priority given to removing variables with the most disruption to the remaining independent variables. A stepwise process was applied until such effects were neutralized. The P-value was used in the MLR model development as an indicator to gauge the disruption of one independent variable to other independent variables in the model. The P-value is a test of significance to the null hypothesis that the variable being tested is not contributing significantly to improving the explanatory or predictive power of the model. When the P-value is larger than 0.05, the null hypothesis cannot be rejected and thus must conclude that the variable is not significantly contributing to the model explanatory power. Therefore, variables with P-values higher than 0.05 were considered redundant and were removed from the model. Using the P-value for stepwise selection identified the variables of significance and multi-collinearity. However, the process provided a model with a low coefficient of correlation (0.084). Thus a different process must be used in the estimation and model development.

Model Development

To explain the model development process and results, the bid item price (i.e., dependent variable) now has 26 explanatory or independent variables. The Multiple Linear Regression model was used to predict the best fitting linear relationship between the bid item price and the independent variables. To avoid model overfitting, the modeling process must achieve the best fitting linear relationship with the least number of independent variables having the least collinearity. Therefore, a stepwise selection process was used in the model development where each of the explanatory variables is added or removed based on a selection criterion checked at each step. The Akaike Information Criterion (AIC) was used as the criterion of choice in the stepwise selection process. The use of AIC mitigates the issue of inflation of the false positive findings commonly associated with using the P-value. When the stepwise selection process was tested using the P-value as the selection criteria, the process provided a model with a low coefficient of correlation (0.084), as mentioned before. The stepwise selection process using the P-value encourages the addition of variables leading to an over selection of variables and false positive results. However, the AIC criterion, when used in the stepwise selection process, optimized the use of linear regression by regulating the addition of variables into the model while penalizing the indiscriminating increase of the number of variables at the same time. Therefore, lower AIC values are better for the model (Akaike 1981).

$$\text{AIC} = 2k - 2\ln(L) \tag{2}$$

k = the number of variables in the model

L = the maximum value of the likelihood function of the model.

As a result, the independent variables and correlation coefficients identified in Table 2 complete the final MLR model for the modeled bid items.

Equation (3) represents the MLR model with the referenced variables from Table 2.

$$y = \beta_0 + \beta_1\, iv_1 + \beta_2\, iv_2 + \beta_3\, iv_3 + \ldots + \beta_n\, iv_n + \epsilon \tag{3}$$

y = the dependent variable (Bid Item Price)

i	=	independent variables identifier (from 1 to n = 26)
ivi	=	independent variable
β_0	=	y-intercept (constant)
β_i	=	coefficients for each independent variable
ϵ	=	the model error (residuals)

ITEM	(Intercept) B_0	Oil (B_1)	LET (B_2)	AM100K (B_3)	Qty (B_4)	Bids (B_5)	DUR (B_6)	Projs (B_7)	GRDG (B_8)	IMPS (B_9)	BRCN (B_{10})	ASPH (B_{11})	SAFE (B_{12})	BRRP (B_{13})	EMER (B_{14})	GEN (B_{15})	ITS (B_{16})	SIGN (B_{17})	CONC (B_{18})	SRFT (B_{19})	MTNC (B_{20})	PVMK	INCI (B_{22})	FF (B_{23})	PPISt (B_{24})	PPICem (B_{25})	Urban (B_{26})
203-05	-17.79	0.08	0.00		0.00	-0.27		-0.31			-1.50														0.11		
209-02	57.71		0.00	0.00	0.00	-0.04	0.03				1.16		2.62	4.81										-0.62	-0.09	-0.16	0.96
209-40	-416.97	2.29	-0.09	0.05	-0.53	1.08	-1.74	-0.48			11.49	39.52		37.57	41.13	-267.78									-1.04	5.04	
402-01	991.96	-0.87	0.12	0.09	-0.04	6.02	-0.74	-0.99	150.38	141.05	160.28	168.50	197.82	169.32		253.91									2.28	-5.37	
403-01	1515.06	-0.78	0.19	0.07	-0.07	1.01	-0.37	-0.45			13.41	-7.31	22.37	19.35	36.71			97.02		-32.62				4.66	1.39	-6.45	
502-04	-59.25	0.08	0.00	0.00		-0.16		0.03	-0.66					3.74					1.59						0.04	0.28	
607-50	-23642	-234.52	4.46	-132.51	-3.97	-1534	1067	79.30	-3636	13446	3738													4094	173.04		
611-12	-11720	8.99	-2.04	-0.45		8.26	-11.88	-2.21	1999	1923	2391		1426			1588		1789						389.95	-25.17	94.32	188.94
611-14	-13844	25.79	-4.45	2.37			-50.04		1479	1644	1203		1282												-35.49	144.41	-588.56
611-38	1184	21.94	2.83	1.60	-26.84	-18.41			1222	1373	1022	2095		1037.4											40.73	-51.08	-639.60
611-51	-96026	603.08	-23.47	4.62	0.00	102.95	47.92	-249.31	2243																-561.46	942.52	
615-01	-290.26		-0.05		0.00		-0.62																	-5.39	0.24	2.37	
615-02	1.80	-2.04	-0.02	0.06	0.00	-1.96		-0.85		18.83	14.36														1.79		-5.83
617-01	-9.03	0.05		0.00	0.00			-0.11		-2.85	-2.65														-0.04	0.14	
702-03	515.75		-0.05		-0.03		-0.45	-2.54	50.30	29.98	42.43	21.62		75.69		-55.00								14.81	-2.30	1.19	8.13
703-01	1763.17			0.07	-0.07	-8.18	-1.67	-3.34	-20.93	-63.21		-150.99	-56.29	-51.25	-116.60									-28.88	-1.86	-4.22	
709-04	-393.85			0.22	-0.88	36.78	-6.10	-14.80			111.64													-227.38	5.57		159.45
710-02	-9.97		0.00	0.00			-0.03	-0.04	0.74	0.96	1.73		2.75											0.38		0.07	0.22
710-06	-37.14	-0.43	-0.05					-0.71	31.81	73.06	55.68		173.31			24.76								-26.31	-0.56	2.85	26.53
713-09	921415	-219.96	110.06	12.45		-270.81	237.95		-12897	-6023	-31995	-23550					35651							24749	90.13	-4863	17365
713-13	8.43	0.03	0.00	0.00		0.02	-0.02	0.00	1.51	1.35	1.24	1.56	1.40	1.03					2.07	3.48	1.50			0.42	0.02	-0.03	
713-14	-0.17	0.02	0.00	0.01	0.00	0.13		-0.07	-1.37	-2.01	-2.36		-1.76			-4.11			-2.37		1.71			1.34	0.10		
714-06	8.39			0.00		0.01	-0.03	-0.02					-1.06	0.33		-1.02	1.91		0.91					0.78	-0.01	-0.02	-0.14
714-08	-18112	85.32		-1.37	-3.88	-93.53	66.47	52.61	-1542	-2890	-2424		-2356			-1753								-1170	-15.28	91.48	1164
730-02	1585	-0.56	-0.11	-0.06			0.96	0.76	-43.90	-41.87	-23.19		-57.78					-93.24	-348.75					17.63	-3.42		6.60
730-23	13154		0.63		-88.97		-35.62	15.24	3085	2349	2771	3686	4140					3146						742.08	-34.04		-746.98
730-40	39417		-3.79			-324.50		-148.08	-1466			-4107												-604.80	-64.31		-886.22
740-10	-0.37	0.01	0.00	0.00	0.00	0.01	-0.01	0.00		-0.06													0.33		-0.01	0.02	
740-11	2.64	0.00	0.00	0.00	0.00		0.01	-0.01	-0.27		-0.17	-0.54	-0.14				-0.15		-0.54					-0.14	0.01		0.10
790-40	0.80		0.00	0.00			-0.05	0.02	0.84	0.86	1.19													0.46			
791-06	183484	-102.78	21.31			180.55	-220.76		2823		8781													-2770	-147.95	-799.27	
795-07	-45198	186.93				54.93	27.08	52.05	-2885	-2043	-1859														-84.49	263.58	
795-08	-6229	16.36	2.31			17.22	-34.05		-758.67	-554.09	-1054													742.55	25.68		232.77
803-01	-5.09	0.00	0.00		0.00		-0.01	-0.01	-0.12	-0.18	-0.11	-0.24	-0.13			-0.68	0.15	-0.30	-0.34						0.01	0.04	-0.03

Table 2 - Selected Bid Items Analyzed Using MLR Models

Time Series Analysis

As mentioned in the methodology section, the Box Jenkins method was used for analysis and selection of the ARIMA model. The identification step of the method shows mostly stationary time series without a pronounced trend that necessitated data transformation. However, the estimation step showed better results (i.e., higher test values) using the Dicky Fuller (DF) and the Augmented Dicky Fuller (ADF) tests and lower P-values for the differenced (transformed) data over the original data for all the selected bid items. The following results present a sample of the 34 bid items selected. The DF and the ADF tests for item 403-01 “Bituminous Material for Tack Coat” is summarized in Table 3.

Table 3 Dicky Fuller and Augmented Dicky Fuller Test Results

Test	Dicky Fuller		Augmented Dicky Fuller	
Parameter	Value	P-value	Value	P-value
Original Data	-3.53	0.06	-2.74	0.29
Differenced Data	-5.62	0.01	-4.93	0.01

The P-value (0.06) of the DF tests is slightly outside the 95% confidence level for the original data series. The P-value of the differenced data is in the 99% confidence level, and thus is a better result. These results indicate that an ARIMA model is likely to have a better fit using the differenced data. By examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), the ACF cuts off after the first lag to within the 95% confidence level envelop using the original data and the differenced data, as shown previously in Figure 3. This is an indication that the ARIMA model should include a Moving Average (MA) component. Using the PACF did not provide a significant change from the ACF result by showing both original and differenced series within the 95% confidence level.

In the Estimation step of the Box Jenkins method, various models were tested using R programming language and the ARIMA model. By using the AIC as the selection parameter, the differenced data model provided better results for 31 (91%) of the 34 selected bid items. Also, 28 (82%) of the 34 bid items showed similar or better results by using differenced data and having a Moving Average (MA) component of the first degree. In most of the bid items, an ARIMA model using the differenced data and including an MA component proved to be the best ARIMA model in 23 (67%) of the 34 selected bid items. These results match the earlier prediction from the Identification step and Table 4 summarizes the results of tested model and the Estimation step.

The resulting ARIMA model is (0,1,1). This means that zero corresponds to the model having no Autoregressive (AR) component. The two ones represent that the model is integrated to the first degree by differencing, and has a Moving Average (MA) component to the first lag value. The ARIMA model (0,1,1), when compared to other models, has the lowest AIC value in 23 (67%) of the 34 bid items. Also, the model is considered the simplest (parsimonious) model including an MA component matching the initial prediction of the Identification step.

Table 4 ARIMA Models Estimation Step Test Results (p,d,q)

ARIMA	(1,0,0)	(2,0,0)	(0,0,1)	(1,0,1)	(1,1,0)	(0,1,1)	(1,1,1)	(1,1,2)	(1,1,3)	(2,1,3)
203-05	113.7	115.6	113.7		114.9	112.8	114.8	116.8	118.7	120.1
209-02	120.6	121.6	120.6	122.1	120.5	117.7	118.9	118.1	116.7	118.4
209-40	284.9	286.3	283.9	284.4	286.8	278.8	278.8	279.6	282.3	281.1
402-01	269.2	270.9	269.5	271.1	270.6	261.6	257.9	254.8	257.3	255.6
403-01	255.5	254.6	254.7	257.3	254.2	248.3	249.3	249.7	250.8	249.0
502-04	140.5	142.0	141.0	142.2	135.7	135.0	135.8	139.0	138.7	140.5
607-50	387.0	389.0	391.7	389.0	370.3	369.6	372.2	370.3	371.8	374.1
611-12	377.0	379.0	379.7	378.9	357.9	357.9	359.3	361.5	363.0	365.0
611-14	147.0	144.4	145.6	146.2	149.1	149.1	145.9	145.4	144.5	145.0
611-38	372.5	373.1	369.7	369.8	368.9	361.0	361.1	359.8	361.8	363.1
611-51	75.5	76.9	84.3	77.1	72.1	72.2	73.9	74.6	74.1	73.4
615-01	210.4	211.9	210.3	212.0	210.1	205.6	206.8	208.7	209.1	208.6
615-02	239.0	240.6	238.9	240.9	238.0	232.2	234.0	235.9	237.9	233.7
617-01	137.4	139.4	137.4	139.4	140.0	130.2	132.1	134.2	128.4	126.8
702-03	255.8	257.7	255.8	257.6	247.2	245.7	247.7	248.7	250.5	252.4
703-01	260.4	262.3	260.4	262.4	258.0	250.8	252.3	253.1	255.1	254.8
709-04	301.2	301.8	304.6	301.7	289.8	289.7	290.5	292.5	294.4	295.0
710-02	109.1	110.7	109.0	110.7	103.8	103.8	105.8	106.8	108.8	110.8
710-06	261.8	263.1	262.6	263.5	253.1	250.5	251.8	253.6	253.7	254.6
713-09	141.2	142.5	142.3	142.5	138.1	137.1	139.1	140.8	139.3	141.3
713-13	99.1	100.9	99.1	100.2	104.4	97.9	99.9	101.8	103.7	102.9
713-14	95.4	97.2	95.5	96.5	99.2	95.6	96.7	97.7	99.1	-0.8
714-06	69.1	70.6	68.9	69.8	75.6	66.2	68.2	68.7	67.1	64.9
714-08	391.7	393.5	391.7	393.6	380.8	378.5	379.9	381.8	382.3	384.3
730-02	269.6	271.4	270.5	271.5	267.9	264.9	263.4	265.4	267.0	268.5
730-23	419.4	421.0	418.9	420.6	408.7	402.5	403.9	405.0	405.8	406.7
730-40	426.2	427.4	426.2	428.1	412.9	408.4	410.4	412.2	410.8	410.9
740-10	33.7	35.7	33.7	34.2	43.0	36.9	38.3	40.2	42.2	44.0
740-11	50.6	51.2	52.9	51.9	50.6	51.1	52.6	52.9	49.0	50.7
790-40	122.3	124.0	116.7	118.3	118.6	106.8	108.5	105.4	101.4	99.4
791-06	123.6	120.5	127.3	123.2	114.9	117.3	116.7	118.5	117.6	119.6
795-07	119.9	121.8	123.6	121.8	117.3	117.3	119.3	119.0	118.8	119.3
795-08	118.3	119.8	117.6	118.9	116.5	116.2	114.2	114.4	116.4	118.2
803-01	31.0	32.9	31.1	32.9	37.6	31.7	32.3	34.0	34.8	35.6

p = Auto Regressive (AR) component

d = Integration by differencing degree

q = Moving Average (MA) component

DISCUSSION

The correlation coefficients between the weighted average letting bid item prices and the chosen numerical independent variables were modeled, as mentioned before, for each bid item. The results of these analyses were presented earlier in Table 2, and indicate the statistically significant linear relationships between bid item prices and the earlier listed independent variables that were identified and further studied.

From the table, notice that there are 34 bid line items and 26 independent variables. Each of the modeled bid line items follow Equation 2 as mentioned earlier with a minimum of 5 independent variables for bid item (Prestressed Concrete I-Beam 615-01 and Undercutting 203-05) up to 17 independent variables for bid item (Bituminous Material For Tack Coat 403-01). Independent variables are counted as significant if the absolute rounded up value of the coefficients is more than 0.01. This is an indication that bid line items vary according to the number of influencing studied independent variables. The bid item identified as being the most independent item is Bituminous Material for Tack Coat and was the sample bid item used to explain several parts of this research.

None of the independent variables are included in all of the bid line items. The Project type of Pavement Marking coefficient (PVMK B_{21}) is not included in any of the bid line items while the Producer Price Index for Steel is included in 32 (~94%) of the 34 bid line items. Only 24 of the approximately 25,000 bid line items used in training the models have Pavement Marking as the project type. This is the second lowest count of independent variables data after Project Type Incidental (INCI B_{22}) with 23 bid line items. The independent variable coefficient INCI B_{22} is included in one bid item model for Geotextile (Type III) (Erosion Control) (740-10). INCI B_{22} is the highest coefficient and is only second to the intercept value in the bid line item model for item 740-10. This is also an indication that the independent variable (iv_{22}) Project Type Incidental is the highest variable with elasticity of the bid item Geotextile (Erosion Control) 740-10.

To illustrate the modeled variable coefficients and explain the developed models in more detail, the Bituminous Material for Tack Coat bid item model is used as an example. Looking at the independent variables, and particularly project types, (iv_{11}) Traffic Signal projects contribute to the increase in the sample bid item price relative to other project types. This can be explained by the typically small quantities of bitumen and asphalt pavement materials in (iv_{17}) Traffic Signal projects. In such projects, the typical fixed costs of hauling and placement of the material is allocated to smaller quantities and therefore are contributing to the increase in the bid item price. The opposite is true for (iv_8) Grading, (iv_9) Integrated Management Practices, (iv_{11}) Asphalt Paving, and (iv_{19}) Resurfacing project types. Such projects have large amounts of bitumen and asphalt pavement materials and are showing an effect of lowering bid item price values. This is due to the large quantities and the allocation of previously mentioned costs over more tons of materials resulting in a lower bid price. Thus, to achieve lower prices for "Bituminous Material for Tack Coat", Traffic Signal Projects are better included in the other identified project types. The same explanation is linked to (iv_4) Bid Item Quantity. The increase in Bid Item Quantity values shows a reduction in the bid item price values. However, the coefficient is very small requiring high change in quantity before causing a significant change in the sample Bid Item Price.

The bid item price is showing a subtle increase over time. This increase is explained by a relatively small positive coefficient for the (iv_2) Contract Letting Date. This increase can be linked to inflation and the increase in oil prices overtime. This conforms to the results of the ARIMA time series analysis as shown later in the Time Series Analysis Results (Figure 6).

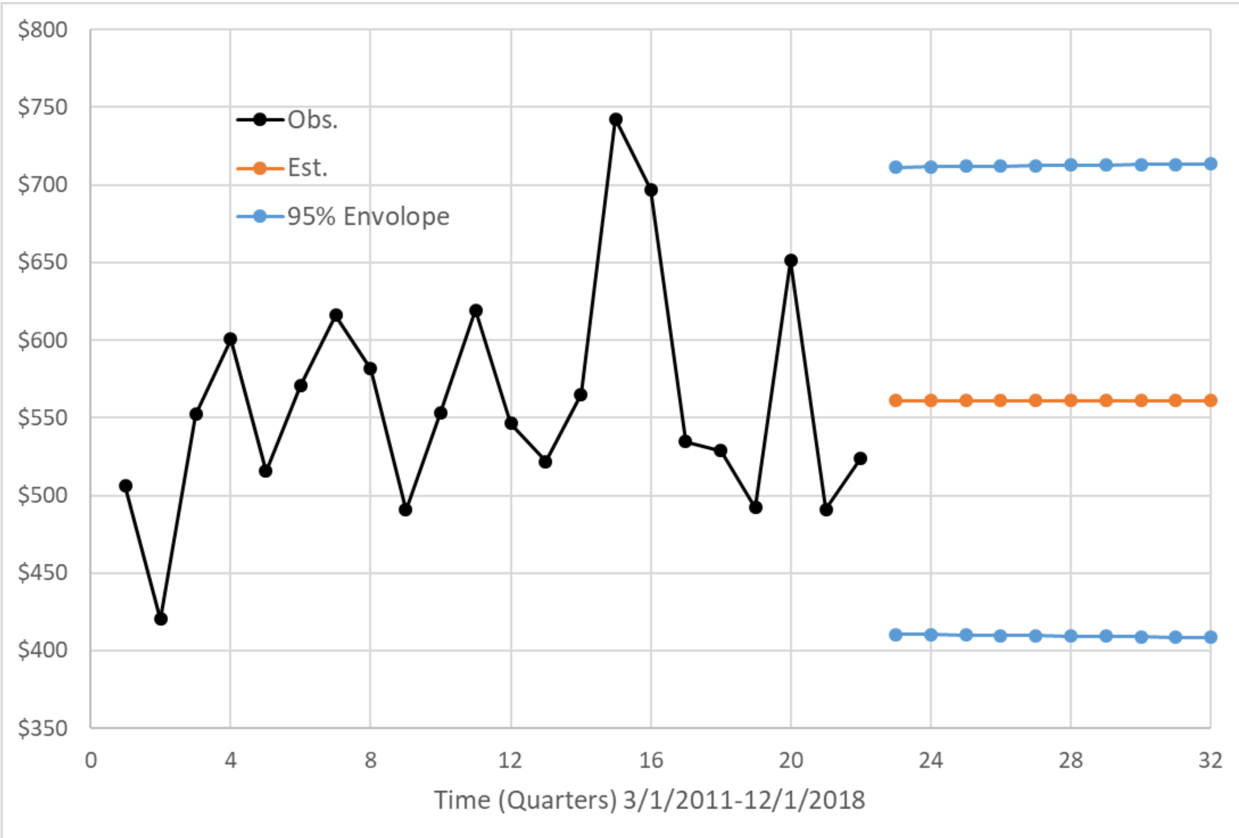

Figure 6 Sample Bid Item Quarterly Price Reported and Estimated Using ARIMA Model

Also, the increase of (iv_3) Project Cost has a positive relationship with the bid item price. This is due to the anticipated increase of material prices and the uncertainty introduced by longer project durations associated with large projects (higher project cost) that require contractors to guarantee the price of materials for a longer project duration. While TDOT has a bituminous price index that accounts for the change in the Bituminous Material for Tack Coat, contractors are having to account for the possible material cost increase by including an added risk cost (i.e., bidding risk) into the bid item prices. This relationship promotes the letting of large projects in phases consisting of smaller sections. Also, the same relationship can be used to justify the costs of project acceleration.

The bid item price for Bituminous Material of Tack Coat has a negative relationship with (iv_7) Number of Projects in the Letting. This can be explained by the link between construction costs and the funding available for highway projects. The increase in the sample bid item price is related to the decrease in (iv_7) Number of Projects in the Letting. In times when highway project costs are increasing, TDOT will reduce the number of projects being let due to the limited funds available for the projects being let. To explain further, the number of projects being let is dependent on the cost of construction including the cost of bituminous items.

The inverse relationships that are most difficult to explain are between the bid item price, (iv_1) Oil Price, and (iv_{25}) Cement Producer Price Index. The bituminous material price is expected to rise with the increase of oil prices as bitumen is a product of the crude oil distillation process. However, the opposite appears to be the case with the negative correlation coefficients. A possible explanation is that at times with lower costs of construction materials, more projects are let to contracts creating higher demands for quantities. Such an increase in demand is influencing the rise in materials costs at the local level, including the bituminous materials. Similarly, the same explanation or rationale can be used with the Cement Producer Price Index.

Model Testing Results

The resulting MLR developed model has the following results:

Coefficient of correlation	= 0.993
R^2 value	= 0.987
Residual standard error	= 1300
Overall F statistic	= 1846759
Significance associated P-value	= 0.000

The R^2 value of 0.987 is very high considering that the maximum possible value of R^2 is 1. This high value indicates that the selected independent variables can be used to estimate the dependent variable (bid item price) using the developed model. Considering the 95% confidence level, the MLR analysis, and the P-value of 0.00, statistically significant relationships between the sample Bid Item Prices and the independent variables are shown. This is for the overall regression data of all bid line items. The R^2 for the bid line item models range from a minimum of 0.12 to a maximum of 1.00.

To illustrate the modeled variable coefficients and explain the developed models in more details, the Bituminous Material for Tack Coat bid item model is used as an example. As previously explained, all models follow the polynomial Equation 3 of

$y = \beta_0 + \beta_1 iv_1 + \beta_2 iv_2 + \beta_3 iv_3 + ... + \beta_n iv_n + \epsilon$

Thus the equation for the sample bid item Bituminous Material for Tack Coat model is the following:

$y = 1515.06 - 0.78\, iv_1 + 0.19\, iv_2 + 0.07\, iv_3 - 0.07\, iv_4 + 1.01\, iv_5 - 0.37\, iv_6 - 0.45\, iv_7 + 13.41\, iv_{10} - 7.31\, iv_{11} + 22.37\, iv_{12} + 19.35\, iv_{13} + 36.71\, iv_{14} + 97.02\, iv_{17} - 32.62\, iv_{19} + 4.66\, iv_{23} + 1.39\, iv_{24} - 6.45\, iv_{25}$

This can also be expressed as the following:

Bid item price for (403-01) = 1515.06 - 0.78 Oil + 0.19 LET + 0.07 AM100K - 0.07 Qty + 1.01 Bids - 0.37 DUR - 0.45 Projs + 13.41 BRCN - 7.31 ASPH + 22.37 SAFE + 19.35 BRRP + 36.71 EMER + 97.02 SIGN - 32.62 SRFT + 4.66 FF + 1.39 PPISt - 6.45 PPICem

The correlation coefficient values and signs show direct relationships between the increase in the sample Bid Item Prices and the independent variables listed in Table 2. These include Letting Date, Award Amount, Number of Contractors Bidding, Project Types (Bridge Construction, Safety, Bridge Repair, Emergency, Signal Improvement, and Surfacing), Federal Funding, and Producer Price Index for Steel.

In order to randomly separate some data for testing, every fifth observation in the bid line item data is separated from the model development. The separated data are used for testing the model fit for the estimated results and the historical observations for the same period. This selection process produced a statically significant model that is able to estimate bid item prices and has a coefficient of correlation of 0.993 and an R^2 value of 0.987.

Using the test data and considering the 95% confidence level, the MLR analysis and the P-values (less than 0.05) show that the developed models explain the change in sample Bid Item

Prices (Observations) using the independent variables listed in Table 2. The independent variables have statistically significant relationships using the coefficients listed earlier in Table 2. Estimated values generated by the developed models are plotted and the plots for the sample bid item were presented earlier in Figure 2 with the observed data for the sample bid item of Bituminous Material for Tack Coat.

The estimated values presented earlier in the scatter plots in Figure 2 show how the model was able to estimate the bid item prices according to each of the plotted independent variables. The plotted estimated values show more consistency and have less outlier values when compared to the plotted observed values. This is an inherent characteristic when developing linear regression models and the linear fitting process whereby observation differences from the fitted linear model are considered as errors.

Time Series Analysis Results

Using the sample bid item of Bituminous Material for Tack Coat to present the application of the ARIMA prediction model (0,1,1) using R programming language, the model produced an estimated (Est.) sample bid item price of $561 per ton for the 10 periods (quarters) ahead of the series data. This estimated model is within a 95% confidence level envelope ($408 to $713 per ton) of the sample Bid Item Prices. This conforms to the time series and the estimated price between the 95% confidence envelope for the 10 quarters of the data series. This step is used as a verification of the developed model. Figure 6 shows the time series and the ARIMA model prediction for the sample bid item price for the future 10 periods. The 95% confidence envelope indicates the range where the predicted sample bid item price can vary.

Tennessee Highway Construction Cost Index (TNHCCI)

The estimated bid item price values generated by the developed estimation models are used to generate a Highway Construction Cost Index as explained earlier in the Indexing subsection. Again, this index is derived using the 34 selected items and can be used for developing a Highway User Fee. The index generated by the estimated data generated by the developed model is presented and plotted in Figure 7 as TNHCCI. The Figure 7 graph that follows has two vertical axes with index values on the left vertical axis and Tennessee Gasoline Tax values on the right. The gasoline tax is one of the taxes used in collecting highway project funds. While this is a representation of the indexing of a Highway User Fee, the developed TNHCCI can be applied to other taxes as necessary to link the change in highway project costs to a representative highway user fee.

Figure 7 is plotted with the assumption that the base period for NHCCI and TNHCCI starts on 2/11/2011. The index is calculated thereafter for every TDOT Highway Project letting date and the corresponding NHCCI quarterly value is matched for the date. The TNHCCI is considered volatile when compared to the NHCCI. This volatility occurs because of the inability of the proposed process to calculate the TNHCCI with high resolution and granularity. But the resulting resolution does produce an index at the bid letting scale. Such resolution can be reduced to a lesser granularity based on further refinement and analyses. For example, when applying the index on the programming level scale, a three-year average is more applicable. Also, using a four or eight periods moving average should be considered in estimating the user fees and adjustment to taxes to allow for annual and semi-annual adjustments. However, when making decisions on a bid letting scale, the presented granularity of the bid letting scale is the most appropriate. The annual cost index is presented next.

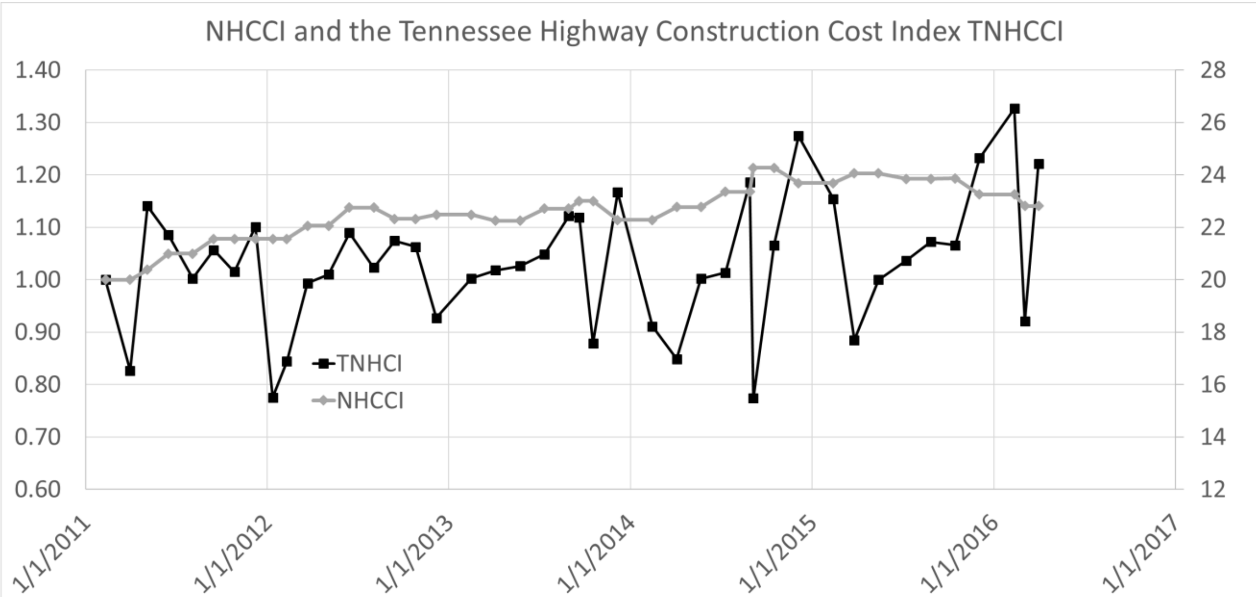

Figure 7 NHCCI and TNHCCI (Jan 2011 - April 2016)

Note that if the index is applied to state gasoline tax starting at the base value of 20 cents on 2/11/2011, changes to the gasoline tax would have varied between 15.5 and 26.53 cents per gallon of fuel sold. Also, note that the ending value of 24.43 cents per gallon of gasoline closely matches the 24 cents gasoline tax rate starting in 2017 following the Public Act of 2016. The act included a 4 cents increase to the gasoline tax applied in July 2017 followed by 1 cent increase in July 2018 and another 1 cent increase in July 2019.

To reduce the volatility of the index change, the following graph in Figure 8 is calculated to represent the equivalent annual index of the NHCCI and the TNHCCI.

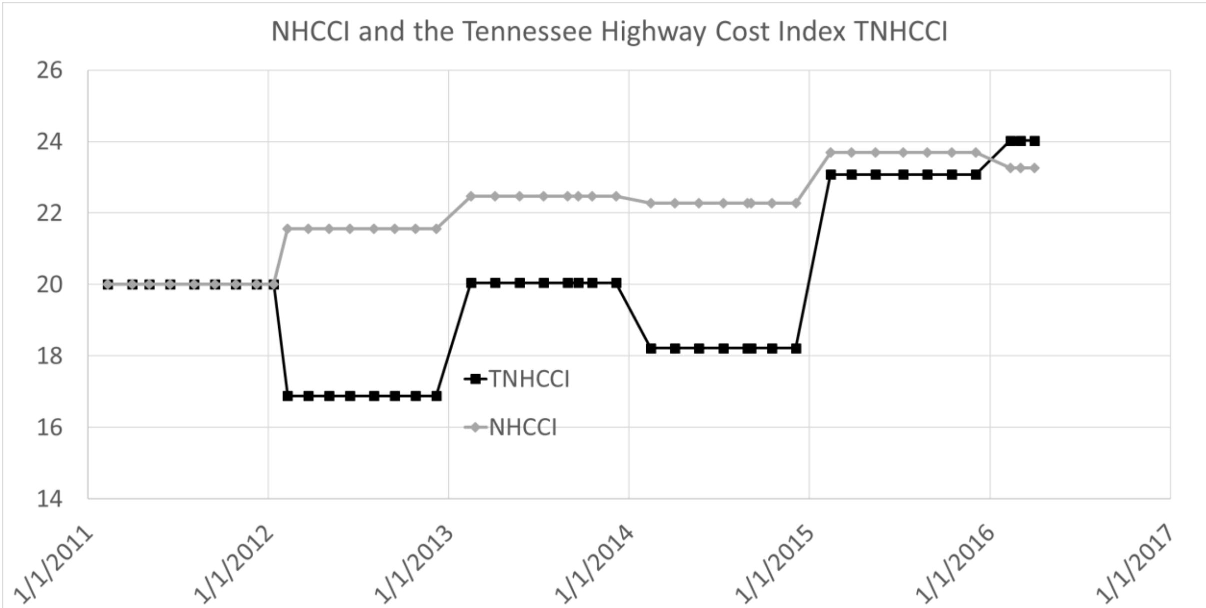

Figure 8 Annual NHCCI and TNHCCI (2011-2016)

Starting with the gasoline tax of 20 cents, applying the TNHCCI shows some decreases in 2012 and 2014 and increases in other years ending with a 24 cent gasoline tax in 2016. When compared to the NHCCI, indexing the gasoline tax results in a subtle increase to the gasoline tax of 23.25 cents in 2016. While this is a representation of the local highway project construction costs, the TNHCCI is based on the highway construction projects actually let in Tennessee. When planning for future highway budgeting, more control can be applied to the index to reduce such volatility. However, this index decrease (in 2012 and 2014) is an actual representation of the project price estimates and not the actual needs of the Tennessee Highway Project Program. In other words, the bid item prices from projects let during the study period are leading the calculated index. In the case of future budgeting, a Highway Project Program would lead the index and therefore the corresponding gasoline tax. Be sure to note that this process does not determine the actual need, but calculates the corresponding index to the estimated bid item prices and highway project costs.

CONCLUSION

Summary of Significant Research Findings

This research developed and tested data selection, modeling techniques, and an indexing process for comparing the change in bid item prices to multiple local and global factors. Such a process can be used in developing future bid item price estimation models and with an indexing mechanism using NHCCI data filters. The bid item price estimation model can be used by Transportation Department administrations in estimating Highway Project Costs and an excise gas tax “Highway User Fee” index. The modeling process can be used to forecast an index that is sensitive to local factors affecting bid item prices. The index can be used to link the forecasted Highway Project costs to a Highway User Fee. This index, when applied to the Highway User Fee, can provide a direct link between actual Highway Project Costs and the taxes collected from the users. The index is more sensitive to local factors affecting the change in bid item prices than comparable national indices similar to NHCCI.

Modeling the factors affecting the change in bid item prices presents some trends that can be used to influence planning choices and how projects are being let. Such choices include optimizing project size, letting times, construction phasing, project duration, acceleration costs, and projects bundling. Transportation Departments can use this process to influence Highway Project Costs. As an example, bid item prices have shown convergence to a narrower envelope with the increase of some independent variables similar to bid item quantities. To further explain, the sample bid item prices of Bituminous Material for Tack Coat tend to show a larger range of prices as quantities vary, and specifically higher prices for lower bid item quantities.

In contrast, certain types of projects with higher total costs showed an associated increase in bid item prices. Such relatively higher cost projects with long durations are associated with the risks in having to guarantee a bid item price for a longer project duration and relatively higher costs. Therefore, projects can be let in phases of shorter durations with smaller project costs to reduce risks with stated project bid item prices over longer time periods. Also, the same relationship can be used to justify the added costs of accelerated projects and the potential gain in decreasing bid item prices associated with shorter project durations. Bid Item Quantities and Project Cost, as independent variables, showed significant relationships with bid item prices in the model development process as presented in Table 2. Interestingly, the relationships between bid item quantities and project costs, numbers, and durations require a balancing process to optimize the most effective project size and bid item quantities.

The project letting date was used to define the time period in calculating several parameters. The selection of the period is necessary and essential to study the presence of any seasonality changes to bid item prices. In addition, the letting period is used to calculate multiple indicators to include the weighted average bid item price, the number of projects being let, and the corresponding material prices at the time of the letting. However, applying the index effects to a longer duration is necessary to reduce volatility that can influence the gasoline tax when only applied monthly.

Using relatively longer durations of time (2002-2016) in developing the MLR model produces estimation models with a forced upward trend that could not explain bid item price changes for the recent years. A relatively shorter duration (2011-2016) eliminates this effect and produces better fitting models. Similarly, the use of relatively longer duration is found (by NHCCI) to affect the resulting index by using an old base year. The older the base year the more the base price values become irrelevant to recent year prices, and the resulting index becomes less representative of the current price values.

The developed bid item prices estimation model presented in this research has some limitations due to the inherent nature of the MLR modeling techniques. With a multitude of variables represented by the numerous bid items considered, MLR represents a modeling technique that is widely used and trusted in construction estimating. This research was then able to further define the majority of the relationships between bid item prices and the independent variables in bid data. The developed model and the independent variables were found to be of significance based on the model testing results.

The ARIMA model produced significant model prediction of a constant value and the 95% confidence level envelop using a moving average parameter and integration by differencing. The selected simple (parsimonious) ARIMA model (0,1,1) presented the best model for the majority of the selected bid items.

The Fisher Price Index produced a superlative and simple approximation index that is capable of producing representative bid item price indices and a state level Highway Construction Cost Index. The resulting estimated bid item prices and the estimated indices closely followed the observed and the calculated bid item prices observations indices. These results give validity to the developed bid item price estimation models, and is indicative of representative models and indices when compared to the studied historical bid item prices.

This research presents a modeling approach to indexing the cost of highway construction and maintenance projects using selected bid items prices and demonstrates the effect of the developed index on the fuel taxes. A comprehensive index utilizing the selected bid items was developed, and was a significant research contribution. In addition, the resulting price index is based on all the identified items of significance and therefore is more representative of collective local changes in bid item prices and highway project costs.

Future Research

The developed modeling approach is dynamic and automation of the process can be developed to provide real time results from the most current letting data. Using such automation can be used to optimize highway construction programs to the parameters of choice similar to total project costs, anticipated changes in the industry, risk mitigation, and other factors tailored to the agency's choice. This ability to identify the factors affecting the change in bid item prices is one of the benefits of using the Multiple Linear Regression model. Using Artificial Intelligence (AI) models techniques may present comparable results, but lacks the latter identified benefit. Future research may wish to compare AI models or other modeling techniques to MLR.

Another recommendation for future research is to test regression by using classes of low, medium, and large Bid Item Quantities. Similarly, Project Cost may present significance in relation to bid item prices in different classes (e.g., less than \$100K, \$100K to \$1M, more than \$1M).

More research could also be conducted to model the variations in resulting bids as a function of the expected number of bidders and number of projects versus the available remaining variables. Such a model is needed, would assist in scheduling the projects over bid lettings, and would reduce the risk effect of increasing item prices due to factors that can be controlled through better budgeting and scheduling decisions.

Bid item prices may show stationarity and seasonal patterns to explain changes in item prices. By identifying such patterns, scheduling projects accordingly can be used to reduce seasonal effects, if any. The developed ARIMA model showed some significance in simulating the data, but no strong significance in the effect on item prices over the period under study. The developed ARIMA model could be used to test different indexing periods expecting different results and significance per time period.

REFERENCES

Akaike, H. (1981). “This week’s citation classic.” *Current Contents Engineering, Technology, and Applied Sciences*, 12(51), 42.

ASCE (2017). “2017 infrastructure report card.” ASCE Reston, VA.

Atkinson, R. D., Commission, N. S. T. I. F., et al. (2009). *Paying Our Way: A New Framework for Transportation Finance: Report of the National Surface Transportation Infrastructure Financing Commission: Executive Summary*. National Surface Transportation Infrastructure Financing Commission.

Board, T. R. (2006). *The Fuel Tax and Alternatives for Transportation Funding: Special Report 285*. The National Academies Press, Washington, DC, <https://www.nap.edu/catalog/11568/thefuel-tax-and-alternatives-for-transportation-funding-special-report>.

Chou, J.-S., Peng, M., Persad, K. R., and O’Connor, J. T. (2006). “Quantity-based approach to preliminary cost estimates for highway projects.” *Transportation Research Record*, 1946(1), 22–30.

Darby, M. R. (1982). “The price of oil and world inflation and recession.” *The American Economic Review*, 72(4), 738–751.

Hillmer, S. C. and Tiao, G. C. (1982). “An arima-model-based approach to seasonal adjustment.” *Journal of the American Statistical Association*, 77(377), 63–70.

Jeong, H. D., Gransberg, D., and Shrestha, K. J. (2017). “Advanced methodology to determine highway construction cost index (hcci).” *Report no.*

Kyte, C. A., Perfater, M. A., Haynes, S., and Lee, H. W. (2004). “Developing and validating a tool to estimate highway construction project costs.” *Transportation research record*, 1885(1), 35–41.

Lowe, D. J., Emsley, M. W., and Harding, A. (2006). “Predicting construction cost using multiple regression techniques.” *Journal of construction engineering and management*, 132(7), 750–758.

Nebraska (2018). “Legislative bill 941.” One Hundred Fifth Legislature,

<https://nebraskalegislature.gov/bills/view_bill.php?DocumentID=34332> (Jan. 9, 2018).

NHCCI, F. (2020). "National highway construction cost index (nhcci)." FHWA, <https://www.fhwa.dot.gov/policy/otps/nhcci/> (Jun. 8, 2020).

Parry, I. W. H., Walls, M., and Harrington, W. (2007). "Automobile externalities and policies." *Journal of Economic Literature*, 45(2), 373–399.

Small, K. A. and Van Dender, K. (2007). "Fuel efficiency and motor vehicle travel: the declining rebound effect." *The Energy Journal*, 28(1).

Wallace, N. (1996). "Hedonic-based price indexes for housing: theory, estimation, and index construction." *Economic Review*, 34–48.

Williams, T. P. (2002). "Predicting completed project cost using bidding data." *Construction Management & Economics*, 20(3), 225–235.

Williams, T. P. (2003). "Predicting final cost for competitively bid construction projects using regression models." *International Journal of Project Management*, 21(8), 593–599.

Wright, M. G. and Williams, T. P. (2001). "Using bidding statistics to predict completed construction cost." *The Engineering Economist*, 46(2), 114–128.