# *Cost Estimation Comparisons Between Least Square Regression and Quantile Regression on Fiber Reinforced Bridge Projects*

***Ankith Reddy PAMULAPARTHY*[1]**
***Robert C. CREESE*[2]**
***Hota V. GANGARAO*[3]**
***Majid JARIDI*[4]**

**Abstract**

This paper focuses on the Least Square (LS) regression using the mean and Quantile (M) regression analysis using median which is based on "well-Known" parametric estimation methodologies. Data from Oregon and California highway bridges were used for the comparison of the two methods. Relationships were developed to predict the unit cost of FRP repair work and FRP cost was found to have a high degree of correlation with FRP area for both Oregon and California. It was observed that the Cost Estimating Relationships (CERs) obtained by Quantile (M) regression method had the smaller Mean Absolute Deviation (MAD) values and lower Mean Absolute Percentage Error (MAPE) values than Least Square (LS) regression. The stuudy showed that Quantile Regression is much less sensitive to outliers than Least Squares Regression.

***Keywords:*** *Quantile regression, cost estimating relationships (cers), fiber reinforced polymer (frp) wraps, outliers*

## 1. Introduction

State transportation departments are faced with a challenge to keep bridges under their jurisdiction in good operating conditions. Corrosion of bridges has been a constant challenge for engineers and new materials (e.g. polymers, metals, ceramics and their composites) are being developed to minimize corrosion related issues. Fiber Reinforced Polymers (FRPs) are composites that combine the strength of fibers with the stability of polymer resins. The strength of FRP materials comes from the type of fibers used, usually they are glass, carbon or aramid fibers. The primary advantages of FRP materials are: lightweight, non-magnetic, high corrosion resistance and high strength to weight ratio. These advantages make FRP materials a viable option for the initial construction or for rehabilitation of current bridges. As a new technology, it is hampered by a lack of approved standards and an effective cost estimation methodology.

Cost Estimating Relationships (CERs) are the parametric equations that are developed between different variables to determine the cost of a project component or the total cost of the project. These relationships are developed by cost estimators and are used by managers to control the variables that impact the cost of a project [4]. The primary statistical method to develop CERs is Least Square (LS) regression which is based upon the mean of the data. An alternative method is the Quantile (Q) regression which is based upon the median of the data [6]. This project compares

[1] *Graduate Student, Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, West Virginia, USA, ankith.redde@gmail.com*
[2] *Professor Emeritus, Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, West Virginia, USA, Robert.Creese@mail.wvu.edu*
[3] *Professor, Civil and Environmental Engineering Department & Director of the Constructed Facilities Center, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, West Virginia, USA Hota.Gangarao@mail.wvu.edu*
[4] *Professor, Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University, West Virginia, USA Majid.Jaridi@mail.wvu.edu*

the two methods using limited data on highway rehabilitation projects in Oregon and California utilizing the software ESTIMFEC [1] to perform the analysis. One of the problems of using historical data is that it may contain outliers and Quantile regression is less sensitive to outliers than the Least Square regression.

## 2. Methodology

Least Square regression is the most commonly used method to develop linear predictive relationships including cost relationships and engineering relationships. Quantile methods are commonly used to determine percentile rankings used on many testing scores and in areas such as finance and sociology. Quantile regression methods have been used within labor or educational economies to study wage determinants, trends in income inequality or effects of socioeconomic characteristics on education attainment [3].

### 2.1 Least Square (LS) Regression

Least Square regression is a statistical method of analysis that estimates the relationship between one or more independent variables and the dependent variable by minimizing the sum of squares of the difference between the observed and predicted values of the dependent variable [8]. A simple Least Square regression model involving only one independent variable (X) predicting a dependent variable (Y) is expressed by Equation 1.

$$Y = a + bX + \varepsilon \quad (1)$$

Here "a" is the intercept that indicates where the straight line intersects Y–axis; "b" is the slope that indicates the degree of steepness of the straight line and "ε" represents the error. The best line or a relationship would be the one with the Least Sum of Squared Errors (SSE) [9].

$$SSE = \sum\left(Y_i - \hat{Y}\right)^2 \quad (2)$$

Where

$Y_i$ = Dependent variable i and where i = 1,......n

$$\hat{Y} = (\sum Y_i)/n \quad (3)$$

#### 2.1.1 Drawbacks of Least Square (LS) Regression

According to Foussier [4], the main drawback is "regression". If there are a large number of observations and they are scattered, the Least Square method underestimates the larger observations and overestimates the smaller observations as the Least Square method regresses towards the average value. It is a major drawback, since the costs of a product or project are always scattered, and there is no necessity for the cost to regresses towards the average cost. The Least Squares regression is based upon the data having a normal distribution and this does not always occur. If the data is truly normal, then the mean and median would be the same. The other major disadvantage of the Least Square method is that it is sensitive to outlier that is due to squaring of the error term. Since the Least Square method is not robust, this might have a tremendous impact on the predicted cost. Foussier [4] indicates it is due to the small break down of the arithmetic average. Hence, the change of just one data point might have a severe impact on the predicted cost. The outliers that are far away from the tentative CER will strongly impact the predicted cost and the $R^2$. A tendency to eliminate the outliers to increase $R^2$ might cause one to eliminate too many data points.

According to Xia [10], the third disadvantage of Least Square method is multi-collinearity. There may be several variables that may be linearly collinear to each other in the data. It may adversely affect the coefficient of the estimates when there is a small change in the data. The relationship thus obtained between these variables does not make sense even though it is mathematically correct and produces a large confidence interval. This is due to the fact that to estimate the values of the parameters one needs to compute the inverse of the matrix based on the inputs. The determinant of a matrix with two collinear variables is equal to zero and hence these matrices cannot be inverted reliably to estimate the parameters [4]. However the relationships developed in this study were concerned with only one variable.

### 2.2 Quantile Regression

According to Koenker [6] Quantile Regression is the estimation of Quantiles of the conditional distribution. It models the relation between a set of predictor variables and the specific Quantiles of the response variable. The prominent form of Quantile regression is the Median regression. Koenker and Bassett stated that for a sample $Y_1, Y_2 \ldots Y_n$ from a population Y, we can compute the sample median $\nu$ by minimizing the sum of absolute deviations.

$$\hat{\nu} = min \sum_{i=1}^{n} |Y_i - \nu| \tag{4}$$

Similarly we can compute the $\tau th$ sample Quantile $\nu(\tau)$ by minimizing the absolute deviations of various Quantiles.

$$\hat{\nu}(\tau) = min \sum_{i=1}^{n} \rho_\tau \, |Y_i - \nu| \tag{5}$$

Here $\rho_\tau(z) = z(\tau - I(z < 0)), 0 < \tau < 1.$ $I(.)$ denotes the Indicator function. Koenker stated that the optimal value of $\tau$ is the median. The estimate of linear conditional Quantile function (i.e. of $Q_\tau\left(\frac{Y_i}{X_i}\right) = X_i\beta_\tau$) can be found by the criterion:

$$\hat{\beta}(\tau) = \arg min \sum_{i=1}^{n} \rho_\tau |Y_i - X_i\beta| \tag{6}$$

Where $\hat{\beta}(\tau)$ is called the $\tau th$ regression Quantile. The minimizer function $\sum_{i=1}^{n} \rho_\tau |Y_i - X_i\beta|$ is differentiable except at $Y_i = X_i\beta$. A simplex based algorithm was developed by Barrodale and Roberts [2] to solve the minimizer function and was later extended by Koenker and d'Orey [7] to Quantile regression estimation. The case where $\tau = 1/2$ corresponds to median regression, which is also known as L1 regression.

## 3. Data Source and Analysis

### 3.1 Oregon Bridges

The state of Oregon had several bridges that have undergone repair and rehabilitation work in recent times. There were seventeen bridges that had undergone repair work using FRP wrappings for which the data was obtained. The area of the bridges ranges from 500 $ft^2$ to 9,204 $ft^2$. There were multiple bids that were quoted for each bridge repair work by several contractors. Table 3.1 gives the total contract cost, FRP area, FRP cost and other calculated parameters.

**Table 3.1:** Summary of Year, Accepted FRP Cost, Accepted FRP Unit Cost and Total Contract of Oregon Highway Bridges

| Location | Year | Bridge No. | FRP Area, $ft^2$ | Accepted FRP Unit Cost, $\$/ft^2$ | Accepted FRP Cost, $ | Total Contract, $ | % of FRP Cost of Total Contract |
|---|---|---|---|---|---|---|---|
| OR | 2007 | 01786A | 3,144 | 50.68 | 159,343 | 12,537,779 | 1.27 |
| | 2008 | 07601B | 712 | 40.82 | 29,062 | 2,548,552 | 1.14 |
| | 2008 | 08381S | 1,280 | 39.30 | 50,299 | 2,548,552 | 1.97 |
| | 2008 | 7458 | 1,955 | 97.20 | 190,020 | 10,193,994 | 1.86 |
| | 2008 | 08383N | 6,103 | 27.47 | 167,665 | 3,324,179 | 5.04 |
| | 2009 | 8843 | 2,345 | 106.82 | 250,480 | 3,442,437 | 7.28 |
| | 2009 | 03173A | 9,204 | 13.48 | 124,069 | 3,885,033 | 3.19 |
| | 2010 | 7530 | 500 | 61.11 | 30,557 | 2,002,807 | 1.53 |
| | 2010 | 7532 | 500 | 61.11 | 30,557 | 2,002,807 | 1.53 |
| | 2010 | 7533 | 500 | 13.10 | 6,548 | 2,002,807 | 0.33 |
| | 2010 | 7534 | 500 | 61.11 | 30,557 | 2,002,807 | 1.53 |
| | 2010 | 06945A | 840 | 38.98 | 32,739 | 3,013,112 | 1.09 |
| | 2010 | 04981A | 1,900 | 31.59 | 60,022 | 976,687 | 6.15 |
| | 2010 | 02233A | 1,900 | 40.20 | 76,391 | 976,687 | 7.82 |
| | 2010 | 04979A | 2,600 | 27.70 | 72,026 | 976,687 | 7.37 |
| | 2010 | 02236A | 3,200 | 27.28 | 87,304 | 976,687 | 8.94 |
| | 2010 | 7392 | 5,800 | 30.11 | 174,609 | 457,704 | 38.15 |

Note: All the cost data has been adjusted for inflation to 2013 dollar value

The accepted FRP cost and FRP area are the two variables that are used to predict the unit cost of FRP repair work. The total FRP cost constitutes a minor fraction of the total contract cost. Hence during the estimation of unit cost of FRP repair work, FRP cost is used instead of the total contract cost. Only one layer of FRP application was considered due to limited availability of the data. Bridges with accepted FRP cost of less than 1% of the total contract cost were omitted from the paper. Bridge No. 7533 was found to have a value of 0.33% (Table 3.1) and was not considered in the study as its cost was more than 70 percent lower than three other bridges with the exact same area. Figure 3.1 shows the scatter plot of FRP cost and FRP area. Figure 3.1 indicates three points (i.e. bridges 7458, 8843, 03173A) to be potential outliers. The $R^2$ value obtained for all the data points included is 0.278 and coefficient of determination ($R^*$) value obtained for Quantile (Median) regression is 0.430. The bridge 01786A was not considered to be an outlier since its FRP cost was within 95% of the confidence interval and no physical characteristic was detected that indicated it to be an outlier.

It can be observed from Figure 3.1 that the Least Square regression method overestimated smaller values and underestimated larger values when compared to Quantile (Median) regression. The slope of the two equations are different and the Least Squares is greatly affected by the outliers as indicated by Equations 7 and 8. The regression equations obtained by Least Sqaure and Quantile (Median) methods for all the data included are:

Accepted FRP Cost (LS-\$) =57,515 + 15.19*FRP Area($ft^2$) (7)

Accepted FRP Cost(Q (M)-\$) =18,321 + 24.47*FRP Are($ft^2$) (8)

The coefficient of determination values obtained are 0.278 (LS, $R^2$) and 0.430 (Q (M), $R^*$) respectively. The equations have very different slopes and some of the data points appear as outliers. When the outliers are detected and removed, the two methods will be in more agreement.

Further analysis for outliers is made by conducting a Tietjen-Moore Test for multiple outlier detections. The number of suspected outliers in the test were considered to be three (i.e. k = 3). The value of test static $E^k$ to be 0.2131 which is less than the critical value of 0.2249 thus confirming the presence of three outliers in the data. When the number of suspected outliers in the test were considered to be four (i.e. k = 4) the test static ($E^k$) value (0.1910) was greater than the critical value (0.1508) and did not confirm four outliers. Further residual analysis wais carried out to identify these outliers.

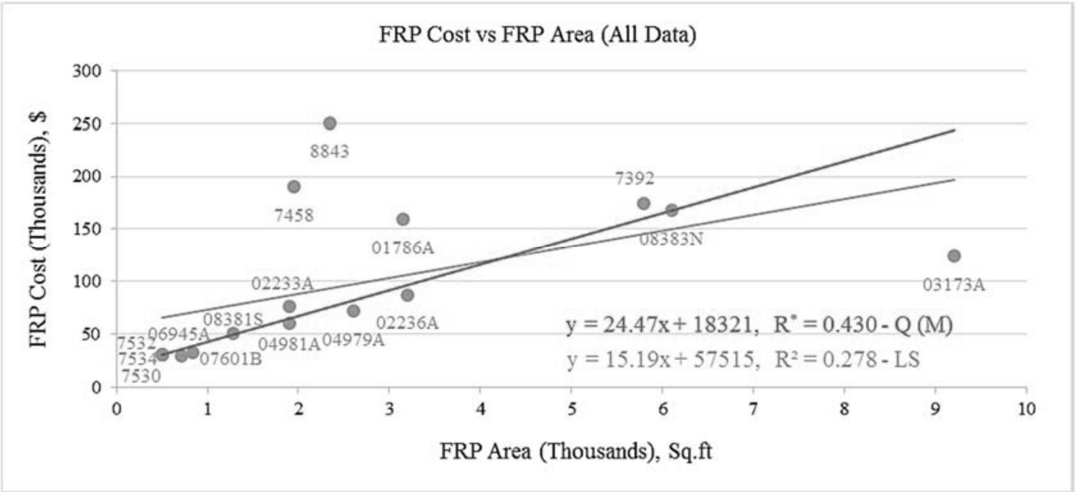


**Figure. 3.1:** Scatter Plot (All Data) of Accepted FRP Cost and FRP Area of Oregon Highway Bridges (Q (M)) & LS) with $R^2$ = 0.278 & $R^*$ = 0.430

The computations to identify the outliers are made in three steps:

1. A linear relationship is obtained using all the data points and the prediction of FRP cost for each bridge is made using this relationship. Deviations of the predicted FRP cost from the actual FRP cost is computed as "Deviations 1".
2. Then each data point is successively discarded and therefore N new relationships are computed. Appendix 1 shows the $R^2$ corresponding to the new relationship without this data point and the deviation, called "Deviation 2" between this data point and the new relationship. A few $R^2$ are much higher (>0.100) than the corresponding data points are identified as outliers since their deletion greatly improves the relationship.
3. The difference between both the deviations is divided by the standard deviation of "Deviation 1" and is multiplied by a factor of 100. The column "Relative Variation" from Appendix 1 gives the resultant value. The data points that have a relative variation of more than ±50% are considered to be outliers in this paper.

Appendix 1 shows that bridges 7458 and 03173A had a relative variation more than the threshold value of ±50% thus indicating them to be possible outliers. Appendix 2 also shows that removing the bridge 8843 greatly improves the correlation between the FRP cost and FRP area.

- Bridge 03173A is a seismic retrofit involving different repair procedure than the rest of the bridges thus having a much higher cost than the rest of the bridges.
- Bridge 7458 involved construction and removal of various temporary work platforms with enclosure. The humidity and temperature in the enclosure was maintained using an HVAC system which is the likely reason for higher unit cost of FRP repair work.
- Bridge 8843 has many elements that contain asbestos. The contractor had to remove these materials before proceeding with the FRP application.

Bridges 03173A, 7458 and 8843 were identified as spurious bids considering their unusual FRP repair cost. These bridges were considered as outliers and were not considered in further analysis of the data.

Equation 9 was obtained for accepted FRP cost against FRP area for all the bridges in Oregon by Least Square method with an $R^2$ of 0.884 is given by:

Accepted FRP Cost (LS-$) =16,736 + 27.03*FRP Area ($ft^2$) (9)

Equation 10 was obtained by performing Quantile (Median) regression for accepted FRP cost against the FRP repair area for all the bridges in Oregon with an $R^*$ value of 0.773 is given by:

Accepted FRP Cost (Q (M)-$) = 18,321 + 24.47*FRP Area ($ft^2$) (10)

The intercept values ( 16,736 (LS), 18,321 (Q (M))) represent the fixed set-up costs such as mobilization, equipment and traffic control.

Figure 3.2 is developed between the accepted FRP cost against FRP area to compare the two methods (LS & Q (M)) of analysis. It shows the extreme observation point (01786A) had minimal impact on the relationship obtained by Quantile (Median) regression method. The slope difference (absolute) between the CERs of Figure 3.1 and Figure 3.2 is 0 (Q (M)) and 11.84 (LS) which indicates the robustness of Q (M) regression with respect to outliers. The slope of CER obtained by Least Square method changed more than 75% indicating a high influence of outlier on the Least Square method whereas the outliers has no effect on the CER obtained by the Qunatile method. When the outliers are removed, agreement is high as indicated by Figure 3.2.

The results from Table 3.2 show that Least Square regression method has the least SSE value and Quantile (Median) regression method has the least MAD and MAPE values. Profits and costs are indicated better by the mean absolute deviation and by the mean absolute percentage error than by the sum of squares of the errors. A relationship between the accepted FRP cost and the FRP area to predict the unit cost ($/ft$^2$) values, can be established. The predicted unit cost obtained from Equations 11 and 12 are given by:

Unit Cost (LS-$/ft$^2$) = 27.03 + 16,736/FRP Area (ft$^2$) (11)

Unit Cost (Q (M)-$/ft$^2$) = 24.47 + 18,321/FRP Area (ft$^2$) (12)

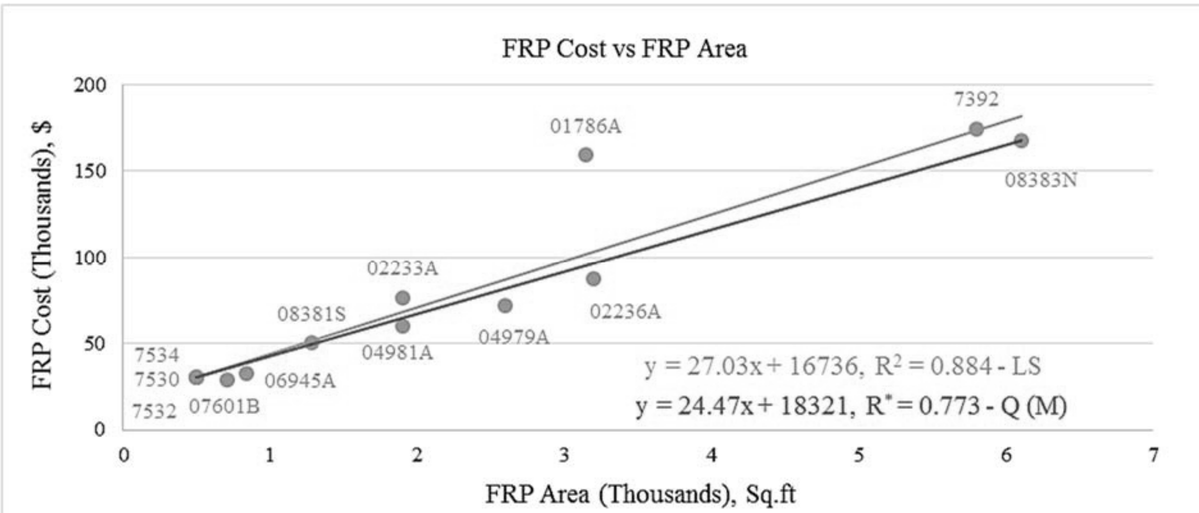


**Figure. 3.2:** Scatter Plot of Accepted FRP Cost and FRP Area of Oregon Highway Bridges (LS & Q (M)) with $R^2 = 0.884$ & $R^* = 0.773$

Table 3.3 shows the predicted FRP unit cost ($/ft$^2$) values obtained by the two methods (LS & Q (M)) of analysis using Equations 11 and 12.. These equations are similar as the outliers have been determined and removed, and thus agreement between the two methods is high. The actual unit costs versus predicted unit costs have good agreement overall, but the predicted equations are approximate only. There is the assumption that the start-up costs, which is indicated by the intercept of the curve, is constant for all project sizes. When the bridges are grouped into two ranges as done in Table 3.3, the agreement is much better for the unit costs.

Table 3.4 gives the weighted average by area of FRP unit cost for the two different methods of analysis for two separate area ranges (i.e. <2000 & >2000). The weighted average values for both the Quantile method and the Least Squares method are near the actual unit cost values. The difference in values between the two area ranges indicate the large impact of area upon FRP unit cost

**Table 3.2:** Residual Analysis Table for Oregon Highway Bridges

| Area, ft$^2$ | Accepted FRP Cost, $ | Predicted FRP Cost, $ | | SSE (X-Xp)^2 | |
|---|---|---|---|---|---|
| | | Least Square | Quantile (Median) | Least Square | Quantile (Median) |
| 712 | 29,062 | 35,989 | 35,744 | 47,981,936 | 44,646,310 |
| 1,280 | 50,299 | 51,347 | 49,643 | 1,097,351 | 431,329 |
| 6,103 | 167,665 | 181,756 | 167,661 | 198,569,710 | 10 |
| 500 | 30,557 | 30,256 | 30,556 | 90,024 | 0 |
| 500 | 30,557 | 30,256 | 30,556 | 90,024 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| 500 | 30,557 | 30,256 | 30,556 | 90,024 | 0 |
| 840 | 32,739 | 39,450 | 38,876 | 45,032,308 | 37,658,713 |
| 1,900 | 60,022 | 68,111 | 64,814 | 65,437,468 | 22,965,764 |
| 1,900 | 76,391 | 68,111 | 64,814 | 68,562,082 | 134,033,976 |
| 3,144 | 159,343 | 101,748 | 95,255 | 3,317,230,332 | 4,107,312,760 |
| 2,600 | 72,026 | 87,038 | 81,943 | 225,369,002 | 98,345,164 |
| 3,200 | 87,304 | 103,262 | 96,625 | 254,639,705 | 86,874,557 |
| 5,800 | 174,609 | 173,563 | 160,247 | 1,093,099 | 206,258,302 |
| **Total** | 1,001,130 | 1,001,144 | 947,289 | 4,225,283,064 | 4,738,526,887 |

| **MAD** | | | | **MAPE, %** | |
|---|---|---|---|---|---|
| **Area (ft²)** | **Least Square** | **Quantile (Median)** | | **Least Square** | **Quantile (Median)** |
| 712 | 6,927 | 6,682 | | 23.84 | 22.99 |
| 1280 | 1,048 | 657 | | 2.08 | 1.31 |
| 6103 | 14,091 | 3 | | 8.4 | 0 |
| 500 | 300 | 1 | | 0.98 | 0 |
| 500 | 300 | 1 | | 0.98 | 0 |
| 500 | 300 | 1 | | 0.98 | 0 |
| 840 | 6,711 | 6,137 | | 20.50 | 18.74 |
| 1,900 | 8,089 | 4,792 | | 13.48 | 7.98 |
| 1,900 | 8,280 | 11,577 | | 10.84 | 15.16 |
| 3,144 | 57,595 | 64,088 | | 36.15 | 40.22 |
| 2,600 | 15,012 | 9,917 | | 20.84 | 13.77 |
| 3,200 | 15,957 | 9,321 | | 18.28 | 10.68 |
| 5,800 | 1,046 | 14,362 | | 0.60 | 8.23 |
| **Average** | 10,435 | 9,811 | **Average** | 12.15 | 10.70 |

**Table 3.3:** Summary of FRP Repair Unit Cost ($/Ft²) Obtained For Oregon Highway Bridges

| **FRP Area, ft²** | **Actual FRP Cost, $** | **Actual FRP Unit Cost, $/ft²** | **Pred. FRP Unit Cost (LS), $/ft²** | **Pred. FRP Unit Cost (Q (M)), $/ft²** |
|---|---|---|---|---|
| 712 | 29,062 | 40.82 | 50.55 | 50.20 |
| 1,280 | 50,299 | 39.30 | 40.11 | 38.78 |
| 6,103 | 167,665 | 27.47 | 29.78 | 27.47 |
| 500 | 30,557 | 61.11 | 60.51 | 61.11 |
| 500 | 30,557 | 61.11 | 60.51 | 61.11 |
| 500 | 30,557 | 61.11 | 60.51 | 61.11 |
| 840 | 32,739 | 38.98 | 46.96 | 46.28 |
| 1,900 | 60,022 | 31.59 | 35.85 | 34.11 |
| 1,900 | 76,391 | 40.21 | 35.85 | 34.11 |
| 3,144 | 159,343 | 50.68 | 32.36 | 30.30 |
| 2,600 | 72,026 | 27.70 | 33.48 | 31.52 |
| 3,200 | 87,304 | 27.28 | 32.27 | 30.20 |
| 5,800 | 174,609 | 30.10 | 29.92 | 27.63 |

**Table 3.4:** Summary of Average FRP Unit Cost ($/Ft$^2$) Obtained For Respective Area Range of Oregon Bridges

| FRP Area Range, ft$^2$ | Average* FRP Unit Cost, $/ft$^2$ (LS) | Average* FRP Unit Cost, $/ft$^2$ (Q (M)) | Average* FRP Unit Cost, $/ft$^2$ (Actual) |
|---|---|---|---|
| <2000 (8) | 43.50 | 42.49 | 41.83 |
| >2000 (5) | 31.05 | 28.87 | 31.70 |

*The area of the respective bridges was considered as weight in calculating the average FRP unit cost

### 3.2 California

There were a total of seven bridges (Table 3.5) for which the contract data was obtained which had undergone FRP repair and rehabilitation work in California. These bridges were spread across Alameda County, Santa Barbara County, Tulare County, Los Angeles County, San Bernardino County, Riverside County and Imperial County. There were multiple contract bids for each bridge. Table 3.5 gives the Contract number, FRP cost, FRP area, Total contract cost and other calculated parameters.

**Table 3.5:** Summary of Year, Accepted FRP Cost and Accepted Unit Cost for California Highway Bridges

| Location | Contract No. | Date | Area, ft$^2$ | Accepted Unit Cost, $/ft$^2$ | Accepted FRP Cost, $ | Total Contract, $ | %FRP Cost |
|---|---|---|---|---|---|---|---|
| CA | 06-0N8504 | 2012 | 170 | 98.61 | 16,763 | 735,961 | 2.28% |
| | 11-264004 | 2010 | 420 | 84.40 | 35,449 | 995,644 | 3.56% |
| | 08-472304 | 2008 | 614 | 62.01 | 38,075 | 39,726,300 | 0.10% |
| | 10-0W0704 | 2012 | 2,620 | 70.81 | 183,974 | 2,559,279 | 7.19% |
| | 08-0G4804 | 2011 | 3,375 | 95.00 | 320,625 | 1,482,509 | 21.63% |
| | 06-0C1304 | 2012 | 3,677 | 65.29 | 238,053 | 3,352,320 | 7.10% |
| | 08-0M94U4 | 2012 | 4163 | 60.12 | 248,157 | 24,155,220 | 1.03% |

Note: All the data has been adjusted for inflation to 2013 dollar values

The California bridge data (Table 3.5) was mainly categorized into FRP repair work. FRP area is the variable which was found to have a significant correlation with the total FRP cost. The regression equations obtained by Least Square and Quantile (Median) methods for all the data included are:

Accepted FRP Cost (LS-$) = 7,202 + 68.53*FRP Area(ft$^2$) (13)

Accepted FRP Cost (Q (M)-$) = 6,036 + 63.10*FRP Area(ft$^2$) (14)

The coefficient of determination values obtained are 0.898 (LS, $R^2$) and 0.806 (Q (M), $R^*$) respectively. Scatter plot (Figure 3.3) is plotted between the accepted FRP cost and FRP area for all the bridges in California.

Bridges with accepted FRP cost of less than 1% of the total contract cost are omitted in this paper. Contract No. 08-472304 (Table 3.5) was found to have value of 0.10% and was not considered in the study to be consistent with the Oregon study, but it did not appear to be an outlier in the California data. Contract No. 08-0G4804 is a seismic retrofit, which had a different repair procedure than the rest of the bridges and much higher unit costs. Hence it was not considered in the analysis. The regression equations are computed again whithout these two bridges.

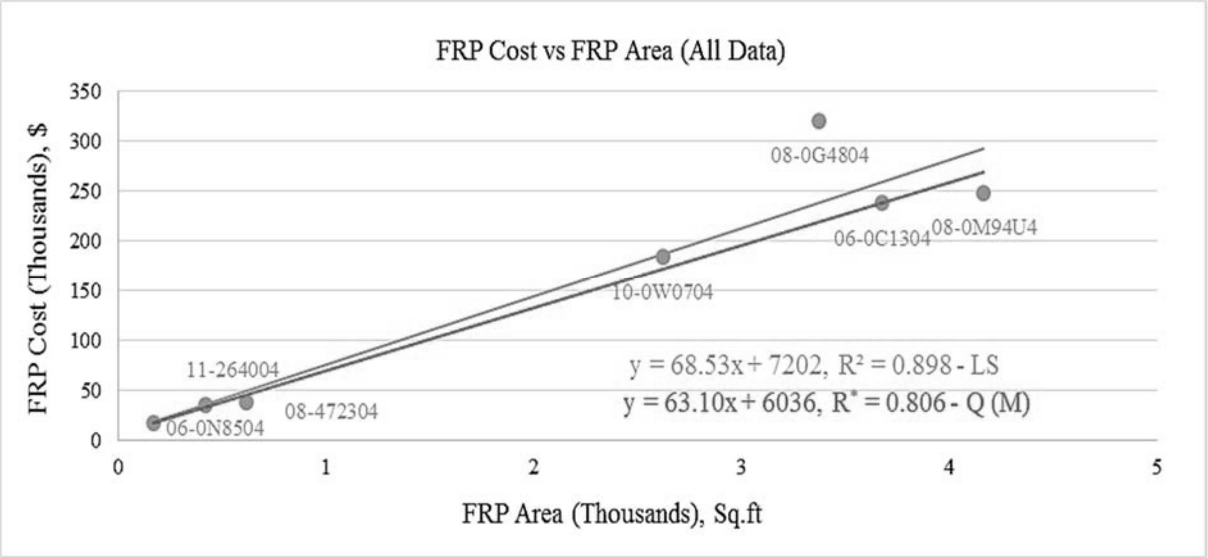


**Figure. 3.3:** Scatter Plot (All Data) of Accepted FRP Cost and FRP Area of California Highway Bridges (LS, Q (M)) with $R^2$ = 0.898 & $R^*$ = 0.806

The regression equation obtained for accepted FRP cost against the FRP area for the bridges in California by LS regression method with an $R^2$ value of 0.991 is given by:

Accepted FRP Cost (LS-$) = 11,609 + 60.12*FRP Area($ft^2$) (15)

The regression equation obtained by Quantile (M) regression method for California bridges with $R^*$ value of 0.926 is given by:

Accepted FRP Cost (Q (M)-$) = 9,322 + 62.20*FRP Area($ft^2$) (16)

Figure 3.4 shows the scatter plot of accepted FRP cost against FRP area of California bridges by the two analysis methods (LS, Q (M)) without the bridges - 08-472304 & 08-0G4804. The high slope value of 62.20 for Quantile (M) and 60.12 for Least Square is due to higher FRP cost associated with the repairs. The slope difference (absolute) between the CERs of Figure 3.3 and Figure 3.4 is 0.90 (Q (M)) and 8.41 (LS) which indicates the robustness of Q (M) regression with respect to outliers.

The results from Table 3.6 show that Least Square regression method has the least SSE value and Quantile (M) regression method has the least MAD and MAPE values. The Quantile (M) regression method reduced the MAPE value from 10.01 to 4.25 than that of Least Square regression method. The major difference between the two methods is the better prediction of the Quantile method for the smallest area.

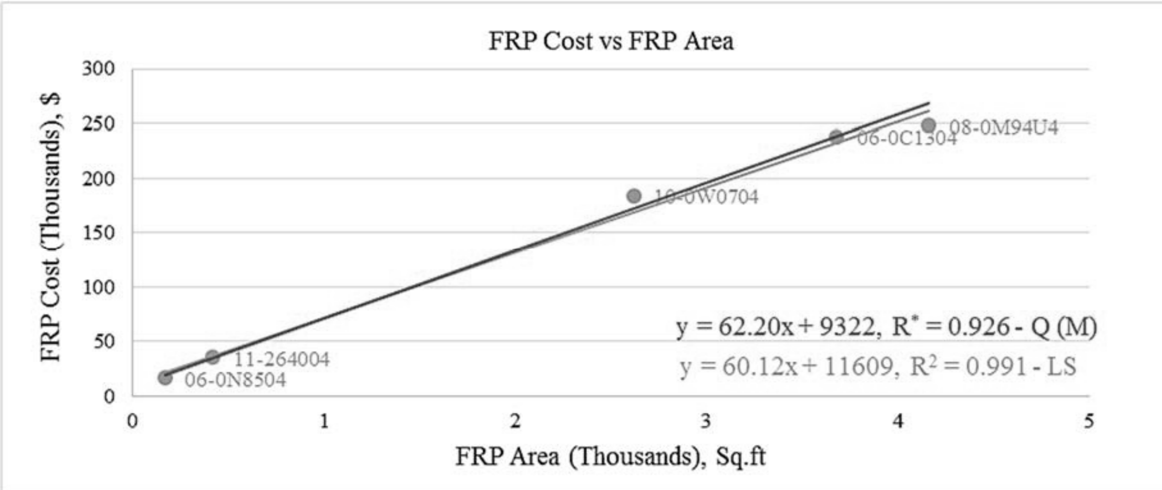


**Figure 3.4:** Scatter Plot of Accepted FRP Cost and FRP Area of California Highway Bridges (LS, Q (M)) with $R^2 = 0.991$ & $R^* = 0.926$

**Table 3.6:** Residual Analysis Table for California Highway Bridges

| Area, ft² | Accepted FRP Cost, $ | Predicted FRP Cost, $ | | SSE (X-Xp)^2 | |
|---|---|---|---|---|---|
| | | Least Square | Quantile (Median) | Least Square | Quantile (Median) |
| 170 | 16,763 | 21,830 | 17,898 | 25,676,719 | 1,287,365 |
| 420 | 35,449 | 36,861 | 35,449 | 1,992,953 | 0 |
| 2,620 | 183,974 | 169,129 | 172,302 | 220,370,462 | 136,228,091 |
| 3,677 | 238,053 | 232,678 | 238,054 | 28,889,830 | 1 |
| 4,163 | 248,157 | 261,897 | 268,286 | 188,797,658 | 405,183,847 |
| **Total** | 722,396 | 722,396 | 731,989 | 465,727,621 | 542,699,304 |
| **Area, ft²** | **MAD** | | | **MAPE, %** | |
| | **Least Square** | **Quantile (Median)** | | **Least Square** | **Quantile (Median)** |
| 170 | 5,067 | 1,135 | | 30.23 | 6.77 |
| 420 | 1,412 | 0 | | 3.98 | 0 |
| 2,620 | 14,845 | 11,672 | | 8.07 | 6.34 |
| 3,677 | 5,375 | 1 | | 2.26 | 0 |
| 4,163 | 13,740 | 20,129 | | 5.54 | 8.11 |
| **Average** | 8,088 | 6,587 | **Average** | 10.01 | 4.25 |

A relationship between the accepted FRP cost and the FRP area to predict the unit cost ($\$/ft^2$) values, can be established. The predicted unit costs obtained from Equations 15 & 16 are given by:

Unit Cost (LS-$\$/ft^2$) = 60.12 + 11,609/FRP Area($ft^2$) (17)

Unit Cost (Q (M)-$\$/ft^2$) = 62.20 + 9,322/FRP Area($ft^2$) (18)

Table 3.7 shows the predicted FRP unit cost ($\$/ft^2$) values obtained by the two methods (LS & Q (M)) of analysis for individual bridges. The actual FRP unit costs versus predicted FRP unit costs have good agreement overall, but the predicted equations are approximate only. There is the assumption that the start-up costs, which is indicated by the intercept of the curve, is constant for all project sizes. The fit is better for the California equations as the intercept values are small than those for the Oregon equations. When the bridges are grouped into two ranges as done in Table 3.8, the agreement is much better for the unit costs.

**Table 3.7:** Summary of FRP Repair Unit Cost ($/Ft$^2$) Obtained for Respective Areas of California Highway Bridges

| Area, ft$^2$ | Actual Cost, $ | Actual FRP Unit Cost, $/ft$^2$ | Pred. Unit Cost, $/ft$^2$ (LS) | Pred. Unit Cost, $/ft$^2$ (Q (M)) |
|---|---|---|---|---|
| 170 | 16,763 | 98.61 | 128.41 | 105.28 |
| 420 | 35,449 | 84.40 | 87.76 | 84.40 |
| 2,620 | 183,974 | 70.22 | 64.55 | 65.76 |
| 3,677 | 238,053 | 64.74 | 63.28 | 64.74 |
| 4,163 | 248,157 | 59.61 | 62.91 | 64.45 |

Table 3.8 shows for California bridges the FRP unit cost decreased as the area increased and the average FRP unit cost is approximately $64/ft$^2$ for areas greater than 2,000 ft$^2$.

**Table 3.8:** Summary Average FRP Unit Cost ($/ft$^2$) Obtained for Respective Area Range of California Bridges

| FRP Area Range, ft$^2$ | Average* FRP Unit Cost, $/ft$^2$ (LS) | Average* FRP Unit Cost, $/ft$^2$ (Q (M)) | Average* FRP Unit Cost, $/ft$^2$ (Actual) |
|---|---|---|---|
| <2000 | 99.47 | 90.42 | 88.49 |
| >2000 | 63.45 | 64.88 | 64.07 |

*The area of the respective bridges was considered as weight in calculating the average FRP unit cost

## 4 . Conclusions

There is a need for an effective cost estimation methodology in the transportation industry especially in the case of construction and rehabilitation of bridges. A brief introduction of Least Square regression and Quantile (M) regression was presented. The data of Oregon and California highway bridges was used to develop linear models for cost estimation of FRP wraps. The linear models obtained were used to compare the properties of the two methods based on graphical analysis and three test statistics (Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE) & Sum of Square of Errors (SSE)). After the comparisons, the Quantile (M) regression had better properties in two aspects:

- Quantile (M) regression was found to have lower MAD and MAPE values, thus indicating the median improved the proximity of the CER to the data (Tables 3.2 & 3.6).
- Graphical analysis of results showed Quantile (M) regression to be robust to outliers. One or more outliers far away from the rest of the data had minimal impact on the predicted values. Quantile (M) regression did not favor high-cost values to the detriment of low-cost values as illustrated in Figures (3.1, 3.2, 3.3 & 3.4). The slope difference (absolute) obtained between the CERs of Figures (3.1 & 3.2) of Oregon bridges is 0 (Quantile (M)) and 11.84 (Least Square) and for California bridges (i.e. between CERs of Figures (3.3 & 3.4)) is 8.41 (LS) and 0.90 (Q (M)) respectively. The low slope difference value for CERs of Quantile (M) regression validates the robustness of the method.

Based on the observations made in this study Quantile (Median) regression would be the preferred method of analysis when the data is large and scattered (Figure 3.1). When the mean and median are closer the differences become small between the two methods as indicated by Figure 3.4.

The analysis of the Oregon and California data was focused on developing relationships between FRP cost and FRP area by using Least Square and Quantile (M) analysis methods to predict the unit cost of FRP repair work. The high coefficient of determination values (0.884 (LS) & 0.773(Q) for Oregon, and 0.991(LS) & 0.926 (Q) for California) indicate FRP area to have a significant influence on the FRP cost. The FRP units cost curves are better predictors when the intercepts of the linear FRP cost curves are small. The coefficient of determination values of Quantile (M) regression are lower than that of Least Square regression since the metric $R^2$ minimizes the sum of square of errors, which is an inherent property of Least Square regression. The effect of outliers is much less with the Quantile (M) regrerssion than with the Least Square regression.

**REFERENCES**


[1] 3f. (2014, november 04). *3f-fr*. Retrieved November 13, 2014, From 3f-fr: http://3f-fr.com/?lang=en.

[2] Barrodale, I. and Roberts, F. D. K., "Solution of an overdetermined system of equations in the L1 norm." *Communications of the Association for Computing Machinery*, 17, pp 319 320. 1974

*[3]* Buhai, S. "Quantile regression: Overview and selected applications," *Ad-Astra-The Young Romanian Scientists' Journal, 4,* 2004

[4] Foussier, P. "Improving CER Building:Basing A CER On The Median". *Journal Of Cost Analysis And Parametrics*, 1-3.24, 2010

[5] Hald, A. "A History Of Parametric Statistical Inference From Bernoulli To Fisher,1713-1935," *Choice Reviews Online* , 44-6209-44-6209.25., 2007

[6] Koenker, R. . *Quantile Regression*. Cambridge: Cambridge University Press, 1-30, 2005

[7] Koenker, R. W., & d'Orey, V. "Algorithm AS 229: Computing regression quantiles," *Applied Statistics*, 383-393, 1987

[8] ORDINARY LEAST SQUARES REGRESSION (Social Science). (n.d.). Retrieved June 30, 2014, From http://what-when-how.com/social-sciences/ordinary-least-squares-regression-social-science/.

[9] Ordinary Least Squares Regression. (2008). In W. A. Darity, Jr. (Ed.), *International Encyclopedia of the Social Sciences* (2nd ed., Vol. 6, pp. 57-61). Detroit: Macmillan Reference USA.

[10] Xia, S. *The Multicollinearity Problem-A Comparison Of Ordinary Least Squares And Some Alternative Regression Methods*. Ann Arbor: Copyright Proquest, UMI Dissertations Publishing 2009.


**Appendix 1:** Identifying The Outliers in Oregon Highway Bridges Data

| Year | Bridge No. | FRP Area, $ft^2$ | Accepted Total FRP Cost, $ | Predicted FRP Cost (ALL Points), $ | Deviation 1 | Deviation 2 | Relative Variation % | $R^2$ (With Data Point Eliminated) 0.278 |
|---|---|---|---|---|---|---|---|---|
| 2007 | 01786A | 3,144 | 159,343 | 105,282 | 54,062 | 57,829 | 6.5 | 0.281 |
| 2008 | 07601B | 712 | 29,062 | 68,332 | -39,270 | -43,855 | -7.91 | 0.251 |
| 2008 | 08381S | 1,280 | 50,299 | 76,962 | -26,662 | -29,094 | -4.19 | 0.265 |
| 2008 | 7458* | 1,955 | 190,020 | 87,217 | 102,803 | 137,545 | -59.93 | 0.352 |
| 2008 | 08383N | 6,103 | 167,665 | 150,238 | 17,427 | 21,641 | 7.27 | 0.23 |
| 2009 | 8843* | 2,345 | 250,480 | 93,142 | 157,337 | 168,017 | 18.42 | 0.449 |
| 2009 | 03173A* | 9,204 | 124,069 | 197,351 | -73,282 | -159,297 | -148.37 | 0.429 |
| 2010 | 7530 | 500 | 30,557 | 65,111 | -34,555 | -39,010 | -7.69 | 0.248 |
| 2010 | 7532 | 500 | 30,557 | 65,111 | -34,555 | -39,010 | -7.69 | 0.248 |
| 2010 | 7534 | 500 | 30,557 | 65,111 | -34,555 | -39,010 | -7.69 | 0.248 |
| 2010 | 06945A | 840 | 32,739 | 70,277 | -37,538 | -41,671 | -7.13 | 0.254 |
| 2010 | 04981A | 1,900 | 60,022 | 86,381 | -26,360 | -28,310 | -3.36 | 0.273 |
| 2010 | 02233A | 1,900 | 76,391 | 86,381 | -9,990 | -10,730 | -1.27 | 0.275 |
| 2010 | 04979A | 2,600 | 72,026 | 97,017 | -24,991 | -26,659 | -2.88 | 0.28 |
| 2010 | 02236A | 3,200 | 87,304 | 106,132 | -18,828 | -20,156 | -2.29 | 0.282 |
| 2010 | 7392 | 5,800 | 174,609 | 145,634 | 28,975 | 35,016 | 10.43 | 0.227 |

*Outliers