Appendix

**A. Candidate models**

Table 1. Candidate Models for Cost Estimation

|  |  |  |  |
| --- | --- | --- | --- |
| Dependent Variable: Average Cost per Finished Block | | | |
| Predictors | (1) | (2) | (3) |
| Units |  |  | -0.0058 |
|  |  |  | (0.0016) |
|  |  | 833.63  (153.34) |  |
| ln(Units) | -2.69 |  |  |
|  | (0.59) |  |  |
| Weight | 5.55 | 5.56 | 5.54 |
|  | (0.34) | (0.34) | (0.34) |
| Complexity | 1.84 | 1.84 | 1.84 |
| (stamp + chisel operations) | (0.06) | (0.06) | (0.06) |
| Intercept | 33.50 | 15.03 | 19.85 |
|  | (3.87) | (1.67) | (1.77) |
| Summary Statistics and Joint Tests | | | |
| PRESS statistic | 12840.28 | 12604.65 | 13027.12 |
| Adjusted R2 | 0.6972 | 0.7021 | 0.6929 |
| AIC | 3033.62 | 3025.49 | 3040.65 |
| BIC | 3054.67 | 3046.54 | 3061.70 |
| F-statistic | 382.5 | 391.5 | 374.8 |
| n | 498 | 498 | 498 |

Notes:

(1) Standard errors are in parentheses.

(2) All of coefficients are statistically significant at the α = 0.001 significance level.

After dropping the 19th and 311th observations[[1]](#footnote-1), 498 observations are used to derive the models. The only difference between these three models is transformation of the unit term[[2]](#footnote-2). The first model uses natural logarithm of units, the second model uses the inverse of units and the third model use units in its original form. Although the three models are significant at all levels, their statistics are slightly different. Since estimation model will be used to estimate the cost for further orders, PRESS statistic is used as criteria to evaluate the models. Among these models, the second model has the smallest PRESS statistic, AIC, BIC and largest adjusted R2. However, it is hard to interpret the meaning of the inverse term of units. Meanwhile, the logarithmic term of units shows how much average cost changes due to the percentage change in number of units produced, holding other variables constant[[3]](#footnote-3). Due to the negative sign of the logarithmic term, the more units that customers order, the less production cost per units. Hence, knowing the change in cost when number of units change, the sale representatives can persuade customers to order more units to get better price as well as increase revenue for the company. Therefore, the first model is the best one among the three.

Table 2. Candidate Models for Cost Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Dependent Variable: Average Cost per Finished Block | | | |
| Predictors | (1) | (2) | (3) |
| Units |  |  | -0.006 |
|  |  |  | (0.001) |
|  |  | 819.24  (122.57) |  |
| ln(Units) | -2.86  (0.47) |  |  |
| Material Cost | 1.06  (0.04) | 1.05  (0.04) | 1.06  (0.04) |
| Labor | 11.15  (1.54) | 10.90  (1.53) | 11.06  (1.56) |
| Complexity  (stamp + chisel operations) | 1.08  (0.08) | 1.09  (0.08) | 1.08  (0.08) |
| Devon  (if Devon is manager on duty) | -3.73  (0.47) | -3.72  (0.47) | -3.73  (0.47) |
| Intercept | 24.04 | 4.94 | 9.70 |
|  | (3.34) | (1.96) | (2.01) |
| Summary Statistics and Joint Tests | | | |
| PRESS statistic | 8217.3 | 8086.0 | 8396.8 |
| Adjusted R2 | 0.8071 | 0.8098 | 0.803 |
| AIC | 2811.11 | 2803.98 | 2821.64 |
| BIC | 2840.58 | 2833.46 | 2851.12 |
| F-statistic | 416.9 | 424.3 | 406.1 |
| n | 498 | 498 | 498 |

Notes:

(1) Standard errors are in parentheses.

(2) All of coefficients are statistically significant at the α = 0.001 significance level

Similarly, due to the benefit of knowing the change in average cost due to the percentage change in number of units produced, the first model is the best among the above three to analyze production costs. This information is very important in order to optimize company’s profit by minimizing production costs and maximizing revenue.

**B. Transformations**

### In Figure 1, when the marginal cost is diminishing, average cost curve is downward slopping but not at constant rate. That means units and average cost per unit are not linearly correlated. Thus, units must be transformed. Using [Tukey’s Bulge Rule](http://www.r-bloggers.com/tukey-and-mostellers-bulging-rule-and-ladder-of-powers/), there are three potential transformations which are inverse square root, logarithm and inverse.

### Figure 2 shows how the three transformations of units fit to the plot of units against average costs. Logarithmic and inverse transformations have better fits than inverse square root. However, as discussed in part A, logarithmic transformation is used in both estimation and analysis models.

### Compare to the plot of average cost against initial form of units in Figure 2, the linear relationship between the two variables is more obvious after the transformation.

Figure 1. Average and Marginal Cost Curves

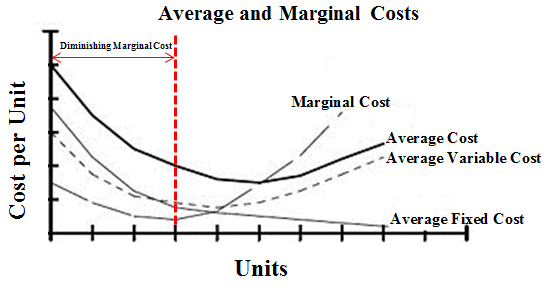


Figure 2. Plot of Average Cost against Units with Fitted Curves

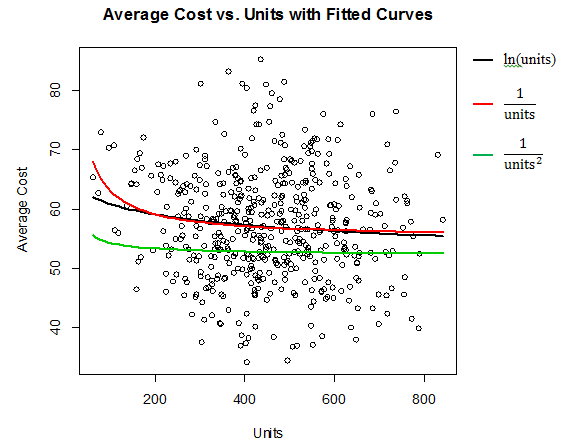


Figure 3. Plot of Average Cost against ln(Units) with Fitted Line



**C. Outliers**

Since the data is random, outliers are expected. Table 3 and 4 show the potential leverage points that need more attention. Although observations 19 and 311 do not have large hii nor having Cook’s distance greater than 1, they have the largest DFFITS, DFBETAS and COVRATIO in term of absolute values. In addition, in the plots of residuals against fitted values and time order, their residuals lie too far away from other observations. Taking a closer look into the data set, average cost of these two observations is abnormally low. While the mean of average cost is 56.98, the average costs of observations 19 and 311 are 6.74 and 5.50, respectively. That happens because they have unusually low total costs. Observation 19 has 499 units and total cost of 3362.97 and observation 311 has 674 units and total cost of 3706.08. With number of units in the range of 450 to 700, the total costs approximately vary from 30,000 to 40,000. It looks like there are errors while recording total costs of these two observations. Thus, eliminating these variables will help to improve the models.

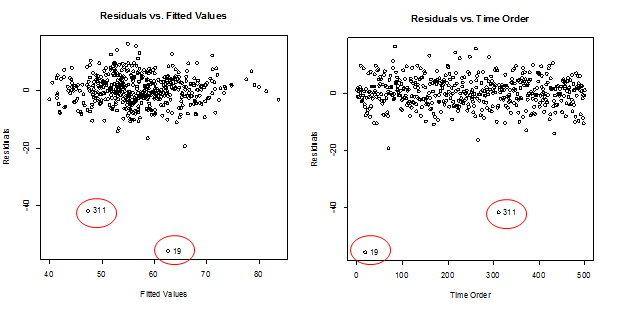
Table 3. Outliers in Cost Estimation Model

|  |  |
| --- | --- |
|  | **Potential Leverage / Influential Points** |
| hii | 400, 204, 149, 21, 215, 123, 227, 453, 487, 347, 515, 489, 216, 94, 350,326, 247, 465, 301, 408, 473, 179, 342, 90, 187, 304, 93, 134, 162, 196, 72, 124. |
| DFFITS | **311**, **19**, 216, 70, 400, 204, 266, 149, 84, 290, 159, 433, 226, 37, 248, 342, 42. |
| DFBETAS | **19**, **311**, 70, 248, 266, 140, 149, 152, 162, 204, 216, 247, 290, 306, 307, 326, 400, 42, 421, 430, 433, 453, 480, 94, 102, 123, 15, 159, 169, 172, 176, 199, 200, 226, 244, 275, 30, 325, 342, 350. |
| COVRATIO | **19**, **311**, 70, 84, 266, 261, 433, 372, 350, 167, 179, 94, 93, 304, 408, 90, 301, 465, 453, 489, 347, 151, 487, 123, 149, 227, 215, 21, 204, 400. |
| Cook’s Distance | Although all of observations have Cook’s distance less than 1, which means there is no potential leverage points by using this criteria, observations **311** and **19** still have largest Cook’s distance. |

Table 4. Outliers in Cost Analysis Model

|  |  |
| --- | --- |
|  | **Potential Leverage / Influential Points** |
| hii | 400, 204, 21, 149, 247, 215, 297, 123, 453, 126, 476, 301, 473, 227, 382, 151, 347, 480, 91, 487, 109, 358, 408, 489, 27, 159, 61, 17. |
| DFFITS | **311**, **19**, 216, 297, 266, 70, 312, 124, 59, 80. |
| DFBETAS | **311**, **19**, 279, 297, 169, 216, 218, 247, 266, 307, 317, 364, 70, 84, 124, 134, 149, 204, 215, 248, 261, 274, 342, 350, 372, 400, 404, 419, 422, 423, 430, 472, 487, 59, 80, 152, 159, 175, 179, 185, 187, 189, 195, 21, 244, 254, 258, 290, 291, 312, 349, 38, 381, 399, 42, 44, 448, 451, 475, 486, 496, 498, 53, 73, 91, 97, 99. |
| COVRATIO | **19**, **311**, 266, 216, 364, 261, 44, 486, 61, 151, 27, 358, 408, 489, 301, 480, 215, 227, 476, 382, 247, 123, 162, 473, 453, 149, 204, 21, 400. |
| Cook’s Distance | Although all of observations have Cook’s distance less than 1, which means there is no potential leverage points by using this criteria, observations **311** and **19** still have largest Cook’s distance. |

Figure 4. Residuals of 9 and 311 Compare with other Observations in Estimation Model

****

**D. Selection of Variables**

**D.1 Selection of Variables for Cost Estimation Model**

The coefficients table shows that only the first four variables are significant even as the last predicted added. That means the three categorical variables (goal.sd, rush and detail) do not help much in estimating average cost while units, weight, chisel and stamp already in the model. However, chisel and stamp variables are linearly correlated. Adding both of them as two separate variables to the model will cause multicollinearity. Hence, since number of chisels and stamps used indicate the complexity of the block produced, adding them together to become one variable called complexity will help to avoid multicollinearity and improve the model. Figure 6 shows the strong upward linear trend between weight and complexity. Therefore, together with units, weight and complexity are good predictors for cost estimation model.

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 34.1322 3.9792 8.578 < 2e-16 \*\*\*

units.log -2.7163 0.5851 -4.642 4.44e-06 \*\*\*

weight 5.4555 0.3420 15.953 < 2e-16 \*\*\*

stamp 2.2921 0.2734 8.385 5.47e-16 \*\*\*

chisel 1.4702 0.2323 6.329 5.60e-10 \*\*\*

goal.sd.cat0.5 -0.4924 0.6890 -0.715 0.475

goal.sd.cat1 0.8091 0.6648 1.217 0.224

rush.catYes 0.4039 0.5864 0.689 0.491

detail.catYes 0.8592 0.6359 1.351 0.177

****



**D.2 Selection of Variables for Cost Analysis Model**

****

****

Not by Devon By Devon

**E. Assumptions Checking**

**E.1 Assumptions Checking for Cost Estimation Model**

= 33.50 – 2.69\*ln(Units) + 5.55\*Weight + 1.84\*(Stamps + Chisels)

The QQ plot of the residuals shows a straight line. Thus, there is no problem with the normality assumption. Also, plots of the residuals against fitted values, time order and predictors do not have any patterns which means residuals are random with constant variance and not correlated with each other and with the predictors.







**E.2 Assumptions Checking for Cost Analysis Model**

= 24.04 – 2.86\*ln(Units) + 1.06\*Material Cost + 11.15\*Labor Hours + 1.08\*(Stamps + Chisels) + 1.84\*Devon

Similar to the estimation model, the QQ plot of the residuals and its plots against fitted values, time order and predictors show that the residuals are random normally distributed with constant variance and are not correlated with each other and with the predictors.









Not by Devon By Devon

1. Outliers will be discussed in part C of the appendix. [↑](#footnote-ref-1)
2. Transformations will be discussed in part B of the appendix. [↑](#footnote-ref-2)
3. In the first model, when number of units increase 1%, average cost will decrease by 0.0269 monetary units, holding other variables constant. [↑](#footnote-ref-3)