ESTIMATING CONTRACT TIMES FOR TRANSPORTATION PROJECTS: CREATING A STATISTICAL MODEL TO ESTIMATE TIMES USING BID QUANTITIES

CSCE, MAY 2017

GUILLERMO NEVETT

PAUL GOODRUM

UNIVERSITY OF COLORADO BOULDER

Presentation Agenda

- Problem statement.
- Current methods for time estimation
- Explain the development process for the statistical model.
- Path forward.

Problem Statement

Low accuracy → Lack of commitment

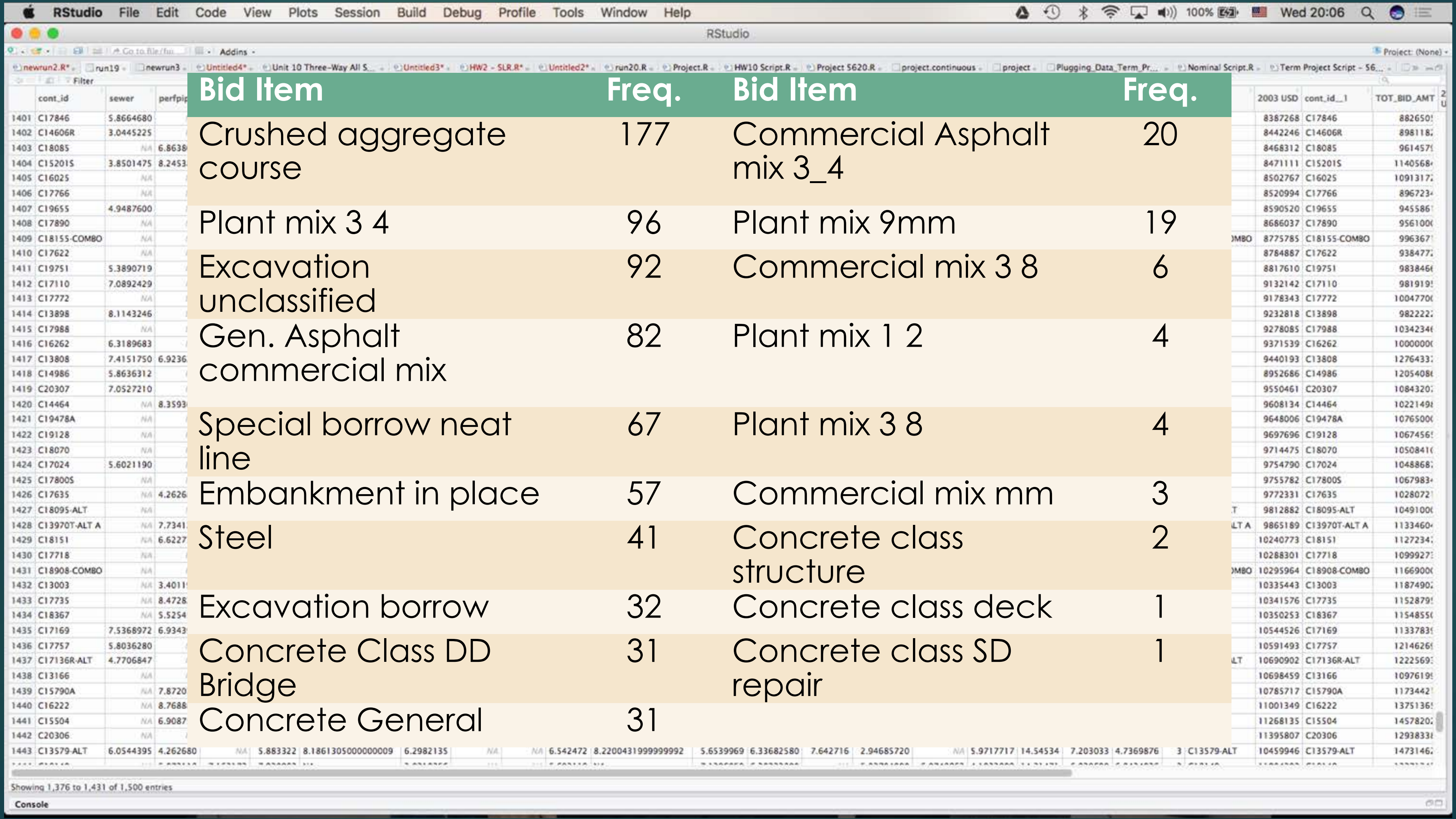
Approved FHWA Contract Time Determination Techniques

- Bar Charts Method
- Critical Path Method
- Estimated Cost Methods

First Steps

- Clean and Organize Data to get same units and grouping similar variables (e.g. piping)
- Convert Engineer's Estimate to 2015 USD using the NHCCI index (https://www.fhwa.dot.gov/policyinformation/nhcci.cfm)

The data



Bid Item	Freq.	Bid Item	Freq.
Crushed aggregate course	177	Commercial Asphalt mix 3_4	20
Plant mix 3 4	96	Plant mix 9mm	19
Excavation unclassified	92	Commercial mix 3 8	6
Gen. Asphalt commercial mix	82	Plant mix 1 2	4
Special borrow neat line	67	Plant mix 3 8	4
Embankment in place	57	Commercial mix mm	3
Steel	41	Concrete class structure	2
Excavation borrow	32	Concrete class deck	1
Concrete Class DD Bridge	31	Concrete class SD repair	1
Concrete General	31		

Multivariate Linear Regression

- MLR is similar to univariate linear regression, but instead of having one independent variable, there can be several estimators.
- Our dependent or response variable is charge days and the independent variables are bid quantities, EE, locations, and project types.

```
> reg.gen<-lm(chgdays ~ sewer + perfpipe + pvc + concpipe + classd + pavmark + muckexc + rockexc +
concrete + concpav + strback10 + emban100 + asphalt10+ unclexc10 + strexc10+ asphre10 + agg10 + lnu
sd10, data = newrun3)
```

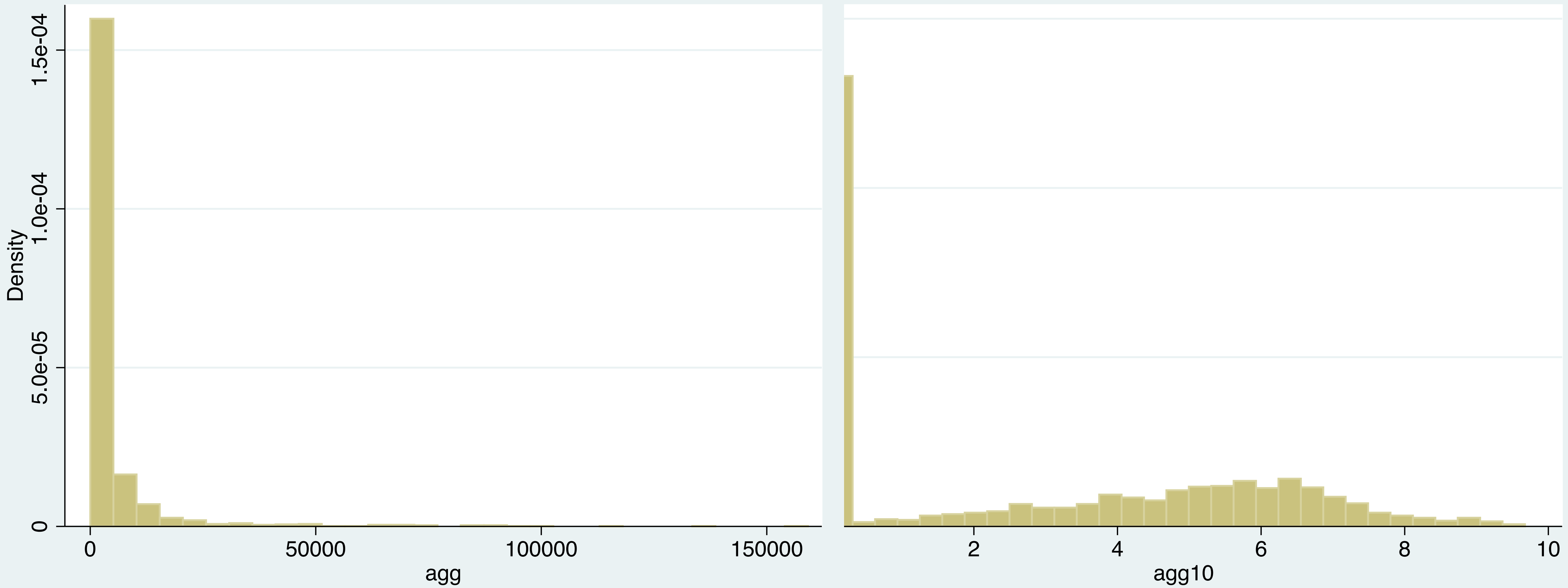
Unusual Observations.

```
##### Unusual Obervations  ################
#--> Outliers (Excessive Residuals, Y direction)
plot(rstudent(run19.lmo),pch=19)          # Plot the Studentized residuals in an Index plot
outlierTest(run19.lmo)                    # Actual statistical test for largest outlier
influenceIndexPlot(run19.lmo)             # As before
marginalModelPlots(run19.lmo, id.n=4)

#--> Leverage (Hat Values, X direction)
plot(hatvalues(run19.lmo))                # Plot of all Hat Values
leveragePlots(run19.lmo, id.n = 4)        # Specific Leverage plots by predictor (and fitted values)like avPlots()
influenceIndexPlot(run19.lmo, id.n=4)     # As before

#--> Influence  (Y and X direction simultaneously)
(crit.cd <- qf(.5, 3,39, lower.tail = FALSE))  # Critical F for assessing Cook's Distances
plot(cooks.distance(medical2.lmo), type="b")            # Plot of Cook's Distances
abline(h=crit.cd, col = "red", lwd=6)          # Cut off line (critical values) for Cook's D
influenceIndexPlot(fuel2.lmo, id.n=4)                # Again, as before
influencePlot(fuel2.lmo)                             # A new one and useful too
```

Variable Transformations



Variable Transformation (cont.)

- More advanced method
  - Suggests the power transformation for a variable in presence of all other variables, which is better than older methods that transformed a variable without considering others.

```
##### Tranformations #####
#predictor#
boxTidwell(chgdays ~ sewer + perfpipe + pvc + concpipe + classd + pavmark + muckexc + rockex

#response#
powerTransform(chgdays ~ sewer + perfpipe + pvc + concpipe + classd + pavmark + muckexc + ro
```

MLR model for time estimation

1. Select project type
2. Select Project Size*
3. Input bid quantities
4. Get Estimated Charge Days

e.g. Resurfacing

e.g. tons of asphalt paving

e.g. 50 charge days

*Based on a standardized Engineer's Estimate using the NHCCI index
https://www.fhwa.dot.gov/policyinformation/nhcci.cfm

Validation

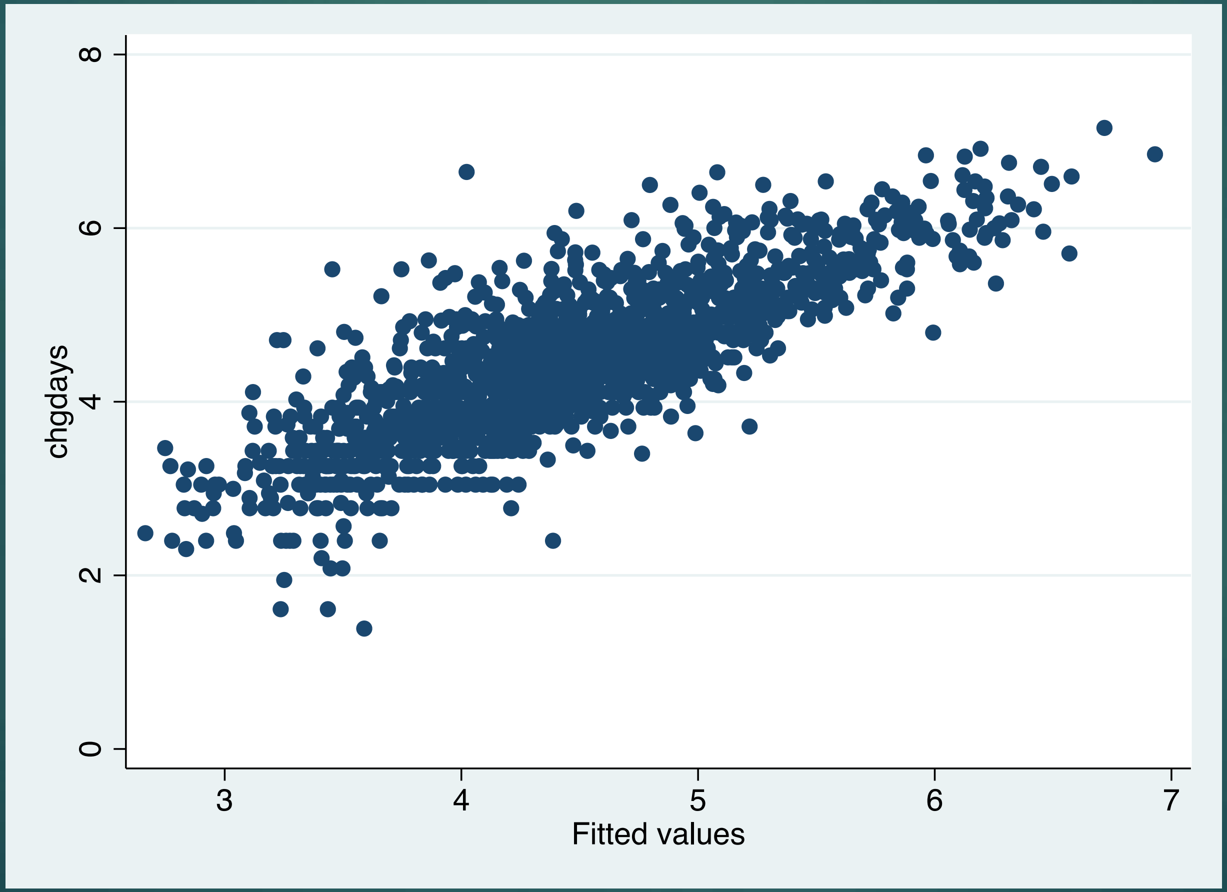
- Create the model using 80% of the data (R^2)
- Use the remaining 20% to validate (median percent error)
- Expl
- K-Fo

$$Percent\ Error = \frac{|Predicted\ Value - Observed\ Value|}{Observed\ Value} \times 100$$

$$Y = 44.532 + 9.253E - 6 * X_1 + .008 * X_2 + .001X_3 + 5.421E - 5 * X_5 + .002 * X_5 + \varepsilon^1$$

Goodness of fit F=121.354, significance =.000, Adjusted R^2=.746, Mean Percent Error= 44.59%, Median Percent Error= 29.54%.

Predicted vs. Actual Durations (log scale)



Path Forward

- Taylor et al. (2013) developed a similar tool for Kentucky Transportation Cabinet.
- Creating a multi-state time determination tool (CO, GA, MS, MT)

Path Forward (cont.)

- Develop a Machine Learning Estimation Model (analyzing Case-Based Reasoning and Neural Networks).

Questions?