**What’s the recommendation?**

* Use this for six months
* Switch evaluation metrics

**What did we do – what was the objective and general method?**

* Build a predictive model
* Used machine learning to select variables

**How’d we do that?**

* Statistical software includes ML packages that penalize regression
* Cross-validation picked the best model specification
* Follow-up regression used that selection to build the equation

**What did the method find?**

* The strongest and most trustworthy predictors of second-lowest bids are project size (particularly for big projects), bidding process, and (obviously) the engineering estimate.

**How does that compare with the existing method?**

* It’s complementary to the existing method. It uses it and builds on it.

**What’s the takeaway?**

* If you have big projects or projects that aren’t publicly bid, pad them.

**Demonstrate need for this project**

The average estimate missed the second-lowest bid by around 6 percent. But in absolute terms it missed by 18.4 percent. A predictive model with interpretable specifications lowers that to 17.6 percent:



Shrinking that variance would:

1.

2.

3.

**Outline next steps**

Programmatic:

I trained the model using more than four years of data.

Here’s what it did on six months’ of withheld observations:

So let’s do this again using Q4 and Q1 (I didn’t have that data and most isn’t done). Involve quickly re-training the model but parameters shouldn’t change much. We can’t tell the estimators or it could contaminate their work.

I’ve been searching for a rule of thumb where:

If it’s an SBE, add X%.

If it’s a big (above $15 million) project, add Z%.

Also, want to talk to others in Engineering and to PMO. Other process-related data sets out there we could use to develop insights. Your blessing.

And Engineering journal. Your blessing.

**Tonight 2019-11-20**

1. Set equation using data withholding 2Q.

2. Use equation to predict on two withheld quarters and compare results (fitted values) in table.

3. Replicate in original excel doc using equation.

**4. Find potential backup in UK Green Book if you can.**

*(Note: use the linear-linear coefficients for each, but apply them on a percentage scale. Meaning the coefficients relate to an average project size of [blank] for those two categories, so apply those percentage increases to the projects meeting those two respective conditions. This isn’t directly tied to the results but follows intuitively and is a bit more accessible and the logic easier to interpret.]*

<https://www.statalist.org/forums/forum/general-stata-discussion/general/1466163-back-transforming-predicted-fitted-values-in-ols>

<https://stats.stackexchange.com/questions/272499/exponentiate-the-predicted-value-for-a-log-transformed-response>

<https://www.dannyadam.com/blog/2011/05/predicting-y-when-the-dependent-variable-is-a-transformation-of-y/>