Free Bonus reprint from The Asset Analyst Report

WHY DOESN'T REGRESSION WORK?

GEORGE DELL, SRA, MAI, ASA, CRE

Originally published in "Appraisal Today"

Why Does Regression Not Work?

Regression is just a math formula. The regression line itself is middle-school. Y = a + bX, where

- Y is predicted (let's say the sale price)
- a is a constant (where the line crosses the vertical axis)
- b is a slope of the line (how price changes with size, let's say SqFt)

The right regression line is the one that best fits the data points. The custom is to find the line which minimizes the squared distances from that line to each data point. (Other ways are possible).

Usually, all appraisers care about is the slope of the line. This is *simple* regression. Multiple regression just looks at several X's at the same time. For example, besides building size, it may consider the site size, and whether it is next to a highway or not. Building size and site size are measure (continuous) variables, while the highway variable is a yes/no (category/discrete) variable. The formula still looks the same.

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

where the X_i's are building size, site size, and yes/no (0/1) numbers. The b_i's are the "coefficient" number. (Each b is multiplied times its X).

The b_i are similar, *but not equal to* what appraisers use as an adjustment. This difference between adjustment amounts and regression coefficients (the b_i's), is **problem #2** in why regression does not work, given the *thinking-training* appraisers have received. (We will get to **problem #1** another day, another blog. Hint -- it has to do with how we pick the data . . . ☺).

But what is the problem we are trying to solve? Regression can be used to solve for Y, or it can also be used to help us to determine, estimate, or approach the right adjustment in traditional comparison. This is a modeling decision! It is important to not pick up a hammer every time, even if we just want to cut the board in half . . . . Regression itself can sometimes be a hammer, and sometimes a saw, and sometimes just a tape measure.

To go further, we need to reconsider what we are after. It is important to know that regression can be used to help hunt down adjustments, as well as the value estimate!

1. We can use multiple regression alone, to directly estimate our Y, the predicted variable.
2. We can use multiple regression to help us determine or estimate adjustments.

In the second use, it is possible to use regression results in conjunction with traditional comparison. (Note that comparison is in all three approaches: cost, income, direct comparison). **Problem #2** requires we have to *first* make a modeling decision. Are we trying to hammer a nail, or saw a board?

The 'regression' model for determining, estimating, or approaching adjustments is a separate, but important topic. For now, what is important is that we have identified the problem (predict Y); we have decided that a useful model might be regression. The solution to this problem #2 has a number of sub-problems.

First, the analyst must decide what data to include in the formula (Problem #1).

Second, the analyst must decide what X's to include in the formula. Oh darn – another modeling decision! Can we just use the three predictors (X_I) building size (X_1), site size (X_2), and highway yes/no (X_3). Or should we also add other predictors, like distance to the city, age, quality . . . ? There are two ways to determine what predictor variables belong in a multiple regression. One is that the experienced appraiser chooses, based on being an expert in the area and in the property type. A second method is to use (computational) classifying tools. This second method is what is used by AVMs. It does not work so well, but works well enough to provide big competition for appraisers, who could otherwise be getting this work. (I can run a regression better than you can!)

So this raises another question: How can appraisers pick comps better than automated systems? The answer is that human judgment is quite good in certain things, like seeing patterns. Yet computer systems *can* pick comps. Is it possible that human appraiser expertise *combined with* computation tools and visualization methods could do better than either one alone?

If you believe as I do – that humans armed with the right tools, hammers, saws, and *power tools* can do a better job than either one alone, then you are in the right place. Is current appraiser education focused on helping the appraiser choose the right tool? Are most software vendors focused on helping the appraiser choose the right data, the right tool – or are they just selling "my regression is better than your regression"?

Regression does not work? It works every time! INPUT -> REGRESS -> OUTPUT. No matter what you put in, REGRESS regresses. It just does the "minimize the squared deviations" thing. It does it well. It does it fast. And it does it perfectly. OUTPUT outputs. So why does regression not work? There is only one thing left in this algorithm. The INPUT. But this was Problem #1.

For now we are dealing with problem #2 – Is regression the right model for the problem at hand? So far, we have a short list of modeling decisions an appraiser must make:

1. What data should I include?
2. What predictors should I include?
3. Should I be using regression on this problem?
4. If yes, should it be simple regression or multivariate regression?
5. How should I explain my use of regression and graphs in my report?

The above questions are basic. Actually, they are not all that hard. Appraisers make similar decisions on each and every appraisal assignment performed. The wording is different, and the problem has changed. The wording is, in order: What comps can I get? What rates/ratios and elements of comparison should I use? Should I use one of the methods suggested in *The Appraisal of Real Estate*? (Such as: paired comparison, grouped comparison, secondary data analysis, statistical analysis, including graphic analysis and scenario analysis, cost-related adjustments or capitalization of income differences). Note that "regression" is not in this list. Perhaps it is considered a sub-method of statistical analysis, taught in several courses and seminars.

What is *not* taught, are the real issues about the use of regression, the tool. The tool is easy. You do not need to memorize the formula. You do not need understand how it is derived. You do not need to

know how it is mathematically 'proved.' You do not even need to be able to explain it in a report. You do not need to explain the calculus involved in 'minimizing' the least squares. What you do need to do is know when to use it, what form of regression to use (simple or multivariate), what data you put in, what predictors to leave in and which to take out. And you need to be able to explain your competent appraiser judgment in each of these things. There is no magic.

So why does regression not work? But wait, there's more. Number 2 above, about using the right predictor variables (elements of comparison), needs another look. What if a critical feature/predictor is not measured or provided in the data source? What if it is noted, but not measured, like "spectacular view." How about predictor variables that interact with each other? And finally, how about those things which are non-linear? And in particular those predictors which are ordinal in nature? (You can tell better or worse, but no measure is possible). This post cannot cover each of these here. Yet they can be the weakest part of reliance on regression.

On the other hand, once an analyst understands about things such as missing variables, interaction variables, correlated variables, and ordinal variables – the appraiser brain can be put to excellent use.

The appraisal profession suffers from two things: over-simplification and over-complication. *Over-simplifying* is due to the idea that you can just throw in a bunch of data, push the button, and get a good answer. If it were that simple, you would be obsolete. The computation is simple. The modeling is not. A machine does the computation. A human is needed for modeling decisions. It need not be very complicated. It just needs to be explained simply. Appraisers are smart enough to make some difficult things work. Doing it a better way maybe easier than trying to explain the "too easy."

Over-complication is the terror inflicted by some statistical methods, imposed by those who may have never used them in practice. Or worse, used by those who wish to win through the clever *appearance* of sophisticated competence, rather than by simple truth and understandability. Much of the over-complication we see today is brought to us via "advanced methods" in appraiser education. The education jumps right to complex theoretical tools -- mostly inferential statistics. What has been neglected is the modeling and using the right tool for the right problem. The 'new' basics.

This last thought brings us right back to the original question. Regression works. Using it in the wrong model to solve the wrong problem is what causes issues.

The appraisal problem requires a predictive solution, not sample statistics. Inferential statistics is designed to help describe large populations of data by drawing random samples. A complex, difficult, but clever-looking "solution" to the wrong question. To learn or teach obsolete (and never needed) inferential statistics for appraisal work is a waste of time. Calling it an "advanced" method sells, and impresses some. But whatever happened to "beginning" and "intermediate" statistical modeling?

The modeling must come first. Then learn and teach and use the tools. Like regression.

For more, click here for GeorgeDell.com. Join a community of appraisers who care. Help act on our future, not react and be pushed around.

Free Reprint

Of an Original George Dell article

First published in Appraisal Today by Ann O'Rourke
Check out Ann's monthly paid newsletter at
AppraisalToday.com