

## Taking the Con out of Econometrics

In econometrics, there is an increasingly large trend of p-hacking, where economists run thousands of models with different combinations of variables to find significance in a certain study. This has led to a lack of credibility in economics, where every paper is suspect and some authors are using less credible methods to find significance. Machine learning can help with this issue, taking the con out of econometrics.

Using the example of a potential drug trial that has about a thousand participants, Susan Athey explains how machine learning can find usual results that may not appear in typical economic models. Using clustering by successes, Athey explains that this is how companies data mine, finding things that these people might have in common. Looking a small subgroup, you can mine the data enough to find significant results for a small population, like 15 out of 1000 participants who see positive results. While at the same time, 15 out of 1000 who had serious side effects would not derail production.

Looking at classic econometrics, measuring the effect of a rise in minimum wage would be to run a multiple regression with a dummy variable for higher minimum wage and control variables for any fixed effects. Machine learning can allow the model to explore the covariates, by using a more complex model, to only use the characteristics that matter. Some characteristics in this minimum wage model might be extremely correlated with each other, which do not parse out all the error effectively. Given the amount of data that exists today, it is very likely that the model might end up with more variables than observations, which can be problematic. These variables could interfere with the precision of the estimates, leading to inconclusive results, but some of these variables can be collapsed together to form more informative results. These machine learning models can reduce the trend of overfitting, that is all too common now in economics.