

# 9 billion regressions:

## A multiverse approach to statistical analysis

Researchers never fully know if the statistical method that they choose provides the best possible model for their data. Their choice of how to analyse the data also influences the results. Dr Cristobal Young, Associate Professor of Sociology at Cornell University is using multiverse analysis to offer a rigorous and transparent framework to tackle the intrinsic issues of model uncertainty.

Model uncertainty is a central challenge of modern science. Researchers do not know what model actually generated their data and they cannot be sure if the statistical method they use is best. Moreover, researchers can explore innumerable analytic variations in their search for results, but readers can only see the handful of estimates curated for publication. These researcher 'degrees of freedom' represent the crisis in science today: are empirical conclusions driven mostly by the data, or mostly by the analytic assumptions made by researchers? How much influence does a researcher actually have over the results?

Dr Cristobal Young, Associate Professor of Sociology at Cornell University, is addressing this conundrum. He believes that "the 'crisis in science' today is rooted in problems of model uncertainty and weak transparency." Dr Young demonstrates the leeway currently afforded to researchers and

authors, enabling them to choose their preferred results. He also offers methods to quantify the influence of the researcher's choice of statistical model.

### MODEL UNCERTAINTY

All empirical estimates require specific modelling assumptions, and the researcher has to choose which assumptions to employ. Decisions such as defining variables, cleaning data, the treatment of outliers, the calculation of standard errors, and the choice of functional form mean that competent researchers can choose different statistical methods and produce different results from the same dataset. Designing the analysis and specifying the model often pose the biggest challenges to researchers. Dr Young points out that the underlying issue here is model uncertainty, in that no one actually knows which model specification is best for a particular study. There are often many credible models that might offer a variety of results that the data could support.

This makes it difficult to say whether a particular conclusion would be reached by all reasonable researchers analysing a particular dataset, or if the empirical results are highly dependent on the researcher's particular choices. This underlines the importance of transparency in model specification.

### MULTIMODEL ANALYSIS

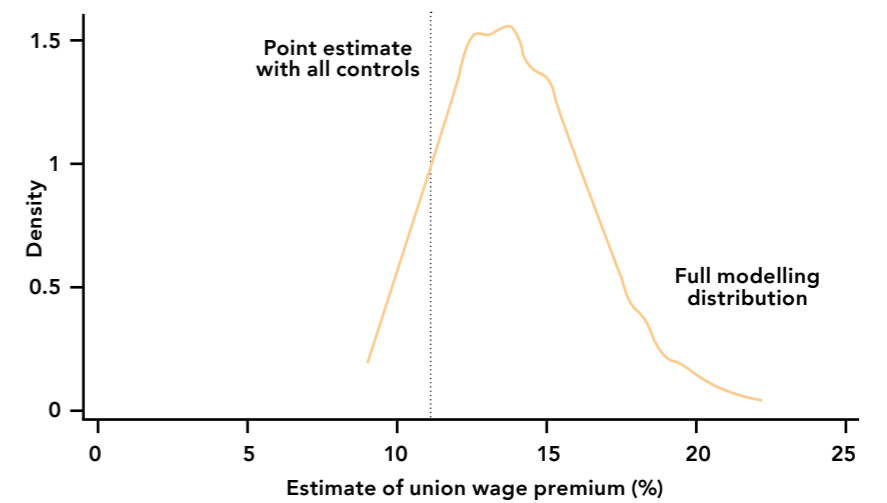
Rather than asking 'which is the best model?' Dr Young encourages researchers to ask 'which models deserve consideration?' Embracing a wider range of models allows authors, journal reviewers and readers a clearer view into how the conclusions rest on a credible, comprehensive, and robust foundation. Multimodel analysis acknowledges alternative combinations of modelling assumptions that different authors could use and offers significant improvements in the transparency and robustness of statistical research. The fundamental challenge in multimodel analysis, however, is defining the set of plausible models.

### THE MULTIVERSE OF MODEL ASSUMPTIONS

Dr Young has developed a computational framework for multimodel analysis. He explains that the goal is "to offer transparent and systematic reporting of alternative results that a researcher could obtain from other plausible model specifications." In a sense, there is a multiverse – multiple universes of analysis – in which all the credible ways of analysing a research question with a given data set all exist at once. A computational approach aims to reduce the author's discretion in choosing their preferred result from the multiverse, while also expanding the range of models and results under consideration.

The method involves specifying a set of plausible model components, such as control variables, variable definitions, estimation commands, functional forms, and standard error calculations, and estimating all possible combinations of these components. This can involve running thousands of small variations of model specifications with the results reported as a graphical distribution of estimates.

### A Robust Result: Modelling Distribution for the Union Wage Premium



Union members earn a large wage premium compared to non-union workers. The premium ranges from 10 to 20 percent depending on what controls are included in the model. However, all models show a substantial benefit of membership. Estimates from 1,024 models. Vertical line indicates the 11.1 percent union wage premium found when all controls are included.

In empirical testing, the multiverse shows three basic patterns of model robustness:

- In some cases, a statistical result holds regardless of how the model is specified. Any combination of the model components will generate the same result.
- Other times, a result depends on only one or two model components, for instance a particular control variable. This prompts follow-up analyses to find out why the control variable is so important.

employs large-scale random noise simulations to investigate the issue of excess false positive errors occurring from model uncertainty. Moreover, it shows how a multiverse analysis can identify and eliminate many false positives.

The simulation strategy involved conducting noise-on-noise regression analysis. This work involved completing 5,000 iterations of the simulation for each of the 17 unique conditions for data size and degree of model

## The 'crisis in science' today is rooted in problems of model uncertainty and limited transparency.

- Most worrisome, some results depend on a 'knife edge' specification, and are supported in only one in a 100 plausible models.

Authors are still free to present their preferred model, albeit within the context of a graph displaying other results that can be found with the set of acceptable model components.

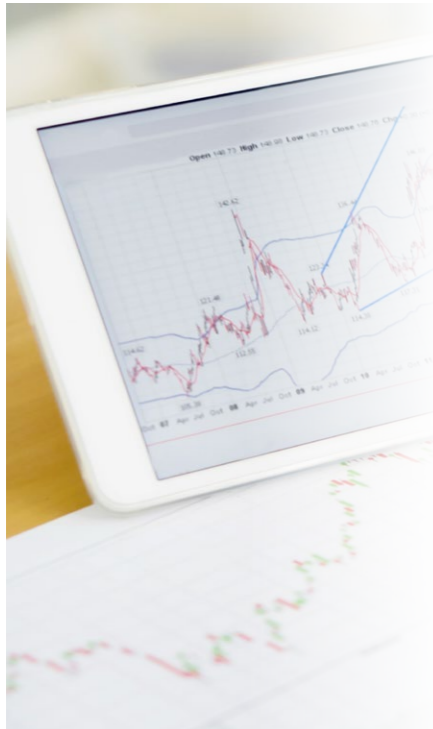
### 9 BILLION REGRESSIONS

False positives – statistically significant results – are common even when analysing data that are made entirely of random noise. Dr Young's research

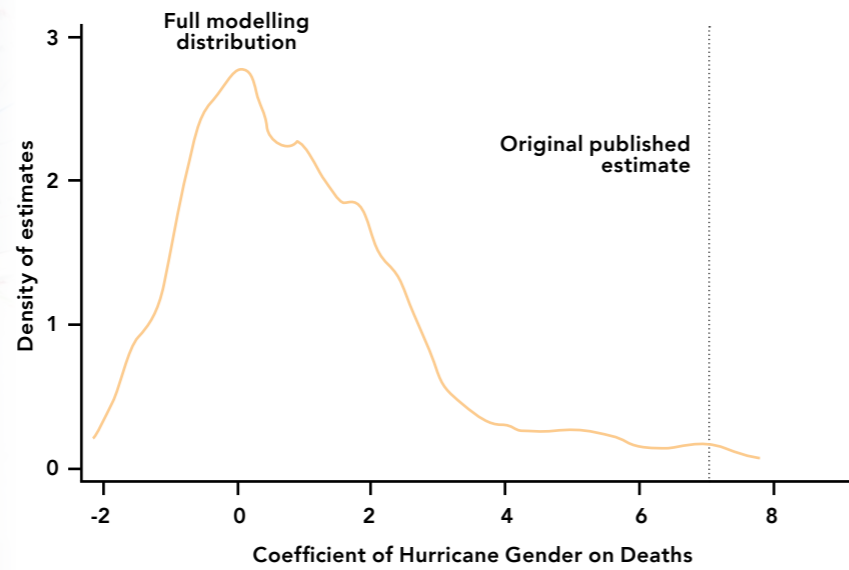
uncertainty. Approximately 550 million regressions were carried out for each condition – more than 9 billion in total. The simulations show that the proportion of false positives is highest when the sample size is low and model uncertainty is high. When the false positive rate is highest, multiverse analysis works best – routinely flagging 'significant' results as being non-robust and continent on using an extremely idiosyncratic model specification.

These simulations indicate that, first, to mitigate the proliferation of false positive results, researchers should





**A False Positive:** The effect of hurricane “gender” on death toll. The original published estimate showed that hurricanes with feminine-sounding names had higher death tolls. A multiverse analysis using 1,152 alternative models largely finds no effect.



aspire to use larger data sets. Second, mature research areas with strong, well-established prior theory – and less model uncertainty – will tend to produce fewer false positives. Finally, multiverse analysis provides a strong check against inflated rates of statistical significance in more novel research areas where there is greater uncertainty in the analysis.

#### MULTIVERSE IN THE REAL WORLD

The informative potential of multiverse analysis is illustrated by revisiting a controversial study reporting that hurricanes with female names are more deadly. Using data on all hurricanes making landfall in the U.S. over six decades, the study reports that hurricanes with feminine-sounding names consistently had higher death tolls. The study inspired a long series of comments and rejoinders published in the original journal, and many alternative estimation strategies were discussed. Dr Young used all possible combinations of those suggestions to populate a multiverse analysis comprising 1,152 unique model specifications. The result was telling: the vast majority of plausible models show a zero effect of hurricane ‘gender,’ and less than 5 percent of specifications supported the original claims. Almost any other way of analysing the data failed to support the study’s claims.

Dr Young recommends carrying out model robustness analysis on empirical data to help distinguish between true relationships and suspect findings in order to have more empirically defensible conclusions – findings based on the data rather than the modelling assumptions.

#### THE NEXT FRONTIER: SOME MODELS ARE MORE EQUAL THAN OTHERS

The multiverse methodology considers all possible combinations of plausible models and produces an associated

weighting of possible models involves a trade-off between transparency and model selection. However, it offers a promising path forward for multiverse analysis.

Dr Young’s computational framework makes the multiverse central to how analyses are conducted and reported. It facilitates more rigorous and comprehensive robustness testing and enables model uncertainty to be embraced as an inherent part of social science. It reduces the fundamental

**When evaluating research findings, a multiverse analysis of alternative models is at least as important as statistical significance.**

distribution of estimates. Each model is treated as equally valid when computing multiverse statistics and distributions. Dr Young notes, however, that some of the plausible models are more compelling than others. This raises the possibility of weighting models by appropriate measures of validity. Young’s work on this has explored the use of “model influence” scores to weight possible models, taking into account potential omitted variable bias. He cautions that the

asymmetry of information between analyst and reader, and addresses the overabundance of false positive research findings. Dr Young stresses that “when evaluating research findings, a multiverse analysis of alternative models is at least as important as statistical significance.”

**Note:** Dr Young’s multiverse analysis software is open source and freely available to researchers. Type `ssc describe mrobust` in Stata.



# Behind the Research

## Cristobal Young

E: [Cristobal.young@cornell.edu](mailto:Cristobal.young@cornell.edu) W: <http://cristobalyoung.com/> @cristobalyoung5

### Research Objectives

In addition to statistical methods Cristobal Young studies the social and economic policies that influence opportunity and wellbeing in America.

### Detail

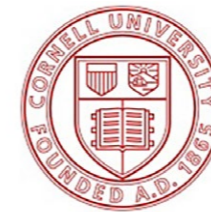
Cristobal Young  
Department of Sociology  
Cornell University  
344 Uris Hall  
Ithaca, NY  
USA

#### Bio

Cristobal Young is an associate professor at Cornell University. He is passionate about evidence-based public policy – and about making the evidence easy to digest. His methodological work focuses on multiverse analysis and model robustness. His applied work is in the areas of economic sociology and inequality, focusing on efforts to increase opportunity and wellbeing in America.

#### Collaborators

- Katherine Holsteen
- John Muñoz



### References

- Muñoz, J. & Young, C. (2018). We Ran Nine Billion Regressions: Eliminating False Positives through Computational Model Robustness. *Sociological Methodology*. 48(1), 1-33. <https://doi.org/10.1177/0081175018777988>
- Muñoz, J. & Young, C. (2018). Rejoinder: Can We Weight Models by Their Probability of Being True? *Sociological Methodology*. 48(1), 43-51.
- Young, C. (2018). Model Uncertainty and the Crisis in Science. *Socius: Sociological Research for a Dynamic World*. 4, 1-7. <https://doi.org/10.1177/2378023117737206>
- Young, C. & Holsteen, K. (2017). Model Uncertainty and Robustness: A Computational Framework for Multi-Model Analysis. *Sociological Methods and Research*. 46(1), 3-40. <https://doi.org/10.1177/0049124115610347>
- Steege, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. (2016). Increasing Transparency through a Multiverse Analysis. *Perspectives on Psychological Science*. 11, 702-712. <https://doi.org/10.1177/1745691616658637>

### Personal Response

**What are your plans for future research involving multiverse analysis?**

“ I am planning to write a book, tentatively titled “Into the Multiverse: Computational Methods for Robust Research.” ”

