Understanding Imperfection: Embracing Uncertainty in Scientific Modeling*

Siqi Fei

March 28, 2024

> [s]ince all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.
>
> George Box (1976)

1 Introduction

Statistics plays a crucial role in our understanding of the world, given that complete and precise data is often out of reach. In practice, we rely on sampled data to make inferences and predictions about larger populations or systems.

Consider the challenge of estimating the total number of ants in a backyard. Directly counting every ant is impractical and impossible. Instead, by counting ants within a smaller, manageable area, one can employ statistical models to estimate the total population in the entire yard.

With different models, we can predict almost all values on Earth. However, as George Box (1976), a renowned statistician said: "All models are wrong." This statement highlights the inherent limitations of models. No model can guarantee that the values it produces are 100% correct. Models can only infinitely approach the true values, but can never reach them. Therefore, when we conduct statistical analysis, we need to accept the imperfect factors that may exist in the models, and instead focus on controlling the factors we can control to minimize the errors generated by model predictions.

The focus of this paper is on the practical wisdom within Box's observation, particularly his caution against "being concerned about mice when there are tigers abroad." We will explore real-world scenarios where minor inaccuracies (mice) distract from significant errors or oversights (tigers), leading to misguided priorities or conclusions. Through examples, we aim to demonstrate the importance of discerning critical flaws in our models and data interpretations,

*Code and qmd file are available at: https://github.com/FXXFERMI/Understanding-Imperfection.git.

thereby enhancing the reliability and utility of statistical analysis in uncovering truths about our complex world.

2 The Misuse of Models in Financial Markets

The 2008 financial crisis stands as a significant event with profound impacts on the global economy. This crisis led to widespread job losses and prolonged unemployment across the globe, devastating many individuals and families. Countless people faced financial ruin, and numerous families lost their homes and futures. A key factor in this crisis was the regulatory reliance on flawed risk models. Avinash Persaud, emeritus professor of Gresham College, points out in his article (CEPR 2023), that financial supervision failed to avert the turmoil because it trusted the very risk models that were sensitive to price fluctuations, contributing to the crisis's outbreak. The economic models prior to the financial crisis focused too intently on immediate, minor variables, such as price fluctuations, neglecting the critical, broader market forces and trends. This narrow perspective, combined with an oversight of the models' limitations and the intricate nature of market dynamics, played a significant role in leading to the financial crisis. This case exemplifies Box's warning against obsessing over the 'mice' of minor data fluctuations while ignoring the 'tigers' of systemic risk factors. It's a stark reminder that in the pursuit of precision, the broader, often more impactful, vulnerabilities can be overlooked, leading to significant consequences.

3 Leveraging Predictive Modeling in Disaster Management

The application of predictive modeling in disaster management exemplifies the effective use of models to focus on significant risks. The prediction model for earthquake can be one of good examples. Based on the report on official website of the United States government (US Geological Survey (USGS) year), the main risk factors aggravate earthquake risks, including but not limited to: Population density, Scales, Poverty, Poorly built and non-engineered buildings. By prioritizing those "tigers" of potential disaster impacts, emergency services deployed resources more strategically, implemented evacuation orders, and communicated risks to the public more effectively. Since we have the predictive model for earthquake, we saved 13,000 lives a year by predicting earthquakes (Marr 2015).This approach underscores the value of using models not just to predict outcomes, but to guide actionable strategies that can save lives and minimize damage, highlighting the critical balance between detail-oriented analysis and big-picture strategic planning.

4 Conclusion

In general, George Box's assertion that "all models are wrong"(Box 1976) serves as a crucial reminder of the inherent limitations within statistical modeling. Through examining the misuse of financial risk models in the 2008 financial crisis and the successful application of predictive modeling in disaster management, we see the importance of focusing on significant risks ('tigers') rather than minor inaccuracies ('mice'). These examples underscore the need for a balanced approach in model use and interpretation, emphasizing the critical role of discernment in enhancing the reliability and utility of models for complex real-world applications.

References

Box, George E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71 (356): 791–99. https://doi.org/10.1080/01621459.1976.10480949.

CEPR, An author from. 2023. "Why Bank Risk Models Failed." *CEPR's Policy Portal*. https://cepr.org/voxeu/columns/why-bank-risk-models-failed.

Marr, Bernard. 2015. "Big Data: Saving 13,000 Lives a Year by Predicting Earthquakes." *Forbes*. https://www.forbes.com/sites/bernardmarr/2015/04/21/big-data-saving-13000-lives-a-year-by-predicting-earthquakes/?sh=57d143775d8d.

US Geological Survey (USGS). year. "Moment Magnitude and the Richter Scale: What Are the Different Magnitude Scales, and Why Are There so Many?" year. https://www.usgs.gov/faqs/moment-magnitude-richter-scale-what-are-different-magnitude-scales-and-why-are-there-so-many.