

**GDN/CERGE-EI RRC GRANT NO RRC12+09**  
**ARUN ADVANI & TYMON SŁOCZYŃSKI**  
**NON-TECHNICAL SUMMARY**

This paper, *Mostly Harmless Simulations? On the Internal Validity of Empirical Monte Carlo Studies*, evaluates the premise from the recent literature on empirical Monte Carlo studies that an empirically motivated simulation exercise is informative about the actual ranking of various estimators when applied to a particular empirical problem (see, e.g., Busso et al. 2013, Huber et al. 2013).

The article's approach to answering this question is based on a unique setting which is provided by the well-known National Supported Work (NSW) data set (see, e.g., LaLonde 1986, Dehejia and Wahba 1999, Smith and Todd 2005). The NSW programme was based on random assignment of its participants, and therefore the "true" experimental estimate of its impact can be easily computed via a comparison of mean outcomes in the treated and control groups. At the same time, there exist so-called NSW-CPS (Current Population Survey) and NSW-PSID (Panel Study of Income Dynamics) data sets which substitute the original control group with nonexperimental control groups from the standard U.S. microeconomic data sets. Therefore, nonexperimental estimators of the average treatment effect on the treated can be judged on the basis of their ability to replicate the experimental benchmark using these nonexperimental data. This approach allows the paper to calculate "true" biases and "true" rankings for the considered estimators. Next, various "empirical Monte Carlo studies" which have been suggested in the recent literature (see, e.g., Abadie and Imbens 2011, Busso et al. 2013, Huber et al. 2013, Lechner and Wunsch 2013) are performed, and they are designed to mimic the NSW-CPS and NSW-PSID data sets. Each experiment allows the authors, of course, to calculate biases of the considered estimators and the corresponding rankings. Finally, the "true" biases and rankings are compared with biases and rankings which are implied by the "empirical Monte Carlo studies".

The authors find that "empirical Monte Carlo studies":

- are based on strong assumptions which are inherently untestable,
- might lead the empirical researcher to choose the "right" or the "wrong" estimators in their empirical context, dependent on this context,
- should not be considered a panacea which always provides information about estimator choice.

The paper suggests empirical researchers should use several different estimators in their analyses, and report these potentially varying estimates as an important robustness check.

*Key words:* empirical Monte Carlo studies, programme evaluation, treatment effects