Research Statement

Nikolay Kudrin

I am an econometrician who is interested in econometric theory, applied econometrics and questions around meta-analysis and research transparency. More specifically, my current research is concerned with developing new statistical tools for detecting and evaluating the extent of p-hacking in research and estimating and correcting for publication bias. The tools that I develop allow researchers and Economics as a profession to draw more accurate conclusions regarding the reliability of empirical findings published in scientific journals. In this way, my research generates significant positive social impact because it allows economic decision makers to choose more optimal policies on the basis of available empirical evidence.

My job market paper, ***"Robust Caliper Tests"*** analyzes the Caliper test, one of the most common statistical tool used by researchers to detect p-hacking and publication bias in economics and other disciplines. The Caliper test is a local test that examines the shape of the distribution of published z-statistics near standard significance cutoffs such as 1.96. In my paper I show that the version of the Caliper test routinely used by practitioners is not valid in a sense that it does not control size over the set of possible distributions of true effects across published studies. I propose an alternative test that I refer to as the Robust Caliper test. The Robust test is constructed on the basis of the worst-case distribution of true effects and hence is uniformly valid. In my paper I show how to derive this worst-case distribution and build a test using it. In addition, I provide the reader with a method of evaluating the extent of p-hacking in the literature that does not require any assumptions regarding the counterfactual distributions of published results absent p-hacking.

In the second paper, ***"Detecting p-hacking"*** (*Econometrica, 2022*; joint with Graham Elliott and Kaspar Wüthrich), we theoretically analyze the problem of testing for p-hacking based on distributions of p-values across multiple studies. We derive general results for when these distributions have testable shape restrictions such as monotonicity (for general tests) and complete mono-

tonicity (for one- and two-sided t-tests), absent p-hacking. We propose new tests that can be used to examine those shape restrictions, which can be substantially more powerful for detecting p-hacking than tests previously suggested in the literature. In the presence of publication bias, our tests can be interpreted as joint tests for p-hacking and publication bias. We demonstrate the usefulness of our tests by applying them to prominent datasets of published results in economics and other sciences.

In the third paper, ***"(When) Can we detect p-hacking"*** (joint with Graham Elliott and Kaspar Wüthrich) we analyze the properties of the tests proposed in the second paper under the alternative (in the presence of p-hacking). We consider likely forms of p-hacking and analyze, both analytically and using Monte Carlo simulations, their implications for the shape of the distribution of reported p-values. We further explore the power properties of our tests for p-hacking under various p-hacking strategies employed by researchers. We also introduce costs of p-hacking in terms of the bias of reported estimates and size distortions of the corresponding tests published in scientific journals. We show that there is a positive relationship between the costs of p-hacking and the power of our tests.

My research in progress is concerned with the problem of non-parametric estimation of publication bias. Existing literature on this topic tends to make strong parametric assumptions on the shape of the publication probability function that can lead to erroneous inference. In my ongoing research, I am trying to develop an estimator of the publication probability function that is agnostic about its parametric form. Finally, one more project that is currently in progress is concerned with the problem of uniform inference in weakly identified discrete choice models.