Short Course on Treatment Effect Heterogeneity and Multiple Testing

R. Vincent Pohl

Fall 2017

Description

This short course has three goals: (1) to think about the importance and different possibilities of allowing for heterogeneous treatment effects in program or policy evaluation, (2) to familiarize students with the basics of correcting inference for multiple testing, and (3) to cover recent papers that use multiple testing methods in the estimation of treatment effects.

In the early days of the program evaluation literature, studies usually estimated a homogeneous treatment effect for the whole sample (e.g., the average treatment effect, ATE). More recently, authors have recognized that individuals may react differently to the same treatment. Treatment heterogeneity may occur by observed characteristics (subgroup-specific treatment effects) or across the outcome distribution (quantile treatment effects).

While it is relatively straightforward to estimate heterogeneous treatment effects, less attention has been paid on how to conduct inference. The main problem lies in the fact that estimating heterogeneous treatment effects implies testing multiple hypotheses. For example, if we estimate a treatment effect for 10 subgroups, we test 10 null hypotheses of no treatment effect. If each of these hypotheses has a false rejection rate of 0.05, the probability of falsely rejecting at least one of the 10 hypotheses equals 0.4. This is problematic because we overestimate the effect of a given policy with a high probability. To control for this issue, several adjustments have been proposed, starting with the Bonferroni correction that suffers from low power to more recent developments such as the step-down method proposed by Romano and Wolf (2005).

In the last part of this course, we apply multiple testing to treatment effect heterogeneity and cover a few recent papers that use multiple testing methods for subgroup-specific treatment effects, multiple outcomes, and quantile treatment effects (Lee and Shaikh, 2014; Lehrer et al., 2016; List et al., 2016).

Reading List

Bitler, M.P., Gelbach, J.B. & Hoynes, H.W., 2006. What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments. American Economic Review, 96(4), pp.988–1012.

Bitler, M.P., Gelbach, J.B. & Hoynes, H.W., 2014. Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment. NBER Working Paper No. 20142

Crump, R.K. et al., 2008. Nonparametric Tests for Treatment Effect Heterogeneity. Review of Economics and Statistics, 90(3), pp.389–405.

Fink, G., McConnell, M. & Vollmer, S., 2014. Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures. Journal of Development Effectiveness 6(1), pp.44–57.

Firpo, S., 2007. Efficient semiparametric estimation of quantile treatment effects. Econometrica, 75(1), pp.259–276.

Lee, S. & Shaikh, A.M., 2014. Multiple Testing and Heterogeneous Treatment Effects: Re-Evaluating the Effect of PROGRESA on School Enrollment. Journal of Applied Econometrics, 29(4), pp.612–626.

Lehrer, S.F., Pohl, R.V. & Song, K., 2015. Targeting Policies: Multiple Testing and Distributional Treatment Effects. NBER Working Paper 22950.

List, J.A., Shaikh, A.M. & Xu, Y., 2016. Multiple Hypothesis Testing in Experimental Economics. NBER Working Paper 21875.

Romano, J.P., Shaikh, A.M. & Wolf, M., 2010. Hypothesis Testing in Econometrics. Annual Review of Economics, 2(1), pp.75–104.

Romano, J.P., & Wolf, M., 2005. Stepwise Multiple Testing as Formalized Data Snooping. Econometrica 73(4), pp.1237–1282.