A Genome Wide Association Study of Educational Attainment

Jonathan Beauchamp, David Cesarini, James H. Fowler, Patrick Groenen, Philipp Koellinger, Matthijs van der Loos, J. Niels Rosenquist, Roy Thurik, Nicholas A. Christakis

Twin and adoption studies have consistently found that genetic variation is an important source of heterogeneity in economic outcomes such as educational attainment and income. The advent of inexpensive, genome-wide scans is now making it increasingly feasible to directly examine specific genetic variants that predict individual differences. In this paper, we conduct a genome wide association study (GWAS) of educational achievement. In the first stage, we used data on over 360,000 genetic markers throughout the genome from the Framingham Heart Study, a family-based sample of nearly 8,500 individuals, and found a number of markers with suggestive associations with educational attainment. The most promising variants were significant at the $5 \cdot 10^{-7}$ level. In the second stage, we attempted to replicate the most significant first-stage associations in the Rotterdam study, an independent sample of over 9,500 individuals. None of the first-stage associations replicated in the Rotterdam study, suggesting that the first-stage results were all false-positives. We conclude by discussing the challenges that arise when doing inference in genoeconomics research, emphasizing the importance of properly correcting for multiple hypothesis testing and of replicating significant results in independent samples. We also discuss issues of power and sample size.