



DIPARTIMENTO DI SCIENZE ECONOMICHE

Prof. Pietro Biroli
Senior Assistant Professor
Department of Economics
University of Bologna
pietro.biroli@unibo.it

Dissertation Feedback

Jaroslav Groero's dissertation is quite impressive and displays great research and analytical skills.

The dissertation presents three empirical chapters on gene-environment interaction models in economics. The candidate uses large-scale data sets from different countries and sophisticated econometric methods to examine how genetic factors interact with educational attainment, macroeconomic shocks, and risk preferences. The dissertation makes several original and significant contributions to the literature and demonstrates the candidate's great analytical skills. The dissertation presents an innovative and promising area of research in economics and makes several original contributions to the literature. It demonstrates the candidate's ability to conduct high-quality research at the intersection of genetics and econometrics.

Let us look at the three chapters independently:

The **first** chapter investigates how education moderates the effect of genetic predispositions on health outcomes. The candidate uses data from the UK Biobank and exploits a natural experiment to identify the causal effect of education on health. The candidate finds that education reduces the impact of genetic risk factors for several chronic diseases and conditions, such as diabetes, hypertension, and obesity. The chapter provides novel evidence on the protective role of education against genetic vulnerabilities and has important implications for public health policies.

Overall, the paper could benefit from being more succinct and having a more focused and cogent statement about the contribution of the paper. This chapter makes two main contributions to the literature, in my opinion. First, it uses a novel method that is free from any specific assumption about the similarity of outcomes in GWAS samples and survey samples used to estimate GxE models. I believe that this is the greatest contribution of the chapter and should be placed front and center from the very beginning. This method has important implications for future research in this area, but the main juice of the paper only comes after the first 10 pages. Second, it examines more medically relevant outcomes and severe medical conditions such as cancer, heart attack, stroke, and diabetes.

I see a couple of potential issues of the new method, that might benefit from a greater discussion in the paper. First, the selection of only K SNPs runs counter to the commonly held belief that most complex traits are polygenic in nature (the so-called fourth law of behavioural genetics, Chabris et al 2015). It would be useful to discuss more this potential limitation. Secondly, the non-linear method might be more computationally intensive to run and, being non-linear, the



DIPARTIMENTO DI SCIENZE ECONOMICHE

minimization algorithm might have convergence problems. Discussing the costs and benefits of the approach might be helpful for the reader.

A more minor comment, is that it could be useful to decompose the bias shown in equation 1.7 into the components that are related to differences in the measures of phenotypes, and maybe connect it to the MTAG method developed by the Turley et al. (2018), and the different penetrance of the same SNP into different cohorts or environments, and maybe connect this to the concept of genetic correlation across cohorts.

Finally, I found the appendix A (extension of section 1.2) quite precise and insightful, albeit a bit verbose.

The **second** chapter examines how macroeconomic conditions affect risk tolerance for individuals with different genetic profiles. The candidate uses data from the German Socio-Economic Panel and exploits exogenous variation in regional unemployment rates to identify the effect of economic shocks on risk preferences. The candidate finds that adverse macroeconomic conditions lower risk tolerance for individuals who are genetically predisposed to be risk tolerant, while they have no significant effect for individuals who are genetically predisposed to be risk averse. The chapter contributes to the literature on the determinants of risk preferences and shows how genetic factors can shape individual responses to economic fluctuations.

I found the approach taken in this paper to be very innovative and ambitious. Jaroslav have clearly put a lot of thought and effort into developing a novel methodology that has the potential to make significant contributions to the field. I commend them for their creativity and hard work in pushing the boundaries of what is possible in this area of research.

One minor suggestion that I have is related to the statistical power of the approach. Given the high level of complexity and non-linearity, it might be helpful to have a monte carlo simulation that estimates the statistical power of this approach and its ability to recover certain levels of gene-environment interactions in a data generating process and a sample size similar to the one available in the HRS.

The **third** chapter discusses some methodological issues in the current gene-environment interaction and proposes an innovative solution. Jaroslav demonstrates that gene-environment interactions can distort the PGS index and proposes a new two-step method to estimate gene-environment models and their non-interacted counterparts. A great innovation of this new method is that it does not rely on genome-wide association study (GWAS) estimates and does not suffer from biases that stem from using GWAS estimates in the PGS index.

This chapter extends and develops further the method implemented in the first chapter, and shows a high level of econometric sophistication and creativity. I believe that this work has the



DIPARTIMENTO DI SCIENZE ECONOMICHE

potential to improve the accuracy of studies that investigate how environments and choices interact with genetic endowments to form important economic, behavioral, or health outcomes.

I have two main suggestions for this last chapter. The first, is that it would be useful to distinguish a bit more the relative contribution of the first chapter vs the last one: the introduction of the problem, the discussion of current GxE research in economics, and the proposed solution seem to have a lot of overlap. For publication purposes, I would suggest Jaroslav to distinguish more these two papers from each other. The second suggestion is to discuss more the literature in biostatistics about Genomewide Interaction Studies (GWIS; sometimes called GxE wide association studies, GxEWAS) especially the more recent work by Miao et al (2022)¹ who propose an alternative solution to a very similar problem.

From an academic perspective, the dissertation exhibits exceptional quality in writing and demonstrates a profound understanding and mastery of the relevant literature, particularly in the field of economics as well as other related disciplines. Jaroslav's adeptness in utilizing technical and statistical tools to address the existing limitations in the literature is commendable. The innovative and creative solutions presented in the dissertation significantly advance the research frontier.

¹ "Reimagining Gene-Environment Interaction Analysis for Human Complex Traits", Jiacheng Miao, Gefei Song, Yixuan Wu, Jiabin Hu, Yuchang Wu, Shubhashrita Basu, James S. Andrews, Katherine Schaumberg, Jason M. Fletcher, Lauren L. Schmitz, Qiongshi Lu bioRxiv 2022.12.11.519973; doi: <https://doi.org/10.1101/2022.12.11.519973>