# External Review for CERGE-EI PhD Dissertation by Jaroslav Groero

Dr Selma Walther

Senior Lecturer

Department of Economics

University of Sussex

## Summary and Comments

Thank you for inviting me to read this fascinating dissertation written for the degree of PhD at CERGE-EI.

The dissertation is broadly in the theme of geno-economics: an emerging field in empirical microeconomics that uses insights from biological genetics to explore the interaction between baseline genetic drivers of economic outcomes, and circumstances external to the individual. The field has gained significant traction in recent years and is an exciting area of research to be working in.

The first chapter explores how education can moderate the negative impacts of genetic predispositions for health outcomes such as heart attack. This is an interesting question as it suggests that education can have benefits beyond earnings and the labour market. It is plausible that educated individuals have healthier behaviours, and this chapter suggests that these behaviours can offset some of the genetic predispositions for negative health outcomes.

I do not find any substantive changes that need to be made in order for this chapter to be presented for defence. However, for future work on this chapter, I have a few suggestions:

> *It would be helpful in the discussion of Tables 1.1-1.3, to explain how large the effects are relative the baseline probabilities of these health outcomes. Also, do these correlations vary for men and women, and by age?*
>
> *Is there existing literature on the impact of the school leaving age policy? If yes, this could be discussed.*
>
> *Some mention of epigenetics may be relevant here. Can education affect genes, and hence the analysis collected among adults in post-education age?*
>
> *A better justification of using the 3$^{rd}$ quartile (isn't this effectively the median?) as a cut-off, e.g. in Table 1.5, is needed.*

The second chapter addresses the following question: do genetics and the environment interact to shape the formation of risk preferences? Specifically, a literature has shown that individuals' risk preferences are sensitive to past experiences; the second chapter extends this literature by allowing an interaction between past macroeconomic unemployment risk, and genetic markers for risk, in the empirical specification for risk attitudes. The results are striking: only individuals who have below median genetic markers for risk are sensitive to past experiences.

I do not find any substantive changes that need to be made in order for this chapter to be presented for defence. My main concern with this chapter is that risk attitudes can affect survival and hence could cause the HRS to be a select sample (in terms of risk); however, this concern is addressed adequately in the robustness section. For future work on this chapter, I have a few suggestions:

> *While the Malmendier and Nagel approach takes care of some of the issues of using past unemployment rates, I wonder whether an even more robust alternative could be to use unemployment rates during a fixed portion of life, e.g. between ages 20-40 (the prime career-building years). This would also remove endogeneity associated with age – and hence number of years of unemployment data used, which will vary with age of the individual.*
>
> *Further, while possibly increasing endogeneity, a more precise measure could be to use unemployment rates by industry – linked to the industry of the individual. This could more precisely capture the risk faced by the individual.*
>
> *Another alternative could be to use the unemployment experiences of individuals' parents during their childhood. These would not have the direct negative effect on own labour market outcomes, but would affect attitudes to risk by learning about risky outcomes through the experiences of parents.*

Chapter 3 takes a step back from these two chapters by revisiting the method used to calculate the polygenic score. Specifically, it is shown that existence of gene x environment interactions can skew the estimation of the polygenic score. One of the main reasons this can occur is because the estimation sample for polygenic scores tends to differ from the estimation sample for gene x environment interactions. A novel two-step procedure is proposed that circumvents some of these issues.

I do not find any substantive changes that need to be made in order for this chapter to be presented for defence. However, for future work on this chapter, I have a few suggestions:

> *The motivating example using mothers' education is helpful, but I would suggest formalising this example and carrying it through the chapter. (That is, instead of only describing it in words, rather defining formally two levels of education, L and H, and showing in simple regression notation the extent of bias in this example.)*
>
> *The two empirical applications (mothers' education and BMI-ROSLA) could be better explained and justified. Why are they a good test of the new model?*
>
> *It would be interesting to discuss why, in some cases a positive bias might occur, and in others a negative bias (as in the two empirical applications).*
>
> *Should we expect smaller bias in the second application, given that the samples for polygenic score and GxE interaction estimation are more similar?*

<u>Recommendation</u>

Overall, I find this dissertation covers interesting questions. The dissertation is convincingly and maturely written. I commend the student on his knowledge of the relevant literature. I recommend that this dissertation is presented for defence, and I have no substantive comments that need to be incorporated ahead of the defence.