Rome 23.12.17

Report on “Essays on Conflicts and Human Capital Accumulation”, a thesis submitted by Dejan Kovac for the award of the title of PhD in Economics at CERGE-EI, Prague.

The thesis is composed of three chapters, two of which co-authored with a senior colleague. The three chapters use in different combinations the following data:

- data on the universe of veterans in the Serbo-Croatian conflict of the early 1990s
- schooling data on children of veterans. These data are longitudinal and refer to a period of over 10 years after the end of conflict. There is also cross sectional data on these children’s hospitalization
- birth records for a very long time period containing names of new-born children as well as of their fathers
- auxiliary data on the intensity of conflict, and other local variables.
- Data on veterans’ suicide.

Chapter 1 examines the effect of father’s death during conflict on children’s’ (long run) schooling and health outcomes. The paper argues that there are pronounced negative effects and that these are largely ascribable to exposure while in utero, which the author argues in turn pointing to an effect via mother’s psychological stress.

Chapter 2 uses data on first names of draftees and volunteers to argue that those carrying “nationalistic” names were more likely to volunteer and to die in conflict. The authors argue that this is evidence of intergenerational transmission of nationalistic values.

Chapter 3 uses data similar to those in chapter 1 to examine the determinants of veterans’ suicide and on children’s outcomes. The authors argue that greater exposure to intense conflict causes increased risk of suicide, and that father’s suicide has a causal negative effect on children’s school performance, although the result is much more nuanced when children of suicides are compared with children of those who died in conflict, perhaps a better control group.

Overall appraisal.

This is extremely good work, well above the average of theses I routinely examine, including from LSE, UCL and Oxford. The thesis is very carefully written, the candidate shows a very high degree of technical competence, high of level of scientific maturity, excellent knowledge of - and an ability to locate his contribution in - the literature, while the narrative is articulated and deep. Equally, the candidate shows a very high level of creativity and resourcefulness in finding novel data to study new questions, or questions studied before but from a new perspective.

Chapter1 has probably good chances of publications in a top field journal, and chapter 2 (which I noticed is R&R at the EJ) has good chances of publication in a general interest journal below the top 5. Chapter 3 is the weakest in my mind and I wonder if there would be no benefit of merging chapters 1 and 3.

Below, I provide some observations/comments and where possible suggestions with a view to publication. Given the status of chapter 2, perhaps my comments might be superfluous. My comments will be necessarily sound a little critical – as they refer to places where the candidate I believe can do better but this does not detract from my positive appraisal of the work.

General comments

One first general comment on the chapters. This applies to all three.

All chapters are extremely rich in terms of analyses. However, It is somewhat hard for a reader to keep track of the material. This is even truer for chapter 2 – which I found the most fascinating. The narrative in this chapter is in my view too loose. Somewhat there are a number of results but not all them are instrumental to the overall story and some detract from the tight narrative that I think a good paper should have. Throughout the thesis, it feels at times that the author has gradually incorporated a number of comments/suggestions from others but often these additions do not show up coherently with the rest of the work and they slow down a reader. All chapters would benefit from more focus, fewer repetitions, a clearer and tighter narrative, where all the material that is presented is instrumental to the one story the authors want to tell.

By converse, while the analysis is of very high standards, the data description including sources, limitations, etc. is not. Often I do not know here variables come from, what the rational for the collecting some administrative data (e.g. on veterans' children) is, at what time some variables are measured (e.g. income in chapter 1 – chapter 1 is the most confusing as data go), how different data sets are linked, what problems there are in terms of selection etc. I looked somewhere in the thesis for a measure of the fraction of men who served, on the criteria for draft eligibility etc. These seem first order statistics/details and these are very important for many of the chapters (primarily chapter 2). It is possible that I missed them but if they are there they are a little hidden (I found something in a footnote at one point). In chapter 1 the model section precedes the data description but that's confusing as the models only make sense in the light of the available data. Footnotes to tables and pictures are somewhat not detailed enough. For example I do not know what "controls" are in the regression tables of chapter 1. To a reader like me all this somewhat detracts from transparency and it does not shed a very good light on the entire empirical exercise, which I think it is ultimately counterproductive for the author.

Below are my main specific comments on the three chapters.

Chapter 1

Clearly, the limitation of this chapter is that death in service is potentially not random. The obvious concern is that children of those who died would have done worse irrespective of their father's fate. The candidate is obviously cognizant about this.

A related concern is that children in utero during the war were somewhat different from other children: ultimately some of them must have been conceived while the father was enlisted - and for a variety of reasons this might refer to a specific set of fathers/households or even children.

I do not think that unit fixed effects actually do the trick, and I think it would be hard to convince a referee for a top field journal of the opposite. Clearly, one way to go is to saturate the model with many more controls, one of which appears to be zip code, which I do not think is there.

Certainly a much more convincing strategy would rely on the timing of incorporation into units coupled with intensity of conflict – and hence risk of death. For this I think the paper – and actually the entire thesis – would greatly befit from an understating and a better description of who was drafted, and conditional on this, on the time of drafting. If one had a good instrument for timing (e.g. letter of surname, date of birth..) that would actually probably do the trick. Clearly some individuals volunteer, but as long as this instrument exists and it generates sufficient variation in the overall probability of enlistment, this should be ok. Clearly assignment to units is not random but I think that at one point the author argues that most people serve (or fight?) locally. So it is effectively local intensity of conflict X variation in the timing/probability of enlistment that is needed.

One concern I have with chapter 1 is that the author argues forcefully that households of deceased veterans were compensated through generous pensions and so income cannot be the channel. I am sceptical about this. The key point is that these pensions must have come well after the war. This is never discussed. Negative income shocks at critical stages of life might have long-term repercussions if households cannot buffer these shocks, even if these shocks are later remedied. Possibly this can also happen in while in utero, as nutritional deprivation in the third trimester is known to affect birth weight and long term outcomes.

This brings me to the issue of timing of effects. Frankly I would not attach much weight to the regressions showing effects by age at time of exposure. For one, these do not see to vary monotonically with time of exposure. If the author looks carefully as the outcomes by ages he will see a clear u-shaped or inverse u-shaped relationship – depending on the outcome. Effects of exposure at later ages are often very similar in magnitude to exposure in utero. The issue there is simply that the estimates of exposure at later ages are very imprecise, and this is due to the small number of observations, something that I think - but I am not 100% sure - is due to the structure of the school data (can this be clarified?).

The author forcefully argues that exposure while in utero has a negative effect due to mother's psychological stress. I would be dubious about this and for me to be convinced I would have expected these effects to show up when exposed in the first trimester of gestation. If instead effects manifest when exposed in the third trimester I would have thought that this is due to nutrition. Perhaps there is no sufficient variation in the data to look at effects by trimester, but in all cases the author should be a little bit careful before jumping to conclusions.

Why the author never looks at drop out in these data?

Effects seem larger for boys. I assume that child labour is really not an issue, but often it is boys who are sent to work when income is low. Even if households received remedial pensions at one point, drop out is often irreversible.

The schooling data are longitudinal. This means that one should be able to tell apart the effect of age of exposure from the effect of time of since exposure (although with time effects the three effects can only be identified with at least one restriction). The strategy can be further refined by examining siblings. Can we see the same specification with and without differentiating across siblings? In this respect I am not completely sure of why siblings outcomes are differentiated rather than simply taking levels and including for household fixed effects. I assume the dyadic model has the advantage of not constraining the household effects to be the same across all siblings, but it allows for pairs differences. It might be worth clarifying this.

Chapter 2

As said, I liked chapter 2. I think perhaps the biggest problem with this is that nationalistic names can be a proxy for low SES. In this case, the greater propensity to volunteer among those with nationalist names has nothing to do with nationalism. Can this be checked, investigated?

When I started reading the paper, I thought the authors were going to match birth records to veteran data. This has the clear advantage of providing baseline characteristics as well as to investigate patterns of volunteering and drafting as a function of characteristics, including first name. I though that the authors would look at the probability of serving as a volunteer or as a draftee among all those at risk, i.e. the adult male population. Instead the authors restrict the analysis to veterans, and then they investigate the relative probability of volunteering vis a vis of being drafted conditional on such (admittedly few, perhaps too few given the available data) characteristics, including name.

Why is the exercise I have in mind not done? Is it because of data limitations? Here is my concern. My bet is that if one did the exercise I have in mind, one would find that also draftees are more likely than an individual at random in the population to carry a nationalistic name, although less so than volunteers. As I suspect that draftees have a lower than average SES, this would confirm my concern that nationalistic names are possibly a proxy for low SES.

The discussion on pages 11 and 12 of the chapter about weighting totally eludes me. I think this would me unintelligible to many readers as the authors make reference to a model they have not specified yet (writing an explicit model would help here).

The authors note a resurgence of nationalistic names. They also note in a separate part of the paper that names are inherited. Can the resurgence be explained by inheritance of a name from the father or a grandfather (the two peaks are around 45 years apart which makes it unlikely)? That would not necessarily change the story (as the recurrence of nationalistic names is not key for the authors' story - and if anything it actually detracts from the narrative).

I did not find the test of Bisin and Verdier's model very important here and actually messing the narrative a bit. Those regressions might simply be interpreted to mean that if there are a lot of Ante's in my town I'd rather not name my son Ante. That can be for a number of reasons (common names denote low SES, people think that everything else equal rare names are more desirable as they uniquely identify individuals etc.).

Chapter 3

I have no major remarks about chapter 3, rather than perhaps saying that is to me the least interesting. It reads a little a bit like a filler chapter, which is perfectly acceptable. Clearly the authors face very similar challenges to those in chapter 1. As said, ultimately I do not see much of a difference between the two chapters. Exposure to intense conflict can lead to immediate death or perhaps long term psychological effects both of which can transmit across generations and one will need to convince a reader that these effects are truly conflict induced.

Once more if the author could come up with an instrument for incorporation/timing of incorporation that would really make it a much more compelling paper.

I confirm that the thesis satisfies formal and content requirements for a PhD thesis in economics, and I recommend the dissertation for a defence.

Marco Manacorda