Posthumus Feedback on 'The Difference A House Makes'

Bas Machielsen

2025-03-19

Introduction

This document details some of the feedback I have on the PhD Project "The difference a house makes: real estate and intergenerational wealth transmission in Belgium (1810-1955)" by Daan Van den bussche. I will focus on two particular aspects: first, on *evaluation*, meaning, whether the candidate has incorporated the feedback from preceding Posthumus sessions, and to which extent. Secondly, on *looking ahead*, meaning, a reflection about whether it is realistic that the project will be finished within due time. In passing, I will also reflect on the original material and the feedback, in the hopes that that will be useful.

Evaluation

My evaluation is generally positive. In general, I think your project has enough structure to continue the way you and your supervisors intend to. But I also want to offer some criticism, which will follow now. I will start this evaluation by providing a short summary of the most important feedback offered to you and consider the major paper as well as the text in sections B, C, and D, and ask whether the feedback has been incorporated. In addition, I'll try to offer my own perspective on both the text and the feedback.

Most important criticism:

- (Richard) "If I understand it well you want to give an overview of the share of the real estate in Belgium and investigate as well how it is distributed. I searched for a main question, though I could only find a goal in the text: focus on 'inequalities in real estate and the changing role of real estate in wealth in the 19th and early 20th century Belgian society'."

Your overarching research question is now "What was the impact of evolutions in real estate property on changing wealth inequality in nineteenth- and early twentieth-century Belgian society?" (p.2) or " "How did the impact of real estate ownership on wealth inequality develop in a society undergoing rapid industrialisation and urbanisation?" (p. 4) **I think that this question is still not precise enough for it to be an overarching research question.** I really like the aspects of (the effects of) redistribution and what difference a house makes. I think those are very precise ideas which your data also allows a convincing focus on. But I also want to stress that I think that, at this stage, not having an overarching question is normal. An essential part of a PhD project is just about looking for the right questions to answer using the data that you have or can obtain.

- "Richard Paping rightly observed that my research questions were not yet sufficiently developed and that I had failed to outline the methodology I intended to apply. He also noted the absence of an overarching main research question in my proposal.

I think that your major paper still doesn't contain a methodology section, even though it is probably needed to bridge the gap between qualitative descriptions of the data and offering quantitative evidence. Even if the point is to illustrate quantitatively a deeper, general historical treatise, I still think that more information and thinking about the method is required. For example, for the research questions pertaining to inequality, I have the impression that you are going to estimate metrics like the Gini coefficient. But it is unclear at what level

or at what frequency, and how you go from the data to constructing a particular metric. And discussions that link the "oversampling of the elite" strategy to the Gini are still missing. And given that something like the Gini coefficient will be a dependent variable, you also need a discussion of what will be the independent variable and control variables, and what model you want to estimate. I have some advice and considerations on that below.

- (Christine) "Coming from this strand of research tradition, I wonder how the design of the PhD project can contribute to this discussion. My impression is that the sources are not very rich with regards to understanding the motives of wealth strategies. This could become a problem, because there is a certain danger that this study will get stuck on a rather descriptive level, instead of answering research questions."

I think that you also have to work on addressing this comment in total. But I don't necessarily think that descriptive analyses and answering research questions are contradictory endeavors: descriptive, historical, and more precise evidence can complement each other. But I think that the research questions have to be more specific and clearer in order to fully use that complementarity.

Data & Methodology: I think in general, you have obviously thought well about the data. But you also mention there is some criticism regarding the feasibility of the data collection process.

In my opinion, *you do need a large sample size*. In my own studies (WPs here and here), I also analyze wealth at the individual level as a dependent variable in a quantitative way. My experience here is that is a *noisy, volatile, high variance type of variable!* And especially when you're doing inequality, you might want to compute estimates of inequality at (e.g.) the municipality-time level, or at the age of death-gender level, for which you need enough observations within a certain cohort to get an estimate with a small enough variance.

As I mentioned, I think a clear overview of methods used is still lacking in your document. I think you can probably get away with something general. When analyzing the data quantitatively, you can probably do something like OLS, but in some cases you do have to take into account the most prominent econometric issues such as **censoring** of your data.[1] If you're interested in the Real Estate Share of Wealth, the minimum observable value is 0 and the maximum value is 1. But some individuals, if they were unconstrained, would probably pick something larger/smaller. Econometricians solve this by using the Tobit model, which essentially poses some latent outcome (y^*), where if this value is large enough but not too large, we observe $y \in [0, 1]$.[2]

Theoretical framework: In my opinion, there are too many theories about inequality to pin down your work as either a contribution to it, or even as a test of one or another theory. I kind of like the juxtaposition of the Kuznets curve against Piketty's $r > g$, but I think it might be too simplistic and ignores a lot of more recent economic literature on the topic (for example Nekoei and Seim, 2023). And since your data is microdata, it might just be best to focus on really specific questions which have intuitive theoretical justifications (I offer some suggestions in what follows), rather than tests of a specific theory.

Looking Ahead

Before reflecting on the planning, I have a couple of suggestions/statements I would like to make because they might be relevant in the future planning.

- **I think to live up to the criticism offered by Christine Fertig and Richard Paping, you would need to think about linking your data to other sources**. I would say that the best investment of your thinking would go out to figuring out how to *link* your data to other kinds of data. I have mentioned the potential link with genealogical data (OpenArch has a lot of Belgian data). This possibly allows you to link parents and children together, allowing you to study the dynamics of inequality much more precisely.

[1] This (Ch. Estimating Selection Models) is a good overview of these issues.

[2] If you want to do this in R/Python, this looks as follows: tobit(y ~ x, left=0, right=1, data=data).

- Concretely, something I am thinking about here is to take parents-children pairs who inherit homes vs. parent-children pairs who inherit a similar sum of money, but no home. That could provide you with a quasi-experiment for the children, i.e. what does a house do with the outcomes of children what "mere money" doesn't do?
- Thinking about experiments more broadly, I think "death shocks" (like the World Wars or the Influenza pandemic) which could have been very different from municipality to municipality, and about which it might be possible to find data, can serve as a nice source of variation that allows you to see what happens to prices of (certain classes of) real estate after the demand "exogenously" declines because of an unexpected spike in mortality that differs from place to place (see e.g. Francke & Korevaar, 2021, Korevaar, 2023). And in addition, there are papers that look at inheritance practices as a determinant for present-day inequality (Bartels et al., 2024). If you're able to track parent-children, a simple but really interesting question could be: who of the children gets the house (see e.g. Bernheim et al., 1985)? And try to predict that on the basis of e.g. gender, marital status, etc. Additionally, if a house is divided between several children, it might have to be sold, which causes redistribution. Then, some places with many of those cases occurring can be compared to place with few of those cases occurring.
- **The "cohort" sampling strategy, I think, is incompatible with intergenerational research**. You acknowledge this on p. 36. So we seem to agree that if you want to look at the intergenerational dynamics, you *need* a dataset with parent-child pairs. In my opinion, the strategy should then proceed to identify parent-child years for whom the parents die in the same year and the children too, and specifically look for these probate inventories. So I'm not really sure you need the *Kadaster* at all.
- **Estimating inequality using (sampled) microdata is really difficult**. You mention that "(...) as for other countries, declarations for Belgium are not representative of the general demographic structure. This is because inheritance declarations were mainly left for the higher socioeconomic strata, resulting in a downward bias in measuring inequality." *But you only need this if you're estimating Gini coefficients*. And I'm not sure whether that is the best use of your time or even the most relevant research purpose given that the main advantage of your data is that it's microdata! And even if you use strategies like Piketty and Rosenthal, I think that due to limited data in the extreme tails and sensitivity to outliers, it can be super difficult to credibly estimate things like Gini coefficients.
- **Geographic variation**: I think a better approach is to just confine yourself to simple random sampling, and focus on (real estate-based) inequality at the geographic (e.g. the municipality-level). For example, you could use the probate inventories to come up with (i) an average real estate share of wealth per municipality-time point, or (ii) the proportion of individuals in your sample with no real estate. And then you could set out to explain that by demographic/socio-economic characteristics.[3]
- Linking your data geographically could also allow you to test some of the things you're mentioning on p. 14 (e.g. "with the advent of industrialisation in the nineteenth century, the proportion of agricultural land as the main component of wealth steadily decreased.") by comparing municipalities that are more dependent on agriculture vs. those that are less dependent on agriculture.
- **Thinking about the granularity of the data sources**. In line with what has been mentioned before about the time it'll take to complete data collection, I think I'd be really pragmatic about *what to extract* from the probate inventories and what to leave out. In principle, you can extract things like Real Estate Share (or Real Estate Value) quite quickly, in addition to metadata of the deceased. Maybe it is also possible to get the addresses and geocode it. Leaving it at that is likely enough for most of the suggestions I've offered here. Potentially, given that you have already collected very granular data for a certain sample, you can use that to conduct robustness or sensitivity tests in the future.

Hence, in line with what I've suggested in the Evaluation part, I'd focus on fewer, simpler, but more analytical

[3] If you succeed in linking your data with genealogical data like death records, you can also calculate the probability that a given deceased person figures in your inheritance database, estimating the coverage in that way.

research questions. I can totally see you want an introduction and also a deeper description of the source (Chs. 1 and 2). And Ch. 3 is probably one of the most relevant and feasible ideas to execute. But for the remaining chapters, I would focus on research questions that are simple, answerable using your data sources, and have a clear intuition as to why your explanation might be true.

- But especially for Ch. 4 (Capital flows between city and countryside), you need extremely granular data for which the opportunity costs of collection might be too large vis-a-vis other things that you could do in the same time (as per the suggestions above).
- In Ch. 5, you "aim to explore how the inheritance of real estate influenced inequality in Belgium during the nineteenth and twentieth centuries." IMO this would really benefit from thinking about counterfactuals (as in Imbens, 2022): i.e., what is the alternative state of the world we're thinking about? What is it that you would ideally want to conduct an experiment on?

Conclusion

All in all, I think it's a great project with a lot of potential. Many things are thought out well. But I would focus on the most feasible and relevant parts, and probably focus on two/three chapters/papers with simple and clear questions you can really write convincingly with this data source as the basis, rather than on many more chapters and sacrifice the focus/simplicity/clarity. And fortunately, I think there is plenty of time and opportunity for that.