Preliminary assessment on the doctoral dissertation of Kristína Hrehová (CERGE, Charles University Prague): "Essays in Empirical Economics of the Family"

The doctoral dissertation of Kristína Hrehová, "*Essays in Empirical Economics of the Family*", consists of an abstract and three research papers in economics. Thus, the dissertation does not have an introductory chapter, but an abstract clearly describes the content and main results of the dissertation. The dissertation aims to study family decisions when they are faced with changes in the work environment as relocations of their employer, access to legal services, and information sets. The empirical papers aim to identify causal effects on family stability, personal bankruptcy rates, and reproductive rates. I will next report specific comments on the individual research papers, followed by a concluding statement.

The first paper, "*Firm Relocations, Commuting and Relationship Stability*", is coauthored with Erika Sandow and Urban Lindgren. It uses Swedish micro-level register data to study the impact of firm relocations on commuting distance and the probability of married and cohabiting couples with children separating. The paper finds that an increase in commuting distance leads to a small but statistically significant negative effect on family stability. This paper is already accepted for publication (https://doi.org/10.1080/21681376.2023.2174042) in the *Regional Studies, Regional Science* -journal, which is a good indicator of the quality of the paper. Here are, nevertheless, comments on it.

The paper contains an extensive discussion of the theoretical framework for the analysis, including the social exchange theory and microeconomic theory of divorce. It also reviews the prior literature on the association of long-distance commuting with family relationships. This section provides useful background information for the interpretation of the empirical results presented in this paper. If any literature is lacking, I would say it would be a brief discussion of the health effects related to long-distance commuting.

The paper utilizes relocation decisions of firms to generate plausibly exogenous variation in commuting distances. The variation created by the relocation of the (larger) firm is useful because it is plausibly exogenous to the employee decisions, unlike the residential relocation decision of the employees. Following a good paper by Mulalic et al. (2014), the study sample only includes those who remain with the firm and do not move their residence in the 3- or 5-year follow-up periods. This choice is understandable, but it also means that the analytical sample only includes individuals who have chosen not to relocate or exit the firm *after* it relocated. Since this sample selection is unlikely to be random, the results cannot be generalized to a larger population and the potentially non-random selection into the sample may influence the results. Having said that, the authors openly document the identifying assumptions and conduct tests of their validity, which I acknowledge.

The study uses both calculated commuting times and distances, and it also discusses the limitations in their measurement. This is all useful. My a priori assumption was that the empirical results with commuting times and commuting distances would be similar because the correlation between the two variables is high (0.96). Thus, it was a bit surprising that the results did not quite turn out to be the same.

Because the study only uses relocations from 2011 and 2012, the sample sizes are not very large. The precision of the estimates would have been better if more relocation years (cohorts) could have been used. It would have been also interesting to observe the results for a longer observation period (than 5 years). Having said that, the modelling framework used in this study is best suited for the estimation of short- and medium-term effects since it restricts the sample to individuals who do not move their location of residence during the study period. I would also note that, in my opinion, both 3 and 5 years after relocation refer to (short or) medium term (rather than medium and long term).

Although the text is easy to read, I would have preferred to see more extensive use of notes at the bottom of tables and figures, as they make them more self-contained and easier to follow. In Table 1.3, it is hard to understand where the number of observations comes from. Just looking at this table, I would have expected that the results in column (3) are based on mean values in columns (1) and (2), but they are not. They seem to refer to figures from "Relocated 5-200 km" and "not relocated", but the means (or medians) are not given for the latter group of observations. The same comment applies to Tables 1.6 and 1.7. I also note that since the differences are generally small between the two groups of relocation distances (5-200 vs. 0-5), they also make a useful comparison.

I do like the empirical model, which is based on Mulalic et al. (2014) and presented in section 1.5.1. The authors also continue the recent trend in the applied econometric literature, which models binary outcomes with linear probability models (LPMs). I also find LPMs appealing, particularly when using large data, because they are flexible and tend to be less sensitive to distributional assumptions. However, I would have appreciated the further discussion of the individual fixed effects part of the model, as they are only implicitly embedded in the time differencing. In the text, a greater emphasis is placed on firm-fixed effects. Also note that, on page 16, the paper seems to accidentally refer to the multinomial logit model, when the appropriate reference would have been to the (binary) logit model. Multinomial logit refers to models that are used when the outcome variable has more than two categories as opposed to only two (0, 1).

In Section 1.5.2, the paper investigates the assumptions of the models by regressing the likelihood of remaining with the firm 3 and 5 years after relocation on the (controlled) family stability rankings. I appreciate the efforts to investigate the assumptions of the model. However, I wonder if there would have been a simpler way to study this, maybe, by modelling the likelihood of including it in the sample. I also would have liked to see a discussion of the size of the estimates in Tables 1.8-1.9, because precise zero estimates are more informative than large estimates without precision.

I acknowledge the candidate's laborious efforts to utilize GIS modelling to estimate the commuting times (in addition to using the distances). As regards the parametrization of the changes in commuting distance, I think it would have been easier to read the results if this variable would have been in logarithmic form. Then the results would have shown how the outcome changes in percentage points if distance changes by one percent. Such a logarithmic specification would have been particularly useful since the paper aims to compare results for distance and time (and found unexpected differences in the results).

I appreciate that the paper studies the heterogeneity of the results across specifications, but the presentation of the results in Table 1.3 could have been clearer. Looking at this table, it is a bit hard to know whether each cell represents a new model and what samples and controls are used in them. Note a similar problem with the extensive set of robustness checks. I appreciate them, as they improve the quality of the paper, but they could have been documented in more detail. Furthermore, appendices could have been augmented with additional figures and tables (e.g., in Appendix B).

Despite some critical comments presented above, I think that the paper provides a valuable contribution to the limited literature on the effects of commuting on family stability. The basic modelling framework is well-suited for the analysis, and the results are interesting, and they provide new insights for future analyses.

The second paper, "*Help Matters: The Effect of Access to Centers for Legal Aid on Bankruptcy Rates*", is coauthored with Štefan Domonkos. It uses a large administrative dataset of personal bankruptcies to study the impact of spatial distance from public Centers for Legal Aid (CLAs) on the regional incidence of personal bankruptcy in Slovakia. The study finds that improved access to free legal aid has a significant impact on the use of personal bankruptcy by the public. Like the first paper, also here the emphasis is on spatial distances but with a direct link to economic policy. Thus, the analyses in the paper are highly policy-relevant and support policymaking more directly.

The early part of the paper positions it well within the prior literature. It was also easy to follow the discussion of the Slovak institutional setting. This discussion is quite long, but it is written well. The writing style is a bit different in this paper compared to the first one, maybe due to different coauthors or target outlets. The graphical map illustrations support the discussion, again illustrating the candidate's skills in GIS modelling.

Related to Figure 2.1, an alternative (but probably unwise) modelling choice would have been simply to compare the outcomes before the reform (2016) and after the reform (2019). However, it is unlikely that all other factors could have been controlled for in such a simple before-after analysis. Thus, I think that the authors have made a good choice in developing a model that is based on the changes in the distance to the nearest Center for Legal Aid (CLA). The efforts put into illustrating this key variable in the analysis are appreciated by the reader.

In this paper, the authors predict the expected distance to the Center for Legal Aid (CLA) using many municipal population characteristics and then look at differences in the bankruptcy rates in municipalities that are near to and far from the CLA. Thus, the analysis assumes that once the predicted distance is controlled for, the remaining variation in the access to the CLA is as good as random. The choice of municipal-level socio-economic and geographic control variables are motivated as variables that are likely to influence the decision about the location of CLAs. Note, however, that the controls should also influence the bankruptcy rates in the municipality because otherwise they would be instruments rather than controls.

The legal framework was reformed in March 2017 and the information on bankruptcies is available from March 2017 to March 2020. Since this information is used to derive the dependent variable of the paper, further details on its quality and representativeness would be helpful. It would be important to know whether the quality of the bankruptcy data varies over time and region, so that it would reflect the true bankruptcy rates rather than just their gradually improved data collection over time. The fact, that data on bankruptcies are not available before the reform for research, is a limitation.

A challenge for the illustration of the credibility of the paper is that the number of personal bankruptcies in the data was close to zero in the early part of the study period, and they started to increase rapidly throughout the reform period (2017-2019). Because there were no data on recorded bankruptcies before the reform, it seems hard to test the validity of the model. There are, however, a couple of additional robustness checks that future work could consider. First, one could consider whether there is some placebo outcome that should not have been affected by the reform in the pre-reform and post-reform periods that could have been used to test the validity of the model assumption. Second, is there any additional outcome that should have been affected by the reform (in the post-reform period, but not in the pre-reform period)? Third, once the expansion of the centers is over, one would expect the estimated increase in the bankruptcy rate to stop, ceteris paribus. Thus, in the future, it would be also interesting to see whether it will be the case. Having said that, the paper includes already many robustness checks on the results, which are useful.

As with the first paper, some figures and tables could have included more extensive notes (see e.g. Figure 1). Occasionally the authors also refer to results from alternative modelling approaches that are not reported (e.g., footnote 10). Appendix 2.B could have included a map illustrating the location of the new CLAs. Finally, note that although Table 2.1 notes contain useful information, the table itself is rather packed with information. Some of the variables used are a bit unusual and the table would have been easier to follow if the variables would have been grouped by similarity along with some subgroup headings.

Despite the comments above, these comments are only suggestive, and I think that the paper provides a valuable contribution to the literature and policy debate on bankruptcies. Future work could consider the effects of CLA on wellbeing because it is not entirely clear whether bankruptcies will improve the wellbeing of these individuals.

The third and last paper, "*Persuasive Campaigns, Abortions and Fertility*", is single-authored. It provides an empirical assessment of the conception and abortion-rate effects of a pro-life leaflet mailing campaign that operated in Slovakia during 2016-2017. The study finds that the campaign did not have a statistically significant effect on conception rates or abortion rates.

In the beginning of the paper, the candidate provides a brief description of an ideal research design that could be used, in theory, to study the effects of a pro-life campaign (a randomly assigned intervention). It is evident that implementation of a randomized field experiment is not feasible here, and another approach is, therefore, taken here. Importantly for the study, the non-governmental organizations (NGOs) running the pro-life campaign picked treated and non-treated districts based on an abortion rate cutoff, which created quasi-random variation in the leaflet distribution across observations with similar abortion rates. Overall, I appreciate the description and motivation of the research design in the introduction.

Prior evidence on this topic is limited to the effects of (one-shot) persuasion campaigns aiming at changing reproductive behavior. The literature review is well-focused, and the key references are published in top journals. The campaign is also clearly described. One important observation from this discussion is that the abortion rates are low in Slovakia compared to most European countries. Therefore, I would expect that the attitudes towards abortions are generally quite negative in the population, and therefore, the expected effect sizes to be low. Fortunately, the campaign was targeted at areas with the highest abortion rates, which is likely to improve the chances of obtaining significant effects.

The proper selection of treatment and control municipalities is a crucial factor for the successful identification of the effects. Although the description of the identification strategy is intuitive, it was a bit hard to understand how the candidate moves from the district-level NGO campaign to the municipal-level analysis. In this respect, for example, the map illustrations and the text could have been clearer and the definition of the districts in the Slovak context comes quite late on page 68. It would have been helpful if the borders of the 79 districts would be more clearly visible, for example, in Figure 3.1. Similarly, Table 3.1 shows the number of municipalities, but the text nearby talks also about districts, thus, it might be useful to also have information on the districts in this table. Further, it is hard to spot the treated and control observation from Figure 3.3.

Because the covariate balance between the treatment and control groups is suboptimal, the study uses the inverse probability weighted (IPW) regression method in the estimation of the treatment effects. I find the choice of the method well suited for the analysis because it can help to reduce bias (covariate misbalance) by adjusting for confounding variables that may be associated with both the treatment and the outcome, and it is not sensitive to the misspecification of the model. However, it is uncertain to me if (or why) only district abortion rates are used in the baseline weighting. Furthermore, the comparison of Tables 3.2 and 3.3 suggest that the weighting used in the analysis did not improve covariate balance noticeably. For example, the difference in abortion rates between control and treatment groups is higher, not lower, in Table 3.3 than in Table 3.2. The candidate has conducted a robustness check that tests whether the district abortion rate is a significant predictor of treatment after using additional control variables in the treatment assignment model. Reassuringly they were not. However, the reader would be more convinced if s/he would see these assignment regression results as well as covariate balance and IPW regression results after using this specification. The candidate briefly notes that other weighting schemes were tried. That is good because the results would be most convincing if the weighted sample would be as balanced in terms of covariate means and distributions as possible (before running the weighted regression model).

As regards the documentation of the results, there is room for improvement in the writing of the results in Tables 3.4–3.8. In Table 3.4 columns 1 and 3 as well as 2 and 4 have the same column headers. Are they all ATEs (Y0, Y1)? What do columns 3-4 report? Furthermore, the table does not say the meaning of the row headers (r1vs0, POmean, TME1, etc.). See also other tables. Columns in Tables 3.4 and 3.5 are in different order. The robustness checks in Tables 3.6 and 3.8 reports average treatment effects for the treated, but I do not think that the candidate has reported ATETs for the baseline results. It also seems to me that the specification is different between Tables 3.4, 3.7, 3.8 and Tables 3.5–3.6 (Abortion rate 2015 included in the table)

Despite some critical comments above, generally, the candidate has conducted many good robustness checks. For example, it makes good sense to try dropping municipalities with the highest concentration of catholic population, because the covariate balance was not perfect in this respect. Even though the results did not change, I recommend showing whether the covariate balance improved after this sample restriction. In this paper, elsewhere where nonsignificant results were found, I would recommend discussing what effect sizes would be feasible and particularly, what type of effect sizes can you rule out.

Conclusion

The doctoral dissertation under review provides clear contributions to the empirical literature on family economics from a geographical perspective. The candidate demonstrates a comprehensive understanding of the related empirical and theoretical literature, and she shows knowledge of different institutional contexts as well as the use of a variety of datasets and research methods.

The research questions are skillfully investigated using well-motivated modern econometric methods and novel identification strategies for the identification of the causal effects. The dissertation also presents useful map illustrations, and more generally the use of spatial distances and approaches are common themes. In each paper, the main results are accompanied by supplementary analyses that investigate the robustness of the results. It is recommended to continue this practice and augment the appendices in future versions of the unpublished papers.

One of the papers has already been published in an international peer-reviewed journal and another paper is single-authored. Thus, the candidate has known that she can conduct independent scientific research. In my opinion, the main limitations of the dissertation relate to the documentation of some findings, rather than its results themselves. Therefore, I regard the my comments as suggestions for improvement, not as binding comments.

Based on the assessment above, we conclude that, in its present form,

a) the dissertation satisfies formal and content requirements for a doctoral dissertation in economics, and
b) I recommend giving Kristína Hrehová the permit to publish and publicly defend her doctoral dissertation.

In Jyväskylä, 28 April 2023

____

Mika Haapanen
Associate Professor of Economics
University of Jyväskylä, Finland