Novel Approaches to Estimating Mobility and Intergenerational Transmission of Well-Being in Brazil[1]

Beatriz Rache[2] and Vinicius Peçanha[3]

1. Introduction

Brazil is one of the most unequal countries in the world and the most unequal country in Latin America (World Bank, 2018). However, research of intergenerational mobility is hindered by the lack of linked data over time. Estimates of intergenerational transmission of education and income in Brazil have been limited to survey data in special supplements of PNAD (Mahlmeister et al., 2017; Ferreira and Veloso, 2013). Importantly, there is growing evidence that intergenerational mobility estimated by surveys can be biased with respect to the true parameter estimated by administrative datasets (Meyer and Mittag, 2019; Gutiérrez, 2020).

In this project, we will gather several sources of administrative, individual-level data from different time periods to create novel and unique data linking parents and children. We aim to document intergenerational mobility of education in Brazil and explore heterogeneities of the estimation with respect to race, gender, neighborhood, and occupation of both parents and children. The dataset we construct overcomes the barrier of the absence of a single, longitudinal administrative dataset that links individuals' outcomes with parental outcomes by using cutting-edge probabilistic linkage techniques (Abramitzky et al, *forthcoming*; Fu et al., 2020).

With this rich data, we are able to precisely identify intergenerational transmission of education for a subset of the population. First, we provide stylized facts about the transition matrix for education in Brazil using a national dataset (*Cadastro Único)* that provides socioeconomic information for individuals and household members that are eligible for social programs. We are particularly interested in differences in intergenerational mobility across geographical regions in Brazil and between rural and urban areas. Moreover, we can analyze heterogeneities with respect to race and gender.

Second, we use several novel administrative datasets from the state of Rio de Janeiro - the third most populous in Brazil, with over 15.9 million residents, and sixth most unequal, out of 27 states -, to discuss possible mechanisms for the patterns of mobility observed in the state. After the probabilistic linkages among the datasets, we are able to examine socioeconomic and health conditions that correlate to the probability that person joins the formal labor market in the future given that this person was enrolled to receive social programs or studied in a public school during childhood and adolescence.

---

[1] Early-stage project

[2] *Instituto de Estudos para Políticas de Saúde* (IEPS)

[3] Vancouver School of Economics, University of British Columbia (UBC)

2. Data

This analysis will draw from different datasets, available for different periods. Nationally, it will rely on *Cadastro Único* between 2012 and 2017 and RAIS (*Relação Anual de Informações Sociais*) between 1985 and 2018. *Cadastro Único* is a centralized registry of individuals and their household members with socioeconomic characteristics, used to determine eligibility for social programs such as the cash transfer program *Bolsa Família.* Families are required to report any changes to their registries, and the Ministry of Social Development claims to run an annual cross-check with other administrative records to ensure all information remains up to date. RAIS comprises a universal registry of all formal employees in Brazil, with information on their salaries and establishments.

For Rio de Janeiro state, we will use the Ministry of Health's System of Information on Births (SINASC), comprising all births in the state. Lastly, for Rio de Janeiro city, we will analyze data from the Municipal Secretariat of Education on all students enrolled in the municipal, public school system between 2003 and 2014. This data contains relevant socio-economic information such as parental education achievement and occupation, school attended and the neighborhood where the individual lives. These 5 datasets are summarized in Table 1.

Table 1

Dataset	**Information**	**Source**	**Period**	**Geography**	**Identifiers**
Cadastro Único	Socioeconomic characteristics on individuals and their household members to assess eligibility for social programs	Ministry of Social Development	2012-2017	National	Name, Mother's name, Father's name, date of birth, Social Security Number
RAIS	Formal workers' history of salary, occupation, and education	Finance Ministry	1985-2018	National	Name, date of birth, Social Security Number
SINASC	Universe of births and birth outcomes	Health Ministry	2003-2017	Rio de Janeiro State	Name, Mother's name, Father's name, date of birth

Administrative data from Municipal Education Secretariat	Universe of students enrolled in Municipal schools	Rio de Janeiro's Municipal Education Secretariat	2003-2014	Rio de Janeiro Municipality	Name, Mother's name, Father's name, NIS, date of birth

3. Methodology

In order to link individuals and parents and individuals over time, we perform probabilistic linkages using the name of the parents, individual's name and date of birth and other invariant variables over time.

We will perform probabilistic linkages between datasets using the identifiers listed in Table 1. For string variables such as an individual's name, date of birth and mother's name, we will allow for up to 0.1 Jaro-Wrinkler string distance in individual and mother's name between datasets. This 0.1 distance threshold between strings allows for misspellings and abbreviations that might prevent exact matches between records and is a standard threshold throughout the literature (Abramitzky et. al, *forthcoming*).

After the linkage process, we will compile a list of stylized facts about intergenerational mobility of education and other measures of well-being from national data, more broadly, as well as Rio de Janeiro state and municipality.

Firstly, for transmission of education, we will split all individuals aged 15 or above in *Cadastro Único* by whether they completed *Ensino Fundamental* (with expected age of completion of 14). Then, analyzing all cohabiting parent-children combinations – considering the highest parental level of education when both live with the child -, we will calculate the share of children who fall in the same education group as their parents. We will also be able to compare these rates across states, and between black and non-black parents.

For the datasets in the city of Rio de Janeiro, we propose the following algorithm to match an individual's presence in the formal labor market with socioeconomic variables. For each cohort defined by the year of birth, we select individuals present in the employee-employer matched data for the formal labor market (RAIS). Using the linkage methodology described above, we link these individuals with the data from the universe of students enrolled in public education at some point in life and with data from *Cadúnico.*

With these linkages, we are able to define a dummy that turns one if the individual joins the formal labor market in the future and, then, we regress this variable on several socio-economic characteristics at the individual level.

A third potentially illuminating analysis relies on SINASC, containing all births in the state of Rio de Janeiro. By linking a cohort's mothers to their respective mothers, we can assess the

intergenerational persistence of teenage pregnancy. This reoccurrence, documented elsewhere (Kahn et al., 1992; Ferraro et al., 2013), is disruptive to young women's likelihoods, and likely further reinforces trends in other aspects of intergenerational mobility.

4. Conclusion

Previous research in intergenerational mobility in Brazil relied on survey data. We use rich administrative datasets to provide stylized facts of intergenerational mobility of education in Brazil and socio-economic factors correlated to mobility related to the formal labor market for the city of Rio de Janeiro, respectively.

This research framework is not without limitations. Given short time spans for some of the data series, some stylized facts will only pertain to a point in time. Secondly, *Cadastro Único* is a registry of individuals below the poverty line who are seeking eligibility to social programs. Any education intergenerational transmission rate catalogued in the project will be biased towards this population. Nevertheless, this is arguably the population with the most to gain from mobility and socioeconomic policies that foster it, reinforcing the importance of our research lens. Moreover, understanding factors correlated to the presence in the formal labor market may shed light to specific mechanisms related to mobility to this subset of population.

References

Abramitzky, R., Boustan, L. Eriksson, K., Feigenbaum, J., and Perez, S. *Forthcoming*. Journal of Economic Literature.

Ferraro, A. A., Cardoso, V. C., Barbosa, A. P., Da Silva, A. A., Faria, C. A., De Ribeiro, V. S., Bettiol, H., & Barbieri, M. A. (2013). Childbearing in adolescence: intergenerational dejà-vu? Evidence from a Brazilian birth cohort. BMC pregnancy and childbirth, 13, 149. https://doi.org/10.1186/1471-2393-13-149

Ferreira, S. G., and Veloso, F. A. 2003. "Mobilidade Intergeracional de Educação no Brasil". Pesquisa e Planejamento Econômico. V.33. n.3

Fu, H., Gosh, A., Hwang, S. and Squires, M. 2021. "Census Linking: A Bounds Approach". *Mimeo.* University of British Columbia.

Gutiérrez, P. C. 2020. "Essays on economic inequality, income taxes, and intergenerational mobility". PhD Thesis. University of British Columbia.

Kahn, J., and Anderson, K. 1992. Intergenerational Patterns of Teenage Fertility. *Demography, 29*(1), 39-57. Retrieved February 22, 2021, from http://www.jstor.org/stable/2061362

Mahlmeister, R., Ferreira, S. G., Veloso, F., Menezes-Filho, N., and Komatsu, B., K. 2019. Revisitando a Mobilidade Intergeracional de Educação no Brasil. *Revista Brasileira de Economia*, *73*(2), 159-180. Epub July 22, 2019.

Meyer, B. D., and Mittag, N. 2019. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net." *American Economic Journal: Applied Economics*, 11 (2): 176-204.