

This “readme” document illustrates the data used and all the steps that are necessary to replicate the main results in the paper “Intergenerational mobility in the very long run: Florence 1427-2011” (included the appendix) by Guglielmo Barone and Sauro Mocetti.

Data Availability and Provenance Statements

In the following, we describe the data used for the results reported in the Tables and Figures of the paper and, for each of them, a “Data Availability Statements”. Further details about the data are included in the online appendix of the article.

1. The **1427 Census** for the city of Florence, including information on the socioeconomic status of the pseudo-ancestors. Data can be downloaded (through an SQL interface) at <http://cds.library.brown.edu/projects/catasto/newsearch/sqlform.php>. More information on the data can be found at <http://cds.library.brown.edu/projects/catasto/overview.html> and, in particular, the list of the variables and their informational content can be found at http://cds.library.brown.edu/projects/catasto/newsearch/catasto_codebook.html.¹ We enriched the 1427 Census with estimates of the earnings, which are attributed to the household heads on the basis of the occupations and the associated skill group. These estimates have been produced and kindly shared by Peter Lindert (University of Davis). To gain access to the data please contact Peter Lindert (phlindert@ucdavis.edu).
2. The **2011 tax records** for the city of Florence, including information on various items of incomes and the main demographic characteristics (age and gender). The income items reported include salaries and pensions, self-employment income, real estate income, and other smaller income items at the surname level. We define as earnings the total income net of real estate income, while real wealth has been estimated from real estate income.² For confidentiality reasons, the data are collapsed at the surname (instead of individual) level and only surname with at least five occurrences are made available. To access the data please contact the statistical office of the municipality of Florence (https://www.comune.fi.it/statistica?language_content_entity=it).
3. Data on **surnames’ frequency of taxpayers by Italian provinces in 2005** have been kindly shared by Giovanna Labartino and has also been used by Pellizzari et al. (2011). To access the data, please contact Giovanna Labartino (g.labartino@confindustria.it) or Michele Pellizzari (michele.pellizzari@unige.ch).
4. Data on **surnames’ frequency of residents in the municipality of Florence in 2010s** are available on the website of the statistical office of the municipality of Florence (https://opendata.comune.fi.it/?testo=cognomi&q=metarepo%2Fdataset_results)

¹ The documentary sources are fully described by D. Herlihy and C. Klapisch-Zuber (1985), *Tuscans and their Families: A Study of the Florentine Catasto of 1427*, New Haven: Yale University Press.

² Such estimation is based on data taken from the Survey of Household Income and Wealth (see point 6 below) according to the procedure outlined in the paper.

5. Data on the number of individuals with certain surnames and employed in certain elite occupations are collected through several data sources. Namely, (i) individuals of the liberal professions are drawn from the archives of the provincial professional organizations: see <http://www.consiglionazionaleforense.it/site/home.html> for lawyers, <http://www.ordine-medici-firenze.it/> for medical doctors and <http://www.ordinefarmacisti.fi.it/> for pharmacists; (ii) individuals belonging to the governing bodies of the banks are drawn from the OR.SO (*Organi Sociali*) archive, managed by the Bank of Italy; please contact the authors to have information on how to access the data; (iii) goldsmiths are drawn from the National Business Register database. To access the data please follow the instruction on the website of the proprietary data at <https://www.infocamere.it/en/accesso-alle-banche-dati>.
6. The **Survey of Household Income and Wealth** is managed by the Bank of Italy and collects information on income and wealth of Italian households. More information on the survey can be found at <https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/index.html?com.dotmarketing.htmlpage.language=1>. As Bank of Italy employees, we have privileged access to confidential records on the province of residence and of birth of each interviewee. To have access to the same data please contact statistiche@bancaditalia.it.

The replication of most of our results need the access to confidential microdata that we cannot provide as part of the replication archive. Information on how to gain access to each data sources are mentioned above. The authors will assist with any reasonable replication attempts for two years following publication.

In the following, we report a description of the data obtained from the association (and anonymization) of the original data sources mentioned above.

Dataset list of the replication package

In terms of **software requirements**, all the elaborations are performed with Stata 16.

The replication package includes five datasets (in dta format) and one do file that allows to replicate most of the tables and the figures contained in the paper. In the following, we report more details.

The **datasets** are obtained as a combination of the original data sources described above and, for confidentiality reasons, surnames have been anonymized. All the datasets included variable with labels describing their content and meaning.

Name	Description of the dataset
<i>restud_1427_micro.dta</i>	This dataset contains information on incomes at the individual level for household heads surveyed in the 1427 Census. The dataset is obtained from the data source (1).
<i>restud_1427.dta</i>	This dataset contains main socioeconomic variables (e.g. earnings, wealth, main occupation, etc.) at the surname level for all surnames included in the 1427 Census and the same information for the subset of these surnames that can be matched with the 2011 tax records. The dataset is obtained from a combination of data sources (1) and (2).
<i>restud_2011.dta</i>	This dataset contains main socioeconomic variables (e.g. earnings, wealth, etc.) aggregated at the surname level for all surnames included both in the 1427 Census and in the 2011 tax records. The dataset is obtained from a combination of data sources (1), (2) and (3).
<i>restud_occupations.dta</i>	This dataset contains information on type of occupation (for a selected set of occupations) at the individual level for taxpayers of the province of Florence whose surname can be matched to those of the 1427 Census. The dataset is obtained from a combination of data sources (1), (3) and (5).
<i>restud_surname_2011.dta</i>	This dataset contains information on the characteristics of the surnames (e.g. length and indicators of complexity) for all the surnames of the current Italian taxpayers of the municipality of Florence. The dataset is obtained from the data source (4).

The **do file** *restud.do* allows to replicate all the tables and the figures with the exception of Table 7 and Figure 1. The do file includes step-by-step instructions on the nature of the exercise and the type of output. Moreover, it produces a log file with a clear description of the output all the references to the Tables and Figures reported in the article.

Instructions to Replicators

Anyone who wants to replicate our results must simply perform the following steps:

- Edit *restud.do* to adjust the default path. This code automatically opens datasets, runs regressions and saves the output in the working directory defined at the beginning of the do file.
- Run *restud.do* that automatically will produce a log file with main results (to be included in the Tables) and the Figures (*.tif format).
- Please see the Table below for more details on the reproduction of Tables and Figures.³

Output	Dataset of the replication package	Lines of the do file <i>restud.do</i>
Table 1	<i>restud_1427</i> and <i>restud_2011</i>	Lines 19 to 53
Table 2	<i>restud_2011</i>	Lines 57 to 70
Table 3	<i>restud_2011</i>	Lines 74 to 218
Table 4	<i>restud_2011</i>	Lines 222 to 263
Table 5	<i>restud_2011</i>	Lines 267 to 341
Table 6	<i>restud_2011</i>	Lines 345 to 433
Table 7	n.a.	n.a.
Table 8	<i>restud_2011</i> and <i>restud_surnames_2011</i>	Lines 437 to 558
Table 9	<i>restud_1427</i>	Lines 562 to 616
Table 10	<i>restud_1427</i>	Lines 648 to 771
Table 11	<i>restud_2011</i>	Lines 775 to 849
Table 12	<i>restud_occupations</i>	Lines 853 to 875
Figure 1	n.a.	n.a.
Figure 2	<i>restud_1427</i>	Lines 622 to 642
Figure 3	<i>restud_1427_micro</i>	Lines 919 to 1021

Acknowledgements

We thank Giovanna Labartino and Michele Pellizzari who kindly shared the data on the distribution of surnames across provinces. We thank Peter Lindert who kindly share estimates of incomes for household heads in 1427. We finally thank Riccardo Innocenti and Massimiliano Sifone of the municipal statistical office of Florence for providing us with tax records for 2011 at the surname level.

³ Table 7 is not replicable because is based on proprietary data. To gain access to this data see the data availability statements for the data source (6). Figure 1 is not replicable for confidentiality reasons (surnames need to be anonymized).