

# *Gatsby in the country of **The Leopard**.*

## **Social Mobility across Italian Provinces in the 1930s**

Giacomo Gabbuti, St. Antony's College, University of Oxford  
(giacomo.gabbuti@sant.ox.ac.uk)  
Supervisor: Dr. Brian A'Hearn, Pembroke College, Oxford

Social mobility is an increasingly important issue in modern societies, the object of a growing economic and historical literature. In this literature, Italy currently features among the least mobile advanced economies; an outlier in the 'Great Gatsby curve' – the empirical association between income inequality and intergenerational elasticity (Corak 2013). Recently, Güell et al. (2018) showed that a similar relationship exists even within Italy, where provinces experience very different degrees of mobility, despite their minimal institutional differences. Almost nothing is known, however, on the *history* of social mobility in Italy. In a methodological chapter of my dissertation (Gabbuti 2021), I discuss the available sources and methodologies to construct historical estimates of inter- and multi-generational income elasticity for modern Italy, for the period between the unification of the country (1861) and WWII. In this chapter, building on the literature that has explored the possibility of using surnames to match information on different generations - and in particular, on the so-called Informative Content of Surnames (ICS) index, developed by Güell et al. (2014) - I estimate the intergenerational elasticity of income across the 92 Italian provinces in the 1930s, in a way that make them comparable with those estimated by Güell et al. (2018) for 2005. I do so by documenting an innovative source: the taxpayers lists, issued by the Italian Ministry of Finance, and reporting assessed incomes, occupation, and locality of residence for more than one million individuals. The present paper briefly documents this new source and showing the first results, revealing the existence of a 'Great Gatsby' curve already in interwar Italy. First, I present the methodology - the ICS index.

### **The Methodology: The Informational Content of Surnames (ICS)**

The ICS is one of several indicators, recently developed by researchers interested in intergenerational research to 'overcome data limitations' (Santavirta and Stuhler, 2020). While not specifically developed for historical applications, these methods clearly appeal to economic historians, since the use of surname is meant to overcome the absence of linked, multi-generational data. According to Güell et al. (2014, p. 694), surnames can be used as 'markers': despite being 'intrinsically irrelevant for the determination of economic well-being', they are transmitted by fathers to the next generation, together with those 'other characteristics that do matter'. The more these inherited characteristics explain individuals' outcomes, 'the more information the surnames will contain on the values of outcomes'. For this reason, even in a cross-sample, it is possible to infer the degree of mobility by measuring the informative content of surname in terms of economic outcomes. The intuition is that surnames should not convey information on intrinsic ability, and if they account for systematic regularities in economic outcomes, they must signal some sort of intergenerational transmission mechanism in place.

In practice, Güell et al. (2014, p. 697-8) define the ICS as 'the difference in the  $R^2$  between two regressions'; a first stage (1) in which economic outcome of the individuals  $i$  on their surnames (a  $k$ -vector of dummy variables,  $D$ ), is regressed on and a vector of additional characteristics:

$$y_{i,k} = b'D + \gamma'X_{i,k} + u_{i,k} \quad [1]$$

Then, a second regression (2) 'mixes up the surnames so that they cannot be informative':

$$y_{i,k} = b'F + \gamma'X_{i,k} + u_{i,k} \quad [2]$$

This time, in fact,  $F$  is a vector of randomly assigned surnames, 'in a manner that maintains the marginal distribution of surname'.

ICS can then be defined as:

$$ICS = R_L^2 - \bar{R}_F^2 \quad [3]$$

Where  $R_L^2$  is the r-square of regression [1], and  $\bar{R}_F^2$  is the average r-square of n regressions of randomly assigned surnames on the same outcomes, as in [2]. Güell et al. (2014, pp. 697-698) theoretically model the mapping between [3] and a classic father-son IGE: as they explain,  $\widehat{ICS}$  is ‘a moment of the joint distribution of surnames and economic well-being that measures the incremental ICS’, and ‘turn out to be monotonically increasing’ in the IGE.

In their insightful discussion of name-based estimators, Santavirta and Stuhler (2020, p. 11) pointed out that the ‘key advantages of the R2 estimator are its modest data requirements and its broad applicability’; indeed, what makes the ICS particularly interesting for economic history, is the possibility of estimating social mobility by means of a single source. It is designed for taking advantage of large contemporary administrative databases, such as a single census, or the 2005 Italian taxpayers lists, made available by the Ministry of Finance in anonymised forms, and exploited by Güell et al. (2018) to estimate income mobility across Italian provinces. Acciari et al. (2019, p. 29) empirically confirmed the reliability of the methodology, comparing their own estimates, based on observed links, to Güell et al. (2018), concluding that ‘the ICS is a very strong proxy of upward mobility that can be used in many contexts where better data are not available’. Indeed, another recent paper, Barone and Mocetti (2020), while using surnames to ‘link’ current taxpayers to the families recorded in the Florentine Catasto of 1427, adopted the ICS to get two separate estimates, showing the sizeable reduction of intergenerational elasticity across centuries. In fact, compared with more common ‘grouping estimators’ such as those adopted by Barone and Mocetti, the ICS is comparatively ‘less sensitive to sampling properties or sample size’ (Santavirta and Stuhler 2020), making it a perfect candidate not only for historical adoption, but also ‘to characterize the geography of intergenerational processes in greater detail’, in line with the influential work by Chetty et al. (2014). This paper contributes to the historical literature on social mobility by presenting one of the first province-level estimates of income mobility, for the 92 provinces of Italy in 1933, based on a very similar source to the one adopted by Güell et al. (2018) for 2005.

### **A New Source for Estimating Social Mobility in Italy: The Taxpayers’ Lists**

In fact, 2005 was not the first year in which the Ministry of Finance published lists of Italian taxpayers. In Gabbuti (2021), I document the existence of several lists of taxpayers, of different coverage and quality, from the first days of the Kingdom of Italy, to the post-WWII years. The most suited for our purposes are the lists of the incomes filed by individuals (as opposed to legal entities) for two categories of the main income tax of the times, the *Imposta di ricchezza mobile*. As most of 19<sup>th</sup>-and early 20<sup>th</sup>-century taxes, the *Imposta di ricchezza mobile* was not a general tax on personal incomes and taxed each source of income separately: categories B and C taxed incomes from business, self-employment (although mainly outside agriculture), and professions. Being considered the least ‘certain’ of all incomes, lists of taxpayers for these categories were issued several times, in an effort of raising public scrutiny and compliance - precisely as in 2005. For instance, the Bank of Italy asked its local branches to make a copy of the lists available to the public in their offices. In particular, a first publication of all taxpayers, in one separate volume for each province, came out in 1889. The publication resumed in the interwar period, when the first Finance Minister of Mussolini’s fascist government, Professor Alberto De’ Stefani, declared his ‘war’ on tax evasion (Gabbuti 2020a). After a first list published in 1922, new and more complete lists were published in 1933, including now more than one million individual taxpayers from all the 92 provinces of the Kingdom of Italy – at the time including the three Croatian cities of Fiume (Rijeka), Pola (Pula), and Zara (Zadar), annexed after the Great War.

The very nature of the source makes clear, if needed, that it should not be considered ‘perfect’: evasion, either total or partial, is definitely an issue when dealing with fiscal data. However, as discussed extensively in other chapters of my dissertation - Gabbuti (2020a) and Gabbuti and Morelli (2021) for, respectively, top incomes and estates tabulations, and Gabbuti (2021) and Gabbuti and Gómez León (2021) with reference to this specific source - Italian historians seem to have exaggerated

in their scepticism for all fiscal sources, at least in comparative terms. While all historical sources have biases that should be acknowledged, and whenever possible, adjusted for, there seems to be no reason to completely discard Italian fiscal evidence, as most economic historians have done so far. The average incomes declared by these taxpayers show a reasonable mark-up with the wages paid in the correspondent sectors, especially when considering that these were ‘net’ of some deductions. Even more importantly for the estimation of inequality and social mobility, in line with anecdotal evidence from qualitative sources, these sources seem to track sufficiently well the relative ranking of professions and individuals between them: at the province level, average declared incomes correlate very well with the industrial value added estimated by Ciccarelli and Fenoaltea (2013); in terms of interpersonal comparisons, Banti (1996) used a very small sample of the lists to show relevant differentials in the income of professionals working in different cities, and in urban versus rural areas; or to follow notable individuals over time, as in the case of the opera composer Giuseppe Verdi, whose income seemed to grow consistently with his fame. It is not surprising, after all, that the statistician Corrado Gini and his collaborators adopted the tabulations of these lists to perform some of the first applications of his inequality index (Gabbuti 2020b).

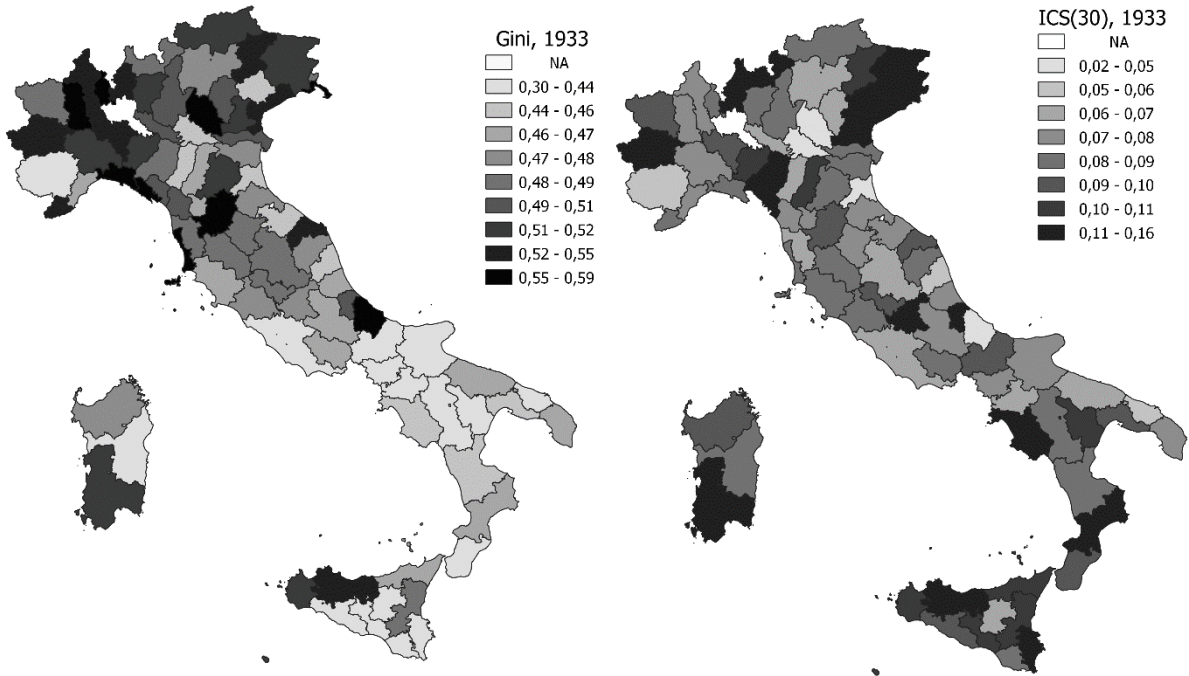
As discussed in Gabbuti (2021), the 1933 lists are not only the most convenient to digitise for the purposes of applying of surname-based names – given the clear structure of their layout, and that surnames are printed in capital letters and in a very clear font – but are also the most complete and reliable, given the inclusion of more categories of taxpayers and the sustained efforts of tackling evasion, both contributing to the increase in the number of incomes reported, but also the period of monetary stability. The paper presents the first result coming from the digitisation of the lists, performed by scanning all the 92 volumes, first applying an OCR-software, and then checking the result. While the potential of the source goes well beyond that – making possible, for instance, to see the presence and status of female entrepreneur, or to estimate municipal-level indicators of average income and concentration - the application of the ICS to these lists is documented in the final part of this paper.

### **Social Mobility in Interwar Italy: First Results and Correlations**

The ICS is particularly suitable for obtaining comparable province-level estimates of intergenerational income elasticity using the taxpayers lists. As discussed by Barone and Mocetti (2020, p. 6-7), Italian surnames are extremely informative, given their long history, often dating back to the Middle Ages (Bizzocchi 2014). Indeed, their distribution is way more skewed than most other European ones, and most of the time they are markedly ‘local’: for this reason, the application of surname-based indicators at the local level is particularly robust, especially when reducing the sample to the surnames not exceeding a given frequency. This paper, consistently with both Barone and Mocetti (2020) and Güell et al. (2018), presents estimates for surnames with less than 30 occurrences at the provincial level.

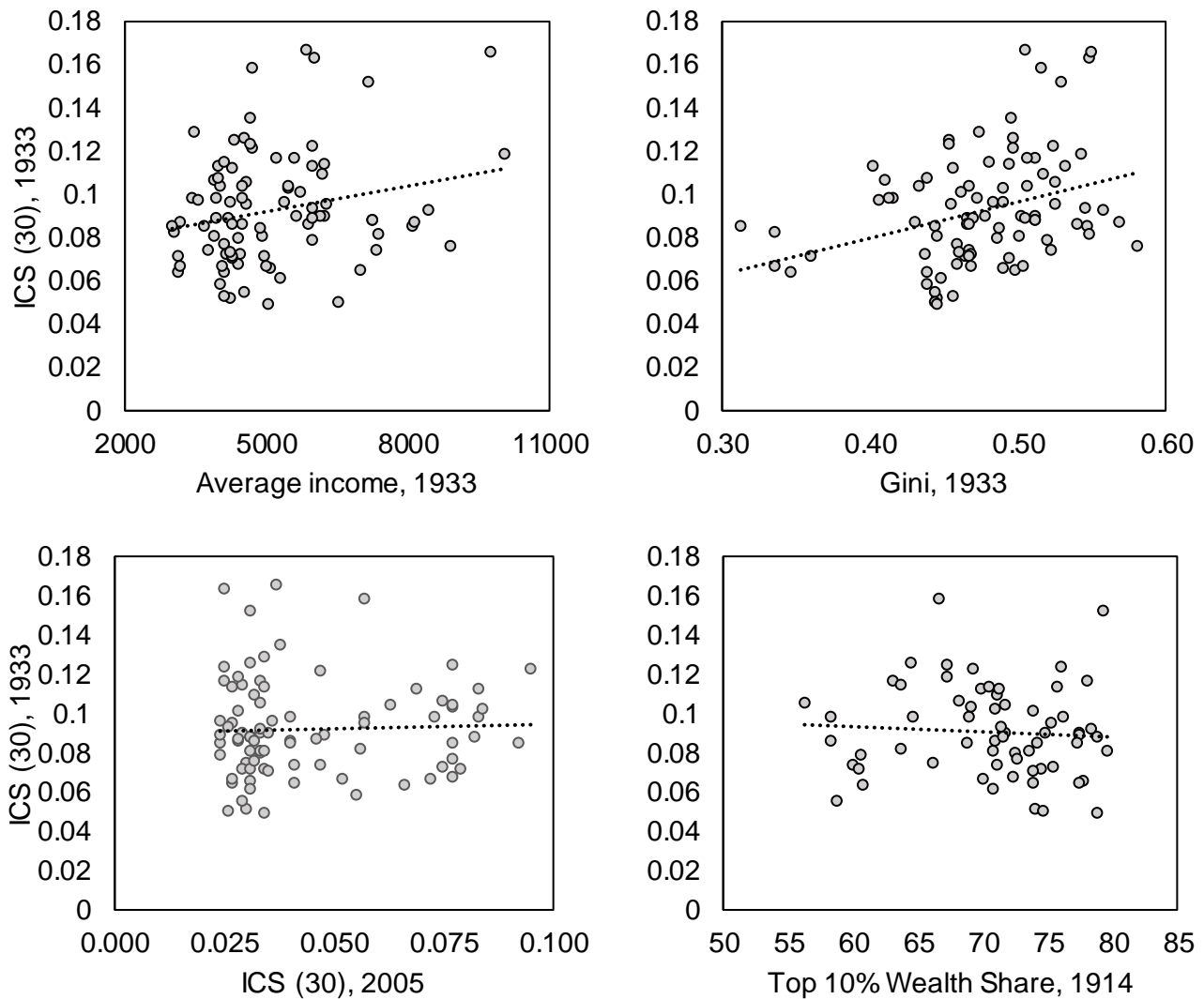
It should also be noted that in the great majority of the times, Italian women maintained their surnames after marriage, so they can be meaningfully included in the estimates. In line with the recommendation of Santavirta and Stuhler (2020), we should remember that more than to the numerosity of the sample, our estimates are very sensitive to the underlying nature of the sample. In this sense, the focus on self-employed and professionals, is a limitation, but also a reason of interest. Extremely relevant in the employment structure of Italy, from 19<sup>th</sup> century to present (Gabbuti 2020c), these groups were historically the most mobile, covering the ‘typical’ trajectories of ‘rag to riches’ experienced by Italian entrepreneurs (Banti 1996, pp. 143-55). On the other hand, in current times, their relevance contributes to explain the ‘stickiness’ of Italian society (Ballarino et al. 2016). Keeping this in mind, figure 1 maps Gini and ICS indices, obtained from the 1933 taxpayers lists; the correlations of the ICS with the Gini, but also the average assessed income, the 2005 estimates by Güell et al. (2018), and top-wealth shares from Gabbuti and Morelli (2021), are shown in figure 2.

**Figure 1 - Income Inequality (Gini) and Social Mobility (ICS) in Italy, 1933**



Source: author's elaborations.

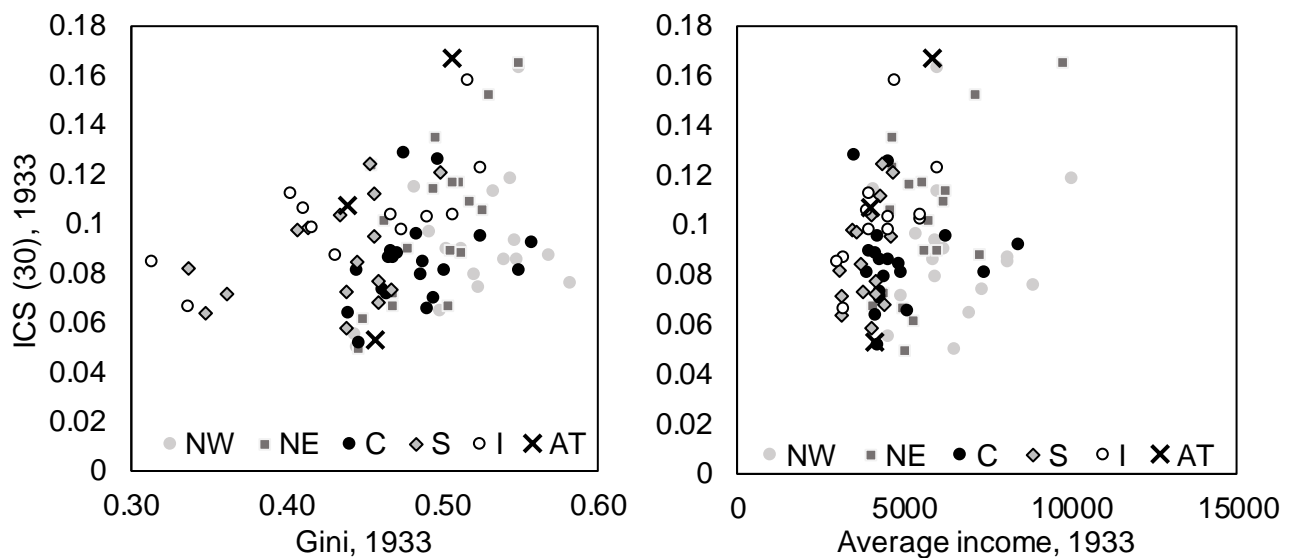
**Figure 2 - ICS in 1933, Mean Incomes, Inequality: First Correlations**



Source: author's elaborations.

The results are still preliminary, since at the moment of writing this paper, Naples and Milan are still being checked, and some of the other biggest provinces, such as Rome are not entirely cleaned. As a result, the total number of observations is 1,026,546, over an expected total of c. 1.3 million. The final paper will also include the robustness checks with other specification and cut-offs, in line with Santavirta and Stuhler (2020) recommendations. Still, the evidence reveals a high degree of variation across provinces. In line with the historical evidence on Italian regional divides, this level of observation reveals within-area differences – in inequality, and even more in ICS. The graphs show the existence of some correlation between the ICS and average income, as well as its concentration – the aforementioned ‘Gatsby curve’. The same does not hold, however, for wealth concentration, nor for present day mobility. Notably, as shown in figure 3, the same relationships seem to be present within each macro-area, making extremely interesting to test, as in Güell et al. (2018), for the correlation with a broader set of socio-economic characteristics, including educational attainment, social capital, and employment structure.

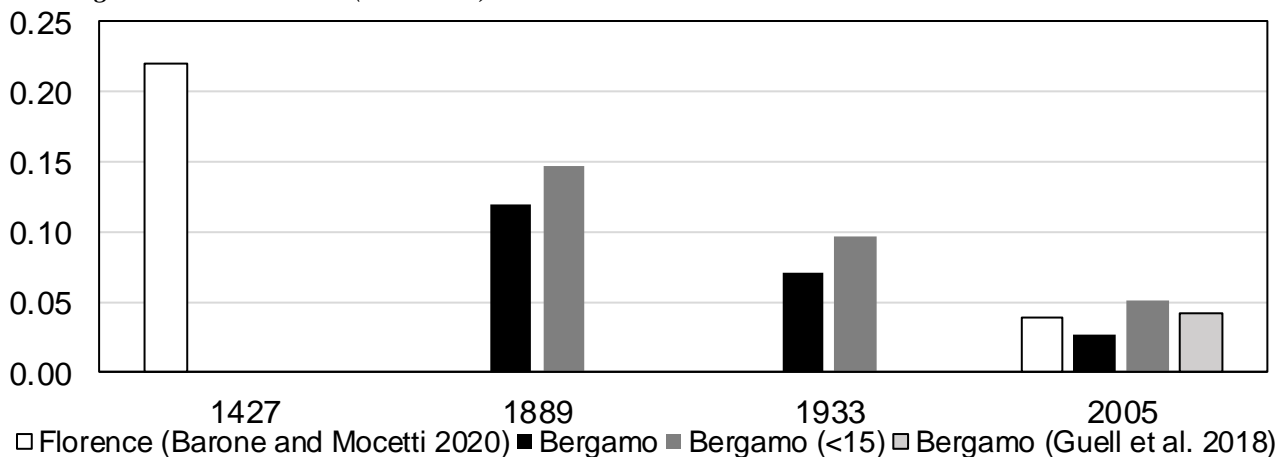
**Figure 3 – Gini, Average Incomes, and ICS by Macro Areas**



Source: author’s elaborations. Macro-areas are defined as North-West, North-East, Centre, South, Island, and Annexed Territories – the provinces of Pola, Fiume, and Zara, annexed by Italy after WWI (not showed in fig. 1).

With the limitations and caveats briefly exposed above, the paper also explores the evolution of social mobility over time, extending the methodology to selected provinces in 1889 and 1922. As shown by figure 4, when compared with Barone and Mocetti’s and Güell et al. (2018), the ICS can be meaningfully adopted to reveal the long-run increase in social mobility associated with economic development - even in the land of Tomasi di Lampedusa’s *The Leopard*, where ‘everything must change for everything to remain the same’.

**Figure 4 - ICS Estimates (1427-2005)**



Source: author’s elaborations.

## Selected References

- Acciari, P., Polo, A., Violante, G.L. (2019). “And yet, it moves”: Intergenerational mobility in Italy. *NBER Working Paper 25732*.
- Ballarino, G., Barone, C., Panichella N. (2016). Origini sociali e occupazione in Italia. *Rassegna Italiana di Sociologia*.
- Banti, A.M. (1996). *Storia della borghesia italiana*, Roma, Donzelli.
- Barone, G., Mocetti, S. (2020). Intergenerational Mobility in the Very Long Run: Florence 1427–2011, *Review of Economic Studies*, rdaa075, <https://doi.org/10.1093/restud/rdaa075>
- Bizzocchi. R. (2014). *I cognomi degli Italiani*, Bari, Laterza
- Chetty, R., Hendren, N., Kline, P., Saez, E. (2014), Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States, *Quarterly Journal of Economics*, 129(4), 1553–1623.
- Ciccarelli, C., Fenoaltea, S. (2013). Through the magnifying glass: provincial aspects of industrial growth in post-Unification Italy, *Economic History Review*, 66(1): 57-85.
- Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3): 79-102.
- Gabbuti, G. (2020a). *A Noi!* Income Inequality and Italian Fascism: Evidence from Labour and Top Income Shares, *Oxford Economic and Social History Working Papers 177*.
- Gabbuti, G. (2020b), Between Pareto and Gini: Notes on the Emergence of Modern Measurement of Personal Income Inequality in Italy (c. 1890-1940), *History of Political Economy*, 52(3): pp. 435-454.
- Gabbuti, G. (2020c), ‘Labor Shares and Inequality: Insights from Italian Economic History, 1895-1970’, *European Review of Economic History*, doi:10.1093/ereh/hea010.
- Gabbuti, G. (2021). Sources and Methods for the History of Intergenerational Mobility in Italy, mimeo.
- Gabbuti G., Gómez-León, M. (2021). Wage Inequality in Italy (1900-1950): An Approach from Dynamic Social Tables, mimeo.
- Gabbuti G., Morelli, S. (2021). Inheritance, Wealth Concentration and Regional Divides in Italy during the First Globalisation, mimeo.
- Güell, M., Pellizzari, M., Pica, G., Rodríguez-Mora J.V. (2018). Correlating Social Mobility and Economic Outcomes. *The Economic Journal*, 128.612: F353-F403
- Güell M., Rodríguez-Mora J.V., Telmer, C.I. (2014). The informational content of surnames, the evolution of intergenerational mobility, and assortative mating, *Review of Economic Studies*, 82(2): 693-735.
- Santavirta T, Stuhler J. (2020). Name-Based Estimators of Intergenerational Mobility, mimeo.