UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

CAPP - Centro di Analisi delle Politiche Pubbliche

CAPP*aper*

n. 181 | aprile 2023

Surnames in local newspapers and social mobility

Massimo Baldini, Andrea Barigazzi

Dipartimento di Economia Marco Biagi
www.capp.unimore.it

Surnames in local newspapers and social mobility

Massimo Baldini[α], Andrea Barigazzi[β]

March 2023

Abstract

This paper exploits an innovative data source, the surnames contained in local newspapers during almost a century in the Italian Province of Modena (NUTS-3 level), to study the phenomenon of social mobility over time. Based on the hypothesis that the surnames that appear in the newspapers have a particular social relevance, the study of changes in the frequency distribution of the set of surnames over time allows to identify periods of greater or lower social mobility. The results show that the periods of greatest change have been the years of transition between democracy and the fascist regime and vice versa, and the 1980s. A strong regression towards the mean in the relative importance of surnames is also observed.

1. Introduction[1]

The study of social mobility between generations has always stimulated academic debate (among others Ganzeboom et al., 1991). However, the scarcity and difficulty in finding data linking generations over time has held back for a long time the development of research on this theme. It is only in recent decades that the empirical literature on intergenerational mobility has made considerable progress, mainly due to the availability of new and better data[2] linking adjacent generations (Solon, 2018).

Long term social mobility, which studies the relationship between three or more generations, has received less attention from scholars because of data limitations (Solon, 1999; Black and Devereux, 2010). This has led in parallel to the use of alternative tools to overcome this obstacle. Among these, there is a growing interest in the exploitation of surnames in studying social mobility, used as pseudo-links connecting households of different generations (Barone and Mocetti, 2021). The

[α] University of Modena and Reggio Emilia, Department of Economics "Marco Biagi", massimo.baldini@unimore.it
[β] University of Modena and Reggio Emilia, Department of Economics "Marco Biagi", andrea.barigazzi@unimore.it
[1] We thank Giovanni Bigi of the Statistical office of the Municipality of Modena and Christian Antonietti, Kashif Riaz and Mattia Savigni for their very appreciated research assistance.
[2] These types of data mainly refer to: (i) cross-sectional data with retrospective surveys on the social status of parents; (ii) panel surveys with detailed and repeated information on the social and economic status of household components followed over extended periods of time; (iii) fiscal/administrative data linking economic information between parents and children (Mooi-Reci, 2020).

present work builds on this literature by introducing a novel data source for the analysis of social mobility: newspapers data.

We investigate the frequency distributions of surnames in newspaper articles, with the hypothesis that those who are frequently mentioned in newspapers have social relevance. The measure derived from these data can be used to analyse social mobility over time, bearing in mind that the informational content of a surname can vary greatly depending on the context and type of the article in which it appears. For example, the political section reports on people with power in collective decision-making processes; in crime-related articles, the surnames of both the victims and the criminals (often poorer persons) appear; those who advertise their products/services may be considered richer than the average member of the reference population, just like those who spend themselves on charity work or those who can afford to buy space for an obituary[3].

To ascertain whether there have been variations in the frequency of surnames over the years, the surnames mentioned in local newspapers in the province of Modena (NUTS-3 level) in Italy from 1921 to 2011 are taken into account, together with census data on the spread of surnames in the provincial population. By way of explanation, if there is an over-representation of certain surnames in pages dealing with local politics compared to their registry distribution, then that surname will most likely belong to individuals/families exercising political power at the local level[4]. If this over-representation persists over time, then power is transmitted from one generation to the next.

The purpose of this work is therefore twofold, on the one hand to study social mobility and on the other to provide evidence on the effectiveness of newspaper data in identifying those who most influence society and belong to power elites (economic, political, cultural, scientific, etc.).

As with other papers that exploit the information content of surnames, it is not possible to state with certainty that a surname observed today is part of the direct descendants of the same surname found in previous decades. But, in line with Barone and Mocetti (2021), by focusing on the local rather than the national level, more precise links between generations are produced[5].

To the best of our knowledge, the contribution of this work to the existing literature is threefold. We are the first to combine the use of newspaper data and pseudo-links of surnames to study social mobility and power transmission of élites. Moreover, while the literature on long-run social mobility has traditionally focused on income, education, and occupation outcomes (Solon, 2018), we implement an analysis that is rather concerned with the transmission of authority, power and social influence. Finally, the use of newspapers as a source of historical data has mainly been developed

[3] Results corroborating this approach have been for example highlighted in the literature studying obituaries that, while becoming more commonly used by different socio-economic classes, still reveal the influence of the dominant Western elites, groups linked by common origins and educational patterns (Matsuda, 1996; Fawler and Bielsa, 2007).

[4] Ban et al. (2019) reinforce this approach, showing that the volume of news coverage devoted to political actors or offices helps to indicate their actual importance.

[5] Moreover, the strong heterogeneity and 'localism' of Italian surnames further reinforce the quality of these pseudo-links.

in the USA and the UK, while in Europe, due to the lack of digitised archives, it is still almost unexplored.

The rest of the paper is organised as follows. In Section 2, a brief review of the literature on both the use of newspapers and the use of surnames in economic research (and in particular in the study of social mobility) is proposed. Section 3 traces the economic and political evolution of Italy and Modena from the end of the First World War to the present day. Section 4 explains the data used and provides a check on newspaper data. Section 5 reports the method of analysis and the observed results. Section 6 offers some extensions and robustness analysis. Finally, Section 7 concludes.

2 Literature review

2.1 The use of surnames in economic research and the study of social mobility

A person's surname can tell a lot about the social status of the ancestors and the geographical origin of the family (e.g., the surnames Le Boulanger in France, Becker in the United States and Bäcker in Germany link their owners to their ancestors' occupation as bakers). Furthermore, the information conveyed by surnames may influence the way a person is judged, a fact that economic research has exploited to study discrimination in various fields, for example in the labour or in the rental markets.

Popular surnames receive a higher and more positive evaluation than uncommon and infrequent ones in the labour market (Pascual et al, 2015). Noble-sounding surnames more frequently hold managerial positions than other surnames (Silberzahn & Uhlmann, 2013). A person with an African American-sounding surname is less likely to be called back for a job interview than a person with a white-sounding name (Bertrand & Mullainathan, 2004), as well as anonymous application procedures increased the chances of both women and individuals of non-Western origin of advancing to the interview stage (Åslund & Skans, 2012). Many studies have also found signs of discrimination in rental housing markets, where foreign-sounding names that have shown interest in the ads are less likely to be contacted by flat owners (Ahmed and Hammarstedt 2008).

As an alternative method of measuring intergenerational social mobility, the use of surnames has received increasing interest. Collado et al. (2012), having data on the socio-economic status of individuals in two Spanish regions at the end of the 19th and the end of the 20th century, use pseudo-links obtained from surnames to link ancestors to descendants. They find that having a high level of education and belonging to a high socioeconomic group is still influenced by the socioeconomic status of the great-great-grandparents. Guell et al. (2015), albeit limited to an intergenerational analysis, estimate a correlation of educational outcomes of 0.60 for Catalonia in Spain. Gregory Clark (2014), in one of the most discussed contributions in recent years, estimates that multigenerational mobility follows an autoregressive process with a high and time-invariant persistence rate. Barone and Mocetti (2021) go even further back in time, exploiting a unique

dataset linking the status of Florentines in 1427 (derived from a tax census) to the status of descendants in 2011. The two scholars find that the elasticity of earnings is positive and statistically significant between generations nearly 600 years apart. They also show evidence of even stronger transmission of real wealth and persistence in some elite professions.

2.2 Newspapers

Newspapers can be the source of an incredibly vast amount of easily accessible historic information (Hansen, 2004). In fact, their pages contain the historical memory collected day by day of major political, cultural, economic, scientific, and other events (Tosh, 2010). However, this rich knowledge potential has so far only been partially explored by the scientific community.

Hanlon and Beach (2022) highlight that research in economics, before the arrival of large-scale digitized historical newspaper databases, has predominantly focused on the use of newspapers: (i) as a way of measuring a certain type of treatment in order to construct a key explanatory variable; (ii) as the basis of many price series and financial and commodity markets studies. They also emphasise that the bulk of existing economic studies have focused on data from the United States or the United Kingdom, where most of the digital archives are located. One of the very promising research perspectives is therefore to expand the use of newspapers outside this geographical area.

Newspaper data can be applied to a wide range of topics, but so far only some of them have been explored. Following the outbreak of the Covid-19 emergency, for example, one theme that has received special emphasis is the study of the 1918-19 influenza pandemic. Markel et al. (2007) demonstrate a strong association between the early and stratified implementation of non-pharmacological interventions such as school closures, isolation or quarantine and the mitigation of the negative consequences of the influenza pandemic in the United States. In addition to administrative data, the authors use information from newspapers to verify the type and date of intervention. Again, based on data from local US newspapers, Ager et al. (2022) investigate the impact of school closures to prevent the spread of influenza on long-term school outcomes.

The study of collective action with newspaper data is another theme that has become commonplace in recent decades. Multiple types of collective action, from racial violence protests to various other types of social movements, have been analysed through newspaper-based event data (for a review see Earl et al., 2004).

Through data obtained from newspapers, it is also possible to monitor economic activity in real time (Shapiro et al., 2020; Aguilar et al., 2021). Textual analysis and news sentiment are indeed exploited to measure in which direction macroeconomy is moving. When journalists write more positive words, then the economy will experience an upward trend and vice versa. Somewhat related to news sentiment is the study by Gentzkow et al. (2006), who construct a historical index of corruption based on mentions of 'fraud' and 'corruption' reported in US newspapers.

3. Economic and political context

The period following the First World War was marked in Italy by a relatively rapid recovery, but also by great social and political tensions. First with the so-called "red biennium" (1919-1920) and then with the advent of fascism (1922), Italy moved from the liberal age to the fascist era. The formal beginning of the dictatorship (1925), the adoption of autarchic policy, and the Great Depression of 1929 had a severe impact on the Italian economy. Growth was then further hit by the tragic participation in the Second World War, which caused a real economic collapse. GDP in 1945 was at the level of 1906.

The reconstruction and the subsequent twenty years were the era of the economic miracle. Convergence achieved its most important steps in these years: the mixed economy model worked, the historical gaps between North and South were partly alleviated, GDP grew at very high rates (never repeated) as well as industrial production and labor productivity. Starting from the mid-1960s, the advantages deriving from backwardness (such as ample labor supply and low labor costs) began to run out. In the early 1970s, growth sharply slowed, also due to international factors such as the change in the monetary order and the first oil crisis. At the same time, the socio-political situation experienced strong tensions throughout the 1970s. Between the late 1970s and the early 1990s progress resumed at sustained rates, and in 1992 Italy had the same per capita GDP as Germany and the United Kingdom. Convergence seemed complete. However, this period was also characterized by contrasting aspects. In no Western country did public debt grew so much, territorial disparities increased again, and the political system entered into crisis.

From the 1990s onwards, Italy has experienced three decades of extremely weak economic growth. The country has been unable to sustain its competitiveness in a globalized world, falling behind in technological innovation, education, and research. The situation was further exacerbated from 2000 to 2010, mainly due to a decline in productivity. By 2011, Italy's GDP was only 1.1% higher than ten years before, and the gap with other developed countries returned.

The economic history of Modena and its province has roughly followed the progress recorded nationally, but with better results in periods of growth. The twenty-year fascist period was very complex for the province and marked by high unemployment. From the post-war period onwards, on the other hand, Modena experienced a prosperity never seen before. The real per capita income of Italians from 1950 to 1990 increased sixfold, while that of the Modenese - initially similar to the national average - grew tenfold. In the per capita income ranking of Italian provinces, Modena, which initially ranked 40th in the early 1950s, rose to first place in 1980 and still occupies the top positions in the ranking. The success is mainly linked to the emergence of small and medium-sized industries with unique products, such as Ferrari or Maserati, or the ceramic, textile and biomedical clusters. From a political point of view, Modena has followed a different path from that of the

country. After the 20-year fascist period, in fact, it has always had left-wing and centre-left majorities in municipal and provincial councils.

4. Data

The analysis is based on a dataset covering surnames mentioned in local newspapers in the province of Modena (Italy) from 1921 to 2011. As digitized archives were not available, data had to be collected manually. This entailed the need to limit the survey to time intervals. We collected surnames from one year per decade (1921, 1931, …, 2011), considering for that year the first 3 days of each month and the entire month of March.

It is significant to collect data from different newspapers active at the local level since media coverage is not independent of the biases of the newspapers themselves and the preferences of the readers (Gentzkow and Shapiro 2010; Larcinese et al 2011). There have been three local newspapers active in the province of Modena in the reference period. Including all three different newspapers reduces this bias.

- *La Gazzetta di Modena*[6], the main daily newspaper in the city and province of Modena, active throughout the time span analyzed. It is a moderate, generalist daily, not politically aligned (except of course during the fascist period).
- *L'Unità (local edition of Modena),* available for the years 1951-1961-1971-1981-1991 of our sample. Historical Italian left-wing daily linked to the Italian Communist Party. Over the years it has gradually embraced more moderate and reformist positions following the evolution of its reference party.
- *Il Resto del Carlino (local edition of Modena)*, used for the years between 1961 and 2011. It is a generalist daily, historically affiliated with the agrarians of Emilia and the sugar industrialists during the first decades of the XX century and usually supporting centre-right parties after the second world war.

In order to distinguish the mentions according to the different spheres of influence of individuals and the type of article, surnames have been classified in pre-established sections: Advertising, Business, Charity, Culture & Public Events, Crime and legal issues, Obituaries, Local Politics, Religion, Science & Technology. Surnames appearing in articles which refer to non-local (for example national politics) and sports events are excluded. Table 1 shows the number of mentions per year and per section. Since many names in the advertising section are present on a daily basis, we have given to each of them a weight of 0.5. Section 6 shows that this choice does not influence the results. All other names have the same unitary weight. The individual sections are also grouped into three macro-categories*: Economy, Local Politics and Social Relevance*. *Economy* comprises Advertising, Business, Charity, Culture & Public Events, and Obituaries. *Local Politics* consists of Local Politics

[6] The newspaper has gone through several name changes: La Gazzetta di Modena (1947-1953); Gazzetta dell'Emilia (1953-1967); Gazzetta di Modena (1968-1977); Nuova Gazzetta di Modena (1981-present).

alone. *Social relevance* finally groups Charity, Culture & Public Events, Obituaries and Science & Technology.

The analysis is finally supplemented by two additional datasets. The first concerns the number of each surname in the population of the city of Modena. We have the surnames of the heads of households in 1936, and of all resident citizens in 1981 and 2001 from the 12th and 14th Italian population census, respectively. The second dataset contains the surnames of professors who taught at the University of Modena in the years 1931, 1941, 1951, 1961, 1971, 1981.

Table 1. Collected surnames by year and section.

Section	1921	1931	1941	1951	1961	1971	1981	1991	2001	2011	Total
Advertising	355	662	363	754	755	311	184	324	325	77	4,110
Business	205	373	181	284	197	403	277	63	394	177	2,554
Charity	359	998	552	434	361	212	73	0	0	0	2,989
Culture & Public Events	212	211	548	665	446	964	762	854	2,334	2,271	9,267
Local News	529	294	413	902	101	402	262	151	268	230	3,552
Obituaries	55	43	384	245	82	282	152	135	287	212	1,877
Local Politics	531	467	476	578	671	790	804	1,228	1,447	1,605	8,597
Religion	0	32	20	58	49	40	63	79	87	72	500
Science & Technology	74	366	111	148	404	313	258	224	395	347	2,640
Total	**2,320**	**3,446**	**3,048**	**4,068**	**3,066**	**3,717**	**2,835**	**3,058**	**5,537**	**4,991**	**36,086**

Before starting the analysis, two crucial checks must be conducted. The first concerns whether the frequency of surnames in newspapers has an independent information content. It is indeed obvious that the most common surnames in the whole population also tend to be very often present in newspapers, but what is important is that the frequency distribution of surnames in newspapers is not a random extraction from the population. To check this point, we can compute for each surname its relative representation (Clark 2014), given by the ratio between the share of the surname in the newspapers and its share in the population in a given year. We know the frequency distribution of total surnames from the registry data for 1936. Since the population composition changes slowly over time, we can use these data to compute the relative representations of surnames in 1931 and 1941 and check whether these values are correlated with the respective relative frequencies in the population. If they are highly correlated, then the names in the newspapers are a simple random extraction from the population and do not provide autonomous information. For the names that are present in the newspapers in 1931, the correlation coefficient between their relative representation and their relative frequency in the population is exactly 0. This same correlation in 1941 is -0.22. For the 100 most frequent names in the registry data of 1936, the correlation is -0.03 for year 1931 and 0.07 for 1941. Many of the surnames that are among the most present in the population have relative representation well below 1. The 100 most frequent surnames in the 1936

registry data have on average a relative representation of 0.67, quite close to the overall average (0.69).

The second check asks whether there is a basis for the assumption that the surnames appearing in local newspapers have a special social relevance. Not all the surnames in this source of course belong to the local elite, but for at least a subgroup of them this may be a concrete possibility. In this sense, observing what happens to the distribution of newspaper surnames over time can give a picture of the changes in the structure of the elite groups. If between two points in time there are low changes in the distribution of surnames, this evidence can be interpreted as a signal that the holders of influential positions did not change very much, and vice versa.

To check the hypothesis that newspaper surnames contain information on the elite groups, one needs to compare the distribution of surnames in newspapers with that of a dataset containing surnames that are surely members of a socially relevant group. To this end, we have collected the surnames of the professors working at the University of Modena from 1931 to 1981. Comparing the frequency distributions of the surnames in the two datasets is not enough, since a high correlation between the two series can be present also in the case that both dataset (the names in newspaper and in university records) are the result of a random extraction from the underlying total population. An alternative is to contrast the relative representations (RR) computed in each of these two datasets, where the RR is defined again as the ratio, for each surname, between its frequency in the dataset and the frequency of the same surname in the total population (from register data). We expect that the surnames with high RR in newspapers - i.e., more likely than others to appear in the local news - have on average also a higher RR in the dataset of university professors than the surnames with low RR in newspapers. Using the registry data from 1936, Tab. 2 shows that this is indeed the case. In each year, the whole set of newspaper surnames is divided into two groups, respectively with RR below and above the median RR for that year. For each group we compute also the average RR in the university data. It turns out that the latter is much higher for surnames with high RR in the news: the surnames with a high relative probability of being present in the news have also - on average - a high probability of being included in the university data, and vice versa. The effect is higher, as expected, if we consider only the surnames appearing in the newspaper sections that we have called “Science” and “Culture and Society”, while it is much lower keeping from the news only the section concerning episodes of incidents, injuries or thefts. Since we do not have the registry data for all years, to compute the RR indices we have used the 1936 registry for the years 1921-1961, the 1981 registry for the years 1971-1991 and the 2001 registry for the years 2001 and *2011.*

Tab. 2 Average relative representation (RR) in the university dataset of names with low or high RR in newspapers.

	1921	**1931**	**1941**	**1951**	**1961**	**1971**	**1981**	**1991**	**2001**	**2011**
whole sample										
low RR in newspapers	0.54	0.53	0.53	0.51	0.53	0.81	0.75	0.84	0.77	0.79
high RR in newspapers	1.98	1.32	1.29	1.35	2.27	5.01	3.80	3.25	1.78	2.73
only "Science" and "Culture and Society"										
low RR in newspapers	0.47	0.57	0.43	0.47	0.53	0.74	0.79	0.74	0.79	0.80

high RR in newspapers	2.19	2.48	1.91	2.65	2.17	6.78	3.79	2.82	1.73	2.88
only "Accident" and "Thief" and "Victim"										
low RR in newspapers	0.31	0.4	0.32	0.43	0.37	0.53	0.78	1.03	0.90	0.58
high RR in newspapers	0.81	0.99	1.24	1.29	0.64	0.90	0.95	1.85	0.77	0.30

5. Results

The aim of this section is to compute some mobility indices of surnames over time. Our hypothesis is that many surnames appearing in newspapers are not a random extraction from the population but contain elements of social relevance/distinction that may signal membership to middle-high social classes. How their distribution changes over time can therefore provide some hints on the dynamics of the society and of the economy. The more stable is the distribution of surnames over time, the lower is mobility since the same persons or families tend to be represented in local news over time. Conversely, the more the distribution of surnames changes across decades, the greater are social change and mobility.

Using the frequency distribution of surnames every ten years, many indices can be imagined. This variety represents a good opportunity since one can check whether the different measures provide a consistent picture. The measures that we compute can be divided into three sets:

1) Mobility indices
2) Emergence of new names
3) Change in the total share of the most common names

5.1 Mobility indices

An intuitive index of persistence over time is the estimated β_1 coefficient in the regression $freq_{i,t+10} = \beta_0 + \beta_1 freq_{i,t} + u_{i,t+10}$, where the unit of analysis is each surname present either in time t or in time t+10, and $freq_{i,t}$ is the relative frequency of the i-th surname at time t. Table 3 shows these regression estimates, where low values of β_1 signal greater mobility. The first column of the table identifies the 10-year period to which each regression refers. For example, the row starting with 1921-1931 contains the results of the regression $freq_{i,1931} = \beta_0 + \beta_1 freq_{i,1921} + u_{i,1931}$. The first column of results provides the estimates computed on the whole sample. This is the more general measure of this kind of persistence of surnames over time that one can compute, but the sample can be decomposed according to the reason of the presence of each surname in the newspaper. We therefore repeat the regression on the subsamples of surnames recorder under the categories “Economy”, “Politics” and “Social relevance”. Their lower size makes the estimates less precise, but they can bring out interesting differences with respect to the more general pattern shown by the regression conducted on the whole dataset.

In the full sample all coefficients are very significantly different from zero (the t-test is always greater than 14). The decades showing more mobility are the 1920s, 1940s, 1960s and 1980s. In each column, we have highlighted with a star the lower numbers (greater mobility). Focusing on the

subsections of the sample, in the "Economy" field a greater mobility is recorded in the 1920s, 1940s and 1980s. Some of these decades show high mobility also in the "Politics" and "Social relevance" categories. The 1920 and 1940s are characterized by a radical change in political regime, first with the rise of fascism to the power in 1922, then with its fall in 1943. Indeed, for the Politics section these two decades record the lowest coefficients while the intermediate period 1931-41 produces the highest one. Also the Social relevance coefficient reaches its lowest value in 1921-31. The shift from democracy to fascism produced therefore a very strong reversal in social and political hierarchies. But during the 1920s the economy was changing too, as the very low coefficient in the "Economy" column shows.

In the period following the end of the second world war, the decades with the major economic and social changes turn out to be the 1960s and the 1980s, while the last twenty years of the sample suggest a reduction in mobility. In particular, the 2001-2011 period has the highest estimated regression coefficients for the whole sample and for two of the three subsamples examined. The 1980s emerge as a period of significant change, not only in the economic dimension.

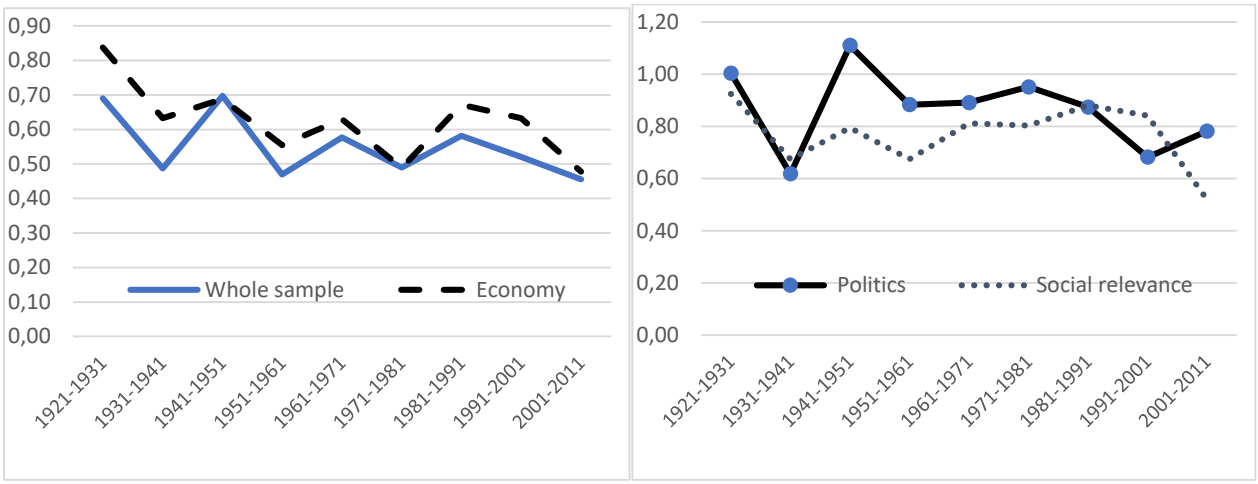
The bottom section of Tab. 3 contains the estimates for the same regression already defined, but covering a longer time span, corresponding to a generation. While in politics there is substantial mobility, reasonably due to the fact that 30 years can be considered the duration of a normal active political career, mobility seems overall decreasing over generations in the other dimensions and in the whole sample. Fig. 1 contains 1 – the estimated β_1, so that the higher points signal greater change. For the whole sample and its three subsections, higher mobility levels are found in the 1920s, 1940s and 1980s, while the two more recent decades appear as periods of declining changes.

Tab. 3 Estimates of β_1 coefficients in the regression of $freq_{i,t+10}$ on $freq_{i,t}$

Time interval	Whole sample	Economy	Politics	Social relevance
1921-1931	0.31*	0.16*	0.00*	0.08*
1931-1941	0.51	0.37	0.38	0.33
1941-1951	0.30*	0.31*	-0.11*	0.21
1951-1961	0.53	0.45	0.12	0.33
1961-1971	0.42*	0.37	0.11*	0.19*
1971-1981	0.51	0.51	0.05*	0.20
1981-1991	0.42*	0.33*	0.13*	0.12*
1991-2001	0.48	0.37	0.32	0.16*
2001-2011	0.55	0.52	0.22	0.48
1921-1951	0.238	0.142	-0.128	0.079
1951-1981	0.28	0.209	-0.128	0.182
1981-2011	0.351	0.333	-0.108	0.207

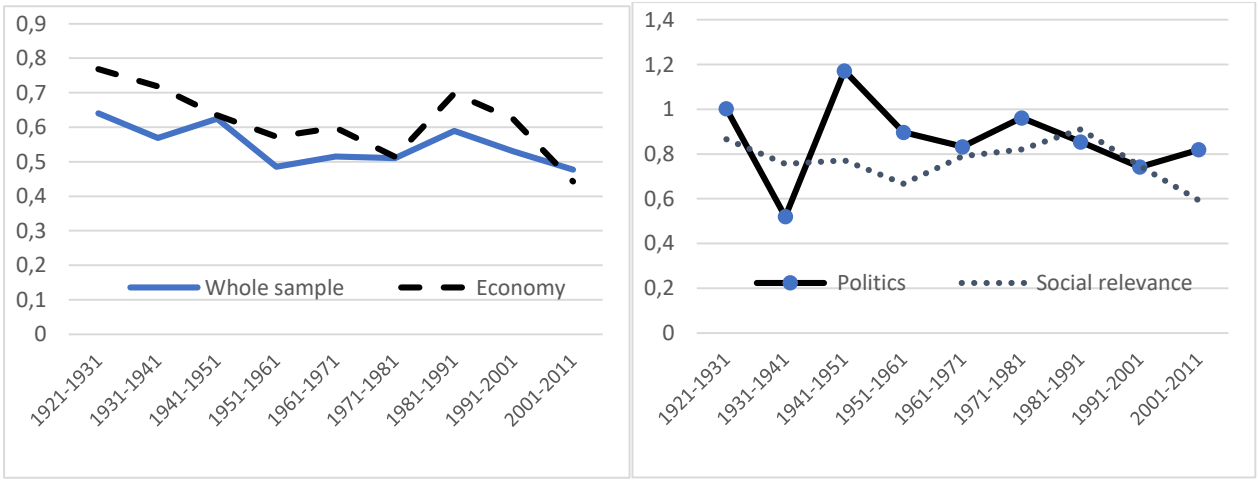
Note: each regression is run on the surnames present in at least one of the two decades

Fig. 1 Ten-year mobility measured by $1 - \beta_1$



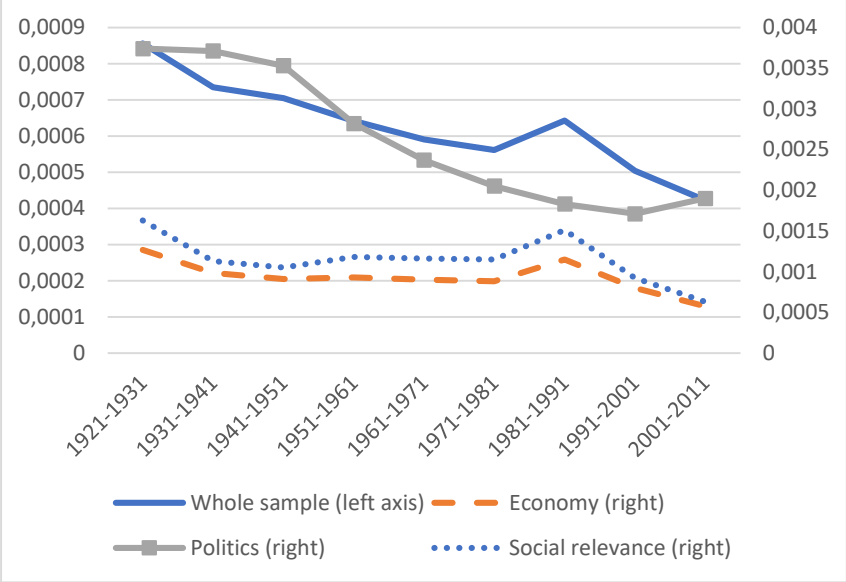
Another very intuitive indicator of the degree of persistence over time of the same surnames is the correlation coefficient of the relative frequencies of surnames present in time t and in time t+10. A high correlation signals low mobility. Fig. 2 provides the difference between 1 and this correlation coefficient. Results are very similar to those obtained with the Ols regression. For the entire sample, the decades with greater mobility are the first section of the whole period and the 1980s. The same pattern characterizes also the various subsections, with some peculiarities. It seems that the "economic miracle" following the end of the second world war was able to produce a surge in new names only in the 1980s. In local politics, the democratic period is characterized, as expected, by lower mobility than the turbulent first three decades under consideration.

Fig. 2 Correlation coefficient between $freq_{i,t}$ *and* $freq_{i,t+10}$



A very simple index used in the field of intergenerational income mobility is the average absolute difference $\frac{\sum|Y_i - X_i|}{N}$, where Y_i is the income or socio-economic position of a person and X_i is the same variable for her parent (Fields and Ok 1996). In the present context, this index can be expressed as $\frac{\sum|freq_{i,t} - freq_{i,t-10}|}{N}$, where $freq_{i,t}$ is the relative frequency of the i-th surname at year t. Since many surnames are present only in time t but not in time t-10 and vice versa, N is the size of the larger of the two samples in years t and t-10. Fig. 3 shows this measure computed on the whole sample and on its subsections. For the whole data set, from this index total mobility seems to be falling over time, with a tip in the 1980s that is evident also in the Economy and Social relevance data. The Politics section shows a clear declining trend as well.

Fig. 3 Average absolute difference between relative frequencies of surnames

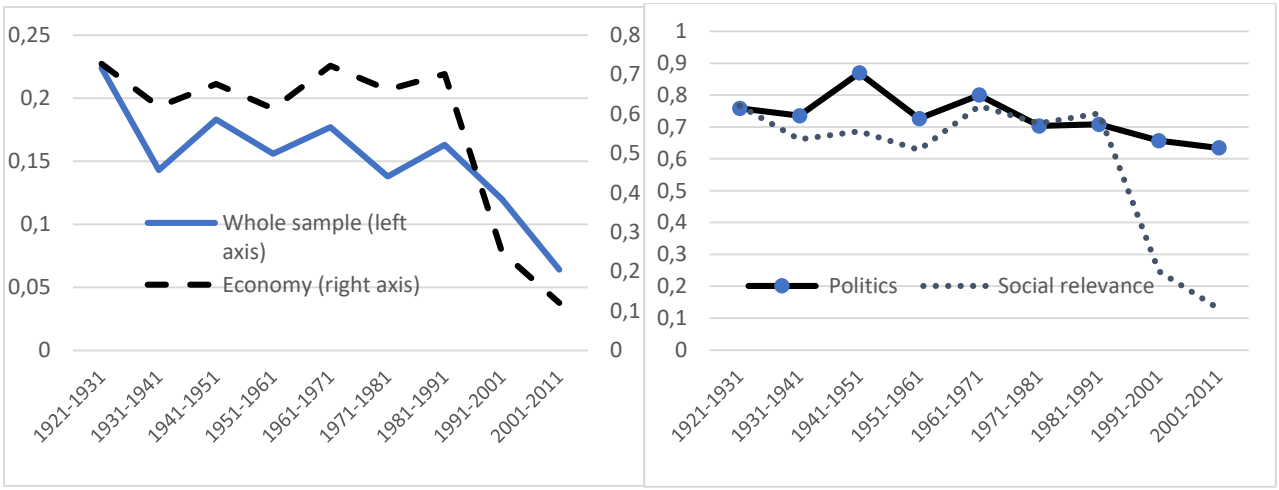


5.2 Emergence of new names

The relevance of socio-economic movements can be inferred also by the emergence of new names in the newspapers. Setting a given relative frequency as a threshold, we can ask how many names are present with at least this relative frequency, that were not present at all in the newspapers a decade before. The result may of course change with the choice of a particular value for the relative frequency threshold. We choose here 0.05%. Fig. 4 shows the share of the surnames that in year t have a relative frequency of at least 0.05%, and where absent from the sample a decade earlier. The results are lower for the whole sample, since it is of greater dimension and therefore it is less likely that a name was totally absent in the previous dataset. In this entire sample mobility appears

higher at the beginning of the century, in the 1940s, 1960s and 1980s, falling sharply at the end. In the Economy section mobility is always high, except in the last 20 years. The same is true for Social relevance, while Politics has a slightly declining pattern.

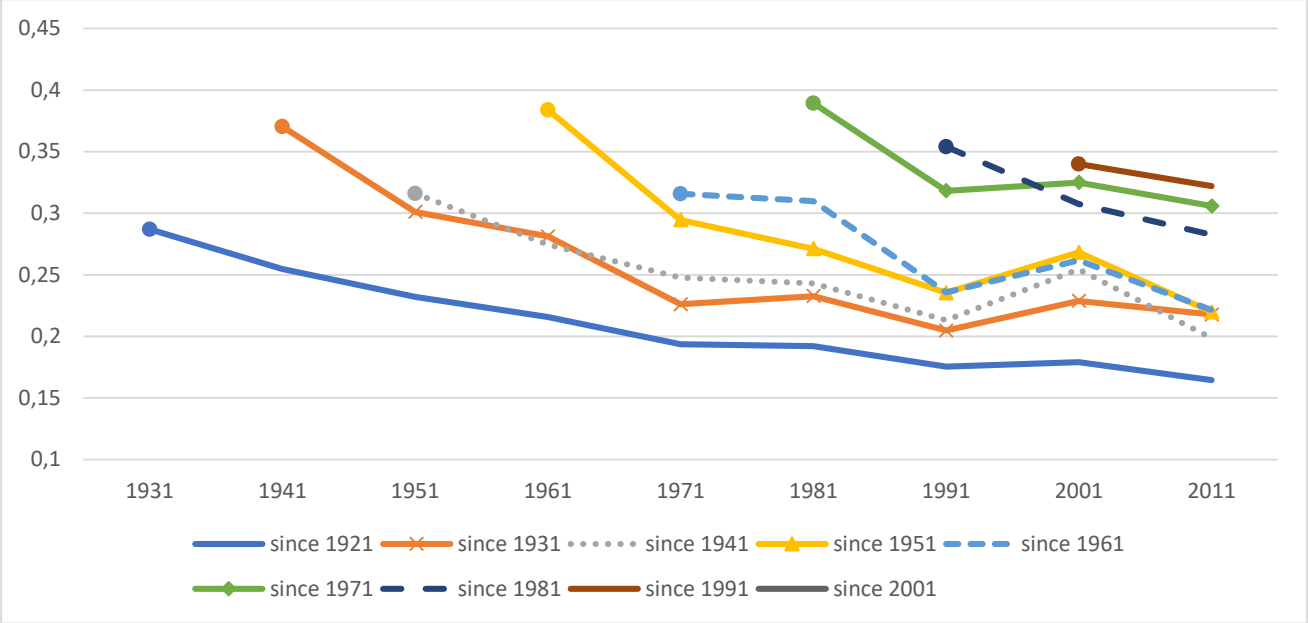
Fig. 4 Share of surnames that in time t have a relative frequency of at least 0.05%, that were absent from the t-10 sample



5.3 Change in the total share of the most common names

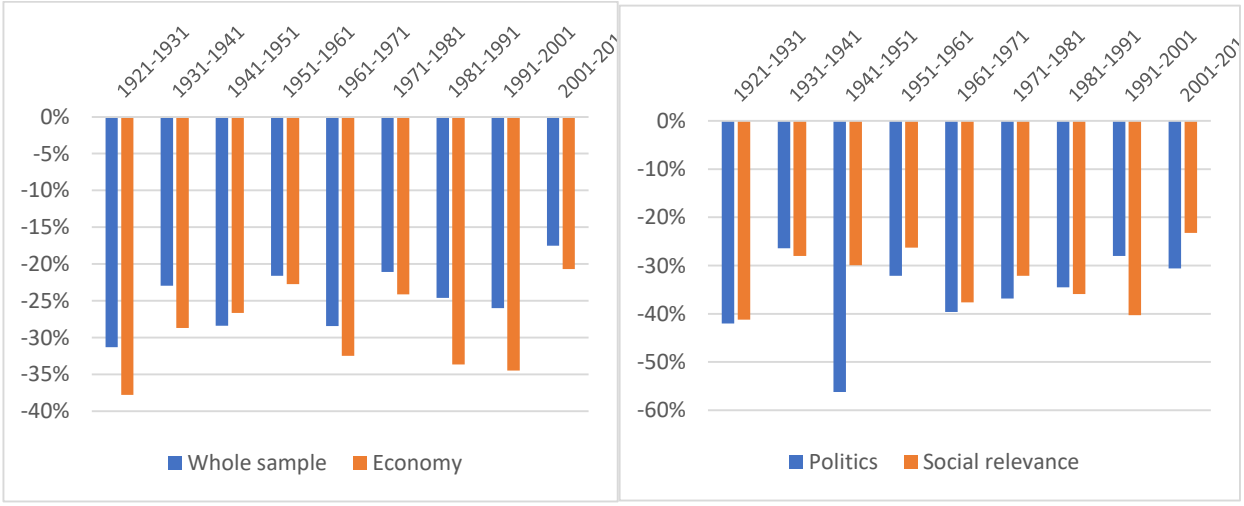
Without imposing arbitrary thresholds, an alternative method to measure how much the names present in the newspapers change over time consists in identifying the most common names in a moment in time, and check how their share falls over time. For example, we select at time t the names that together represent 60% of the total number of names in the newspapers; after ten years it is very likely that this group of names will represent a share that is lower than 60%, and so on for the following decades. The more rapid is the fall of this share, the higher is social and economic mobility represented in newspapers. In Fig. 5, for example, the names that in 1921 represented 60% of total frequency had, ten years later, a cumulative frequency of 29%. The dotted points at the beginning of each curve are the total share of these names ten years after they represented together 60% of the data. The lower these points are, the greater is mobility in that decade. It is also interesting to note that the names representing in 1921 60% of the sample had a share, 90 years later, of only 16%. There is a clear common tendency to regress toward lower values over time at a decreasing rate, with local exceptions like in 2001, with a rebound in the importance of some names which were very frequent in faraway decades.

Fig. 5 Share of the names representing in the initial year 60% of the total distribution



Using the same approach, and considering only the movement in one decade, Fig. 6 shows the reduction in the total share of the names. For example, in the whole sample from 1931 to 1941 the total share occupied by the names that in 1931 represented 60% of the data fell by 23 percentage points, i.e. from 60% to 37%. The lower the bar, the higher is mobility. The decades with greater change for the total sample have been the 1920s, 1940, 1960s and the period from 1981 to 2001, while low mobility is recorded in the 1930s, 1950s, 1970s and in the new millennium. The same pattern is provided by the Economy subsection. In Politics, as expected, the biggest changes depend on the story of the fascist regime, but important movements also characterize the period 1960-1990, followed by two decades of lower mobility.

Fig. 6 Reduction, in percentage points, of the total share of the most common names from 60% at the beginning of each decade



6. Some extensions and checks

6.1 What is the effect of the most common names in the population?

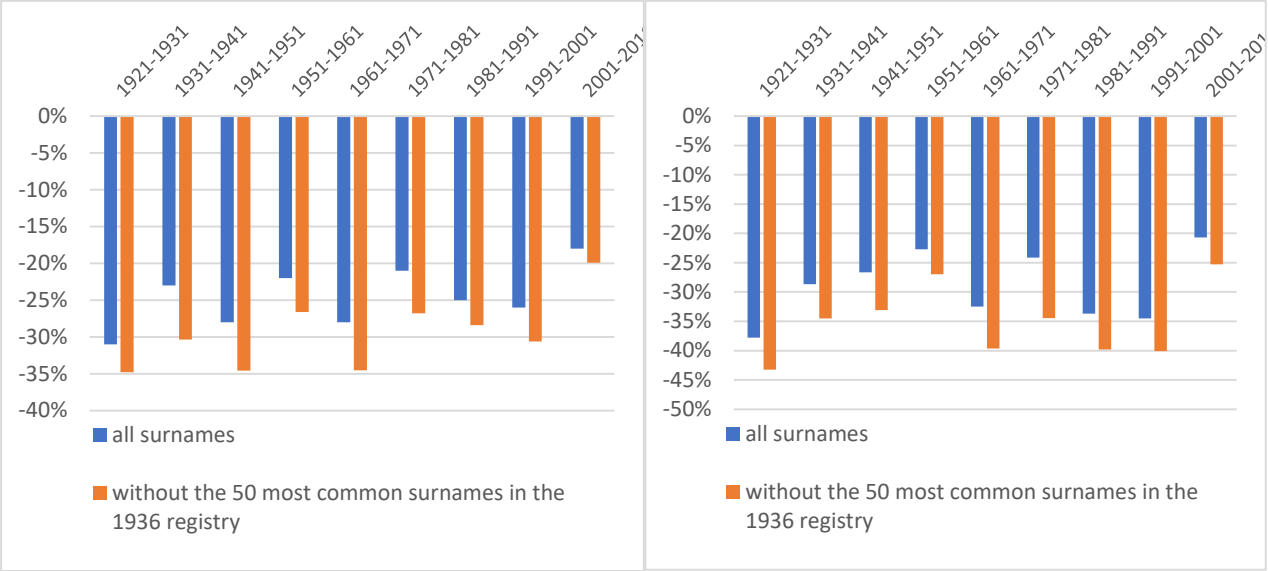
There may be a legitimate suspicion that the results of section 5 could be heavily influenced by the presence in the sample of the surnames that are very common among the resident population, and that inevitably tend to be significantly represented also in the newspapers. We therefore have recalculated the indices shown above after having excluded from the sample the 50 most common names in the registry data of 1936. The choice of the number of excluded names is arbitrary, but a greater number would risk reducing too much the size of the sample. The estimates become less precise, but all results obtained above are confirmed.

For example, the correlation between the estimated β_1 coefficients of the regression of $freq_{i,t+10}$ on $freq_{i,t}$ (see Tab. 3) with and without the most common 50 names in the 1936 registry is 0.84 for the whole sample, 0.61 for the Economy section, 0.99 for Politics and 0.70 for Social relevance. The lower correlation among the results for the Economy section still allows to observe two periods with high mobility: the 1920s and the 1980s, while small mobility is again found in the 1950s and in the first decade of the new century.

Turning to the indicator shown in section 5.3, i.e. the reduction over a decade of the share occupied by the names representing 60% of the total at the beginning of each decade, Fig. 7 shows that, for the whole dataset and the Economy subsample, results are strongly correlated and substantially similar. The correlation coefficient between the two series in the graph for the whole sample is 0.95, and for the Economy section is 0.96. We can therefore conclude that the whole analysis could be carried out on the sample obtained after the exclusion of the more common names in the population.

Fig. 7 Reduction, in percentage points, of the total share of the most common names from 60% at the beginning of each decade

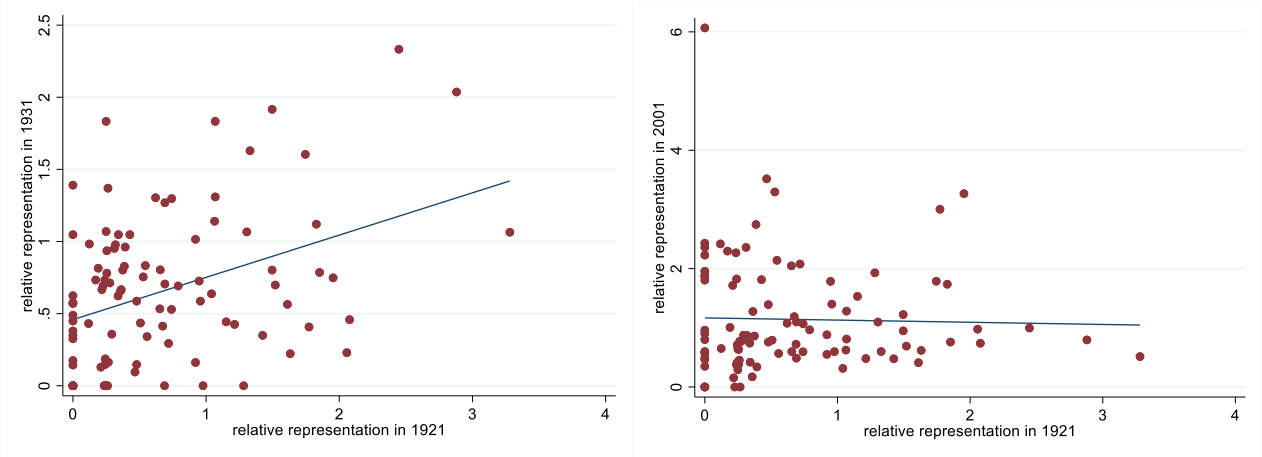
Whole sample Economy



6.2 Does the relative advantage persist over time?

The availability of registry data for the entire population of residents in the town of Modena for the year 1936 allows not only to rule out the suspect, as seen in section 3, that the surnames present in newspapers are a simple random draw from the population, but also to check what happens over time to surnames that, relatively to their presence in the total population, are more or less numerous in the newspapers in a given initial date. For each of the most common surnames in the registry data of 1936, we compute its relative representation, given by the ratio between the share of the surname in the newspaper data and its share in the population. Since we have the distribution of the names in the registry data of 1936, 1981 and 2001, we repeat this exercise for the years 1921, 1931, 1981 and 2001, using the share in 1936 as the denominator in the computation of the index for 1921 and 1931. The question is how the relative representations of the 100 most frequent names in the population at the start of the period changes over time. From Fig. 8, the correlation between the relative representations a decade apart - in 1921 and 1931 - is significant, but disappears after 80 years.

Fig. 8 The relative representations of the 100 most common names in the registry data of 1936 over time



The same pattern can be observed with the other pairings available: Tab. 4 shows the estimated Ols coefficients between the relative representations among different time distances. For short time spans of 10 or 20 years the coefficients are positive and significant, denoting persistence, but in the long run regression towards the mean is dominant. Still there are, among the most frequent names in the population, some names that over these 80 years are always under or never represented in the newspapers, and some names that in all years have high relative representation, but a tendency towards overall mobility is prevalent. It would be interesting to repeat this exercise with other local contexts, to check if this is a more general pattern or it is due to more local reasons. According to Acciari et al. (2022), the provinces of northern Italy are characterized by substantial levels of economic mobility, so a comparison with a local reality in different parts of Italy could be very interesting if one wants to separate the influence of the common historical changes to that of local factors.

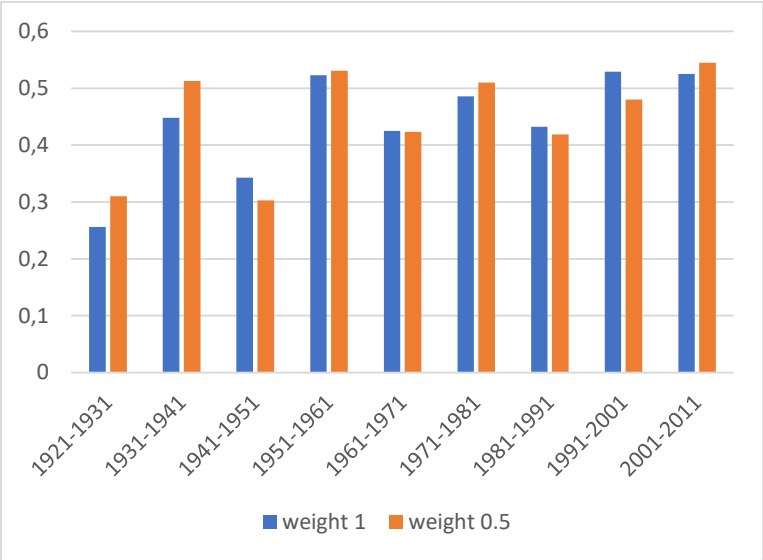
Tab. 4 Ols coefficients of relative representation in year t on relative representation in year t-x

Time interval	β_1	t	R^2
1931 on 1921	0.293	4.2	0.155
1981 on 1921	-0.027	-0.27	0.0007
2001 on 1921	-0.037	-0.27	0.0007
1981 on 1931	0.131	0.96	0.009
2001 on 1931	0.060	0.33	0.001
2001 on 1981	0.309	2.34	0.050

6.3 The effect of advertising

As previously explained, we have attributed to the names that appear in the advertisements a weight of 0.5, since, particularly in the early decades, some names are present in this section on a daily basis. From Fig. 9, it turns out that the effect of this choice is modest.

Fig. 9 β_1 coefficient of the regression of $freq_{i,t+10}$ on $freq_{i,t}$ estimated on the whole sample with different weights for the names in advertisements



7. Conclusions

In this work we propose a new source of information, the surnames appearing in local newspapers, to study phenomena such as the social mobility of elites and family generations. We have also verified, using complementary sources, that the surnames present in newspapers are not a simple random extraction from the population and that they present traits of social relevance. Surnames are thus a great source of information, already widely exploited by economic research, but not from newspapers. If one also keeps track of the context in which each surname is mentioned, it is possible to study not only the evolution of overall social mobility, but also that of specific areas, such as the political or economic one.

The main results confirm both a persistence of the same surnames over time and a tendency for this persistence to diminish over the long term. We have also identified certain periods in which mobility seems to have been more pronounced. First, the break caused by fascism emerges, both in the 1920s and in the period of the regime's dissolution in the early 1940s. In the post-war period, the 'economic miracle' was slow to show its effects on the mobility of surnames, which – after some signs in the 1960s, increased markedly in the 1980s. Periods of strong economic growth might in fact be positively correlated with absolute mobility, whereby entire social classes see their living standards change, but less so with relative mobility, whereby new people emerge, and others fall down the social hierarchy. Further, a period of economic boom, such as the 30 years after the end of the Second World War, might lead to a strong growth in the average income of low-skilled workers, but this might not emerge in the newspapers of the same years, until some of them, or their descendants, manage to make a substantial leap and are socially recognized as successful entrepreneurs. Add to this the fact that Modena reached its period of greatest prosperity in the

country - in comparison with other local areas - precisely during the 1980s, when social mobility according to our estimates seems to have been very high.

We believe that data of this nature can be a good source for the analysis of socio-economic mobility. Further applications may include the comparison between the newspapers of distant local realities, to see whether the differences in mobility rates obtained from other sources are confirmed by this kind of data. It would also be interesting to compare closer local realities, for example in the same region, because they should not be assumed to have similar mobility stories, either in the whole society or in specific subsections.

References

Acciari P., Polo A., Violante G. (2022), And yet it moves: integenerational mobility in Italy. American Economic Journal: Applied Economics, vol. 14, n. 3, pp.118-63.

Ager, P., Eriksson, K., Karger, E., Nencka, P., & Thomasson, M. A. (2022), School closures during the 1918 flu pandemic. Review of Economics and Statistics, 1-28.

Ahmed A. M., Hammarstedt M. (2008), Discrimination in the rental housing market: a field experiment on the Internet. Journal of Urban Economics, vol. 64, n. 2, pp. 362.372.

Åslund, O., & Skans, O. N. (2012). Do anonymous job application procedures level the playing field?. ILR Review, 65(1), 82-107.

Ban, P., Fouirnaies, A., Hall, A. B., & Snyder, J. M. (2019). How newspapers reveal political power. Political Science Research and Methods, 7(4), 661-678.

Barone, G., & Mocetti, S. (2021). Intergenerational mobility in the very long run: Florence 1427–2011. The Review of Economic Studies, 88(4), 1863-1891.

Beach, B., & Hanlon, W. W. (2022). Historical newspaper data: A researcher's guide and toolkit (No. w30135). National Bureau of Economic Research.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American economic review, 94(4), 991-1013.

Black, S. E., & Devereux, P. J. (2010). Recent developments in intergenerational mobility.

Buckman, S. R., Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). News sentiment in the time of COVID-19. FRBSF Economic Letter, 8(1), 5-10.

Clark, G. (2014). The son also rises. Princeton University Press.

Collado, M. D., Ortuño-Ortín, I., & Romeu, A. (2012). Long-run intergenerational social mobility and the distribution of surnames. manuscript, Universidad de Alicante.

Bortoli, C., Combes, S., & Renault, T. (2018). Nowcasting GDP growth by reading newspapers. Economie et Statistique, 505(1), 17-33.

Earl, J., Martin, A., McCarthy, J. D., & Soule, S. A. (2004). The use of newspaper data in the study of collective action. Annu. Rev. Sociol., 30, 65-80.

Felice E. (2016), Ascesa e declino. Storia economica d'Italia, Il Mulino editore

Fowler, B., & Bielsa, E. (2007). The lives we choose to remember: A quantitative analysis of newspaper obituaries. The Sociological Review, 55(2), 203-226.

Ganzeboom, H. B., Treiman, D. J., & Ultee, W. C. (1991). Comparative intergenerational stratification research: Three generations and beyond. Annual Review of sociology, 17(1), 277-302.

Gentzkow, M., Glaeser, E. L., & Goldin, C. (2006). The rise of the fourth estate. How newspapers became informative and why it mattered. In Corruption and reform: Lessons from America's economic history (pp. 187-230). University of Chicago Press.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. Econometrica, 78(1), 35-71.

Güell, M., Rodríguez Mora, J. V., & Telmer, C. I. (2015). The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. The Review of Economic Studies, 82(2), 693-735.

Hansen, P. (2004), “Writing business history without an archive. newspapers as sources for business history – possibilities and limitations”, in Gadd, C.-J., Granér, S. and Jonsson, S. (Eds), Markets and Embeddedness. Essays in Honour of Ulf Olsson, pp. 99-120.

Hao, Y. (2021). Social mobility in China, 1645–2012: A surname study. China Economic Quarterly International, 1(3), 233-243.

Larcinese, V., Puglisi, R., & Snyder Jr, J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers. Journal of public Economics, 95(9-10), 1178-1189.

Markel, H., Lipman, H. B., Navarro, J. A., Sloan, A., Michalsen, J. R., Stern, A. M., & Cetron, M. S. (2007). Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. Jama, 298(6), 644-654.

Matsuda, M., (1996), The Memory of the Modern, Oxford, Oxford University Press.

Mooi-Reci, I. (2020). Intergenerational Mobility Research: Current Challenges and Future Directions. Australian Economic Review, 53(2), 241-246.

Pascual, A., Guéguen, N., Vallée, B., Lourel, M., & Cosnefroy, O. (2015). First name popularity as predictor of employability. Names, 63(1), 30-36.

Silberzahn, R., & Uhlmann, E. L. (2013). It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers than as employees. Psychological Science, 24(12), 2437-2444.

Solon, G. (1999). Intergenerational mobility in the labor market. In Handbook of labor economics (Vol. 3, pp. 1761-1800). Elsevier.

Solon, G. (2018). What do we know so far about multigenerational mobility?. The Economic Journal, 128(612), F340-F352.

Tosh, J. (2010), The Pursuit of History: aims, Methods, and New Directions in the Study of Modern History, 5th ed., Pearson Longman, Harlow.

Zamagni, V. (1993). The economic history of Italy 1860-1990. Clarendon Press.