Correlating Social Mobility and Economic Outcomes*

Maia Güell[†] Michele Pellizzari[‡] Giovanni Pica[§]
José V. Rodríguez Mora[¶]

May 11, 2016
First version: February 2015
This version: March 2016

Abstract

We apply a novel measure of intergenerational mobility (IM) developed by Güell, Rodríguez Mora, and Telmer (2014) to a rich combination of Italian data allowing us to produce comparable measures of IM of income for 103 Italian provinces. We then exploit the large heterogeneity across Italian provinces in terms of economic and social outcomes to explore how IM correlates with a variety of outcomes. We find that (*i*) higher IM is positively associated with a variety of "good" economic outcomes, such as higher value added per capita, higher employment, lower unemployment, higher schooling and higher openness and (*ii*) that also within Italy the "the Great Gatsby Curve" exists: in provinces in which mobility is lower cross-sectional income inequality is larger. We finally explore the correlation between IM and several socio-political outcomes, such as crime and life expectancy, but we do not find any clear systematic relationship on this respect.

Key words: Surnames, intergenerational mobility, cross-sectional data analysis.
JEL codes: C31, E24, R10

*We thank Cristina Blanco, Nicola Solinas, Alessia de Stefani and Robert Zymeck for superb research assistance. We benefited from the very useful comments of seminar participants at Stanford University, Universidad de Alicante, EUI, Tinbergen Institute, IZA and EALE meetings. We are also very grateful to Daniele Checchi, Carlo Fiorio and Marco Leonardi for sharing their program files to generate the traditional measures of mobility using the Bank of Italy SHIW data. Financial support from the Spanish Ministry of Education and Science under grants ECO2011-28965 (MG) and ECO2011-25272 (JVRM) is gratefully acknowledged. Michele Pellizzari also gratefully acknowledges the financial support of NCCR-LIVES.

[†]University of Edinburgh, CEPR, FEDEA and IZA. Email: maia.guell@gmail.com
[‡]University of Geneva, CEPR, fRDB and IZA. Email: michele.pellizzari@unige.ch
[§]Università degli Studi di Milano, LdA, CSEF, Baffi and fRDB. Email: giovanni.pica@unimi.it
[¶]University of Edinburgh and CEPR. Email: sevimora@gmail.com

1 Introduction

The vast literature analysing intergenerational mobility (IM) (see Solon (1992), Haider and Solon (2006), Hertz (2007), Lee and Solon (2009) and the extensive literature surveys available in Solon (1999) and Black and Devereux (2011)) is often pervaded by a belief that more mobility is a desirable feature of the economy or society at large. However, due to the well-known difficulties in producing reliable measures of IM (Solon, 2002) we still know very little about how it correlates with meaningful economic and social outcomes.

In this paper we adopt a novel methodology and a very rich set of data to compute measures of IM that are highly comparable across very diverse geographical areas. We can then explore the correlation between IM and a variety of interesting social and economic outcomes, such as income per capita, employment, crime, life expectancy and many more. To our knowledge our analysis provides the most reliable empirical support so far to the common claim that more IM is a desirable feature of an economic system and complements the evidence of Chetty, Hendren, Kline, and Saez (2014) in which they compare mobility across U.S. regions – probably the closest paper to ours.[1]

So far the most popular way to measure IM is by means of a regression with some meaningful outcome of sons as a dependent variable and the same outcome for the parents (or one of the parents) as explanatory variable (possibly with additional controls). The coefficient on the outcome of the parents, the so-called intergenerational elasticity, is considered the best summary measure of IM. However, estimating the above regression is extremely problematic (Solon, 1992). For once, it requires data linking parents' and children's outcomes, typically very long panel datasets but such data are only available for a limited number of countries and, when they exist they are usually not very easily comparable across countries. Moreover, the point in the life cycle at which outcomes are observed matters a lot for the estimates of the intergenerational elasticity and it is very difficult to choose the appropriate timing, especially because the exact shape of the life cycle profiles may change across generations. For example, young people now enter the labour market much later than their parents, often with lower

[1] Among others, Björklund and Jäntti (1997), Couch and Dunn (1997), Checchi, Ichino, and Rustichini (1999), Björklund, Eriksson, Jäntti, Raaum, and Österbacka (2002), Comi (2003) and Grawe (2004) also compare mobility patterns across countries.

earnings and experience steeper growth later on.

As a result of these measurement problems, we are still uncertain on, for example, whether IM is higher in rich or in poor countries. On how IM correlates with growth or the degree of state intervention in the economy, particularly with the degree of investment in key areas such as education. We do not know how it correlates with crime or corruption. Thus, even after the enormous improvements of the last decades we know very little about the economic meaning of intergenerational mobility.

Güell, Rodríguez Mora, and Telmer (2014) recently propose a new method to measure mobility that overcomes most of these difficulties and that does not require panel data. The minimal data required for this methodology are cross-sections of individual records with only two variables, an interesting outcome, such as income, education or occupation, and the surname of the individual. In fact, it is not even necessary to know the full surname of the person, which might be difficult to obtain for confidentiality reasons, and it is sufficient to anonymously recode it in a way that allows identifying individuals who share the same surname. The data are then used to construct an indicator of the capacity of family names to capture the variance of the outcomes. Güell, Rodríguez Mora, and Telmer (2014) baptise this indicator *Informational Content of Surnames* (ICS) and they use it to study changes in mobility over time in Catalonia, Spain. An important assumption of the empirical exercise carried out in Güell, Rodríguez Mora, and Telmer (2014) is that all factors affecting the distribution of surnames change very slowly over time so that changes in the ICS can be attribute exclusively – or at least primarily – to changes in intergenerational mobility.

The obvious next step in this line of research is the comparison of IM across countries using the ICS, but the method requires surname distributions to be comparable across countries. Such an assumption is problematic and in this paper we take a first step in this direction by comparing different provinces within a country. Italy is an ideal setting for this type of exercise because, while the surname conventions have been the same over its entire territory for a long period of time, there is enormous variation in almost all relevant economic and social outcomes across different parts of the country. In other words, Italian provinces, which are the units of observations in our analysis, all share very similar distributions of surnames and at the same time they perform very differently in a wide range of economic and social indicators. According to the official (PPP adjusted) Eurostat data in 2004 (the year of reference of our data), per capita income in the poorest Italian region was comparable to that of Hungary while the richest was second only to Luxembourg, the richest country of the European Union.

The main empirical analysis in this paper is based on the universe of all tax declarations submitted in Italy in 2005 (referring to incomes earned in 2004). This dataset contains the full names and surnames of the persons submitting the declaration, their taxable income and their province of residence. We use these data to compute the ICS for each single Italian province and we then correlate the resulting indicators with a variety of measures of economic and social development that we obtain from official sources, mostly the Italian national statistical institute.

Our results show that higher mobility correlates positively with "good" economic outcomes, such as value added per capita, income, wealth, employment rates and participation rates; instead "bad" economic outcomes, such as unemployment rates of different socio-economic groups and the shares of low educated young individuals, are related to lower mobility. Also social capital proxies are positively related to social mobility. Patterns are less clear-cut for other socially relevant outcomes (suicide rates, life expectancy, intensity and quality of public sector activity and crime rates). Interestingly, we find that also within Italy the "the Great Gatsby Curve" exists: in provinces in which social mobility is lower cross-sectional income inequality, measured as the 90/10 percentile ratio, is larger.

Notice the focus of our investigation. We do not look after causal relationships. We report the co-movements between intergenerational mobility and a large array of interesting economic and social outcomes. Nevertheless, we can make one claim related to causality: institutional differences are **not** the cause of the observed differences in mobility across Italian provinces, and of the correlations with economic outcomes.

The reason is that, within Italy, institutional differences across provinces are small, while the observed differences in social and economic outcomes are large. Thus, in this paper we do not measure the impact of policies and of the institutional set-up on mobility or on other outcomes. The differences across provinces and the correlation between social mobility and the level of economic activity are equilibrium outcomes, not the result of differences in policies. Interestingly, we do detect wide ranging differences in mobility across provinces, with clear patterns in the correlations with meaningful economic outcomes even under an encompassing political and institutional setting.

The paper is organised as follows. Section 2 describes the methodology based on the informational content of surnames used to measure intergenerational mobility across Italian provinces. Section 3 provides information on the rules governing the transmission of surnames in Italy. Section 4 describes the data used; section 5 and 6 discuss the results of the analysis. Section 7 concludes.

2 Measuring Mobility

In this paper, we use the measure of intergenerational mobility proposed by Güell, Rodríguez Mora, and Telmer (2014), the *Informational Content of Surnames* (ICS). The ICS is a moment of the joint distribution of surnames and economic outcomes. Unlike traditional measures of mobility, this measure does not require panel data nor any explicit links between children and their parents. One cross-sectional data of surnames and economic outcomes is enough.

The basic idea is simple. Surnames are intrinsically irrelevant for the determination of economic well-being, but they get passed from one generation to the next, alongside other characteristics that *do* matter. The more important are these characteristics in determining outcomes, the more "inheritance" matters for economic outcomes, and, therefore, the more information surnames contain on the values of outcomes. Thus, surnames can be used to measure the importance of inheritance and identify the degree of social mobility: the more surnames matter the lower the degree of social mobility.

The reason why this approach works is that surnames establish a partition of the population which is informative about family links. Family members inherit genes and cultural background from their ancestry. Insofar as ancestry does determine economic outcomes (i.e. mobility is low), the expected variance of income of family members is bound to be smaller than the variance of income in the population at large.

Saying that surnames contain information is the same as saying that the variance of income conditional on sharing a surname is smaller than the unconditional variance of income. Given a certain mapping between the surname partition and family linkages, the more prevalent inheritance is, the larger the difference between the expected variance of income of family members and the unconditional variance of income (for any definition of "family"), and consequently, the more information surnames contain.

Thus, the key of the method is that surnames are informative about family linkages. They do so because surname distributions are very skewed. If there were only a few surnames, the mapping between the surname partition and family relationship would be extremely blurred, and conditioning on surnames would not change the variance for any degree of inheritance. Fortunately, the western surname convention insures that the surname distribution is bound to be very skewed. Despite a few surnames being very abundant and their members being very unlikely to have common ancestors, there are very many uncommon surnames whose members are likely to have close family relationship. In those infrequent surnames lies the power of the methodology.

Thus, before comparing the ICS measures across Italian provinces, we first need to make sure that the mapping from the surname partitions to family links is very similar across provinces. We do so by making sure that the distributions of surnames (for the population under scrutiny) is very similar across provinces. We can then apply this method – further described below – to Italy and its provinces and interpret any differences in the ICS across provinces as driven by differences in mobility.

2.1 The Informational Content of Surnames

The *Informational Content of Surnames* is a measure of how much surnames inform about economic outcomes of individuals, after controlling for other factors. The definition of the ICS is as follows.

Consider a cross-section in which each individual is associated with a surname s, a measure of their economic well-being y_{is}, and a vector of additional demographic characteristics X_{is}, such as age and gender. Güell, Rodríguez Mora, and Telmer (2014) define the ICS as the difference between the R^2 of two sets of regressions. A first regression, whose R^2 is denoted as R_L^2, estimates economic well-being for the *average* individual with surname s as follows:

$$y_{is} = \gamma' X_{is} + b' D + \text{residual} \quad , \tag{1}$$

where D is an S-vector of surname-dummy variables with $D_s = 1$ if individual i has surname s and $D_s = 0$ otherwise.

Since the number of surnames is very large and they may happen to explain the variance of y_{is} even if they do not carry any information on family linkages, a second set of regressions is performed to insure that we do not spuriously attribute informativeness to surnames. In each of the regressions we include a different S-vector of 'fake' dummy variables F that randomly re-assign surnames to individuals in a manner that maintains the marginal distribution of surnames but destroys the informativeness of surnames about familial linkages. The regression is:

$$y_{is} \quad = \quad \gamma' X_{is} + b' F + \text{residual} \quad , \tag{2}$$

The R^2 from this regression is denoted as R_F^2. We replicate the regression in (2) M times and calculate the average over the M R^2 obtained. Denoting such an average as $\overline{R}_F^2$, the ICS is defined as

$$\text{ICS} \equiv R_L^2 - \overline{R}_F^2 \quad . \tag{3}$$

The ICS measure has a number of important advantages. It has value zero if there is one surname per person or if there is only one surname for everyone. More generally, it captures the information that surnames contain due to family linkages and measures how much of the variance of the dependent variable is explained by the variance of the surnames.

An additional advantage is that it is comparable with the traditional measure based on father-sons regressions as Güell, Rodríguez Mora, and Telmer (2014) provide a model that maps the ICS into the traditional measure and show that the former is monotonically increasing in the latter.

2.2 Comparability across regions

Our goal in this paper is to measure the ICS for every province in Italy to obtain comparable measures of mobility and correlate them with a battery of macro-economic outcomes. Key for our exercise is that the distributions of surnames across provinces is comparable so that any differences in provincial ICS reflect differences in mobility and not something else.

In order to address this potential issue we will provide measures of the ICS based on infrequent surnames in all regions. The tail of the surname distribution that contains infrequent surnames identifies family linkages with less noise and is therefore more comparable across provinces. In particular, we will concentrate on individuals whose surname contains less than 15, 20, 25 or 30 people. The idea behind this is that these sub-populations have the same degree of “family connectivity” in different provinces.

2.3 Migration

Migration to an Italian province from other countries as well as from other Italian provinces is another potential challenge for our exercise. Migrants may have both very different surnames and very different economic outcomes as compared to natives in the recipient region (at least initially), making their surnames very informative. However, we want our measure to reflect social mobility and not differential migration patterns across provinces. This ethnicity issue is also discussed in Güell, Rodríguez Mora, and Telmer (2014). As a solution, they propose to construct an index of how local a surname s in province r is, as follows:

$$LocalDegree(s,r) = \frac{\text{Number of people with surname } s \text{ in province } r}{\text{Number of people with surname } s \text{ in Italy}} \tag{4}$$

To the extent that migrants have very different surnames from natives, they will have a very low value of the index in the recipient province. Therefore, to clean our mobility measure from migration effects we also calculate the ICS measure for the fraction of individuals in the top 50 percent of the distribution of the $LocalDegree(s,r)$ Index in every province.

Finally, as an additional robustness check, to further improve the comparability of the surname distributions across provinces, we also calculate the ICS for the top 50 percent of the distribution of the $LocalDegree(s, r)$ Index in every province within the tails of the surname distribution, as explained in section 2.2.

3 Italian Surnames

In Italy, surnames follow the standard western naming convention. Most people inherit their surname from their father. At the same time there can be some surname innovations as it is possible, although not easy, to change surname. The procedure to do so is quite complex and it can take up to 1 year. As discussed in Güell, Rodríguez Mora, and Telmer (2014), this will imply that the surname distribution in Italy is necessarily very skewed, implying that surnames are extremely informative about family links.

Unlike most other countries, in Italy women do not change their official surnames upon marriage. While in everyday life it may happen that married women use the husbands' surname, the law requires everyone to use their inherited surnames in all official documents regardless of marriage status. Indeed, in Italy the government identifies tax payers through a unique fiscal code, which is given to each person at birth and does not change with marriage. It is a code which depends on the name, the surname at birth, date and place of birth. So, the state identifies taxpayers through the surname at birth. Furthermore, in the instructions of the income tax forms it is said explicitly that married women should use their maiden surname.

As mentioned, it is possible to change one's surname, in which case also one's fiscal code is changed. This same procedure applies also to married women who want to officially add their husbands' surnames to their original ones or even replace their maiden surnames with their husbands'. Hence, in the vast majority of cases both men and women file their tax reports using their inherited surname. This means that technically we can calculate the ICS for the whole of the society using tax data for both males and females. In practice, we will do it only for males (as most of the literature).

4 Data

In this paper we exploit very rich individual-level micro-data from Italy with information on taxable income as well as the (anonymised) surnames of individuals. From these we obtain different measures of the ICS at the provincial level. We then link such measures with macroeconomic variables at the same level of geographical aggregation.[2]

4.1 Tax records

Our main indicators of mobility are the ICS computed using data from the universe of all the Italian official tax declarations of the year 2005. These declarations were submitted between the beginning of May and mid June 2005 and refer to all incomes (excluding capital incomes) earned between January 1st and December 31 2004. Unfortunately only this year of data is available for research purposes [3]

Despite covering the entire universe of submitted declarations, our data do not necessarily include the whole Italian population. Although in principle every resident in Italy is required to submit a tax declaration, there are exceptions. The first and most important exception are children (and any other dependent family members) who are not required to submit their own tax forms but appear in the forms of their parents (either one or both) who may be eligible for family allowances.[4] The second important category are persons whose income falls below

[2]The exact number and boundaries of the provinces have changed a few times over the recent decades. We use the definition of provinces as of 2004, which is the reference year of our tax data although the current (2015) definitions are slightly different.

[3]The origin of these data is a bit peculiar. They were published online on the website of the Italian Ministry of Finance on April 30th 2008. This was the first time non-anonymised individual tax declarations were made available through the internet in Italy and it was supposed to be part of a general strategy against tax evasion via decentralized social control and stigma. Formally these data had always been legally accessible but the law regulated very strictly the procedures to access this information and the Italian authority for the protection of privacy considered online publication to be illegal. The Ministry was then requested to remove the data from their website. The Authority also clarified that whoever had obtained the data through the Ministry's website had done so legally and were thus allowed to use them. However, the norms regulating access to the data still applied and it was (and still is) prohibited to publish these data online, at least in their original format. For this project we have produced a fully anonymised version of the data with individual names and surnames replaced by numerical codes (still allowing the identification of individuals sharing the same names or surnames) which we use to produce all the results in the paper and that can be distributed for replication. The same data have been used by Braga, Paccagnella, and Pellizzari (2014); Anelli and Peri (2013). Researchers at some institutions, such as the Ministry of Finance or the Bank of Italy, might have access to more detailed data covering longer time periods under special agreements.

[4]Technically, one is considered depended family member if one's income is below a fixed threshold (€2,840.51 in 2004). Submitting one's own declaration separate from that of the household head is, however, always possible.

a given threshold who are exempted from submitting a tax declaration. The exact threshold depends on the composition of the income sources and varies between €3,000 and €7,500 in the year of our data. Among this second group of exemptions are also those who earn exclusively capital income, which is taxed separately in Italy and does not enter the calculation of taxable income.

There are three different forms of tax declarations in Italy. Persons who only have incomes from dependent employment are deducted their taxes directly from their monthly salaries and their employers submit a summary tax report for them. So technically these persons do not submit any form themselves. The second form is used by those who have incomes from both dependent employment and other sources. Finally, the third form is for all those who don't fall in any of the first to groups, namely the self-employed and those with incomes from rents and dividends. In our data each of these forms is used by about one third of the taxpayers.

All three tax forms are quite voluminous, from 6 to 30 pages depending on the exact situation of the taxpayer. However, our data contain only a limited subset of this information, namely the names of the person submitting the file, their dates of birth, the province of residence, total taxable income, the most prevalent source of income (e.g. dependent employment, self-employment, rents and dividends), the amount of the tax due and the form used for the declaration.

In the original data the first name and the surname of the taxpayer are coded in a single string variable and in order to separate them we have used the following procedure. First, we considered only those cases in which the original string contained only two separate words, indicating that the person only has one name and one surname. For these cases we know that the first word is the first name and the second is the surname. About 70% of cases in our data were settled in this simple way. For the others we created an archive of first names using those derived in the first step of our procedure complemented by a number additional lists of typical Italian first names.[5] Next, we consider records with more than two words in the original string variable and we code as surnames the continuous sequences of words that do not appear in our archive of first names. The sequences are continuous in the sense that the algorithm

[5]For this we use a number of websites and books providing guidance to parents choosing a name for their newborn.

takes into account the fact that the original string must be formed by a sequence of first names followed by a sequence of surnames and the two cannot be mixed. We then code as first names the remaining sequences of words. Our archive of first names also allows classifying them by gender, although about 7.5% of the records cannot be unambiguously assigned to a gender.[6]

Overall, there are 38,514,292 records in the original tax files, which compares with about 50 million residents in Italy aged 15 and over in 2004 or about 80% of the entire population who could legally earn incomes. After dropping 2,932,851 observations for which the information on gender is not reliable we are left with 35,581,441 observations. In order to limit complications due to the process of labour market participation, we focus exclusively on men aged between 16 and 100 years, that is, working age and retired men, excluding outliers. [7] This leaves us with 19,247,222 men. Finally, we also exclude individuals with unique surnames in their province (356,331 observations) as the ICS is not defined for them. This leaves us with 18,890,891 observations, of which 18,884,811 have non missing taxable income.

The latter, as recorded in the tax declarations, is our main indicator of economic success and the basis for our analysis of mobility. According to the Italian legislation as of 2005, taxable income is the sum of all gross earned incomes (excluding capital income) minus deductions which are granted for a number of reasons (e.g. number of children, mortgage interests on first homes, some medical and educational expenses, etc.). Importantly, there are no differences across geographical areas in the rules defining fiscal deductions. Due to these allowances and to the fact that self-employed can report losses, taxable income can be zero. The existence of the allowances also implies that individuals with the same taxable income may end up paying different amounts of taxes. For robustness, in Appendix **??** we present results based on the net tax paid.

Table 1 (Panel A) reports some descriptive statistics for our data. The final working population contains about 19 millions taxpayers with an average annual gross income of about 15,500 Euros and a standard deviation of almost 43,000 Euros, approximately 2.8 times the average. A non negligible fraction of individuals declare zero income, around 18% in our population.

[6] Note that these ambiguities are much more likely to arise for foreigners than Italians.

[7] The 2004 tax records refer to earnings in 2003. That year in Italy education was compulsory until the age of 15, which means that our youngest working men in 2004 are 16 years old.

Table 1. Tax records: descriptive statistics

Variable	Mean	Std. Dev.	Min	Max	N
	Panel A: individual-level				
Taxable income	15,737.21	42,993.09	0	101,255,692	18,884,811
	Panel B: surname/province-level				
Number of individuals in the province per surname (a)	16.32	60.43	2	18,684	1,157,740
Number of individuals in the province (b)	334,004.3	35,3625.6	30,632	1,249,617	1,157,740
Frequency of surname (a/b) (× 10,000)	0.890	2.815	0.016	237.199	1,157,740

Source: 2005 Italian tax records. Sample: males aged 16-100 years old.

Given the size of this fraction, we to keep them in the estimation population and take the log of (1+taxable income). As it is common with most distributions of incomes, there is a relatively long right tail with the 95% percentile at around 50,000 Euros and the 99% percentile just over 100,000 Euros.

For the purpose of constructing the ICS, the distribution of surnames is perhaps more interesting than the distribution of individuals (Table 1, Panel B). We have slightly less than one million surnames (treating the same surname in different provinces as different units) with 16.4 individuals holding the same surname on average in the same province. Considering that the average province has about 335,000 residents, each surname covers slightly less than one (0.88) every ten thousand persons. Additionally, the distribution of surnames is very skewed, as predicted by the rules of surname transmission. The median frequency of surnames is one every 40 thousands and the 25% percentile is one every 90 thousands. This indicates that the probability that any two persons selected at random in the population share the same surnames is extremely low and, thus, that the probability of their being linked by some family tie is extremely high.

4.2 Macrodata

For each of the 103 provinces we collect various aggregate economic and social outcomes from the Italian National Institute of Statistics (ISTAT), unless otherwise explicitly specified. Our ICS indicators are produced using data on incomes earned in 2004, hence we consider macro variables at the provincial level around the same time period. Moreover, we think about the correlations between intergenerational mobility and the aggregate outcomes as a structural relations, therefore whenever possible we average the outcomes over the years 1999-2003 to limit cyclical fluctuations and concentrate on long-run structural correlations.

For the sake of expositional clarity, we organise the provincial variables into 2 main different categories: economic outcomes and socio-political outcomes. Tables 2 and 3 provide a full list of such variables along with their descriptive statistics.

Without going into the details of each variable, it is worth noticing the great deal of heterogeneity that characterises the Italian provinces. For example, value added per capita is on average equal to €18,830 (Table 2). However, the province at the 90^{th} percentile (Brescia) is 30% above the average, namely €24,717, and the province at the 10^{th} percentile (Trapani) is 37% below, namely €11,930. Thus, value added per capita is twice as large in Brescia as in Trapani. The same very heterogeneous pattern across provinces arises for all economic variables in Table 2 including labour market outcomes, trade openness, education and cross-sectional inequality.[8]

Table 3 reports descriptive statistics for our selected social outcomes, that can ge grouped into 5 subcategories: life expectancy, suicides, crimes , social capital (such as voters turnout and newspaper sales) and public sector activity. The latter consists of variables capturing the degree of intervention of both the central and the local governments (value of public works started and completed, either by the central or the local government) and the efficiency of local governments (delay of payments to suppliers, measured by the ratio between paid and committed outlays in the municipal budget within the year, schooling level of the local politicians and the budget deficit). As for economic performance, the data display a very large degree of variability across provinces, with the exception – perhaps not surprisingly – of life expectancy.

5 Surname distributions and ICS

We use Italian tax records described in (4.1) to obtain the surname distributions of Italian taxpayers for each province. To our knowledge, this is the most complete data set with

[8]Our data of course confirm the well-known fact that provinces in southern Italy perform worse than those in the centre and in the north in terms of economic outcomes. We omit to report the detailed geographical breakdown to save space.

TABLE 2. Macro outcomes: descriptive statistics

	N	mean	Percentiles 10	50	90
Economic activity					
Value added per capita	103	18,830	11,932	19,378	24,717
Value added per full time equivalent worker	103	46,118	38,233	45,340	50,566
Protested cheques per 1000 inhabitants	103	564.5	211.9	460	1,034
Labour market					
Unemployment rate	103	9.322	3.237	5.854	21.53
Unemployment rate - Males	103	6.725	1.933	3.921	16.39
Unemployment rate - Females	103	13.75	4.931	8.877	31.40
Unemployment rate in the age group 15-24 years	103	25.95	8.715	18.20	54.33
Long-term unemployment rate (12 months or more) - Total	103	3.850	0.962	2.136	9.238
Employment rate	103	45.22	34.92	47.40	52.70
Employment rate - Males	103	56.73	48.97	57.68	63.62
Employment rate - Females	103	34.52	21.75	37.40	42.42
Employment rate aged 15-24 years	103	28.46	13.43	31.23	41.37
Employment rate of individuals aged 25-64 with a diploma	103	73.69	60.43	76.87	82.02
Employment rate of individuals aged 25-64 years with a degree or doctorate	103	79.61	72.53	80.18	85.48
Participation rate in the age group 15-64 years	103	61.24	52.03	63.37	68.57
Participation rate in the age group 15-64 years - Males	103	73.82	69.61	74.11	77.44
Participation rate in the age group 15-64 years - Females	103	48.64	33.30	51.31	59.75
Participation rate in the age group 15-24 years	103	32.92	24.05	33.03	40.92
Openness					
Imports to value added	103	172.9	38.74	152.9	315.9
Exports to value added	103	204.0	35.46	194.9	412.9
Education					
Individuals 25-26 with at most secondary school per 100 same age individuals	103	52.84	44.96	52.61	61.58
Early school dropout aged 18-24 per 100 same age individuals	103	22.26	14.32	21.54	31.88
Inequality					
Log 90/10 income percentile (based on tax records)	103	3.037	2.684	2.912	3.526

Table 3. Socio-political outcomes: descriptive statistics

	N	mean	Percentiles 10	50	90
Life Expectancy					
Life expectancy at birth, males	103	77.45	76.27	77.53	78.60
Life expectancy at 65, males	103	17.05	16.37	17.07	17.70
Life expectancy at birth, females	103	83.22	82.27	83.30	84.13
Life expectancy at 65, females	103	20.99	20.33	21.07	21.67
Suicides					
Suicides per 100,000 - Total	103	7.272	3.887	6.954	10.99
Suicides per 100,000 population - Males	103	10.19	2.361	9.788	17.14
Suicides per 100,000 population - Females	103	2.950	0.474	2.645	5.583
Suicide attempts per 100,000 - Males	103	7.129	2.043	5.607	16.86
Suicide attempts per 100,000 - Total	103	7.621	3.213	6.393	13.58
Suicide attempts per 100,000 - Females	103	7.401	1.211	5.163	18.55
Crimes					
Total crimes	103	3,520	2,409	3,284	5,106
Violent crimes	103	162.1	110.4	146.6	219.8
Thefts	103	1,932	1,013	1,775	3,106
Other crimes	103	1,467	1,040	1,410	1,948
Murders per 100,000 inhabitants	103	1.217	0	0.919	2.439
Sleight of hand per 100,000 inhabitants	103	163.1	21.03	105.7	368.9
Theft with tear per 100,000 inhabitants	103	27.03	4.798	15.65	62.87
Burglaries per 100,000 inhabitants	103	425.0	225.1	398.4	588.2
Theft of parked cars per 100,000 inhabitants	103	355.5	152.2	304.4	622.5
Car thefts per 100,000 inhabitants	103	231.7	68.66	149.1	496.0
Scams per 100,000 inhabitants	103	123.8	73.07	117.3	168.9
Smuggling offenses per 100,000 inhabitants	103	12.54	0.319	1.114	28.57
Drug production and sale for 100,000 inhabitants	103	63.07	27.98	52.59	97.00
Exploitation of prostitution per 100,000 inhabitants	103	4.767	1.729	3.611	8.146
Distraints per 1,000 inhabitants aged 18 years and older	103	8.026	3.434	7.238	13.47
Distraints per 1,000 families	103	17.06	7.393	15.12	27.59
Social Capital					
Voters per 100 voters in the Chamber of Deputies	103	82.05	74.86	83.23	87.47
Voters per 100 voters in the Senate of the Republic	103	82.17	74.54	83.18	87.58
Voters per 100 voters in the European Parliament	103	73.94	63.09	75.12	81.09
Newspaper sales per capita	103	0.234	0.0540	0.130	0.481
Public Sector Activity					
Value of public works started (pct of VA)	103	17.36	4.517	10.23	25.22
Value of public works started by Provincial institutions (pct of VA)	103	0.867	0	0.267	1.764
Value of public works started in the construction sector (pct of VA)	103	3.113	1.042	2.477	5.525
Value of public works completed (pct of VA)	103	12.39	5.151	9.825	20.30
Value of public works completed by Provincial institutions (pct of VA)	103	0.644	0	0.295	1.631
Percentage politicians with at least secondary education (2001)	103	0.0232	0.0200	0.0230	0.0271
Ratio of paid to committed expenses	102	77.58	73.89	77.82	80.49
Deficit per capita in Euros	103	12.17	3.889	11.66	22.82
Growth rate of deficit per capita in Euros ($\times$100)	103	-5.030	-108.1	-0.717	14.05

Source: ISTAT unless otherwise specified. Information on *Newspaper sales* has been drawn from `dati.adsnotizie.it`. The *Ratio of paid to committed expenses* and the *Deficit per capita* is from Gagliarducci and Nannicini (2013); the *Percentage of politicians with at least secondary education* is available from the Ministry of Internal Affairs.

(anonymised) surnames available for Italy, the closest to a census. As discussed, surname distributions are very skewed. To the extent that those distributions – the complex result of fertility processes, (assortative) mating and migration patterns – are similar, any differences in the ICS will reflect differences in mobility.

As it is well known that the Pareto distribution – completely characterized by two moments, the Gini coefficient and the number of persons per surname – provides a good approximation of the surname distributions in many societies (Fox and Lasker (1983)), Figure 1 plots, for every province, the Gini coefficient and the average number of persons per surname.[9] While the Gini indices seem relatively homogeneous within the range $[0.6, 0.9]$, the average number of persons per surname spans between 10 and 50. To enhance cross-province comparability, as explained in Section 2.2, we also calculate the ICS measures concentrating on the right tail of the distribution of surnames, i.e. in each province we focus on the individuals whose surname contains less than a certain amount of people (we experiment with 15, 20, 25 and 30). The idea behind this strategy is that, for these sub-populations, surnames measure the same degree of "family connectivity" in different provinces. Thus, the mapping from these measures into family relationship is very similar (if not identical) across provinces. In other words, by insuring the same surname distributions we insure that we are measuring the same across all provinces. We are comparing alike with alike, and the differences in ICS reflect differences in the intergenerational persistence of the income process across provinces. Figures 1(a) to 1(e) show the Gini coefficient and the number of persons per surname for the whole population and for the tails of the distribution. These figures show that indeed we gain in cross-provincial comparability when focusing on the tails of the distribution as the surname distribution becomes essentially identical in all of them. We are thus quite confident that (at least when using the tails) surnames map family relationship in the same manner in all provinces, and that the mapping from ICS to income persistence is the same in all of them.

In any case, just for the possibility that differences in migration patterns could account for difference in this mapping from surnames to families, we also do the exercise of calculating for each province the ICS of the tails of the surname distribution but only for the 50% of the

[9] Appendix ?? shows the surname distributions for the 20 Italian regions. The 103 provincial surname distributions provide a very similar visual impression.

Table 4. ICS measures based on taxable income: descriptive statistics

	N	Mean	St.Dev.	Percentiles 10	50	90
ICS based on taxable income,	103	0.0247	0.0087	0.0151	0.0236	0.0370
ICS based on taxable income, tail 30	103	0.0456	0.0171	0.0289	0.0389	0.0724
ICS based on taxable income, tail 25	103	0.0478	0.0179	0.0311	0.0406	0.0751
ICS based on taxable income, tail 20	103	0.0505	0.0190	0.0332	0.0426	0.0802
ICS based on taxable income, tail 15	103	0.0540	0.0205	0.0351	0.0456	0.0842

Source: 2005 Italian tax records. Sample: males aged 16-100 years old.

population with the most local surnames as defined in section 2.3.

5.1 Empirical measures of the ICS

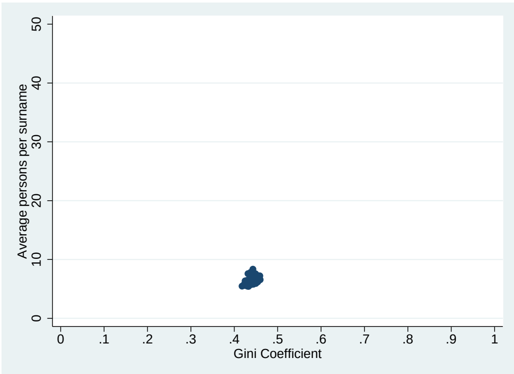
This section presents the ICS measures calculated on the population of Italian males taxpayer described in Section 4.1. We first calculate, for each province, the ICS measure on the full male population taking the difference between the R^2 of a regression of (log) taxable income on a full set of surname dummies (and age dummies) and the average R^2 of M identical regressions[10] in which surnames are previously randomly reshuffled across individuals. This allows to control for the fact that (part of) the variance of income may be mechanically explained by the very inclusion of such a large set of dummies.

Descriptive statistics for ICS measures based on taxable income are reported in Table 4. The first row refers to the ICS calculated on the full population. Subsequent rows report the ICS restricting the population to the individuals with the least frequent surnames, i.e. those containing less than 30, 25, 20, and 15 persons. Overall, the Table shows that there is substantial variation in the ICS across provinces: the baseline ICS of the province at the 90^{th} percentile (Latina) is almost 2.5 higher than the ICS of the province at the 10^{th} percentile (Macerata). Table 4 also shows – not surprisingly – that the ICS monotonically increases when focusing on more and more infrequent surnames.

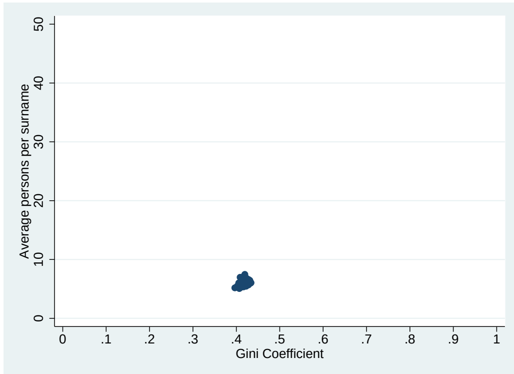
Figures 2 and 3 provide a geographical breakdown of the estimates of social mobility and show, consistently across the different ICS measures, that social mobility is higher in the Centre and North (highest in the North-West) and lower in the South (lowest in the South-West).

[10] We set $M = 10$ and experiment with more replications without changes in the results.

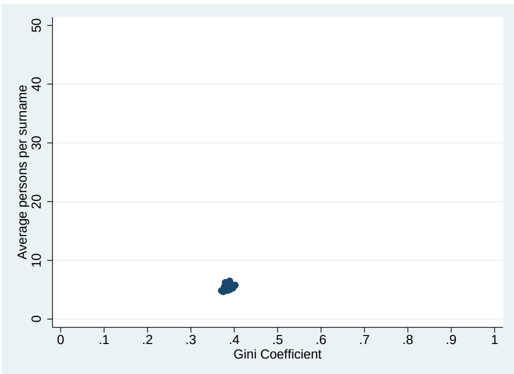
(a) All Individuals



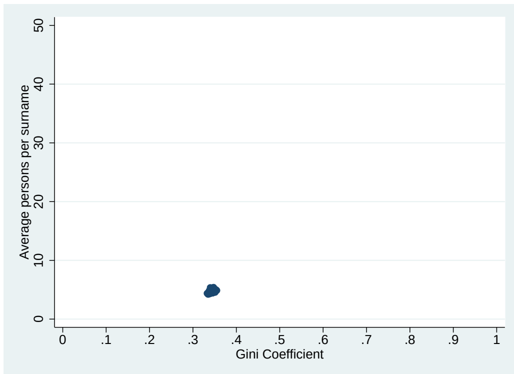
(b) Individuals with surnames with < 30 people



(c) Individuals with surnames with < 25 people



(d) Individuals with surnames with < 20 people



(e) Individuals with surnames with < 15 people

FIGURE 1. Comparability of surname distributions across provinces.

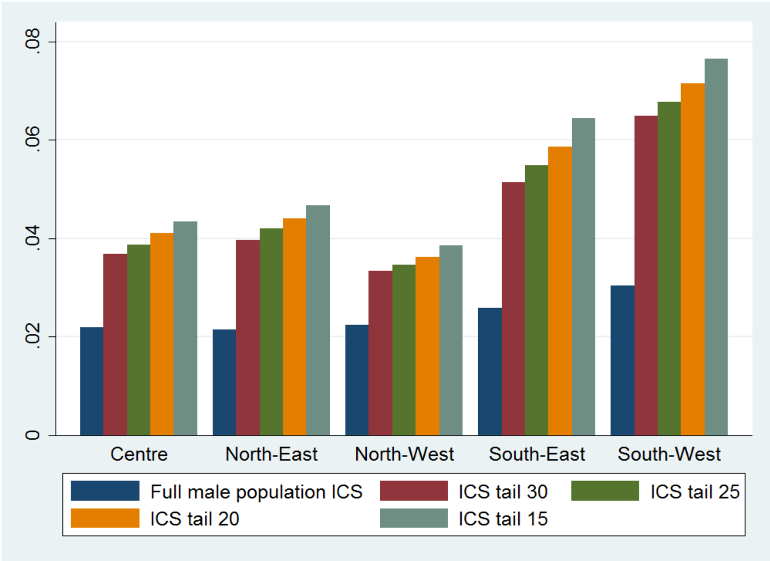


FIGURE 2. ICS measures based on taxable income by geographical areas

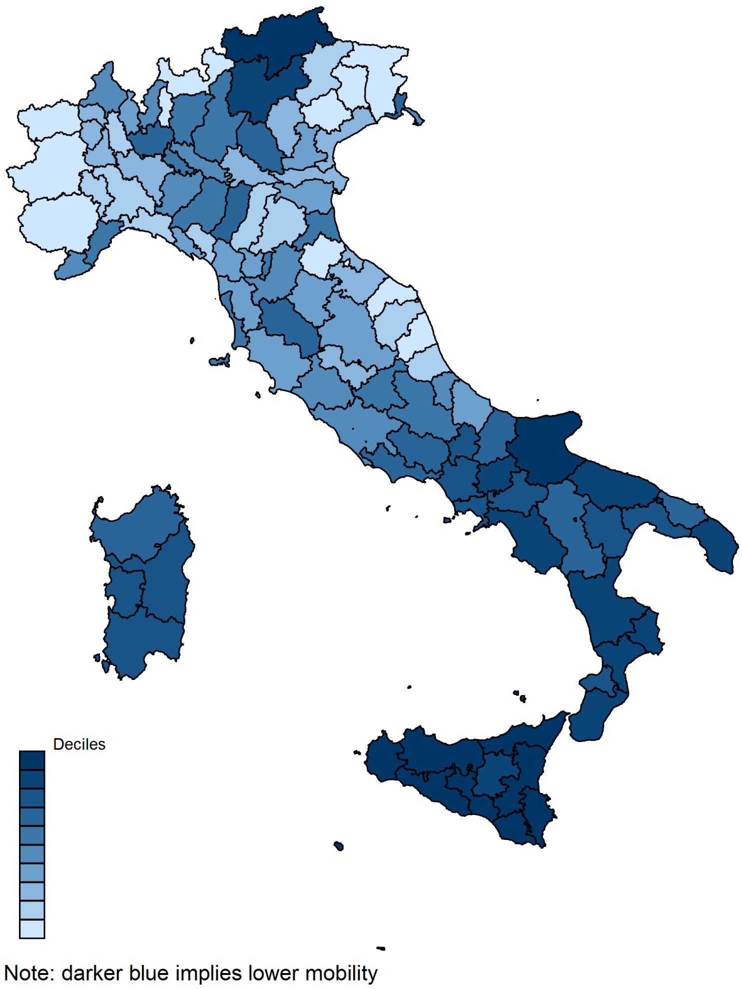


FIGURE 3. ICS (tail 30) based on taxable income across Italian provinces (deciles)

Table 5. ICS measures based on taxable income exploiting how local surnames are: descriptive statistics

				Percentiles		
	N	Mean	St.Dev.	10	50	90
ICS based on taxable income, local	103	0.0243	0.0102	0.0124	0.0219	0.039
ICS based on taxable income, local and tail 30	103	0.0507	0.0195	0.0326	0.0463	0.072
ICS based on taxable income, local and tail 25	103	0.0546	0.0209	0.0340	0.0501	0.074
ICS based on taxable income, local and tail 20	103	0.0587	0.0221	0.0385	0.0525	0.082
ICS based on taxable income, local and tail 15	103	0.0643	0.0250	0.0414	0.0586	0.089

Source: 2005 Italian tax records. Sample: males aged 16-100 years old.

Table 6. Pairwise correlations across ICS measures

	Full ICS	ICS-15	ICS-20	ICS-25	ICS-30	Local ICS	Local ICS-15	Local ICS-20	Local ICS-25	Local ICS-30
Full ICS	1.0000									
ICS-15	0.6923	1.0000								
ICS-20	0.6967	0.9934	1.0000							
ICS-25	0.6984	0.9893	0.9956	1.0000						
ICS-30	0.7025	0.9870	0.9941	0.9960	1.0000					
Local ICS	0.9096	0.5339	0.5316	0.5299	0.5369	1.0000				
Local ICS-15	0.5781	0.8548	0.8475	0.8436	0.8442	0.5076	1.0000			
Local ICS-20	0.6193	0.8721	0.8715	0.8693	0.8713	0.5495	0.9849	1.0000		
Local ICS-25	0.6373	0.8673	0.8698	0.8718	0.8737	0.5679	0.9774	0.9923	1.0000	
Local ICS-30	0.6482	0.8672	0.8739	0.8750	0.8805	0.5779	0.9686	0.9875	0.9935	1.0000

Notes: Full ICS refers to the ICS calculated with the full male population ICS. All other ICS are calculated with the relevant tail of the surname distribution. Local ICS is calculated with only the 50% of the population with the most local surnames. Source: 2005 Italian tax records. Sample: males aged 16-100 years old.

Table 5 shows descriptive statistics for ICS measures based on taxable income and calculated for the fraction of individuals in the top 50 percent of the distribution of the $LocalDegree(s, r)$ Index in every province, as described in Section 2.3. From the second row (as in Table 4) we further restrict the population to the most infrequent surnames. Overall, we again see marked variation across provinces and a monotonically increasing pattern of the ICS as we restrict to more and more infrequent surnames. The geographical breakdown (not reported) of the local ICS provides a picture similar to the one that emerges from Figure 2.

Table 6 displays the pairwise correlations between all the ICS measures shown in Tables 4 and 5. Correlations are all very high (and all significantly different from zero). In particular, it is reassuring to see that the *Full male population ICS* and the *Local ICS* are very correlated (0.9068), implying that differential migration patterns across provinces are not likely to be a major source of concern.

Table 7. Pairwise correlations between ICS and traditional intergenerational elasticity

	Full ICS	ICS-30	ICS-25	ICS-20	ICS-15
Traditional IM measure	0.7878	0.7310	0.7225	0.7233	0.7297
	(0.1136)	(0.1605)	(0.1680)	(0.1673)	(0.1617)
Level aggregation & observations	5 areas	5 areas	5 areas	5 areas	5 areas
Traditional IM measure	0.2314	0.2527	0.2623	0.2836	0.2799
	(0.3264)	(0.2824)	(0.2639)	(0.2257)	(0.2320)
Level aggregation & observations	20 regions	20 regions	20 regions	20 regions	20 regions
Traditional IM measure	0.4752	0.6626*	0.6875*	0.6930*	0.6980*
	(0.0734)	(0.0071)	(0.0046)	(0.0042)	(0.0038)
Level aggregation	20 regions	20 regions	20 regions	20 regions	20 regions
Observations (exclude 5 regions with least observations)	15 regions	15 regions	15 regions	15 regions	15 regions

Pairwise correlations and p-values in parentheses. (*) indicates significance at the 5% level or better. The traditional IM elasticity as in Checchi, Fiorio, and Leonardi (2013). ICS measures as in tables 4 and 5. Full ICS refers to the ICS calculated with the full male population ICS. All other ICS are calculated with the relevant tail of the surname distribution. Local ICS is calculated with only the 50% of the population with the most local surnames.

5.1.1 Correlation ICS and traditional measure of IM

In this section we compare our ICS measure with a traditional measure of intergenerational mobility. However, for Italy there is no extensive longitudinal data set that would allow us to measure mobility at the province level using traditional methods. Some papers have used data from the *Survey on Household Income and Wealth* (SHIW) from the Bank of Italy, which consists of repeated cross-sections with some retrospective information on fathers characteristics to obtain measures of mobility.[11] However, given the limited number of observations in the SHIW data at the province level, it is not possible to obtain reliable estimates of the traditional measure at such a detailed geographical level. For this reason, we calculate both the traditional measure – following Checchi, Fiorio, and Leonardi (2013) – and ours at a more aggregate levels, namely 20 regions and 5 broad areas, in order to be able to make a meaningful comparison.

Table 7 reports the correlation between the traditional measure and our surname-based measure. Before interpreting the reported correlations we need to be aware of the several caveats here. First, the very small number of observations and the fact that one measure is on income and the other one on years of education. Still, despite this, our surname-based measure

[11] See Piraino (2007), Mocetti (2007) and Checchi, Fiorio, and Leonardi (2013). The only other data source used to estimate mobility in Italy is a survey conducted in 1985 on occupations with retrospective information on parents (Checchi, Ichino, and Rustichini, 1999).

and the traditional one are positively correlated and some of them (even if not significant) have high values. When we drop 25% of the regions with least observations in the SHIW, then the correlations become significant (these are very small regions with a small number of observations). Overall, this is reassuring as it shows that indeed our approach is capturing the different mobility patterns across geographical areas. We can, thus, confidently use use our province-level ICS to explore how social mobility correlates with a number of meaningful macro outcomes.

6 Intergenerational mobility and macroeconomic outcomes

We now turn to the analysis of the correlations between the ICS measures and the battery of (log) macroeconomic outcomes described in Section 4.2. To organize the analysis we divide the macro variables described in section 4.2 in two groups, the first one relating to purely economic variables, and the second to socio-political ones.

6.1 Correlating ICS and Economic Outcomes

Table 8 presents, in each column, the pairwise correlations between different ICS measures (full male population in column 1 and tail-based measure from column 2) and a number of meaningful economic outcomes. Recalling that a higher ICS implies lower mobility, the table shows that "good" economic outcomes measured at the province level, such as value added, wealth, income, employment rates, participation rates, imports and exports, are consistently positively and significantly related to higher mobility; instead "negative" outcomes, such as unemployment rates of different socio-economic groups and shares of low-educated young individuals, are related to lower mobility. This pattern emerges consistently across the different types of ICS displayed in the different columns.

This is our first qualitative result: intergenerational mobility correlates positively with "good" economic outcomes, even controlling for identical institutional set-up. Where (economic) things are good, mobility is high.

In order to provide a visual representation of this result we divide the economic variables in

Table 8. Pairwise correlations between ICS from Taxable Income and Macro outcomes, Men

	Full ICS	ICS-30	ICS-25	ICS-20	ICS-15
Economic Activity					
Value added per capita	-0.3353*	-0.5618*	-0.5618*	-0.5626*	-0.5681*
	(0.0005)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Value added per full time equivalent worker	-0.2260*	-0.2446*	-0.2411*	-0.2338*	-0.2419*
	(0.0217)	(0.0128)	(0.0141)	(0.0175)	(0.0138)
Protested cheques per 1000 inhabitants	0.2240*	0.3724*	0.3646*	0.3817*	0.3692*
	(0.0229)	(0.0001)	(0.0002)	(0.0001)	(0.0001)
Labour market					
Unemployment rate	0.4303*	0.6490*	0.6434*	0.6498*	0.6551*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Unemployment rate (males)	0.4352*	0.6560*	0.6512*	0.6582*	0.6636*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Unemployment rate (females)	0.4245*	0.6487*	0.6427*	0.6488*	0.6539*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Unemployment rate (age 15-24)	0.4414*	0.5775*	0.5720*	0.5799*	0.5879*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Long-term unemployment rate (> 12 months)	0.4071*	0.5776*	0.5685*	0.5803*	0.5815*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Employment rate	-0.5076*	-0.6679*	-0.6642*	-0.6751*	-0.6781*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Employment rate (males)	-0.5053*	-0.5791*	-0.5763*	-0.5880*	-0.5939*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Employment rate (females)	-0.4914*	-0.7068*	-0.7026*	-0.7125*	-0.7133*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Employment rate (age 15-24)	-0.4808*	-0.6338*	-0.6328*	-0.6441*	-0.6524*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Employment rate (high school aged 25-64)	-0.4780*	-0.7168*	-0.7111*	-0.7194*	-0.7208*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Employment rate (college degree aged 25-64)	-0.3912*	-0.5902*	-0.5895*	-0.6029*	-0.6115*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Participation rate (age 15-64)	-0.4327*	-0.6576*	-0.6593*	-0.6690*	-0.6702*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Participation rate (males aged 15-64)	-0.3551*	-0.4141*	-0.4181*	-0.4286*	-0.4331*
	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Participation rate (females aged 15-64)	-0.4323*	-0.7009*	-0.7016*	-0.7106*	-0.7103*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000
Participation rate (age 15-24)	-0.5307*	-0.5386*	-0.5333*	-0.5501*	-0.5543*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Openness					
Imports to values added	-0.2581*	-0.3466*	-0.3427*	-0.3491*	-0.3541*
	(0.0085)	(0.0003)	(0.0004)	(0.0003)	(0.0002)
Exports to values added	-0.4260*	-0.5761*	-0.5714*	-0.5753*	-0.5798*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Education					
Individuals with at most secondary school (aged 25-26)	0.1741	0.4597*	0.4567*	0.4498*	0.4488*
per 100 same age individuals	(0.0786)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Early school dropout (aged 18-24)	0.1843	0.4589*	0.4437*	0.4352*	0.4177*
per 100 same age individuals	(0.0624)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Inequality					
Log 90/10 income percentile	0.0381	0.2311*	0.2435*	0.2510*	0.2473*
	(0.7022)	(0.0188)	(0.0132)	(0.0105)	(0.0118)

Notes: Pairwise correlations and *p*-values in parentheses. (*) indicates significance at the 5% level or better. Full ICS refers to the ICS calculated with the full male population ICS. All other ICS are calculated with the relevant tail of the surname distribution. Local ICS is calculated with only the 50% of the population with the most local surnames.

Table 9. Good and Bad Economic Outcomes

Good Economic Outcomes	**Bad Economic Outcomes**
Value added per capita	*Unemployment rate*
Value added per full time equivalent worker	*Unemployment rate (males)*
Imports to value added	*Unemployment rate (females)*
Exports to value added	*Unemployment rate in the age group 15-24 years*
Employment rate	*Long-term unemployment rate (12 months or more) - Total*
Employment rate (males)	*Protested cheques per 1000 inhabitants*
Employment rate (females)	*Individuals aged 25-26 with at most secondary school per 100 same age individuals*
Employment rate aged 15-24 years	*Early school dropout aged 18-24 per 100 same age individuals*
Employment rate (high school aged 25-64)	
Employment rate (college degree aged 25-64)	
Participation rate in the age group 15-64 years	
Participation rate in the age group 15-64 years (males)	
Participation rate in the age group 15-64 years (females)	
Participation rate in the age group 15-24 years	

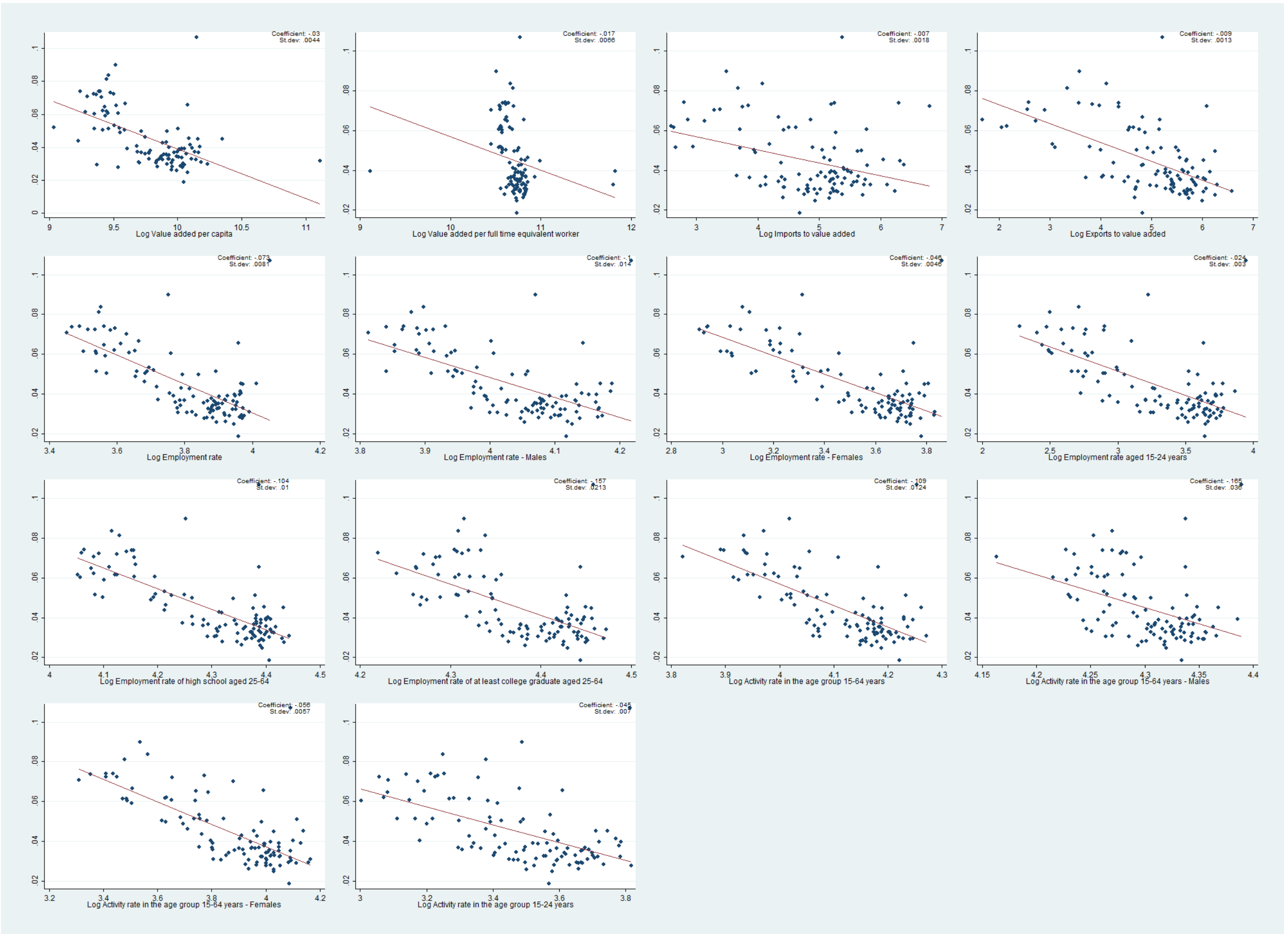


Figure 4. Scatter plots of ICS-30 and "good" economic outcomes

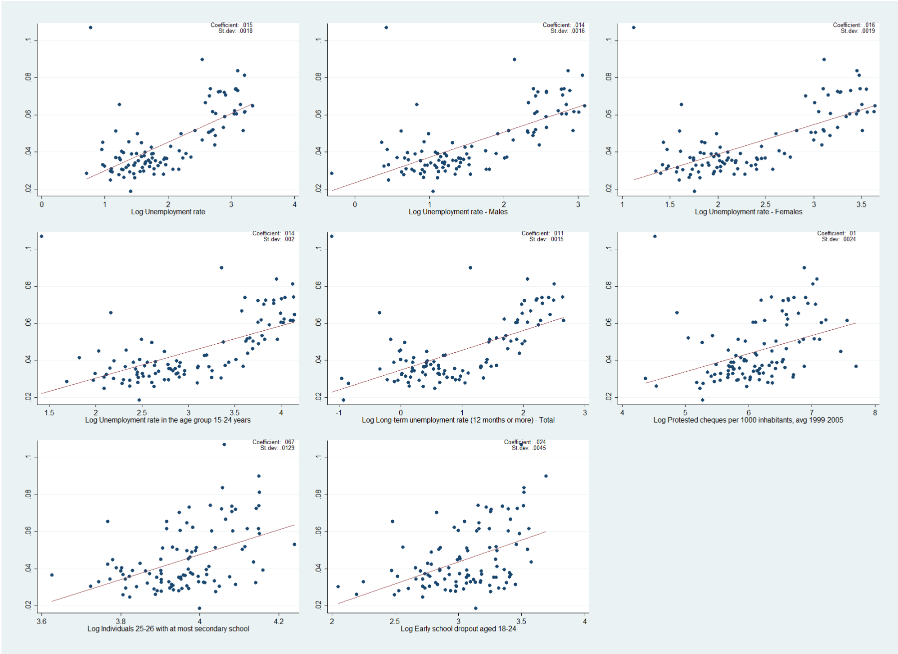


FIGURE 5. Scatter plots of ICS-30 and "bad" economic outcomes

two groups of "good" and "bad" variables (see table 9 for the categorization of each variable). We then plot the scatter plots of the "good" and "bad" variables in Figures 4 and 5 respectively. Moreover, in Figure 6 we plot the value of the correlations (for ICS-30) and their P-values for those variables. It is clear from those graphs that effectively, high mobility happens in places where good (economic) things happen.

The relationship between intergenerational mobility and inequality has a special interest on its own. A clear positive correlation between the intergenerational elasticity of earnings and the degree of cross-sectional inequality – named "the Great Gatsby Curve" – exists across countries. This correlation has become the focus of a large public debate (Corak, 2013; Krueger, 2012) which often interprets it as the result of institutional differences: inequality and the prevalence of inheritance being low in countries with more government intervention as the Nordic countries, and high in *laissez-faire* societies like the Anglo-Saxon countries.

We explore the existence of a Great Gatsby Curve within Italy using as a measure of cross-sectional inequality the 90/10 percentile ratio calculated from our tax data. We correlate it with our ICS measures (see the last row of Table 8) and find our second qualitative result: in provinces where income inequality is lower, inheritance is less prevalent. We plot the Italian Great Gatsby curve in Figure 7, and report the correlation coefficient (and it P-value) in Figure

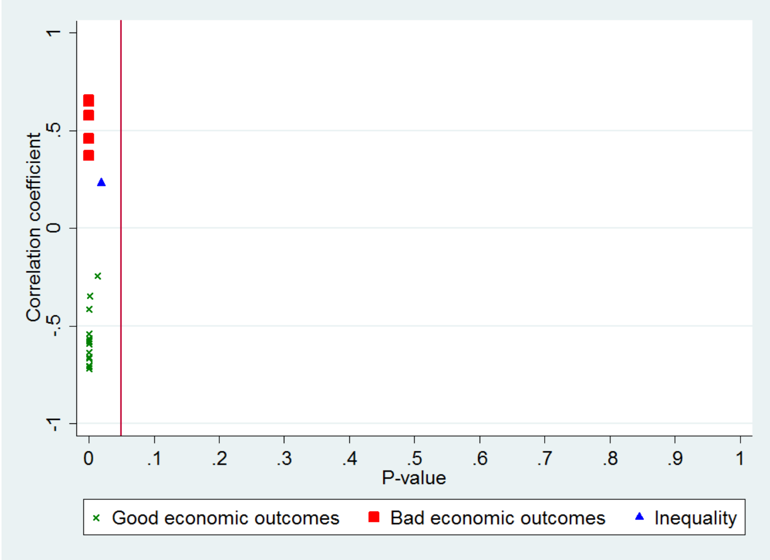


FIGURE 6. Pairwise correlations between ICS-30 economic outcomes and their P-values

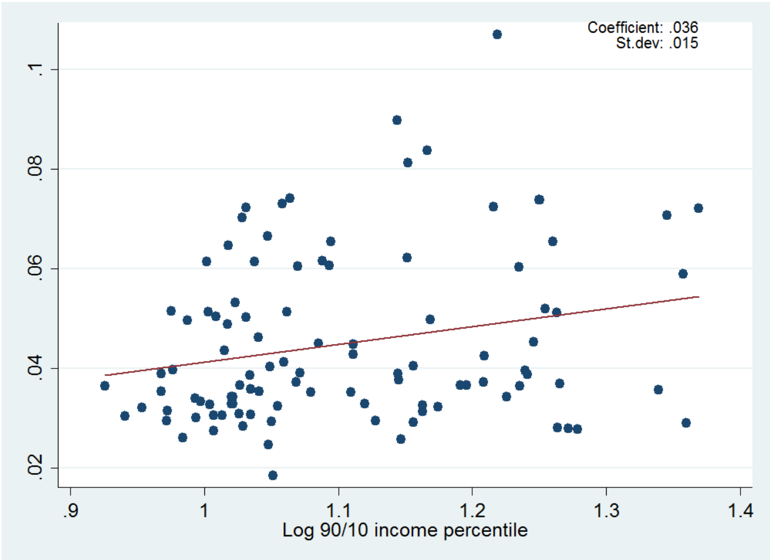


FIGURE 7. The Italian Great Gatsby Curve. Scatter plot of ICS-30 and inequality

6 together with the other economic outcomes mentioned above.

6.2 Socio-Political Variables

We now turn to how socio-political aggregate variables correlate with intergenerational mobility at the province level. In table 10 we present the correlation coefficients between our ICS measures and a battery of socio-political variables categorized in five groups.

A mere glimpse at the table shows that we can not make the same claim as with the economic variables. Thus, our third qualitative result: it is not the case that more social

Table 10. Pairwise correlations of ICS from Taxable Income and Socio-Political variables, Men

	Full ICS	ICS-30	ICS-25	ICS-20	ICS-15
Life Expectancy					
Life expectancy at birth (males)	-0.1297	-0.0197	-0.0169	-0.0010	-0.0115
	(0.1916)	(0.8438)	(0.8657)	(0.9917)	(0.9079)
Life expectancy at 65 (males)	-0.0395	0.0725	0.0765	0.0912	0.0914
	(0.6919)	(0.4669)	(0.4426)	(0.3593)	(0.3587)
Life expectancy at birth (females)	-0.4136*	-0.3695*	-0.3662*	-0.3590*	-0.3616*
	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Life expectancy at 65 (females)	-0.4213*	-0.4957*	-0.4910*	-0.4863*	-0.4882*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Suicides					
Suicides per 100,000 individuals	-0.2296*	-0.4801*	-0.4673*	-0.4860*	-0.4754*
	(0.0197)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Suicides per 100,000 individuals (males)	0.0209	-0.3120*	-0.3061*	-0.3197*	-0.3174*
	(0.8354)	(0.0015)	(0.0018)	(0.0011)	(0.0012)
Suicides per 100,000 individuals (females)	-0.0711	-0.3094*	-0.3068*	-0.3247*	-0.3226*
	(0.4956)	(0.0024)	(0.0026)	(0.0014)	(0.0015)
Suicides attempts per 100,000 individuals	-0.0395	-0.3095*	-0.3057*	-0.3232*	-0.3286*
	(0.6924)	(0.0015)	(0.0017)	(0.0009)	(0.0007)
Suicides attempts per 100,000 individuals (males)	0.0199	-0.2871*	-0.2861*	-0.3034*	-0.3008*
	(0.8438)	(0.0038)	(0.0039)	(0.0021)	(0.0024)
Suicides attempts per 100,000 individuals (females)	0.0395	-0.2116*	-0.2116*	-0.2274*	-0.2420*
	(0.6982)	(0.0355)	(0.0355)	(0.0236)	(0.0158)
Crimes					
Total crimes	-0.2444*	-0.2297*	-0.2288*	-0.2317*	-0.2499*
	(0.0128)	(0.0196)	(0.0201)	(0.0185)	(0.0109)
Violent crimes	-0.1378	0.0365	0.0364	0.0324	0.0331
	(0.1652)	(0.7146)	(0.7154)	(0.7454)	(0.7399)
Thefts	-0.3534*	-0.2496*	-0.2475*	-0.2558*	-0.2781*
	(0.0003)	(0.0110)	(0.0117)	(0.0091)	(0.0045)
Other crimes	0.0676	-0.0967	-0.0991	-0.0947	-0.0950
	(0.4974)	(0.3312)	(0.3195)	(0.3415)	(0.3400)
Murders per 100,000 inhabitants	0.3235*	0.2949*	0.2860*	0.2934*	0.2839*
	(0.0017)	(0.0043)	(0.0057)	(0.0045)	(0.0061)
Sleight of hand per 100,000 inhabitants	-0.2562*	-0.3893*	-0.3888*	-0.3908*	-0.4041*
	(0.0090)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Breaking and entering per 100,000 inhabitants	-0.0453	0.2351*	0.2315*	0.2284*	0.2060*
	(0.6498)	(0.0168)	(0.0186)	(0.0203)	(0.0368)
Burglaries per 100,000 inhabitants	-0.4595*	-0.5286*	-0.5292*	-0.5306*	-0.5521*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Theft of parked cars per 100,000 inhabitants	-0.4320*	-0.4204*	-0.4118*	-0.4169*	-0.4324*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Car thefts per 100,000 inhabitants	0.0019	0.2886*	0.2785*	0.2787*	0.2651*
	(0.9846)	(0.0031)	(0.0044)	(0.0044)	(0.0068)
Scams per 100,000 inhabitants	-0.1815	-0.2657*	-0.2605*	-0.2704*	-0.2782*
	(0.0666)	(0.0067)	(0.0079)	(0.0057)	(0.0044)
Smuggling offences per 100,000 inhabitants	0.2948*	0.3096*	0.3114*	0.3139*	0.3127*
	(0.0026)	(0.0015)	(0.0014)	(0.0013)	(0.0014)
Drug production and sale for 100,000 inhabitants	0.0173	-0.1124	-0.1174	-0.1034	-0.1159
	(0.8626)	(0.2585)	(0.2374)	(0.2985)	(0.2436)
Exploitation of prostitution per 100,000 inhabitants	-0.3502*	-0.4711*	-0.4739*	-0.4645*	-0.4697*
	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Distraints per 1,000 inhabitants aged $\geq$18 years	0.0821	0.0949	0.0789	0.0905	0.0853
	(0.4095)	(0.3406)	(0.4280)	(0.3632)	(0.3915)
Distraints per 1,000 families	0.0989	0.1597	0.1446	0.1560	0.1510
	(0.3205)	(0.1072)	(0.1449)	(0.1157)	(0.1278)
Social Capital Measures					
Voters per 100 voters in the Chamber of Deputies	-0.4287*	-0.6199*	-0.6224*	-0.6278*	-0.6372*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Voters per 100 voters in the Senate	-0.2912*	-0.3857*	-0.3831*	-0.3902*	-0.3943*
	(0.0028)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
Voters per 100 voters in the European Parliament	-0.4703*	-0.5941*	-0.5939*	-0.5928*	-0.5943*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Newspaper sales per capita	-0.2372*	-0.4156*	-0.4151*	-0.4136*	-0.4269*
	(0.0159)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Public Sector					
Value of public works started (pct of VA)	0.1866	0.1935	0.1847	0.1730	0.1745
	(0.0592)	(0.0502)	(0.0618)	(0.0805)	(0.0780)
Value of public works started by Provincial institutions	0.2260*	0.3401*	0.3467*	0.3482*	0.3542*
(pct of VA)	(0.0496)	(0.0026)	(0.0022)	(0.0021)	(0.0017)
Value of public works started in the construction sector	0.1659	0.3324*	0.3348*	0.3319*	0.3477*
(pct of VA)	(0.0939)	(0.0006)	(0.0005)	(0.0006)	(0.0003)
Value of public works completed	0.2823*	0.2325*	0.2305*	0.2289*	0.2383*
(pct of VA)	(0.0039)	(0.0181)	(0.0192)	(0.0200)	(0.0153)
Value of public works completed by Provincial institutions	0.2504*	0.3478*	0.3613*	0.3548*	0.3734*
(pct of VA)	(0.0281)	(0.0019)	(0.0012)	(0.0015)	(0.0008)
Percentage of politicians with at least secondary education	0.0419	0.2468*	0.2531*	0.2565*	0.2366*
	(0.6740)	(0.0120)	(0.0099)	(0.0089)	(0.0161)
Ratio of paid to committed expenses	0.0333	-0.0802	-0.0708	0.0619	-0.0477
	(0.7385)	(0.4209)	(0.4775)	(0.5345)	(0.6321)
Deficit per capita in Euro	-0.0987	0.0177	0.0164	0.0042	0.0051
	(0.3361)	(0.8635)	(0.8737)	(0.9672)	(0.9608)
Growth rate of deficit per capita in Euro	0.1032	0.1980	0.2004	0.2011	0.2101
	(0.5101)	(0.2032)	(0.1977)	(0.1960)	(0.1763)

Notes: Pairwise correlations and p-values in parentheses. (*) indicates significance at the 5% level or better.

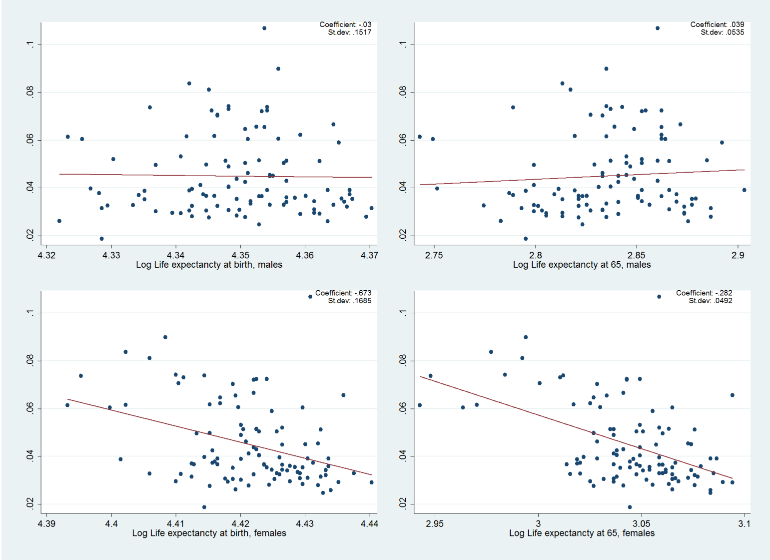


FIGURE 8. Scatter plots of ICS-30 and life expectancy

mobility is systematically associated with better social outcomes.

This can be also seen in the scatter plots for each socio-political group of variables (grouped as explained in section 4.2): figures 8, 9 10, 11, and 12.

Looking at them one by one, the following evidence emerges:

- Social mobility correlates with higher life expectancy for females, but not for males.
- On the other hand, social mobility correlates with higher suicide rates (in essentially any possible categorization of suicides).
- as to crime, while higher mobility is associated to higher total crime rates and higher property crime rates (thefts, sleight of hand, burglaries, exploitation of prostitution), murders and smuggling are related to lower mobility. It is not possible to get any clearcut conclusion on this set of outcomes. For instance, "theft of parked cars" and "theft of cars" exhibit opposite correlations.
- Higher social mobility correlates positively with all the "social capital" proxies available: voters turnout in different types of elections and newspaper sales per capita. For these variables, as for the economic ones, good outcomes are related to higher social mobility while bad ones are related to lower mobility.
- Finally, we also correlate our measures of mobility with measures of public sector activity

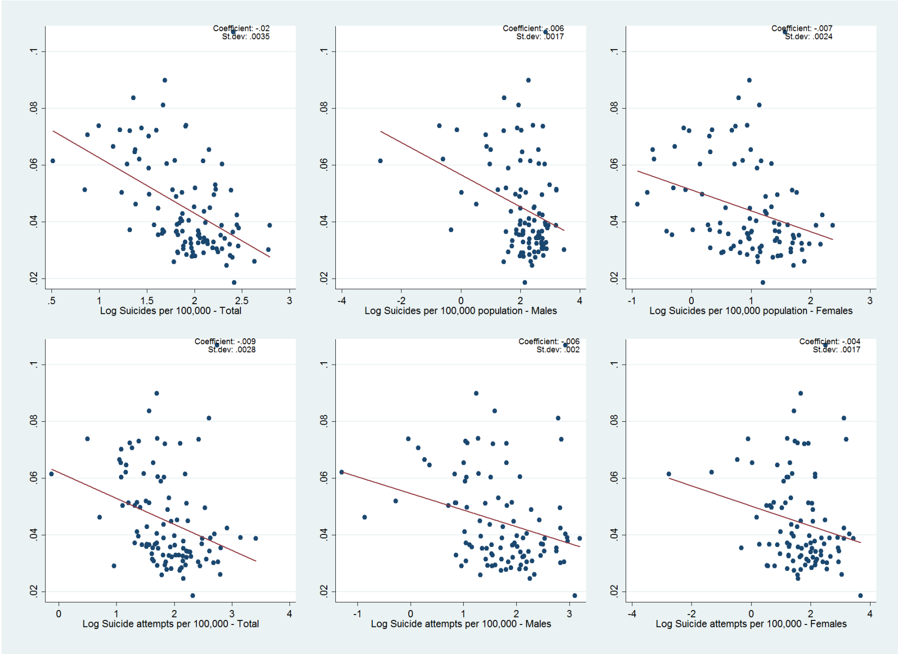


FIGURE 9. Scatter plots of ICS-30 and suicides

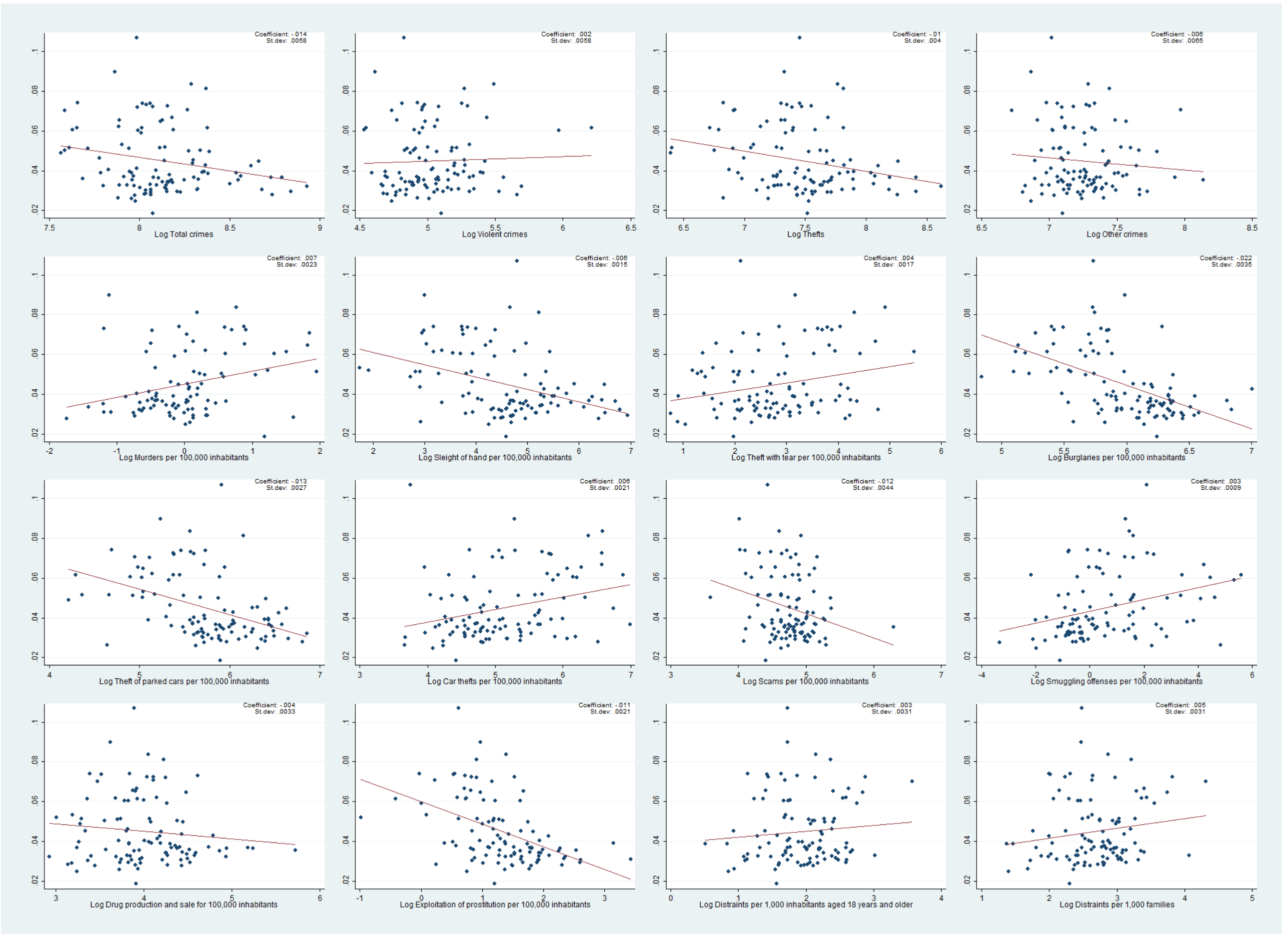


FIGURE 10. Scatter plots of ICS-30 and crime

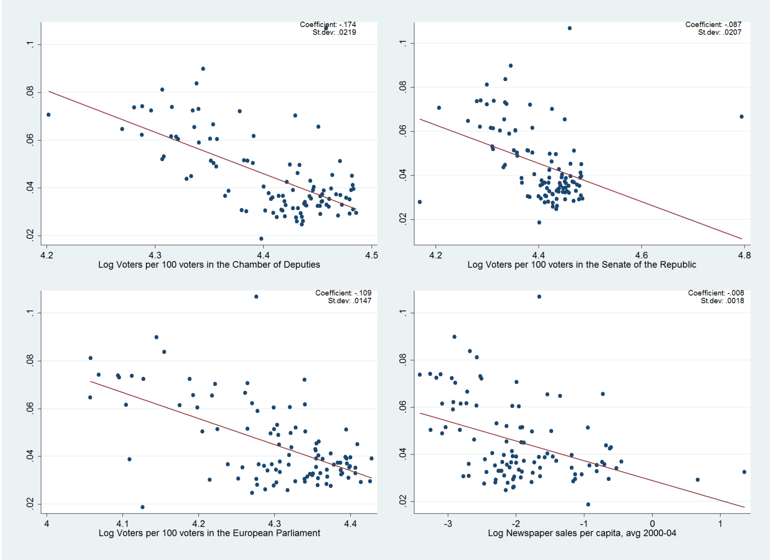


FIGURE 11. Scatter plots of ICS-30 and social capital

at the provincial level. We include them here, and not with the economic variables, because they also capture the reaction of the local governments to the economic conditions of the local areas (intensity of public sector activity measured as the value of the public works started), and the quality of the political system as measured by the schooling level of the politicians and by the budget deficit.

As we said, the picture is much less clear-cut than in the case of the economic variables. In figure 13 we plot the correlations and P-value for all socio-political variables grouped in the 5 groups just described. While some variables which are obviously desirable (like life expectancy for females) are clearly positive related with mobility, others that are obviously negative (like suicide rates) are also positively correlated to intergenerational mobility, and crime is all over the place.

Perhaps not surprisingly, the interaction between mobility and these social variables appears much more complex and unpredictable the interaction with the economic ones, where mobility is systematically associated to desirable outcomes.

7 Conclusions

Is intergenerational mobility a good thing? Well, of course it is. It has to do with equality of opportunities and it is clearly a desirable feature of any society. But things do not often

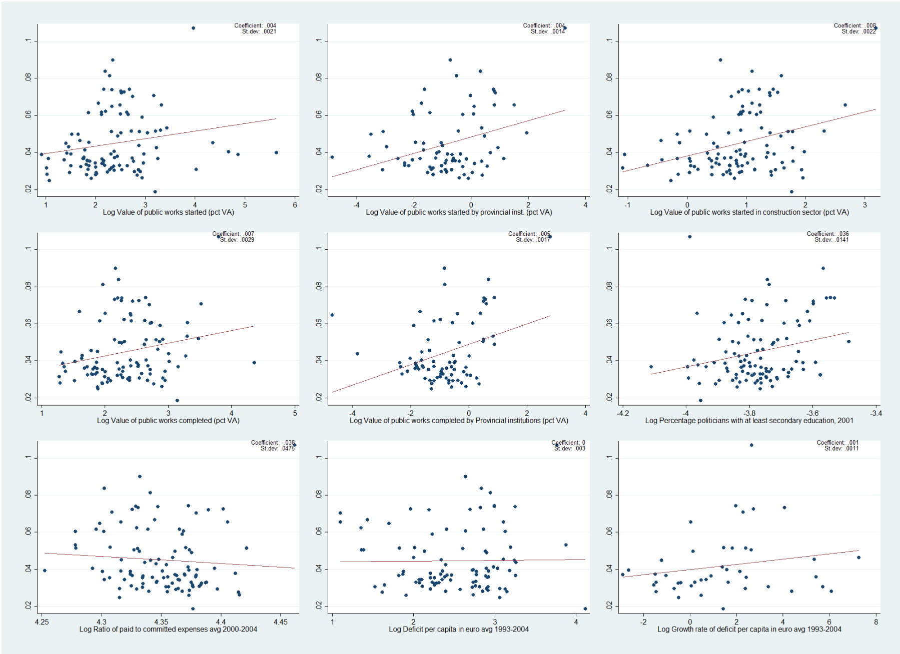


FIGURE 12. Scatter plots of ICS-30 and public sector activity

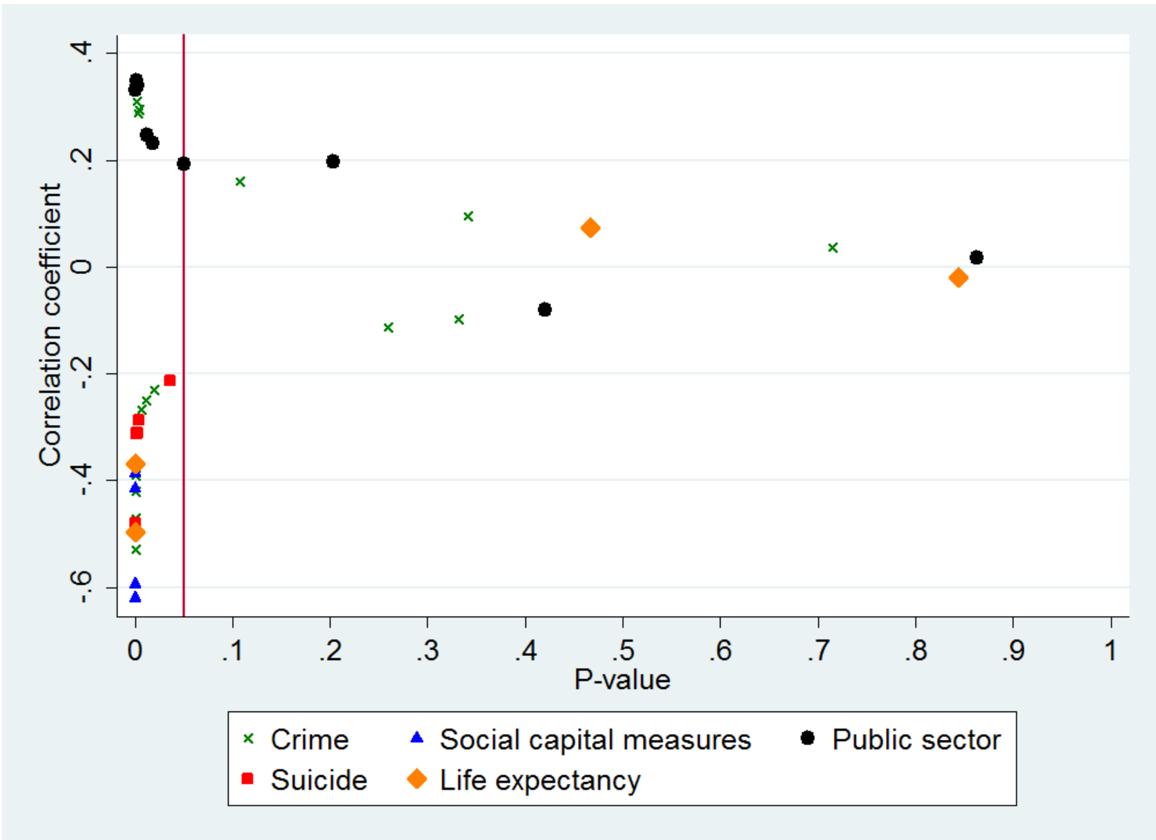


FIGURE 13. Pairwise correlations between ICS-30 and Socio-Political Variables.

come for free. It could be well the case that societies that show high mobility could in principle suffer from undesirable outcomes. Actually, it is often suggested that this is the case, given the evidence that intergenerational mobility in the US is low (albeit with large regional variation, as we know since recently (Chetty, Hendren, Kline, and Saez, 2014)), while income per-capita (along with many other desirable economic dimensions) is high.

In this paper we use Italian data to correlate intergenerational mobility with macroeconomic and social variables at the provincial level. We believe that the advantages of using Italy are manifold:

1. Italy is a relatively large country, composed of many different local entities which are extraordinarily heterogeneous in their economic and social performances.
2. The institutional framework is the same in all provinces. Whatever makes mobility and other outcomes high in one place and low in others is not the difference in the rules of the game.
3. The surname distribution is very similar across provinces, and when concentrating on the tails of the surname distributions, it is essentially identical. Thus, we can reliably use the ICS (a measure of intergenerational mobility developed by Güell, Rodríguez Mora, and Telmer (2014)) to compare intergenerational mobility across provinces because in all of them the mapping from surnames to family relationships is very similar. Thus, in all of them the mapping between the persistence of the income process and the ICS is identical. Therefore, comparing the ICS across provinces is akin to compare the degree of persistence of the income process across generations.

The measurement exercise shows that intergenerational mobility is larger in the North, and lower in the South of Italy. Reassuringly, our ICS measure correlates pretty well with existing measures of mobility (albeit these are based on scarce data and are unsuitable to explore the correlation between mobility and other outcomes).

Our key findings are as follows:

1. Mobility correlates positively with desirable economic outcomes and negatively with undesirable ones.

2. The latter includes economic inequality. Thus, there exists a Great Gatsby curve for Italian provinces even if there are no institutional differences across them.

3. The clear and systematic pattern that shows up for economic outcomes does not exist for socio-political outcomes.

We see as an advantage of our approach that the institutional framework is the same across all units of observation. Nevertheless, it is reasonable to expect that part of the observed differences in economic (and socio-political) outcomes and intergenerational mobility across countries may be due in differences in the institutional framework. More or less redistribution. More or less meritocracy. More or less availability of public education. All these institutional characteristics are extremely likely to affect the degree of intergenerational mobility and the performance of the economy. Our intention is to use the methodology that we have developed here in order to look at the effect of the differences in the institutional setting in further research focused on cross-country comparisons.

References

Anelli, M. and G. Peri (2013). Peer Gender Composition and Choice of College Major. NBER Working Papers 18744, National Bureau of Economic Research, Inc.

Björklund, A., T. Eriksson, M. Jäntti, O. Raaum, and E. Österbacka (2002). Brother correlations in earnings in Denmark, Finland, Norway, and Sweden compared to the United States. *Journal of Population Economics 15*(4), 757–772.

Björklund, A. and M. Jäntti (1997). Intergenerational income mobility in Sweden compared to the United States. *American Economic Review 87*(5), 1009–1018.

Black, S. E. and P. J. Devereux (2011). *Recent Developments in Intergenerational Mobility.* in Orley C. Ashenfelter and David Card (eds.), Handbook of Labor Economics, Volume 4B, Amsterdam: North-Holland, pp. 1487-1541.

Braga, M., M. Paccagnella, and M. Pellizzari (2014). The academic and labor market returns of university professors. *The Journal of Labor Economics.*

Checchi, D., C. V. Fiorio, and M. Leonardi (2013). Intergenerational persistence in educational attainment in Italy. *Economics Letters 118*, 229–232.

Checchi, D., A. Ichino, and A. Rustichini (1999). More equal but less mobile? Education financing and intergenerational mobility in Italy and in the U.S. *Journal of Public Economics 74*(3), 351–93.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the United States. *Quarterly Journal of Economics 129*(4), 1553–1623.

Comi, S. (2003). Intergenerational mobility in Europe: evidence from ECHP. Departmental Working Papers 2003-03, Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano.

Corak, M. (2013). *Inequality from Generation to Generation: The United States in Comparison.* Robert Rycroft editor, The Economics of Inequality, Poverty, and Discrimination in the 21st Century, ABC-CLIO.

Couch, K. A. and T. A. Dunn (1997). Intergenerational correlations in labor market status: A comparison of the United States and Germany. *Journal of Human Resources 32*(1), 210–32.

Fox, W. R. and G. W. Lasker (1983). The distribution of surname frequencies. *International Statistical Review 51 (1)*, 81–87.

Gagliarducci, S. and T. Nannicini (2013, April). Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association 11*(2), 369–398.

Grawe, N. D. (2004). *Intergenerational mobility for whom? The experience of high- and low-earnings son in international perspective.* in Miles Corak (ed.), Generational Income Inequality, Cambridge University Press, pp.58-89.

Güell, M., J. V. Rodríguez Mora, and C. I. Telmer (2014). The informational content of surnames, the evolution of intergenerational mobility and assortative mating. *The Review of Economic Studies*, Forthcoming.

Haider, S. and G. Solon (2006). Life-cycle variation in the association between current and lifetime earnings. *The American Economic Review 96*(4), 1308–1320.

Hertz, T. (2007). Trends in the intergenerational elasticity of family income in the United States. *Industrial Relations 46 (1)*, 22–50.

Krueger, A. B. (2012). The rise and consequences of inequality in the United States. Speech of the Chairman of Council of Economic Advisers at the Center for American Progress on January 12th, 2012.

Lee, C.-I. and G. Solon (2009). Trends in intergenerational income mobility. *Review of Economics and Statistics 91*(November), 766–772.

Mocetti, M. (2007). Intergenerational earnings mobility in Italy. *The B.E. Journal of Economic Analysis & Policy 7,*(Iss. 2 (Topics)), 1935–1682.

Piraino, P. (2007). Comparable estimates of intergenerational income mobility in Italy. *The B.E. Journal of Economic Analysis & Policy 7,*(Iss. 2 (Topics)), 1935–1682.

Solon, G. (1992). Intergenerational income mobility in the United States. *American Economic Review 82*(3), 393–408.

Solon, G. (1999). *Intergenerational Mobility in the Labor Market.* in Orley C. Ashenfelter and David Card (eds.), Handbook of Labor Economics, Volume 4B, Amsterdam: North-Holland, pp. 1761-1800.

Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *Journal of Economic Perspectives 16*(3), 59– 66.