Rule breaking, honesty and migration

Rule breaking, honesty and migration*

Massimo Anelli† Tommaso Colussi§ Andrea Ichino‡

October 19, 2022
Click here for most updated version.

Abstract

Using Census data we study false birth date registrations in Italy, a phenomenon well known to demographers, in a setting that allows us to separate honesty from cheating benefits and deterrence. By comparing migrants and remainers within locality×biennium cells we then illustrate the tendency of Italians to sort themselves across geographic areas according to their honesty. Overtime, this tendency has modified the average honesty in each locality, with relevant consequences for the distribution across geographic areas of outcomes like human capital, productivity, earnings growth and the quality of local politicians and government.

JEL-Code: J61, C93, R23 Keywords: Migration, Aversion to Breaking Rules, Productivity

*We are grateful to Istat for giving us access to the individual observations of the restricted Census at the protected Adele sites. When the Covid pandemic started, access to the Adele sites was no longer permitted, but Istat allowed us to export group averages of the data, as described in the paper, in order for us to be able to continue our research. We are particularly indebted to several colleagues who shared with us their data: Ethan Ilzetzki and Saverio Simonelli (Vote Counting Rates); Josh Angrist, Eric Battistin, Daniela Vuri (school cheating); Lorenzo Casaburi and Ugo Troiano (property tax evasion). We also benefited from conversations with Vittorio Bassi, Diogo Britto, Adriano De Falco, Alice Dominici, Roberto Galbiati, Diego Gambetta, Giulia Giupponi, Joseph Heath, David Levine, Moti Michaeli, Massimo Morelli and Yannick Reichlin as well as from seminar presentations at Berkeley, Berlin, Bocconi, Boston College, Cornell, Dartmouth, Princeton, the 2020 Labor-ski seminar, University of Florida, University of Milan and UC Davis. The Appendix for this paper can be found online at https://www.dropbox.com/s/7bsn9x4ecn1n2d2/ACI_Online_Appendix_Sorting_ABR.pdf?dl=0

†Bocconi University, CESifo and IZA; massimo.anelli@unibocconi.it.

‡European University Institute, University of Bologna, CEPR, CESifo, and IZA; andrea.ichino@eui.eu.

§Catholic University of Milan, IZA tommaso.colussi@unicatt.it.

1 Introduction

Economists think of observed rule breaking as an outcome of a decision process in which subjects compare their aversion to break rules (honesty for brevity) with the gain they derive from rule breaking and with the level of public deterrence against this type of actions. It is therefore surprising that observed rule breaking is often used as an indicator of social capital, because what should enter in such measure is only intrinsic honesty, not the effects of cheating benefits or deterrence.

Using Census data, in this paper we study false birth date registrations in Italy, a phenomenon well known to demographers, in a setting that allows us to separate honesty from cheating benefits and deterrence. We confirm the vital statistics evidence documented by Livi (1929), Maroi (1954) and Breschi, Giovanna, and Gabriele (2018), suggesting that Italians, in some localities more than in others, have a significant tendency to register a false date of birth for their children. Starting in early December of each year, the frequency of registered births per day declines substantially, while an abnormally large mass of registered births is concentrated in the first five days of the following month of January.

The demography literature cited above has described the main motives for this tendency to lie in the registration of birth dates, in particular delaying school entry, compulsory military service, marriage and the age of emancipation. Irrespective of the motive, a parent who registers a false date of birth for a child violates the Italian penal code according to which, at least since 1889, any false declaration in a public or private deed is punishable by imprisonment from 3 to 10 years (*Codice Zanardelli*: art. 278 R.D. 30 June 1889, n. 6133).[1] What makes this indicator particularly interesting for our purposes is that, to our knowledge, it is the only cheating measure that can be computed for groups of the Italian population observed within small localities at different points in time during the 20th century. As in the case of the unpaid parking violations of United Nations diplomats studied by Fisman and Miguel (2007), this setting allows us to compare subjects facing the same degree of local deterrence and deriving similar

[1] Falsification of the date of birth is explicitly punished in the Penal Code of 1930 (Codice Rocco), as discussed in Maroi (1954), p. 414.

benefits from breaking a rule. Differences in birthday cheating between these groups can therefore capture the different inclinations of their members towards honest behaviour.[2]

We then exploit this setting to compare the honesty of migrant and remainer families within small locality×biennium cells, showing that it differs across the two groups in a way that illustrates the tendency of Italians to sort themselves across geographic areas according to their honesty. We cannot nail down the precise reasons of this sorting process. For example cheaters in a locality could obtain material advantages that put them in better economic conditions, thus reducing their need to migrate. Honest citizens, instead, may pay a prize for respecting rules, which in the long run constitutes a push factor for migration. Or it could be that the honest population prefers to move elsewhere if it is continuously free-rided by too many peers who do not care about rules.[3]

Independently of the specific reason that drives sorting based on honesty, this phenomenon may have important long term consequences because, overtime, it can alter the degree of honesty prevailing in a locality and thus change the distribution of social capital across geographic areas. Our data allow us to measure the extent to which average honesty in different Italian localities has changed over the 20th century as a result of migration movements. In this respect, a collateral contribution of our paper is to suggest that sorting based on honesty may be one of the reason of the unequal distribution of social capital and economic prosperity across nearby localities, that has been extensively documented around the world and particularly in Italy.[4]

We then use our measure of the honesty drain or gain experienced by different southern Italian localities to show that those characterized by a more severe honesty drain display today lower indicators of human capital, productivity and wage growth. These localities also appear to be governed by politicians that are themselves characterized

[2]While we cannot exclude that our measure of cheating conditional to location and observable characteristics capture also other dimensions than only intrinsic honesty, for simplicity in the rest of the paper we use the word honesty to refer to low-cheating subjects.

[3]Some of these mechanisms are discussed in Michaeli et al. (2022).

[4]See, for example, Rupasingha, Goets, and Freshwater (2006), Braeseman and Stephany (2017), Cohn et al. (2019), Fisman and Miguel (2007) and Lowes et al. (2017). With specific reference to Italy, the path-breaking book of Putnam, Leonardi, and Nanetti (1993) explores systematically the heterogeneity of social capital measures, and rule breaking specifically, within and between regions. For more recent studies, see Ichino and Maggi (2000), Guiso, Sapienza, and Zingales (2004), Durante and Buggle (2010), Bigoni et al. (2016), Buonanno et al. (2015) and Michaeli et al. (2022).

by higher birthday cheating, and, interestingly, their city councils are more frequently dismissed by the central government, before their normal electoral term, because of corruption or bad functioning. These estimates cannot, be interpreted as causal but, relying on Oster (2019), we show that they are likely to be robust to potential unobserved confounders.

Our paper is organized as follows. We start in Section 2 by describing the historical and the Census evidence on birthday cheating. Section 3 presents evidence on the propensity of Italians to sort themselves across local areas on the basis of their honesty. Section 4 measures the honesty drain or gain at the level of localities and shows that it correlates with important economic outcomes. Section 5 concludes.

2 Historical and Census evidence on the falsification of birth dates in Italy

Using data from the 1991 Italian Census for the cohorts born between 1921 and 1954,[5] the left panel of Figure 1 displays, for the North of the country and, the histogram of the number of births (in thousand) over days of the calendar year grouped in bins of five days. The right panel does the same for the South.[6] While we would expect an almost uniform distribution of birth dates over the year, it is evident that around mid December the frequency of births declines abnormally in the South, to then suddenly increase with a large spike in the first five days of January. In the North the pattern is similar, although less pronounced.[7]

[5]1991 is the first year in which publicly available Census data contain complete dates of birth, as well as cities of birth and residence. See the Online Appendix Section OA1 for a description of the data that completes the information provided in this section. The evidence described below is present also in more recent Census data, as we show in the Online Appendix Figures OA1,OA2.

[6]The South is defined as the set of localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies", for reasons that will be made clear below. The North is the complement set.

[7]This is not the only form of birthday cheating that emerges from the Census.
Italians are also abnormally less likely to be born on the 17th of each month, a form of birthday cheating that is clearly driven by superstition because the number 17 is associated to "La Disgrazia" (The Misfortune) in the traditional game of the Neapolitan Tombola. See Liccardo (2019) and Maroi (1954).

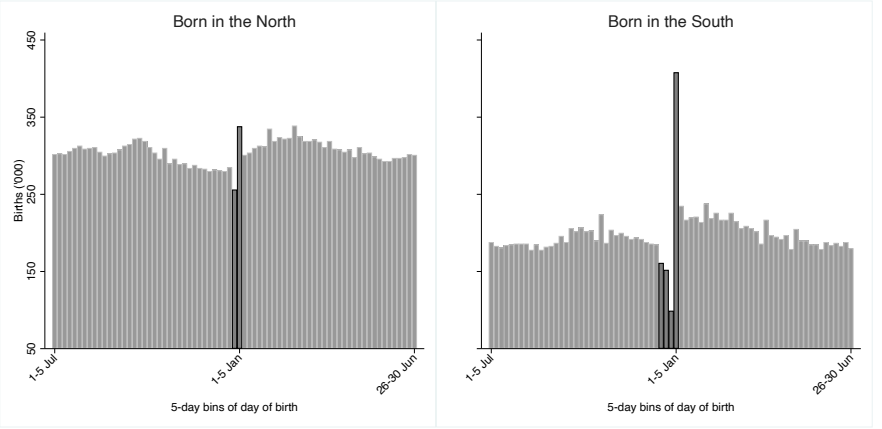

Figure 1: The distribution of birth dates over the days of a calendar year

Livi (1929), Maroi (1954) and Breschi, Giovanna, and Gabriele (2018) document this phenomenon using vital statistics and show that it is due to false birth dates declarations so that it can be classified as a form of rule breaking.[8] It is remarkable that the evidence collected by these authors using official birth registries matches closely our census-based evidence, which refers only to subjects who survived until 1991. This comparison is performed in Figure 2 and allows us to show that the spikes in Figure 1 are not due to differential mortality rates for subjects born immediately before or after January 1.

In Figure 2, we use the methodology described in the Online Appendix Section OA2 to obtain the share Π_{glt} of birth dates that can be considered as falsified in a population group g of a locality l at a time t from the information in Figure 1.[9]

The first ten light-grey bars on the left of Figure 2 plot these shares observed in the Italian provinces for which Livi (1929) reports daily vital statistics for the months of December 1924 and January 1925. The remaining two dark-grey bars on the right of the figure display the same statistics for the entire country, constructed using the data of Maroi (1954) for the days around New Year's Eve of 1950 and 1951. The black bars measure instead the analogous shares Π_{glt} of falsified birth dates that we

[8]Table OA3 in the Online Appendix shows that birthday cheating correlates well with more traditional rule-breaking indicators like cheating in school exams, excessive absenteeism and property tax evasion.

[9]We assume that births should naturally be distributed according to a uniform distribution around January first. We thus consider the excess number of recorded births in the first five days of January relative to the ones predicted by the uniform distribution as the result of false date declarations.

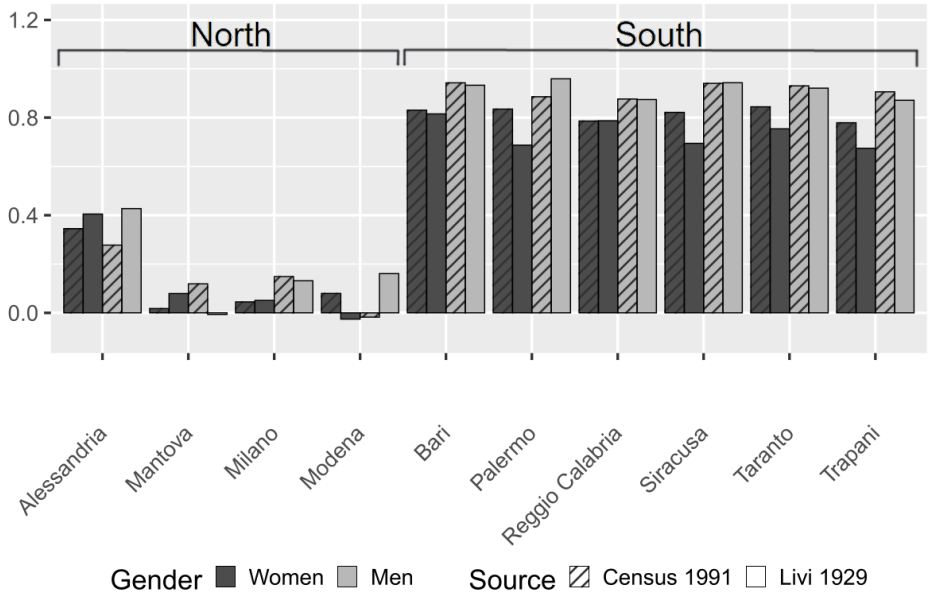

Figure 2: Birthday cheating around January 1925, January 1950 and January 1951 measured with different data sources

computed using the 1991 Census. In each geographical unit the black and grey bars have about the same height, as one would expect in the absence of differences in mortality between cheaters and non-cheaters.[10] Interestingly, Π_{glt} is substantially lower in the four northern provinces (Modena, Mantova, Milano and Alessandria) than in the remaining six southern ones. Therefore, the South-North difference in birthday cheating emerges similarly from historical vital statistics and recent census data.

The map in Figure 3 is based on the 1991 Census data and shows that birthday cheating is markedly more frequent in the localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies", whose historical borders are indicated in the map with the black thick line. What is particularly striking is that around the northern border of this kingdom there is a sharp discontinuity in the frequency of false birth dates, which are almost absent in municipalities located just north of the border. Interestingly, this border does not correspond to the official boundaries of some modern regional administrations (specifically, Lazio, Umbria and Marche, as shown by D'Adda

[10]Considering the greater incentive to falsify the birth date for male children to delay military service, in Figure OA3 of the Online Appendix we replicate the same exercise separately by gender for the provinces considered by Livi (1929).

and De Blasio, 2017).

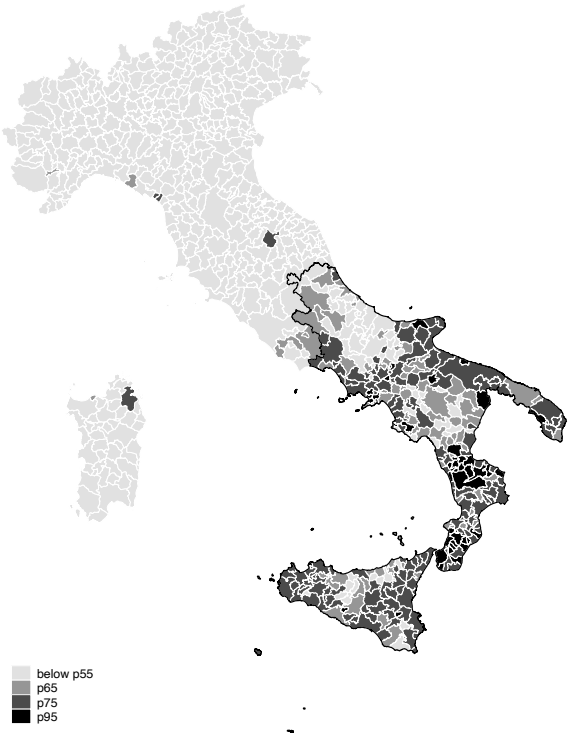

Figure 3: Birth date cheating and the "Kingdom of the two Sicilies"

So birthday cheating appear to be related more to institutions of the past than of the present, and specifically to state authorities known to be characterized by a less efficient administration and by lower levels of deterrence against rule breaking.[11] For example, birthday cheating is almost absent in the insular region of Sardinia that is usually included in the standard definition of "South" of Italy, but that was historically part of the northern Kingdom of Piedmont and Sardinia, ruled by the Savoy dynasty. Historians typically credit this kingdom with an efficient administration and with high levels of deterrence against crime.[12] For these reasons, in Figure 1 and in the rest of the paper, the "South" is defined as the set of localities that historically were part of the

[11]See for example Putnam, Leonardi, and Nanetti (1993), Di Liberto and Sideri (2015) and Bosker et al. (2008).

[12]See again Putnam, Leonardi, and Nanetti (1993). The Savoy unified Italy under their power in the second half of the 19th century.

"Kingdom of the two Sicilies".[13]

There is a well defined set of motives that may induce parents to register a false date of birth for their children, as suggested by Livi (1929), Maroi (1954) and Breschi, Giovanna, and Gabriele (2018). The likely most important one has to do with the fact that the activities of children typically take place within cohorts defined by the calendar year of birth. Therefore, a child born in December is always among the youngest in the groups of mates with whom she competes, physically or intellectually. If the same child is instead registered as being born in early January, she will be the oldest in her cohort. This is particularly relevant in the case of school activities, sport competitions and army enrollment, which was compulsory in Italy for cohorts born until 1985.[14]

Another relevant motive for shifting the birth of a child born in December to early January, is that it makes the child available at home for a longer time. If the child is male, military service will start one year later, while in the case of a female there will be more time to find a husband. As shown in the Online Appendix Figure OA4, birthday cheating is observed for both females and males, although it is more pronounced for the latter. It is also observed for children that later reach high levels of education (more than high school) as well as for those who instead do not go beyond compulsory education or are dropouts (Online Appendix Figure OA5). Therefore, birthday cheating does not seem to be specifically related to family affluence. This is not surprising because motives like being older in a cohort, delaying the military service or having more time to find a husband are largely unrelated to family affluence.

Independently of the motive, for the purpose of this study, a major advantage of birthday cheating as a measure of rule breaking is that it can be estimated, using Census information, for small population groups in given localities and at different points in time during the 20th century. As in Fisman and Miguel (2007), who compare parking violations of United Nations diplomats of different nationality in the city of New York

[13]More precisely, we follow D'Adda and De Blasio (2017) in defining "South" as the set of localities belonging to the modern regions of Sicily, Calabria, Apulia, Campania, Basilicata, Molise, Abruzzo and to the provinces of Frosinone, Latina, and Rieti in the region of Lazio, the province of Ascoli Piceno in the Marche region, and a few municipalities in the province of Perugia.

[14]Practices aimed at shifting the activities of children to later born cohorts take the name of "Red Shirting" in the USA.

in two periods characterized by different levels of deterrence, we can compare birthday cheating of different groups in the Italian population who live in the same city and at the same time and hence arguably face the same deterrence and derive similar benefits from this form of rule breaking. Therefore, conditioning on a locality and a time period, differences in birthday cheating may reflect differences in the proclivity to cheat in general and thus in the honesty of the population groups that we compare. Moreover, we can do it using Census data for an entire country covering about a century.[15]

There are, in particular, two groups of subjects that can be interestingly studied in such a context: migrants between localities and remainers in each locality. The reason why these two groups are interesting is that rule breaking has been shown to be unequally distributed across nearby geographic areas in different parts of the world and in Italy in particular.[16] Comparing migrants and remainers in terms of their honesty, as revealed by differences in birthday cheating for given deterrence and cheating benefits, allows us to explore the hypothesis that migration movements explain at least part of this heterogeneity across localities.

There are many reasons why migrants and remainers may differ in terms of honesty, and we cannot tell them apart with our data. For example cheaters in a locality could obtain material advantages that put them in better economic conditions, thus reducing their need to migrate. Honest citizens, instead, may pay a prize for respecting rules, which in the long run constitute a push factor for migration. In the specific case of birthday cheating, a child registered as being born in January instead of December will start school later becoming one of the oldest in his/her cohort. Campaniello, Aparicio Fenoll, and Monzon (2022), among others, have shown that children who start school later perform better in terms of future test scores. If these educational achievements translate into higher earnings the probability of migration is likely to decline. Alternatively, Michaeli et al. (2022) has suggested that members of a community who, ceteris paribus, dislike breaking rules may prefer to move elsewhere if they are continuously free-rided by too many peers who instead do not care about rules.[17]

[15] According to Breschi, Giovanna, and Gabriele (2018) birthday cheating is common also to other countries.

[16] See the references in footnote 4.

[17] At the same time, subjects who are dishonest may prefer to leave a community if it becomes poor

Whether migration generates a drain or a gain of honesty for a given locality is therefore an important empirical question that we address in Section 3. Answering this question is also important because honesty drains or gains may have, in the long run, relevant economic consequences on the affected localities, as we show in Section 4.

Before proceeding with the analysis, however, we must highlight one caveat of our measure of rule breaking. Cheating in the registration of the date of birth of a newborn child is a decision of the parents, while the decision to later migrate may be a decision of parents, if the entire family move, or just a decision of the child when she becomes an adult and decides to migrate alone. This is because we observe the place of birth and the place of residence in 1991, but we have no information on when migration took place. However, given the extensive evidence of inter-generational transmission of ethical values,[18] an agent in our analysis should be thought of as a family, which may or may not cheat and may or may not migrate at different points in time. For brevity, we will use occasionally the words "migrants" and "remainers" to refer, respectively, to migrant and remainer families.

3 Geographic sorting based on honesty

Starting from census data for the entire Italian population, our observations are groups of migrants (from South to North or viceversa) and remainers (in the corresponding macro-region), born in late December or early January in a narrowly defined locality (a Local Labor Market as defined by Istat) during one of the 17 bienniums of the 1921-1954 period.[19] As explained in the Online Appendix Section OA3, we restrict the analysis to localities and bienniums in which a minimum number of migrants and remainers are born (at least 6 of each type). This constraint is particularly binding for the North

because of excessive free-riding. In the model of Michaeli et al. (2022), the prevailing outcome depends on the risk attitudes and on the beliefs about deterrence in the place of origin vs. the place of destination that characterize subjects with high or low honesty.

[18] A large literature on parenting styles suggests that this is the case. See, for example, Tabellini (2008), Algan and Cahuc (2010), Houser et al. (2016), Lowes et al. (2017) and Doepke and Zilibotti (2017).

[19] In 1954, a major reform of civil registries made doctors and obstetricians responsible for the registration of birth dates so that after this year the phenomenon rapidly disappears. See also Breschi, Giovanna, and Gabriele (2018).

because migration flows from North to South are rare and small in size. For the South we end up with 294 localities out of the 327 that are defined by Istat; for these 294 localities we have the minimum number of migrants and remainers in 6,432 cells defined by locality×biennium×migration-status out of the $294 \times 17 \times 2 = 9,996$ cells that are theoretically possible. The corresponding figures for the North are 17 out 454 and 206 out of 578. Figure 4 shows the location, along the Italian peninsula, of these southern and northern localities.[20]

Figure 4: Local labor markets with enough emigration for our analysis

If within the same locality and biennium migrants and remainers face the same deterrence and have the same cheating benefits, the probability of birthdate cheating in the two groups is informative about their respective honesty. Figure 5 reports preliminary descriptive evidence about migration sorting based on honesty. Aggregating data

[20]Black areas indicate local labor markets with a minimum number of migrants and remainers (at least 6 of each type). White areas refer to local labor markets with insufficient emigration and are therefore excluded from the analysis. The thick white solid line indicates the border of the "Kingdom of the two Sicilies".

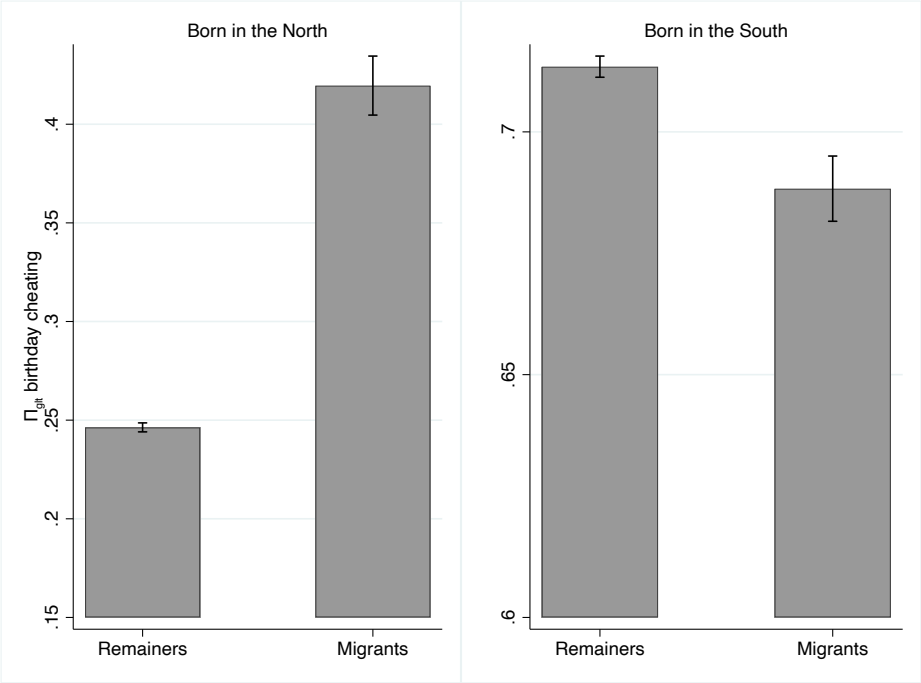

Figure 5: Descriptive evidence on geographic sorting based on honesty

over the entire 1921-1954 period, the left panel compares unconditional probability of cheating of the families of migrants and remainers born in the 17 northern localities in which this comparison is possible. The share of cheaters is 42% among migrants from North to South while it is 17 percentage points lower among remainers in the North and the difference is statistically significant. On the contrary, the right panel shows that in the 294 southern localities for which the comparison is possible, the share of cheaters is higher among remainers in the South than among migrants from South to North. The difference is relatively small (2.51 percentage points) but statistically significant. This descriptive evidence suggests that migrants from South to North and viceversa were not randomly selected from the respective populations with respect to honesty.

Given the large migration rates from South to North (29% on average in our dataset), this non-random selection may have induced a drain of families with high honesty out of the South even if the average difference between the cheating probability of migrant and remainer families is small in this macro-region. We will quantify precisely the size of this drain in Section 4. In the opposite direction, migration rates were significantly

less intense (7% on average in our dataset), but in the few localities in which there was some migration, the cheating probability of migrant families is substantially higher than that of remainer families, suggesting the possibility of a localized loss of low honesty families from North to South in specific geographical contexts.[21]

Table 1: JBD cheating of Migrant and Remainer families

	(1)	(2)	(3)	(4)
Migrant∗South (β_4)	-0.024***	-0.013***	-0.015***	-0.015***
(Migrants South-Remainers South)	(0.009)	(0.004)	(0.004)	(0.005)
Migrant∗North (β_3)	0.181**	0.071***	0.058**	0.104***
(Migrants North-Remainers North)	(0.074)	(0.023)	(0.026)	(0.030)
South (β_2)	0.466***			
(Remainers South)	(0.063)			
β_1	0.248***			
(Remainers North)	(0.062)			
Observations	6,638	6,638	6,638	6,638
Number of SLL	311	311	311	311
Population represented by cells	354817	354817	354817	354817
R-squared	0.411	0.921	0.923	0.923
LLM x Biennium FE	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Controls interacted	No	No	No	Yes
Oster δ for $\beta_4 = 0$		9.642	15.13	13.25
Oster δ for $\beta_3 = 0$		3.750	2.807	4.007
p-value of F-test for controls			0	0

Notes: The table reports OLS estimates based on data for 294 Local Labor Markets (LLM) in the South of Italy and 17 in the North, observed for at most 17 bienniums between 1921 and 1954 (*** p<0.01, ** p<0.05, * p<0.1). The Population represented by cells is the total number of births observed in the ten days around January first in the localities and during the period covered by our analysis. The p-value in the last row refers to an F-test for the joint significance of all included controls.

To dig more deeply into the sorting pattern displayed in Figure 5, Table 1 reports controlled estimates of the shares of cheating families in the four groups defined by migration status (migrant or remainer) and macro-region of birth (North or South). An observation in this analysis is a cell of migrants or remainers (g), in a locality (l) and a

[21]It is possible that the North–to–South migrants are return migrants, but we do not have the possibility to ascertain this hypothesis in our data

biennium (t), weighted by the population that the cell represents (namely the number of births around New Year's Eve), to give more weight to cells in which our measure of cheating is more precise.[22] The estimates of Table 1 are obtained with the following regression:

$$\Pi_{glt} = \beta_1 + \beta_2 South_{glt} + \beta_3 Migrant_{glt} * North_{glt} + \beta_4 Migrant_{glt} * South_{glt} + \gamma X_{glt} + \psi_{lt} + \epsilon_{glt}, \quad (1)$$

that we estimate on the sample of the 6,638 group×locality×biennium cells (6,432 in the South and 206 in the North) that we can use for the analysis, as explained above. Standard errors are clustered at the locality level.[23] ψ_{lt} is the interaction between locality and biennium fixed effects.[24] X_{glt} are average characteristics of migrants or remainers in these cells (share of females, average year of birth, share with primary education and share with tertiary education)[25] and ϵ_{glt} contains unobservable characteristics of groups, localities and periods. Note that, based on the discussion in Section 2, gender, education and year of birth are likely to capture the most important determinants of the benefits of birthday cheating, while LLM×biennium fixed effects should control almost perfectly for the level of deterrence.

As a benchmark, in column 1 no control is included.[26] Among remainers in the North the share of cheaters is 24.8% while it is 46.6 percentage points higher among remainers in the South. Migrants born in the North add 18.1 percentage points of cheating probability to the level of remainers born in the same region, while migrants

[22] These weights are hence also a proxy for the population size in each locality.

[23] This specification is equivalent to running the regression on individual data with clustering at the cell level and attributing to each subject the birthday cheating of its cell (Angrist and Pischke, 2009). If the regression had been run at the individual level, the number of observations would have been 354,817, which corresponds to the total Italian population in the 1991 census who was born around New Year's eve in the localities we consider and in the 1921-1954 period.

[24] When this interaction is included in equation 1 the β_1 and β_2 are of course not identified.

[25] Secondary education, corresponding to a junior-high (8 years) or a high school degree (13 years), is the omitted category. Primary education corresponds to 5 years of elementary schools. Tertiary education corresponds to more than high shool.

[26] Therefore, like in Figure 5, these results are still not informative about underlying differences in honesty because deterrence and the distribution of benefits are not controlled for in this specification, which is reported to be compared below with the remaining columns. Note also that these results are not numerically equal to those of Figure 5, because in the figure observations are aggregated over the entire 1921-1954 period, while in the table they are dis-aggregated by LLM×biennium cells that are weighted by number of births.

born in the South have a lower probability of cheating (by 2.4 percentage points) with respect to remainers in the same regions. All these differences are statistically significant at conventional values.

In columns 2-4 of Table 1 we progressively add more stringent controls which are meant to purge the comparison of migrant and remainer families from confounders related to differences in benefits and deterrence.[27] Specifically, in column 2 we include the interaction between LLM and biennium fixed effects (ψ_{lt}), which is possible given that observations are weighted by the number of subjects in each cell; column 3 adds the X_{glt} controls linearly; and column 4 adds a fully saturated specification of the same X_{glt} controls.[28] Even in these more demanding comparisons, migrant families from North to South have a cheating probability that is 6–10 percentage points higher than that of remainer families in the North, while for migrants in the opposite direction the analogous probability is 1.3–1.5 percentage points lower than the one of remainers in the South.

These controlled estimates give further support to the conclusion that sorting based on honesty has occurred in Italy. Depending on the intensity of migration flows out of the different LLMs, this sorting may have induced, locally, a drain of honest families from South to North and a drain of low-honesty families in the opposite direction.

This conclusion rests, however, on the identifying assumption that, within a given LLM×biennium cell and controlling for what we can observe, the distribution of benefits and the level of deterrence are similar for migrants and remainers. The δ statistics proposed by Oster (2019) and reported in the table for the parameters β_3 and β_4 provide evidence in favor of this assumption. To interpret this parameter in our context, note that comparing the last 3 columns of the table with the first column, the R^2 indicates that the locality and biennium fixed effects together with the observed characteristics of migrants and remainers in each LLM×biennium cell explain more than 50 additional percentage points of the variability of the probability of cheating, on top of the 41% explained by the uncontrolled specification in the first column. Therefore, these controls must capture a good part of the variability of deterrence and cheating benefits. However,

[27] The β_1 and β_2 coefficients in columns 2,3 and 4 are not reported because they do not have a meaningful interpretation given the inclusion of the interaction between LLM and biennium fixed effects.

[28] In Online Appendix Table OA4 we report coefficients for all controls

the estimated coefficients β_3 and β_4, that indicate the existence of sorting related to honesty, remain relatively stable when these controls are included.

In light of this evidence, the δ statistic proposed by Oster (2019) measures by how many times the remaining characteristics of localities that we do not observe (namely the unobservable determinants of deterrence and benefits) should be correlated with migration status in each macro-region of birth in order to bring down to zero the β_3 and β_4 coefficients, given that these unobservables can only explain the small remaining variability of the cheating probability. For example, with reference to the comparison between columns 4 and 1, a $\delta = 13.25$ for β_4 says that if the unobservable characteristics (which explain less than 8% of the variability of the outcome) could be included, they would have to be about 13.25 times more correlated with migration status than the observed ones in order to conclude that migrants and remainers born in the South have the same cheating probability. As for β_3, the analogous correlation would have to be 4 times higher to conclude that there has been no drain of low honesty families out of the North. Such high correlations between unobservable characteristics and migration status are arguably implausible, in a context in which they explain only a relatively small part of the variability of the outcome.

4 Longterm consequences of a honesty drain

Having established that migrants between the South and the North of Italy are non-randomly selected with respect to their honesty, our next goal is to measure the honesty drain (or gain) that localities have experienced because of these internal migration movements. Such measure reflects not only the difference in honesty of migrant and remainer families, but also the size of migration flows. A small outflow of very different migrants as well as a big outflow of randomly selected migrants would obviously not generate a relevant drain of honesty. We also want to assess to what extent this phenomenon is heterogeneous across localities and whether it is quantitatively large enough to possibly have detectable economic consequences.

To this end, we restrict the analysis to the 294 LLMs of the South, where emigration

to the North has significantly characterized the entire macro region.[29] Moreover, we abstract from time differences along the 20th century and for each LLM we collapse bienniums to a single period from 1921 to 1954 (and thus we omit the time subscript t in the remaining part of the paper); we keep the first biennium (1921-22) to measure baseline characteristics of each locality and we use recent years (after 1954) to measure outcomes.

4.1 A measure of honesty drain or gain

Consider a locality l of the South and two overlapping sets of agents: the set of all those who are born in l, denoted with b, and its subset containing those who are born and also remain in l (the remainers considered so far in the paper and denoted by r). The complement of the set r in the b set are migrants to the North, denoted by m. Consider the quantity

$$\theta_l = \Pi_{rl} - \Pi_{bl} \tag{2}$$

which measures the difference in the probability to falsify a birth date of remainers in l versus born in l. If $\theta_l = 0$, obviously the cheating probability of remainer and migrant families must be identical and both groups are random samples of the population of families giving birth in l with respect to birthday cheating. Therefore, even in the presence of a large migration outflow there would be no drain or gain of honesty in this case. If instead $\theta_l > 0$, it must be the case that remainer families cheat more frequently than migrant families and therefore more frequently than the average family of children born in l. In this case, θ_l captures the honesty drain suffered by locality l because it measures how average honesty has declined in the remaining population as a result of the South–North emigration process that took place between 1921 and 1954. Viceversa, $\theta_l < 0$ indicates that locality l experienced a gain of honesty for the opposite reason.

Note that θ_l does not consider other types of migration in and out of a locality l. The focus on South to North migration is justified by three empirical observations. First, until the late '70s of the past century, immigration from abroad was essentially absent

[29] For completeness, we report in the Online Appendix Table OA5 also results that include the 17 localities of the North for which a drain can be computed. Results are similar, both in terms of point estimates and significance.

in Italy. Second, migration flows from North to South were so small in the period to be practically irrelevant. Third, migration within the South was practically irrelevant as well in the same period.[30][31]

θ_l	
Mean	.003
Std. Dev.	.017
Min	-.214
10^{th} percentile	-.012
Median	.0037
90^{th} percentile	.0194
Max	.22
Skewness	-.252
Share with $\theta_l < 0$	.452

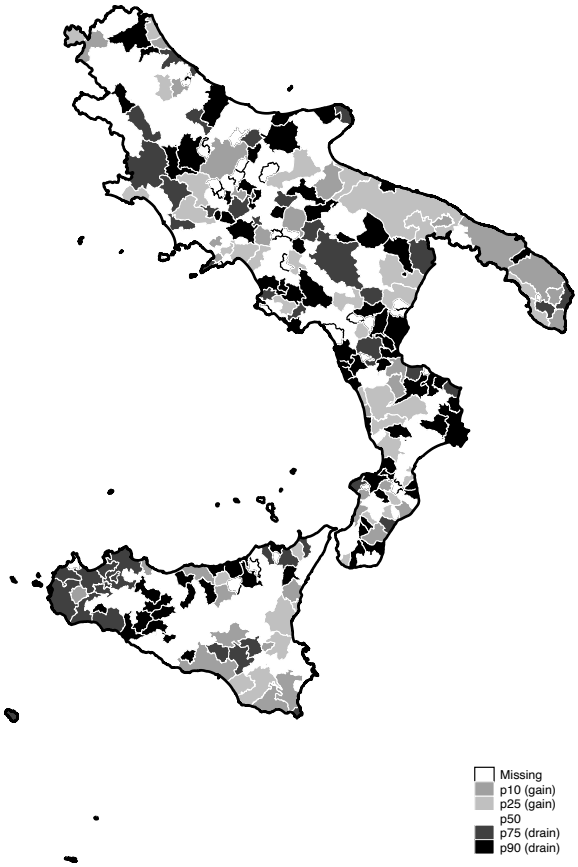

Figure 6: Honesty drain or gain across LLMs during the 20th century

Figure 6 shows the honesty drain θ_l for the 294 southern LLMs for which we can compute it. The coloring type and intensity indicate whether a locality experiences a honesty drain ($\theta_l > 0$) in dark grey or a gain ($\theta_l < 0$) in light grey. The table to the left of the figure provides the correspondent descriptive statistics, weighted by the size of each locality (total number of birth around New Year's Eve between 1921 and 1954).

[30]For the first observation, see Del Boca and Venturini (2005); for the second, see Online Appendix Section OA3; for the third, see Bonifazi (2009) who shows that South–North migration rates are more than four times larger than within South rates. Using Istat migration matrices (Annali di Statistica Serie VIII - Vol.17, pag. 684, table 11.VII), South–North migration rates are 3.4 times larger.

[31]This period was characterized also by emigration to other countries. However, our census data does not contain information on individuals that by 1991 were living abroad.

The figure shows that θ_l is highly heterogeneous across southern LLMs. The support of the distribution of θ_l ranges between -21.4 and +22.2 percentage points, with the 10th and 90th percentiles respectively equal to -1.2 and +1.9 percentage points. On average, the drain θ_l is small in size (the mean is 0.3 percentage points) but its variability across localities is substantial and our goal is to exploit it in order to understand whether it may have had relevant consequences.

To this end, we estimate variants of this equation:

$$Y_l = a + b\ \theta_l + c\ \Pi_{l21} + g\ X_l + \psi_l + \epsilon_l, \tag{3}$$

where: Y_l is an outcome for locality l, θ_l is the standardized honesty drain measured across all bienniums between 1921 and 1954,[32] X_l is a set of locality controls and ψ_l is a set of fixed effects for the seven current administrative regions that partition the South. Ideally we would like to regress *the change* in the outcomes Y_l on the honesty drain θ_l, but information on these outcomes for 1921 is not available, except for the case of the educational outcome described below. We address this limitation by including instead the probability of birthday cheating in l measured in the biennium 1921-22 (Π_{l21})[33] to capture differences in initial birthday cheating conditions.[34] Similarly to the regressions considered in previous sections, also here we weight observations by the population size that they represent (the total number of births in each locality during the 1921-1954 period).

The parameter of main interest is b, which in general does not have a causal interpretation and offers only a suggestive controlled correlation. However, using again Oster (2019), we can determine whether it is plausible that the unobservable confounders for which we cannot control are sufficiently correlated with θ_l, relative to how the observable controls X_l are, to bring down to zero the estimate of b in equation (3).

[32] One standard deviation of θ_l corresponds to 1.7 percentage points.

[33] The coefficient for this cheating initial condition is presented in Online Appendix Table OA6 together with those for all of the included controls.

[34] Since $\theta_l = \pi_{rl} - \pi_{bl}$, a more flexible specification would allow to estimate different coefficients for π_r and π_b. As discussed below, when we estimate this more flexible specification, we reject the hypothesis that the two coefficients of π_r and π_b differ.

4.2 Economic outcomes

Table 2 reports OLS estimates of equation (3), based on data for Local Labor Markets (LLMs) in the South of Italy. In Columns (1) and (2) the outcome Y_l is a measure of human capital: the log of the average math test score of high-school students of a locality l in the standardized national exams, averaged over the 2012-2021 period.[35]

In the first column, in which no control is included, a one standard deviation increase of honesty drain reduces the math test scores by 2.5% and this estimate is significantly different from zero at conventional values. In the second column all controls are included and the estimated coefficient is slightly larger in absolute size and significance. The controls are: the probability of birthday cheating for the same locality in the cohort born in the 1921–22 biennium, to control for initial conditions of honesty; a set of predetermined labor force and geographic characteristics of localities,[36] and fixed effects for the 7 modern administrative regions corresponding to our definition of South. To exclude the possibility that our measure of honesty drain captures some form of brain drain, we include also a variable constructed with the same methodology of Section 4.1 but using, in place of the probability of cheating, the share of secondary and tertiary graduates among remainers versus born in a locality. A proxy for initial condition of this outcome is included as well, measured by the share of illiterates from the 1921 Census. As already mentioned we cannot do the same for the other outcomes, but it is important to observe that our conclusions are qualitatively similar with or without controlling for the initial condition of this outcome.[37]

The stability of the honesty drain coefficient across columns 1 and 2 of Table 2 is remarkable given the increase of the R^2 from 0.022 to 0.338. As a result of this stability, the estimates of the Oster (2019) δ parameter reported at the bottom of column 2

[35]The test is implemented by Invalsi, the Italian Agency for the evaluation of the school system (see: Falzetti, 2021). High Schools for which the average test can be constructed exist in only 245 LLMs of the South out of the 294 in our analytical sample. Estimates based on the literacy test score in the Invalsi dataset are qualitatively similar to those reported here for the math test score.

[36]Employment rate, share of employment in agriculture, share of employment in manufacturing from the 1936 census (service omitted); total population in the LLM and population density from the 1921 census; dummies for coastal land, low lands, low mountains, high mountains, flood risk, rock slide risk.

[37]The Online Appendix Table OA6 reports the complete table with coefficients and standard errors of all the covariates.

support the claim that the characteristics of localities that we do not observe would not bring down to zero the coefficient of honesty drain if they were observed.

Table 2: Honesty drain and economic outcomes

	Math Score Invalsi		Firm Value added		VCR Dec 2016		Earnings Growth	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Honesty Drain (standardized)	-0.025** (0.010)	-0.026*** (0.009)	-0.029*** (0.010)	-0.025*** (0.009)	-0.064** (0.026)	-0.045** (0.023)	-0.002** (0.001)	-0.002** (0.001)
Observations	245	245	186,507	186,507	294	294	294	294
R-squared	0.022	0.338	0.002	0.281	0.017	0.332	0.017	0.529
Controls:								
Region FE	No	Yes	No	Yes	No	Yes	No	Yes
Initial Period	No	Yes	No	Yes	No	Yes	No	Yes
Empl. and Geo	No	Yes	No	Yes	No	Yes	No	Yes
Drain mean	0.003	0.003	0.004	0.004	0.003	0.003	0.003	0.003
Drain S.D.	0.015	0.015	0.053	0.053	0.017	0.017	0.017	0.017
Outcome Mean	181.638	181.638	22.954	22.954	186.439	186.439	0.018	0.018
Outcome S.D.	9.088	9.088	14.791	14.791	31.404	31.404	0.006	0.006
Oster δ		-806.1		-16.78		6.465		7.783

Notes: In columns (1) and (2) the dependent variable is the log of the Invalsi test-score in Math measured in grade 13 (high-school). This variable is only available for 245 LLMs in the South. Observations are weighted by the number of births in each locality l. Standard errors are robust for heteroskedasticity. In columns (3) and (4) the dependent variable is the logarithm of firm value added per employee. Observations are weighted by the employment share of each firm within a locality l. Standard errors are clustered at the LLM level. In columns (5) and (6) the dependent variable is the logarithm of the average vote counting rate per hour in a locality for the 2016 constitutional referendum. In columns (7) and (8) the dependent variable is the log of per capita yearly labor earnings growth in a locality. Standard errors are, again, robust for heteroskedasticity. The honesty drain (θ_l) is the difference in the probability of birthday cheating of remainers in l versus born in l. θ_l is standardized. (*** p<0.01, ** p<0.05, * p<0.1).

The second outcome that we consider is instead a measure of labor productivity averaged over the 2009-2018 period, and the correspondent estimates are reported in columns 3 and 4 of Table 2. Specifically, this measure is a conventional indicator of firm-level value added constructed with Bureau van Dijk data, which cover all firms required to register their balance sheets.[38] Since in the case of this outcome an observation is a

[38]In the construction of this indicator we follow Ilzetzki and Simonelli (2017). The data source is the ORBIS database from Bureau van Dijk, which provides accounting books information for the universe of companies who are required to register their balance sheets at the Chambers of Commerce (that are all companies, except sole proprietorship enterprises and partnerships). After selecting the subsample of firms with non-missing employment, value added and capital information, this data set covers 656,518 firms that correspond to roughly 40% of all firms in Italy. Labor productivity is calculated as the value added per employee, averaged over the 2009-2018 period and measured in 2019 Euros.

firm, denoted by f, we estimate the following modified version of equation (3):

$$Y_{lf} = a + b\ \theta_l + c\ \Pi_{l21} + g\ X_{lf} + \psi_l + \epsilon_{lf}, \tag{4}$$

In this specification, as in Ilzetzki and Simonelli (2017), we cluster standard errors at the locality level and we weight observations by the employment share of each firm within locality l. In this way we give more weight to larger firms and obtain estimates that are informative at the locality level.

Also for labor productivity we estimate a substantial loss of at least 2.9% induced by one additional standard deviation of the honesty drain in column 3, where no control is included. This loss declines only slightly in absolute value (2.5% in column 4) when we include the same controls of column 2 as well as industry fixed effects, the log of physical capital per employee and a measure of human capital in the LLM. Following Ilzetzki and Simonelli (2017), these last two controls are meant to isolate the effect of honesty drain on labor productivity. To give a sense of the economic relevance of these estimates, the overall North-South labor productivity gap (which is equal to 34%) would decrease by about 7.4% in association with a decrease of one standard deviation of the honesty drain. Also in the case of columns 3 and 4 the stability of the estimates of the honesty drain coefficient is remarkable, given the increase of the R^2 from 0.002 to 0.281 and the estimates of the Oster (2019) δ parameter reach a high absolute value.

Next we consider a less conventional but novel and informative measure of pure labor productivity proposed by Ilzetzki and Simonelli (2017): the per hour vote counting rate (VCR) in an election.[39] These authors observe that ballot counting in Italian elections is a labor intensive task that must be performed in the same way over the entire country and that does not require any substantial piece of physical capital to be performed. The technology and the tools used by vote counters are also the same in all electoral polling stations and the citizens who perform this job receive a monetary compensation that is independent of the time spent to complete the job. Moreover, they are allowed to take a paid leave from their employers for all the time that is needed to complete the votes' count, so that also the opportunity cost is controlled for. According to Ilzetzki

[39]We are very grateful to Ethan Ilzetzki and Saverio Simonelli for sharing their data with us.

and Simonelli (2017) this measure of "pure" labor productivity accounts for about half of the North-South gap in the firm-level value added constructed with the Bureau van Dijk data that constitute our second economic outcome described above (see footnote 38).

While in political or administrative elections ballots may differ across localities because of the number of competing candidates, this problem does not exist for national referenda, in which only Yes/No ballots need to be counted. For this reason we use the VCR measure that Ilzetzki and Simonelli (2017) provide for the 2016 Italian referendum on the constitutional amendments proposed by the Renzi government.[40]

For the purpose of our analysis, the VCR is an informative indicator not only because it is a "pure" measure or labor productivity, but also because it is likely to partly capture reciprocal trust. Consider two persons with opposite preferences about the referendum outcome, who have to count votes in a context in which both of them are sure that the other respects rules and does not cheat. In this case, they could speed up vote counting by splitting the ballots between them without any double or joint checking. If instead they do not trust each other, they will want to double check and jointly assess any ballot. This extreme example makes it clear why we can expect to observe a negative correlation between the honesty drain θ_l and the VCR in the same locality, for given initial conditions.

The estimates for this less conventional but interesting outcome are reported in columns 5 and 6 of Table 2. A one standard deviation increase of the honesty drain reduces the VCR by 6.4% when no controls are included (column 5), and the effect is statistically significant. When all the controls are included (column 6), the estimated loss caused by the honesty drain declines slightly in absolute size, to 4.5%, but remains statistically significant. The Oster (2019) δ parameter reaches the reassuring value of 6.5. Finally, in columns 7 and 8 of Table 2 we report estimates in which the outcome is a measure of yearly labor earnings growth based on tax returns data.[41] For each

[40]This referendum attracted a lot of attention and raised a very intense debate in the country, which generated a very large turnout (65.5%). It ended up becoming a referendum in favor or against Matteo Renzi himself, who in fact lost the battle and his quick political decline started immediately after. Vote counting in this referendum was perceived as extremely important by all citizens and political parties.

[41]We obtained this information from the website of the Italian Ministry of Finance.

locality we have information on total labor earnings and on the number of workers with positive earnings. We can thus construct a measure of per capita yearly earning growth in a locality, averaged over the period 2005-2019. A one standard deviation increase of honesty drain reduces by 0.2% the average earnings growth, and this estimate remains unchanged between the two columns, independently of whether controls are included or not. Once again, the Oster (2019) δ parameter reaches a very high level.

For all the outcomes discussed so far, in Online Appendix Table OA7 we test a more flexible specification in which we separately include cheating for the remainers in a locality π_{rl} and cheating of all those born in it, π_{bl}, instead of $\theta_l = \pi_{rl} - \pi_{bl}$. This specification allows the coefficients of π_{rl} and π_{bl} to be different and thus to capture the effect of the post-migration level of cheating separately from the effect of the pre-migration level. Results in the Appendix Table OA7 show that we cannot reject the null hypothesis that the two coefficients of π_{rl} and π_{bl} have the same value and that this value is the coefficient b of θ_l in our baseline specification (3).

Although none of these estimates can be considered as causal, they are jointly compatible with the possibility that the more severe honesty drain experienced by some localities during the 20th century has induced lower levels of human capital, labor productivity and wage growth in recent times.

4.3 Quality of politicians and of political outcomes

Depending on how close an electoral system is to being proportional, politicians elected with that system should, to some extent, reflect the characteristics of the population that elects them. It is then natural to wonder whether, among localities with similar initial conditions, those that experienced a more severe honesty drain between 1921 and 1954, and thus had a less honest electorate of remainers after this period, also elected less honest local politicians and for this reasons suffered worse political outcomes. An advantage of measuring honesty with birthday cheating Π_{gl} within a locality is that it can be measured in the same way for both the electorate and the elected politicians.[42]

Columns 1 and 2 of Table 3 report estimates of equation (3) in which the outcome

[42]We are grateful to an anonymous referee for alerting us on the possibility of this kind of analysis.

is birthday cheating Π_{rl} of the electorate of remainers in 282 LLM's of the South.[43]. As expected, given how we defined the honesty drain in Section 4.1, independently of whether our usual set of controls is excluded (column 1) or included (column 2), one standard deviation of honesty drain increases by about 35 percent of a standard deviation the birthday cheating of the electorate of remainers, and the coefficient is statistically significant.

The estimate in column 3 of Table 3 is more surprising. In this specification, the outcome is the standardized birthday cheating Π_{pl} of local politicians ($g = p$) born before 1954 and elected between 2010 and 2019 in the 282 southern localities in which this analysis is possible.[44] We employ administrative data on all Italian local government officials over the 2010-2019 period, provided by the Italian Ministry of Interiors and containing their exact date of birth. These local politicians are mayors ("Sindaci"), members of local councils ("Consiglieri"), and members of the local executive governing bodies, ("Giunta", the municipality board composed by the deputy-mayor and executive members called "Assessori"). This outcome is regressed on birthday cheating among remainers, Π_{rl}, without controls.

It is remarkable that the estimated coefficient in this regression is very close to one (0.951), and statistically significant, as one would expect in an electoral system that were almost perfectly proportional. This is, however, not the electoral system that prevails in Italian local elections, where the coalition receiving more votes obtains a majority premium in the elected council and determines who are the mayor and the members of the executive governing body. The inclusion of controls in column 4 reduces the size of this coefficient to 0.687, without reducing its statistical significance. These estimates confirm that the average honesty of Italian local politicians reflects to a large extent the average honesty of their electorate, inasmuch as honesty is measured by birthday cheating.[45]

[43]We had to drop 12 southern localities because the number of elected politicians born around New Years Eve was not large enough to conduct the analysis.

[44]To facilitate the comparability of the estimates, the standardization is computed subtracting the average birthday cheating of politicians and then dividing by the standard deviation of birthday cheating among the population of remainers.

[45]In the Online Appendix Table OA8 we show the robustness of these results to the inclusion of the 17 LLM in Northern Italy. In Table OA9, we instead report the coefficients for all controls of this

Table 3: Honesty drain and political outcomes

	Birthday cheating Remaining population		Birthday cheating Politicians		City Councils Dismissal	
	(1)	(2)	(3)	(4)	(5)	(6)
Honesty Drain (standardized)	0.356***	0.350***				
	(0.119)	(0.068)				
B-day cheating (standardized)			0.951***	0.687***		
			(0.147)	(0.208)		
Politicians' B-day cheating (standardized)					12.760*	2.170***
					(7.559)	(0.689)
Observations	282	282	282	282	282	282
R-squared	0.029	0.748	0.238	0.372	0.108	0.967
Controls:						
Region FE	No	Yes	No	Yes	No	Yes
Initial Period	No	Yes	No	Yes	No	Yes
Employment and Geo	No	Yes	No	Yes	No	Yes
Drain mean	-0.026	-0.026	0.714	0.714	0.821	0.821
Drain S.D.	0.300	0.300	0.115	0.115	0.224	0.224
Outcome Mean	0.714	0.714	0.821	0.821	37.654	37.654
Outcome S.D.	0.115	0.115	0.224	0.224	47.277	47.277
Oster δ		-231.7		1.233		6.082

Notes: Observations are weighted by the number of births in each locality l and standard errors are robust for heteroskedasticity. In columns (1) and (2) the dependent variable is the standardized birthday cheating among remainers. In columns (3) and (4) the dependent variable is the standardized birthday cheating among local elected politicians in office between 1985 and 2019 and born between 1921 and 1954. The independent variable is the standardized birthday cheating among remainers (as in columns 1 and 2). In columns (5) and (6) the dependent variable is the number of city councils that were dismissed due to serious malfunctioning or corruption in a LLM over the 1990-2015 period. (*** p<0.01, ** p<0.05, * p<0.1)

Finally, in column 5 and 6 of Table 3 we show that the honesty of local politicians correlates with an important political outcome, that is the frequency of dismissals of city councils in a LLM, decided by the central government before their normal electoral term (Anelli and Peri, 2017). A city council may be dismissed if some form of severe malfunctioning has happened or if its members are put under investigation for corruption or crime, including involvement in some forms of organized crime, such as *Mafia*. The dependent variable in these columns is the total number of city council dismissals in the 282 southern LLMs that we consider over the 1990-2015 period. When controls are not included in column 5, an increase of birthday cheating of politicians equal to a standard deviation of the birthday cheating of remainers, increases the number of city council dismissals by almost 13, although this estimate is borderline significant. Note that in the average LLM there are about 7 cities observed for 25 years; the average number of council dismissals (weighted by the number of births in the LLM) is about 37. When the same controls of Table 2 are included in column 6, the estimated coefficient reduces to about 2, but becomes strongly significant.

As in the case of the economic outcomes examined in Section 4, also the estimates reported in this section cannot be considered as causal. But in all cases the Oster (2019) δ parameter reaches reassuring values, suggesting that unobservable confounders, if they could be included, should not distort excessively these estimates. All together, the results presented in this section support the plausibility of the hypothesis that in localities experiencing a more severe honesty drain during the 20th century, the honesty of recently elected politicians was lower and city local administrators performed less well.

specification.

5 Conclusion

We have studied the historical tendency of Italian parents to register a false date of birth for their children if they are born near the end of a year, shifting this date to early January. This phenomenon is well known to demographers and is a form of rule breaking.

With respect to other rule breaking indicators used in the literature, birthday cheating has the advantage that, using Census data, it can be constructed for migrants and remainers born in narrowly defined localities and at different points in time during the 20th century. Within our detailed space×time cells, the two groups are likely to share similar cheating benefits and deterrence, so that observed rule breaking in the two groups can be considered as an indicator of their average intrinsic honesty.

Based on this information, we show that migrants between Italian regions during the 20th century were non-randomly selected with respect to their intrinsic honesty. This is particularly true for migration flows from South to North that are significantly more sizeable than those in the opposite direction. As a result, some localities of the South experienced a reduction of the average honesty of their remaining population. We find a large heterogeneity in honesty drain across localities in the South, of which some gained and some lost honest families.

Finally, we measure the size of the honesty drain or gain experienced by each locality and we correlate this measure with a set of economic and political outcomes. Results are suggestive that a non-random selection of migrants based on honesty may have had important negative consequences in localities that experienced a more intense honesty drain.

We believe our results are of general relevance beyond Italy because they give a warning about the possibility that some localities may be caught in a vicious circle of increasing diffusion of cheating attitudes and emigration of families that are more averse to break rules. Finding ways to cut this vicious circle is crucial in some Italian localities and maybe elsewhere.

References

Algan, Yann and Pierre Cahuc. 2010. "Inherited trust and growth." *American Economic Review* 100:2060–2092.

Anelli, Massimo and Giovanni Peri. 2017. "Does emigration delay political change? Evidence from Italy during the great recession." *Economic Policy* 32 (91):551–596. URL https://doi.org/10.1093/epolic/eix006.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics. An Empiricist's companion.* Princeton University Press.

Bigoni, Maria, Stefania Bortolotti, Marco Casari, Diego Gambetta, and Francesca Pancotto. 2016. "Amoral Familism, Social Capital, or Trust? The Behavioural Foundations of the Italian North-South Divide." *Economic Journal* 126:1318–1341.

Bonifazi, Corrado. 2009. "Le migrazioni interne meridionali: vecchi e nuovi ritardi." *neodemos. it* .

Bosker, Maarten, Steven Brakman, Harry Garretsen, Herman De Jong, and Marc Schramm. 2008. "Ports, plagues and politics: explaining Italian city growth 1300–1861." *European Review of Economic History* 12 (1):97–131.

Braeseman, Fabian and Fabian Stephany. 2017. "United in Diversity? An Empirical Investigation on Europe's Regional Social Capital." *Agenda Austria Working Paper* 7.

Breschi, Marco, Gonano Giovanna, and Ruiu Gabriele. 2018. "Le nascite tra la fine e l'inizio dell'anno: dagli artifici del passato al presente artificiale." *Popolazione e storia* 19 (1). URL https://popolazioneestoria.it/article/view/884.

Buonanno, Paolo, Ruben Durante, Giovanni Prarolo, and Paolo Vanin. 2015. "Poor Institutions, Rich Mines: Resource Curse and the Origins of the Sicilian Mafia." *Economic Journal* 125:175–202.

Campaniello, Nadia, Ainoa Aparicio Fenoll, and Ignacio Monzon. 2022. "Parental Love is Not Blind: Identifying Selection into Early School Start." *Working paper* .

Cohn, Alain, Michel André Marechal, David Tannenbaum, and Christian Zund. 2019. "Civic Honesty around the Globe." *Science* .

D'Adda, Giovanna and Guido De Blasio. 2017. "Historical Legacy and Policy Effectiveness: the Long-Term Influence of Preunification Borders in Italy." *Journal of Regional Science* 57 (2):319–341.

Del Boca, Daniela and Alessandra Venturini. 2005. "Italian migration." *European migration: What do we know* :303–336.

Di Liberto, Adriana and Marco Sideri. 2015. "Past dominations, current institutions and the Italian regional economic performance." *European Journal of Political Economy* 38:12–41.

Doepke, Matthias and Fabrizio Zilibotti. 2017. "Parenting with style: altruism and paternalism in intergenerational preference transmission." *Econometrica* 85:1331–1371.

Durante, Ruben and Johannes C. Buggle. 2010. "Climatic risk, cooperation and the co-evolution of culture and institutions." *CEPR Discussion paper 12380* .

Falzetti, Patrizia. 2021. *INVALSI data: methodologies and results: III Seminar" INVALSI data: a reasearch tool"*. FrancoAngeli.

Fisman, Raymond and Edward Miguel. 2007. "Corruption, norms, and legal enforcement. Evidence from diplomatic parking tickets." *Journal of Political Economy* 115:1065–1091.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "The role of social capital in financial development." *American Economic Review* 94 (3):526–556.

Houser, Daniel, John A List, Marco Piovesan, Anya Samek, and Joachim Winter. 2016. "Dishonesty: From parents to children." *European Economic Review* 82:242–254.

Ichino, Andrea and Giovanni Maggi. 2000. "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm." *Quarterly Journal of Economics* 115 (3):1057–1090.

Ilzetzki, Ethan and Saverio Simonelli. 2017. "Measuring Productivity Dispersion: Lessons From Counting One-Hundred Million Ballots." CEPR Discussion Papers 12273, C.E.P.R. Discussion Papers. URL https://ideas.repec.org/p/cpr/ceprdp/12273.html.

Liccardo, Giovanni. 2019. *La smorfia napoletana: origine, storia e interpretazione*. Newton Compton Editori.

Livi, Livio. 1929. "Sulle false dichiarazioni della data di nascita per i nati alla fine dell'anno, e rettifica della distribuzione mensile delle nascite nel triennio 1923-1925." *Annali di Statistica* 3 (6):41–109.

Lowes, Sara, Nathan Nunn, James Robinson, and Jonathan Weigel. 2017. "The evolution of culture and institutions: evidence from the Kuba kingdom." *Econometrica* 85:1020–1048.

Maroi, Lanfranco. 1954. "La periodicità giornaliera dei nati e dei morti in Italia." *Statistica* XIV:412–423.

Michaeli, Moti, Marco Casari, Andrea Ichino, Ginevra Marandola, Maria De Paola, and Scoppa Vincenzo. 2022. "Civicness drain." *The Economic Journal* URL http://www.andreaichino.it/wp-content/uploads/2019/06/civicness_drain.pdf.

Oster, Emily. 2019. "Unobservable selection and coefficient stability: Theory and evidence." *Journal of Business & Economic Statistics* 37 (2):187–204.

Putnam, Robert D., Robert Leonardi, and Raffaella Y. Nanetti. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Rupasingha, Anil, Stephan J. Goets, and David Freshwater. 2006. "The production of social capital in US counties." *Journal of Socio-Economics* 35:83–101.

Tabellini, Guido. 2008. "The scope of cooperation: Values and incentives." *Quarterly Journal of Economics* 123 (3):905–950.