# Aversion to breaking rules and migration

Massimo Anelli[†] Tommaso Colussi[§] Andrea Ichino[‡]

## ONLINE APPENDIX

October 11, 2021

# Contents



# List of Figures



[†]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.

# List of Tables

# Online Appendix to Section 2

## Data description

**Italian Population Census.** For our empirical analysis, we employ micro-level data from the Italian Population and Housing Censuses. These data are restricted-use and have been accessed through the Laboratory for the Analysis of Elementary Data (ADELE) at the National Statistical Office (Istat), in compliance to the laws on the protection of statistical confidentiality and of personal data.[1]

The data include the universe of individuals residing in Italy in the year of the Census and provide information on a variety of social, demographic and economic characteristics; in particular, it contains information on the respondents' municipality and date of birth, municipality of residence, gender, and highest educational attainment. The information on the municipality and date of birth is only available from Census years 1991 and 1981, respectively; for this reason the main sample comes from the $13^{th}$ Population and Housing Census of 1991, which comprises 56,778,031 individuals residing in any of the 8,100 Italian municipalities.[2]

We combined this information with several municipality characteristics coming from historical censuses, including the resident population and the literacy rate in 1921, as well as the resident population and the employment sectoral distribution in 1936.[3] Moreover, we collected for each municipality a set of indicators for flood, and landslide risks along with geographical information, such as altitude and land area.[4] Based on the information provided by d'Adda and de Blasio (2017), we constructed a variable indicating if a municipality lay within the historical border of the "Kingdom of the Two Sicilies", which represents our definition of Southern Italy; overall, 28% of the Italian municipalities historically belonged to the South.

**Vote counting rate and workers' value added.** In section 7 we analyse the effect of ABR drain on two economic outcomes: vote counting rate per hour and firm value added per employee. We refer to Ilzetzki and Simonelli (2017) for an accurate description of the

[1] Information to access the ADELE laboratory and its data-sets can be found at the following link:https://www.istat.it/en/information-and-services/researchers/laboratory-for-elementary-data-analysis.

[2] Specifically, ADELE laboratory provides micro-level census data for the years 1971-2011. The exact date of birth is available from 1981; however, 1981 Census provides the municipality of birth only for individuals who reside in the same municipality as the one of birth; the information on the municipality of birth and residence becomes fully available in 1991.

[3] We are thankful to Giampaolo Lecce for sharing this information with us. The original volumes are available in pdf (ocr-optimized) format at the ISTAT Digital library (https://ebiblio.istat.it).

[4] The whole set of indicators is available at the ISTAT page "Mappa dei rischi nei comuni Italiani", at the following link: https://www.istat.it/it/mappa-rischi.

data.

**Descriptives.** We collapsed this information at the level of local labor market (LLM).[5] As of 1991, ISTAT divided the Italian territory in 781 Local Labor Market commuting zones ("Sistemi Locali del Lavoro").[6] As described in Section 2, our empirical analysis uses information coming from 343 LLMs, of which 305 are in the South and 38 in the North. Table A–1 shows the main characteristics of the local labor markets in the South for the whole and selected sample, respectively. The first rows provide the average characteristics for individuals born in the last five days of December and the first five days of January over the period 1920-1954. The table also reports descriptive statistics for our measure of January birthday cheating at the LLM level ($\Pi_l$) and for the number of individuals born in the December-January window (i.e. Cell Size) along with some of their characteristics, such as gender and education. We also show the LLM population in 1991 and LLM characteristics in 1921 and 1936, such as the share of illiterates and the employment rate. In the last two panels of the Table, we report statistics for and hydro-geological risks as well as the vote counting and value added variables (in log). One can easily notice that the sample selection we applied in Section 2 essentially drops small local labor markets and does not affect the average characteristics of the sample of Southern LLMs.

Table A–2 instead focuses on Northern local labor markets. The sample selection here only retains large LLMs in which the share of migrants is larger and comparable to the one in the South.

Moreover, the table includes statistics for the LLM population as of 1991 and for the one recorded in 1921. Historical censuses from 1921 and 1936 further provides information on the share of illiterates and employment shares. Finally, the table reports statistics on a set of geographic characteristics, such as the population living in areas characterized by a high risk of flood or landslide.

[5]There are a few cases in which a LLM includes municipalities located both in the South and North of Italy; we therefore classified a LLM as belonging to the South if the share of the LLM population living within the historical boundaries of the Kingdom is higher than 50%.

[6]Local labour markets are geographical areas where the bulk of the labour force lives and works, and they are based on the analysis of commuting patterns. The exact number of 1991 LLMs is 784; however, because of the laws on the protection of statistical confidentiality, we could only extract data on LLMs with more than 10 observations, ultimately loosing information for three local labor markets in the North.

Table A–1: Descriptive statistics of the Local Labor Markets - South

| | All LLMs: 327 | | | | Selected LLMs: 305 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean (1) | Std. Dev. (2) | Min. (3) | Max. (4) | Mean (5) | Std. Dev. (6) | Min. (7) | Max. (8) |
| Census 1991 | | | | | | | | |
| $\Pi_l$ | 0.694 | 0.126 | 0.103 | 0.936 | 0.696 | 0.124 | 0.115 | 0.936 |
| Cell Size | 1040 | 2573 | 58 | 37556 | 1104 | 2653 | 99 | 37556 |
| Shr. Migrant | 0.498 | 0.013 | 0.100 | 0.529 | 0.499 | 0.006 | 0.400 | 0.529 |
| Shr. Female | 0.468 | 0.028 | 0.339 | 0.611 | 0.468 | 0.028 | 0.339 | 0.611 |
| Shr. Primary Edu. | 0.752 | 0.070 | 0.497 | 0.950 | 0.751 | 0.070 | 0.497 | 0.950 |
| Shr. Tertiary Edu. | 0.086 | 0.035 | 0.000 | 0.232 | 0.086 | 0.035 | 0.000 | 0.232 |
| Population | 59111 | 164022 | 4071 | 2381483 | 62793 | 169254 | 4121 | 2381483 |
| Census 1921 | | | | | | | | |
| Population | 35396 | 76694 | 0 | 1092627 | 37503 | 78999 | 3037 | 1092627 |
| Shr. Illiterates | 0.472 | 0.087 | 0.140 | 0.724 | 0.472 | 0.087 | 0.140 | 0.724 |
| Density | 324.267 | 427.761 | 0.000 | 1443.422 | 326.247 | 429.235 | 20.267 | 1443.422 |
| Census 1936 | | | | | | | | |
| Employment Rate | 0.365 | 0.046 | 0.269 | 0.527 | 0.364 | 0.046 | 0.269 | 0.527 |
| Shr. Manufacturing | 0.242 | 0.098 | 0.063 | 0.436 | 0.243 | 0.097 | 0.069 | 0.436 |
| Shr. Agriculture | 0.535 | 0.203 | 0.167 | 0.877 | 0.533 | 0.203 | 0.167 | 0.869 |
| Geography | | | | | | | | |
| Landslide risk | 34.371 | 49.743 | 0.000 | 271.266 | 34.593 | 49.928 | 0.000 | 271.266 |
| Flood risk | 24.582 | 31.957 | 0.000 | 135.012 | 24.800 | 32.040 | 0.000 | 135.012 |
| Ilzetzki and Simonelli (2017) | | | | | | | | |
| Value Added (log) | 3.131 | 0.117 | 2.717 | 3.555 | 3.131 | 0.116 | 2.717 | 3.547 |
| VCR (log) | 5.215 | 0.175 | 4.499 | 5.795 | 5.215 | 0.173 | 4.534 | 5.795 |

Notes: The table reports descriptive statistics for the observable characteristics of all the LLMs in the South (columns 1-4) and the 305 selected LLMs in the South (5-8). All the statistics are weighted by LLM cell size, i.e. the number of individuals born in each local labor market over the period 1920-1954.

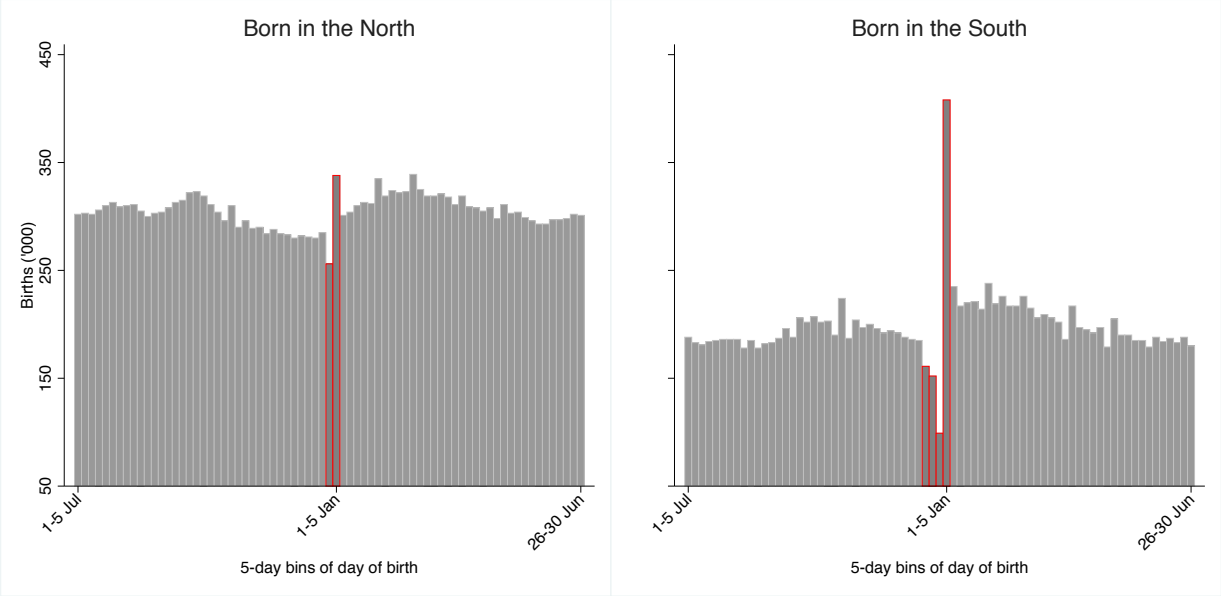
Table A–2: Descriptive statistics of the Local Labor Markets - North

| | All LLMs: 454 | | | | Selected LLMs: 38 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean (1) | Std. Dev. (2) | Min. (3) | Max. (4) | Mean (5) | Std. Dev. (6) | Min. (7) | Max. (8) |
| Census 1991 | | | | | | | | |
| $\Pi_l$ | 0.178 | 0.137 | 0.000 | 0.704 | 0.198 | 0.138 | 0.016 | 0.589 |
| Cell Size | 832 | 1545 | 34 | 21656 | 3393 | 4246 | 422 | 21656 |
| Shr. Migrant | 0.298 | 0.155 | 0.000 | 0.500 | 0.464 | 0.050 | 0.280 | 0.500 |
| Shr. Female | 0.528 | 0.043 | 0.330 | 0.682 | 0.530 | 0.034 | 0.455 | 0.631 |
| Shr. Primary Edu. | 0.733 | 0.098 | 0.461 | 0.950 | 0.668 | 0.091 | 0.461 | 0.926 |
| Shr. Tertiary Edu. | 0.076 | 0.047 | 0.000 | 0.232 | 0.108 | 0.046 | 0.017 | 0.232 |
| Population | 82427 | 235349 | 1390 | 3314237 | 419391 | 707109 | 15633 | 3314237 |
| | 0.694 | 0.126 | 0.103 | 0.936 | 0.696 | 0.124 | 0.115 | 0.936 |
| Census 1921 | | | | | | | | |
| Population | 46852 | 83981 | 0 | 1095819 | 186309 | 228414 | 19906 | 1095819 |
| Shr. Illiterates | 0.195 | 0.141 | 0.004 | 0.706 | 0.188 | 0.138 | 0.042 | 0.606 |
| Density | 211.867 | 217.147 | 0.000 | 1122.047 | 334.150 | 308.574 | 40.354 | 1122.047 |
| Census 1936 | | | | | | | | |
| Employment rate | 0.448 | 0.053 | 0.275 | 0.806 | 0.451 | 0.044 | 0.345 | 0.576 |
| Shr. Manufacturing 1936 | 0.305 | 0.148 | 0.065 | 0.754 | 0.348 | 0.129 | 0.089 | 0.569 |
| Shr. Agriculture 1936 | 0.465 | 0.223 | 0.069 | 0.855 | 0.320 | 0.220 | 0.069 | 0.831 |
| Geography | | | | | | | | |
| Landslide risk | 46.491 | 100.593 | 0.000 | 1327.178 | 73.230 | 128.213 | 0.000 | 510.924 |
| Flood risk | 47.575 | 65.306 | 0.000 | 394.420 | 86.003 | 89.738 | 0.000 | 394.420 |
| Ilzetzki and Simonelli (2017) | | | | | | | | |
| Value Added (log) | 3.407 | 0.141 | 2.468 | 3.902 | 3.385 | 0.116 | 3.080 | 3.584 |
| VCR (log) | 5.536 | 0.184 | 4.545 | 6.335 | 5.420 | 0.140 | 5.261 | 5.974 |

Note: The table reports descriptive statistics for the observable characteristics of all the LLMs in the North (columns 1-4) and the 38 selected LLMs in the North (5-8). All the statistics are weighted by LLM cell size, i.e. the number of individuals born in each local labor market over the period 1920-1954.

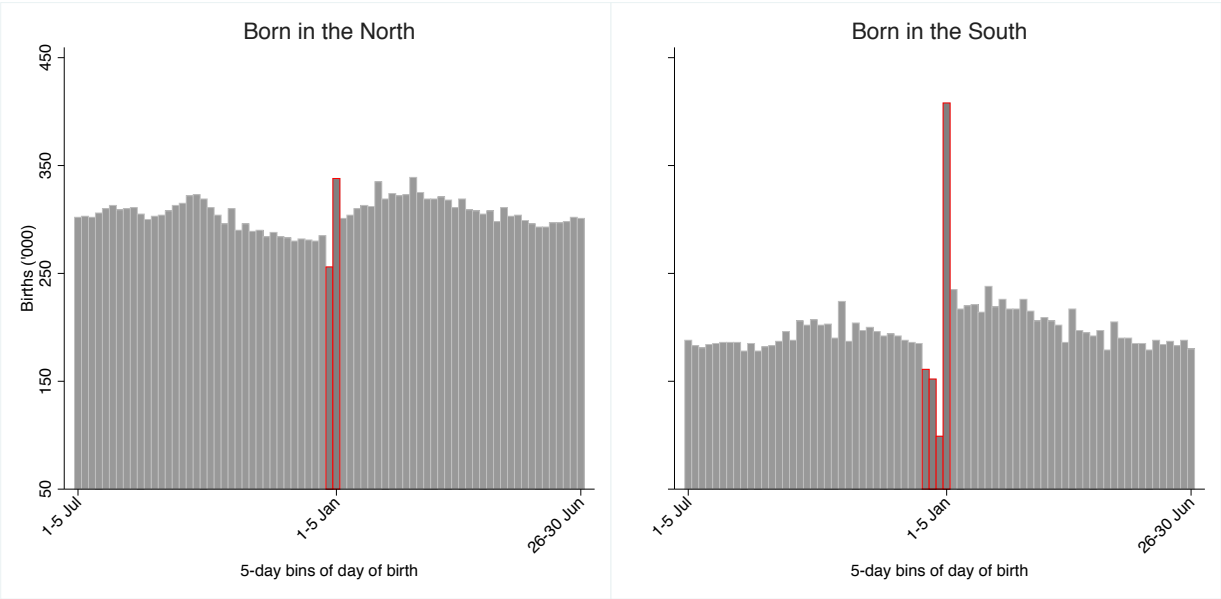
## Birthday cheating in other censuses

Figure A–1: The distribution of birth dates over the days of a calendar year - 2001 Census



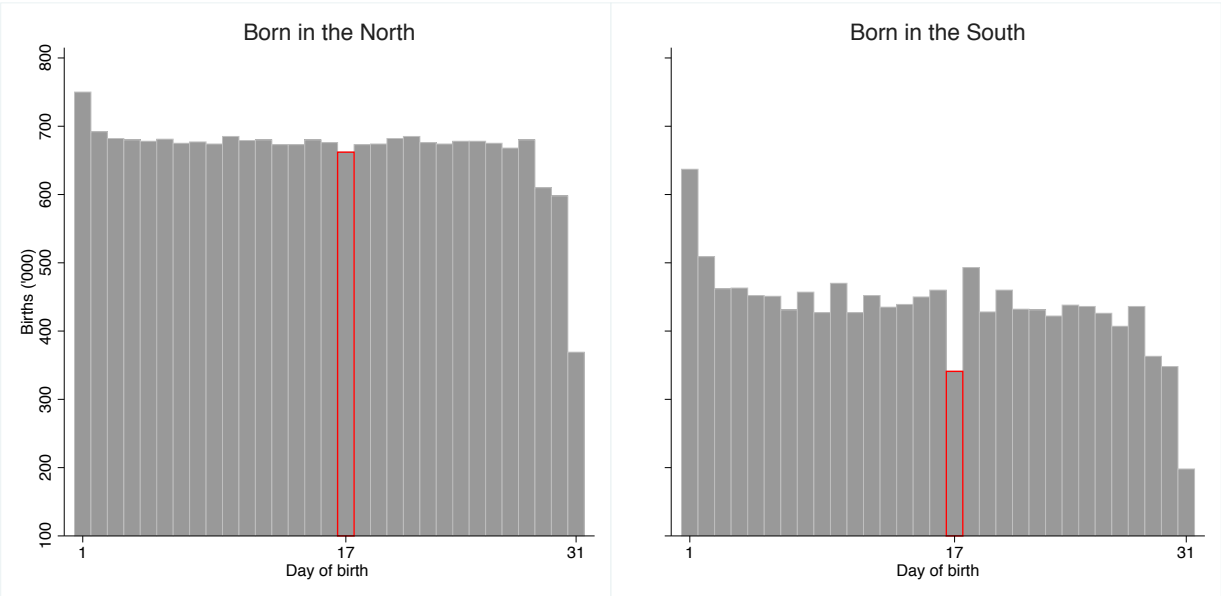

Notes: Restricted Census 2001 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the year, grouped in 5-day bins. The South is defined as the localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies".

Figure A–2: The distribution of birth dates over the days of a calendar year - 2011 Census



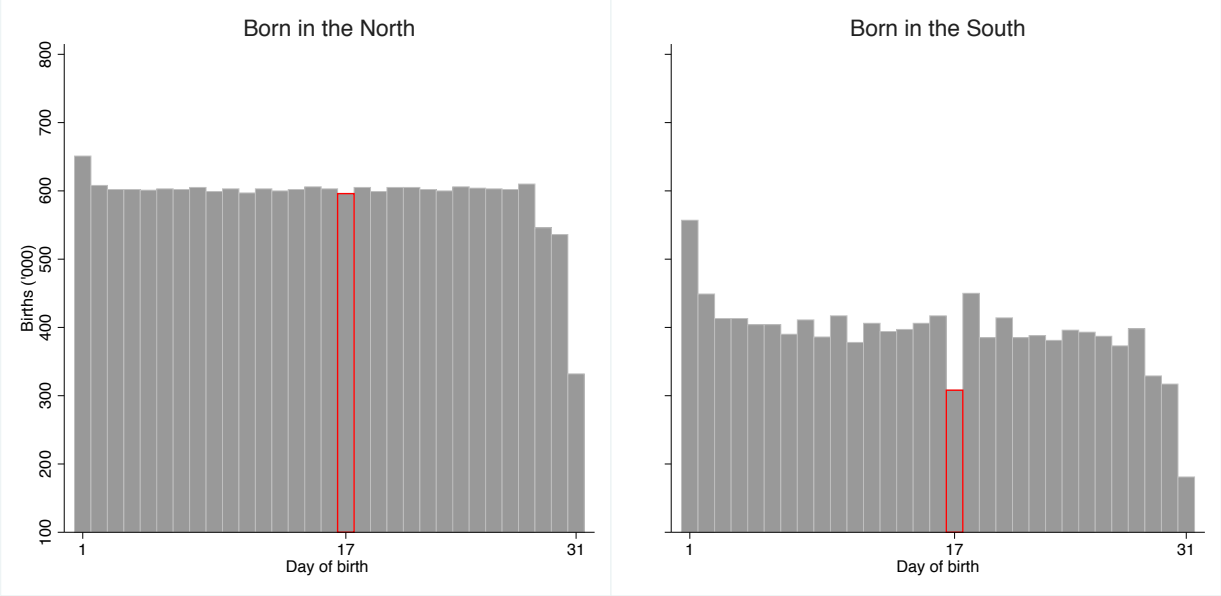

Note: restricted Census 2011 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the year, grouped in 5-day bins. The South is defined as the localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies".

Figure A–3: The distribution of birthdays over the days of a calendar month - 2001 Census




Note: restricted Census 2001 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the calendar month. The South is defined as the localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies".

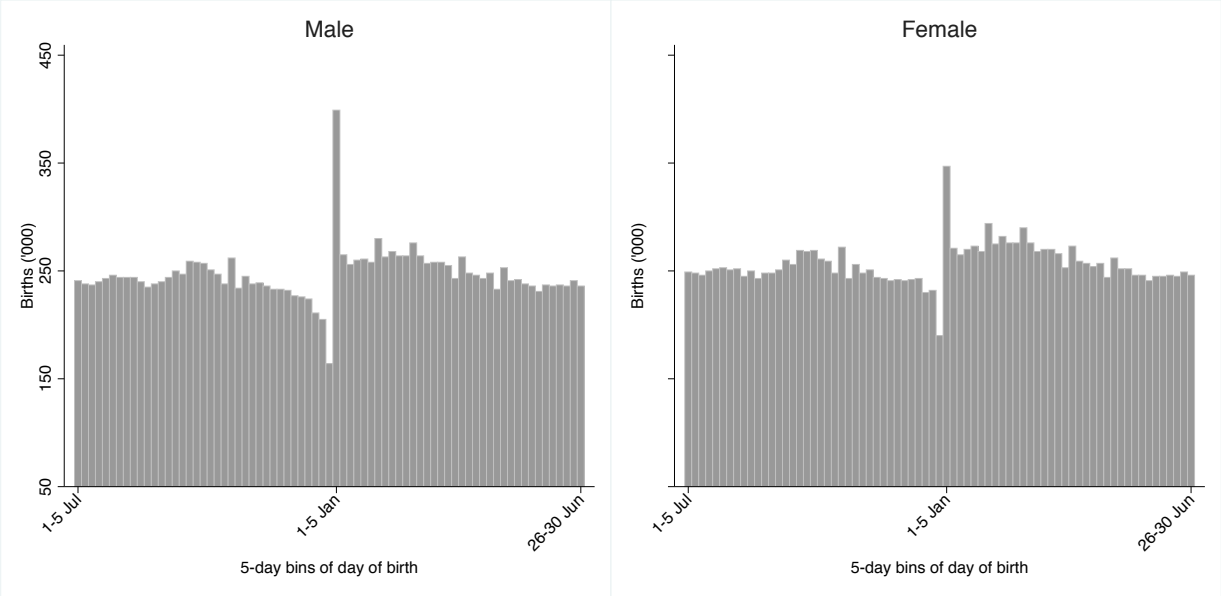
Figure A–4: The distribution of birthdays over the days of a calendar month - 2011 Census




Note: restricted Census 2011 data, with exact birth date for the 1921-1954cohorts. The figure plots the total births by day of the calendar month. The South is defined as the localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies".

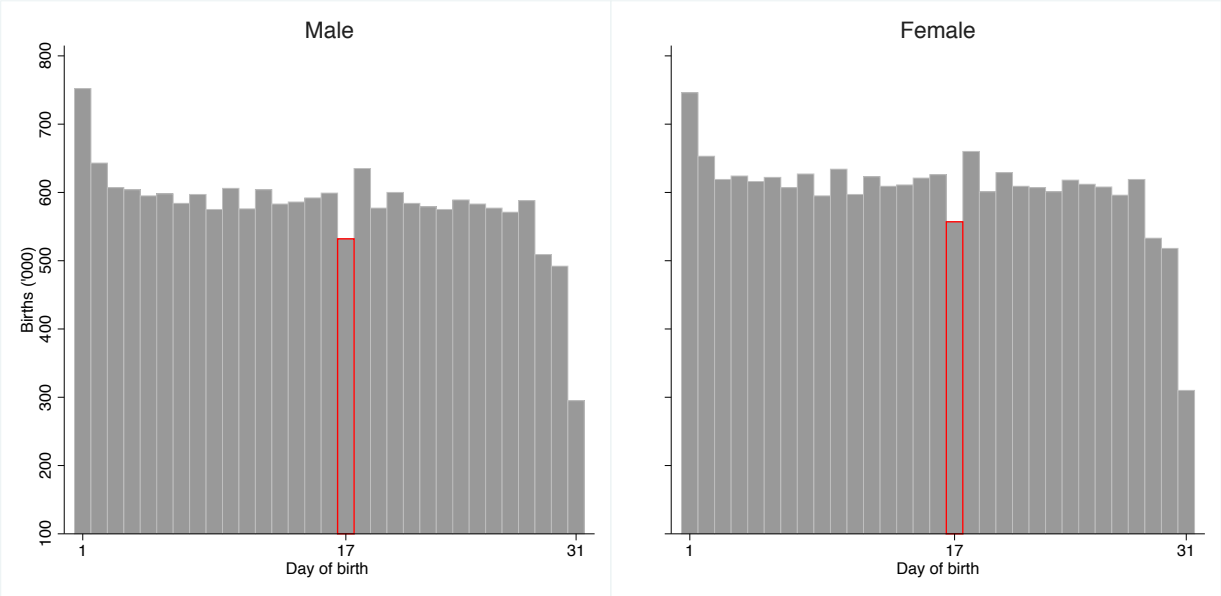
## Birthday cheating by gender and education

Figure A–5: JBD and gender of the child



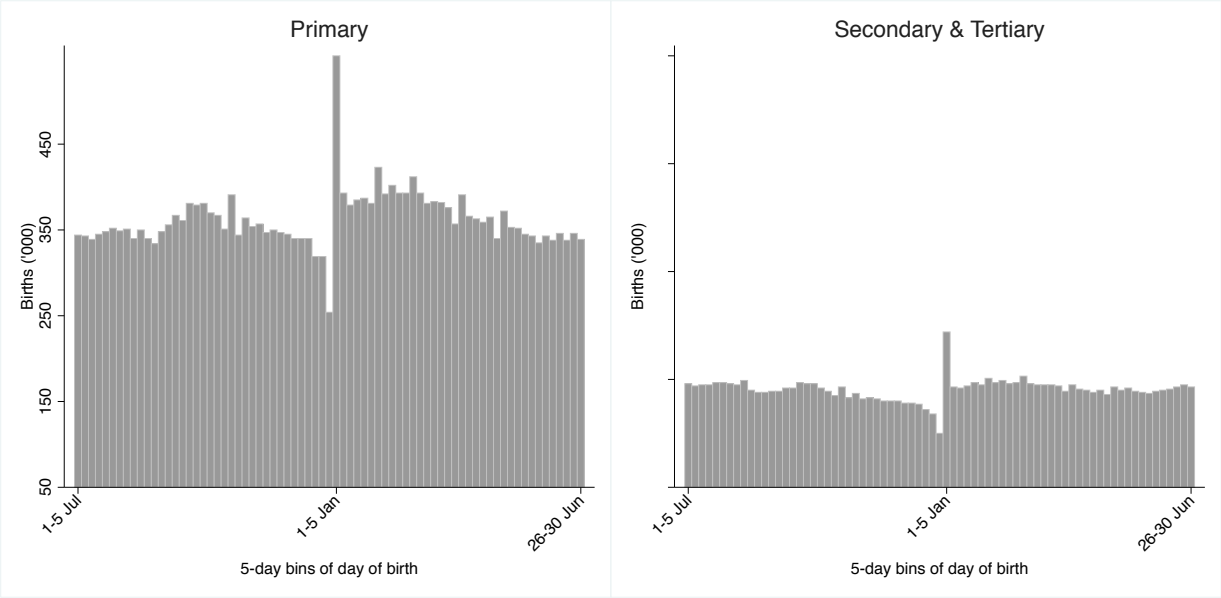

Notes: Restricted Census 1991 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the year, grouped in 5-day bins, separately for males and females.

Figure A–6: 17BD and gender of the child



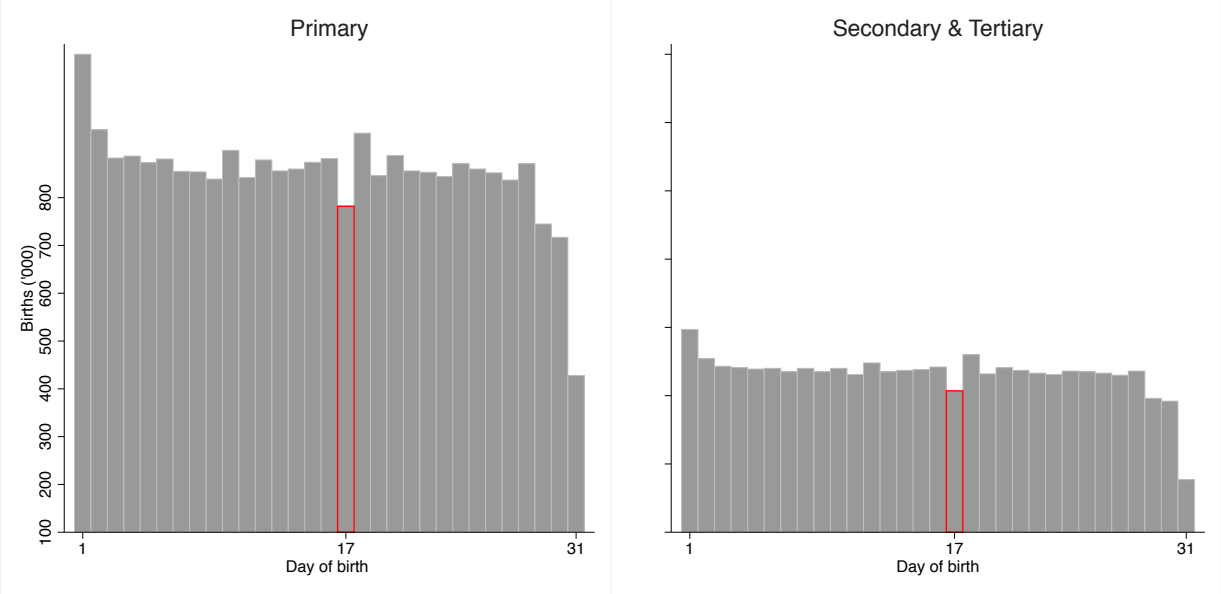

Notes: Restricted Census 1991 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the calendar month separately for males and females.

Figure A–7: JBD and education of the child




Notes: Restricted Census 1991 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the year, grouped in 5-day bins, separately for primary (and less) and secondary/tertiary educated individuals.

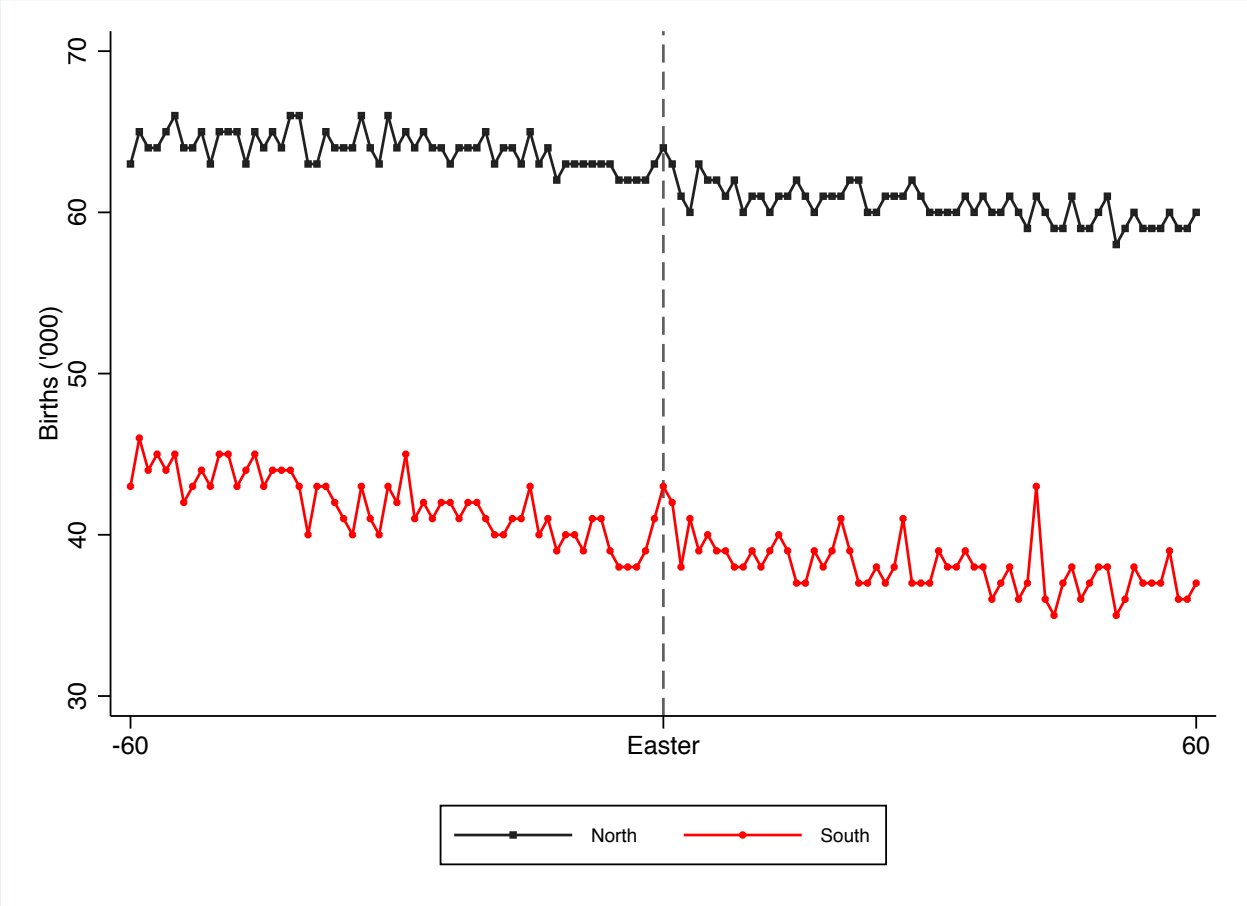
Figure A–8: 17BD and education of the child




Note: Restricted Census 1991 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total births by day of the calendar month separately for primary (and less) and secondary/tertiary educated individuals.

## Evidence excluding irrelevant explanations for birthday cheating

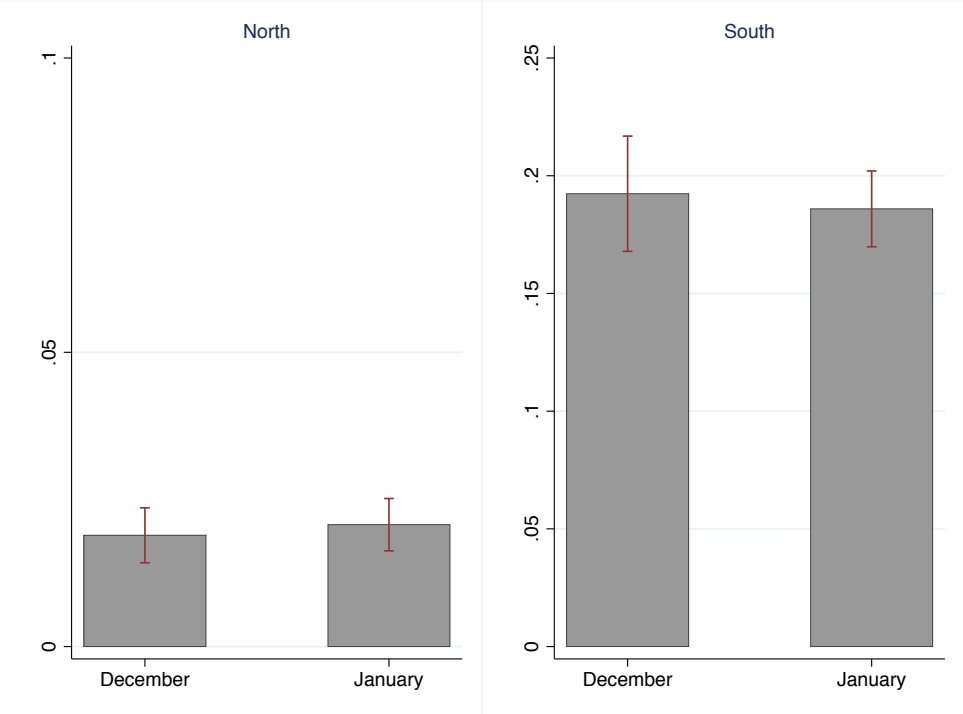
Figure A–9: BD Cheating around Easter day




Notes: Restricted Census 1991 data, with exact birth date for the 1921-1954 cohorts. The figure plots the total daily births by distance from/to the Easter day of the year of birth. The South is defined as the localities that between 1816 and 1861 were part of the "Kingdom of the two Sicilies".

# Online Appendix to Section 5

Figure A–10: Migration probability by date of birth




Notes: The figure reports averages and 95% confidence intervals of migration probability for January and December born in the 38 Local Labor Markets (LLM) in the North of Italy (left) and in the 305 LLM in the South (right). We only consider indivuals born in the last and first five days of the year for the period 1955-1965. Observations are weighted by the number of births in the cell.

Table A–3: JBD cheating of Migrants and Remainers, showing covariates: Prediction (A)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Migrant∗South ($\beta_4$) | -0.021** | -0.011*** | -0.013*** | -0.013*** |
| *(Mig.S-Rem.S)* | (0.009) | (0.004) | (0.004) | (0.004) |
| Migrant∗North ($\beta_3$) | 0.197*** | 0.104*** | 0.096*** | 0.114*** |
| *(Mig.N-Rem.N)* | (0.050) | (0.024) | (0.026) | (0.020) |
| South ($\beta_2$) | 0.492*** | | | |
| *(Remainers,South)* | (0.040) | | | |
| $\beta_1$ | 0.212*** | 0.591*** | 0.729*** | 0.724*** |
| *(Remainers,North)* | (0.037) | (0.001) | (0.023) | (0.062) |
| Share primary edu. | | | -0.099*** | 11.187** |
| | | | (0.028) | (5.217) |
| Share tertiary edu. | | | 0.002 | 15.029* |
| | | | (0.040) | (8.590) |
| Share women | | | -0.136*** | -6.080 |
| | | | (0.022) | (4.465) |
| South x share prim. edu. | | | | -0.090 |
| | | | | (0.109) |
| South x share tert. edu. | | | | 0.195 |
| | | | | (0.131) |
| South x share women | | | | -0.126 |
| | | | | (0.083) |
| Cohort x share prim. edu. | | | | -0.006** |
| | | | | (0.003) |
| Cohort x share tert. edu. | | | | -0.008* |
| | | | | (0.004) |
| Cohort x share women | | | | 0.003 |
| | | | | (0.002) |
| Share prim. edu. x share tert. edu. | | | | -0.112 |
| | | | | (0.220) |
| Share prim. edu. x share women | | | | 0.095 |
| | | | | (0.143) |
| Share tert. edu. x share women | | | | 0.014 |
| | | | | (0.242) |
| Observations | 10,360 | 10,360 | 10,360 | 10,360 |
| Sum of weights | 430,709 | 430,709 | 430,709 | 430,709 |
| R-squared | 0.506 | 0.801 | 0.805 | 0.807 |
| LLM x Biennium FE | No | Yes | Yes | Yes |
| Controls | No | No | Yes | Yes |
| Controls interacted | No | No | No | Yes |
| Oster $\delta$ for $\beta_3 = 0$ | | 5.618 | 4.510 | 4.795 |
| Oster $\delta$ for $\beta_4 = 0$ | | 6.854 | 11.51 | 11.41 |
| p=value of F-stat for controls | | | 0 | 0 |

Notes: The table reports OLS estimates based on data for 305 Local Labor Markets (LLM) in the South of Italy and 38 in the North (p<0.01, ** p<0.05, * p<0.1). The dependent variable is $\Pi_{glt}$, the share of births that occurred before December 31 but are declared as occurring in the first five days of the following year in group $g$, locality $l$ and biennium $t$. $\beta_1$ and $\beta_2$ in column 1 are estimates of the share of cheaters in the group of remainers in the north and in the South, respectively. $\beta_3$ captures the difference in cheating between the migrants from North to South and the remainers in the North and $\beta_4$ the difference between the migrants from South to North and the remainers in the South. LLM fixed effects capture differences in unobserved characteristics across localities, such as the level of deterrence. Controls include the average year of birth, the share of female, the share of people with primary education and share with tertiary education. In column 4 we add all possible interactions among the controls. $\delta$'s are the statistics proposed by Oster (2019) capturing how much the unobservable characteristics would have to be correlated with migration status more than the observed ones in order to conclude that migrants and remainers born have the same cheating probability. The p-value in the last row refers to an F-test for the joint significance of all controls included.

# Online Appendix to Section 6

**Reactions to deterrence in the North.** We repeat the same analysis of Table A–4 for agents born in the North keeping in mind that, as explained in Section 4, this exercise can be conducted for only 38 localities of the North and for a total of 728 group×locality×biennium cells. Again in line with the evidence of Section 5, the emerging pattern suggests that in the North, differently than in the South, migrants constitute the group that reacts more, in absolute terms, to the changes of deterrence induced by the surge and collapse of Fascism. Focusing for example on our preferred specification in column 4, which implies a comparison within locality and with a full set of interacted controls, the probability of JBD cheating for remainers in the North outside of the Fascist period is 49.2% and it increases by 10.5 percentage points among migrants born in the same region. The effect of the higher Fascist deterrence between 1926 and 1940 is a reduction of 10.6 percentage points of cheating probability for remainers, but the estimate of the $\beta_4$ coefficient suggests that the reaction of migrants is about the same.

Also in the case of this table we rely on Oster (2019) to assess the plausibility that our results would change if we could include unobservable characteristics in the specification. While for the parameters $\beta_3$ the estimates of $\delta$ are reassuring, $\beta_4$ is already estimated to be close to zero and this result is unlikely to be changed by the hypothetical inclusion of unobservable controls.

A possible conclusion, based on this finding about $\beta_4$ in Table A–5 for the North, is that migrants and remainers have the same underlying ABR and for this reason they react in the same way to the same change of deterrence. But this interpretation would be incompatible with the evidence in Section 5, which indicates that in the North the proportion of high ABR families is lower among migrants towards South.

An alternative interpretation is that deterrence against JBD cheating was already high in the North even before Fascism. This is the case in which our model of Section 3 suggests that the cheating probability is completely determined by the common high level of deterrence, independently of the possibly different underlying levels of ABR in the two groups. However, also this conclusion would not be compatible with the evidence in Section 5 because if the North had been in the case of high deterrence of Table 1, not only the reactions to changes in deterrence but also the levels of the cheating probability would have had to be the same in the two groups. Moreover, under this interpretation the cheating probabilities would not be informative about the underlying levels of ABR, and they would therefore be useless for inference on the possibility of a ABR drain or gain.

Table A–4: Reactions to deterrence, South, showing covariates: Predictions (B) and (C)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Fascism $*$ Migrant ($\beta_4$) | 0.024*** | 0.024*** | 0.018** | 0.017** |
| *(R.NF. -R.F.)–(M.NF.-M.F.)* | (0.009) | (0.008) | (0.008) | (0.008) |
| Migrant ($\beta_3$) | -0.031*** | -0.020*** | -0.019*** | -0.020*** |
| *(Mig–Rem, No Fascism)* | (0.007) | (0.004) | (0.004) | (0.004) |
| Fascism ($\beta_2$) | -0.201*** | | | |
| *(Rem Fascism–Rem No Fascism)* | (0.009) | | | |
| $\beta_1$ | 0.776*** | 0.702*** | 0.841*** | 0.871*** |
| *(Remainers, No Fascism)* | (0.013) | (0.001) | (0.024) | (0.065) |
| Share primary edu. | | | -0.101*** | 9.954* |
| | | | (0.029) | (5.540) |
| Share tertiary edu. | | | 0.013 | 16.969* |
| | | | (0.042) | (9.017) |
| Share women | | | -0.139*** | -7.277 |
| | | | (0.023) | (4.743) |
| Cohort x share prim. edu. | | | | -0.005* |
| | | | | (0.003) |
| Cohort x share tert. edu. | | | | -0.009* |
| | | | | (0.005) |
| Cohort x share women | | | | 0.004 |
| | | | | (0.002) |
| Share prim. edu. x share tert. edu. | | | | -0.183 |
| | | | | (0.241) |
| Share prim. edu. x share women | | | | 0.140 |
| | | | | (0.166) |
| Share tert. edu. x share women | | | | 0.082 |
| | | | | (0.253) |
| Observations | 9,632 | 9,632 | 9,632 | 9,632 |
| R-squared | 0.205 | 0.830 | 0.833 | 0.833 |
| LLM x Biennium FE | No | Yes | Yes | Yes |
| Controls | No | No | Yes | Yes |
| Controls interacted | No | No | No | Yes |
| Oster $\delta$ for $\beta_4 = 0$ | | 117.9 | 9.023 | 8.738 |
| Oster $\delta$ for $\beta_3 = 0$ | | 6.555 | 6.622 | 7.098 |
| p=value of F-stat for controls | | | 0 | 0 |

Notes: The table reports OLS (difference-in-difference) estimates based on data for 305 Local Labor Markets (LLM) in the South of Italy, observed for at most 17 bienniums between 1921 and 1954 (*** p<0.01, ** p<0.05, * p<0.1). Observations are weighted by the number of births in the cell defined by a group $g$ of migrant or remainers in a locality $l$ and biennium $t$. Standard errors are clustered at the locality level. The dependent variable is $\Pi_{glt}$, the share of births that occurred in the five days before December 31 but are declared as occurring in the first five days of the following year in group $g$, locality $l$ and biennium $t$. $\beta_1$ in column 1 estimates the share of cheaters in the group of remainers in the South in the periods 1921-1926 and 1940-1954. $\beta_2$ estimates the difference in cheating for remainers in the South between the period of Fascist deterrence (1927-1939) and the periods 1921-1926 and 1940-1954. $\beta_3$ captures the difference in cheating between the migrants from South to North and the remainers in the South in the periods 1921-1926 and 1940-1954. $\beta_4$ is the diff-in-diff estimate that captures the difference in reaction to Fascist deterrence between migrants from South to North and remainers in the South. The $\beta_1$ coefficients in columns 2, 3 and 4 are not reported because they do not have a meaningful interpretation given the inclusion of LLM fixed effects. The Population represented by cells is the total number of birth on which the regression would be run if individual observations had not been collapsed at the cell level. Controls include the average year of birth, the share of female, the share of people with primary education and the share with tertiary education. In column 4 we add all possible interactions among the controls. $\delta$s are the statistics proposed by Oster (2019) capturing how much the unobservable characteristics would have to be correlated with migration status more than the observed ones in order to conclude that migrant and remainer families have the same cheating probability. The p-value in the last row refers to an F-test for the joint significance of all controls included.

Table A–5: Reactions to deterrence, North, showing covariates: Predictions (B) and (C)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Fascism $*$ Migrant ($\beta_4$) | 0.023 | 0.037 | 0.033 | 0.038 |
| *(M.F.-M.NF) – (R.F. -R.NF)* | (0.035) | (0.040) | (0.043) | (0.043) |
| Migrant ($\beta_3$) | 0.185*** | 0.089*** | 0.100*** | 0.099*** |
| *(Mig-Rem, No Fascism)* | (0.048) | (0.017) | (0.021) | (0.021) |
| Fascism ($\beta_2$) | -0.110*** | | | |
| *(Rem Fascism-Rem No Fascism)* | (0.016) | | | |
| $\beta_1$ | 0.259*** | | | |
| *(Remainers, No Fascism)* | (0.042) | | | |
| Share primary edu. | | | 0.010 | 18.381 |
| | | | (0.105) | (15.575) |
| Share tertiary edu. | | | -0.111 | 2.076 |
| | | | (0.112) | (30.263) |
| Share women | | | -0.027 | 9.555 |
| | | | (0.083) | (16.342) |
| Cohort x share prim. edu. | | | | -0.009 |
| | | | | (0.008) |
| Cohort x share tert. edu. | | | | -0.001 |
| | | | | (0.016) |
| Cohort x share women | | | | -0.005 |
| | | | | (0.008) |
| Share prim. edu. x share tert. edu. | | | | 0.193 |
| | | | | (0.550) |
| Share prim. edu. x share women | | | | -0.133 |
| | | | | (0.259) |
| Share tert. edu. x share women | | | | -0.208 |
| | | | | (0.658) |
| Observations | 728 | 728 | 728 | 728 |
| Sum of weights | 98,005 | 98,005 | 98,005 | 98,005 |
| R-squared | 0.127 | 0.730 | 0.740 | 0.752 |
| LLM x Biennium FE | No | Yes | Yes | Yes |
| Controls | No | No | Yes | Yes |
| Controls interacted | No | No | No | Yes |
| Oster $\delta$ for $\beta_4 = 0$ | | -9.970 | -17.26 | 49.06 |
| Oster $\delta$ for $\beta_2 = 0$ | | 3.494 | 3.861 | 4.423 |
| p-value of F-test for controls | | | 0.167 | 0.034 |

Notes: The table reports OLS (difference-in-difference) estimates based on data for 38 Local Labor Markets (LLM) in the North of Italy, observed for at most 17 bienniums between 1921 and 1954 (*** p<0.01, ** p<0.05, * p<0.1). Observations are weighted by the number of births in the cell defined by a group $g$ of migrant or remainers in a locality $l$ and biennium $t$. Standard errors are clustered at the locality level. The dependent variable is $\Pi_{glt}$, the share of births that occurred in the five days before December 31 but are declared as occurring in the first five days of the following year in group $g$, locality $l$ and biennium $t$. $\beta_1$ in column 1 estimates the share of cheaters in the group of remainers in the North in the periods 1921-1926 and 1940-1954. $\beta_2$ estimates the difference in cheating for remainers in the north between the period of Fascist deterrence (1927-1939) and the periods 1921-1926 and 1940-1954. $\beta_3$ captures the difference in cheating between the migrants from North to South and the remainers in the North in the periods 1921-1926 and 1940-1954. $\beta_4$ is the diff-in-diff estimate that captures the difference in reaction to Fascist deterrence between migrants from North to South and remainers in the North. The $\beta_1$ coefficients in columns 2, 3 and 4 are not reported because they do not have a meaningful interpretation given the inclusion of LLM fixed effects. The Population represented by cells is the total number of birth on which the regression would be run if individual observations had not been collapsed at the cell level. Controls include the average year of birth, the share of female, the share of people with primary education and the share with tertiary education. In column 4 we add all possible interactions among the controls. $\delta$s are the statistics proposed by Oster (2019) capturing how much the unobservable characteristics would have to be correlated with migration status more than the observed ones in order to conclude that migrant and remainer families have the same cheating probability. The p-value in the last row refers to an F-test for the joint significance of all controls included.

# Online Appendix to Section 7

Table A–6: ABR drain and Vote Counting Rate for the 2016 referendum - Including North

| Log(Vote Counting Productivity) - Referendum 2016 December | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| ABR Drain (standardized) | -0.058*** | -0.055*** | -0.043*** | -0.029* | -0.032** |
| | (0.017) | (0.018) | (0.016) | (0.016) | (0.016) |
| Brain Drain (standardized) | | -0.022 | -0.005 | -0.005 | -0.008 |
| | | (0.021) | (0.018) | (0.017) | (0.015) |
| Observations | 343 | 343 | 343 | 343 | 343 |
| R-squared | 0.027 | 0.033 | 0.168 | 0.387 | 0.449 |
| Region FE | No | No | Yes | Yes | Yes |
| Initial Period Controls | No | No | No | Yes | Yes |
| Employment and Geography Controls | No | No | No | No | Yes |
| Drain mean | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| Drain S.D. | 0.016 | 0.016 | 0.016 | 0.016 | 0.016 |
| Outcome Levels Mean | 194.500 | 194.500 | 194.500 | 194.500 | 194.500 |
| Outcome Levels S.D. | 35.544 | 35.544 | 35.544 | 35.544 | 35.544 |
| Oster $\delta$ for ABR drain | | | 5.882 | 2.851 | 3.471 |

Notes: The table reports OLS estimates based on data for 305 Local Labor Markets (LLM) in the South of Italy (*** p<0.01, ** p<0.05, * p<0.1). Observations are weighted by the number of births in each locality $l$. Standard errors are robust for heteroskedasticity. The dependent variable is the logarithm of vote counting rate per hour in the 2016 constitutional referendum. The ABR drain ($\theta_l$) is the difference in the probability of JBD cheating of remainers in $l$ versus born in $l$ and measures how the the fraction of high ABR agents has changed in the remaining population after the emigration process has taken place. $\theta_l$ is standardized. The Brain Drain is standardized as well and measures how the the fraction of agents with secondary or tertiary education has changed in the remaining population after the emigration process has taken place. The initial period controls are the probability of JBD cheating for the same locality in the cohort born in the 1920–26 period and the share of illiterates from the 1921 Census. Region fixed effects are for the 7 current administrative units partitioning the South, as defined by the Kingdom of the two Sicilies. The employment controls are: employment rate, share of agricultural employment, share of manufacturing employment, share of service employment from the 1936 census, total population in the LLM and population density from the 1921 census. The geography controls are dummies for: coastal land, low lands, low mountains, high mountains, flood risk, rock slide risk.

Table A–7: ABR drain and Firm Labor Productivity - Including North

| Firm value added per worker | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| ABR Drain (standardized) | -0.017** | -0.017** | -0.015* | -0.015** | -0.015** |
| | (0.009) | (0.009) | (0.009) | (0.007) | (0.007) |
| Brain Drain (standardized) | | -0.001 | 0.012 | 0.002 | 0.003 |
| | | (0.011) | (0.011) | (0.008) | (0.008) |
| Observations | 415,907 | 415,907 | 415,907 | 415,907 | 415,907 |
| R-squared | 0.001 | 0.001 | 0.012 | 0.287 | 0.289 |
| Initial Period Controls | No | No | yes | Yes | Yes |
| Region FE | No | No | No | Yes | Yes |
| Industry FE | No | No | No | Yes | Yes |
| Capital Controls | No | No | No | Yes | Yes |
| Employment and Geography Controls | No | No | No | No | Yes |
| Drain mean | -0.002 | -0.002 | -0.002 | -0.002 | -0.002 |
| Drain S.D. | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 |
| Outcome Levels Mean | 23.519 | 23.519 | 23.519 | 23.519 | 23.519 |
| Outcome Levels S.D. | 15.263 | 15.263 | 15.263 | 15.263 | 15.263 |
| Oster $\delta$ for ABR drain | | | 28.81 | -13.15 | -10.32 |

Notes: The table reports OLS estimates based on data for 305 Local Labor Markets (LLM) in the South of Italy (*** p<0.01, ** p<0.05, * p<0.1). Observations are weighted by the employment share of each firm within a locality $l$. Standard errors are clustered at the locality level. The dependent variable is the logarithm of firm value added per employee, averaged over the 2009-2018 period. The ABR drain ($\theta_l$) is the difference in the probability of JBD cheating of remainers in $l$ versus born in $l$ and measures how the the fraction of high ABR agents has changed in the remaining population after the emigration process has taken place. $\theta_l$ is standardized. The Brain Drain is standardized as well and measures how the the fraction of agents with secondary or tertiary education has changed in the remaining population after the emigration process has taken place. The initial period controls are the probability of JBD cheating for the same locality in the cohort born in the 1920–26 period and the share of illiterates from the 1921 Census. Region fixed effects are for the 7 current administrative units partitioning the South, as defined by the Kingdom of the two Sicilies. Industry fixed effects are defined as 2-digit NACE classification. Capital controls are the logarithm of capital per employee and the share of high education individuals in the SLL. The employment controls are: employment rate, share of agricultural employment, share of manufacturing employment, share of service employment from the 1936 census, total population in the LLM and population density from the 1921 census. The geography controls are dummies for: coastal land, low lands, low mountains, high mountains, flood risk, rock slide risk, volcanic risk.

Table A–8: ABR drain and Vote Counting Rate for the 2016 referendum, showing covariates

| Log(Vote Counting Productivity) - Referendum 2016 December | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| ABR Drain (standardized) | -0.041*** | -0.040** | -0.037** | -0.025* | -0.027* |
| | (0.016) | (0.016) | (0.015) | (0.015) | (0.015) |
| Brain Drain (standardized) | | -0.009 | -0.010 | -0.004 | -0.006 |
| | | (0.020) | (0.018) | (0.016) | (0.015) |
| ABR 1920 | | | -0.130* | -0.087 | -0.108 |
| | | | (0.078) | (0.086) | (0.105) |
| Share Illiterates 1921 | | | 0.117 | 0.299* | -0.103 |
| | | | (0.167) | (0.177) | (0.185) |
| Employment Rate 1936 | | | | | 0.062 |
| | | | | | (0.330) |
| Agriculture Emp. Share 1936 | | | | | 0.800** |
| | | | | | (0.330) |
| Manufacture Emp. Share 1936 | | | | | 1.130** |
| | | | | | (0.503) |
| Population 1921 | | | | | 0.000 |
| | | | | | (0.000) |
| Population density 1921 | | | | | -0.000 |
| | | | | | (0.000) |
| Rock slide risk medium | | | | | -0.001 |
| | | | | | (0.001) |
| Rock slide risk low | | | | | -0.000 |
| | | | | | (0.000) |
| Rock slide risk high | | | | | 0.000 |
| | | | | | (0.000) |
| Flood risk high | | | | | 0.001 |
| | | | | | (0.001) |
| Flood risk medium | | | | | -0.004*** |
| | | | | | (0.001) |
| Flood risk low | | | | | 0.003** |
| | | | | | (0.001) |
| High Mountains | | | | | -0.067 |
| | | | | | (0.048) |
| Low Mountains | | | | | -0.038 |
| | | | | | (0.045) |
| Costal | | | | | 0.001 |
| | | | | | (0.044) |
| Observations | 305 | 305 | 305 | 305 | 305 |
| R-squared | 0.019 | 0.020 | 0.037 | 0.192 | 0.283 |
| Drain mean | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| Drain S.D. | 0.018 | 0.018 | 0.018 | 0.018 | 0.018 |
| Outcome Levels Mean | 186.657 | 186.657 | 186.657 | 186.657 | 186.657 |
| Outcome Levels S.D. | 32.759 | 32.759 | 32.759 | 32.759 | 32.759 |
| Oster $\delta$ for ABR drain | | | 5.158 | 3.666 | 4.303 |

Notes: The table reports OLS estimates based on data for 305 Local Labor Markets (LLM) in the South of Italy (p<0.01, ** p<0.05, * p<0.1). The dependent variable is the logarithm of vote counting rate per hour in the 2016 constitutional referendum. The ABR drain ($\theta_l$) is the difference in the probability of JBD cheating of remainers in $l$ versus born in $l$ and measures how the the fraction of honest agents has changed in the remaining population after the emigration process has taken place. $\theta_l$ is standardized. The Brain Drain is standardized as well and measures how the the fraction of agents with secondary or tertiary education has changed in the remaining population after the emigration process has taken place. The initial period controls are the probability of JBD cheating for the same locality in the cohort born in the 1920–26 period and the share of illiterates from the 1921 Census. Region fixed effects are defined for the 7 current administrative units partitioning the South, as defined by the Kingdom of the two Sicilies. The employment controls are: employment rate, share of agricultural employment, share of manufacturing employment, share of service employment from the 1936 census, total population in the LLM and population density from the 1921 census. The geography controls are dummies for: coastal land, low lands, low mountains, high mountains, flood risk, rock slide risk.

## Table A–9: ABR drain and Firm Labor Productivity, showing covariates

| Firm value added per worker | | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| ABR Drain (standardized) | -0.017** | -0.017** | -0.016* | -0.014** | -0.014** |
| | (0.009) | (0.008) | (0.009) | (0.007) | (0.007) |
| Brain Drain (standardized) | | 0.007 | 0.010 | 0.002 | 0.004 |
| | | (0.010) | (0.011) | (0.008) | (0.008) |
| ABR 1920 | | | -0.117* | 0.066 | -0.004 |
| | | | (0.060) | (0.049) | (0.056) |
| Share Illiterates 1921 | | | 0.038 | 0.074 | 0.191* |
| | | | (0.123) | (0.101) | (0.101) |
| Log of capital per worker | | | | 0.122*** | 0.122*** |
| | | | | (0.004) | (0.004) |
| Years of educ. in SLL | | | | 0.263* | 0.002 |
| | | | | (0.137) | (0.145) |
| Employment Rate 1936 | | | | | -0.102 |
| | | | | | (0.216) |
| Agriculture Emp. Share 1936 | | | | | -0.219 |
| | | | | | (0.214) |
| Manufacture Emp. Share 1936 | | | | | 0.094 |
| | | | | | (0.297) |
| Population 1921 | | | | | -0.000 |
| | | | | | (0.000) |
| Population density 1921 | | | | | 0.000 |
| | | | | | (0.000) |
| Rock slide risk medium | | | | | -0.000 |
| | | | | | (0.000) |
| Rock slide risk low | | | | | -0.000 |
| | | | | | (0.000) |
| Rock slide risk high | | | | | 0.000 |
| | | | | | (0.000) |
| Flood risk high | | | | | -0.000 |
| | | | | | (0.001) |
| Flood risk medium | | | | | 0.003* |
| | | | | | (0.002) |
| Flood risk low | | | | | -0.003* |
| | | | | | (0.001) |
| High Mountains | | | | | 0.006 |
| | | | | | (0.025) |
| Low Mountains | | | | | 0.063* |
| | | | | | (0.034) |
| Costal | | | | | 0.018 |
| | | | | | (0.021) |
| Observations | 187,389 | 187,389 | 187,389 | 187,389 | 187,389 |
| R-squared | 0.001 | 0.001 | 0.002 | 0.280 | 0.283 |
| Drain mean | -0.002 | -0.002 | -0.002 | -0.002 | -0.002 |
| Drain S.D. | 0.034 | 0.034 | 0.034 | 0.034 | 0.034 |
| Outcome Levels Mean | 22.789 | 22.789 | 22.789 | 22.789 | 22.789 |
| Outcome Levels S.D. | 14.642 | 14.642 | 14.642 | 14.642 | 14.642 |
| Oster $\delta$ for ABR drain | | | 22.68 | 43.93 | -31.98 |

Notes: The table reports OLS estimates based on data for 305 Local Labor Markets (LLM) in the South of Italy (p<0.01, ** p<0.05, * p<0.1). The dependent variable is the logarithm of firm value added per employee, averaged over the 2009-2018 period. The ABR drain ($\theta_l$) is the difference in the probability of JBD cheating of remainers in $l$ versus born in $l$ and measures how the the fraction of honest agents has changed in the remaining population after the emigration process has taken place. $\theta_l$ is standardized. The Brain Drain is standardized as well and measures how the the fraction of agents with secondary or tertiary education has changed in the remaining population after the emigration process has taken place. The initial period controls are the probability of JBD cheating for the same locality in the cohort born in the 1920–26 period and the share of illiterates from the 1921 Census. Region fixed effects are defined for the 7 current administrative units partitioning the South, as defined by the Kingdom of the two Sicilies. Industry fixed effects are defined as 2-digit NACE classification. Capital controls are the logarithm of capital per employee and the share of high education individuals in the SLL. The employment controls are: employment rate, share of agricultural employment, share of manufacturing employment, share of service employment from the 1936 census, total population in the LLM and population density from the 1921 census. The geography controls are dummies for: coastal land, low lands, low mountains, high mountains, flood risk, rock slide risk.

# References

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.

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.

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