STUCK IN THE WRONG PLACE: LONG-TERM CONSEQUENCES OF PIRATE ATTACKS IN ITALY•

Antonio Accetturo, Michele Cascarano, Guido de Blasio
Bank of Italy

April, 2017

Abstract
From the VIII to early XIX century, coastal areas in Italy (especially, the South) were raided by pirates coming from the shores of Northern Africa. This paper documents that to protect themselves, residents of coastal places relocate inwards to mountainous and rugged areas. It also shows that the relocation had persistent effects, as it constrained local economic development for a long period after the pirates' threat was over. We also present evidence that – by lowering the human capital endowment and by hampering the growth of major urban centers – mislocation had negative aggregate effects on the economy.
JEL Classification: R1, N9, O1
Keywords: City structure; Historical shocks; Local development

• We wish to thank Andrea Fracasso, Vernon Henderson, Francesco Manaresi, Paolo Sestito and the participants to workshops/presentations at the Bank of Italy, University of Trento, and Free University of Bolzano for stimulating discussions. The views and the opinions expressed in this paper are those of the authors and do not necessarily reflect those of the institution they are affiliated.

1. Introduction

The spatial distribution of population within a country is far from being homogenous. All countries are characterized by core areas, with high levels of income and wealth, and peripheral regions, often specialized in low value added sectors. One of the possible explanations for these patterns is the fact that places are inherently different in terms of productivity; core urban are frequently characterized by the presence of some natural advantages (i.e. first nature advantages: large and deep harbors or the location at the center of large plains with a highly productive agriculture) that eventually evolved in agglomeration economies (i.e. second nature).

However, the distribution of population is also subject to (non-economic) historical shocks that can have long lasting effects (Rosenthal and Ross, 2015; Schumann, 2014). A possible consequence is that the distribution of population is *mislocated*, that is not optimally allocated across areas, and some high-(low)productivity sites end up to be suboptimally under-(over) populated. Spatial mislocation might also have negative effects on the aggregate growth of a country (Hsieh and Moretti, 2015).

The aim of this paper is to analyze the causes and the effects of the mislocation of population in Italy. We consider, in particular, the role of the attacks from the north-African pirates that coastal places (especially in the Southern-West coast) experienced in Italy over two different waves for a very long period of time (i.e. from VIII century with the fall of the Byzantine Empire as a naval power to early XIX century when northern Africa, Tunis in particular, fell under the French influence). Due to the fear of being attacked, coastal places lost their attractiveness for residents who moved towards inner locations, far from the costs and difficult to assault because positioned on mountainous and rugged territories. As a result, these low productivity territories ended up in being relatively overpopulated. These findings are robust to alternative definitions of the security features of the havens and using different measures for the likelihood of being targeted by pirates. More importantly, our results are obtained by using a very detailed set of geographical fixed effects and by controlling for subsoil characteristics, which take into account the productivity features of each location. We also show that the concentration of population over the space was very different before the VIII century: Roman cities were not located in places that ensured protection from the pirates, as there were no raids before the VIII century.

The effects of pirates' attack on the distribution of Italy's population are shown to persist overtime, a feature highlighted in the literature on population changes over space (Rosenthal and Ross, 2015). The impact measured for 1871 (the first year for which a complete census of Italian cities is available) is still evident, tough its magnitude is reduced by a half, in the 1951 distribution of population. The effect persisted notwithstanding two world wars, the exceptional wave of outward migration from the end of the XIX century and the twenties of the XX century. The impact ceases to exist after 1981, wiped away by the massive south-to-north and rural-to-urban migration, which went hand in hands with the Italian industrialization process up to middle 1970s.

We then analyze the consequences of spatial mislocation of population induced by the pirate attacks on a number of economic outcomes. Due to the data constraint, all estimates refer to post-WWII censuses when measured mislocation was quantitatively smaller than the one registered 80 years before. This notwithstanding, we find that overpopulation derived by pirates' attacks in areas less suitable for economic activities determined a slower accumulation of human capital and an over specialization in subsistence agriculture; in the period 1951-1981, escaping from the wrong places also determined a sizable and persistent increase in the aging index with possible long-run consequences on the future development of those areas.

We finally provide evidence of the effects of the historical shock on more aggregate income levels. We show that more mislocated provinces (the equivalent of a county for the US) registered a lower value added per capita in 1951, with persistent effects until 1981. This is partly explained the fact that more affected provinces registered a lower endowment of human capital; we also show that mislocation had also a negative spillover on the provincial urban system by preventing the emergence of major urban centers and, as a consequence, the full development of agglomeration economies.

This paper relates to the literature on the consequences of historical shocks on city development and growth. Bleakley and Lin (2012) show that portage paths in the

Appalachian region stimulated the concentration of people nearby when they were relevant trade routes. However, they remain today - when portage paths do not provide location advantages any longer - places relatively overpopulated. Similar stories are provided by Jedwab and Moradi (2014), with reference to the colonial railroads in Africa, which became quickly obsolete, and Michaels and Rauch (2016), who suggest that French cities were trapped in Roman locations with no costal access. Differently from the previous literature on path-persistence of population settlements, we do not exploit a natural or infrastructural original location advantage, but we refer to the consequences of predatory behaviors like in Nunn (2008) and Nunn and Puga (2012); we also provide evidence of long-run effects not only on population but also on other economic variables.[1]

The paper is structured as follows. Next Paragraph provides the reader with the main historical accounts. Paragraph 3 introduces our baseline estimations and illustrates a number of robustness exercises, intended to probe the validity of our measures for the exposure to the attacks, and a placebo experiment, which consider the Roman sites referring to pre-VIII century. Paragraph 4 looks at the persistence of the effect from the end of the attacks to the period post WWII. Paragraph 5 illustrates the consequences of being trapped in inefficient places for the local economic fortunes. Paragraph 6 provide an attempt to analyze the impact of the attacks at a less detailed level of geographic stratification to show the possible aggregate effects of our historical shock. Some implications of our evidence are discussed in the concluding paragraph.

2. Historical background

From VIII to early XIX centuries Italy's coasts were subject to pirate attacks coming from the shores of Northern Africa. There were two main waves. The first relates to the period VIII-XI centuries; the second started in the XVI century and finished in the early decades of the XIX century. These two waves were inherently different in terms of intensity and purposes.

[1] In the same vein, Schumann (2014) exploits an historical shock to show persistency. In particular, he shows how the relocation of German refugees after WWII determined a permanent increase in population in those areas even after the ban for relocation was lifted.

The first was carried out during the great expansion of Arab domination in Africa and, in part, in Europe. The aim of these attacks in Italy was to create strategic outposts, which may serve as bases for a subsequent conquest by Arab Caliphs (Amari, 1933). Regular raids against Italian shores started after the fall of Carthage (modern Tunisia) under Arab conquerors in 698 AD. Between 703 and 752 Sicily was raided nine times by Arab fleets. An organized invasion of the island started in 827 and was finally accomplished 75 years later. While Sicily was occupied and colonized, in IX and XI centuries raids were registered on the entire Tyrrhenian coast (west coast of Italy) from ships leaving from Sicilian ports. Attacks were mostly aimed at pillaging villages and capturing people to be sold on slave markets but other settlements were chosen as bases for further expansion (Gosse, 1933).[2] Some raids went deep in the interior parts of the peninsula; for example, Arab predators were registered in inner Latium and Umbria along with some areas of western Alps. Attacks on the Adriatic coast were rarer and mostly concentrated in the Southern parts: the most Northern point of an Arab attack on the Adriatic shore was Ancona (848 AD). Arab raids finished after the Christian *Reconquista* of Sicily (1061-1091) by the Normans.

The second wave started at the end of the XVI century; it was not aimed at occupying or colonizing territories but just predate kidnap people for ransom (Colley, 2004). This phase started on 1587 when the Barbary coast (i.e. modern Western Libya, Tunisia, Algeria, and Morocco) fell under the nominal Ottoman sovereignty but was actually governed by local rulers which chose to live by plunder (Encyclopædia Britannica, 1911). Starting ports for raiders were generally Tripoli, Tunis, and Algiers; Konstam (2016) reports that, in 1620s, out of roughly 45 vessels devoted to raids in the entire Western Mediterranean, 34 were based in Tunis, 6 in Algiers and the remaining sailed from Tripoli.[3] Even in these cases main targets in Italy were Sicily and the Tyrrhenian Italian coast, while the Adriatic coast was relatively sheltered by the presence of the Venetian navy. At the end of the XVIII century, Barbary piracy was still considered a

[2] These settlements included southern France (La Garde - Freinet), Northern Campania (Traetto), and Bari.

[3] At the same time Algiers mostly served as the main arsenal for Barbary navy. In 1620s the total number of vessels harbored in this city skimmed 60.

major problem in Western Mediterranean.[4] This called for several military interventions: in several instances British, Dutch, Sardinian, Neapolitan, and (even) US Fleets bombed Algiers, Tunis, or Tripoli. Corsair activity based in Algiers did not entirely cease until France conquered the state in 1830; in the same period both modern-day Tunisia and Morocco fell under the French influence, thus greatly reducing the attacks on Italian shores (Encyclopædia Britannica, 1911).

3. Estimating mislocation

In this section we first show the empirical specification and the baseline results (section 3.1); then we present robustness checks on both the ports of departure (section 3.2) and the geographical features of the units of observations (section 3.3).

3.1. *Baseline specification and results*

The first set of data available on population at the municipality level refers to the period of the unification of Italy.[5] At that time the attacks were over by less than half a century. We start by regressing municipality population in 1871 on a measure for the likelihood for being attacked, the characteristics of the territory that provide protection, and an interaction term between the two.[6] The estimating equation (where m stands for municipality) is the following:

$$(1)\ \log(POP_{1871m}) = \kappa + \alpha \Pr(attack_m) + \beta Protection_m + \gamma \Pr(attack_m) * Protection_m + \sigma X_m + \eta GeoDummies_m + \varepsilon_m$$

[4] For example in 1798 the town of Carloforte (Sardinia) was raided and 900 inhabitants were enslaved and taken as prisoners for 5 years in Tunis (Paoletti, 2011).

[5] Data constraints therefore do not allow us to distinguish between the two waves of the pirate attacks. Our estimates have to be interpreted are reflecting both waves. To the extent that the relocations occurring during the first wave (that ended at the end of the XI century) were completely reversed by the end of the XVI century (were raids started aging with the second wave), our estimates will be capturing only the impact of the second wave only. We believe that it is unlikely, however, that there was a complete reversal, given the results on persistency that we document in the paper.

[6] We also have 1861 data on a limited subsample of municipalities since the process of Italian unification was still incomplete. Results obtained using this subsample are similar to those referring 1871 data (see Section 4).

As for the likelihood of being a target of a raid (Pr(Attack)), we use the inverse geodesic distance between the municipality and the shores of departures. Measuring the probability of receiving an attack with distance is consistent with Nunn (2008) that shows that, with reference to the slave exports, deportations were decreasing in the distance to the final destination. We start by using the distance from Tunis, which is our baseline as the bulk of the attacks departed from there (see Section 2) and then provide robustness by using both alternative measures of being attacked (see Section 3.2). Our index for Protection takes into account several security features of the havens. For attacks coming from the seaside protection is higher for internal and high-altitude locations. Moreover, ruggedness – sloped and irregular terrains – might provide additional defense from being raided (see: Nunn and Puga, 2012). The three elements are combined into a single measure derived with Principal Component Analysis (PCA). However, we also present evidence referring the single components (see Section 3.3). Parameters α and β (main effects) capture, respectively, the influence of geographic location (like, for example, market access) and orographic characteristics on the size of the city.

X_m include a set of city level controls to proxy for the productivity characteristics of a site. We use the subsoil characteristics provided by the European Soil Database (ESDB). ESDB provides very detailed data (1-km-by-1-km) on a number of geological characteristics, we aggregate this information at municipality level.[7] As in Combes et al. (2010), we consider 10 geological characteristics which predict the development of urban centers: Topsoil mineralogy (7 classes), Subsoil mineralogy (6 classes), Parent material hydrogeological type (8 classes), Topsoil available water capacity (4 classes),[8] Subsoil available water capacity (5 classes), Depth to rock (5 classes), Soil profile differentiation (4 classes), Soil erodibility (5 classes), Carbon content (3 classes), and Hydrogeological class (7 classes). In addition, we also use the seismic hazard of the area provided by the Italian Statistical Office that proxies for the likelihood that the housing

[7] Given that soil characteristics are usually discrete, we use the value that appears more often in each area.

[8] This feature proxies also for the presence of marshlands and swamps where malaria was particularly common in some areas of Sardinia, Tuscany, Latium, and Romagna. Until 1950s, malaria was a major determinant for the location of population in Central and Southern Italy

stock could be destroyed and resident population killed by an earthquake.[9] $GeoDummies_m$ include a set of geographical fixed effects. These are aimed at controlling for possible omitted variables (e.g. common economic conditions or institutions) that might influence the location of individuals in certain areas.[10]

Our coefficient of interest is γ, the parameter for the interaction term. It captures the extent to which the concentration in sheltered localities is due to the fear of attacks, rather than by other factors.

In equation (1), pirates-driven overpopulation is captured by the interaction between a location characteristic (inverse distance from Tunis) and an orographic characteristic of the city. Antecedents for this specification can be found in the comparative advantages literature (Rajan and Zingales, 1998; Romalis, 2004; Nunn, 2007; Accetturo et al., 2015). There are two main identification assumptions for a consistent estimate of γ. The first is that the two characteristics are not correlated thus implying that the probability of being attacked is as good as randomly assigned with respect to the orographic conditions. Figure 1 shows that this condition is fulfilled in our data.

[Fig. 1]

The second assumption relates to the reduced form flavor of equation (1). We hypothesize, in particular, that the inverse distance from Tunis is correlated with the (i) probability to be attacked while (ii) it does not determine any other location determinant of population. As for assumption (i) we show that the inverse distance from Tunis is positively correlated with the presence of medieval defense facilities like towers and forts (see Appendix B: *Fortifications and distance from Tunis*) thus implying that, even at a narrow geographical level, threats from security were always coming from the shores of Northern Africa. Regarding assumption (ii) we provide a number of

[9] Seismic hazard is a discrete measure ranging from 1 (high seismicity) to 4 (low seismicity).

[10] An example for the need of these fixed effects is crime rate in the south. After Italian unification southern Italy (i.e. the area that is closer to North Africa) was characterized by widespread peasant revolts (so-called Brigandage) and harsh military repressions until mid-1860s (Accetturo et al., 2016). The need for defense from such attacks might have induced the population to choose more sheltered locations. By comparing Northern and Southern locations without geographical fixed effects, parameter γ could, in principle, capture these omitted variables.

placebo exercises on Mediterranean ports (see Section 3.2), departing points for either pirates or goods and people for trade purposes.

Baseline results are provided in Table 1 (see: Appendix A, Table A1 for the descriptive statistics). In Column 1, we provide the results we obtain from estimating equation (1), omitting to control for territorial fixed effects. This means that the probability of attack actually captures the difference between Southern (high probability) and Northern (low probability) locations. Main effects have the expected sign. The higher the protection granted from an (inhospitable) location the lower the population; cities with a high probability of attack are instead larger implying that southern locations were (in 1871) on average larger than Northern ones. Crucially, the variable of interest is positive and enters with high significance. The standardized coefficient suggests that a s.d. increase of the interaction term is associated with a 37% rise in the s.d. of the municipal population. Obviously, these correlations might capture aspects related to differences across Italian territories that only by chance happen to be correlated with our variable of interest; for example, in 1871 – 10 years after the unification of Italy – differences between southern and northern localities were likely even more pronounced than the current ones (see Felice, 2014), as regions were previously part of different national entities. To control for these aspects Columns 2, 3, and 4 add, respectively a dummy for southern municipalities, regional (NUTS2 in the European Union Nomenclature; 19 areas as Trentino-Alto Adige was not part of Italy in 1871), and provincial dummies (NUTS3; 91 areas roughly of the same size of US counties). In this last specification, which is the most conservative one and it is taken to be our preferred, the provincial-common factors in determining city population size is differentiated away. According to the specification of Column 4, the standardized coefficient on the interaction term is equal to 0.26 and is highly significant.

[Tab. 1]

To the extent that our evidence can correctly be attributed to the attacks of the pirates - occurring from the VIII to the XIX century - the effect of our variable of interest on the spatial distribution of population *before the start of the raids* should be different; in particular, in historical ages in which threats were coming from the North while

Mediterranean shores were relatively safe, we should find that the estimates γ is negative. By using historical data from Pleaides database of ancient places (Bagnall et al., 2016), we are able to select among the 7,152 Italian municipalities, those that were Roman sites. 583 municipalities can be classified as previously Roman spots. As we do not have information on the respective populations, we use as outcome a dummy equal to 1 for the 1871 municipalities with Roman origins and re-estimated eq. (1) by linear probability model. The results depicted in Column 5 of Table 1 show that the coefficient of interest enters with a negative, rather than positive, sign and it is highly significant. This is not surprising as traffics with northern Africa under the Roman Empire were limited to commercial flows, while the Barbarian invasions of the III century followed a path from north Europe to Italy.

Appendix C provides a first robustness experiment, in which we exploit the circumstance that compared to the Adriatic coast, the Tyrrhenian coast was over-targeted by the pirates. In what follows, we focus on illustrating that our proxies for the probability of being targeted and the protective features of the sites are relatively insensitive to the way we measure them.

3.2. *Robustness with respect to the definition of* Pr(Attack)

As explained in Sect. 2, the majority of the boats raised anchor from Tunis both in the first and the second wave. However, other ports in northern Africa were likely involved in pirate's departures. This is the case of Algiers and Tripoli, according to Kontsam (2016). In Table 2 (Column 1 and Column 2) we show what happens to our results if we measure the probability of being attacked by using the inverse distance from Algiers and Tripoli instead of Tunis. As the two alternative sailing paths are highly correlated with our baseline path, this check should be taken *cum grano salis*. The results, in any case, nicely survive. We also calculate Pr(Attack) by using a weighted average of the probabilities of being attacked with respect to these three main ports of departure, with the weights suggested by the historical accounts of Kontsam (2016). Results are depicted in Column 3.

During the period in which boats from northern Africa carried pirates, other ships sailing the Mediterranean. They were peaceful vessels, transporting goods and people for trade purposes (they were also targeted by the pirates). The main ports of departures for peaceful ships were Barcelona and Marseille; since trading with those ports was a source of comparative advantage for less rugged locations, we should observe that, when we measure the probability of receiving an attack by the inverse distance from these two places, the interaction term should be negative and not positive. Results in Column 4 and 5 are very supportive.

[Tab. 2]

3.3. *Robustness with respect to the definition of* Protection

Historical accounts suggest that adequate protective sites were found in inward spaces with a high altitude and on a rugged territory. Our variable Protection therefore considers these three elements jointly through a PCA routine. The first component, which we take to proxy for Protection, is the only one with an eigenvalue larger than one and explains 59.9% of the common variance (see Table A2). It is highly correlated with altitude and slope, while the correlation with the distance to the coast is lower. As Italy has a very assorted territory a relevant question is whether the three elements have to be jointly considered. Table 3 provides the regression results obtained for the baseline specification of Table 1, Column 4 when the single components of our composite index for Protection are used. The coefficients on the interaction enter with very high significance. Note also that our results suggest that altitude plays the biggest role in driving the coefficient thus hinting that altitude alone is generally enough to discourage attacks.

[Tab. 3]

4. Persistency

Table 4 provides the estimation results we obtain by using as dependent variable municipal population measured at Census dates from 1861 to 2001. The results are striking. The impact of the pirates' attacks is still detectable in 1981 (even though statistically significant up to 1971), more than one century and a half after the raids were terminated. The pattern of point estimates we obtain is monotonically (and slowly) decreasing overtime: for instance, in 1951 the effect is estimated to be almost 50% of that measured in 1871. This persistency is noteworthy and consistent with the vast literature on population dynamics across cities (see, Rosenthal and Ross, 2015, for a review). During the period covered by Table 4, the spatial distribution of the Italian population was shocked by the two world wars. In addition, there was an exceptional wave of outward migration from the end of the XIX century and the twenties of the XX century. Finally, at the beginning of the XX century the country experienced a first wave of industrialization (Castronovo, 1995), led by the regions of the north (with related rural-to-urban migration flows). On the other hand, the persistency might have been helped by the overall structural characteristics of Italy's economy from the Unification to the aftermath of WWII, mainly agricultural with a reduced need of migrants, and the prolonged period of Fascist dictatorship, which banned migration (see: Andini et al, 2016). The effect we estimate ceases to exist after 1981. This is not surprising as internal migrations, manly south to north and rural to urban, were a distinctive feature of the Italian industrialization process up to middle 1970s.

[Tab. 4]

5. Effects on local economic development

So far we have consistently shown that, due to the fear of attacks by pirates coming from the Northern Africa, population tended to concentrate relatively more in easy-to-protect locations in regions that were more exposed to raids. Unfortunately, highly defensible locations are also generally less suitable for the economic activity. In general, inner locations have a lower market access; high altitude determines a reduction in the availability of crops; sloped terrains are generally characterized by a lower productivity in agriculture and, once again, by a more limited access to external markets.

This might imply that the fear of pirates' attacks has determined an abnormal growth for locations that were not able to sustain such a large population. In other words, people were mislocated with respect to the actual productivity of such areas.

To test this hypothesis we have to check whether economic outcomes in those areas were worse than those registered in other parts of the country. We estimate the following equation:

(2) $\mathrm{Y_m} = \kappa + \alpha \Pr(Attack_m) + \beta Protection_m + \gamma \Pr(Attack_m) * Protection_m + \sigma X_m + \eta GeoDummies_m + \varepsilon_m$

where $\mathrm{Y_m}$ is now an indicator for local economic development. Data availability constrains our choices in terms of indicators and time span. Information on wages and rents, for example, are available only for the last 15 years for Italian cities but overpopulation was registered only until 1981. From 1951 (the first year when economic censuses are available) our dependent variables are:

- share of employees in agriculture
- number of non-agricultural plants per capita
- human capital
- ageing index
- employment rate

For the sake of simplicity Tables 5 to 9 report the estimated coefficient (and its standardized version) for the interaction term only.

We first analyze whether mislocation had consequences on the productive structure of local economies. Table 5 presents the estimates when we use the share of agricultural employment as outcome variable. Estimation results show that in 1951 overpopulated municipalities were relatively more specialized in agriculture; the magnitude is relevant, with a standardized coefficient of 0.18. This feature persisted and deepened in subsequent years; in 2001 the standardized coefficient rose to 0.37. The fact that high-protection areas are also less suitable for high value-added agricultural production, due to altitude and slope of terrain, implies that those areas mostly lived on subsistence farming.

[Tab. 5]

We subsequently analyze the impact of overpopulation in areas with bad geography on entrepreneurship by using non-agricultural plants per capita as dependent variable. In 1951 the interaction term was negative and significant (-0.12 the standardized coefficient); it grew (in absolute terms) in subsequent years reaching -0.36 in 2001.

[Tab. 6]

Quite interestingly, interaction terms for both agricultural share and plants per capita remain negative and significant even when estimated overpopulation disappears after 1981. This result can be read through the lens of a core-periphery model in a New Economic Geography framework. As explained by Ottaviano and Thisse (2004), industries characterized by imperfect competition, increasing returns to scale, and trade costs tend to concentrate in more densely populated areas (so called, Home Market Effect); the depopulation pattern that we observe in previously overpopulated areas (see table 5) might have determined a rise in the comparative advantage of core areas in those industries. This implies that previously overpopulated municipalities had to rely even more on the agricultural sector and observed a further decay in the number of (non-agricultural) plants per capita. This might explains why the impact survive and get bigger even after the end of the treatment (i.e. overpopulation).

Sluggish economic conditions in affected areas also reduced the incentives to accumulate human capital. Table 7 presents the results by using the share of individuals with at least a secondary school diploma that is the only outcome variable that is consistently available from 1951 on.[11] Still in 1951 overpopulated localities were characterized by a lower average schooling than other areas (-0.09 in standard deviation terms). The effect basically disappears from 1981 probably due to the fact that migration flows in the 1950s and 1960s mostly involved unskilled workers.

---

[11] Data on tertiary education are available from 1971 on with results much in line with those of table 9 (results available upon request).

[Tab. 7]

Outmigration from overpopulated areas also involved a change in the age structure of the staying individuals. Table 8 presents the results by using the Ageing Index (i.e. the share of individuals above 65 years old) as outcome variable. Still in 1951, age structure in affected municipalities was not, *coeteris paribus*, different from other cities. Migration flows that started in 1950s however mostly involved younger individuals thus determining a rise in the ageing index by almost 20% in terms of standard deviation.

[Tab. 8]

We finally check whether overpopulation had an effect on the employment opportunities of local population by using the employment rate (i.e. number of employees over population) as outcome variable. The effect is not ex-ante clear. On the one hand, if wages are downward rigid the effect of overpopulation on employment rate should be negative, since labor market adjustment is on quantities. On the other hand, even in a rigid labor market the effect on employment rate could be zero if there is a marginal sector that can employ all workers that cannot be absorbed in other industries. Results are displayed in table 9 and refer to the years 1951-1991 due to data availability data constrains. For all census waves (except for 1961) the share of employed individuals is not influenced by overpopulation; considering the well-known wage rigidities in the Italian labor market, this results can be rationalized by the presence of a large subsistence agricultural sector that is able to absorb workers in excess.

[Tab. 9]

6. Evidence on aggregate effects

Overpopulation due to the fear of pirates' attacks determined that, as we have seen in the previous section, individuals concentrated in areas that were less suitable for economic development. This implied overspecialization in subsistence agriculture, lower entrepreneurship rate, slower accumulation in human capital, and, in the long run, permanent outmigration of younger population.

Did these features have aggregate effects for the development of the entire country?
The question whether the internal distribution of economic activities fosters or hampers aggregate growth has been widely investigated under a theoretical point of view (Baldwin et al., 2001; Fujita and Thisse, 2002; Baldwin et al., 2002; Accetturo, 2010; Fujishima, 2013). Theoretical models suggest that if knowledge spillovers in innovation are localized, the aggregate growth of a country might benefit from the spatial concentration of innovative activities in core areas; in principle, higher aggregate growth light also be beneficial for peripheral areas. Empirical evidence is scanter (Brülhart and Sbergami, 2009; Gardiner et al., 2010) and supports the idea that the emergence of a core-periphery pattern correlates with a more intense aggregate growth at least at the initial stages of development.

In this section, we quantify the aggregate impact of mislocation on value added per capita. The empirical approach we use is adapted from Hornbeck and Moretti (2015). The idea is that the estimation of aggregate effects of economic and historical shocks should take into account not only the economic impact on directly affected areas but also the spillover on other territories. In our context pirates' attacks not only had depressing consequences on the local economic development of overpopulated municipalities but it might have also prevented the rise of primary cities and – as a consequence – the development of agglomeration economies.

We proceed in two steps. First, we move to a lower level of geographic detail. We calculate – for each province – the average contribution to overall national mislocation (APCM), this is the provincial pervasiveness of mislocation. The higher ACPM, the more municipalities of the province in rugged territories were overpopulated due to the fear of the attacks. In the second step we correlate provincial value added per capita with APCM and we disentangle the main channels at work.

We use the province (i.e. the equivalent of a US county in terms of size) as the unit of analysis following a long tradition in macroeconomics that uses regions or cities as laboratories to understand sources of differences across countries (for example, Barro and Sala-i-Martin, 1991 and 1992; Gennaioli et al., 2013). Obviously, there can be some drawbacks in this approach. For example, province could not be the "right" unit of analysis since the aggregate economic consequences of pirates' attacks might spill over the provincial border. The reason why we concentrate on the province is that it is the only geographic unit for which we are able to construct a reliable series on value added per capita from 1951 (source: Tagliacarne Institute). Moreover, if spillovers are at work, this should attenuate the estimated effects; in other words we are presenting a lower bound estimate of the overall aggregate effect.

6.1. *The construction of the average provincial contribution to overall national mislocation (APCM)*

In order to detect the contribution of each observation in the national average mislocation we use the decomposition of the least squares estimator. Consider a simple univariate regression with one dependent variable (y) and one regressor (x). Gelman and Park (2007) show that OLS estimate can be decomposed as follows:

$$(3)\ \beta_{ls} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_{i<j} (y_i - y_j)(x_i - x_j)}{\sum_{i<j} (x_i - x_j)^2} = \frac{\sum_{i<j} \frac{y_i - y_j}{x_i - x_j} (x_i - x_j)^2}{\sum_{i<j} (x_i - x_j)^2}$$

where $\bar{y}$ ($\bar{x}$) is the sample average of y (x). Equation (3) shows that the contribution of each observation *i* to the OLS estimate is equal to the weighted average of each individual slope between *i* and *j*; weights are equal to the distance (in terms of x) between the two observations. This formula is used to calculate the contribution of each municipality (γ_{mt}) to the estimate of γ in equation (1) for each year $t = 1951, \dots, 2001$.[12]

With γ_{mt} at hand, we are able to compute the contribution of each larger geographical unit (province, regions, etc.) to overall mislocation exploiting the fact that $\sum_p \sum_{m \in p} \gamma_{mp} = \gamma$, where p indicates the province. It should be noted that Italian provinces have different sizes and summing up municipal contributions for each

[12] Regression of equation (1) is multivariate while formula (3) is for a univariate setting. We use the Frisch-Waugh-Lovell (projection) theorem to collapse the multivariate framework in a univariate one.

different-sized province could determine that larger provinces may mechanically contribute more to overall mislocation than smaller ones. To avoid this problem, the indicator that we use is the provincial average (rather than sum) of all γ_m belonging to the same province. In formulas:

(4) $APCM_{pt} = \frac{1}{n_p} \sum_{m \in p} \gamma_{mt}$

where n_p is the number of municipalities in province p.

6.2. *Estimated aggregate economic consequences*

We now turn to the estimation of a correlation between mislocation at provincial level with value added per capita.

As we said the Hornbeck and Moretti (2015)'s idea is that aggregate effects are the weighted average between direct and indirect effects. In our context, as we have seen in section 5, directly affected municipalities registered generally worse economic conditions in terms of human capital, industry mix, and age composition of population. If mislocation is particularly pervasive at provincial level, provincial aggregates may be mechanically (and *directly*) affected by the presence of a large share of "unlucky" locations. However, indirect effects might also be important; mislocation might have had an influence on the entire urban system of an area, favoring the growth of locations that – in absence of the pirates' threat – would have remained small. This means that overpopulation also hampered the development of primary cities and, as a consequence, of agglomeration economies. In these circumstances, the estimation of the consequences of mislocation on income is particularly difficult since usual determinants of value added per capita are influenced by mislocation itself through the direct channel.

Given the presence of such difficulties, we first use a reduced form approach by directly regressing log value added per capita (VA_{pt}) on the pervasiveness of mislocation at provincial level ($APCM_{pt}$) for all census years t=1951, …, 2001:

(5a) $VA_p = \alpha + \beta_{NC}APCM_p + \kappa GeoDummies_p + \varepsilon_p$

where estimated β_{NC} summarize both direct and indirect channels. $GeoDummies_p$ is a set of orthogonal geographical dummies ranging from the NUTS1 (4 dummies) to the NUTS2 (19 dummies) level.

Then, we try to assess the relative importance of direct and indirect channels. We insert a set of controls (X_p) and we estimate:

(5b) $VA_p = \alpha + \beta_C APCM_p + \delta X_p + \kappa GeoDummies_p + \varepsilon_p$

X_p is a set of "bad controls" (Angrist and Pischke, 2009), that is a set of variables that are influenced by $APCM_p$. It includes for the *direct* effects:

- human capital;
- share of agricultural employment;
- number of plants per capita
- employment rate;
- ageing index;

For the *indirect* effect we use a measure of urbanization computed as spatial concentration of population among the municipalities of the same province.[13]

In order to assess the relative importance of each channel we use the omitted variable formula:

(6) $\widehat{\beta_{NC}} = \widehat{\beta_C} + \widehat{\delta}'\widehat{\eta_X}$

where $\widehat{\eta_X}$ is the vector of coefficients from regressions of elements of X_p on $APCM_p$:

[13] We use the Zipf coefficient that is obtained from a regression of $Log(rank_{ip}) = a_p + b_p Log(POP_{ip}) + e_{ip}$. b_p is the value of the Zipf coefficient for province p. It should be equal to -1 if the Zipf law is respected; however, since we are working on the entire population of Italian cities, we expect that it is always different from -1 (Eeckhout, 2004). The Zipf coefficient can be interpreted as a measure of concentration of population across different municipalities within each province. The lower the index the more dispersed is population and the less important is the major urban center in the province.

(7a) $X_{1p} = d_1 + \eta_{1X} APCM_p + e_{1p}$

(7b) $X_{2p} = d_2 + \eta_{2X} APCM_p + e_{2p}$

(…)

(7c) $X_{Kp} = d_K + \eta_{KX} APCM_p + e_{Kp}$

High values of η_{jX} indicate that $APCM_p$ determines control X_{jp}; when this is associated with a large δ_j this implies that control X_{jp} is a relevant channel through which mislocation affects value added per capita.

6.3. *Results on aggregate effects*

The estimation of reduced form effects (equation (5)) are displayed in Table 10. In the upper panel we use NUTS1 dummies as geographical controls, while in the bottom one we use the more demanding specification with regional (NUTS2 dummies). Coefficients for APCM are negative and significant until 1981; after that year the effect disappears. In terms of magnitude a standard deviation increase in the APCM determines a fall of 0.12-0.13 standard deviations of the log value added per capita when we consider NUTS1 dummies. The magnitude is slightly larger (0.17-0.18, at least until 1971) when we use finer geographical units (NUTS2).

The results suggest that mislocation had relevant consequences in terms of aggregate income. Table 11 presents evidence on which channel was more relevant for the more conservative specification with NUTS1 dummies (in this table all coefficients are standardized). When we include controls the point estimate of the coefficient of interest becomes smaller in modulus while the standard error remains roughly stable; in all specification except for 1981 β_C is never statistically different from zero. This indicates that – by including regressors X_p – we are considering the most important sources of variability of APCM in determining the outcome variable. The bottom of the table reports the product between $\hat{\delta}$ and $\widehat{\eta_X}$ coefficients; summing up these products we obtain the difference between $\widehat{\beta_{NC}}$ and $\widehat{\beta_C}$ (see equation (6)). Each product $\hat{\delta}$*$\widehat{\eta_X}$ can be interpreted as the contribution of each channel to the effects of pirates' attacks on provincial value added per capita. The first five lines (share of agriculture, plants per capita, human capital, employment rate, and ageing index) refer to the direct effects; the last is the indirect effect, that is the effect that goes through the urban system. As for the direct effects, the most important channel is human capital and, in more recent years, the employment rate. The urbanization parameter is always important, in particular in the early decades. Urbanization explains away roughly 30% of the differences in value added per capita.

7. Conclusions

In this paper we have shown how historical shocks might have persistent effect on the spatial distribution of population and, in turn, on the economic development of an area. We first presented evidence that, due to the fear of pirates' attacks, population in some areas of Italy (especially south-west) concentrated relatively more in locations that were easier-to-defend but less productive. The result is that those areas were registered worse economic outcomes in terms of human capital and industry specialization; in the long run those cities were characterized by a marked out-migration. We also presented evidence on the fact that overpopulation in low productivity area prevented the emergence of important urban centers with negative effects on aggregate incomes.

These results have relevant consequences from both a positive and normative perspectives.

From the positive side, we have shown the importance of first-nature advantages in shaping the economic outcomes. In the areas affected by pirates' attacks the advantages of overpopulation in terms of agglomeration economies were widely overcome by the disadvantages in terms of productivity; as a result, once the historical event that determined concentration was over, those locations slowly depopulated.

On the normative side, this paper cast some doubts on the economic foundations of many policies for local development aimed at peripheral areas. Many countries have policies aimed at the development of areas characterized by low productivity and weak fundamentals (see Accetturo and de Blasio, 2012, for a review). Italian tradition is particularly strong; quite recently (2012), for example, the Italian government has proposed a project for the development of Internal Areas (Aree Interne) with the aim to resist depopulation and attract economic activities to trigger agglomeration economies.[14] The results of this paper are quite pessimist on the sensibleness of such policies: public resources in those areas are probably not able to overcome the lack of first-nature advantages with possible negative consequences on the development not only at local level but also at aggregate one.

[14] See: http://www.dps.tesoro.it/aree_interne/ml.asp

References

- Accetturo, A, M. Bugamelli, and A. Lamorgese (2016). " Law Enforcement and Political Participation: Italy, 1861-65", Bank of Italy, mimeo.
- Accetturo, A. (2010). "Agglomeration and growth: the effects of commuting costs." *Papers in Regional Science* 89.1: 173-190.
- Accetturo, A. and G. De Blasio (2012). "Policies for local development: An evaluation of Italy's “Patti Territoriali”." *Regional Science and Urban Economics* 42.1: 15-26.
- Accetturo, A., A. Linarello, and A. Petrella (2015). "Legal Enforcement and Global Value Chains: Micro-Evidence from Italian Manufacturing Firms." Bank of Italy, mimeo.
- Amari, M. (1933). "Storia dei Musulmani di Sicilia. 2nd edn, modificata e accresciuta dall’autore, pubblicata con note a cura di CA Nallino. 3 vols." *Catania,* 1933.
- Andini, M., A. Dalmazzo, and G. de Blasio (2016). "The size of political jurisdictions: A model with some evidence from a fascist consolidation." Forthcoming *Journal of Comparative Economics*.
- Angrist, J. and J.-S. Pischke (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton university press.
- Baldwin, R. E., P. Martin, and G. Ottaviano (2001). "Global income divergence, trade, and industrialization: The geography of growth take-offs." *Journal of Economic Growth* 6.1: 5-37.
- Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud (2003). *Economic geography and public policy*. Princeton University Press.
- Barro, R. J., and X. Sala-i-Martin (1991). "Economic Growth in a Cross Section of Countries." *The Quarterly Journal of Economics* 106.2: 407-443.
- Barro, R. J., and X. Sala-i-Martin (1992). "Convergence." *Journal of political*

Economy: 223-251.

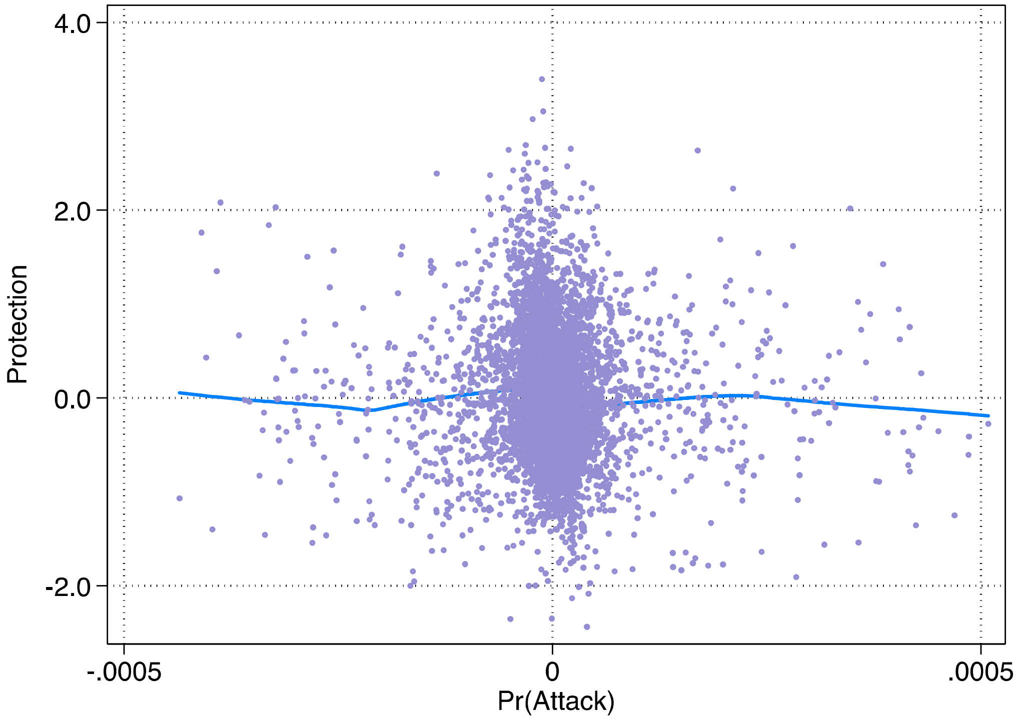
- Bleakley, H. and J. Lin. (2012). "Portage and path dependence." *The Quarterly Journal of Economics* 127.2: 587.
- Brülhart, M. and F. Sbergami (2009). "Agglomeration and growth: Cross-country evidence." *Journal of Urban Economics* 65.1: 48-63.
- Castronovo, V. (1995). *Storia economica d'Italia: dall'Ottocento ai giorni nostri*. Vol. 619. Einaudi.
- Colley, L. (2004). *Captives: Britain, Empire, and the World, 1600–1850*, Anchor Books Edition, New York.
- Eeckhout, J. (2004). "Gibrat's law for (all) cities." *The American Economic Review* 94.5: 1429-1451.
- Encyclopædia Britannica (1911). *Barbary Pirates.*
- Faber, B. and C. Gaubert (2016). "Tourism and Economic Development: Evidence from Mexico's Coastline." No. w22300. National Bureau of Economic Research.
- Felice, E. (2013). *Perché il Sud è rimasto indietro.* Il Mulino.
- Fujishima, S. (2013). "Growth, agglomeration, and urban congestion." *Journal of Economic Dynamics and Control* 37.6: 1168-1181.
- Fujita, M. and J.-F. Thisse (2002). *Economics of agglomeration: cities, industrial location, and globalization*. Cambridge university press.
- Gardiner, B., R. Martin, and P. Tyler (2010). "Does spatial agglomeration increase national growth? some evidence from Europe." *Journal of Economic Geography*.
- Gelman, A. and D.K. Park (2009). "Splitting a predictor at the upper quarter or third and the lower quarter or third." *The American Statistician*, 63(1), 1-8.
- Gennaioli, N., R. LaPorta, F. Lopez-de-Silanes, and A. Shleifer (2014). "Growth in regions." *Journal of Economic growth* 19.3: 259-309.
- Gosse, P. (1933). *The history of piracy*. Tudor Publishing.
- Hornbeck, R. and E. Moretti (2015). "Who Benefits from Productivity Growth?

The Local and Aggregate Impacts of Local TFP Shocks on Wages, Rents, and Inequality." Mimeo.

- Hsieh, C. and E. Moretti (2015). *Why do cities matter? Local growth and aggregate growth*. No. w21154. National Bureau of Economic Research.
- Jedwab, R. and A. Moradi (2016). "The permanent effects of transportation revolutions in poor countries: evidence from Africa." *Review of economics and statistics* 98.2: 268-284.
- Kline, P. and E. Moretti (2014). "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics* 129.1: 275-331.
- Konstam, A. (2016). *The Barbary Pirates 15th-17th Centuries*. Bloomsbury Publishing.
- Michaels, G. and F. Rauch (2016). "Resetting the Urban Network: 117-2012." Forthcoming *Economic Journal.*
- Nunn, N. (2007). "Relationship-specificity, incomplete contracts, and the pattern of trade." *The Quarterly Journal of Economics*: 569-600.
- Nunn, N. (2008). "The Long-term Effects of Africa's Slave Trades." *The Quarterly Journal of Economics* 123.1: 139-176.
- Nunn, N. and D. Puga (2012). "Ruggedness: The blessing of bad geography in Africa." *Review of Economics and Statistics* 94.1: 20-36.
- Ottaviano, G. and J.-F. Thisse (2004). "Agglomeration and economic geography." *Handbook of regional and urban economics*, 4, 2563-2608.
- Paoletti, C. (2011). *La marina sabauda dal 1798 alla restaurazione*. Bollettino d'Archivio dell'Ufficio Storico della Marina Militare.
- Rajan, R. G. and L. Zingales (1998). "Financial Dependence and Growth." *American Economic Review*, LXXXVIII, 559-586.
- Roger Bagnall, et al. (eds.) (2016). *Pleiades: A Gazetteer of Past Places.* <http://pleiades.stoa.org/> [Accessed: September 26, 2016 1:00 pm].

- Romalis, J. (2004). "Factor proportions and the structure of commodity trade." *The American Economic Review* 94.1 (2004): 67-97.
- Rosenthal, S. and S. L. Ross. (2015). "Change and Persistence in the Economic status of Neighborhoods and Cities." *The Handbook of Regional and Urban Economics* 5.
- Schumann, A. (2014). "Persistence of Population Shocks: Evidence from the Occupation of West Germany after World War II." *American Economic Journal: Applied Economics* 6.3: 189-205.

Figure 1. Correlation between Pr(Attack) and Protection



Notes: Authors' calculations on Istat data.

Each dot represents a municipality. The continuous line is the local polynomial estimate. *Pr(Attack):* inverse geodesic distance between the city and Tunis. *Protection*: synthetic measure (first component) of Altitude, Slope of the ground, and Distance to the sea; see table A2. All values in this graph represent the residuals of a regression of, respectively, *Protection* and *Pr(Attack)* on provincial and soil characteristics dummies.

Table 1. The impact of pirate attacks on population, baseline estimates

Dependent variable:	Log population in 1871				Roman settlement
	(1)	(2)	(3)	(4)	(5)
Protection*Pr(Attack)	195.4225***	202.6459***	137.0627***	139.2846***	-27.9725***
	[21.1427]	[21.0033]	[20.7286]	[24.0481]	[10.5431]
Stand. coef.	*0.3682****	*0.3818****	*0.2582****	*0.2624****	*-0.1690****
Protection	-0.3035***	-0.3168***	-0.1964***	-0.1727***	0.0222*
	[0.0291]	[0.0290]	[0.0290]	[0.0353]	[0.0132]
Stand. coef.	*-0.4376****	*-0.4566****	*-0.2831****	*-0.2490****	*0.1028**
Pr(Attack)	114.2786***	183.3233***	424.0311***	127.4538	35.6996
	[25.1349]	[31.2555]	[68.8731]	[152.8810]	[69.1619]
Stand. coef.	*0.0715****	*0.1147****	*0.2653****	*0.0797*	*0.0716*
Constant	5.2748***	5.1393***	5.3054***	6.4190***	-0.6056***
	[0.3631]	[0.3611]	[0.3446]	[0.4136]	[0.1509]
Area Dummies	NO	South	NUTS2	NUTS3	NUTS3
Quality of soil dummies	YES	YES	YES	YES	YES
R^2	0.197	0.198	0.267	0.310	0.120
No. Obs.	7152	7152	7152	7152	7152

Notes: Authors' calculations on Istat, Pleaides and ESDB data.

OLS regressions, see equation (1). Unit of observation: municipality. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Roman settlement: dummy equal to one if the municipality hosted a roman settlement as recorded in the Pleaides database. Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard. Area dummies: "NO" for no area dummies; "South" for a dummy variable for Mezzogiorno (1 dummy); "NUTS2" for regional dummies (18 dummies); "NUTS3" for provincial dummies (91 dummies).

Table 2. The impact of pirate attacks on population, robustness w.r.t. Pr(Attack)

Dependent variable: Log population in 1871	(1) Algiers	(2) Tripoli	(3) Weigh. Avg.	(4) Barcelona	(5) Marseille
Protection*Pr(Attack)	260.7744***	206.7197***	166.3969***	-119.0833***	-50.7595***
	[81.4466]	[40.9039]	[28.7287]	[39.6452]	[11.1488]
Stand. coef.	*0.9153****	*0.7256****	*0.5840****	*-0.4180****	*-0.1782****
Protection	-0.2218***	-0.1721***	-0.1917***	0.1638***	0.1225***
	[0.0767]	[0.0395]	[0.0383]	[0.0506]	[0.0277]
Stand. coef.	*-0.3198****	*-0.2480****	*-0.2763****	*0.2361****	*0.1766****
Pr(Attack)	1688.7352***	-106.6065	172.2871	1446.7306***	491.3656***
	[481.9196]	[442.5013]	[190.1719]	[330.2426]	[85.9952]
Stand. coef.	*1.6821****	*-0.1062*	*0.1716*	*1.4410****	*0.4894****
Constant	4.7909***	6.5998***	6.3777***	4.1062***	4.7218***
	[0.6142]	[0.4843]	[0.4238]	[0.6614]	[0.4924]
Area Dummies	NUTS3	NUTS3	NUTS3	NUTS3	NUTS3
Quality of soil dummies	YES	YES	YES	YES	YES
R^2	0.309	0.310	0.310	0.309	0.311
No. Obs.	7152	7152	7152	7152	7152

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (1). Unit of observation: municipality. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from the localities described in the column. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard. Area dummies: "NUTS3" for provincial dummies (91 dummies).

Table 3. The impact of pirate attacks on population, robustness w.r.t. Protection

Dependent variable: Log population in 1871	Protection: individual components		
	(1) Distance to the coast	(2) Altitude	(3) Slope
Protection*Prob(attack)	0.0028**	0.5572***	0.1749***
	[0.0014]	[0.0881]	[0.0479]
Stand. coef.	*0.0029***	*0.5634****	*0.1768****
Protection	-0.0000***	-0.0013***	0.0001*
	[0.0000]	[0.0001]	[0.0001]
Stand. coef.	*-0.0036****	*-0.9196****	*0.0819**
Prob(attack)	-14.2593	-133.2612	-0.5579
	[159.1022]	[158.2273]	[156.4659]
Stand. coef.	*-0.0089*	*-0.0834*	*-0.0003*
Constant	6.8366***	6.8626***	6.2716***
	[0.4333]	[0.3981]	[0.4032]
Area Dummies	NUTS3	NUTS3	NUTS3
Quality of soil dummies	YES	YES	YES
R^2	0.307	0.322	0.329
No. Obs.	7152	7152	7152

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (1). Unit of observation: municipality. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Pr(Attack) is taken to be the geodesic distance from Tunis. Protection is taken to be the single component of the composite index, as reported in the column. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard. Area dummies: "NUTS3" for provincial dummies (91 dummies).

Table 4. The impact of pirate attacks on population, various Census dates

Dependent variable: Log population in…	Protection*Pr(Attack)	Standard error	Stand. coef.
1861	133.7465***	[25.2386]	0.2526***
1871	139.2846***	[24.0481]	0.2624***
1881	127.5478***	[23.9442]	0.2387***
1901	122.4453***	[24.1949]	0.2235***
1911	118.6349***	[24.5140]	0.2127***
1921	118.2498***	[25.0515]	0.2063***
1931	112.2972***	[24.3042]	0.1905***
1936	108.5154***	[24.3075]	0.1806***
1951	88.2465***	[24.9451]	0.1410***
1961	70.2933***	[25.9056]	0.1070***
1971	57.5309**	[27.3623]	0.0812**
1981	41.1855	[28.2402]	0.0553
1991	23.3705	[28.8204]	0.0304
2001	17.2897	[29.4776]	0.0220

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (1). Unit of observation: municipality. Number of observations: 6325 for the year 1861 and 7152 for all remaining years. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. All regressions include NUTS3 (provincial dummies) and quality of soil dummies. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard.

Table 5. Share of Agricultural employment

Dependent variable: Share of agricultural empl.	Interaction	Standard error	Standardized coef.
1951	2620.9736***	[515.3350]	0.1770***
1961	1917.9856***	[501.1968]	0.1362***
1971	2367.1755***	[464.7167]	0.1925***
1981	4858.0982***	[636.5243]	0.4179***
1991	2508.5271***	[355.8549]	0.3318***
2001	2057.2668***	[264.8592]	0.3658***

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (2). Unit of observation: municipality. Number of observations: 7152. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Share of agriculture: share of employees in agricultural activities. Interaction is the product between Protection and Pr(Attack). Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. All regressions include NUTS3 (provincial dummies) and quality of soil dummies. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard.

Table 6. Non agricultural plants per capita

Dependent variable: Plants per capita	Interaction	Standard error	Standardized coef.
1951	-0.8721***	[0.3325]	-0.1158***
1961	-3.7587***	[1.1517]	-0.0838***
1971	-2.1320***	[0.5202]	-0.2023***
1981	-4.2705***	[0.6862]	-0.2649***
1991	-5.7472***	[0.7480]	-0.3614***

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (2). Unit of observation: municipality. Number of observations: 7152. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Non agricultural plants per capita: ratio between non agricultural plants and total population. Interaction is the product between Protection and Pr(Attack). Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. All regressions include NUTS3 (provincial dummies) and quality of soil dummies. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard.

Table 7. Human Capital

Dependent variable: Share of individuals with at least a secondary school diploma degree	Interaction	Standard error	Standardized coef.
1951	-69.7276**	[34.8979]	-0.0923**
1961	-103.9413**	[44.3348]	-0.1144**
1971	-198.1165***	[68.8105]	-0.1429***
1981	-71.4689	[106.0990]	-0.0321
1991	92.6350	[151.1612]	0.0279
2001	-251.2971	[185.0418]	-0.0652

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (2). Unit of observation: municipality. Number of observations: 7152. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Human capital: share of individuals with at least a secondary school diploma. Interaction is the product between Protection and Pr(Attack). Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. All regressions include NUTS3 (provincial dummies) and quality of soil dummies. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard.

Table 8. Ageing Index

Dependent variable: Ageing rate	Interaction	Standard error	Standardized coef.
1951	44.6751	[84.1381]	0.0178
1961	303.3885***	[110.0071]	0.0891***
1971	914.2975***	[169.2944]	0.1779***
1981	1086.8413***	[237.1077]	0.1561***
1991	1210.4096***	[267.3151]	0.1600***
2001	1689.0567***	[305.5261]	0.2017***

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (2). Unit of observation: municipality. Number of observations: 7152. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Ageing index: share of individuals with more than 65 years old over total population. Interaction is the product between Protection and Pr(Attack). Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. All regressions include NUTS3 (provincial dummies) and quality of soil dummies. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard.

Table 9. Employment rate

Dependent variable: Employment rate	Interaction	Standard error	Standardized coef.
1951	0.0510	[2.2619]	0.0008
1961	-16.0826***	[5.9365]	-0.0430***
1971	2.7033	[2.7718]	0.0249
1981	-1.0467	[3.4048]	-0.0103
1991	5.5968*	[3.3123]	0.0569*

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (2). Unit of observation: municipality. Number of observations: 7152. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Employment rate: ration between number of employed individuals and total population. Interaction is the product between Protection and Pr(Attack). Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. All regressions include NUTS3 (provincial dummies) and quality of soil dummies. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard.

Table 10. Direct effect of mislocation on Value Added per capita

Dependent variable: Log Value Added per capita	APCM	Standard error	Standardized coef.
		NUTS1 CONTROLS	
1951	-1.4090**	[0.6911]	-0.1222**
1961	-1.5664***	[0.5552]	-0.1387***
1971	-0.8809***	[0.3179]	-0.1153***
1981	-0.9714***	[0.2464]	-0.1223***
1991	-0.6033	[0.4098]	-0.0749
2001	-0.5806	[0.4487]	-0.0649
		NUTS2 CONTROLS	
1951	-2.0788***	[0.6985]	-0.1802***
1961	-1.9468***	[0.5975]	-0.1723***
1971	-1.3641***	[0.3650]	-0.1785***
1981	-0.9707***	[0.3454]	-0.1223***
1991	-0.6521	[0.4498]	-0.0810
2001	-0.6255	[0.4403]	-0.0699

Notes: Authors' calculations on Istat and Tagliacarne data.

OLS regressions, see equation (5a) Unit of observation: province. Number of observations: 91. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. APCM is the average contribution of the municipalities of a province to overall Mislocation as estimated in equation (4). See section 6.1 for details. NUTS1 controls: macroarea dummies; NUTS2 controls: regional dummies.

Table 11. Possible channels

Dependent variable: Log value added p.c. in …	1951	1961	1971	1981	1991	2001
β_{NC}	-0.1222**	-0.1387***	-0.1153***	-0.1223***	-0.0749	-0.0649
	[0.0596]	[0.0489]	[0.0414]	[0.0308]	[0.0506]	[0.0499]
β_{C}	-0.0493	-0.0557	-0.0446	-0.0827**	0.0089	-0.0002
	[0.0686]	[0.0415]	[0.0414]	[0.0406]	[0.0390]	[0.0399]
	$\hat{\delta}*\widehat{\eta_X}$	$\hat{\delta}*\widehat{\eta_X}$	$\hat{\delta}*\widehat{\eta_X}$	$\hat{\delta}*\widehat{\eta_X}$	$\hat{\delta}*\widehat{\eta_X}$	$\hat{\delta}*\widehat{\eta_X}$
Share of agriculture	0,0038	-0,0052	-0,0045	-0,0031	-0,0142	-0,0072
Plants per capita	0,0003	0,0002	0,0003	0,0001	0,0000	
Human Capital	-0,0385	-0,0291	-0,0236	-0,0118	-0,0452	-0,0350
Employment rate	-0,0157	-0,0173	-0,0220	-0,0226	-0,0264	-0,0312
Ageing index	-0,0091	-0,0083	0,0064	0,0160	0,0177	0,0220
Urbanization	-0,0137	-0,0232	-0,0273	-0,0181	-0,0158	-0,0131

Notes: Authors' calculations on Istat and Tagliacarne data.

OLS regressions. β_{NC} is estimated with equation (5a); β_C is estimated with equation (5b). $\hat{\delta}*\widehat{\eta_X}$ are the contribution to the difference between β_{NC} and β_C, see equation (6) and the procedure described in section 6.2. Standardized coefficients. Unit of observation: province. Number of observations: 91. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. All regressions include NUTS1 (macroarea) dummies.

Appendix A: Descriptive statistics and Principal Component Analysis

Table A1. Descriptive statistics

	Mean	Standard Deviation
Protection:		
Distance to the coast (continuous, in meters)	68905	55480
Altitude (continuous, in meters)	341.211	284.995
Slope (continuous, in meters)	611.980	628.881
Inverse distance from Tunis (continuous, 1/km)	0.001	0.001
Roman settlement (dummy 0-1)	0.082	0.273
Log(population-1871) (continuous)	7.653	0.877
Log(population-1951) (continuous)	7.928	1.033

Table A2. Principal Components results

	PC1	PC2	PC3
Scoring coefficients:			
Distance to the sea	0.410	0.912	0.009
Altitude	0.646	-0.283	-0.296
Slope	0.644	-0.296	0.703
Eigenvalues	1.797	0.839	0.363
Explained variance	0.599	0.280	0.121

Appendix B: Fortifications and distance from Tunis

In this appendix we show that the inverse distance from Tunis – our proxy for the probability to be attacked – is correlated the presence of an actual military threat. To this aim, we construct a new variable starting from the official list of all name places in Italy. In particular, we construct a dummy equal to one ($Fort_m$) if municipality m has a name place (neighborhood, locality, or the name of the town itself) that hints at the (even past) presence of defense facility. Defensive infrastructures are identified by the following Italian words: *Torre* (Tower and its variation *Tor*), *Forte* (Fort and its variation *Fort*), and *Castello* (Castle and its variation *Castel*).[15] Roughly 15% of Italian municipalities has a name place hint at a defense facility.
We can now estimate the following equation:

(A1) $\mathrm{Fort_m} = \kappa + \alpha \Pr(Attack_m) + \sigma X_m + \eta GeoDummies_m + \varepsilon_m$

where $\Pr(Attack_m)$ is the inverse distance from Tunis, X_m and $GeoDummies_m$ include the same set of regressors of equation (1). If α is positive, the inverse distance from Tunis positively correlates with an actual military threat. Equation (A1) is estimated with a linear probability model.[16] Results are displayed in table A3. When no geographical controls are included, α is not statistically different from zero (column (1)); this is not surprising since – when we consider Italy as a whole – military threats were coming from both North and South. Once we include geographical controls, α turns positive and significant; the result holds – though more imprecisely – even when we include very detailed NUTS3 (provincial) dummies.

[15] For example, Torre a Mare – a neighborhood of the city of Bari – is characterized by the presence of a XVI century Tower.
[16] Using a probit or logit model delivers very similar results, with smaller standard errors.

Table A3. Correlation between fortifications and probability of attack

Dependent variable: City with a fortification name place	(1)	(2)	(3)	(4)
Pr(Attack)	-4.7765	35.8167***	54.2670*	129.4004*
	[9.4756]	[12.4684]	[28.3083]	[67.6421]
Stand. coef.	-0.0073	0.0544***	0.0824*	0.1966*
Constant	-0.5081***	-0.5851***	-0.4799***	-0.5458***
	[0.1263]	[0.1275]	[0.1335]	[0.1678]
Area Dummies	NO	South	NUTS2	NUTS3
Quality of soil dummies	YES	YES	YES	YES
R^2	0.036	0.040	0.075	0.099
No. Obs.	7152	7152	7152	7152

Notes: OLS regressions, see equation (A1). Unit of observation: municipality. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Dependent variable: dummy equal to one if the municipality has a name place (city name, neighborhood, locality) with a fortification name (e.g. Forte, Torre, Tor, Castello, Castel, etc.). Pr(Attack) is taken to be the geodesic distance from Tunis. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard. Area dummies: "NO" for no area dummies; "South" for a dummy variable for Mezzogiorno (1 dummy); "NUTS2" for regional dummies (18 dummies); "NUTS3" for provincial dummies (91 dummies).

Appendix C: Additional robustness on overpopulation

Section 2 describes the exposition of the Italian territories to the raids. The Tyrrhenian coast was massively targeted. The north Adriatic one was instead relatively preserved because of the protection granted by the Republic of Venice, a worldwide naval power. Table A4 presents the results we obtain by splitting the sample between areas targeted and areas mostly unspoiled. As the two areas correspond to the two sides of the Italian peninsula, for this exercise we refer to a bandwidth of 50 kilometers from the coast, to avoid a wrong attribution to untargeted areas of exposed territories. The results for the most affected areas are in Column 1. The regions we include in this experiment are Liguria, Tuscany, Lazio, Campania, Basilicata, Campania and Sicily. The findings for the untargeted areas are presented in Column 2. In this case we include in the estimation sample: Marche, Emilia-Romagna, Veneto and Friuli Venezia Giulia. That is, the coastal areas that relatively sheltered by the Venetian naval power. The results we obtain confirm our insight. No impact is found for untargeted areas, while for the most affected areas the estimated effect is positive and significant.

TABLE A4. The impact of pirate attacks on population: most vs. least affected areas

Dependent variable: Log population in 1871	Most affected	Most affected	Least affected	Least affected
Protection*Pr(Attack)	200.8981***	181.5830***	-482.9688	609.8630
	[40.1046]	[44.4183]	[487.2718]	[489.3105]
Stand. coef.	*0.4150****	*0.3751****	*-0.5818*	*0.7346*
Protection	-0.3512***	-0.2832***	0.6000	-0.6296
	[0.0842]	[0.0944]	[0.6084]	[0.6063]
Stand. coef.	*-0.3298****	*-0.2660****	*0.6492*	*-0.6812*
Pr(Attack)	315.1402***	174.5309	-591.8350	-2512.5732
	[48.6349]	[167.4778]	[647.0759]	[2545.6118]
Stand. coef.	*0.2204****	*0.1221*	*-0.1065*	*-0.4521*
Constant	4.5327***	6.1805***	4.5181***	6.3907**
	[0.5790]	[0.5805]	[1.3713]	[3.0731]
Area Dummies	NO	NUTS3	NO	NUTS3
Quality of soil dummies	YES	YES	YES	YES
R^2	0.206	0.318	0.097	0.246
No. Obs.	1955	1955	623	623

Notes: Authors' calculations on Istat and ESDB data.

OLS regressions, see equation (1). Unit of observation: municipality. Robust standard errors are in parenthesis. * (**) [***] denotes significance at the 10% (5%) [1%] level. Protection is the first PC, as described in Table A2. Pr(Attack) is taken to be the geodesic distance from Tunis. Most affected regions (Tyrrhenian coast): Liguria, Tuscany, Lazio, Campania, Basilicata, Campania and Sicily. Least affected regions (Adriatic coast): Marche, Emilia-Romagna, Veneto and Friuli Venezia Giulia. All observations are within 50 km from the coast. Quality of soil dummies include: Topsoil Mineralogy, Subsoil Mineralogy, Parent Material Hydrogeological type, Topsoil Available Water Capacity, Subsoil Available Water Capacity, Depth to Rock, Soil Profile Differentiation, Soil Erodibility Class, Topsoil Organic Carbon Content, Hydrogeological Class, and Seismic Hazard. Area dummies: "NO" for no geographical controls; "NUTS3" for provincial dummies.