CSEF

Centre for Studies in Economics and Finance

WORKING PAPER NO. 751

Internal Migration, Local Development and Structural Change: Evidence from the Italian Golden Age

Paolo Croce, Matteo Filippi, Paolo Piselli, Andrea Ramazzotti

June 2025

University of Naples Federico II

University of Salerno

Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 751

Internal Migration, Local Development and Structural Change: Evidence from the Italian Golden Age

Paolo Croce[*], Matteo Filippi[†], Paolo Piselli[‡], Andrea Ramazzotti[§]

Abstract

Internal migration facilitates an efficient allocation of labor within the economy, but are its sending and receiving areas affected differently? We address this question through the lens of Italy during the Golden Age (1950s-1970s), a period of population reshuffling with no parallel in the country's history. Exploiting detailed spatial data on migratory flows, we characterize the impact of short- and long-distance migration on economic development and structural change in the provinces of origin and destination. To tackle endogeneity of migration flows, we build on recent advances in the shift-share IVs literature: we interact past interwar government-authorized migrations with employment growth during the Golden Age to estimate exogenous short-distance migrations; origin-destination railway distances with provinces' employment growth for long-distance ones. We find that short-distance emigration negatively affected origin provinces' value added per capita mostly through lower business creation and productivity, while it determined even larger productivity gains in destination provinces. Similarly, short-distance immigration boosted structural change away from agriculture in favor of the industrial sector, while emigration curbed it in the provinces of origin, by reducing employment, value added and productivity in industry. We do not find comparably strong results for long-distance flows, which are shown to negatively affect origin provinces mostly through the employment rate, while the effects on productivity are limited; receiving provinces are also not as affected. We attribute the difference between short and long-distance effects to selection by type of migrants, where the most productive ones tend to favor nearby destinations.

JEL Classification: J61, N14, O12, O15.

Keywords: Internal migration, Regional development, Economic growth.

Acknowledgments: This work benefited from comments from Federico Barbiellini Amidei, Matteo Gomellini, Gaetano Basso, Mauro Rota, the participants in Ninth ASE Annual Meeting (12-13 Sept 2024, Turin), University of Zurich PhD workshop and the EHS Annual Conference 2025 (4-6 Apr 2025, Glasgow, UK). Any remaining errors are ours. The views expressed in the article are those of the authors and do not involve their affiliation Institutions.

[*] Bank of Italy, Economics, Statistics and Research DG, Economic History Division. Email: paolo.croce@bancaditalia.it

[†] University of Zurich. Email: matteo.filippi@econ.uzh.ch

[‡] Bank of Italy, Economics, Statistics and Research DG, Economic History Division. Email: paolo.piselli@bancaditalia.it

[§] CSEF, University of Naples Federico II. Email: andrea.ramazzotti@unina.it

1 Introduction

Even though migration has traditionally received a great deal of attention in the economics literature, most of the research has focused on international migration, looking in particular at the labor market impacts on destination areas. In comparison to this large literature, the economic effects of internal migration are relatively less studied, even though there is consensus that internal and international migration are likely to differ both in terms of determinants and impacts. The costs of migrating internally are lower than the costs of international migration in terms of geographical, cultural, and institutional distances; this affects the number and characteristics of migrants and allows migration to respond more quickly to labour market conditions (Kleemans and Magruder, 2017; Jia et al., 2023; Franceschin and Görlach, 2024). In particular, internal migration facilitates an efficient allocation of labor within the domestic economy, which allows to absorb local shocks, but what are its effects on the sending and receiving areas? Is internal migration growth-enhancing or does it delay development, and what are the implications in terms of structural change? Despite the importance of these questions, very limited reduced-form evidence exists to-date linking internal migration to growth and the process of structural change, possibly also due to the difficulty of finding suitable empirical strategies.

The present work aims to fill this gap in the literature by studying the role of internal emigration and immigration rates as possible drivers of local (under)development and structural transformation. Our empirical analysis focuses on Italy during the decades of the Italian Golden Age (1950s-1970s), an ideal setting to study this phenomenon due to the contemporaneous substantial surge in internal migration rates, economic growth and structural change. Between 1946 and 1975, approximately 42 million people where involved in residence transfers between municipalities, a population reshuffling with no parallel in the Italian history. This trend was not limited to short-distance and temporary residence: for instance, the population living in a different macro-region from their place of birth increased from 2.6 to 5.6 million between 1951 and 1971 (Bonifazi 2013, p.191).[1] In parallel, Italy experienced an average annual rate of GDP growth of 5.1% between 1951 and 1973 (Baffigi, 2013); over this period, the Italian rates of growth were only second (but very close) to West Germany in Europe, and to Japan among the OECD countries.

Motivated by this evidence, the paper aims to investigate the causal effect of internal migration on the economy of origin and destination provinces, through the analysis of a large set of economic indicators related to economic growth (value added, productivity, employment, firm dynamics) and structural change (sectoral shares of value added and employment, sectoral productivity). However, estimating the causal effect of migration on

[1]There are five statistical macro-regions in Italy: North-West, North-East, Centre, South, and Islands.

local economic outcomes is challenging, mainly due to selection and omitted variable bias. People may be more likely to leave provinces with lower economic activity, and several omitted factors pushing people to emigrate may also affect at once economic activity in the province of origin. Likewise, people might immigrate to provinces with faster income and productivity growth, which might correlate with similar omitted factors.

Although some literature has overcome this difficulty by studying episodes of exogenous or quasi-exogenous migration such as moves in response to natural disasters or wars (e.g. Ager et al. 2017; Deryugina et al. 2018; Becker and Ferrara 2019), the vast majority has relied on instrumental variable strategies. In this paper, we follow the latter approach and adopt a shift-share empirical design, which is increasingly used in the research on international migration (Card 1990, 2005; Anelli and al., 2023) but also internal migration (Boustan et al, 2010; Kleemans and Magruder, 2017; Imbert et al, 2022). However, compared to the previous literature, we build on recent advances in the literature on shift-share designs (Adao et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022) and characterize the effects of internal migration according to the spatial extent of migration flows.

The first main source of data we exploit is migration tables derived from annual publications of the Italian National Statistical Institute (Istat), recording changes of residence at the province of origin - province of destination level. These data were originally collected and digitized by Ramazzotti (2024) for 1961-1981, which we extend and harmonize to include the years 1955-1960. The bilateral nature of these data crucially allows us to i) consider how the spatial extent of migration flows mediates their effects on origin and destination areas, and ii) use different empirical designs for different types of internal migration. In particular, we distinguish between short and long-distance internal migrations, noting that the two differ qualitatively with respect to the types of migrants and their motivations (Jia et al., 2023). In our baseline results, we define as short-distance all migrations that happened within the same macro-region—i.e. between provinces in the same region or across regions in the same macro-region—, and we conversely define as long-distance all migrations that happened across macro-regions. This distinction is commonly used in the Italian literature on this topic (Biagi et al., 2011; Bonifazi, 2013), which has dedicated much attention to long-distance flows—particularly to those from the South to the North of the country. Short-distance migration, instead, has received relatively little attention, despite the fact that two thirds of the Italian internal mobility between 1955 and 1975 happened within the same macroarea. Our paper addresses this gap in the literature by separately analysing both types of migration flows, and finds unexpected heterogeneity in their local effects.

In order to construct an instrument for short-distance migration flows, we newly digitize historical records of workers' migration within Italy for the period 1929-1938 as explicitly authorized by the Fascist government. These data are derived from annual publications of the *Commissariato per le migrazioni e la colonizzazione interna*, a special commission established by the Fascist regime with the aim of regulating the internal reallocation of the labor force. Due to the peculiarities of the Fascist rules about internal migration, these migrations flows exhibited a spatial reach which was mostly local; they were directed to different destinations than flows to the fast-growing provinces of the North, in the government's hopes of favoring the so-called "return to the land"; finally, these migration flows were meant to induce future permanent moves of workers by temporarily exposing them to new destination provinces. Thanks to these characteristics, we can interpret these data as an appealing source of quasi-exogenous Intention-to-Treat (ITT) variation in later local migration flows during the Italian Golden Age, in line with Goldsmith-Pinkham et al. (2020)'s IV framework, related to the exogenous resettlement policies discussed for instance in Jia et al. (2023) or Becker and Ferrara (2019). To operationalize these ideas, we exploit the bilateral structure of the data and compute emigration shares of any given province to each destination; we then construct our instrument by interacting these exposure shares with pull and push factors in nearby provinces.

For long-distance migration, we adopt a similar approach based on the observation that most of the long-distance migrants during the Golden Age used to travel by train, implying that railroad connectivity is likely to be a good measure of provinces' exposure to each other. We therefore digitize the railroad network at 1955 as derived from maps published by the Italian Ministry of Transports. Next, we construct exposure shares based on the railroad distance between province capitals and interact these shares with push/pull factors. By weighting more a pull/push factor of a relatively more distant province according to the railroad network, we both i) match the empirical evidence of large migration flows towards the fast-growing and far-located cities of the North and the Center (relevance), ii) use identifying variation which is likely exogeneous with respect to any given province of interest (exogeneity). The idea of the instrument is to give more weight to those provinces which are less likely to be subject to the same local shocks. As opposed to our short-distance instrument, whose identifying variation is based on the shares, we stress that for long-distance migration the identifying variation comes instead from the shocks, in line with Borusyak et al. (2022)'s framework.

We find substantial evidence of heterogeneous effects across long and short-distance migration. With respect to short-distance flows, we find that internal migration negatively affected origin areas' income, mostly through lower productivity, employment rate and business creation. Our estimates suggest that a 1% increase in short-distance emigration reduces value added per-capita by 0.29% in origin provinces, almost entirely due to a 0.17% drop in productivity and a 0.14% drop in the employment rate; in parallel, the number of firm establishments per-capita shrinks by 0.19%. However, in destination provinces productivity gains are positive and larger: a 1% increase in short-distance immigration increases value added per-capita by 0.39%, which is almost entirely driven by a 0.35% increase in productivity. A sectoral analysis also reveals that overall short-distance migration boosted structural change away from agriculture in favour of the industrial sector. On one side, short-distance emigration negatively affect the share of value added and employment (both -0.3%) in industry in origin provinces; positive effects of the same magnitude are shown by immigration on the share of industrial value added in destination provinces (+0.31%). On the other side, emigration strongly reduces productivity in the service sector at origin, while immigration strongly increases industry and services productivity in destination provinces: we estimate that a 1% increase in short-distance immigration leads to a 0.52% productivity boost in industry, and a 0.31% productivity increase in the service sector at destination.

We do not find comparably strong results for long-distance flows. A 1% increase in long-distance emigration only reduces value added per-capita by about 0.1% in origin provinces, a magnitude equivalent to a third of the short-distance effect; moreover, this effect is mostly driven by the employment rate and the number of firm establishments per capita, with very limited effects on productivity. Additionally, receiving areas are almost unaffected, except through the demographic dividend. In summary, we provide evidence that long-distance emigration contributed mostly on the *extensive* margin of economic growth, while short-distance emigration impacted origin and destination provinces on the *intensive* margin (i.e. on productivity). We argue that our results are consistent with a model of distance-based selection of emigrants, where the most productive individuals tend to favor nearby destinations possibly based on a preference for proximity to origin areas, as in Monras (2020).

We finally confirm our baseline results with a number of robustness exercises. First, we test for robustness of our main results additional controls and to changes in the definition of the exposure shares; then, we relax our distinction between short and long distance migration, looking at heterogeneity in effects by distance thresholds. In the Appendix, we further expand our battery of robustness checks, as aplpy different clustering strategies, we allow for alternative transformations of our push and full factors, we check for exposure robust inference, and we formally test for the exogeneity of our instruments: exogeneity of the Fascist migration shares with respect to provinces' subsequent development, following Goldsmith-Pinkham et al. (2020) in the short-distance migration case; in the context of long-distance migration, exogeneity of the growth rates of employment in destination provinces, relying on the framework provided in Borusyak et al. (2021). The Appendix also provides some additional historical background.

There are three main areas in which this paper contributes to the literature. The first area is the economics literature on the effects of migration flows within the same country.

Potentially due to the intrinsic difficulty of identifying causal effects both on the areas of origin and destination, most contributions in this area belong to the literature on ex-ante evaluation of policies using structural econometric models.[2] Within this literature, our work is closest to Hao et al. (2020), who structurally quantify the contribution of internal migration to growth and structural change in China between 2000 and 2015 and find that internal migration accounted for the majority of the reallocation of workers out of agriculture.

Within the reduced-form literature, most of the research on migration has typically focused on international migration (e.g. Card and di Nardo, 2000; Borjas, 2003; Card, 2009). Most of the existing literature on internal migration has looked at developing countries (Maystadt et al., 2016; Kleemans and Magruder, 2017; El Badaoui et al, 2017; Imbert and Papp, 2020) or at the US (e.g. Wozniak, 2010; Molloy et al., 2011), mainly studying the labor market impacts in terms on employment and wages.[3] Overall, we are not aware of reduced-form contributions linking internal migration, local development, and measures of structural change. A paper close to ours is Imbert et al. (2022), who also use a similar shift-share design to study how rural-urban migration shapes urban production structure, and find that immigration induced an increase in labor-intensive production and a productivity decline. However, due to the nature of their data, they can only look at effects within the manufacturing sector; moreover, the scope is limited to rural-urban migration and receiving areas, making it hard to assess the contribution of internal migration to sectoral and geographical reallocation of economic activity. Compared to this paper, we employ a richer set of data to estimate the effects of internal migration on local economic development and sectoral reallocation, looking both at sending and receiving areas.

The second area of research concerns the historical literature of Italy's internal migration and economic development during the Golden Age. According to the historiographical consensus, mass internal migration was a prominent characteristic of Italy's fast economic transformation between the 1950s and the 1960s (Ginsborg 1990; Zamagni 2003; Fauri 2015; Gomellini et al. 2017; Ciocca 2020). General historical overviews agree that the movement of workers from rural to urban areas—particularly from Southern

[2] The papers' range include looking at the role of self-selection in returns to education (Dahl, 2002), income prospects at destination (Kennan and Walker, 2011), the interaction between capital and labor to explain convergence across US states (Kleinman et al., 2023), skill sorting and the college/school wage gap (Diamond, 2016), barriers to workers' relocation (Heise and Porzio, 2019; Schmutz and Sidibè, 2019; Tombe and Zhu, 2019), childhood location and adulthood economic outcomes (Ishimaru, 2024), welfare effects of migration (Ma and Tang, 2020), and the absorption of low-skilled immigration shocks (Monras, 2020).

[3] Boustan et al. (2010) studied the effects of internal workers' relocation on local labor markets during the Great Depression, Cadena and Kovak (2016) the responses of Mexican-born immigrants to local demand shocks, Blackburn (2010) the effects of internal migration on the earnings of married couples, Derenoncourt (2022) the effect of the Great Migration on the gains from growing up in specific locations for Black families.

regions to the main manufacturing centres of the North—fostered the expansion of the industrial sector and thus underpinned structural transformation and productivity growth (Eichengreen 2007, pp. 112-118). It is often argued that migrants offered cheap manual labor which eased pressure on the labor market. This translated into wage moderation that allowed firms to reinvest profits for capital accumulation, benefiting productivity growth (Eichengreen 1996; Graziani 2000; Borjas 2003). Migration from backward areas would also reduce rural underemployment without loss of productivity for the agricultural sector (Cafiero and Marciani 1972, pp. 276-278; Pugliese and Rebeggiani 2004, pp. 52-60). Due to the large regional differences in income between rural and urban areas, this process of structural transformation would also affect spatial convergence, an implication that appears particularly important given the historically dualistic nature of the Italian economy (Bodo and Sestito 1991, pp. 110-114, Gomellini and Ó Gráda 2011). Despite the preminent role of internal migration in the Italian historiography, we are not aware of quantitative estimates of its impact on local economic development. Hence, our paper complements the historiography by offering the first identification of the effects of internal migrants on the structural transformation of Italian provinces during the 1950s-1970s.[4]

The third area we contribute to is the empirical literature on shift-share IV methods. We build on recent innovations in the literature (Goldsmith-Pinkham et al. (2020); Borusyak et al., 2022) to provide two competing identification procedures for short and long distance migration, exploiting exogeneity of past migration shares in order to build the instrument for the former and the exogeneity of destinations' push/pull factors (weighted with railway distances) in order to compute the instrument for the latter. To the best of our knowledge, we are the first to combine these two competing empirical designs within the same paper.

We also marginally contribute to a literature studying the link between transportation networks and growth (Banerjee et al., 2020; Donaldson, 2018; Duranton and Turner, 2012). In particular, we draw on Sequeira et al. (2022) who use a county's connection to the railway network to instrument for the number of immigrants that settled in the county. Our work also relates to a small subset of papers studying the effects of migration on origin areas (Dustmann et al., 2012; Elsner, 2013; Giesing and Laurentsyeva, 2017; Anelli and Peri 2017; Dicarlo 2022; Dodini 2022; Anelli et al. 2023).

The rest of the paper is organized as follows. Section 2 reviews the literature on Italy's economic development and internal migration during the Golden Age. Section 3 describes in detail data and sources. Section 4 introduces the empirical specification and the identification strategy, while Section 5 presents our estimates and the main results.

[4] Within economic history outside of Italy, much attention has been given to the topic of internal migration for market integration and wage convergence, but not so much to the effects on local productivity, and the overall impact on aggregate growth. See for instance Boyer and Hatton (1997), Collins (1998), Rosés and Sánchez-Alonso (2004); Boustan et al (2010); Enflo et al. (2014).

Section 6 reports the main robustness checks while Section 7 concludes. Further robustness checks, analyses of instruments validity and details on the historical background are provided in the Appendix.

2 Internal migrations during the Italian Golden Age

Between 1950 and 1973, in the "Golden Age" of the Italian Economy, per-capita GDP grew annually by 5.3 percent, industrial production by 8.2 percent, labor productivity by 6.2 percent, while the employment share in agriculture dropped from $42,2\%$ to 17%, as the country was becoming a modern industrial power (Toniolo, 2013). Within the same period, per-capita real income tripled, life expectancy increased from 63 to 69 years, and the share of households owning a refrigerator and a washing machine skyrocketed from 3% and 1% to 94% and 76% respectively.

In parallel, Italy witnessed very large internal migration flows, which responded to and provided support for the structural transformations of the economy. The flows of migrants were mostly concentrated in fast-growing industrial areas in Northwestern Italy (the "industrial triangle" of Milan, Turin and Genoa) and Rome, coming from adjacent provinces and regions, but also from the regions of Northeastern Italy and the South. Such large flows are commonly attributed to the creation of new jobs in the industrial sector (North) and public sector (Rome), proving a very strong pull factor for low-skill individuals in backward areas.[5]

However, it is widely accepted that both internal and international migration was aided by institutional changes. With respect to migration outside of Italy, the post-WWII period was underscored by a gradual reopening of the international economic relations, which also affected migration opportunities: first, Italy signed bilateral agreements with European countries in need of workers for reconstruction and industrial expansion; later, the creation of the European Economic Community further reduced legal barriers to free movement (Corti 2003, ch. 5). With respect to internal migration, the democratization of the political sphere and the liberalisation of the economy eased formal restrictions to movement. In particular, the "anti-urbanisation laws" which had been established by the Fascist government (see next section) were first dis-applied and then abolished (Tortorici 2023). Thus, institutional change in the post-war period helped relax formal barriers to internal migration.

At the same time, changes in the country's transport infrastructure reduced costs to move. Maggi (2003, pp. 168-69) notes that both international and national migration was largely carried out through the railway: new *direttissimi* trains connected Rome to

[5] After all, Southerners and Italians in general were particularly receptive to distant job opportunities, as it had been the case during the Age of Mass Migration (abroad) and again in the post-WWII period, when migration to northern European countries reached historical highs.

the main European destinations with relative comfort, taking passengers directly to the places with the highest demand for labor; internally, the *Treno del Sole* and the *Freccia del Sud* linked Southern cities with Genoa, Turin, and Milan—the country's industrial hub. These direct trains made travelling not only cheaper, but also easier, reducing the perceived distance from home and thus decreasing both the monetary and non-monetary costs of migration.

Due to these factors, migratory intensity increased in the 1950s, exceeded the threshold rate of 30 per thousand since 1960, and peaked in 1962, reaching 43.2 per thousand, equivalent to 2.2 million transfers (Bonifazi and Heins, 2000).[6] Using our migration data, Figure 1 displays emigration rates over our study periods based on the origin-destination distance. Not surprisingly, most migration moves were confined within the same province; however, inter-areas migration flows were also substantial, peaking at 30.6% in 1963 (Bonifazi and Heins, 2000). Figure 2 shows that the most quantitatively relevant migration flows were directed from the South and the Islands to Turin and Rome during the 1960s, and to Milan during 1970s. However, despite particular salience within the public debate,[7] limiting the discussion of internal migration only to South-North flows is a reductionist approach. First, not all long-distance migration was pulled by the 'industrial triangle' and by jobs in manufacturing, with a sizeable number of migrants remaining employed in traditional sectors (agriculture, construction) and replacing native workers who transitioned to industrial jobs (Grilli and Zanotelli 2015, pp. 60-67). Second, according to our data, while 55% of migration flows took place between macro-regions, 45% of migrants moved within the same macro-region, and 31% within the region of origin. In particular, Table 1 shows the emigration rates within and across the areas and confirms the importance of short-distance migration: for all macro-regions but the South, the within-area migration rate is higher than the flows outward, peaking at 10.04 per 1,000 inhabitants per year in the North-West between 1955 and 1961. Third, these flows differed qualitatively from long-distance, South-North migration: while South-North moves were mostly directed from poorer/agriculture-based to richer and industrializing areas, short-distance migrants were often qualified workers looking for higher-skill (white-collar) jobs or entrepreneurial and professional opportunities in province capitals and larger urban areas (Gallo 2012, pp. 134-155). Overall, even though long-distance migration was a salient phenomenon with sizeable social and cultural repercussions, it represented only one aspect of a much broader spatial reallocation of the Italian population, which modified traditional urban hierarchies and led to a greater concentration of people in the fastest-growing places (Ramazzotti 2021, pp. 146-171). Hence, for the purposes of our

[6] The peak year however also reflects the effects of post-census adjustments and the repeal of the fascist law on urbanisation in 1961 (Tortorici, 2023).

[7] The image of the young, typically male Southerner leaving his ancestral rural home for a big city in the North is stereotypical of the time and has been reinforced by novels, films, music and other popular media (Pizzolato, 2012).

analysis, we will distinguish between short and long-distance migration flows, depending on whether migrants moved within or across macro-regions.

What role did internal migration play for economic growth in these years? Few quantitative analyses are available. In Toniolo's (2013) Handbook of Italian Economy, there is some analysis of international migrations (Gomellini and Ó Gráda 2013), but the topic of internal migration is not addressed. In the chapter by Iuzzolino et al. (2013) on regional convergence, a simple accounting exercise shows that internal migration from the Mezzogiorno to the Centre-North contributes to the convergence of GDP per capita, by reducing population in the former and increasing it in the latter.[8] Cappelli et al. (2018) carried out a standard decomposition of regional GDP per capita into labor productivity and a demographic component (employment to population), in turn divided into participation rate (employment to working-age population) and the age structure (working-age population to total population). Their analysis shows that the demographic component, including changes in migration flows, remained largely stable for a large time span until 1971 as a consequence of a divergence in the participation rate and a reduction in the age structure dispersion. Hence, the authors conclude that internal and outward migration helped the South to catch up with North income per capita , but as an effect of a reduced population, while the influx of labour force to the Northern regions "offset the structural change that was taking place in the South, thus partially reducing the positive impact *[on regional divides]* of increased capital-to-labour ratios brought about by the government's regional industrial policy" (Cappelli et al. 2018, p. 19). In the same vein, Felice (2019) also partly questions the role of internal migration from South to North, suggesting that the reduction in participation rates in the South in the period might be a consequence of emigration, which drew away the most productive labour force. However, we are not aware of systematic studies quantifying the contribution of internal migration to local economic development and structural transformation.

3 Data and Sources

The analysis is performed on data that combines and extends multiple historical datasets on internal migration flows and economic outcomes at the province level from the 1950s to the 1970s. This section describes the sources and methodologies applied to create the variables used throughout the analysis.[9]

[8]Gomellini and Ó Gráda 2013 extend their analysis of international migrations to internal migration in Italy in a short footnote and their very preliminary findings suggest that internal migration was a win-win game. Migrants would have stayed unemployed in the regions of departure while they contributed to productivity growth in the regions of destination.

[9]The number of Italian provinces has slightly changed over the study period. In order to keep the sample as homogeneous as possible, we work with an homogeneous sample of 91 provinces. Hence, we consider Pordenone (created in 1974) as part of Udine, Isernia (created in 1970) as part of Campobasso

3.1 Migration rates

We define as internal migrants all individuals that canceled their residence status in one province and registered as resident in another Italian province during the solar year. This is a common definition in Italy's demographic and historical literature and is coherent with most reconstructions of internal migration flows (see for instance Bonifazi and Heins 2000; Piras 2017). Despite sharing the common limitations of administrative sources (e.g. missing unreported and temporary migrations, observing migrations with a lag due to the administrative process), changes in residence status are the most systematic and comprehensive source to track changes in the spatial distribution of Italy's population with annual frequency (Istat 2011: 82-83) and remains to-date the main source used by Istat to quantify internal migration flows.[10]

Our primary sources for the number of internal migrants are statistical publications by the Italian National Statistical Institute (Istat). Istat regularly collected information on the changes of residence status from the municipalities' civil registries and published the resulting statistics at different levels of geographical aggregation and detail. For our analysis, we use the tables of cancellations and registrations at the province level, which report the number of migrants between all couples of provinces in matrix format. This data format, available from 1955 on, allows us to reconstruct all migration flows (both short and long distance) by origin and destination.[11] The tables regarding the years from 1961 to 1981 were originally collected and digitized by Ramazzotti (2024) while the tables regarding 1955-1960 were specifically collected for this project and harmonized with the existing dataset.[12]

We exclude from the analysis all changes of residence taking place within the same province (i.e. the diagonal of the matrix), for two reasons. Methodologically, it would not be possible to build a shift-share instrument for internal migration following what we do with other flows. Secondly, mixing migration within province with out-of-province flows might bias estimates due to differences in provinces' extension and degree of spatial concentration in economic activities (as provinces with greater extension and/or more spatially concentrated activities would plausibly show a higher level of internal migration).

and Oristano (created in 1974) as part of Cagliari. We exclude also Caserta, created with the present borders in 1945, to make our cross section consistent with data on internal migration from the 1930s.

[10] See for instance the 'internal migrations' section of the official Istat website on Italy's demography: https://demo.istat.it/ (last retrieved 11/01/2024).

[11] In 1955, Istat started producing a new statistics of population movement, based on a new Civil Registry Regulation, which established the contemporaneity of the registration and the cancellation due to transfer of residence. This allowed start collecting data on migratory exchanges between various territorial realities and thus have a more precise vision of mobility processes (Bonifazi 2013, p.190).

[12] The tables have been digitized from the following ISTAT sources: *Annuario di statistiche demografiche* (1955-68, 1980-81); *Supplemento straordinario al Bollettino Mensile di Statistica n. 8 - Agosto 1972 – Dati sommari sulle statistiche demografiche* (1970); *Popolazione e movimento anagrafico dei comuni* (1969, 1971-1979).

In our main analysis, we use these data to compute the province-specific emigration (immigration) rates to (from) long and short distance provinces. More formally, denoting by S the set of all provinces and $R(j)$ the macro-region province j belongs to, we can define for each province p

$$S_p^{sd} = \{j \in S : R(j) = R(p)\} \quad \text{and} \quad S_p^{ld} = S \setminus \{j \in S : R(j) = R(p)\}$$

the set of all provinces which are located at short and long distance with respect to p, respectively[13]. In the case of emigration, we next compute the total number of emigrants from each province p to short and long distance provinces in each year τ as

$$emig_{p,\tau}^{sd} = \sum_{d \in S_p^{sd}} emig_{p,d,\tau} \quad \text{and} \quad emig_{p,\tau}^{ld} = \sum_{d \in S_p^{ld}} emig_{p,d,\tau}$$

Finally, we cumulate the total number of emigrants from p over each decade, and re-scale by the population at the start of the decade (per 1,000 people), in order to get the short and long distance emigration rates $em\_rate_{p,T}^{sd}$ and $em\_rate_{p,T}^{ld}$ in the decade preceding each year $T \in \{1961, 1971, 1981\}$[14]:

$$em\_rate_{p,T}^{sd} = \frac{\sum_{\tau=T-k}^{T} emig_{p,\tau}^{sd}}{pop_{p,T-10}} \quad \text{and} \quad em\_rate_{p,T}^{ld} = \frac{\sum_{\tau=T-k}^{T} emig_{p,\tau}^{ld}}{pop_{p,T-10}}$$

We follow the same methodology to compute the number of immigrants from short and long distance provinces of origin, and compute short and long distance immigration rates in a similar manner as above.

Nonetheless, it is important to note that during the earlier period of our analysis, changes to residence status continued to be regulated by laws that had been passed by the fascist government in the 1930s in an attempt to stifle internal migration. Whilst the effectiveness of these laws was probably limited, they affected the reliability of the official statics as a possibly large share of internal migrations went unreported. The fascist

[13]Note that in the current iteration of the project $D = 91$ for we exclude the province of Caserta, as it is missing from the computation of the instrument (see section 4). When we perform the analysis on long (short) distance migration we restrict the computation only to destinations outside (inside) the province's macroregion.

[14]Note that due to constraints in the data availability, the emigration rate for $T = 1961$ (ie, for the 1950s) only includes data from 1955 to 1961, that is $k = 6$ for $T = 1961$ and $k = 9$ otherwise. To ensure the comparability of the 1950s with the subsequent decades, we annualize the estimates by dividing the emigration rates by the number of available years in each period, that is:

$$em\_rate_{p,T} = \begin{cases} \frac{\sum_{\tau=T-6}^{T} emig_{p,\tau}}{pop_{p,T-10}}/7, \text{when } T = 1961 \\ \frac{\sum_{\tau=T-9}^{T} emig_{p,\tau}}{pop_{p,T-10}}/10, \text{when } T = 1971 \\ \frac{\sum_{\tau=T-9}^{T} emig_{p,\tau}}{pop_{p,T-10}}/10, \text{when } T = 1981 \end{cases} \tag{1}$$

laws were abolished only in February of 1961, leading to a temporary spike in changes to residence status when people that had emigrated in previous years regularized their situation with the local registries. Tortorici (2023: 71-75) estimates that this process affected about one million people, leading to adjustments to the provinces' population in the order of 4-7%. This event might influence our estimates through two channels: first, by underestimating internal migration in the 1950s, when the 1939 Fascist law were still in place and some migrations went unreported; second, by overestimating migration in the early 1960s (as regularizations were computed among new migrations). To address this potential threat, we follow Tortorici (2023: 73) who suggests that most regularizations happened in 1961, thanks to the migrants' high responsiveness to the repeal of the fascist laws. By including the emigrants of 1961 to the stock of the 1950s, we effectively shut down both channels of potential bias (we add regularizations back to the previous years and avoid attributing them to the 1960s).

3.2 Socio-Economic variables

As dependent and control variables, we mainly use socio-economic variables measured at the province level and observed at Census ten-year intervals. The original sources of the data for population and employment are population censuses from 1951 to 1981 while the number of establishments in the province is taken from the industrial censuses of the same years (Istat 1998). Sectoral value added and GDP at the provincial level is obtained from Unioncamere (2011). Labour productivity is computed as value added over the number of employees, total and by sector. Given that our main variable of interest is the emigration/immigration rate, we express all level variables in per capita terms (dividing by the population at time t) or for 1,000 residents, except for sector shares (in percentages). Table 2 shows descriptive statistics for our main outcomes of interest by decade.

3.3 Railway network

Historical accounts of the Italian Golden Age suggest that the train was the main means of transport used by internal migrants, especially those leaving the Southern regions for the Northern ones (Maggi, 2003)[15]. Emigrants mostly travelled at night with the *Treno del Sole* and the *Freccia del Sud* from Palermo, Syracuse and Reggio Calabria to Genoa-Turin and Bologna-Milan, while *I treni del levante* mainly started from Apulia. The large flow of internal migration during this period is also apparent from data on railway usage: while in

[15]By the end of the Second World War, the situation on the railways was disastrous: around 7,000 km of track were destroyed (30% of the entire length), mainly in the central area of Italy from Genoa and Bologna to Naples, Benevento and Foggia, where almost all the lines were interrupted. Despite the limited resources of the post-war period, reconstruction was rapid, and in 1955 the network was only 1000 km shorter than before the war (22.000 Km).

1939, before the outbreak of war, 167 million passengers and 11,8 billion passengers/Km were recorded[16], in 1951 they doubled (348 million and 21 billion pass/Km), with similar passenger levels and even higher travelled distances recorded in 1971 (347 million and 34 billions pass/Km)[17]. Drawing from this historical context, and exploiting the fact that all Italian provinces were linked to the railway network already in 1955, we reconstruct the railway lines at 1955 to build a measure of provincial connectivity; we later use this measure to build our instrument for long-distance migration. This reconstruction has been performed by first geolocalizing a map of the railways that were active in 1955, as obtained from the *Ministry of Transport*, and then manually digitizing the railway lines[18]. Figure 3 shows the resulting railroad network. We next combine our network with a least-cost-path algorithm based on physical distance to compute the minimal distance between any two provinces along the network, which we later use to construct our measure of exposure of a province to other provinces pull/push factors.

3.4 Migrations during the interwar period

In order to build our instrumental variable for short-distance migration, we resort to information on migratory flows during the interwar period. The rationale for doing so is that pre-existing migration experience (either personal or collective) to a certain province make potential emigrants better informed about, and more susceptible to, local shocks in employment opportunities in that specific provice than any other potential destination.Thus, bilateral migratory flows during the interwar period provide a measure of exposure to local push/pull factors during the Golden Age, which we can use to build our shift-share instrument. Crucially, the exposure measure must be relevant for short-distance flows and exogenous to later provincial development. To ensure that our exposure measure presents these characteristics, we exploit some peculiar institutional features of internal migration under the fascist regime.

During the interwar period, the fascist regime officially maintained a strong restrictive stance towards internal migration, although the government's influence was *de facto* very limited. According to the regime's approach, large-scale migration was a cause of concern for public order and it ought to either be repressed or directed by the authorities. The repression side of the fascist approach intended to hinder urbanisation and it was formally established by a series of laws that made increasingly difficult for Italians to legally change their residence. In fact, this legislation remained largely ineffective and internal migration

[16]Passengers/Km is a standard unit of measurement for transport demand; it is equal to the product passengers by kilometers travelled.

[17]Source: Istat, Statistiche storiche; https://seriestoriche.istat.it, Table 17.3, last retrieved 17/07/2024

[18]The source is the map 'Ferrovie, tranvie e filovie extraurbane, 1955' by Ministero dei Trasporti, Ispettorato Generale Motorizzazione Civile e Trasporti in Concessione, available online on https://www.stagniweb.it/mappe/fer955.jpg (last retrieved 18/01/2023).

increased significantly over the interwar period (Lemmi 1965; Treves 1976; Gallo 2012, p. 104).[19]

The interventionist side of the fascist approach intended instead to govern the collective movement of workers from areas of high unemployment to areas susceptible to economic development. To this end, Law No 358, 9th April 1931 established a Commission for migration and internal colonization (*Commissariato per le migrazioni e la colonizzazione interna*) to centralize control of migrations.[20] The Commission was to receive reports on the provincial labor market from local employment offices and decide on the extent and direction of inter-provincial flows. The main activity of the Commission was to authorize flows of groups of workers or farming households (*famiglie coloniche*) for agriculture and public works or land assignment. Many flows, especially short-distance ones, were temporary in nature because they followed seasonal agricultural activities (Treves, 1976).

As Gallo (2012, p. 104) emphasizes: "contrary to a common opinion, also consolidated in the historiographic field, Commissariato's activities were not intended to hinder periodic internal mobility,[21] nor did its statistical surveys have inquisitorial intentions, especially on the movement of agricultural labour force; on the contrary, they were judged as a useful remedy to unemployment. Control was in fact exercised not so much to prevent movements as to avoid clashes between national projects and the defence of local labour markets: the main objective remained to prevent the unemployed to roam in search of jobs and avoid their concentration in hotspots of public works." From 1929 to 1938, the Commissions authorized about 330,000 displacements per year on average until 1936, and more than 400,000 in 1937 and 1938.

Among its activities, the Commission also published data on the bilateral flows of workers authorized to move for employment purposes between any couple of Italian provinces.[22] We use this data, which we have collected and digitized with annual frequency for the years 1929-1938 (Commissariato 1930, 1932, 1938), in order to build our measure of exposure to local economic shocks for short-distance migration during

[19]Section A in the Appendix provides additional details on the repression side of the fascist approach to internal migration and its ineffectiveness.

[20]The Commission replaced the former "Standing Committee for Internal Migrations" (*Comitato permanente per le migrazioni interne*), established in 1926 (Royal Decree no. 440 of 4 March 1926), which had had the aim of "to study and propose the necessary measures to facilitate migration flows from the Provinces of the Kingdom with an overabundant population towards the less populated provinces of the Mezzogiorno with higher industrial and land productivity", and which had been created as part of a vast program of public works, with a special attention for less-developed Southern region. Here and in the rest of the paper, translations from Italian are our own unless otherwise stated.

[21]Actually Treves (1976), quoting Auletta (1960), reminded that the control of migration to the cities wasn't among the tasks of Commissariato.

[22]The Office's sources of information were concession railway tickets reserved for industrial and agricultural workers and labourers that travelled in group—known as 'Concessione ferroviaria XI - Operai, braccianti e lavoratori agricoli in comitiva.'—as well as (from 1931) records from local job centres (Uffici di collocamento, see Commissariato per le migrazioni e la colonizzazione interna 1932, p. iii).

the Golden Age.[23] The economic rationale for using these interwar data is the widely-accepted belief during the Fascist period that definitive displacements would follow temporary ones, i.e. that periodic migrations, by creating a sort of habit of moving to certain places, would paved the way for the consolidation of the relocation and the stable transplantation of the worker to the new location (Treves 1976, p. 103). This allows us to interpret these migration flows as an Intention-to-treat (ITT) mutual measures of exposure between provinces, where a stronger migration flow during the Fascism would be able to generate larger transfers of residence later on.

Moreover, what makes the Commission's data particularly appealing is the observation that the workers' migratory flows authorized by the Commission were plausibly exogenous to later provincial development. The plausibiliy of their exogeneity is supported by a comparison between the Commission's data and evidence on all migratory flows computed by Lemmi (1965) at the regional level from population censuses. Figure 4 panel A presents the share of workers moving between any couple of regions over the total cross-regional out-migration, which we have computed from the Commission's data over the period 1929-1938. This graph effectively gives a measure of how strong was the migratory link between each couple of Italian regions in the interwar period according to the workers' flows authorized by the Commission. Panels B and C, instead, show the same measure using all changes of residence between 1921 and 1931 and between 1931 and 1951, respectively. The comparison shows that the workers' flows authorized by the Commission were much more local (they are mostly located on the diagonal of the matrix), with very limited North-South and South-North moves. Moreover, while there are quantitatively-important migration shares towards Piedmont and Lazio, we also find comparatively similar ones towards Southern Regions. In comparison, the population changes of residence at 1931 and 1951 show very similar patterns to the massive moves of population from the South to the North recorded in the Golden Age, in line with what reported by Treves (1976).

These characteristics suggests that the workers' flows authorized by the Commission are plausibly orthogonal to underlying factors affecting migration in the Golden Age and better suited to capture exposure to shocks for short-distance flows. The next section presents in greater detail the identification strategy and explains how this measure of exposure is used to construct the instrumental variable.

4 Empirical Specification and Identification Strategy

We estimate a panel model at census decades (1961-1981) where our unit of analysis is a province-decade cell. Our baseline OLS equation takes the following form, which we

[23] The data are organised in double-entry matrices of annual transfers (origin/destination), a structure practically identical to the one that will be adopted by Istat and used in publications from 1955 onwards.

estimate separately for long and short-distance migration:

$$ln(y_{p,T}) = \alpha_p + \gamma_T + \sum_{i=em,im} \beta_i ln(i\_rate_{p,T}) + \sum_{t\in\{1971,1981\}} GDP\_pc_{p,1951} I(t = T) + \epsilon_{p,T} \qquad (2)$$

where p denotes a typical province, $T \in \{1961, 1971, 1981\}$, and $I(\cdot)$ denotes the indicator function. We denote a typical outcome by $y_{p,T}$. Those outcomes denoting stocks are divided by the current population $pop_{p,T}$ (see Table 2); those outcomes intrinsically defined as ratios remain unchanged (e.g. productivity, sector shares). The main explanatory variables are the annualized short and long distance emigration and immigration rates over the previous decade ($em\_rate^{sd}_{p,T}$, $em\_rate^{ld}_{p,T}$, $im\_rate^{sd}_{p,T}$, $im\_rate^{ld}_{p,T}$), computed as explained in Section 3 and similar to Anelli et al. (2023). We also transform all variables in percentages, and then take the logs.[24]

We include province fixed effects, that capture time-invariant unobserved geographic and institutional factors, and time fixed-effect to account for common trends over time. Moreover, in order to take into account that different provinces may already be on differential development trajectories in 1951, and to better isolate the role of internal migration, our baseline specification also controls for GDP per-capita in 1951, and interacts it with time fixed-effects to generate within-province variation over time. Section 6 explores robustness of our baseline results to the inclusion of further controls related to the sectoral composition of provinces in 1951. Finally, in our baseline specification we cluster standard errors at the province level, but explore robustness to alternative clustering strategies in Appendix B.

However, as previously noted, estimating the causal effect of emigration/immigration on sending/receiving areas using Equation (2) poses empirical challenges due to the endogeneity of migration decisions. For instance, provinces with the highest emigration (immigration) rates could be negatively (positively) selected among the least(most)-performing and slowest (fastest) growing areas, which would imply that naive OLS estimates would tend to overestimate the magnitude of the migration effect. We tackle the endogeneity concerns of province-specific migration rates through a shift-share identification strategy similar to Anelli and Peri (2017) and Anelli et. al (2023), and build on recent advances in the literature on shift-share designs (Adao Kolesar Morales, 2019; Borusjack Hull Jaravel, 2022; Goldsmith-Pinkham Sorkin Swift, 2020).

In particular, we build a province-specific instrument as the weighted-average of the other provinces' pull/push factors, where the weights measure the exposure of the

[24] An alternative specification of equation (2) is also possible where all variables are in levels and the *stock* of emigration/immigration and initial population are included as distinct covariates. In this specification, more common in gravity models of migration (e.g. Piras 2015; Ramazzotti 2024), the elasticity of the outcome variable to migration and population size are free to vary indipendently of each other. Results remain similar, but we prefer the current specification, where we can read effects in terms of migration rates.

province to each of the others and the push/pull factor captures the attractiveness of the provinces' labour markets. In the case of short-distance migration, we use the interwar flows of workers' migrations as measure of exposure and the change in total employment in the province as push/pull factor, using data from the Industrial Censuses. In the case of long-distance migration, instead, the measure of exposure is the degree of connection through the railway network and the push/pull factor is the change in employment as recorded in the Population Censuses. The rationale for building the instruments differently between long- and short-distance migration is detailed in the following sections.

Here it is important to note that, in order to ensure comparability of the effects of emigration and immigration, and to guarantee the same LATE-like interpretation of the results, we choose to use the same measure of exposure when constructing the instruments for emigration and immigration; additionally, we choose to use the same variable as push/pull factor, the only difference being that our push and pull factors are the inverse of each other. However, due to the symmetric construction of the instruments for emigration and immigration rates, the two instruments suffer from collinearity, implying that we cannot jointly estimate the effects of emigration and immigration within the same equation. Hence, our strategy consists in simply controlling for the emigration (immigration) rate when instrumenting for the immigration (emigration) rate[25]. For instance, when estimating the causal effect of emigration on the sending provinces, the first-stage takes the following form ($Z_{p,T}$ being the instrumental variable) :

$$ln(em\_rate_{p,T}) = \alpha_p + \gamma_T + \delta_Z Z_{p,T} + \delta_{im} ln(im\_rate_{p,T}) + \sum_{t \in \{1971,1981\}} GDP\_pc_{p,1951} I(t = T) + \nu_{p,T} \tag{3}$$

While the second stage is:

$$ln(y_{p,T}) = \alpha_p + \gamma_T + \beta_{em} \widehat{ln(em\_rate_{p,T})} + \beta_{im} ln(im\_rate_{p,T}) + \sum_{t \in \{1971,1981\}} GDP\_pc_{p,1951} I(t = T) + \nu_{p,T} \tag{4}$$

When we instrument for the immigration rate, the equations take a similar form but substitute $im\_rate_{p,T}$ for $em\_rate_{p,T}$ and viceversa. In the next 2 sections, we detail the construction of two different instruments depending on the spatial extent of recorded migration flows.

[25] Deuchert and Huber (2017) discuss the case of IV regressions where control variables help get conditional independence of the instrument. When the control variable is measured after the instrument assignment, exclusion restriction could not be valid, unless treatment variable and control are independent (Deuchert and Huber, 2017, p.418). Preliminary analyses show that emigration and immigration rates are uncorrelated at provincial level, so that there is no indirect effect of the treatment variable on the outcome variable via the control variable and the causal effect is correctly identified by the instrument (results are available upon request).

4.1 Short-distance migration

Focusing on short-distance migration is likely to introduce additional empirical challenges. In fact, not only migration is concentrated in slow-growing areas and directed towards fast-growing areas, but in the case of short-distance migration origin and destination areas are likely to be subject to the same local shocks. As a consequence, the typical assumption of exogeneous pull and push factors is not credible in the context of short-distance migration flows. Additionally, to the extent that these shocks are persistent, any measure of exposure to nearby provinces, even if predetermined, is likely to capture the effect of local shocks, thereby introducing an additional source of bias through the exposure shares.

We circumvent these issues by using our novel data on origin-destination authorized migrations of workers during the Fascist regime (1929-1938). Appendix C also shows that the exposure shares constructed through these data are unable to predict a large set of provincial socio-economic characteristics at the beginning of our study period (1951), confirming their orthogonality with respect to provincial development trajectories. Our framework builds on a recent contribution by Goldsmith-Pinkham et al. (2020) who show that exogeneity of exposure shares is sufficient to achieve identification, while allowing for endogeneous shift factors. As such, our strategy allows for common regional shocks affecting both origin and destination provinces. For each origin p and destination p', define

$$Fascist\_emigration_{p,p'} = \sum_{t=1929}^{1938} emigration_{p,p',t}$$

as the total stock of emigration from p to p' during the fascist period. In our baseline specification, in order to focus only on short-distance destinations without artificially overstating their role as push/pull drivers, we compute our measure of exposure as:

$$exposure^{sd}_{p,p'} = \begin{cases} \frac{Fascist\_emigration_{p,p'}}{\sum_{d \in S \setminus p} Fascist\_emigration_{p,d}} \text{if } p' \in S^{sd}_p \\ 0 \text{ otherwise} \end{cases}$$

capturing the share of the short-distance migrants between p and p' over the total out-migration from p. Section 6 shows that our results are robust to alternative definitions of the exposure shares. Following the Fascist narrative, these shares can be interpreted as a Intention-to-Treat measure of exposure, aiming at permanent moves by temporarily exposing workers from province p to province p'.

Next, as previously introduced, we make sure to retain the same LATE-like interpretation of our results on emigration and immigration by using the same variable as push/pull factor, which are simply computed as the inverse of each other. In particular, we choose to use the growth rate of the stock of employed workers in industry and private services as registered by the Italian Industrial Censuses. The choice of this source is motivated by the observation that it best captures recent structural change away from agriculture. Migrants that responded to opportunities offered by local structural change were suitable to capture new jobs in these fast-growing sectors, and thus they were likely more skilled and specialized than migrants that moved long-distance.

In order to construct our push and pull factors, we first take some preliminary steps to ensure comparability and limit the influence of outliers: first, in order to make the computation of inverses meaningful, we recenter the growth rates to be defined on the positive axis; next, in order to limit the influence of outliers, we winsorize the data at 5% and 95% level; finally, in order to ensure a smooth distribution, we take logs. The results are robust to using alternative transformations (Appendix B)[26]. We can then build our shift-share IV for emigration and immigration as

$$Shift\_share^{em}_{p,t} = \sum_{p' \in S} exposure^{sd}_{p,p'} * Pull_{p',t}$$

$$Shift\_share^{im}_{p,t} = \sum_{p' \in S} exposure^{sd}_{p,p'} * Push_{p',t}$$

where $Pull_{p',t}$ is the decadal growth rate of the stock of employed workers in province p', and $Push_{p',t}$ is its inverse. We interpret our instrument as a combination of push/pull factors of nearby provinces, where the weight measures the intensity of exposure of province p to p' during the Fascist regime.

4.2 Long-distance migration

The case of long-distance migration is conceptually different, as it allows us to treat shocks at destination provinces as exogenous with respect to the local development of origin provinces. This assumption seems to be particularly credible in the Italian context, characterized by little economic integration across macro-regions. As a result, our identifying assumption in the case of long-distance emigration is that shocks at destination affect the local development of origin areas only through destination-specific emigration rates; similarly, we assume that shocks in the provinces of origin only affect destination areas through destination-specific immigration rates. In support of this claim, in Appendix C we show that shocks at destination are unable to predict lagged outcomes at origin provinces, using a wide array of socio-economic characteristics.

To estimate the effect of long-distance migration on origin and destination areas, we again construct a shift-share instrument similarly to the previous section. However, in this case our measure of exposure reflects the proximity of each origin-destination pair

[26]In particular, we test for robustness to an inverse hyperbolic sine transformation and a cubic root transformation

along the railroad network at 1955 (the initial year of our analysis period). The idea of the instrument is to exploit the stylized fact that most long-distance migration at the time took place by train; by weighting more push/pull factors of relatively more distant provinces, we both i) weight more push/pull factors which are likely to be more relevant for long-distance moves (relevance) and ii) neglect the role of local economic shocks, which may independently affect local development (exclusion). In particular, the instrument builds on a recent contribution by Borusyak et al. (2021), where they show that exogeneity of the "shifts" (or shocks) in the context of shift-share designs is sufficient to achieve identification. Importantly, the framework in Borusyak et al. (2021) allows for endogenous railroad connectivity, which is likely to be the case if railroad proximity predicts origin areas' development through channels other than migration (for example, trade)[27]. In our baseline specification, for each couple of provinces p and p', we define our measure of exposure as

$$exposure^{ld}_{p,p'} = \begin{cases} \frac{Rail\_distance_{p,p'}}{\sum\limits_{d \in S^{ld}_p} Rail\_distance_{p,d}} \text{if } p' \in S^{ld}_p, \\ 0 \text{ otherwise} \end{cases}$$

which can be interpreted as the distance between p and a long-distance province p', as a share of the total distance from p to each long-distance province p'. Again, as in the previous section, we test for robustness to alternative definitions of the exposure shares, and show the results in Section 6[28].

Along the same lines as in the previous section, we choose to use the same variable as our push/pull factor for emigration/immigration, and perform exactly the same preliminary steps to ensure comparability and limit the influence of outliers. However, compared to the previous section, we choose to use the growth rate of the total stock of employed people as recorded by the Population Censuses instead of the Industrial Censuses. The rationale for choosing this other source hinges on the observation that it better captures total changes in employment opportunities, across all sectors. Greater opportunity for any employment was plausibly a more crucial determinant of migration for low-skill individuals, who were less suited to find good local jobs. In fact, historical evidence suggests that long-distance migration—particularly from the South to the North—was mostly chosen by unskilled young workers from rural areas, who responded to opportunities of low-specialized jobs in the main urban centres of the country. We then construct our

[27]Our specification is closely related to Sequeira et al. (2020), who interact the timing of a county's connection to the railway network with the aggregate inflow of immigrants over time to predict (exogenous) immigration.

[28]In particular, we test for robustness to $\frac{Rail\_distance_{p,p'}}{\sum\limits_{d \in S \setminus p} Rail\_distance_{p,d}}$, and $\frac{Rail\_distance_{p,p'}}{\sum\limits_{d \in S \setminus p} Rail\_distance_{p,d}}$ if $p' \in S^{ld}_p$, 0 otherwise.

shift-share IV for emigration and immigration as

$$Shift\_share^{em}_{p,t} = \sum_{p' \in S} exposure^{ld}_{p,p'} * Pull_{p',t}$$

$$Shift\_share^{im}_{p,t} = \sum_{p' \in S} exposure^{ld}_{p,p'} * Push_{p',t}$$

where again $Pull_{p',t}$ is the decadal growth rate of the stock of employed workers in province p', and $Push_{p',t}$ is its inverse. We interpret our instrument as a combination of push/pull factors of long-distance provinces, where the weight measures the relative distance of a destination based on its connectivity to the railroad network at 1955.

5 Main Results

The empirical analysis starts with OLS estimates of equation 2 for short-distance emigration. To better interpret our results, we first decompose the change in VA per capita in the usual sum of logs

$$\Delta VA\_pc = \Delta Productivity + \Delta Empl\_rate + \Delta DD. \tag{5}$$

where $Empl\_rate = \dfrac{employment}{pop_{15-64}}$ and $DD = \dfrac{pop_{15-64}}{population}$ is generally called *demographic dividend* and measures the contribution of population dynamics to (potential) labor force (Bloom et al., 2003). The contributions to growth in equation 5 helps us understanding the channel of the migration effect on the local economy.

Results are reported in panel A of Table 3. Our naive OLS estimates find a negative association between emigration and local per-capita value added (-.116, column 1) which is driven by the lower employment rate (column 3), while productivity remains unaffected (column 2). In fact, any possible loss of productivity is compensated by a positive effect on the demographic dividend (+.011, column 4). Moving away from value added data and into the industrial censuses, the overall negative effect of emigration on the economic performance is confirmed: we find a negative impact of emigration on firm dynamics (-.140). On the other hand, immigration is weakly associated with higher value added (+.082, column 1), which is mostly driven by greater productivity (+.103, column 2) while the effect on the demographic dividend appears to be negative (-.009, column 4). The estimates for firm dynamics are positive but not statistically significant. These OLS results seem to suggest that internal migration negatively affects provinces of origin and benefits provinces of destination.

However, we need to account for the known threats of endogeneity, hence we leverage the short-distance instrument created above to separately identify the causal effects of the emigration and immigration flows on local development. Our first-stage fit is quite satisfactory in both cases: the F-test adjusted for clustered-autocorrelated errors (Kleibergen-Paap F) is about 23 for the emigration and 20 for immigration. Panel B of Table 3 reports the results from the second stage for emigration. We find a stronger negative effect of emigration on local per-capita value added than in our baseline model (-.292, column 1) which is now also explained by lower productivity (-.172) and employment rate (-.141). The positive effect on the demographic dividend is confirmed and larger (+.021). Finally, the negative effect on the number of establishments per capita is also statistically significant and larger.

Panel C of Table 3 reports instead the results from the second stage for immigration on provinces of destination. We find again a larger positive effect of immigration on local value added (+.390, column 1) which is now mostly driven by higher productivity (+.348, column 2). The negative effect on the demographic dividend is again confirmed, while we do not find significant results by using the census data.

Not surprisingly, IV coefficients are (in absolute value) larger the OLS ones. There are at least two plausible explanations for this difference. First of all, it is likely that migrations are measured with error, as they underestimate emigration flows in case of failure of cancellation from municipalities' civil registries or immigration ones in case of failure of registration in the province of destination. This error produces an attenuation bias in the OLS estimates, which IV estimates do not suffer from. Secondly, IV estimates the local average treatment effect (LATE), the effect of a subgroup of migrants, shifted by the instruments, while OLS estimates the average treatment effect (ATE) over the entire population of migrants. In our case, the instrument in short-distance flows, is capturing the movements of workers (shares in the instrument are workers) and, among them, the more productive and skilled, as the shift is driven by growth rate of the stock of employed workers in industry and private services. As a result, the IV estimates should measure a stronger effect of migrations on economic variables.

Finally, when comparing the effects of emigration and immigration we find that the latter has a larger positive effect on the provinces of destinations than the former's negative effect on the provinces of origin. This suggests that, even though internal migration might depress areas of origins, the relocation of workers provides a net benefit for the national economy.

In order to better identify the sources of this productivity effects, we move to a sectoral analysis looking at the effect of emigration and immigration on the local share of workers employed, of value added and of productivity in agriculture, industry and services. Panel A of Table 4 reports the OLS estimates, which find a negative association between emigration and both the share of workers employed in industry and value-added in the sector. Employment is equally redistributed between agriculture and services while value added is compensated by services. We also find a negative association with productivity in agriculture. Immigration, on the other hand, is negatively associated with employment and value added in agriculture and weakly positively associated with productivity in services.

This heterogeneity by sector is amplified when we apply our IV approach. Looking at panel B, our second-stage estimates find a large and significant negative effect of emigration on industry shares, both in terms of employment and value added, which are compensated by large positive effects in services. These results suggest that emigration influenced the evolution of structural change in the provinces of origin, weakening or impeding industrialization. This interpretation is supported also by the negative coefficients that we estimate for productivity in services. In particular, despite the relative expansion of employment and value added in the service sector, its productivity decreased, which suggests that the sector's growth was concentrated in low-productivity activities.

We replicate the analysis for immigration, reporting the second-stage coefficients in panel C of 4. We find effectively the opposite effects of emigration: provinces that receive more migrants show a higher share of valued added (+.31, column 5) and possibly employment (+.14 but not statistically significant, column 2) in industry. We also find a strong negative effect of immigration on value added in agriculture. Moreover, immigration is associated with significantly higher productivity both in industry (+.52, column 8) and services (+.31, column 9).

Overall, our analysis suggests that short-distance migration accompanied structural transformation in favour of high-productivity industrial activities in the provinces of destination. These results are particularly important for interpreting the effect of internal migrations on local economic development, because recent growth accounting reconstructions (Antonelli and Barbiellini Amidei (2011); Giordano and Zollino (2021)) find that productivity growth in the industrial sector was the largest contributor to Italy's economic growth during the Golden Age (1951-1973).

On the other hand, we also found that emigration negatively affected the sending province's value added, mainly through reducing the share of industrial employment and productivity in services. In that respect, emigration would delay structural change in favour of industry while reducing the benefits of growing employment in services. These findings are also coherent with the recent historiography: Cappelli et al. (2018, pp. 15-17) argue that during the Golden Age 'the strong migration of the labour force from the south to the northern regions partly offset productivity convergence due to labour reallocation.'

To further test our hypothesis, we replicate the analysis for long-distance migrations, i.e. considering only migrants to provinces in other macro-regions. Table 5 presents the usual comparison between OLS and IV estimates, both for emigration and immigration. Similarly to the previous analysis, we find a negative effect of emigration on value added in provinces of origin, which is largely explained by a lower employment rate. Looking at census data, we also find a negative association between the emigration rate and the number of firm establishments per capita in the provinces of origin, which appears even larger with the IV estimates. Differently from short-distance movements, however, we do not find any significant effect of emigration on productivity in the province of origin. Immigration, on the other hand, does not appear to be associated with any significant improvement in destination provinces: both in the OLS and IV estimates we find insignificant negative coefficients across all outcomes, except for the demographic dividend which is positive and statistically significant.

Moving to the sectoral analysis, Table 6 shows a significant negative effect of emigration on industry employment share in the origin provinces, but no effect of immigration on the same sector. Services remain unaffected, while there is some unexpected effects on agriculture: emigration and immigration are both associated with higher employment share in agriculture in both origin and destination provinces (in the IV estimates, +.33 and +.87 respectively). We attribute this counter-intuitive result to migrants substituting for native workers in the agricultural sector of destination provinces—thus allowing provinces that received a larger number of migrants to preserve its agricultural sector while expanding industrial employment (Grilli and Zanotelli, 2015)—, and to provinces of origin relatively specializing in agricultural activities with respect to destinations. However, we do not exclude that this results might also be driven by migrants being classified under their occupation of origin in statistical sources, thus inducing classification errors.

Overall, our results point to significant heterogeneity in effects between short and long-distance migration. We attribute this heterogeneity to migrant selection as it is plausible that migrants self-selected differently into destinations, with the most productive ones favoring nearby destinations—possibly based on a preference for proximity to origin areas—, while the least productive ones were forced to engage in long-distance migration (Monras 2020). In order to assess the robustness of our estimates and the plausibility of our interpretations, we next perform a string of checks and additional analyses.

6 Robustness checks

In this section, we test for robustness of our main results to a variety of changes to our main specification. First, we include additional controls meant to capture the provinces' sectoral composition in 1951, and interact those with time fixed effects; these additional controls are meant to better isolate the response of our outcomes to internal migration from differential trends due to the start-of-period sectoral composition. Second, we test for robustness of our main results to changes in the definition of the exposure shares.

Third, we test for robustness to different definitions of short and long distance migration. Additional robustness checks (alternative clustering strategies, alternative transformation of the push/pull factors, exposure-robust inference of Adao et al. (2019)) are provided in Appendix B.

6.1 Controlling for the sectoral composition in 1951

A potential concern to our main specification is that heterogeneous sectoral composition of the provinces in the base period (1951) may differentially drive the subsequent evolution of income, productivity and employment, both in terms of levels and sectoral shares, independently from the province-specific patterns in internal migration. For instance, if provinces with a larger agricultural share are those which are mostly subject to migration patterns and structural change, our estimates may erroneously ascribe the observed changes in outcomes to in and out-migration. To address this concern, we enlarge the set of controls to include the agricultural and industrial shares of value added and employment in 1951, and interact these controls with time fixed effects to generate time variability within-province. Tables 7 and 8 show the results for short-distance migration, while Table 9 and 10 show the results for long-distance migration. Overall, we confirm the main results of section 5: while short-distance emigration affects origin provinces' value added per capita, mostly through a productivity and employment rate channel, immigration positively affects destination provinces' value added through productivity, with a positive net effect; on the contrary; long-distance emigration has limited effects on origin provinces value added, mostly due to an employment rate channel, while we find limited effects on destination provinces.

6.2 Alternative definition of the exposure shares

In this section, we explore robustness of our results to alternative definitions of the exposure shares. In particular, our baseline shift-share instrument for the short-distance migration scaled the destination-specific Fascist migration flow over the entire out-migration flow, and then set the weights of the destinations outside the macro-region to zero. This was motivated in order not to overstate the role of within-macro-region provinces in predicting within-macro-region in and out migration, and entails that the exposure shares do not sum to 1. In the context of sectoral employment shares, Borusyak et al. (2022) show that shift share-designs with exposure shares not summing to a constant may be affected by bias if units displaying a higher sum of shares (and therefore a higher value of the instrument) are also different in terms of unobservables. Again, we believe this may be less than an issue in our context, since: i) our identifying variation for the effects of short-distance migration directly comes from the Fascist migration shares, according to Goldsmith-Pinkham et al. (2020) framework; ii) as the weight of destinations outside *any* given province' macro-region are set to zero, this is unlikely to create imbalances in the intrument's value across provinces. However, for completeness we explore robustness to two alternative specifications: a first one where the exposure shares sum to 1 within each province's macro-region (and are zero outside), and a second one where the exposure shares sum to 1 across all destinations. Again, we only report the results for our main outcomes for brevity. Tables 11 and 12 show the results. Also with these specifications, the instruments for short and long-distance migration exhibit considerable predictive power, and the results are qualitatively similar to our baseline ones, although our second alternative specification shows effects of larger magnitude.

For completeness, we also report alternative definitions of the instrument for long-distance migration. In particular, our baseline instrument weights the push and pull factors of provinces located outside each given province' macro-region according to a weighted average, where the weight is computed as the origin-destination distance rescaled by the total distance from a given origin to all destinations outside its macro-region. Similarly to above, we can test for robustness to alternative exposure share definitions: in the first one, we first scale all the origin-destination distances by the total distance from the origin to all destinations, and then set the weight of destinations within the macro-region to zero; in a second specification, we simply scale the origin-destination distance by the total distance from the origin to all destinations. One implication is that the exposure shares no longer sum to 1 across provinces outside macro-region in specification 1, while they sum to 1 across all destinations in specification 2. Tables 13 and 14 show the results. Even though the instrument for immigration is sligthly weak in specification 1, it is still above the 10 threshold in specification 2; the instrument for emigration is strong in both cases. The results show again limited effects on destination provinces, and negative effects on origin provinces' income, mostly through an employment rate channel; we still find negative effects on the number of establishments per capita in origin provinces.

6.3 Alternative definition of short and long distance

In this section, we explore alternative definitions of short and long distance migration. In particular, our baseline definition is not based on a measure of actual distance between provinces but rather on whether they belong to the same macro-region or not. Although a natural baseline definition and in line with the historical tradition, this definition may suffer from some limitations. First, migration flows between two provinces located on opposite sides of a macro-region's boundary would be considered long-distance, while in fact it would be short-distance by definition. This would especially affect the quantitatively-important migration flows from the South to the Center and from the North-East to the North-West. Moreover, given the considerable spatial extent of some macro-regions compared to others (for instance, the South vs the Center or the Islands), this definition may create imbalances across macro-regions in the type of migration flows we are capturing. In particular, the short-distance migration in the South would involve considerably larger distances compared to the Center or the Islands. In order to address these points, we consider alternative definitions of short and long-distance migration based on the actual physical distance between provinces along the railway network at 1955. For short-distance emigration (immigration), we consider the emigration (immigration) stock to (from) provinces located between 100 and 400 km from the origin (destination) province. Similarly, for long-distance emigration (immigration), we consider the emigration (immigration) stock to (from) provinces located between 400 km and 800 km from the origin (destination) province. In both cases, estimates are performed considering the stock of migrants at the baseline distance—respectively, those coming from (going to) provinces under 100 km for short-distance migration and under 400 km for long-distance migration—and then augmenting the number to include the stock of migrants coming from (going to) farther provinces, with 50 km increments. For each distance threshold, we construct a corresponding instrument giving positive weight only to those provinces located within or outside the threshold, for short-distance and long-distance migration respectively.

Figures 5 and 6 show the results for short distance migration, which confirm large negative income effects of short-distance migration on origin provinces, mostly driven by productivity within the first 200 km and later mostly driven by the employment rate; similarly, we confirm large positive income effects of immigration on destination provinces, which are particularly driven by productivity effects. Figures 7 and 8 show instead the results for long distance migration. We find negative effects of emigration on origin provinces' per-capita value added, but unlike the case of short-distance flows, these results appear mostly driven by the employment rate and the number of firm establishments per capita, especially for distances over 600 km. Long-distance immigration is confirmed to have a negligible effect on productivity in destination provinces, especially for migrants coming farther away than 600 km; the contribution of long-distance migration to local value added appears to be explained mostly by the employment rate and by the number of establishments per capita, for which the effects are positive and precisely estimated after 600 km.

The results are again consistent with a model of localized productivity where the most productive workers stay close and negatively affect origin provinces, whose damage is more than compensated by destination provinces' benefits; on the contrary, the least productive workers engage in long-distance migration, reducing the demographic pressure on origin provinces and supporting extensive growth in destination provinces, but with little productivity gains.

7 Conclusions

Internal migration is a crucial mechanism underlying an efficient allocation of labor within the domestic economy. Its evolution underpins economies of agglomeration, structural transformation, urbanization and demographic change, especially in periods of fast economic growth and technological change. However, the effects of internal migration on local economic development are not well known, for the literature mostly concentrates on the effects of international migration, often restricting the attention to labor market outcomes.

In this paper, we address this gap in the literature by studying the role of internal migration on local development and sectoral reallocation of sending and receiving areas in Italy during the Golden Age (1950s-1970s). This was the period of fastest economic growth, structural change and regional convergence in the country's history and it was characterized by large flows of internal migration, both short- and long-distance. The significance of such large flows is well recognized by the historiography for their social, cultural and economic effects. Contemporaries also highlighted the extraordinary shifts in Italian society that accompanied this explosion of internal migration. However, a quantitative analysis of its impact on local economic development did not exist to-date.

To address this problem, our analysis has focused on changes in GDP per capita, productivity, economic structure, employment and firm dynamics at roughly ten-year census intervals. We have thus combined information from the industrial and population censuses with reconstructions of migration flows across provinces (both emigration and immigration) from the 1950s through 1970s, obtained from changes in residence status. We have opted to conduct the analysis at the province level because we are able to match a variety of historical records with origin-destination migration flows at the lowest available level of spatial aggregation.

Our research design has been geared towards addressing the endogeneity which is intrinsic to migration decisions. In order to cope with this problem, we have adopted a shift-share instrumental variable approach, which combines push/pull factors driving internal migration (employment growth in our case) with a measure of exposure to them. Exploiting recent advances in the literature on shift-share designs, we have been able to construct instruments for immigration and emigration flows showing strong predictive power, retaining the same LATE-like interpretation of the results, and passing key validity tests. In particular, we have distinguished between long- and short-distance migration, due to its different nature and evolution over time, and correspondingly employed two competing but different research designs. For short-distance migration, which is more likely to be affected by the same local economic shocks, our identification has relied on a measure of exogenous provincial exposure based on authorized flows of workers during

Fascism years, while allowing push and pull factors to be endogenous. On the contrary, for long-distance migration we have exploited the exogeneity of shocks at distant destinations with respect to the local development of origin provinces, a credible assumption in the Italian context of that years characterized by little economic integration across macro-regions.

Our analysis has found that internal migration had heterogeneous effects between sending and receiving areas, and they differed between short- and long-distance. With respect to short-distance, emigration reduces per-capita value added in the sending areas—which is mostly explained by lower productivity and employment—and curbes structural change measured by lower employment, value added and productivity in industry. Short-distance immigration, on the other hand, has a positive effect on the provinces of destination, with a particularly strong association with value added and productivity, both in industry and services. With respect to long-distance migration, we find a similar negative effect on value added, employment and firm dynamics in the provinces of origin, but not on productivity. Moreover, we do not find a clear causal effect on economic outcomes in the provinces of destination. We speculate that these differences between types of destination are due to selection of migrants. The results are consistent with a model of distance-based selection of emigrants, where the most productive ones tend to favor nearby destinations possibly based on a preference for proximity to origin areas (in line with the theoretical model in Monras, 2020). It is also plausible that the effect of long distance migration acted mostly by impeding overheating, an effect not captured by productivity measures.

We confirm our results through a large battery of robustness checks. We control for the sectoral composition in 1951, we use alternative exposure measure and employ alternative definitions of short and long distance. Additionally, our results hold under alternative clustering strategies, inference methods and transformations of our push/pull factors. Finally, we perform a string of tests (presented in Appendix C) to ensure the instruments' validity as required in Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022). Potential extensions of the paper include a deeper analysis of the underlying mechanisms and an overall assessment of the effects of migrations on regional convergence.

8 Tables

Table 1: Emigration rates by macro-region - per 1,000 inhabitants in the macro-region of origin

	Area					
Origin/Dest	North-West	North-East	Centre	South	Islands	Total
Period: 1950s						
North-West	10.04	1.82	0.99	0.80	0.47	14.12
North-East	7.21	7.65	1.46	0.58	0.34	17.24
Centre	2.29	1.63	8.96	1.48	0.63	15.00
South	5.06	1.00	3.29	4.91	0.61	14.87
Islands	4.47	0.99	2.30	1.12	6.88	15.75
Period: 1960s						
North-West	8.72	2.17	1.50	2.21	1.48	16.08
North-East	4.21	7.00	1.49	0.78	0.46	13.94
Centre	2.56	1.52	8.24	1.92	0.85	15.09
South	7.56	1.23	3.31	4.87	0.58	17.55
Islands	8.54	1.22	2.68	1.14	6.11	19.69
Period: 1970s						
North-West	5.56	1.52	1.22	2.27	1.41	11.97
North-East	1.70	4.72	0.97	0.74	0.38	8.52
Centre	1.43	1.01	5.15	1.66	0.69	9.93
South	4.72	1.26	2.43	4.11	0.49	13.00
Islands	4.98	1.06	1.84	0.94	4.73	13.54

Notes: The table shows the emigration rates by macro-region, which are computed as the number of annual emigrants from a given origin macro-region to a given destination macro-region cumulated over the decade, per 1,000 inhabitants in the macro-region of origin at the beginning of the period (i.e. 1951, 1961 and 1971). For each decade, we annualize the resulting emigration rates by diving by the number of available years: 7 for the 1950s (migration data are only available for the years 1955-1961), 10 for 1960s (data available for 1962-1971), 10 for the 1970s (data available for 1972-1981). Migration within the same province is excluded from the computations.

Table 2: Descriptive statistics - Economic and social variables

	1961		1971		1981	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Population	549167.49	500793.26	587457.01	613388.55	613200.91	647060.63
PC GDP	7,627.88	2,267.55	12,564.57	2,934.66	17,915.86	4,259.74
PC value added - total	7,227.31	2,138.77	11,505.12	2,783.50	16,943.33	4,098.73
PC value added - agriculture	1,221.59	586.44	1,243.31	607.79	1,214.04	640.90
PC value added - industry	2,607.91	1,464.41	4,289.89	1,867.69	6,036.92	2,723.08
PC value added - services	3,397.81	1,114.30	5,971.91	1,693.00	9,692.37	2,124.79
Productivity	18.45	4.74	32.82	6.56	46.58	6.34
Productivity - agriculture	10.16	3.72	19.00	7.04	28.20	11.99
Productivity - industry	16.79	4.90	28.55	6.87	41.79	7.79
Productivity - services	30.68	3.42	46.28	4.66	56.17	4.70
Employment - agriculture	132.76	64.28	74.28	40.68	47.59	26.65
Employment - industry	148.51	52.14	147.14	45.32	140.09	49.82
Employment - services	109.96	32.00	128.01	29.38	172.00	33.32
Share employment agriculture	33.98	15.15	21.60	11.76	13.90	8.47
Share employment industry	37.70	10.98	41.62	9.64	38.04	9.25
Share employment services	28.32	8.24	36.78	8.17	48.06	7.76
Firm establishments	41.45	8.41	46.81	9.87	58.15	14.33
Observations	91		91		91	

Notes: Summary statistics (mean, standard deviation) for our 91 Italian provinces at Census years (1961, 1971, 1981).

Table 3: Short-distance migration effects on provinces' economic development

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094***	0.011**	-0.140***
	(0.030)	(0.036)	(0.023)	(0.004)	(0.035)
Short-distance immig. rate	0.082*	0.103**	-0.012	-0.009*	0.022
	(0.042)	(0.046)	(0.022)	(0.005)	(0.041)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.292***	-0.172**	-0.141***	0.021***	-0.187**
	(0.073)	(0.074)	(0.044)	(0.008)	(0.077)
Rmse	0.063	0.063	0.035	0.007	0.058
Kleibergen-Paap F	22.921	22.921	22.921	22.921	22.921
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.390***	0.348***	0.071	-0.028**	0.105
	(0.122)	(0.125)	(0.071)	(0.012)	(0.109)
Rmse	0.068	0.066	0.035	0.007	0.059
Kleibergen-Paap F	19.990	19.990	19.990	19.990	19.990
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. Panel A shows OLS estimates, Panel B shows IV estimates. The instrument is computed as a weighted average of destinations' occupation growth rate, where the weight is given by the rail-based proximity index. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Short-distance migration effects on provinces' sectoral composition

	Endogenous: local emigration/immigration rate								
	Employment (share)			Value added (share)			Productivity		
	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS									
Short-distance emig. rate	0.174**	-0.250***	0.094**	-0.012	-0.145***	0.092***	-0.219***	0.074	-0.036
	(0.079)	(0.050)	(0.039)	(0.066)	(0.041)	(0.023)	(0.081)	(0.047)	(0.029)
Short-distance immig. rate	-0.173**	0.049	0.058	-0.301***	0.020	0.031	-0.033	0.076	0.076*
	(0.072)	(0.055)	(0.052)	(0.074)	(0.053)	(0.037)	(0.099)	(0.055)	(0.041)
Rmse	0.129	0.082	0.068	0.147	0.092	0.056	0.182	0.121	0.063
N	273	273	273	273	273	273	273	273	273
Panel B: IV emigration									
Short-distance emig. rate	0.196*	-0.302***	0.135**	0.144	-0.307***	0.138**	-0.227	-0.177	-0.169***
	(0.114)	(0.071)	(0.064)	(0.134)	(0.092)	(0.053)	(0.173)	(0.128)	(0.062)
Rmse	0.104	0.067	0.055	0.121	0.078	0.046	0.148	0.104	0.054
Kleibergen-Paap F	22.921	22.921	22.921	22.921	22.921	22.921	22.921	22.921	22.921
N	273	273	273	273	273	273	273	273	273
Panel C: IV immigration									
Short-distance immig. rate	-0.211	0.140	-0.014	-0.575***	0.307**	-0.049	-0.018	0.518**	0.311***
	(0.173)	(0.125)	(0.122)	(0.217)	(0.153)	(0.087)	(0.281)	(0.222)	(0.115)
Rmse	0.105	0.067	0.056	0.123	0.082	0.047	0.148	0.111	0.058
Kleibergen-Paap F	19.990	19.990	19.990	19.990	19.990	19.990	19.990	19.990	19.990
N	273	273	273	273	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. Panel A shows OLS estimates, Panel B shows IV estimates. The instrument is computed as a weighted average of destinations' occupation growth rate, where the weight is given by the rail-based proximity index. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Long-distance migration effects on provinces' economic development

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053***	0.019	-0.080***	0.008**	-0.116***
	(0.020)	(0.022)	(0.012)	(0.003)	(0.028)
Long-distance immig. rate	-0.022	0.028	-0.050***	0.000	-0.025
	(0.028)	(0.035)	(0.018)	(0.005)	(0.031)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.095**	-0.036	-0.078***	0.020***	-0.150***
	(0.038)	(0.044)	(0.019)	(0.005)	(0.041)
Rmse	0.061	0.062	0.032	0.007	0.057
Kleibergen-Paap F	83.584	83.584	83.584	83.584	83.584
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.169	-0.164	-0.044	0.042***	-0.144
	(0.128)	(0.144)	(0.057)	(0.014)	(0.121)
Rmse	0.063	0.066	0.032	0.009	0.058
Kleibergen-Paap F	17.508	17.508	17.508	17.508	17.508
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. Panel A shows OLS estimates, Panel B shows IV estimates. The instrument is computed as a weighted average of destinations' occupation growth rate, where the weight is given by the rail-based proximity index. The standard errors are clustered at the province level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 6: Long-distance migration effects on provinces' sectoral composition

	Endogenous: long-distance emigration/immigration rate								
	Employment (share)			Value added (share)			Productivity		
	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS									
Long-distance emig. rate	0.108**	-0.070**	0.084***	-0.025	-0.060**	0.051***	-0.115**	0.029	-0.013
	(0.046)	(0.028)	(0.020)	(0.042)	(0.024)	(0.013)	(0.044)	(0.030)	(0.022)
Long-distance immig. rate	0.089*	-0.006	0.096***	0.061	0.047	0.044*	0.004	0.082*	-0.026
	(0.048)	(0.053)	(0.034)	(0.086)	(0.038)	(0.023)	(0.094)	(0.041)	(0.029)
Rmse	0.131	0.091	0.064	0.153	0.094	0.056	0.184	0.121	0.064
N	273	273	273	273	273	273	273	273	273
Panel B: IV emigration									
Long-distance emig. rate	0.329***	-0.130**	0.056	0.020	-0.000	0.048	-0.336***	0.091	-0.043
	(0.086)	(0.056)	(0.037)	(0.079)	(0.045)	(0.034)	(0.117)	(0.060)	(0.030)
Rmse	0.115	0.075	0.052	0.124	0.078	0.046	0.155	0.099	0.052
Kleibergen-Paap F	83.584	83.584	83.584	83.584	83.584	83.584	83.584	83.584	83.584
N	273	273	273	273	273	273	273	273	273
Panel C: IV immigration									
Long-distance immig. rate	0.867***	-0.215	0.001	0.219	0.256	0.031	-0.771**	0.299	-0.132
	(0.327)	(0.194)	(0.101)	(0.197)	(0.160)	(0.100)	(0.361)	(0.233)	(0.108)
Rmse	0.150	0.079	0.054	0.126	0.082	0.046	0.183	0.103	0.054
Kleibergen-Paap F	17.508	17.508	17.508	17.508	17.508	17.508	17.508	17.508	17.508
N	273	273	273	273	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. Panel A shows OLS estimates, Panel B shows IV estimates. The instrument is computed as a weighted average of destinations' occupation growth rate, where the weight is given by the rail-based proximity index. The standard errors are clustered at the province level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 7: Short-distance migration, Main Outcomes: Inclusion of agricultural and industrial shares in value added and employment in 1951

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.089***	0.001	-0.100***	0.010***	-0.112***
	(0.030)	(0.034)	(0.022)	(0.004)	(0.030)
Short-distance immig. rate	0.047	0.062	-0.002	-0.014**	-0.003
	(0.046)	(0.047)	(0.020)	(0.006)	(0.043)
Rmse	0.067	0.068	0.039	0.008	0.068
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.303***	-0.184**	-0.146***	0.027***	-0.198**
	(0.073)	(0.071)	(0.046)	(0.008)	(0.082)
Rmse	0.061	0.060	0.032	0.007	0.056
Kleibergen-Paap F	19.171	19.171	19.171	19.171	19.171
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.490***	0.446***	0.093	-0.049***	0.176
	(0.158)	(0.165)	(0.078)	(0.016)	(0.132)
Rmse	0.073	0.069	0.033	0.007	0.058
Kleibergen-Paap F	13.046	13.046	13.046	13.046	13.046
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. In panel A, we additionally control for GDP per-capita in 1951, the share of employment in agriculture in 1951, and the share of value added in agriculture in 1951, and their interaction with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8: Short-distance migration, Sectoral Outcomes: Inclusion of agricultural and industrial shares in value added and employment in 1951

	Endogenous: local emigration/immigration rate								
	Employment (share)			Value added (share)			Productivity		
	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS									
Short-distance emig. rate	0.114**	-0.170***	0.155***	-0.014	-0.143***	0.097***	-0.131*	0.030	-0.058**
	(0.056)	(0.040)	(0.025)	(0.063)	(0.034)	(0.015)	(0.074)	(0.044)	(0.028)
Short-distance immig. rate	-0.151**	0.026	-0.001	-0.195*	-0.015	0.006	0.012	0.023	0.070
	(0.059)	(0.040)	(0.029)	(0.099)	(0.048)	(0.023)	(0.121)	(0.065)	(0.045)
Rmse	0.105	0.061	0.047	0.138	0.081	0.042	0.166	0.110	0.058
N	273	273	273	273	273	273	273	273	273
Panel B: IV emigration									
Short-distance emig. rate	0.139	-0.280***	0.176***	0.198	-0.270***	0.135***	-0.134	-0.174	-0.224***
	(0.108)	(0.058)	(0.043)	(0.146)	(0.084)	(0.042)	(0.182)	(0.114)	(0.062)
Rmse	0.085	0.051	0.038	0.114	0.067	0.034	0.133	0.092	0.052
Kleibergen-Paap F	19.171	19.171	19.171	19.171	19.171	19.171	19.171	19.171	19.171
N	273	273	273	273	273	273	273	273	273
Panel C: IV immigration									
Short-distance immig. rate	-0.203	0.254*	-0.045	-0.634**	0.247	-0.073	0.019	0.444**	0.415***
	(0.217)	(0.141)	(0.085)	(0.272)	(0.170)	(0.085)	(0.372)	(0.222)	(0.142)
Rmse	0.085	0.055	0.038	0.121	0.071	0.035	0.133	0.100	0.061
Kleibergen-Paap F	13.046	13.046	13.046	13.046	13.046	13.046	13.046	13.046	13.046
N	273	273	273	273	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. In panel A, we additionally control for GDP per-capita in 1951, the share of employment in agriculture in 1951, and the share of value added in agriculture in 1951, and their interaction with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 9: Long-distance migration, Main Outcomes: Inclusion of agricultural and industrial shares in value added and employment in 1951

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.045**	0.036*	-0.086***	0.005	-0.114***
	(0.020)	(0.020)	(0.012)	(0.003)	(0.028)
Long-distance immig. rate	-0.028	0.007	-0.036**	0.001	-0.014
	(0.037)	(0.042)	(0.016)	(0.005)	(0.033)
Rmse	0.068	0.068	0.036	0.008	0.065
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.059	0.006	-0.082***	0.017***	-0.122***
	(0.038)	(0.042)	(0.018)	(0.005)	(0.038)
Rmse	0.055	0.055	0.029	0.007	0.052
Kleibergen-Paap F	69.559	69.559	69.559	69.559	69.559
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.078	-0.104	-0.019	0.046***	-0.043
	(0.115)	(0.128)	(0.055)	(0.017)	(0.106)
Rmse	0.055	0.057	0.029	0.009	0.052
Kleibergen-Paap F	12.418	12.418	12.418	12.418	12.418
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. In panel A, we additionally control for GDP per-capita in 1951, the share of employment in agriculture and industry in 1951, and the share of value added in agriculture and industry in 1951, and their interaction with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 10: Long-distance migration, Sectoral Outcomes: Inclusion of agricultural and industrial shares in value added and employment in 1951

	Endogenous: long-distance emigration/immigration rate								
	Employment (share)			Value added (share)			Productivity		
	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS									
Long-distance emig. rate	0.099***	-0.044**	0.099***	-0.012	-0.039*	0.042***	-0.077*	0.041	-0.021
	(0.035)	(0.021)	(0.014)	(0.040)	(0.022)	(0.012)	(0.040)	(0.026)	(0.019)
Long-distance immig. rate	0.025	-0.036	0.073***	0.067	-0.000	0.017	0.055	0.044	-0.050
	(0.041)	(0.034)	(0.023)	(0.079)	(0.040)	(0.021)	(0.102)	(0.052)	(0.031)
Rmse	0.106	0.066	0.045	0.140	0.084	0.043	0.166	0.109	0.059
N	273	273	273	273	273	273	273	273	273
Panel B: IV emigration									
Long-distance emig. rate	0.169***	-0.072**	0.099***	-0.023	0.033	0.004	-0.179*	0.108*	-0.089***
	(0.057)	(0.032)	(0.027)	(0.077)	(0.041)	(0.023)	(0.103)	(0.063)	(0.032)
Rmse	0.087	0.054	0.036	0.113	0.069	0.036	0.135	0.089	0.049
Kleibergen-Paap F	69.559	69.559	69.559	69.559	69.559	69.559	69.559	69.559	69.559
N	273	273	273	273	273	273	273	273	273
Panel C: IV immigration									
Long-distance immig. rate	0.279	-0.137	0.072	0.028	0.261*	-0.121	-0.316	0.288	-0.298**
	(0.190)	(0.125)	(0.089)	(0.201)	(0.152)	(0.091)	(0.287)	(0.217)	(0.133)
Rmse	0.092	0.055	0.036	0.113	0.076	0.039	0.142	0.094	0.058
Kleibergen-Paap F	12.418	12.418	12.418	12.418	12.418	12.418	12.418	12.418	12.418
N	273	273	273	273	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects. In panel A, we additionally control for GDP per-capita in 1951, the share of employment in agriculture and industry in 1951, and the share of value added in agriculture and industry in 1951, and their interaction with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 11: Short-distance migration, Main Outcomes: Exposure shares summing to 1 within the macro-region

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094***	0.011**	-0.140***
	(0.030)	(0.036)	(0.023)	(0.004)	(0.035)
Short-distance immig. rate	0.082*	0.103**	-0.012	-0.009*	0.022
	(0.042)	(0.046)	(0.022)	(0.005)	(0.041)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.266***	-0.120*	-0.170***	0.024***	-0.240***
	(0.060)	(0.060)	(0.042)	(0.008)	(0.067)
Rmse	0.062	0.061	0.036	0.007	0.059
Kleibergen-Paap F	29.473	29.473	29.473	29.473	29.473
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.425***	0.302**	0.162*	-0.040***	0.250**
	(0.132)	(0.125)	(0.097)	(0.014)	(0.109)
Rmse	0.071	0.064	0.040	0.008	0.064
Kleibergen-Paap F	21.645	21.645	21.645	21.645	21.645
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). The exposure shares are computed as the Fascist migration flow from a given origin to a given destination within the origin province's macro-region, over the total out-flow from the same origin province within macro-region. All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Short-distance migration, Main Outcomes: Exposure shares summing to 1 over all destinations

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094***	0.011**	-0.140***
	(0.030)	(0.036)	(0.023)	(0.004)	(0.035)
Short-distance immig. rate	0.082*	0.103**	-0.012	-0.009*	0.022
	(0.042)	(0.046)	(0.022)	(0.005)	(0.041)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.339***	-0.130*	-0.239***	0.030***	-0.344***
	(0.084)	(0.076)	(0.064)	(0.010)	(0.103)
Rmse	0.066	0.061	0.040	0.007	0.064
Kleibergen-Paap F	21.559	21.559	21.559	21.559	21.559
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.532***	0.299**	0.281*	-0.049***	0.434**
	(0.180)	(0.145)	(0.148)	(0.018)	(0.169)
Rmse	0.078	0.064	0.048	0.008	0.075
Kleibergen-Paap F	16.565	16.565	16.565	16.565	16.565
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). The exposure shares are computed as the Fascist migration flow from a given origin to a given destination, over the total out-flow from the same origin province. All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Long-distance migration, Main Outcomes: Exposure shares not summing to 1 outside the macro-region

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053***	0.019	-0.080***	0.008**	-0.116***
	(0.020)	(0.022)	(0.012)	(0.003)	(0.028)
Long-distance immig. rate	-0.022	0.028	-0.050***	0.000	-0.025
	(0.028)	(0.035)	(0.018)	(0.005)	(0.031)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.122***	-0.046	-0.098***	0.022***	-0.152***
	(0.045)	(0.049)	(0.021)	(0.005)	(0.045)
Rmse	0.062	0.062	0.032	0.007	0.057
Kleibergen-Paap F	90.497	90.497	90.497	90.497	90.497
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.340	-0.270	-0.135	0.067**	-0.188
	(0.205)	(0.214)	(0.084)	(0.025)	(0.183)
Rmse	0.074	0.073	0.034	0.011	0.060
Kleibergen-Paap F	9.080	9.080	9.080	9.080	9.080
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). The exposure shares are computed as the origin-destination railway distance, over the total railway distance from the given origin to all destinations; the weights of destinations within macro-regions are next set to zero. All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

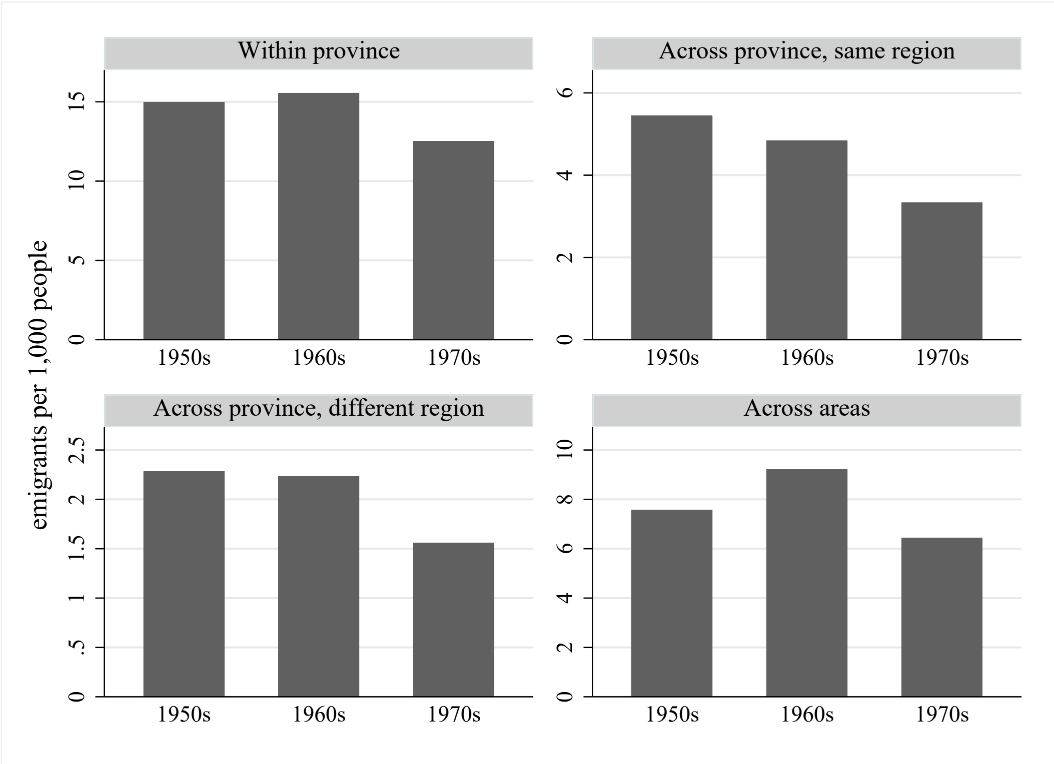
Table 14: Long-distance migration, Main Outcomes: Exposure shares summing to 1 over all destinations

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053***	0.019	-0.080***	0.008**	-0.116***
	(0.020)	(0.022)	(0.012)	(0.003)	(0.028)
Long-distance immig. rate	-0.022	0.028	-0.050***	0.000	-0.025
	(0.028)	(0.035)	(0.018)	(0.005)	(0.031)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.085**	-0.023	-0.078***	0.017***	-0.140***
	(0.039)	(0.044)	(0.020)	(0.005)	(0.038)
Rmse	0.060	0.062	0.032	0.007	0.056
Kleibergen-Paap F	61.720	61.720	61.720	61.720	61.720
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.137	-0.124	-0.046	0.034**	-0.113
	(0.135)	(0.153)	(0.065)	(0.015)	(0.111)
Rmse	0.062	0.064	0.032	0.008	0.057
Kleibergen-Paap F	14.097	14.097	14.097	14.097	14.097
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). The exposure shares are computed as the origin-destination railway distance, over the total railway distance from the given origin to all destinations. All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

9 Figures

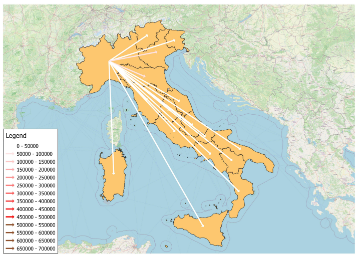
Figure 1: Internal migration by distance



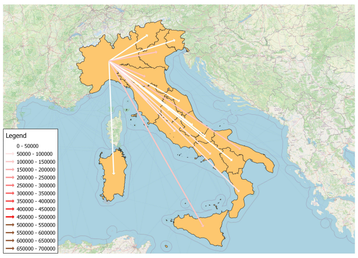
Notes: The figure shows annualized emigration rate by type of destination and decade. The number of emigrants is the number of individuals trasferring their residence: 1) within the province, 2) to another province in the same region, 3) to another region in the same area, or 4) to another area. The number of emigrants is cumulated over the periods 1955-1961 ('1950s'), 1962-1971 ('1960s'), and 1972-1981 ('1970s'). The emigration rate by type of destination is obtained dividing the number of emigrants by the total population of the country at the beginning of the period (1951, 1961, 1971). The emigration rate is annualized by dividing the ratio by 7 for the 1950s and by 10 for the other two decades. The population is expressed in thousands.

Figure 2: Long-distance emigration from Italy's macro-areas across time

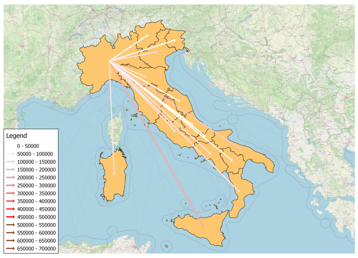
(a) North-west: decade 1



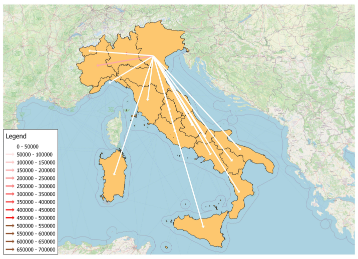
(b) North-west: decade 2



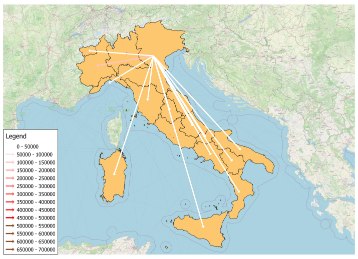
(c) North-west: decade 3



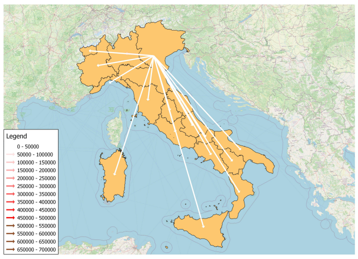
(d) North-east: decade 1



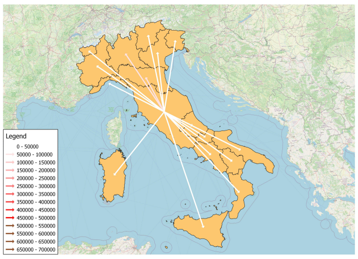
(e) North-east: decade 2



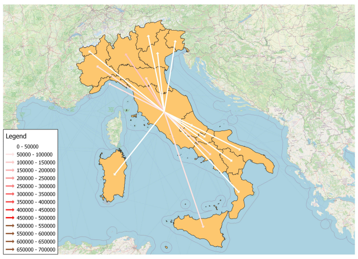
(f) North-east: decade 3



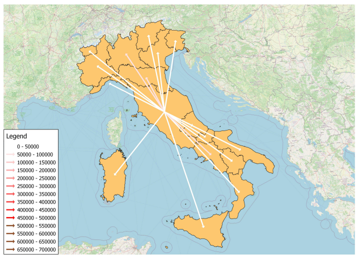
(g) Center: decade 1



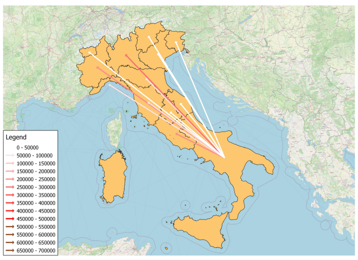
(h) Center: decade 2



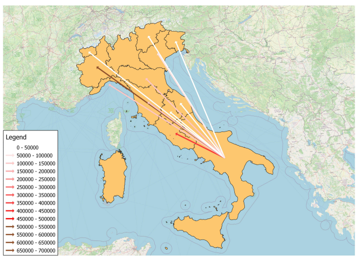
(i) Center: decade 3



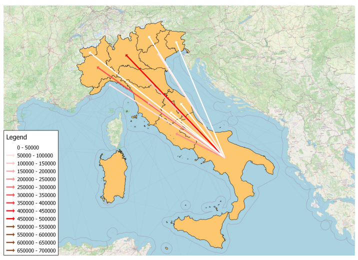
(j) South/islands: decade 1



(k) South/islands: decade 2



(l) South/islands: decade 3



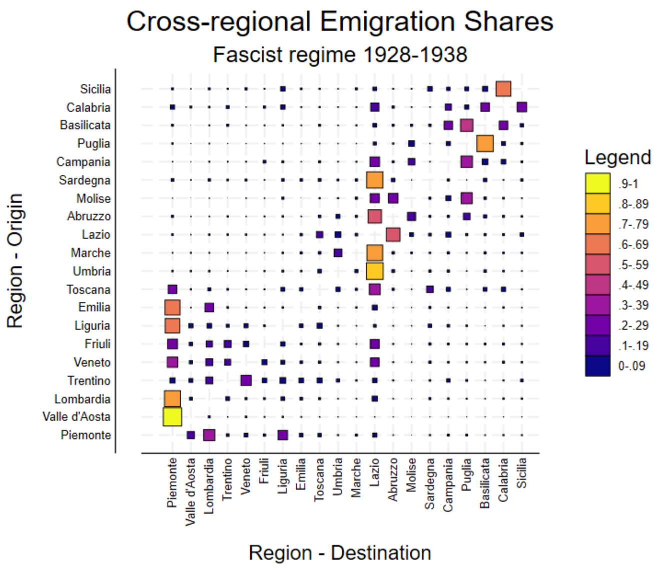
Notes: The figure shows the cumulated *long-distance* emigration flows over the 3 periods of analysis (1955-1961, 1962-1971, 1972-1981). Origin areas are represented at the macro-area level, while destination areas are represented at the region level. Each arrow indicates the total emigration from a given macro area to a given destination region over the relevant period. The color-based categorization of flows is constant across macro-areas and time periods.

Figure 3: Italian extraurban railroad at 1955

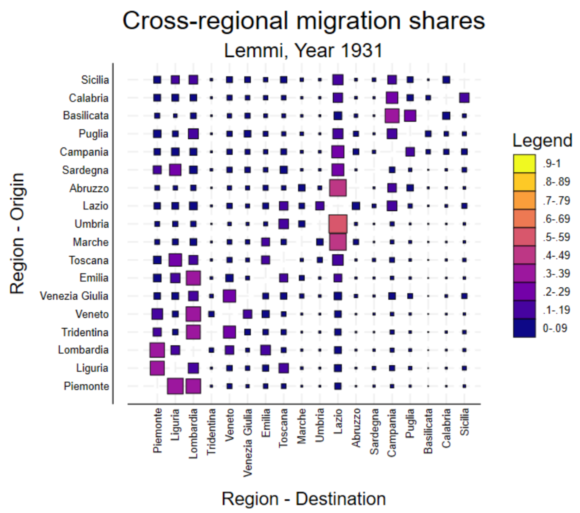
Notes: The figure shows the Italian extraurban railroad at 1955 as dygitized from Historical maps from the Italian Ministry of Transports. Continuous lines denote rail connections, dotted lines denote ferry connections.

Figure 4: Comparison of interwar workers' flows and changes of residence

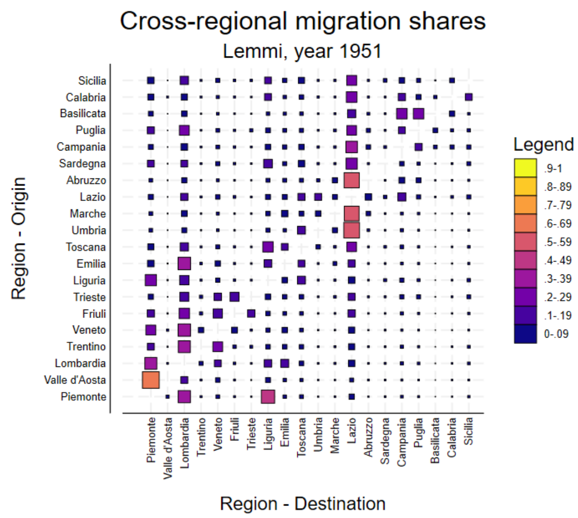
Panel A)



Panel B)

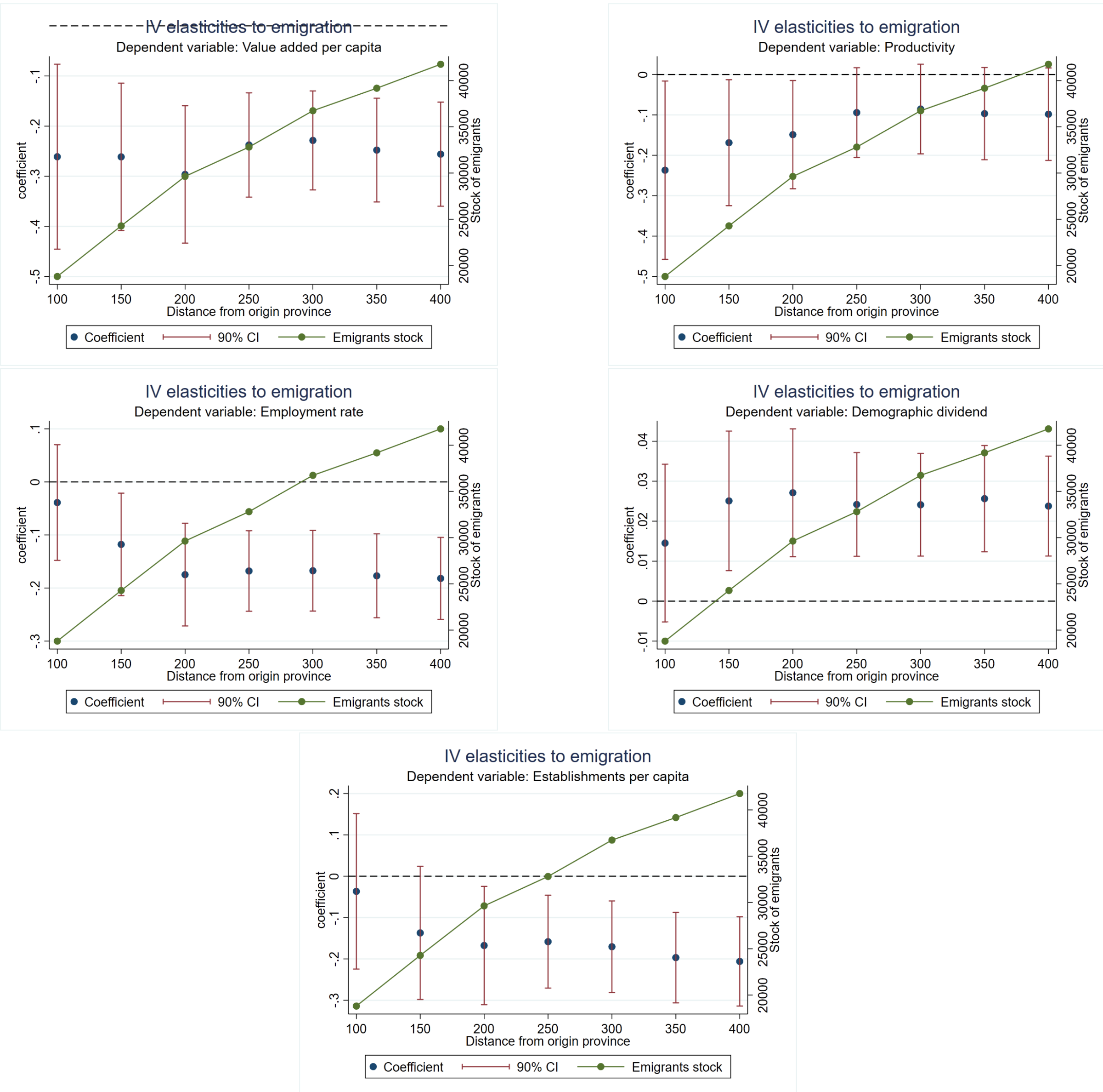


Panel C)



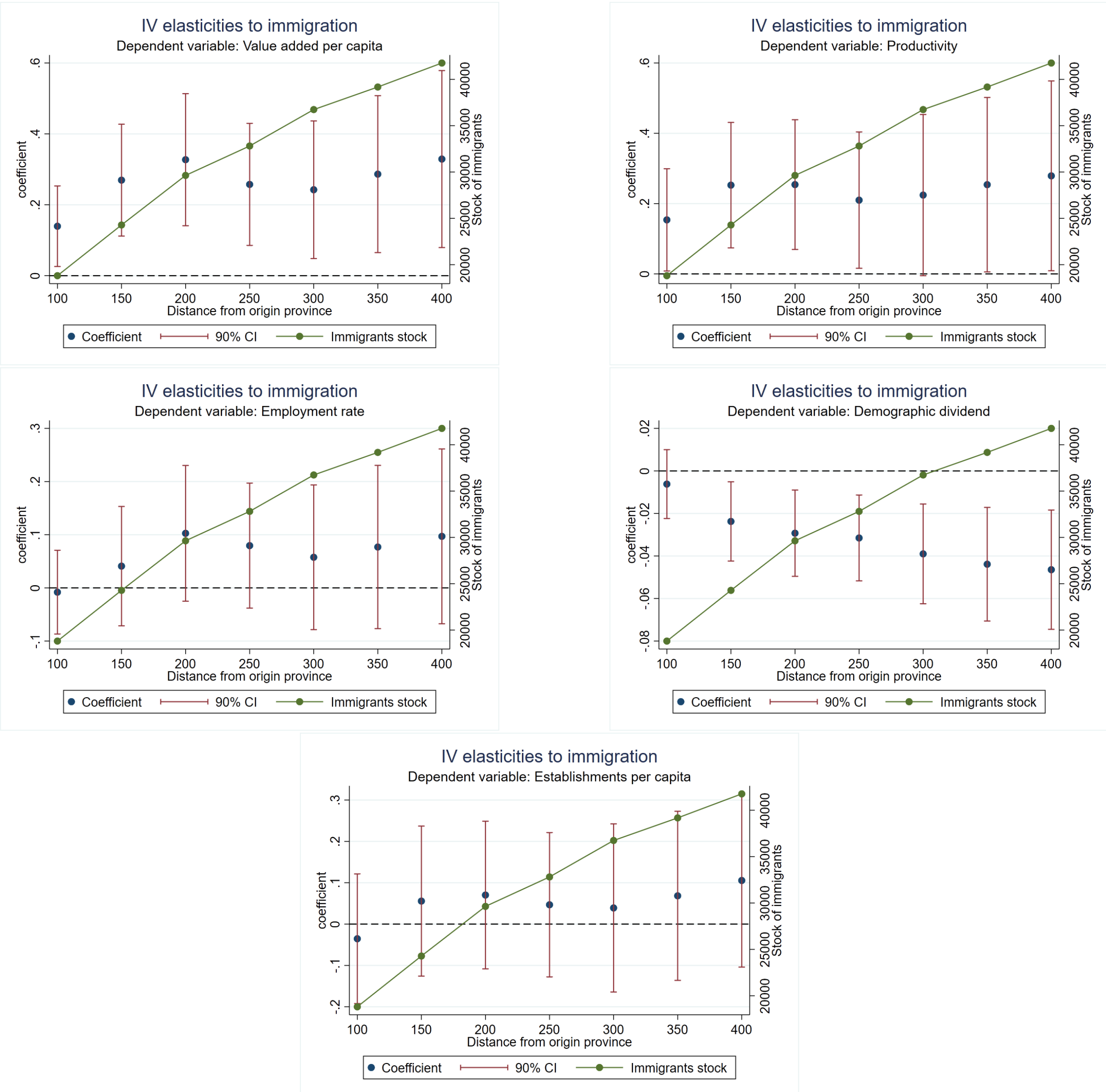
Notes: Panel A shows the migration shares resulting from the flows of workers authorized by the fascist regime, averaged over the period 1929-1938. Rows refer to the origin province, columns refer to the destination provinces. For a given origin (row), the migration shares sum to 1. For readability and comparison purposes with Lemmi (1965), the matrix only shows cross-regional migration shares. Panel B and C show the shares resulting from changes of residence according to population censuses from 1921 to 1931 and from 1931 to 1951, respectively, as reported by Lemmi (1965).

Figure 5: Short-distance emigration effects



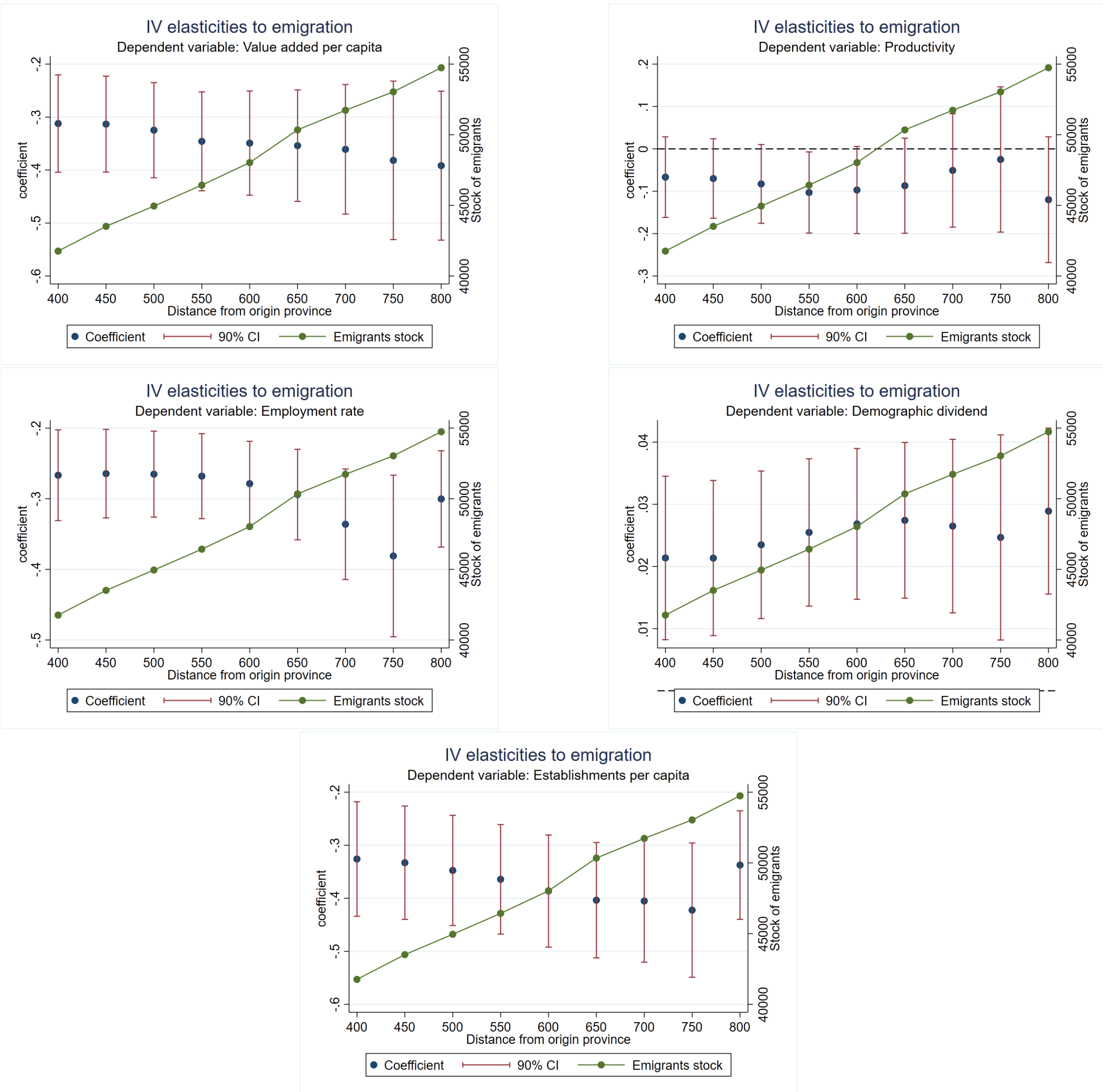
Notes: The figure shows the effects of short-distance emigration on origin provinces. In order to define short-distance migration, we consider the emigration stock of emigrants to provinces located within 100 km and then with 50 km increments until 400 km from the origin province. For each distance threshold, we construct a corresponding instrument where the weight of provinces located outside the threshold is set to zero. For each threshold and each outcome, we report the regression coefficient and 90% confidence interval from a regression of the outcome on the instrumented threshold-specific emigration rate, controlling for province fixed effects, time fixed effects, the threshold-specific immigration rate, and the interaction of GDP in 1951 with time fixed effects. The figures also report the corresponding average stock of emigrants per-province across the three decades.

Figure 6: Short-distance immigration effects



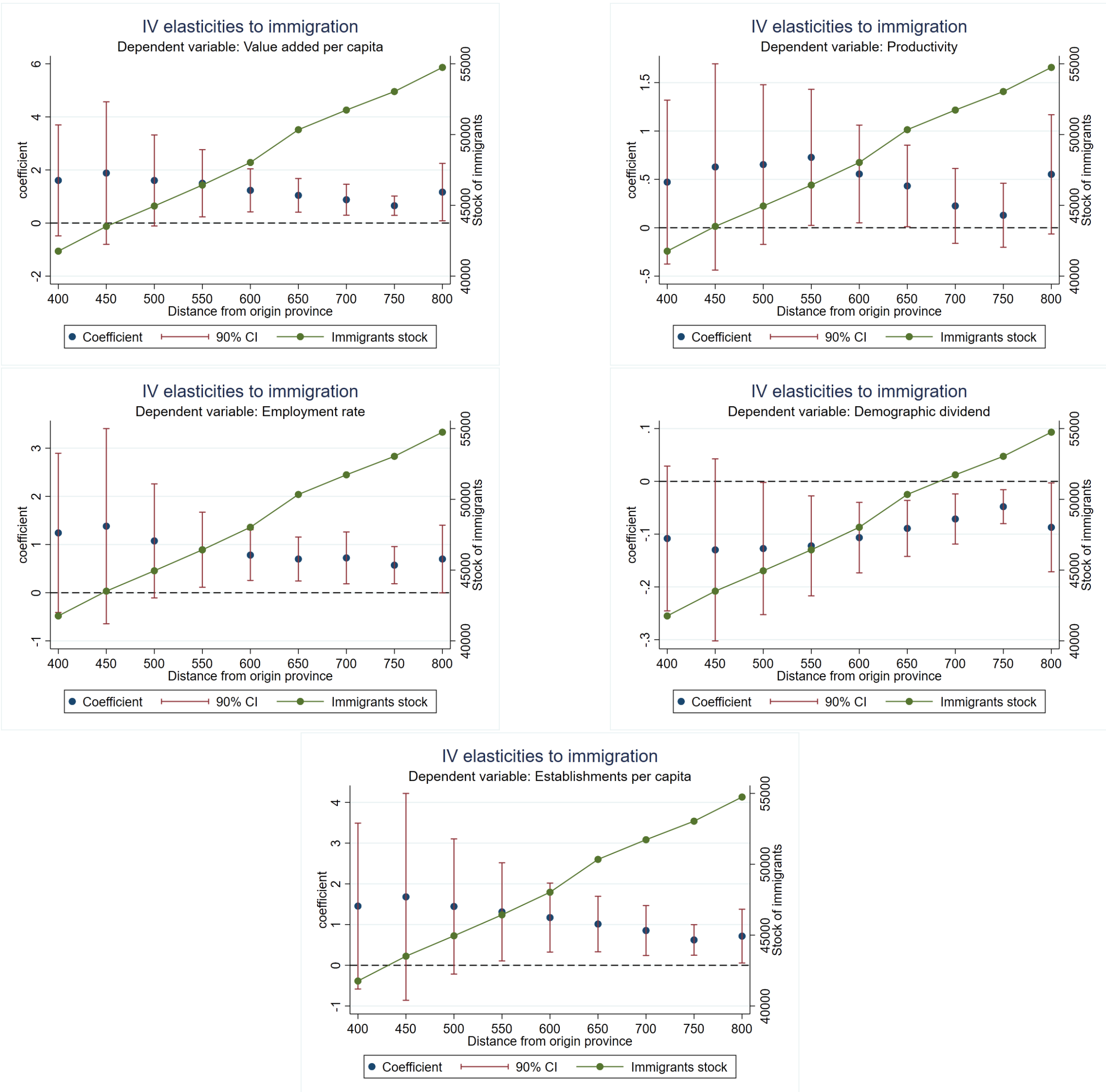
Notes: The figure shows the effects of short-distance immigration on destination provinces. In order to define short-distance migration, we consider the immigration stock from provinces located within 100 km and then with 50 km increments until 400 km from the destination province.. For each threshold, we construct a corresponding instrument where the weight of provinces located outside the threshold is set to zero. For each distance threshold and each outcome, we report the regression coefficient and 90% confidence interval from a regression of the outcome on the instrumented threshold-specific immigration rate, controlling for province fixed effects, time fixed effects, the threshold-specific emigration rate, and the interaction of GDP in 1951 with time fixed effects. The figures also report the corresponding average stock of immigrants per-province across the three decades.

Figure 7: Long-distance emigration effects



Notes: The figure shows the effects of long-distance emigration on origin provinces. In order to define long-distance migration, we consider the emigration stock to provinces located within 400 km and then with 50 km increments until 800 km from the origin province. For each distance threshold, we construct a corresponding instrument where the weight of provinces located within the threshold is set to zero. For each threshold and each outcome, we report the regression coefficient and 90% confidence interval from a regression of the outcome on the instrumented threshold-specific emigration rate, controlling for province fixed effects, time fixed effects, the threshold-specific immigration rate, and the interaction of GDP in 1951 with time fixed effects. The figures also report the corresponding average stock of emigrants per-province across the three decades.

Figure 8: Long-distance immigration effects



Notes: The figure shows the effects of long-distance immigration on destination provinces. In order to define long-distance migration, we consider the immigration stock from provinces located within 400 km and then with 50 km increments until 800 km from the origin province. For each distance threshold, we construct a corresponding instrument where the weight of provinces located within the threshold is set to zero. For each threshold and each outcome, we report the regression coefficient and 90% confidence interval from a regression of the outcome on the instrumented threshold-specific immigration rate, controlling for province fixed effects, time fixed effects, the threshold-specific emigration rate, and the interaction of GDP in 1951 with time fixed effects. The figures also report the corresponding average stock of immigrants per-province across the three decades.

References

Abhijit, B., D. Esther, and Q. Nancy (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics 145*, 102442.

Adao, R., M. Kolesar, and E. Morales (2019). Shift-share designs: Theory and inference. *Quarterly Journal of Economics 134*(4), 1949–2010.

Ager, P., M. Brueckner, and B. Herz (2017). The boll weevil plague and its effect on the southern agricultural sector, 1889–1929. *Explorations in Economic History 65*, 94–105.

Anelli, M., G. Basso, G. Ippedico, and G. Peri (2023). Emigration and entrepreneurial drain. *American Economic Journal: Applied Economics 15*(2), 218–52.

Anelli, M. and G. Peri (2017). Does emigration delay political change? evidence from italy during the great recession. *Economic Policy 32*(91), 551–96.

Antonelli, C. and F. Barbiellini Amidei (2011). *The Dynamics of Knowledge Externalities.* Cheltenham, UK-Northampton, MA, US: Edward Elgar Publishing.

Auletta, G. (1960). La disciplina legale all'urbanesimo (ostacoli ed incentivi alle migrazioni). Atti parlamentari, camera, discussioni, leg. iii, 17 feb. 1960, «XXXIII Settimana sociale dei cattolici d'Italia» cit.

Baffigi, A. (2013). National accounts, 1861-2011. In G. Toniolo (Ed.), *The Oxford Handbook of the Italian Economy since Unification.* Oxford: Oxford University Press.

Barro, R. T. and X. Sala-I-Martin (1992). Regional growth and migration: A Japan-United States comparison. *Journal of the Japanese and International Economies 6*(4), 312–346.

Becker, S. O. and A. Ferrara (2019). Consequences of forced migration: A survey of recent findings. *Labour Economics 59*, 1–16.

Biagi, B., A. Faggian, and P. McCann (2011). Long and short distance migration in italy: The role of economic, social and environmental characteristics. *Spatial Economic Analysis 6*(1), 111–131.

Bloom, D. E., D. Canning, and J. Sevilla (2003). *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change.* Population Matters Monograph MR-1274. Santa Monica: RAND.

Bodo, G. and P. Sestito (1991). *Le vie dello sviluppo. Dall'analisi del dualismo territoriale una proposta per il Mezzogiorno.* Bologna: Il Mulino.

Bonifazi, C. (2013). *L'Italia delle migrazioni*. Bologna: Il Mulino.

Bonifazi, C. and F. Heins (2000). Long–term trends of internal migration in Italy. *International Journal of Population Geography 6*, 111–131.

Borjas, G. (2003, November). The labor demand curve is downward sloping: reexamining the impacts of immigration on the labor market. *Quarterly Journal of Economics 118*(4), 1335–1374.

Borusyak, K., P. Hull, and X. Jaravel (2022, 06). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Borusyak, K., P. Hull, and X. Jaravel (2025, February). A practical guide to shift-share instruments. *Journal of Economic Perspectives 39*(1), 181–204.

Boustan, L. P., P. V. Fishback, and S. Kantor (2010). The effect of internal migration on local labor markets: American cities during the great depression. *Journal of Labor Economics 28*(4), 719–746.

Boyer, G. R. and T. J. Hatton (1997). Migration and labour market integration in late nineteenthcentury england and wales1. *Economic History Review 4*, 697–734.

Cadena, B. C. and B. K. Kovak (2016, January). Immigrants equilibrate local labor markets: Evidence from the great recession. *American Economic Journal: Applied Economics 8*(1), 257–90.

Cafiero, S. and G. E. Marciani (1972). Le zone povere nella politica di sviluppo. In A. Graziani (Ed.), *L'economia italiana: 1945-1970*, pp. 271–279. Bologna: Il Mulino.

Cappelli, G., F. Emanuele, M.-G. Julio, and T. Daniel (2018). Still a long way to go: decomposing income inequality across Italy's regions, 1871 – 2011. Working Papers 0123, European Historical Economics Society (EHES).

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Industrial and Labor Relations Review 43*(1), 245–257.

Card, D. (2005). Is the new immigration really so bad? *Economic Journal 115*(507), 300–323.

Card, D. (2009). Immigration and inequality. *American Economic Review 99*(2), 1–21.

Card, D. and J. DiNardo (2000). Do immigrant inflows lead to native outflows? *The American Economic Review 90*(2), 360–367.

Ciccarelli, C. and P. Groote (2018). The spread of railroads in italian provinces: A gis approach. *Scienze Regionali, Italian Journal of Regional Science* (2), 189–224.

Ciocca, P. (2020). *Ricchi per sempre? Una storia economica d'Italia (1796-2020)* (2 ed.). Torino: Bollati Boringhieri.

Collins, W. J. (1999). Labor mobility, market integration, and wage convergence in late 19th century india. *Explorations in Economic History 36*, 246–277.

Coluccia, D. M. and L. Spadavecchia (2021). The Economic Effects of Immigration Restriction Policies - Evidence from the Italian Mass Migration to the US. Technical report.

Commissariato (1930). Le migrazioni interne nell'anno 1929. Technical report, COMMISSARIATO PER LE MIGRAZIONI INTERNE E LA COLONIZZAZIONE.

Commissariato (1932). Le migrazioni interne in italia nell'anno 1931. Technical report, COMMISSARIATO PER LE MIGRAZIONI INTERNE E LA COLONIZZAZIONE.

Commissariato (1938). Le migrazioni nel regno e nell'africa italiana 1937-1938. Technical report, COMMISSARIATO PER LE MIGRAZIONI INTERNE E LA COLONIZZAZIONE.

Corti, P. (2003). *Storia delle migrazioni internazionali.* Roma-Bari: Laterza.

Crafts, N. and G. Toniolo (1996). *Economic Growth in Europe since 1945.* Cambridge: Cambridge University Press.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica 70*(6), 2367–2420.

Derenoncourt, E. (2022, February). Can you move to opportunity? evidence from the great migration. *American Economic Review 112*(2), 369–408.

Deryugina, T., L. Kawano, and S. Levitt (2018). The economic impact of hurricane katrina on its victims: Evidence from individual tax returns. *American Economic Journal: Applied Economics 10*(2), 202–233.

Deuchert, E. and M. Huber (2017). A cautionary tale about control variables in iv estimation. *Oxford Bulletin of Economics and Statistics 79*(3), 411–425.

Diamond, R. (2016, March). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review 106*(3), 479–524.

Dicarlo, E. (2022, January). How do firms adjust to negative labor supply shocks? evidence from migration outflows. IZA Discussion Paper 14994, Institute of Labor Economics (IZA).

Dodini, S., K. Løken, and A. Willén (2022, November). The effect of labor market competition on firms, workers, and communities. Working Paper SAM 17/2022, Institutt for samfunnsøkonomi, Department of Economics.

Donaldson, D. (2018, April). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4-5), 899–934.

Dustmann, C., T. Frattini, and A. Rosso (2015). The effect of emigration from poland on polish wages. *Scandinavian Journal of Economics 117*(2), 522–64.

Eichengreen, B. (1996). Institutions and economic growth: Europe after world war ii. In N. Crafts and G. Toniolo (Eds.), *Economic Growth in Europe since 1945*, pp. 38–72. Cambridge: Cambridge University Press.

Eichengreen, B. (2007). *The European Economy since 1945: Coordinated Capitalism and Beyond*. Princeton and Oxford: Princeton University Press.

El Badaoui, E., E. Strobl, and F. Walsh (2017). The impact of internal migration on labour market outcomes of native males in thailand. *Economic Development and Cultural Change 66*(1).

Elsner, B. (2013). Does emigration benefit the stayers? evidence from eu enlargement. *Journal of Population Economics 26*(2), 531–53.

Enflo, K., C. Lundh, and S. Prado (2014). The role of migration in regional wage convergence: Evidence from sweden 1860-1940. *Explorations in Economic History 52*, 93–110.

Fauri, F. (2015). *Storia economica delle migrazioni italiane*. Bologna: Il Mulino.

Felice, E. (2019). The roots of a dual equilibrium: Gdp, productivity, and structural change in the italian regions in the long run (1871–2011). *European Review of Economic History 23*(4), 499–528.

Franceschin, R. and J.-S. Görlach (2024, November). Asymmetric shocks and heterogeneous worker mobility. Working Paper.

Gallo, S. (2012). *Senza attraversare le frontiere: Le migrazioni interne dall'Unità a oggi*. Roma-Bari, Laterza.

Ganong, P. and D. W. Shoag (2017). Why has regional income convergence in the u.s. declined? Working Paper 23609, NBER.

Giesing, Y. and N. Laurentsyeva (2017). Firms Left Behind: Emigration and Firm Productivity. CESifo Working Paper Series 6815, CESifo.

Giordano, C. and F. Zollino (2021). Long-run factor accumulation and productivity trends in italy. *Journal of Economic Surveys 35*(3), 741–803.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *The American Economic Review 110*(8), pp. 2586–2624.

Gomellini, M. and C. O'Grada (2011). Outward and inward migrations in italy: A historical perspective. (8).

Gomellini, M. and C. Ó Gráda (2013). Migrations. In G. Toniolo (Ed.), *The Oxford Handbook of Italian Economy since Unification* (2nd ed.)., Chapter 10. Oxford University Press.

Gomellini, M., C. Ó Gráda, and G. Vecchi (2017). Migration. In G. Vecchi (Ed.), *Measuring Wellbeing: History of Italian Living Standards in Italy*, pp. 215–253. Oxford: Oxford University Press.

Graziani, A. (2000). *Lo sviluppo dell'economia italiana. Dalla ricostruzione alla moneta europea* (2 ed.). Torino: Bollati Boringhieri.

Grilli, S. and F. Zanotelli (2015). Il contributo delle migrazioni interne alle trasformazioni produttive e sociali: il caso della toscana meridionale dagli anni cinquanta ad oggi. *SIDeS, «Popolazione e Storia»* (1), 57–80.

Imbert, C. and J. Papp (2019, 03). Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India. *Journal of the European Economic Association 18*(2), 927–963.

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2022, June). Migrants and firms: Evidence from china. *American Economic Review 112*(6), 1885–1914.

Ishimaru, S. (2024). Geographic mobility of youth and spatial gaps in local college and labor market opportunities. *Journal of Labor Economics 0*.

Istat (1957a). Annuario di statistiche demografiche. 1955, Istituto Centrale di Statistica.

Istat (1957b). Istruzioni per la rilevazione statistica del movimento della popolazione. Metodi e norme b.2, Istituto Centrale di Statistica.

Istat (1969). Annuario di statistiche del lavoro e dell'emigrazione. Vol. x – 1969, Istituto Centrale di Statistica.

Istat (1998). I censimenti delle attività produttive dal 1951 al 1991. Informazioni 92, Roma.

Istat (2011). *L'Italia in 150 anni. Sommario di statistiche storiche.* Roma: Istat.

Iuzzolino, G., G. Pellegrini, and G. Viesti (2013). Regional convergence. In G. Toniolo (Ed.), *The Oxford Handbook of Italian Economy since Unification* (2nd ed.)., Chapter 20. Oxford University Press.

Jia, N., R. Molloy, C. L. Smith, and A. Wozniak (2023, forthcoming). The economics of internal migration: Advances and policy questions. *Journal of Economic Literature*.

Kennan, J. and J. R. Walker (2011). The effect of expected income on individual migration decisions. *Econometrica 79*(1), 211–251.

Kleemans, M. and J. Magruder (2017). Labour Market Responses to Immigration: Evidence from Internal Migration Driven by Weather Shocks. *The Economic Journal 128*(613), 2032–2065.

Kleinman, B., E. Liu, and S. J. Redding (2023). Dynamic spatial general equilibrium. *Econometrica 91*(2), 385–424.

Lemmi, C. (1965). Movimento migratorio interno -luogo di nascita dei censiti. Annali di Statistica Series VIII - Vol. 17, Istat.

Ma, L. and Y. Tang (2020). Geography, trade, and internal migration in china. *Journal of Urban Economics 115*, 103181.

Maggi, S. (2003). *Le ferrovie.* Bologna: Società editrice il Mulino.

Maystadt, J.-F., V. Mueller, and S. Ashwini (2016). Environmental Migration and Labor Markets in Nepal. *Journal of the Association of Environmental and Resource Economists 3*(2), 417–452.

Mocetti, S. and C. Porello (2010). How does immigration affect native internal mobility? new evidence from italy. *Regional Science and Urban Economics 40*(6), 427–439.

Molloy, R., C. L. Smith, and A. Wozniak (2011, September). Internal migration in the united states. *Journal of Economic Perspectives 25*(3), 173–96.

Monras, J. (2020). Immigration and wage dynamics: Evidence from the mexican peso crisis. *Journal of Political Economy 128*(8), 3017–3089.

Piras, R. (2017). A long–run analysis of push and pull factors of internal migration in italy. estimation of a gravity model with human capital using homogeneous and heterogeneous approaches. *Papers in Regional Sciences 96*(3), 571–602.

Pizzolato, N. (2012). 'I terroni in Cittá': Revisiting Southern Migrants' Militancy in Turin's Hot Autumn. *Contemporary European History 21*(4), 619–634.

Porzio, T. and S. Heise (2019, December). Spatial wage gaps in frictional labor markets. Discussion Paper DP14197, CEPR.

Pugliese, E. and E. Rebeggiani (2004). *Occupazione e disoccupazione in Italia. Dal dopoguerra ai giorni nostri.* Roma: Edizioni Lavoro.

Ramazzotti, A. (2021). *Il lento avvicinamento. Popolazione, ferrovie e territorio nell'Italia contemporanea.* Soveria Mannelli: Rubbettino.

Ramazzotti, A. (2024). Wage setting institutions and internal migration: The effect of regional wage equalization in italy after 1969. CSEF Working Paper 716, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.

Rosés, J. R. and B. Sánchez-Alonso (2004). Regional wage convergence in Spain 1850–1930. *Explorations in Economic History 41*(4), 404–425.

Schmutz, B. and M. Sidibé (2018, 09). Frictional Labour Mobility. *The Review of Economic Studies 86*(4), 1779–1826.

Sequeira, S., N. Nunn, and N. Qian (2020). Immigrants and the Making of America. *Review of Economic Studies 87*(1), 382–419.

Tombe, T. and X. Zhu (2019, May). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review 109*(5), 1843–72.

Tongtong, H., S. Ruiqi, T. Trevor, and Z. Xiaodong (2020). The effect of migration policy on growth, structural change, and regional inequality in china. *Journal of Monetary Economics 113*, 112–134.

Toniolo, G. (2013). *The Oxford Handbook of Italian Economy since Unification.* Oxford University Press.

Tortorici, G. (2023). Irregular citizens: Internal migrations and anti-urbanism in Italy (1955-1965). *Rivista di storia economica* (1), 63–88.

Treves, A. (1976). *Le migrazioni interne nell'Italia fascista : politica e realtà demografica.* Einaudi, Torino.

Unioncamere (2011). *Italia 150. Le radici del futuro. Il sistema camerale tra imprese e istituzioni a 150 anni dall'Unità d'Italia.* Roma: Istat.

Wozniak, A. (2010, August). Are college graduates more responsive to distant labor market opportunities? *The Journal of Human Resources 45*(4).

Zamagni, V. (2003). *Dalla periferia al centro. La seconda rinascita economica dell'Italia (1861-1990).* Bologna: Il Mulino.

Appendix A Fascist legislation on internal migration

This section provides some additional details on the legislation of internal migration in the fascist period, with special reference to the repressive side of the fascist approach.

After the immigration restrictions put in place in the USA with the "emergency quota act" of 1921 and 1924, Italian migratory flows could only have turned to the European continent or the African colonies.[29] In reality, neither one nor the other destinations prevailed. The people who opted for these foreign destinations totalled only a few hundred thousand against the millions of individuals who chose, instead, to move within the national territory from one province to another (Bonifazi 2013).

Consequently, the fascist regime was worried that this internal migratory pressure would destabilize public order and ultimately its own power. Moreover, the regime's stance on internal migration was influenced by its own demograpic campaigns for a *return to land* and a contemporary opposition to urbanization, as industrial urbanization was seen by fascism as a "destructive factor that sterilizes the population", as Mussolini claimed in a famous speech in 1927 (*"Discorso dell'Ascensione", 26th of May 1927*).

The first migration restriction saw the light just in 1927 (Royal Decree: Regulations for the Establishment of Industrial Establishments, 3 November 1927) and aimed at preventing the creation of new industrial plants in urban centres with a population of over 100,000 inhabitants. But the first law directed against immigrants towards cities was enacted at the end of 1928 and was made of a single lapidary, albeit powerless, article: "It is given faculty to the Prefect (. . .) to issue orders, having compulsory force, in order to limit the excessive increase in the city's resident population." (Law No. 2961 of 24 December 1928).

Later, in 1939, Law 1092/1939 banned anyone from migrating to an urban centre without a signed employment contract in the destination town. The '39 Act shifted the focus from groups of workers (as the 1931 Law had done, see Section 3.4) to individual workers in order to prevent emigration to urban areas where unemployment was already high. Citizens coud not freely relocate anywhere in Italy and those living in rural areas could not register on civil rolls in larger cities without having a work contract there; those who moved anyway were excluded from formal labour markets and could not rent a house nor claim any social benefits (Tortorici 2023). The latter rule remained formally in force until 1961, when it was finally superseded by Law no. 5 of 10 February 1961.

Today, historians consider these legislative interventions essentially ineffective. A

[29]The 1921 Emergency Quota Act restricted the annual number of immigrants admitted into the US to no more than 3% of the number of residents from that origin country, as recorded in the 1910 census. The 1924 Johnson-Reed Act reduced the quota to 2% and pegged the reference date to the 1890 census. These laws explicitly targeted Southern and Eastern European countries, which until the early 1900s hardly took part in the Age of Mass Migration and whose immigrants were perceived by the public as a threat to US's economic welfare and culture (Coluccia and Spadavecchia, 2021).

study by Lemmi (1965) already showed how migration flows, as measured by changes of residence between decennial censuses (1901, 1911, 1921, 1931 and 1951), were very substantial in the interwar period, in the order of millions of people. Treves (1976) speaks of "huge migration flows" of about 18 million movements between 1923 and 1939, with a peak in 1937 of 1,5 million people registered in a municipality other than that of residence. These flows were mostly directed towards the largest cities (see Chianese 2017 for Rome). In fact, the Law of 1928 remained largely inapplied and limited to some public order interventions on a few thousands of people (Treves 1976, p. 98; Gallo 2012, p. 117). Between the law of 1928 and that of 1939, there were no notable attempts to directly hinder immigration in the cities.

Appendix B Further robustness checks

In this section, we provide additional robustness tests of our main results. First, we allow for correlation in outcomes at broader geographical levels compared to provinces (our unit of analysis), testing for clustering of the standard errors at the appropriate level. Second, we allow for non-standard correlation in outcomes due to correlation in exposure shares, which is not accounted for by standard clustering strategies (Adao et al., 2019). Third, we modify the way we construct our push and full factors.

B.1 Alternative clustering strategies

Next, we allow for spatial correlation in outcomes by exploring alternative clustering strategies. In particular, our main specification, which clusters the standard errors at the province level, only allows for within-province serial correlation in outcomes, while different provinces are treated as independent from each other. This might not be true in case there are local shocks introducing positive correlation in outcomes for provinces belonging to the same geographical cluster. As such, we explore how inference is affected by clustering the standard errors at higher geographical levels. As a first exercise, we create within-region clusters of provinces by considering the regional capital as an independent cluster and creating two or three other clusters depending on the region size. This first exercise delivers 53 geographical clusters, which is above the standard rule of thumb of 40. For completeness, we also report the results when we cluster at the regional level, although we note upfront that the small number of regions (20) may artificially reduce the effective sample size. We report for brevity only our main outcomes; the results for the sectoral outcomes are qualitatively and quantitatively similar to our baseline results. Tables A1 and A2 show the results for the 1st exercise, while Tables A3 and A4 show the results for the regional clustering (for short-distance and long-distance migration respectively). The results are again in line with our baseline results, both in terms of sign and magnitude.

B.2 Exposure robust inference

In a recent contribution, Adao et al. (2019) show that inference in shift-share designs may be affected by a overrejection problem due to correlated regression residuals across regions with similar sectoral shares, independent of their geographic location. We notice upfront that in our setting this issue may be less relevant than in other contexts, as the exposure shares are defined in terms of local migration shares (short-distance) and railway distances (long-distance), which means that shift-share exposure and geographical exposure are highly correlated. As the previous exercise showed robustness of our results to alternative geographical clustering strategies, we do not anticipate substantial differences. For completeness, we perform inference using the novel methods developed in Adao et al. (2019), which are shown to be valid under arbitrary cross-regional correlation in the regression residuals. In particular, we first perform both AKM0 and AKM1 inference methods (Adao et al.; 2019). Next, we also take advantage of the possibility to cluster the standard errors at *arbitrary* level; as the definition of our baseline instruments for short and long distance migration is based on the macro-region borders, we cluster the standard errors at this level. Again, we only report for brevity the results for our main outcomes. Tables A5 and A6 show the results and indicate that inference is even more precise when we take into account the cross-province correlation in exposure shares.

B.3 Alternative transformation of our pull/push factors

Finally, we test for robustness to alternative transformations of the push and pull factors in the construction of the shift-share instruments. In particular, our baseline instruments for emigration and immigration used the direct and inverse growth rates as pull and push factors respectively, and then applied a log transformation to those in order to achieve a smooth distribution; in this section we explore robustness to alternative transformations: i) an inverse hyperbolic sine transformation; ii) a cubic root transformation. Tables A7 and A8 show the results for short distance migration, Tables A9 and A10 for long distance migration. The results are almost identical to our baseline results, both in qualitative and quantitative terms.

Appendix C Instruments' validity

In this section, we validate our shift-share instruments for short and long distance migration. In particular, we stressed that, in the context of short-distance migration, identification comes from the exogeneity of the Fascist migration shares with respect to later provinces' development; for this reason, we rely on the framework discussed in Goldsmith-Pinkham et al. (2020). On the contrary, in the context of long-distance migration, identification comes from the exogeneity of the growth rates of employment in destination provinces, which motivates relying on the framework provided in Burusyak et al. (2021).

C.1 Short-distance migration

In this section, we evaluate the plausibility of our shift-share research design for short-distance migration according to Goldsmith-Pinkham et al. (2020)'s framework. The main result of the paper is a numerical equivalence result: a two-stage least squares (TSLS) estimator with a Bartik-like instrument is numerically equivalent to a generalized method of moments (GMM) estimator with the local exposure shares as instruments and a weight matrix constructed from the shocks. The authors interpret this result as saying that using a Bartik-like instrument is equivalent, in a causal identification sense, to using the local exposure shares as instruments, so that the exogeneity condition should be interpreted in terms of the shares. In our context, the implied empirical strategy is an exposure research design, where the province-specific migration share to a given destination during the Fascist period measures the differential exposure to the common destination-specific shock.

In order to build credibility of our exposure design, we first draw on the results in Goldsmith-Pinkham et al. (2020), in turn building on Rotemberg (1983), who show that a Bartik-like estimator can be decomposed into a weighted sum of the just-identified instrumental variable estimators that use each exposure share (here, each migration share) as a separate instrument. These weights, referred to as Rotemberg weights, tell us which migration share gets more weight in the overall estimate, and thus which of these identifying assumptions is most worth testing. If the high-weight provinces (those migration shares associated with the largest Rotemberg weight) pass basic specification tests, researchers should feel reassured about the overall empirical strategy (Goldsmith-Pinkham et al., 2020). As a result, after summarizing the distribution of the Rotemberg weights, we re-perform our main regressions using overidentified IV estimators with the high-weight migration shares as separate instrument, and show the results of overidentification tests; additionally, we perform Hausman-like tests studying the correlation between the 1st stage residual and our main outcomes, once our posited channel (the emigration/immigration rates during the period 1955-1981) is controlled for. For completeness, we also show the visual diagnostic tests summarizing the distribution of First-stage F-statistic, just identified estimates and Rotemberg weights developed in Goldsmith-Pinkham et al. (2020). Next, we build on these results to look at correlates of the high-weight shares to study whether they predict outcomes through channels other than migration flows over the period 1955-1981 (our posited channel).

C.1.1 Rotemberg weights

In this section, we describe the distribution and characteristics of the Rotemberg weights (α_k). First, we notice that we have few negative Rotemberg weights, which only represent 10% of the total weights. This is reassuring, since a common concern of Goldsmith-pinkham et al. (2020) is that negative weights raise the possibility of non-convex weights on the location-specific treatment effects, in which case the overall Bartik estimate does not have a LATE-like interpretation as a weighted average of treatment effects. As suggested by Goldsmith-Pinkham et al. (2020), in A11, Panel E, we also split the instruments into those with positive and negative Rotemberg weights and compare the weighted sum of their just-identified estimates. We confirm that the weighted sum of the just-identified estimates with the negative α_k is relatively small in magnitude, so that it is unlikely that negative weights on the location-specific treatment effects end up being important in the overall estimate.

Next, in Panel B we note that the sensitivity to misspecification α_k is negatively correlated with the pull factors g_k. This is also reassuring, since it means that those provinces with the largest pull factors are the least problematic in terms of sensitivity to misspecification. Conversely, we note a relatively large positive correlation between the Rotemberg weights α_k and the variation in destination-specific emigration shares across provinces $Var(z_k)$: in other words, those provinces most subject to misspecification tend to be those whose emigration share varies more across provinces. This again confirms that indeed our identifying variation comes from variation in the emigration shares.

Panel C reports variation in the Rotemberg weights across years. We note here that most of the sensitivity to misspecification (around 0.76) comes from the period 1972-1981, which is notably the less relevant for our purposes in terms of internal migration and structural change; conversely, the previous two decades seem to be relatively robust to misspecification, with Rotemberg weights summing to about 0.25. This is again reassuring, as misspecification looks relatively innocuous for our central periods of analysis.

However, in panel D we note that the distribution of the α_k is skewed, with the top 5 provinces accounting for around 50 percent (0.57/1.127) of the positive weight, which results in a small number of instruments having a large share of the weight. This is not surprising, especially when looking at what provinces have the largest weight: Turin, Milan, Pavia (North Italy), Rome (Center Italy) and Campobasso (Molise). As a result, one of the important comparisons in our empirical design is across places within the same macro-region with greater and smaller migration shares to Milan, Rome and Turin during the Fascist period. These are mostly large provinces, with major urban centres, a developing industrial or public sector, and most of all able to attract millions of immigrants within our period of analysis. Therefore, it is not surprising that the validity of our instrument hinges on the exogeneity of the dynamics of these provinces with respect to the other less dynamic provinces within their macro-region. We address this point in the next two sections.

C.1.2 Overidentification and Hausman tests

In this section, we build on the previous results on Rotemberg weights to perform alternative estimation strategies allowing for overidentification tests and Hausman-like tests. In particular, we first consider the migration shares to the top-5 Rotemberg weights destinations (Turin, Milan, Pavia, Rome, Campobasso) and then use these migration shares (interacted with time fixed effects) as separate instruments of our main regression in place of the Bartik instrument. This specification entails 15 excluded instruments (5 provinces, 3 time periods), allowing for overidentification tests. We perform several types of estimators: overidentified TSLS, Limited-information Maximum Likelihood, and Fuller-modified LIML (together with their heteroskedasticity-robust versions). Moreover, we perform a Hausman-like test of endogeneity by using these separate instruments to generate a 1st stage residual, and then regress our main outcomes on this residual after controlling for the emigration rate. The results of this test provide a direct falsification of the exclusion restriction by checking whether the unexplained variation in our endogenous regressor predicts our outcomes, once the main posited channel is controlled for. Table A12 shows the results. The results indicate that, irrespective of the specific estimator, we fail to reject the null of exogenous instruments for most of our main outcomes (value added per-capita, productivity, employment rate, demographic dividend); we also find that our Hausman test fails to reject the null of exogenous instruments. Finally, we find more mixed results for establishments per-capita, where non-robust IV estimators reject but robust estimators do not, and where our Hausman test find marginal evidence against the exclusion restriction. Overall, we consider these results as strong evidence in favor of exogeneity of those migration shares identified as most important for misspecification concerns by the Rotemberg weights.

C.1.3 Correlates of the Fascist migration shares at 1951

In order to further support our empirical exposure design, we follow Goldsmith-Pinkham et al. (2020) and look at how the migration shares towards the top-5 Rotemberg weights destinations (Turin, Milan, Pavia, Rome and Campobasso) correlate with provinces' characteristics at 1951. In particular, we are interested in whether provinces with higher/lower exposure to a given top-5 destination according to its migration share during the Fascist period correlate with other relevant characteristics that may predict subsequent levels of our main outcomes of interest. For instance, provinces highly exposed to Milan or Turin during Fascism could also have different demographic or labour market characteristics in our base year (1951), potentially affecting income, productivity and employment over the next three decades. In order to perform our exercise, we first use data compiled from Istat's 8milaCensus, a dataset which contains detailed information on municipalities' education levels, gender differentials, demographics and labour markets at census years since 1951;[30] we combine these data with additional census data on productivity and the banking sector. For each top-5 Rotemberg weight destination, we run a provincial cross-sectional regression of the migration share towards this destination on each characteristic of interest at 1951. We notice that, due to the way we construct our baseline migration shares, we need to take into account: i) the own migration share of a province is set to zero, which motivates controlling for an indicator for the relevant top-5 destination (for instance, when regressing the migration share towards Milan, we control for an indicator for Milan); ii) the migration shares towards destinations outside the macro-region are set to zero, which is likely to introduce residuals' correlation at the macro-region level; although this would motivate clustering the standard errors at the macro-region level, due to the low number of clusters (5) we cluster at the immediate lower level (region, 20 clusters). We perform our exercise for a large variety of provinces' characteristics: education (illiteracy, share of college graduates), gender (sex ratio, gender differentials, family size), age and demographics (incidence of population below age 6, elderly and youth dependency indexes, old age index, demographic density), socio-economic indicators (incidence of owned house, utilities' availability index, rooms' settlement index), labour markets (labour market participation and employment rates, both for males and females, and sectoral employment shares), productivity (both in total and by sector), and the banking sector (branches, bank assets and deposits per capita). Table A13 shows the results and confirms very little evidence against the shares' exogeneity: out of 150 regressions, we are only able to reject the null of no correlation at 5% for one share and one outcome, namely higher productivity in the service sector for those provinces with higher exposure towards Rome. Overall, we conclude the section with enhanced confidence into the credibility of our research design.

[30]The dataset can be dowonloaded from the following website https://ottomilacensus.istat.it/ (last retrieved June 2025).

C.2 Long-distance migration

As opposed to short-distance migration, we evaluate the plausibility of our research design for long-distance migration according to the framework in Borusyak et al. (2022). This is in line with Borusyak et al. (2025) shift-share taxonomy, recommending to apply their framework in all those cases where shocks are tailored to a specific question (in our case, migration push and pull factors) while the shares are "generic," in that they could conceivably measure an observation's exposure to multiple shocks (in our case, bilateral distances along the railway network at 1955). The main result of the paper is again an equivalence result: the orthogonality between a shift-share instrument and an unobserved residual can be represented as the orthogonality between the underlying shocks and a shock-level unobservable, which captures the average unobserved determinants of the original outcome among observations most exposed to a given shock. When applied to our setting, it would mean that the push/pull factors in a given province must be uncorrelated with their average unobserved effect on provinces most exposed to that same province. We believe that this condition is a-priori more likely to hold in the context of long-distance migration, where economic conditions of distant provinces are more likely to evolve independently. In particular, we stress that our empirical strategy only hinges on exogenous variation in the shocks, allowing the variation in exposure shares, here represented by railway distances, to be endogenous. In line with Borusyak et al. (2022), in the next sections we analyse the distribution of push and pull factors to assess the plausibility of (conditional) uncorrelation in shocks, use balance tests to land credibility on quasi-random shock assignment with-respect to provinces of immigration/emigration, and use equivalent shock-level IV regressions to obtain exposure-robust inference.

C.2.1 Properties of push/pull factors and exposure shares

Our research design places particular emphasis on the variation in the shocks and their average exposure s_{nt} across provinces and periods. Table A14 reports summary statistics for our shocks computed with importance weights s_{nt}, and characterizes these weights. Recalling that our push and pull factors are computed respectively as the (log) inverse and direct growth rates of total employment over the previous decade, we first notice that due to this symmetric construction the distribution of the push and pull factors looks identical but with an opposite-sign mean. In both cases, we have significant variation with a std. dev. of 0.282 and an interquartile range of 0.360; this variation is only slightly smaller when residualizing the push and pull factors on time fixed effects. We also see low concentration of the railway-based exposure shares: when looking at the inverse Herfindal index (HHI), which corresponds to the effective sample size of our equivalent shock-level regressions, we have substantial independent information with an effective sample size around 240; this also holds when aggregating the exposure shares over provinces and provincial clusters (and periods), with effective sample sizes around 80 and 40, respectively. Finally, we notice that the largest shock weights s_{nt} are small across province-periods, and also so when aggregating at the level of province and provincial clusters, with the maximal exposure amounting to 0.8%, 2.3% and 6.5% respectively. Overall, it appears that we have substantial variation in shocks and exposure shares not only across province-years, but also within provinces and provincial clusters.

However, besides the condition on the effective sample size, finite-sample inference in the Borusyak et al. (2025) framework also requires sufficient mutual uncorrelation in the shocks. To assess its plausibility and choose the appropriate level of clustering for exposure-robust standard errors, we next analyse the correlation patterns of push and pull factors across provinces at various level of geographical aggregation. In particular, we follow Borusyak et al. (2025) and compute shocks' intra-class correlation coefficients (ICCs) using a random effect model with period fixed effects, which provides a hierarchical decomposition of residual within-period shock variation. Following convention, we estimate the model as a hierarchical linear model by maximum likelihood, assuming Gaussian residual components. Again, due to the symmetric construction of our push and pull factors, their ICC are identical, so we only report the results once; for comparison, we also report the intra-class correlations for push/pull factors constructed in a similar manner using the growth rates of GDP per-capita, workers and establishments. Table A15 shows our results. For our baseline push/pull factors, we find evidence of substantial (serial) within-province correlation, with an ICC coefficient equal to 0.313 and significant at the 1% level; however, when moving to the immediate higher level (provincial clusters), the ICC coefficient drops by almost 30% and in fact it can no longer be distinguished from zero, which also applies to higher aggregation levels. In comparison, we find little evidence of intra-class correlation for the growth rates of GDP per-capita and workers, even at the province level, while we find evidence of significant intra-class correlation in the growth rates of establishments at the level of the 53 provincial clusters (ICC coefficient 0.759, significant at 5%). Overall, the results indicate that, although we cannot reject within-province serial correlation in employment growth, we can safely consider each province as an independent source of employment shocks. In other words, it will be sufficient to cluster the standard errors at the level of individual provinces, and Table A14 confirms that at this level of aggregation we still have an adequate effective sample size.

C.2.2 Falsification tests

Following Borusyak et al. (2022), we next implement falsification tests of the shocks' orthogonality. The key concern we highlight in this section is similar to the standard pre-trends concern in a diff-in-diff setting: a province exposed to destination-level employment growth may have independently evolved along similar trends as the observed ones, even in the absence of these shocks. In particular, this may happen for two reasons: i) provinces most exposed to other provinces' employment growth may have a different start-of-period level of the outcome; ii) provinces most exposed to other provinces' employment growth may have different start-of-period levels of covariates which are able to later generate higher outcome growth. Given that our main specification is in levels (rather than growth rates), we re-calibrate the standard pre-trends test to our setting, and test whether employment growth over a given decade is able to predict start-of-period outcomes' or covariates' levels of those provinces most exposed to these shocks. We perform these tests using equivalent shock-level regressions where each lagged outcome/covariate is regressed on the shift-share instrument, which in turn is itself instrumented with province-level push/pull factors; all the regressions are weighted by the exposure weights computed as in Borusyak et al. (2022). The specification choice is guided by Borusyak et al. (2022)'s discussion of shift-share designs in a panel setting with time-invariant exposure shares summing to one: in this case, while the unit fixed effects are able to eliminate the time-invariant component of the shocks, the period fixed effects isolate within-period shock variation. Recalling that our main specification include both province and decade fixed effects, we estimate the equivalent shock-level regressions with the inclusion of province and time fixed effects. Moreover, given the results on intra-class correlations in push/pull factors reported in Table A15, we cluster the standard errors at the province level. In order to perform our exercise, we use all the outcomes reported in Tables 5 and 6 and all the covariates obtained from 8000 Census and also used in Table A13. For completeness, we report the results for all the three share computation methods explained in section 6. Again, we notice that, due to the symmetric construction of the push/pull factors, their correlation with the lagged outcomes/covariates are identical in magnitude but of opposite sign. Table A16 shows the results. Again, we find very limited evidence of systematic correlation between employment growth and lagged outcomes' or covariates' levels: out of 163 independent regressions (41 outcomes, 3 methods for the shares' computation), we only find strong evidence of some correlation in the case of bank deposits per capita: provinces most-exposed to other provinces' large employment growth also tend to have a smaller number of bank deposits per-capita at the beginning of the period. Overall, we again consider these results as strong evidence in favor of exogenous employment growth with respect to those provinces most exposed to these shocks.

C.2.3 Main estimates using exposure-robust inference

Finally, we report versions of our main estimates of Tables 5 and 6 but computed using equivalent shock-level regressions as in Borusyak et al. (2022), which has the main advantage of delivering valid exposure-robust standard errors and measure of first-stage predictive power. In particular, we obtain shift-share IV coefficients, standard errors and first-stage F statistics from shock-level IV regressions of the aggregated outcome and treatment, which is directly instrumented with our pull/push factors; all the regressions are weighted by the exposure weight s_{nt} as computed in Borusyak et al. (2022). Consistent with the above analysis of shock ICCs, we cluster standard errors at the province level. Tables A17 and A18 show the corresponding results. By construction, the coefficients have identical magnitude as in Tables 5 and 6. Reassuringly, the F-statistics are above the conventional threshold of ten, especially in the case of emigration, and standard errors are if anything smaller. The results are almost unchanged compared to our baseline ones presented in Section 5, but due to the smaller standard errors, we are now able to identify additional effects: we indeed find marginal evidence of a negative effect on productivity; when looking at sectoral outcomes, we find symmetric effects of internal migration on origins and destinations, with positive (negative) productivity effects on industrial (agricultural) productivity, and increases (decreases) in the agricultural (industrial) employment share. Overall, the results are in line with those presented in Section 5 and increase our confidence in the validity of our main estimates.

Appendix D Additional Tables

Table A1: Short-distance migration, Main Outcomes: Alternative clustering

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094***	0.011**	-0.140***
	(0.032)	(0.037)	(0.029)	(0.005)	(0.042)
Short-distance immig. rate	0.082*	0.103**	-0.012	-0.009	0.022
	(0.046)	(0.049)	(0.023)	(0.007)	(0.043)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.292***	-0.172**	-0.141***	0.021**	-0.187*
	(0.091)	(0.076)	(0.051)	(0.008)	(0.099)
Rmse	0.063	0.063	0.035	0.007	0.058
Kleibergen-Paap F	17.342	17.342	17.342	17.342	17.342
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.390***	0.348***	0.071	-0.028**	0.105
	(0.125)	(0.117)	(0.071)	(0.013)	(0.134)
Rmse	0.068	0.066	0.035	0.007	0.059
Kleibergen-Paap F	16.943	16.943	16.943	16.943	16.943
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	53	53	53	53	53

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the level of the 53 geographical clusters described in the text. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Long-distance migration, Main Outcomes: Alternative clustering

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053***	0.019	-0.080***	0.008*	-0.116***
	(0.017)	(0.021)	(0.012)	(0.004)	(0.029)
Long-distance immig. rate	-0.022	0.028	-0.050**	0.000	-0.025
	(0.030)	(0.038)	(0.022)	(0.006)	(0.034)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.095**	-0.036	-0.078***	0.020***	-0.150***
	(0.045)	(0.052)	(0.022)	(0.006)	(0.047)
Rmse	0.061	0.062	0.032	0.007	0.057
Kleibergen-Paap F	49.995	49.995	49.995	49.995	49.995
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.169	-0.164	-0.044	0.042**	-0.144
	(0.143)	(0.163)	(0.071)	(0.017)	(0.136)
Rmse	0.063	0.066	0.032	0.009	0.058
Kleibergen-Paap F	14.508	14.508	14.508	14.508	14.508
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	53	53	53	53	53

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the level of the 53 geographical clusters described in the text. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Short-distance migration, Main Outcomes: Regional clustering

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094**	0.011**	-0.140***
	(0.038)	(0.039)	(0.035)	(0.004)	(0.049)
Short-distance immig. rate	0.082	0.103	-0.012	-0.009	0.022
	(0.067)	(0.070)	(0.024)	(0.009)	(0.051)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.292**	-0.172*	-0.141**	0.021*	-0.187
	(0.112)	(0.090)	(0.055)	(0.010)	(0.125)
Rmse	0.063	0.063	0.035	0.007	0.058
Kleibergen-Paap F	13.299	13.299	13.299	13.299	13.299
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.390***	0.348***	0.071	-0.028	0.105
	(0.129)	(0.114)	(0.065)	(0.017)	(0.192)
Rmse	0.068	0.066	0.035	0.007	0.059
Kleibergen-Paap F	14.088	14.088	14.088	14.088	14.088
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	20	20	20	20	20

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the level of the 20 regions. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Long-distance migration, Main Outcomes: Regional clustering

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053**	0.019	-0.080***	0.008	-0.116***
	(0.020)	(0.025)	(0.012)	(0.006)	(0.019)
Long-distance immig. rate	-0.022	0.028	-0.050*	0.000	-0.025
	(0.033)	(0.041)	(0.025)	(0.007)	(0.028)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.095	-0.036	-0.078***	0.020**	-0.150**
	(0.063)	(0.072)	(0.024)	(0.008)	(0.054)
Rmse	0.061	0.062	0.032	0.007	0.057
Kleibergen-Paap F	22.184	22.184	22.184	22.184	22.184
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.169	-0.164	-0.044	0.042**	-0.144
	(0.179)	(0.207)	(0.079)	(0.019)	(0.154)
Rmse	0.063	0.066	0.032	0.009	0.058
Kleibergen-Paap F	14.855	14.855	14.855	14.855	14.855
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	20	20	20	20	20

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the level of the 20 regions. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Short-distance migration, Main Outcomes: Adao et al. (2019) Inference Methods

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: Emigration IV					
	-0.292	-0.172	-0.141	0.021	-0.187
AKM0 (no clustering)	(-1.72116322570e-09)	(0)	(0)	(0)	(-1.21704619500e-09)
AKM1 (no clustering)	(9.06313860595e-24)	(1.67844510724e-23)	(1.59799697057e-24)	(2.14794926723e-25)	(7.29981073968e-24)
AKM0 (clustering)	(-1.72116322570e-09)	(0)	(0)	(0)	(-1.21704619500e-09)
AKM1 (clustering)	(1.97989638391e-24)	(3.67146415479e-24)	(3.57508588441e-25)	(3.48966997875e-26)	(1.54951593055e-24)
# of clusters	5	5	5	5	5
Panel B: Immigration IV					
	0.390	0.348	0.071	-0.028	0.105
AKM0 (no clustering)	(0)	(0)	(0)	(0)	(0)
AKM1 (no clustering)	(1.36497272375e-22)	(5.30431147261e-23)	(1.34949484743e-24)	(1.24590000622e-24)	(1.28785571216e-23)
AKM0 (clustering)	(0)	(0)	(0)	(0)	(0)
AKM1 (clustering)	(8.50904267090e-23)	(3.27780393909e-23)	(9.55616298945e-25)	(7.46369225501e-25)	(8.07620870501e-24)
# of clusters	5	5	5	5	5
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
N	273	273	273	273	273

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are produced using the exposure-robust inference methods in Adao et al. (2019). Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Long-distance migration, Main Outcomes: Adao et al. (2019) Inference methods

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: Emigration IV					
	-0.107	-0.036	-0.092	0.021	-0.171
AKM0 (no clustering)	(6.76401713050e-10)	(0)	(6.76401713050e-10)	(0)	(0)
AKM1 (no clustering)	(7.94289374166e-19)	(1.03679044419e-18)	(4.72707629269e-19)	(5.30602888124e-21)	(1.63790690408e-19)
AKM0 (clustering)	(6.76401713050e-10)	(0)	(6.76401713050e-10)	(0)	(0)
AKM1 (clustering)	(6.28567153435e-20)	(7.94290254240e-20)	(3.49133966324e-20)	(3.82321425839e-22)	(1.39220941635e-20)
# of clusters	5	5	5	5	5
Panel B: Immigration IV					
	-0.172	-0.125	-0.083	0.037	-0.179
AKM0 (no clustering)	(1.30567066838e-09)	(.)	(.)	(0)	(0)
AKM1 (no clustering)	(1.81056615084e-18)	(6.30340398366e-20)	(2.03555556124e-20)	(2.58734997625e-20)	(5.22292909376e-19)
AKM0 (clustering)	(1.30567066838e-09)	(.)	(.)	(0)	(0)
AKM1 (clustering)	(1.00165309046e-19)	(4.06598649630e-21)	(1.19208361158e-21)	(1.41772062111e-21)	(2.87535032928e-20)
# of clusters	5	5	5	5	5
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
N	273	273	273	273	273

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are produced using the exposure-robust inference methods in Adao et al. (2019. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Short-distance migration, Main Outcomes: Inverse hyperbolic sine transformation

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094***	0.011**	-0.140***
	(0.030)	(0.036)	(0.023)	(0.004)	(0.035)
Short-distance immig. rate	0.082*	0.103**	-0.012	-0.009*	0.022
	(0.042)	(0.046)	(0.022)	(0.005)	(0.041)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.255***	-0.139**	-0.132***	0.016**	-0.095*
	(0.065)	(0.063)	(0.043)	(0.007)	(0.054)
Rmse	0.061	0.062	0.035	0.007	0.058
Kleibergen-Paap F	40.894	40.894	40.894	40.894	40.894
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.359***	0.334***	0.053	-0.028***	0.172
	(0.109)	(0.119)	(0.072)	(0.010)	(0.106)
Rmse	0.066	0.066	0.035	0.007	0.060
Kleibergen-Paap F	28.516	28.516	28.516	28.516	28.516
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). In the construction of the instrument, we apply a inverse hyperbolic sine transformation to the growth rates (emigration) and inverse growth rates (immigration). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Short-distance migration, Main Outcomes: Cubic root transformation

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Short-distance emig. rate	-0.116***	-0.033	-0.094***	0.011**	-0.140***
	(0.030)	(0.036)	(0.023)	(0.004)	(0.035)
Short-distance immig. rate	0.082*	0.103**	-0.012	-0.009*	0.022
	(0.042)	(0.046)	(0.022)	(0.005)	(0.041)
Rmse	0.071	0.074	0.042	0.008	0.071
N	273	273	273	273	273
Panel B: IV emigration					
Short-distance emig. rate	-0.262***	-0.143**	-0.137***	0.017**	-0.122**
	(0.066)	(0.063)	(0.042)	(0.007)	(0.056)
Rmse	0.061	0.062	0.035	0.007	0.058
Kleibergen-Paap F	40.358	40.358	40.358	40.358	40.358
N	273	273	273	273	273
Panel C: IV immigration					
Short-distance immig. rate	0.303***	0.285**	0.048	-0.031***	0.286***
	(0.082)	(0.116)	(0.080)	(0.009)	(0.088)
Rmse	0.064	0.063	0.035	0.007	0.066
Kleibergen-Paap F	32.041	32.041	32.041	32.041	32.041
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). In the construction of the instrument, we apply a cubic root transformation to the growth rates (emigration) and inverse growth rates (immigration). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Long-distance migration, Main Outcomes: Inverse hyperbolic sine transformation

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053***	0.019	-0.080***	0.008**	-0.116***
	(0.020)	(0.022)	(0.012)	(0.003)	(0.028)
Long-distance immig. rate	-0.022	0.028	-0.050***	0.000	-0.025
	(0.028)	(0.035)	(0.018)	(0.005)	(0.031)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.103***	-0.040	-0.082***	0.020***	-0.155***
	(0.038)	(0.044)	(0.019)	(0.005)	(0.041)
Rmse	0.061	0.062	0.032	0.007	0.057
Kleibergen-Paap F	84.739	84.739	84.739	84.739	84.739
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	0.106	-0.020	0.096	0.032**	0.058
	(0.162)	(0.168)	(0.070)	(0.015)	(0.141)
Rmse	0.062	0.061	0.037	0.008	0.057
Kleibergen-Paap F	13.188	13.188	13.188	13.188	13.188
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). In the construction of the instrument, we apply a inverse hyperbolic sine transformation to the growth rates (emigration) and inverse growth rates (immigration). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Long-distance migration, Main Outcomes: Cubic root transformation

	Endogenous: long-distance emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Long-distance emig. rate	-0.053***	0.019	-0.080***	0.008**	-0.116***
	(0.020)	(0.022)	(0.012)	(0.003)	(0.028)
Long-distance immig. rate	-0.022	0.028	-0.050***	0.000	-0.025
	(0.028)	(0.035)	(0.018)	(0.005)	(0.031)
Rmse	0.074	0.075	0.039	0.008	0.069
N	273	273	273	273	273
Panel B: IV emigration					
Long-distance emig. rate	-0.121***	-0.049	-0.092***	0.020***	-0.169***
	(0.038)	(0.043)	(0.019)	(0.005)	(0.041)
Rmse	0.062	0.062	0.032	0.007	0.057
Kleibergen-Paap F	84.265	84.265	84.265	84.265	84.265
N	273	273	273	273	273
Panel C: IV immigration					
Long-distance immig. rate	-0.069	-0.112	0.007	0.038***	-0.070
	(0.135)	(0.151)	(0.059)	(0.014)	(0.125)
Rmse	0.060	0.064	0.032	0.008	0.056
Kleibergen-Paap F	15.928	15.928	15.928	15.928	15.928
N	273	273	273	273	273
Province FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	91	91	91	91	91

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). In the construction of the instrument, we apply a cubic root transformation to the growth rates (emigration) and inverse growth rates (immigration). All the specifications are log-log and include province and year fixed effects, as well as the GDP per-capita in 1951 interacted with time fixed effects. Panel B additionally controls for the immigration rate, Panel C additionally controls for the immigration rate. The emigration rate is computed as the total emigration stock to all other provinces outside the province's macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces outside the economic macro-region over the previous decade, over the province's population at the previous census year, annualized with the same method. The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Summary of Rotemberg weights: Emigration

Panel A: Negative and positive weights					
	Sum	Mean	Share		
Negative	-0.127	-0.005	0.101		
Positive	1.127	0.017	0.899		
Panel B: Correlations of Province Aggregates					
	α_k	g_k	β_k	F_k	Var(z_k)
α_k	1				
g_k	-0.224	1			
β_k	-0.044	0.292	1		
F_k	0.108	0.065	-0.040	1	
Var(z_k)	0.455	-0.366	-0.097	-0.019	1
Panel C: Variation across years in α_k					
	Sum	Mean			
1961	0.194	0.002			
1971	0.048	0.001			
1981	0.758	0.008			
Panel D: Top 5 Rotemberg weight provinces					
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	Province Share
Torino	0.224	0.783	-0.070	(-0.30,0.10)	11.488
Milano	0.076	0.639	0.084	N/A	6.530
Pavia	0.084	0.191	-0.196	N/A	7.818
Roma	0.109	1.404	-0.069	(-0.50,0.50)	17.727
Campobasso	0.073	1.668	-0.036	(-0.30,0.40)	5.717
Panel E: Estimates of β_k for positive and negative weights					
	α-weighted Sum	Share of overall β	Mean		
Negative	-0.036	0.118	1.765		
Positive	-0.267	0.882	3.211		

Notes: This table reports statistics about the Rotemberg weights. Panel A reports the sum, mean and share of positive and negative weights. Panel B reports correlations between the weights ($\hat{\alpha}_k$), the destination-level growth rates (g_k), the just-identified coefficient estimates ($\hat{\beta}_k$), the first-stage F-statistic of the migration shares ($\hat{F}_k$), and the variation in the migration shares across locations (var(z_k)). Panel C reports variation in the weights across years. Panel D reports the top 5 provinces according to the Rotember weights. The g_k is the pull factor, $\hat{\beta}_k$ is the coefficient from the just-identified regression, the 95% confidence interval is the weak instrument robust confidence interval using the method from Chernozhukhov and Hansen (2008) over a range from -10 to 10, and *Province share* is the migration share (multiplied by 100 for legibility). Panel E reports statistics about how the values of $\hat{\beta}_k$ vary with the positive and negative Rotemberg weights.

Table A12: Short-distance migration: Overidentification and Hausman Tests

	Endogenous: short-distance emigration rate				
	Value added per-capita	Productivity	Empl. rate	Dem. dividend	Plants per-capita
	(1)	(2)	(3)	(4)	(5)
Panel A: overidentified IV					
Short-distance emig.rate	-0.040	0.006	-0.045	-0.000	-0.015
	(0.047)	(0.047)	(0.028)	(0.005)	(0.049)
Sargan	2.893	3.580	5.619	5.330	12.454
P-value	0.968	0.937	0.777	0.805	0.189
Panel B: Het. overidentified IV					
Short-distance emig.rate	-0.040	0.006	-0.045	-0.000	-0.015
	(0.036)	(0.036)	(0.029)	(0.004)	(0.041)
Hansen J	6.384	8.024	7.460	6.702	14.160
P-value	0.701	0.532	0.589	0.668	0.117
Panel C: LIML IV					
Short-distance emig.rate	-0.038	0.006	-0.041	-0.001	0.004
	(0.048)	(0.048)	(0.029)	(0.006)	(0.053)
Anderson-Rubin	2.907	3.604	5.654	5.364	12.615
P-value	0.968	0.936	0.774	0.801	0.181
Panel D: Het. LIML IV					
Short-distance emig.rate	-0.038	0.006	-0.041	-0.001	0.004
	(0.037)	(0.037)	(0.030)	(0.004)	(0.048)
Hansen J	6.363	8.022	7.453	6.497	13.648
P-value	0.703	0.532	0.590	0.689	0.135
Panel E: Fuller IV					
Short-distance emig.rate	-0.039	0.006	-0.042	-0.001	0.001
	(0.048)	(0.048)	(0.029)	(0.006)	(0.052)
Anderson-Rubin	2.907	3.604	5.654	5.364	12.615
P-value	0.968	0.936	0.774	0.801	0.181
Panel F: Het. Fuller IV					
Short-distance emig.rate	-0.039	0.006	-0.042	-0.001	0.001
	(0.037)	(0.037)	(0.030)	(0.004)	(0.047)
Hansen J	6.375	8.023	7.456	6.564	13.730
P-value	0.702	0.532	0.590	0.682	0.132
Panel G: Hausman test					
First stage residual	-0.068	-0.006	-0.076	0.014*	-0.134*
	(0.064)	(0.064)	(0.052)	(0.008)	(0.079)
N	273	273	273	273	273

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log and include province and year fixed effects; all specifications also control for gdp per-capita, the share of employment in agriculture and industry, and the share of value added in agriculture and industry in 1951 (interacted with time fixed effects). The emigration rate is computed as the total emigration stock to all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized by dividing for 7 in the case of 1961 and for 10 for 1971 and 1981. The immigration rate is computed as the total immigration stock from all other provinces within the macro-region through the previous decade, over the province's population at the previous census year, annualized with the same method. Panels A, C and E use overidentified TSLS, LIML IV and Fuller-modified LIML IV estimators; panels B, D and F use their heteroskedasticity-robust versions. Panel G performs a Hausman-like test. The instruments are the migration shares to the top-5 Rotemberg weights provinces (Turin, Milan, Pavia, Rome, Enna) interacted with time fixed effects. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Short-distance migration: Correlates of Fascist migration shares at 1951

	Dependent: migration shares				
	Turin	Milan	Pavia	Rome	Campobasso
	(1)	(2)	(3)	(4)	(5)
Panel A: Education and gender					
Analfabetism	-0.316	-0.213	-0.181	0.078	0.162
	(0.220)	(0.165)	(0.135)	(0.218)	(0.100)
College graduates	-0.176	0.227	1.022	-1.697	-0.457
	(0.394)	(0.221)	(1.034)	(1.795)	(0.376)
Sex ratio	0.049	-0.516	-0.279	1.411	-0.314
	(0.173)	(0.455)	(0.186)	(0.943)	(0.283)
Educ. gender differential	-0.066	-0.030	-0.029	-0.045	0.041*
	(0.052)	(0.024)	(0.024)	(0.051)	(0.022)
Avg. family size	-7.831	-1.494	-0.101	5.395	1.470
	(4.953)	(1.086)	(0.823)	(3.535)	(1.215)
Panel B: Age and demographics					
Incidence pop. less than 6	-1.215	-0.452	-0.550	-0.256	0.524*
	(0.941)	(0.356)	(0.445)	(0.887)	(0.254)
Elderly dependence index	2.194	0.156	-0.229	-0.436	-0.392*
	(1.422)	(0.160)	(0.462)	(0.705)	(0.207)
Youth dependence index	-0.387	-0.159	-0.188	-0.129	0.159*
	(0.305)	(0.126)	(0.152)	(0.269)	(0.077)
Old age index	0.292	0.055	0.042*	-0.058	-0.070*
	(0.169)	(0.046)	(0.022)	(0.109)	(0.036)
Demographic density	-0.001	0.000	0.003	-0.003	0.001
	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)
Incidence owned houses	0.112	-0.098	-0.175	0.026	0.044
	(0.133)	(0.084)	(0.151)	(0.219)	(0.055)
Utilities' availability index	-0.015	0.077	0.075	-0.091	-0.078
	(0.068)	(0.059)	(0.068)	(0.160)	(0.051)
Rooms' ettlement index	-0.068	-0.028	-0.017	-0.061	0.038*
	(0.053)	(0.023)	(0.013)	(0.046)	(0.019)
Panel C: Labour markets					
Labor mkt participation	0.338	0.228	0.113	0.315	-0.038
	(0.302)	(0.224)	(0.084)	(0.410)	(0.078)
Labor mkt participation (M)	1.915	-0.045	-0.055	2.391*	-0.078
	(1.769)	(0.136)	(0.138)	(1.243)	(0.295)
Labor mkt participation (F)	0.110	0.142	0.069	0.048	-0.005
	(0.092)	(0.135)	(0.048)	(0.223)	(0.039)
Employment rate	0.359	0.128	0.054	0.371	-0.006
	(0.353)	(0.140)	(0.056)	(0.409)	(0.064)
Employment rate (M)	1.198	-0.212	-0.100	1.303	0.002
	(1.194)	(0.200)	(0.120)	(0.799)	(0.231)
Employment rate (F)	0.111	0.112	0.047	0.074	0.010
	(0.104)	(0.111)	(0.036)	(0.230)	(0.035)
Incidence in employment (Agric.)	0.013	-0.155	-0.105	0.240	0.064
	(0.060)	(0.123)	(0.077)	(0.156)	(0.041)
Incidence in employment (Ind.)	0.022	0.332	0.179*	-0.276	-0.081
	(0.063)	(0.207)	(0.087)	(0.195)	(0.047)
Incidence in employment (Comm.)	0.014	0.191	0.496	-1.410*	-0.239
	(0.128)	(0.157)	(0.470)	(0.784)	(0.189)
Incidence in employment (Extra-comm.)	-0.206	-0.098	-0.050	-0.358	-0.109
	(0.239)	(0.131)	(0.044)	(0.262)	(0.103)
Panel D: Productivity and technology					
Productivity	-0.011	0.464	0.559	-0.840	-0.274
	(0.256)	(0.364)	(0.481)	(0.530)	(0.185)
Productivity (Agric.)	-0.241	0.845	0.652	-1.604*	-0.176
	(0.545)	(0.730)	(0.591)	(0.891)	(0.193)
Productivity (Ind.)	0.351*	0.265	0.330	-0.719	-0.211
	(0.193)	(0.197)	(0.297)	(0.461)	(0.140)
Productivity (Serv.)	-0.216	0.028	0.273	1.541**	-0.426*
	(0.386)	(0.091)	(0.268)	(0.684)	(0.237)
Branches per capita	5.2e+04	1.8e+04	1.9e+04	6.3e+04*	-2.3e+04*
	(4.1e+04)	(1.7e+04)	(1.5e+04)	(3.3e+04)	(1.2e+04)
Assets per capita	26.917	80.039	147.957	-55.711	-43.595
	(21.635)	(73.537)	(100.153)	(74.143)	(26.601)
Deposits per capita	51.280	70.745	95.244	-87.109	-30.648
	(41.778)	(60.978)	(64.853)	(59.722)	(19.142)
Model	OLS	OLS	OLS	OLS	OLS
Std. errors	cluster	cluster	cluster	cluster	cluster
# of clusters	20	20	20	20	20

Notes: The sample consists of the 91 provinces in 1951. Data for panels A-C come from 8000 Census, while data for panel D come from our Census data. Each cell represents the coefficient of a separate cross-sectional regression where we regress the migration towards a given top-5 Rotemberg weight destination on a given province characteristic at 1951. Each column refers to a different top-5 Rotemberg weight destination (Turin, Milan, Pavia, Rome, Enna). The standard errors are robust to heteroskedasticity and clustered at the region level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A14: Long-distance migration: distribution of push and pull factors

	Growth rate of total employment			
	Pull factor		Push factor	
	Without time FE	With time FE	Without time FE	With time FE
	(1)	(2)	(3)	(4)
Mean	1.327	0.000	-1.327	-0.000
Std. dev.	0.282	0.239	0.282	0.239
IQR	0.360	0.291	0.360	0.291
HHI (total)	239.973	239.973	239.973	239.973
HHI (province)	79.991	79.991	79.991	79.991
HHI (provincial clusters)	38.927	38.927	38.927	38.927
HHI (region)	13.008	13.008	13.008	13.008
HHI (ripartition)	4.874	4.874	4.874	4.874
Max share (total)	0.008	0.008	0.008	0.008
Max share (province)	0.023	0.023	0.023	0.023
Max share (provincial clusters)	0.065	0.065	0.065	0.065
Max share (region)	0.184	0.184	0.184	0.184
Max share (ripartition)	0.242	0.242	0.242	0.242
# Provinces	91	91	91	91
# Provincial clusters	53	53	53	53
# Regions	20	20	20	20
# Ripartitions	5	5	5	5

Notes: This table summarizes the distribution of push and pull factors across provinces n and decades t. Push and pull factors are computed as the (log) inverse and direct growth rates of total employment over the previous decade, respectively. All statistics are weighted by the average exposure shares s_{nt} as computed in Burusjack et al. (2022); shares are computed based on bilateral distances along the 1955 railway network, as described in Section 4. Columns 2 and 4 residualize the pull and push factors on period indicators. We report the effective sample size (the inverse renormalized Herfindahl index of the s_{nt} weights, as described in Burusjack et al. (2022)) at various levels (aggregated across periods): province-by-period, province, 53 provincial clusters, region, macro-region; for each level, we also report the largest s_{nt}.

Table A15: Long-distance migration: Push/pull factors intra-class correlation

	Push/Pull factors in growth rate			
	Total Employment	GDP per capita	Workers	Plants
	(1)	(2)	(3)	(4)
Economic ripartition	0.156	0.142	0.118	0.110
	(0.175)	(0.124)	(0.200)	(0.151)
Region	0.070	0.115	0.036	0.010
	(0.252)	(0.515)	(0.244)	(0.259)
Provincial clusters	0.223	0.243	0.140	0.759**
	(0.302)	(0.365)	(0.152)	(0.385)
Province	0.313***	0.083	0.014	0.017
	(0.110)	(0.076)	(0.035)	(0.020)
sep				
Model	Mixed	Mixed	Mixed	Mixed
Covariance matrix	Exchangeable	Exchangeable	Exchangeable	Exchangeable
Time FE	YES	YES	YES	YES
N	273	273	273	273

Notes: This table reports intra-class correlation coefficients for our push/pull factors, computed as the (log) inverse/direct growth rates of total employment. For comparison, we also report intra-class correlations for push/pull factors constructed in a similar manner using the growth rate of GDP per-capita, workers and establishments. The estimated coefficients come from a hierarchical mixed model with period fixed effects estimated using a maximum likelihood procedure and assuming an exchangeable covariance structure for each administrative level's random effect. Robust standard errors are reported in parentheses.

Table A16: Long-distance migration: Shock balance tests

	Pull factor			Push factor		
	Version 1	Version 2	Version 3	Version 1	Version 2	Version 3
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Lagged outcomes						
Value added per capita	0.745	0.614	-1.465	-0.745	-0.614	1.465
	(0.526)	(0.412)	(4.281)	(0.526)	(0.412)	(4.281)
Productivity	0.617	0.504	-1.255	-0.617	-0.504	1.255
	(0.536)	(0.483)	(4.096)	(0.536)	(0.483)	(4.096)
Employment rate	0.107	0.095	-0.196	-0.107	-0.095	0.196
	(0.240)	(0.181)	(0.201)	(0.240)	(0.181)	(0.201)
Demographic dividend	0.022	0.016	-0.015	-0.022	-0.016	0.015
	(0.045)	(0.032)	(0.015)	(0.045)	(0.032)	(0.015)
Workers per capita	0.209	0.179	-0.637	-0.209	-0.179	0.637
	(0.447)	(0.378)	(2.071)	(0.447)	(0.378)	(2.071)
Plants per capita	-0.219	-0.168	0.180	0.219	0.168	-0.180
	(0.438)	(0.313)	(0.336)	(0.438)	(0.313)	(0.336)
Employment share (agric.)	-0.594	-0.472	1.137	0.594	0.472	-1.137
	(0.509)	(0.328)	(2.836)	(0.509)	(0.328)	(2.836)
Employment share (ind.)	1.487	1.211	-2.592	-1.487	-1.211	2.592
	(1.299)	(0.790)	(6.183)	(1.299)	(0.790)	(6.183)
Employment share (serv.)	0.242	0.201	-0.576	-0.242	-0.201	0.576
	(0.409)	(0.363)	(2.152)	(0.409)	(0.363)	(2.152)
Value added share (agric.)	-1.492	-1.235	2.753	1.492	1.235	-2.753
	(1.235)	(0.769)	(6.712)	(1.235)	(0.769)	(6.712)
Value added share (ind.)	1.654	1.374	-3.182	-1.654	-1.374	3.182
	(1.431)	(0.888)	(7.543)	(1.431)	(0.888)	(7.543)
Value added share (serv.)	0.827	0.676	-1.023	-0.827	-0.676	1.023
	(0.790)	(0.482)	(2.378)	(0.790)	(0.482)	(2.378)
Productivity (agric.)	-0.278	-0.257	0.355	0.278	0.257	-0.355
	(0.888)	(0.693)	(0.495)	(0.888)	(0.693)	(0.495)
Productivity (ind.)	0.780	0.662	-1.835	-0.780	-0.662	1.835
	(0.592)	(0.506)	(5.429)	(0.592)	(0.506)	(5.429)
Productivity (serv.)	1.204	0.980*	-1.706	-1.204	-0.980*	1.706
	(0.962)	(0.558)	(4.326)	(0.962)	(0.558)	(4.326)
Panel B: Education and gender						
Analfabetism	-59.677	-49.338	106.287	59.677	49.338	-106.287
	(51.790)	(31.405)	(249.858)	(51.790)	(31.405)	(249.858)
College graduates	0.323	0.347	-1.672	-0.323	-0.347	1.672
	(4.082)	(3.352)	(10.794)	(4.082)	(3.352)	(10.794)
Sex ratio	0.842	0.728	0.675	-0.842	-0.728	-0.675
	(0.873)	(0.682)	(0.835)	(0.873)	(0.682)	(0.835)
Educ. gender differential	-333.702	-275.265*	578.644	333.702	275.265*	-578.644
	(279.937)	(166.888)	(1379.422)	(279.937)	(166.888)	(1379.422)
Avg. family size	1.607	1.316	-1.789	-1.607	-1.316	1.789
	(2.079)	(1.380)	(3.364)	(2.079)	(1.380)	(3.364)
Panel C: Age and demographics						
Incidence pop. less than 6	-7.232	-6.058*	13.186	7.232	6.058*	-13.186
	(5.415)	(3.302)	(33.205)	(5.415)	(3.302)	(33.205)
Elderly dependence index	-2.058	-1.569	1.551	2.058	1.569	-1.551
	(5.217)	(3.801)	(2.200)	(5.217)	(3.801)	(2.200)
Youth dependence index	8.533	6.448	-10.427	-8.533	-6.448	10.427
	(11.775)	(7.599)	(18.964)	(11.775)	(7.599)	(18.964)
Old age index	-39.688	-31.489	50.543	39.688	31.489	-50.543
	(53.979)	(35.545)	(90.135)	(53.979)	(35.545)	(90.135)
Demographic density	1930.969	1539.163	-3.0e+03	-1.9e+03	-1.5e+03	3020.501
	(1882.395)	(1133.944)	(6800.369)	(1882.395)	(1133.944)	(6800.369)
Incidence owned houses	-49.532	-40.437	74.737	49.532	40.437	-74.737
	(56.352)	(36.204)	(154.193)	(56.352)	(36.204)	(154.193)
Utilities' availability index	-24.172	-18.924	27.307	24.172	18.924	-27.307
	(81.636)	(61.615)	(40.850)	(81.636)	(61.615)	(40.850)
Rooms' settlement index	-572.133	-478.336	1002.126	572.133	478.336	-1.0e+03
	(611.961)	(396.706)	(2157.314)	(611.961)	(396.706)	(2157.314)
Panel D: Labour markets						
Labor mkt participation	3.329	2.878	-4.740	-3.329	-2.878	4.740
	(4.628)	(3.247)	(9.394)	(4.628)	(3.247)	(9.394)
Labor mkt participation (M)	-26.612	-22.207	48.745	26.612	22.207	-48.745
	(25.725)	(16.356)	(109.636)	(25.725)	(16.356)	(109.636)
Labor mkt participation (F)	30.737	25.848	-54.465	-30.737	-25.848	54.465
	(31.307)	(20.136)	(119.267)	(31.307)	(20.136)	(119.267)
Employment rate	-0.629	-0.419	3.001	0.629	0.419	-3.001
	(3.851)	(3.247)	(12.804)	(3.851)	(3.247)	(12.804)
Employment rate (M)	-28.828	-24.010	52.576	28.828	24.010	-52.576
	(25.345)	(15.664)	(123.109)	(25.345)	(15.664)	(123.109)
Employment rate (F)	25.228	21.216	-43.161	-25.228	-21.216	43.161
	(27.773)	(18.196)	(90.936)	(27.773)	(18.196)	(90.936)
Incidence in employment (Agric.)	-25.609	-20.885	49.624	25.609	20.885	-49.624
	(17.870)	(13.376)	(141.613)	(17.870)	(13.376)	(141.613)
Incidence in employment (Ind.)	26.799	21.633	-47.160	-26.799	-21.633	47.160
	(22.457)	(13.742)	(116.164)	(22.457)	(13.742)	(116.164)
Incidence in employment (Comm.)	-15.121	-12.415	22.323	15.121	12.415	-22.323
	(18.817)	(12.411)	(43.350)	(18.817)	(12.411)	(43.350)
Incidence in employment (Extra-comm.)	13.921*	11.658**	-24.769	-13.921*	-11.658**	24.769
	(8.001)	(4.972)	(68.094)	(8.001)	(4.972)	(68.094)
Panel E: technology						
Branches per capita	0.000	0.000*	-0.000	-0.000	-0.000*	0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Assets per capita	-0.293	-0.237	0.399	0.293	0.237	-0.399
	(0.292)	(0.178)	(0.901)	(0.292)	(0.178)	(0.901)
Deposits per capita	-0.233***	-0.196***	0.334	0.233***	0.196***	-0.334
	(0.071)	(0.042)	(1.086)	(0.071)	(0.042)	(1.086)
Destination FE	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES
Std. errors	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
# of clusters	91	91	91	91	91	91
Model	IV	IV	IV	IV	IV	IV
N	273	273	273	273	273	273

Notes: This table reports coefficients from shock-level regressions of lagged outcomes/covariates on the shift-share instrument, which is itself instrumented with province-level push/pull factors. All the regressions include province and time fixed effects and are weighted by the exposure weights computed as in Borusjack et al. (2022). The standard errors are clustered at the province level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Long-distance migration: Exposure-robust inference (Main outcomes)

	Endogenous: local emigration/immigration rate				
	Value-added pc	Productivity	Employment rate	Demographic dividend	Establishments pc
	(1)	(2)	(3)	(4)	(5)
Panel A: IV emigration					
Long-distance emig. rate	-0.095**	-0.036	-0.078***	0.020***	-0.150***
	(0.037)	(0.024)	(0.023)	(0.002)	(0.033)
Exposure robust 1st stage F	47.160	47.160	47.160	47.160	47.160
Panel B: IV immigration					
Long-distance immig. rate	-0.169	-0.164	-0.044	0.042***	-0.144
	(0.132)	(0.112)	(0.081)	(0.012)	(0.104)
Exposure robust 1st stage F	10.703	10.703	10.703	10.703	10.703
Destination FE	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES
Std. errors	Cluster	Cluster	Cluster	Cluster	Cluster
# of clusters	91	91	91	91	91
Model	IV	IV	IV	IV	IV
N	273	273	273	273	273

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log, include province and year fixed effects and control for the interaction of GDP per capita in 1951 and time fixed effects; additionally, panel A controls for the immigration rate, panel B controls for the emigration rate. All the coefficients, standard errors and first stage F statistics are obtained from equivalent shock-level regressions controlling for province and year fixed effects and weighted using the exposure weights computed as in Borusjack et al. (2022). The standard errors are clustered at the province level. The emigration (immigration) rate is computed as the total emigration (immigration) stock to (from) long-distance destinations (origins) over the previous decade, over a province's population at the previous census year. The instrument is computed as a weighted average of destinations' (origins') occupation (inverse) growth rates, where the weight is given by the rail-based proximity index. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

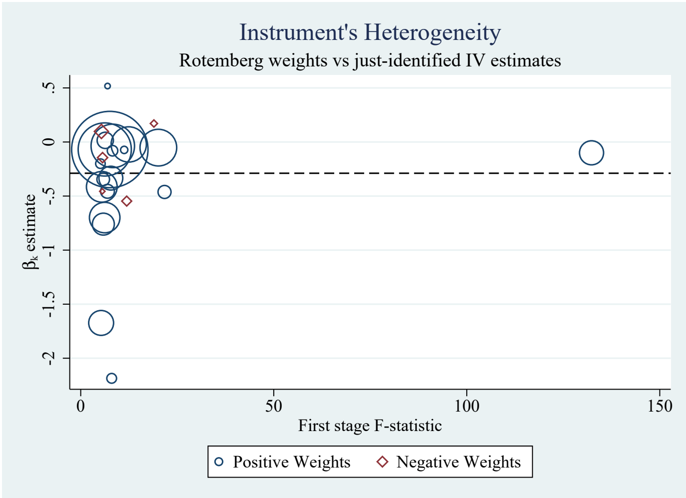
Table A18: Long-distance migration: Exposure-robust inference (Sectoral outcomes)

	Endogenous: local emigration/immigration rate								
	Employment (share)			Value added (share)			Productivity		
	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: IV emigration									
Long-distance emig. rate	0.329***	-0.130***	0.056*	0.020	-0.000	0.048*	-0.336***	0.091**	-0.043
	(0.059)	(0.025)	(0.032)	(0.046)	(0.038)	(0.026)	(0.073)	(0.039)	(0.029)
Exposure robust 1st stage F	47.160	47.160	47.160	47.160	47.160	47.160	47.160	47.160	47.160
Panel B: IV immigration									
Long-distance immig. rate	0.867***	-0.215**	0.001	0.219*	0.256*	0.031	-0.771**	0.299**	-0.132
	(0.244)	(0.105)	(0.128)	(0.130)	(0.153)	(0.090)	(0.385)	(0.145)	(0.102)
Exposure robust 1st stage F	10.703	10.703	10.703	10.703	10.703	10.703	10.703	10.703	10.703
Destination FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Std. errors	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
# of clusters	91	91	91	91	91	91	91	91	91
Model	IV	IV	IV	IV	IV	IV	IV	IV	IV
N	273	273	273	273	273	273	273	273	273

Notes: The sample consists of the 91 provinces at census years (1961, 1971, 1981). All the specifications are log-log, include province and year fixed effects and control for the interaction of GDP per capita in 1951 and time fixed effects; additionally, panel A controls for the immigration rate, panel B controls for the emigration rate. All the coefficients, standard errors and first stage F statistics are obtained from equivalent shock-level regressions controlling for province and year fixed effects and weighted using the exposure weights computed as in Borusjack et al. (2022). The standard errors are clustered at the province level. The emigration (immigration) rate is computed as the total emigration (immigration) stock to (from) long-distance destinations (origins) over the previous decade, over a province's population at the previous census year. The instrument is computed as a weighted average of destinations' (origins') occupation (inverse) growth rates, where the weight is given by the rail-based proximity index. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

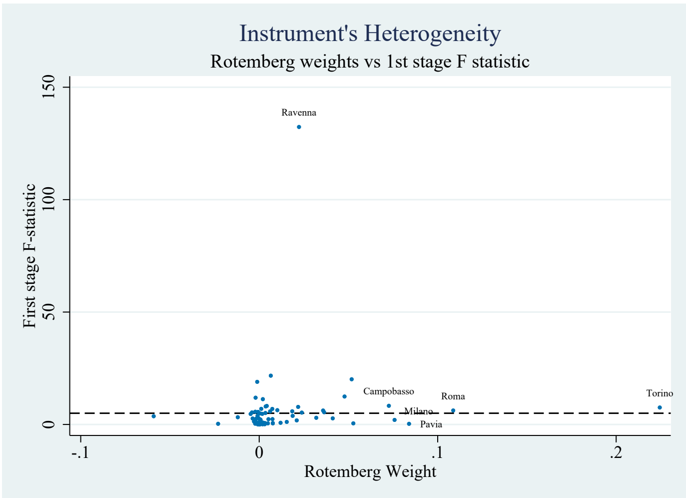
Appendix E Additional Figures

Figure A1: F-statistics vs just-identified IV estimates



Notes: The figure shows the cross-sectional variation between the 1st stage F-statistic and the second-stage coefficient obtained from the just-identified estimates using the migration shares to each individual province as a separate instrument. The outcome variable used for the second-stage regression is GDP per-capita. The horizontal line shows the second-stage coefficient estimated from the shift-share regression (Table 5). Following Goldsmith-Pinkham et al. (2020), the figure excludes just-identified estimates with a 1st stage F-statistic below 5.

Figure A2: F-statistics vs Rotemberg weights



Notes: The figure shows the cross-sectional variation between the 1st stage F-statistic and the Rotemberg weights obtained from the just-identified estimates using the migration shares to each individual province as a separate instrument. The horizontal line shows the second-stage coefficient estimated from the shift-share regression (Table 5). In line with Figure A1, the horizontal line is plotted at level 5.