Accounting for the Duality of the Italian Economy

Jesús Fernández-Villaverde
University of Pennsylvania

Dario Laudati
University of Southern California

Lee Ohanian
University of California, Los Angeles

Vincenzo Quadrini
University of Southern California[*]

May 25, 2023

Abstract

After 162 years of political unification, Italy still displays large regional economic differences. In 2019, the per capita GDP of Lombardia was 39,700 euros, but Calabria's per capita GDP was only 17,300 euros. We build a two-region, two-sector model of the Italian economy to measure the wedges that could account for the differences in aggregate variables between the North and the South. We find that the largest driver of the regional disparity in per capita output is the difference in total factor productivity, followed by fiscal redistribution. These two factors, together, account for more than 70 percent of the output disparity between the North and the South.

Keywords: Italian economy; macroeconomic wedges; regional fiscal redistribution; regional convergence.
JEL codes: E10, E60.

[*]Corresponding author: quadrini@usc.edu.

1 Introduction

Italy's political unification was accomplished in 1861 (except for the rump Papal States around Rome, which survived until 1870). Over 162 years later, large regional economic differences remain within Italy without evidence of convergence. The current income in certain Italian regions is less than half that of other regions. In 2019, the per capita income of the northern region of Lombardia was 39,700 euros, but for the southern region of Calabria it was only 17,300 euros.[1] To put these figures in perspective, the difference between Lombardia and Calabria is similar to the difference between the per capita income of Germany (41,500 euros in 2019) and Greece (17,500 euros in 2019). Thus, the income per capita heterogeneity within Italy is as large as the heterogeneity among some of the key members of the eurozone.

The main goal of this paper is to identify the major drivers of the regional income differences in Italy using the macroeconomic approach based on the measurement of various wedges pioneered by Chari et al. (2007). This approach has gained popularity in the macro literature and identifies distortions (called wedges) affecting factor inputs and productivity. For example, we have learned much about the historical process of China (Cheremukhin et al., 2015) and Russian economic growth (Cheremukhin et al., 2017) through the measurement of these wedges. Even though the wedge analysis does not reveal the ultimate causes of the distortions, it highlights the sectors or segments of the economy where allocations appear especially problematic. We can then focus our attention on these sectors or segments with a deeper analysis that goes beyond the measurements of the wedges.

We start by presenting some stylized facts about regional differences in Italy. We show that regional differences are large, and there is no evidence of regional convergence over the last several decades. The lack of convergence is surprising and cannot be attributed to differences in official institutions, since economic policies in Italy are quite centralized, and the formal regulatory environment is homogeneous across the country.[2] Of course, this might not be the case for informal institutions or the actual implementation of formal institutions, which could be part of the reason why regional differences are so large. In our analysis, the measured wedges capture reduced-form differences in the implementation of policies and informal institutions. But to measure the wedges, we need a structural model.

The model consists of two integrated regions. The first is representative of the Northern and Central regions. The second is representative of the Southern and Island regions. Each region in the model produces two types of goods: tradable and nontradable. The presence of these two

[1] We use 2019, the last year before COVID-19, to avoid contaminating the comparison with this exogenous shock. Recall the North of Italy was the original focus of COVID-19 in Europe.

[2] The Regional Authority Index of Hooghe et al. (2016) ranks Italy's decentralization in the 1950s and 1960s as very low, but with increases in 1970, when the first regional elections were held, and the constitutional reform of 2001. Even after those reforms, Italy is much less decentralized than Germany or Spain and about as decentralized as France. See https://www.arjanschakel.nl/index.php/regional-authority-index for details.

sectors allows us to distinguish regional income differences when income is measured in euros from its actual purchasing power. This could be important for capturing the impact of national transfer policies such as pensions and public wages that, in euro terms, are about the same across regions but with different purchasing power across regions.

Another feature of the model is that tradable and nontradable goods could be produced in the official and unofficial economy. The unofficial economy is also referred to as the irregular or underground economy. The unofficial economy plays an important role in Italy, especially in the South, because it is large. In more recent years, irregular occupations in the North are estimated to be around 11 percent, while in the South, they are about 20 percent.

Using the model and the macroeconomic data constructed by the Italian Statistical Agency (ISTAT), we measure three types of wedges that distort the optimal decisions of households and firms from 2000 to 2020: those that distort the input of labor, those that distort investment, and those that distort total factor productivity. All three types of wedges are higher in the South than in the North, which is not surprising: When looked at through the lens of the model, the fact that the South has lower income must be because the wedges in the South are higher than those in the North. But our goal is to understand *which* of the measured wedges are especially important for generating lower income in the South.

We conduct a battery of counterfactual exercises with our model to answer that question. For instance, we compare the baseline steady-state equilibrium with a counterfactual steady-state equilibrium in which some of the wedges in the South are set equal to their corresponding values in the North. In this way, we quantify the importance of each wedge in generating regional income disparity. We find that differences in labor and investment wedges contribute somewhat to income disparities, but the main contributors are differences in total factor productivity. We also find that inter-regional fiscal transfers contribute significantly to regional income differences. The combined contribution of productivity differences and inter-regional fiscal transfers accounts for more than 70 percent of the income gap between Southern and Northern regions.

The finding that inter-regional fiscal transfers contribute to regional income disparities is the most interesting result of the paper, and the intuition is straightforward. First, inter-regional fiscal transfers are large. In the baseline calibration, the size of the transfers the South receives from the North is 6.56 percent of the total output produced in the South. Conversely, the North pays transfers to the South that are 2.08 percent of the value of total output produced in the North. The elimination of these transfers has a positive income effect on the supply of labor in the South and a negative income effect on the supply of labor in the North. In the counterfactual steady-state equilibrium without fiscal transfers, the output gap between the South and the North is reduced by one-fourth.

Our investigation has important policy implications beyond understanding Italy's economic performance. Like many other countries, Italy has invested large funds in regional development

for decades. Were these monies well spent? After 1975, the European Union (still named the European Economic Community) made regional policies one of its core missions. Nowadays, the European Structural and Investment Funds account for more than one-third of the whole budget of the European Union, with a forecasted expenditure of 392 billion euros in 2021-2027.[3] Will these funds make a difference? Our paper's results cast doubt on these regional policies' efficacy.

In terms of the literature, our paper is related first to the many studies documenting the lack of regional economic convergence in Italy, going back, at the very least, to Clough and Livi (1956) and Eckaus (1961). See, for more recent examples, Iuzzolino et al. (2011, 2013). More in general, there is a vast literature on the convergence of European regions and the effectiveness of regional policies (Boldrin and Canova, 2001; Rosés and Wolf, 2018). Second, our paper is close to an extensive literature that has built regional models of growth and development for the South of Italy, which started with the classic work by Chenery (1962). Finally, our work stresses the importance of heterogeneity in regional productivity, which has recently been studied by Boeri et al. (2021) in connection with wage policies in Italy and Germany.

The rest of the paper proceeds as follows. Section 2 documents four stylized facts about the Italian economy. Section 3 presents the model, which we calibrate in Section 4. Section 5 describes how we measure the wedges. Section 6 conducts the counterfactual exercises and Section 7 concludes. The appendix provides extra details about data and calibration.

2 Four empirical facts

The main goal of this paper is to quantify the sources of regional income disparity in Italy using the well-established methodology of wedge analysis. The wedge analysis will focus on the last two decades, for which richer data are available. Before delving into the technical analysis, however, we would like to provide an overview of some of the most salient features of the Italian economy to set the stage for the technical analysis we will conduct later in the paper. In particular, we want the reader to keep in mind: i) the large regional income disparities in Italy; ii) their persistence over time; iii) the significant size of the informal economy, in particular in the South; and iv) the large fiscal transfers from rich to poor regions. These four facts will motivate our modeling choices in Section 3. Additional statistics are provided in the appendix.

2.1 Regional income disparities and their persistence

Figure 1 presents choropleth maps showing per capita GDP in each of the twenty Italian regions. The left panel plots the 1995 values, while the right panel plots the 2019 values. There are two patterns worth emphasizing. The first is that the levels of per capita GDP in the Southern

[3] See https://ec.europa.eu/regional_policy/funding/available-budget_en.

regions are significantly lower than those in the Northern regions. The wealthier regions in the North (dark areas) have incomes twice as large as those of the poorer regions in the South (lighter areas). The second fact is that there are minimal differences between the left and right panels, both in terms of regional differences and levels of income in the same regions. This indicates that income disparities have not changed much during the last twenty-five years. It also reflects the fact that the Italian economy has not experienced any significant growth over this period. Although the lack of growth is also an important fact about the Italian economy, our focus is to understand the regional income disparities.

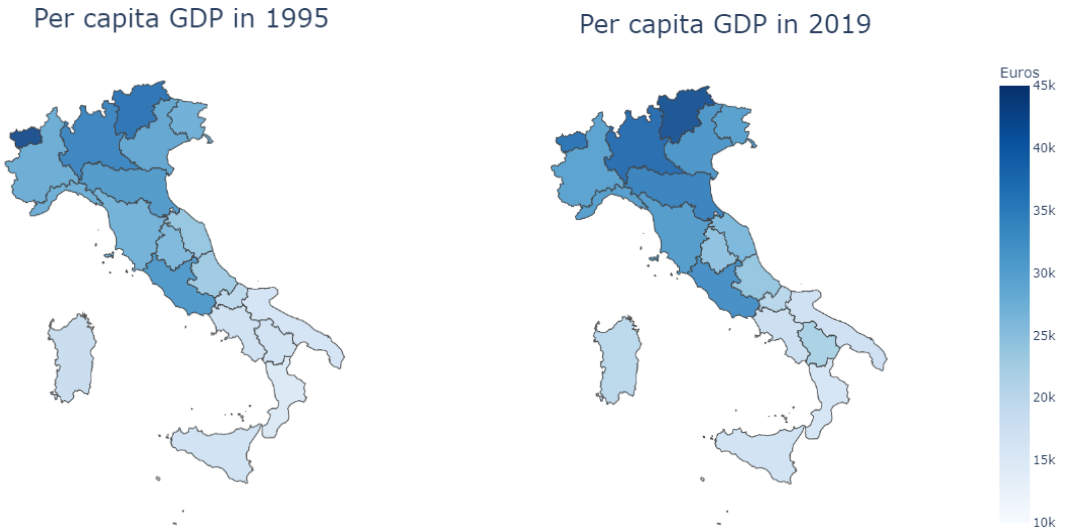

Figure 1: Regional per capita GDP in 1995 and 2019. Real values in 2015 chain prices. *Source*: Istituto Nazionale di Statistica, https://www.istat.it

The lack of regional convergence is not limited to the last twenty-five years. Looking at a longer historical perspective, Figure 2 plots the per capita GDP ratio between Southern and Island regions (the "South" for short) and Center and Northern regions (the "North") after the unification in 1861. The first eighty years of unification witnessed regional divergence: the North grew faster as it completed a structural transformation and industrialization, in particular around the "industrial triangle" of Milan, Turin, and Genoa.[4] The process of regional divergence was particularly acute from the start of World War I to the end of World War II. The South remained largely rural, and economic policies during the Fascist period were unfavorable to this sector.

The decades after World War II saw a brief period of convergence, mainly in the 1950s and 1960s. The South experienced productivity increases, with large declines in the population engaged in agriculture and also a sizable outward migration, mainly toward the North and the

[4]The bibliography of Italian economic history in English is somewhat limited, but the interested reader can get more details in Zamagni (1993) and Toniolo (2013). A good reference, although in Italian, is Daniele and Malanima (2011).

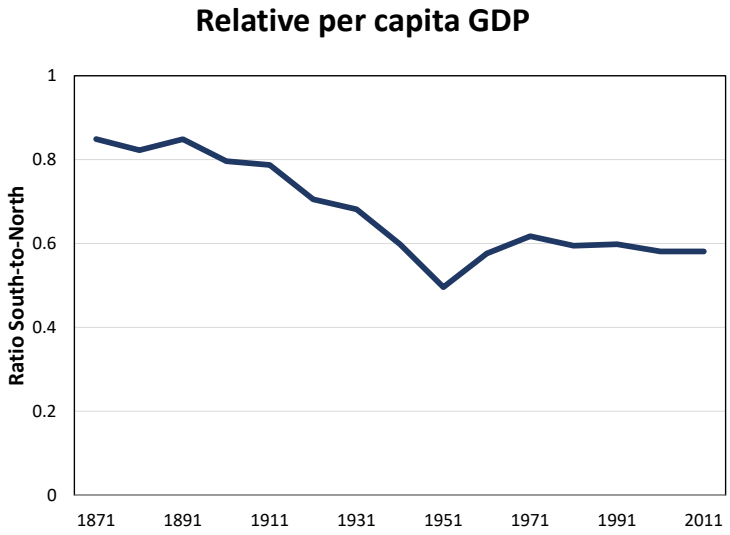

Figure 2: Regional relative per capita GDP from 1871 to 2011. Ratio of per capita GDP in the South over per capita GDP in the North. **Northern regions**: Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombadia, Marche, Piemonte, Trentino-Alto Adige, Toscana, Umbria, Val d'Aosta, Veneto. **Southern regions**: Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, and Sicilia. *Source*: Felice (2019), online supplementary data on journal web site.

rest of Europe, which reduced the income gap from around 50 percent to 40 percent. Also, the central government funded significant investments in the South starting in 1950 with the *Cassa per il Mezzogiorno*. But the slowdown in economic growth at the start of the 1970s coincided with the end of the regional convergence. If anything, the South has experienced some divergence compared to the North. Nowadays, the per capita income ratio between the South and the North is at the same level as in the early 1930s. This suggests that the inefficiencies and distortions that afflicted the South in the early years of the Italian unification are still present today.

Figure 3 provides another illustration of the lack of convergence. It plots the relative position of each of the twenty regions in the distribution of per capita income in two different years: 1871 and 2011. Each point represents a region and indicates the 1871 income relative to the national average (horizontal axis) and the 2011 income, also relative to the national average (vertical axis). If the relative distance from the mean of all regions had not changed between 1871 and 2011, all points would be perfectly aligned along the 45-degree line. Clearly, there have been some movements. In particular, we can see that Aosta Valley and Trentino-Alto Adige significantly improved their relative income. On the other hand, Campania –home to Naples, the largest city in the South– has become relatively poorer. However, most regions are around the 45-degree line. More importantly, the dispersion of regional income in 2011 (the horizontal axis range) is the same as in 1871 (the vertical axis). This indicates that there has been no convergence over

this long period.[5]

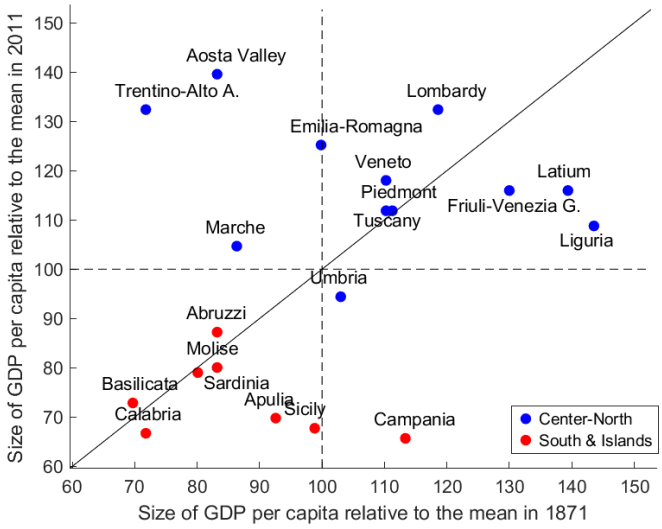

Figure 3: Ratio of regional GDP per capita relative to the national average in 1871 and 2011
Source: Felice (2019), online supplementary data on journal web site.

Italy's lack of regional convergence appears remarkable compared to other similar European countries. For example, in 1871, Spain was around 10 percent poorer than Italy in per capita income, and its regional inequality in income per capita was 4 percent higher than Italy's. In 2005, Spain was still around 10 percent poorer than Italy in per capita income, but its regional inequality was now 49 percent lower than Italy's (Iuzzolino et al., 2013, Table 20.1).

Differences in per capita income could derive from many sources. Here we find it useful to conduct a simple decomposition based on the following accounting identity:

$$\text{Per capita GDP} = \left(\frac{\text{Total GDP}}{\text{Population employed}}\right) \times \left(\frac{\text{Population employed}}{\text{Population 15-64}}\right).$$

The first term on the right-hand side is labor productivity (GDP per employed person), and the second is labor participation, the fraction of employed people. Both terms could be important for generating lower income in the South.

The top panels of Figure 4 are choropleth maps of labor productivity for 1995 and 2019. Comparing 1995 (left panel) to 2019 (right panel), we observe only minor differences for most regions. Looking at a single year, we observe that the productivity differences across regions are quite large and highly correlated with differences in per capita GDP we showed earlier. This indicates that differences in labor productivity could be an important factor for understanding cross-regional income differences. However, this is only part of the story.

[5]Interestingly, Iuzzolino et al. (2011) document a convergence between Western and Eastern Italian regions after a period of growing divergence, but not between Northern and Southern regions.

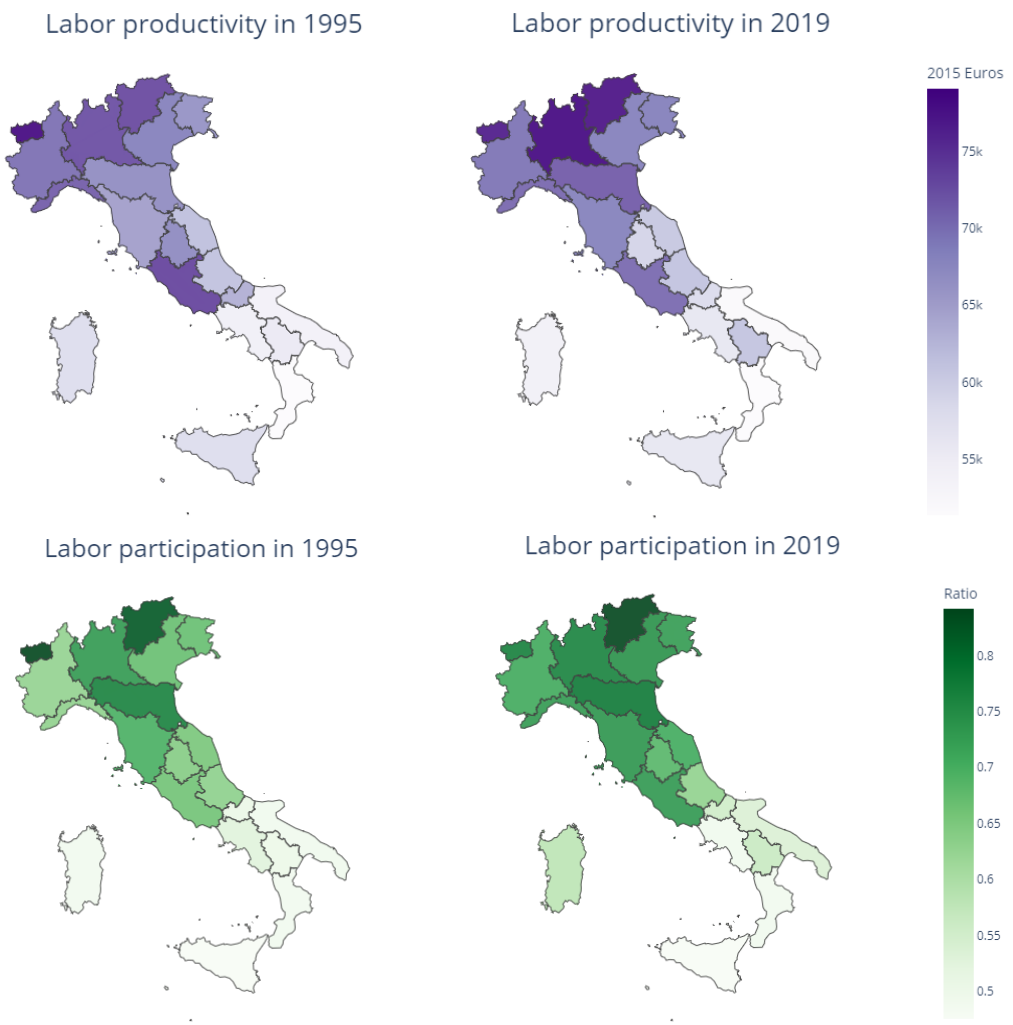

Figure 4: GDP per employed person and fraction of the working age population employed in 1995 and 2019. GDP is in 2015 chain prices. *Source*: Istituto Nazionale di Statistica, https://www.istat.it

The bottom panels of Figure 4 are choropleth maps for regional labor participation, that is, the fraction of the working-age population employed. In this case, we also notice that the differences between 1995 and 2019 are sizable only for a very few regions. At the same time, we observe significant differences in labor participation across regions, which are highly correlated with regional per capita GDP. Thus, labor participation is also important for understanding Italy's regional income disparities. The goal of the wedge analysis we will conduct later in the paper is to understand the contribution of the various wedges to generating differences in labor productivity and participation.

2.2 The informal economy

A characteristic of the Italian economy is the exceptionally large size of the informal sector, at least compared to other countries at a similar stage of economic development (proxied by income

per capita). The Italian statistical agency provides estimates of the informal economy, also called the "unobserved" economy, from 2011 through 2020. The estimates are shown in the first panel of Figure 5. Over this period, the informal sector contributed to Italy's GDP by an average of 12 percent.

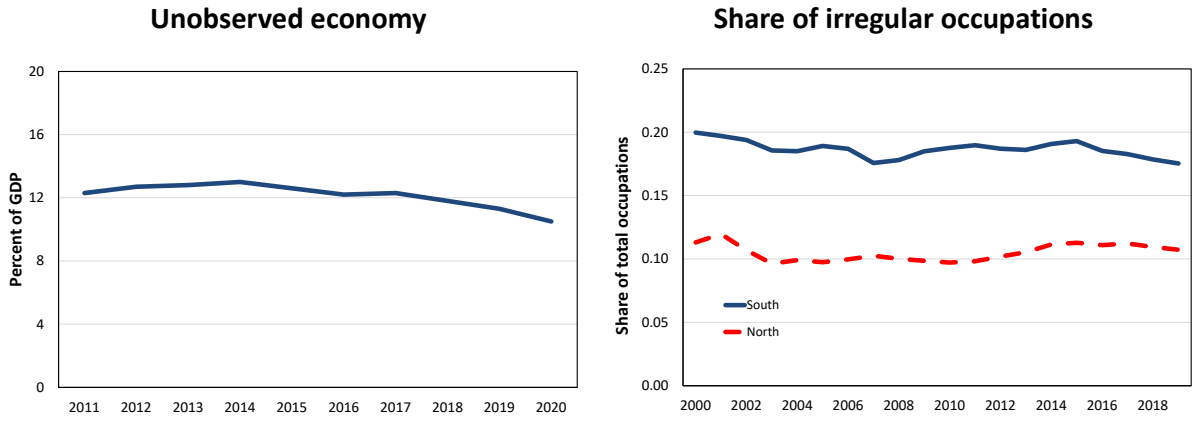

Figure 5: Size of the informal economy in Italy and shares of irregular occupations in Northern and Southern regions. **Northern regions**: Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombadia, Marche, Piemonte, Trentino-Alto Adige, Toscana, Umbria, Val d'Aosta, Veneto. **Southern regions**: Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, and Sicilia. *Source*: Istituto Nazionale di Statistica, https://www.istat.it

The official statistics, unfortunately, do not provide estimates of the size of GDP generated by the informal economy separately for each region. However, it provides disaggregated regional measures for "regular" and "irregular" occupations. Broadly speaking, regular occupations represent the labor force employed in the formal economy, and irregular occupations the labor force employed in the informal sector.

The second panel of Figure 5 plots the share of irregular occupations over the total in the North and South from 2000 to 2019. While the share of irregular occupations is sizable in both regions, it is almost twice as large in the South (two out of ten workers vs. one out of ten workers in the North). Perhaps, the larger informal sector in the South could be important for explaining regional income disparities. We explore this hypothesis with the structural wedge analysis in the next section.

2.3 Inter-regional transfers

Public finances are highly centralized in Italy. Generally speaking, most taxes are paid to the central government, which then reallocates the funds across the country. Because of the regional income disparity just documented, it is no surprise that the system is characterized by sizable inter-regional fiscal redistribution.

Figure 6 separates the regions with a fiscal surplus in 2019 (darker areas) and a fiscal deficit (lighter areas). All Southern regions have a fiscal deficit, while most Northern regions have a fiscal surplus. The only exceptions are the central region of Umbria and the Northern region of Friuli-Venezia Giulia. Liguria, another Northern region, also has a deficit but it is close to zero. It is also worth emphasizing that Lombardia, the Northern region with the highest per capita income level, is also the region with the highest fiscal surplus: close to 6,000 euros per person living in the region. Considering that Lombardia is also the most populous region, its contribution to the government budget is quite significant.

Figure 6: Per capita fiscal balance in 2019 for each Italian region: Surplus (darker areas) and deficit (lighter areas). Source: *Conti Publici Territoriali* (CPT), www.agenziacoesione.gov.it/sistema-conti-pubblici-territoriali. CPT provides consolidated expenses and revenues in current and capital accounts for the whole public sector in Italy.

Suppose we aggregate the Northern regions on the one hand and the Southern regions on the other. In that case, we find that in 2019 the per capita fiscal surplus of the North was about 3,000 euros, and the fiscal deficit of the South was about 2,000 euros. These are large numbers: for the North, the fiscal surplus is about 8.5 percent of the value of its per capita GDP, while for the South, the fiscal deficit is about 11 percent of its per capita GDP. These large inter-regional transfers could have significant economic effects that we will explore with the structural analysis conducted in the remaining sections of the paper.

3 Model

We postulate a model that formalizes the four stylized facts outlined in the previous section. In particular, the model has the following features:

- Two economic areas: Southern Italy and Northern Italy.
- Two sectors of production: Tradable and nontradable.
- Two segments of the economy: Official and unofficial.
- Fiscal transfers that redistribute resources between the North and the South.

The motivation for having a two-region model is evident since the paper's main purpose is to understand income differences and their persistence between the North and the South. The presence of two distinct sectors, tradable and nontradable, allows us to have different final goods prices in the two regions. Differences in final goods prices imply differences in the cost of living, affecting the real value of public transfers to Southern and Northern regions. The distinction between the official and unofficial economies is justified by the magnitude of the underground economy and the sizable differences between the two regions documented in the previous section. Finally, the presence of a government that redistributes resources geographically allows us to investigate how inter-regional transfers affect income disparity between the North and the South.

3.1 Model details

There are two economic areas indexed by $j \in \{N, S\}$, where N denotes Northern Italy and S denotes Southern Italy. Region $j \in \{N, S\}$ is populated by a continuum μ_j of homogeneous households. Each region has two sectors of production: a tradable sector and a nontradable sector. Both sectors produce intermediate goods in two segments of the economy: "official" and "unofficial." The unofficial or underground economy can avoid some taxes and certain types of regulatory restrictions, such as labor regulations. However, the unofficial economy could also face limitations (such as limited access to financial markets) that could adversely affect productivity and the effective cost of productive inputs. The impact on productivity and the effective cost of inputs will be captured by wedges specific to the official and unofficial segments of the economy.

Households have one unit of time that can be supplied to the official segments, $h_{j,t} \geq 0$, or to the unofficial segments, $\tilde{h}_{j,t} \geq 0$. Households cannot change residency, and the population μ_j is constant.

The lifetime utility of a representative household in region $j \in \{N, S\}$ is:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U\Big(c_{j,t} + G_{j,t}, h_{j,t}, \tilde{h}_{j,t}\Big),$$

where β is the discount factor, $c_{j,t}$ is private consumption, $G_{j,t}$ is government purchases, $h_{j,t}$ is hours allocated to the official economy, and $\tilde{h}_{j,t}$ is hours allocated to the unofficial economy. We assume that government purchases (public consumption) enter the households' utility additively to private consumption. Otherwise, changes in government purchases would have large income effects on the labor supply. More specifically, if $G_{j,t}$ increases in region j funded locally by region j with higher taxes, this must be compensated by lower consumption and/or investment. But lower consumption increases the labor supply through the typical income effect and leads to more production. Instead, if $G_{j,t}$ is additive to $c_{j,t}$, what matters for labor supply is the sum of private and public consumption. Then, an increase in $G_{j,t}$ compensated by a decrease in $c_{j,t}$ does not affect the labor supply.

The period utility takes the standard form:

$$U\Big(c_{j,t} + G_{j,t}, h_{j,t}, \tilde{h}_{j,t}\Big) = \ln(c_{j,t} + G_{j,t}) + \alpha \ln(1 - h_{j,t} - \tilde{h}_{j,t}).$$

Final goods are produced by competitive firms that combine tradable and nontradable intermediate inputs using the production function:

$$y_{j,t} = F_{j,t}(m_{T,j,t}, m_{NT,j,t}) \equiv A_{j,t} m_{T,j,t}^{\eta} m_{NT,j,t}^{1-\eta}.$$

The variable $m_{T,j,t}$ denotes the input of tradable goods in region $j \in \{N, S\}$ at time t, while $m_{NT,j,t}$ denotes the input of nontradable goods in region $j \in \{N, S\}$ at time t. Productivity in the final sector, A_j, could differ between the two regions.

Both tradable and nontradable intermediates are produced by competitive firms operating in official and unofficial segments of the economy. Intermediate inputs are produced with the use of capital and labor according to the technologies:

$$x_{i,j,t} = z_{i,j,t} k_{i,j,t}^{\theta_i} l_{i,j,t}^{1-\theta_i}, \tag{1}$$

$$\tilde{x}_{i,j,t} = (1 - \tau_{i,j,t}^{x}) z_{i,j,t} \tilde{k}_{i,j,t}^{\theta_i} \tilde{l}_{i,j,t}^{1-\tilde{\theta}_i}. \tag{2}$$

The subscript i denotes the production type, tradable ($i = T$) or nontradable ($i = NT$), and j identifies the region, North ($j = N$) or South ($j = S$). The variable $x_{i,j,t}$ is the production of intermediate good $i \in \{T, NT\}$ in the official sector of region $j \in \{N, S\}$, and $\tilde{x}_{i,j,t}$ is the production of intermediate good $i \in \{T, NT\}$ in the unofficial sector of region $j \in \{N, S\}$. Thus, variables related to intermediate production without a tilde are for the official economy, and variables with a tilde are for the unofficial economy.

The variable $z_{i,j,t}$ is the productivity in the official segment of the sector. The productivity

in the unofficial segment is distorted by the wedge $\tau^x_{i,j,t}$. We assume the wedge has the form:

$$\tau^x_{i,j,t} = \kappa_{i,j,t} + \nu \cdot \left(\frac{\tilde{X}_{i,j,t}}{X_{i,j,t} + \tilde{X}_{i,j,t}} \right),$$

where capital letters denote aggregate variables. The wedge is the sum of two terms. The first term, $\kappa_{i,j,t}$, is exogenous. The second term, $\nu(\tilde{X}_{i,j,t}/(X_{i,j,t} + \tilde{X}_{i,j,t}))$, is endogenous because it is determined in equilibrium by the size of unofficial production relative to the overall production in the sector. Furthermore, the endogenous component increases with the aggregate share of unofficial production, not individual production. This implies that an individual firm takes the productivity wedge between official and unofficial productions as given.

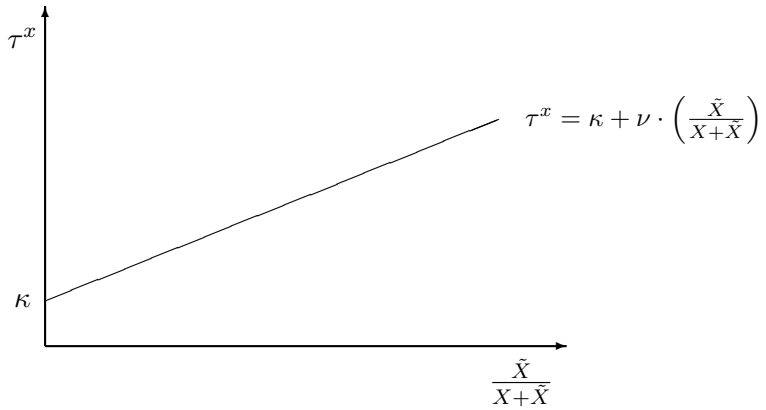

Figure 7: Productivity wedge in the unofficial segment of the sector relative to the official segment.

The dependence of the wedge from the share of unofficial production in a particular sector i is shown in Figure 7. When total production arises in the official segment of the sector, the wedge is κ. After that, the wedge increases linearly with the share of unofficial production, with the slope governed by the parameter ν.

The dependence of unofficial productivity on its share is necessary to ensure that production occurs in both segments of the sector. Without this dependence, the assumption of constant returns to scale in production would imply that production occurs either in the official segment of the sector or in the unofficial segment, but not in both.[6]

The factor share θ_i is the same for the official and unofficial segments of sector i and in both regions. This can be justified by the fact that input shares are related to technological

[6] An alternative approach would be to assume that the two segments of the sector produce different and imperfectly substitutable goods. In this case, we would not need to have "endogenous" productivity differences between the two segments of the sector: the role of endogenous relative productivity would be played by the endogenous relative prices of the different goods produced by the two segments. The substitutability between the goods produced in the official and unofficial segments of the sector will play a role similar to that of the parameter ν.

knowledge, and there is no reason why it should differ between official and unofficial production or between North and South. If there are differences in input factors or productivity, they are likely to derive from distortions captured by wedges that differ between official and unofficial productions and between Northern and Southern regions (as we will describe shortly).

Using capital letters to denote per capita aggregate variables, the equilibrium in the labor market requires:

$$\begin{aligned} L_{T,j,t} + L_{NT,j,t} &= H_{j,t}, \\ \tilde{L}_{T,j,t} + \tilde{L}_{NT,j,t} &= \tilde{H}_{j,t}. \end{aligned}$$

The first equation states that, in each region, the demand for labor in the official economy (both tradable and nontradable sectors) must be equal to the official supply of labor from households. The second equation is for the nonofficial segment of the economy.

For capital, we assume that there is full mobility between the two sectors. This is a strong assumption but allows us to simplify the analysis considerably. The equilibrium in the market for capital satisfies:

$$K_{T,j,t} + K_{NT,j,t} + \tilde{K}_{T,j,t} + \tilde{K}_{NT,j,t} = \overline{K}_{j,t}.$$

The market clearing conditions for intermediate inputs are:

$$\begin{aligned} X_{NT,N,t} + \tilde{X}_{NT,N,t} &= M_{NT,N,t}, \\ X_{NT,S,t} + \tilde{X}_{NT,S,t} &= M_{NT,S,t}, \\ (X_{T,N,t} + \tilde{X}_{T,N,t})\mu_N + (X_{T,S,t} + \tilde{X}_{T,S,t})\mu_S &= M_{T,N,t}\mu_N + M_{T,S,t}\mu_S. \end{aligned}$$

While nontradable intermediate goods clear at the regional level, tradable intermediate goods clear at the national level. Since the variables are in per capita terms, when we aggregate at the national level, we multiply them by the population size of each region, that is, μ_j.

Households can hold bonds issued by the other region. The bonds purchased at time t by a household in region j are denoted by $b_{j,t+1}$. Given the bond price Q_t, the household pays $Q_t b_{j,t+1}$ at time t and will get repaid $(1 - \tau^b_{t+1})b_{j,t+1}$ at time $t+1$. Bonds $b_{j,t}$ are expressed in units of tradable intermediate goods, and negative values represent borrowing.

The variable τ^b_{t+1} is a wedge that limits inter-regional lending or borrowing. We assume that this wedge is endogenous and depends on the region's aggregate bond holdings according to:

$$\tau^b_{j,t} = \zeta_{j,t} + \chi \cdot B_{j,t}. \tag{3}$$

As before, this wedge is the sum of two components: the first, $\zeta_{j,t}$, is exogenous, while the

second, $\chi \cdot B_{j,t}$, is endogenous. The endogeneity of the wedge guarantees that the accumulation of bonds from the other region is bounded as their return decreases in $B_{j,t}$. Similarly, extra-regional borrowing is bounded since more debt (negative values of $B_{j,t}$) increases its cost. This guarantees that the net external asset positions of the two regions are stationary, which is a common assumption in open economy models with multiple regions or countries.

The budget constraint for the representative household, expressed in units of the final good produced in the corresponding region, is:

$$c_{j,t} + i_{j,t} + Q_t b_{j,t+1} P_{T,j,t} \quad = \quad W_{j,t} h_{j,t} + \tilde{W}_{jt} \tilde{h}_{jt} + R_{jt} k_{jt} + (1 - \tau^b_{j,t}) b_{j,t} P_{T,j,t} + T_{j,t} P_{T,j,t},$$

where $i_{j,t} = k_{j,t+1} - k_{j,t}$ is net investment.

Capital can be reallocated freely from one sector to another (tradable and nontradable) and between the official and unofficial segments of each sector. Because of this, investment is not sector-specific. However, capital cannot be reallocated from one region to the other. Still, cross-regional borrowing and lending allow for capital mobility. Since bonds $b_{j,t}$ are expressed in units of tradable goods, while the budget constraint is expressed in units of final goods, we multiplied $b_{j,t}$ by the price of tradables in units of region j's final goods, $P_{T,j,t}$. Thus, $b_{j,t} P_{T,j,t}$ represents the value of bonds in units of region j's final goods.

The variable $T_{j,t}$ denotes transfers net of taxes received from the government. They are expressed in units of tradable goods. If we multiply the transfers by the price of tradable goods, $P_{T,j,t}$, we obtain its value in units of region j's final goods. This highlights an important difference between the Northern and Southern regions. Transfers to the two regions could be the same, that is, $T_{S,t} = T_{N,t}$. However, since the price of tradable goods in units of local final goods differs in the two regions, that is, $P_{T,S,t} \neq P_{T,N,t}$, the real value of transfers (purchasing power) also differs. The bond market clears at the national level, that is, $B_{N,t+1}\mu_N + B_{S,t+1}\mu_S = 0$.

The public sector is centralized at the national level but makes purchases and transfers at the regional level. Denote by $G_{j,t}$ the per capita purchases of final goods and services in region j, and $T_{j,t}$ the per capita transfers net of taxes also in region j. While the purchases of goods and services are in units of local final goods, the transfers are in units of tradable goods. Thus, the national government budget is:

$$\left(\frac{G_{N,t}}{P_{T,N,t}} + T_{N,t} \right) \mu_N + \left(\frac{G_{S,t}}{P_{T,S,t}} + T_{S,t} \right) \mu_S = 0.$$

Since the government budget is defined in units of tradable goods, we converted government purchases $G_{j,t}$ to tradable goods using the price of tradables $P_{T,j,t}$.

3.2 Optimization and first-order conditions

Firms producing intermediate goods $i \in \{T, NT\}$ solve the profit maximization problem,

$$\max_{\substack{l_{i,j,t},k_{i,j,t} \\ \tilde{l}_{i,j,t},\tilde{k}_{i,j,t}}} \Big\{(x_{i,j,t}+\tilde{x}_{i,j,t})P_{i,j,t} - (1+\tau^l_{i,j,t})W_{i,j,t}l_{i,j,t} - (1+\tau^k_{i,j,t})(R_{i,j,t}+\delta)k_{i,j,t} \\ -(1+\tilde{\tau}^l_{i,j,t})\tilde{W}_{i,j,t}\tilde{l}_{i,j,t} - (1+\tilde{\tau}^k_{i,j,t})(\tilde{R}_{i,j,t}+\delta)\tilde{k}_{i,j,t}\Big\},$$

subject to the production functions (1) and (2).

The variables $\tau^l_{i,j,t}$, $\tilde{\tau}^l_{i,j,t}$, $\tau^k_{i,j,t}$, $\tilde{\tau}^k_{i,j,t}$ are the wedges that affect firms' decisions. The wedges can be interpreted as including formal taxes as well as other costs. For example, it could reflect red tape, ransoms collected by criminal organizations, or bribes to local administrators. Generally, the part captured by formal taxes is similar between the two regions since the regulatory and fiscal systems are centralized in Italy (although there are some geographical exemptions). However, the nonlegal component of the wedges could be very different across regions. Since we expressed all quantities in units of final goods produced in the corresponding region j, the quantities of intermediate goods are multiplied by the price $P_{i,j,t}$ expressed in units of final goods. For instance, $x_{T,N,t}P_{T,N,t}$ is the quantity of final goods that can be purchased in the North with $x_{T,N,t}$ units of tradable goods.

The optimality conditions for firms producing intermediate inputs are:

$$(1-\theta_i)z_{i,j,t}K^{\theta_i}_{i,j,t}L^{-\theta_i}_{i,j,t}P_{i,j,t} = (1+\tau^l_{i,j,t})W_{i,j,t} \quad (4)$$

$$(1-\theta_i)(1-\tau^x_{i,j,t})z_{i,j,t}\tilde{K}^{\theta_i}_{i,j,t}\tilde{L}^{-\theta}_{i,j,t}P_{i,j,t} = (1+\tilde{\tau}^l_{i,j,t})\widetilde{W}_{i,j,t} \quad (5)$$

$$\theta_i z_{i,j,t}K^{\theta_i-1}_{i,j,t}L^{1-\theta_i}_{i,j,t}P_{i,j,t} = (1+\tau^k_{i,j,t})(R_{i,j,t}+\delta) \quad (6)$$

$$\theta_i(1-\tau^x_{i,j,t})z_{i,j,t}\tilde{K}^{\theta_i-1}_{i,j,t}\tilde{L}^{1-\theta_i}_{i,j,t}P_{i,j,t} = (1+\tilde{\tau}^k_{i,j,t})(\tilde{R}_{i,j,t}+\delta). \quad (7)$$

The profit maximization problem solved by final good producers is:

$$\max_{m_{T,j,t},m_{NT,j,t}} \Big\{F_{j,t}(m_{T,j,t},m_{NT,j,t}) - m_{T,j,t}P_{T,j,t} - m_{NT,j,t}P_{NT,j,t}\Big\}.$$

Because there is free trade, the price of tradable intermediate goods will be equalized in the two regions when expressed in the same units. However, since we expressed the tradable price $P_{T,j,t}$ in units of final goods produced in region j, the price $P_{T,N,t}$ is not equal to $P_{T,S,t}$.

The optimality conditions for final good firms are

$$\frac{\partial F_{j,t}(M_{T,j,t},M_{NT,j,t})}{\partial M_{i,j,t}} = P_{i,j,t}, \quad (8)$$

where $i \in \{T, NT\}$.

The optimality conditions for households are:

$$\frac{W_{i,j,t}}{c_{j,t}+G_{j,t}} = \frac{\alpha}{1-h_{j,t}-\tilde{h}_{j,t}}, \tag{9}$$

$$\frac{\widetilde{W}_{i,j,t}}{c_{j,t}+G_{j,t}} = \frac{\alpha}{1-h_{j,t}-\tilde{h}_{j,t}}, \tag{10}$$

$$\frac{1}{c_{j,t}+G_{j,t}} = \beta\mathbb{E}_t\frac{1+R_{i,j,t+1}}{c_{j,t+1}+G_{j,t+1}} \tag{11}$$

$$\frac{1}{c_{j,t}+G_{j,t}} = \beta\mathbb{E}_t\frac{1+\widetilde{R}_{i,j,t+1}}{c_{j,t+1}+G_{j,t+1}} \tag{12}$$

$$\frac{Q_tP_{T,j,t}}{c_{j,t}+G_{j,t}} = \beta\mathbb{E}_t\frac{(1-\tau^b_{j,t+1})P_{T,j,t+1}}{c_{j,t+1}+G_{j,t+1}}. \tag{13}$$

The first two equations imply that the wage paid in the official and unofficial segments of the economy will be equal in equilibrium. The third and fourth equations imply that, without uncertainty, the return from capital in the official and unofficial segments of the economy will be equalized. However, the fact that wages and returns from capital are equalized does not mean that firms face the same costs in official and unofficial productions since they face the same wedges.

3.3 Equilibrium conditions

We now impose equilibrium conditions for which individual variables (expressed in small letters) are equal to per capita aggregate variables (expressed in capital letters). Combining the first-order conditions of intermediate firms with the first-order conditions of households, we obtain:

$$\frac{(1-\theta_i)X_{i,j,t}P_{i,j,t}}{C_{j,t}+G_{j,t}} = (1+\tau^l_{i,j,t})\left(\frac{\alpha L_{i,j,t}}{1-H_{j,t}-\tilde{H}_{j,t}}\right), \tag{14}$$

$$\frac{(1-\theta_i)\widetilde{X}_{i,j,t}P_{i,j,t}}{C_{j,t}+G_{j,t}} = (1+\tilde{\tau}^l_{i,j,t})\left(\frac{\alpha \tilde{L}_{i,j,t}}{1-H_{j,t}-\tilde{H}_{j,t}}\right), \tag{15}$$

$$\frac{1}{C_{j,t}+G_{j,t}} = \beta\mathbb{E}_t\left[\frac{1+\left(\frac{1}{1+\tau^k_{j,t+1}}\right)\theta_i\frac{X_{i,j,t+1}P_{i,j,t+1}}{K_{i,j,t+1}}-\delta}{C_{j,t+1}+G_{j,t+1}}\right] \tag{16}$$

$$\frac{1}{C_{j,t}+G_{j,t}} = \beta\mathbb{E}_t\left[\frac{1+\left(\frac{1}{1+\tilde{\tau}^k_{j,t+1}}\right)\theta_i\frac{\widetilde{X}_{i,j,t+1}P_{i,j,t+1}}{\widetilde{K}_{i,j,t+1}}-\delta}{C_{j,t+1}+G_{j,t+1}}\right] \tag{17}$$

$$\frac{Q_tP_{T,j,t}}{C_{j,t}+G_{j,t}} = \beta\mathbb{E}_t\left[\frac{(1-\tau^b_{j,t+1})P_{T,j,t+1}}{C_{j,t+1}+G_{j,t+1}}\right]. \tag{18}$$

With perfect foresight, we can eliminate the expectation operator and rewrite them as:

$$\frac{(1-\theta_i)X_{i,j,t}P_{i,j,t}}{C_{j,t}+G_{j,t}} = (1+\tau^l_{i,j,t})\left(\frac{\alpha L_{i,j,t}}{1-H_{j,t}-\tilde{H}_{j,t}}\right), \tag{19}$$

$$\frac{(1-\theta_i)\widetilde{X}_{i,j,t}P_{i,j,t}}{C_{j,t}+G_{j,t}} = (1+\tilde{\tau}^l_{i,j,t})\left(\frac{\alpha \tilde{L}_{i,j,t}}{1-H_{j,t}-\tilde{H}_{j,t}}\right), \tag{20}$$

$$\frac{C_{j,t+1}+G_{j,t+1}}{C_{j,t}+G_{j,t}} = \beta\left[1+\left(\frac{1}{1+\tau^k_{j,t+1}}\right)\theta_i\frac{X_{i,j,t+1}P_{i,j,t+1}}{K_{i,j,t+1}}-\delta\right] \tag{21}$$

$$\frac{C_{j,t+1}+G_{j,t+1}}{C_{j,t}+G_{j,t}} = \beta\left[1+\left(\frac{1}{1+\tilde{\tau}^k_{j,t+1}}\right)\tilde{\theta}_i\frac{\tilde{X}_{i,j,t+1}P_{i,j,t+1}}{\tilde{K}_{i,j,t+1}}-\delta\right] \tag{22}$$

$$\frac{C_{j,t+1}+G_{j,t+1}}{C_{j,t}+G_{j,t}} = \beta\left(\frac{P_{T,j,t+1}}{Q_tP_{T,j,t}}\right)(1-\tau^b_{j,t+1}). \tag{23}$$

The last condition implies that in the steady state, the wedge $\tau^b_{j,t+1}$ must be the same in the two regions. Since the wedge is endogenous, this condition pins down the steady-state value of the net foreign asset position for the two regions.

We also have the first-order condition for final good firms:

$$\frac{\partial F_{j,t}(M_{T,j,t},M_{NT,j,t})}{\partial M_{i,j,t}} = P_{i,j,t}, \tag{24}$$

the market clearing equilibrium in the final goods market,

$$C_{j,t}+G_{j,t}+K_{j,t+1}-(1-\delta)K_{j,t} = M_{T,j,t}P_{T,j,t}+M_{NT,j,t}P_{NT,j,t} = Y_{j,t} \tag{25}$$

and the market clearing condition in the bond market,

$$\mu_N B_{N,t+1}+\mu_S B_{S,t+1} = 0. \tag{26}$$

The wedge $\tau^b_{j,t}$ does not appear in the market clearing condition for final goods because we assume that it is not a deadweight loss but is redistributed to households as transfers.

The trade balance in region j, expressed in units of tradable goods, is:

$$TB_{j,t} = \mu_j X_{T,j,t}+\mu_j\widetilde{X}_{T,j,t}-\mu_j M_{T,j,t}.$$

The nontrade component of the current account (net factor income) in region j, also in units of tradable goods, is:

$$NFI_{j,t} = (1-Q_{t-1})B_{j,t}+\frac{G_{j,t}}{P_{T,j,t}}+T_{j,t}.$$

Notice that $T_{j,t}$ represents the transfers received by residents of region j minus the taxes they pay. Thus, $G_{j,t}+T_{j,t}$ represents the fiscal balance of region j. The net factor incomes received

by the region are given by the interest residents earned on bonds (capital incomes) plus the fiscal transfers received from the other region (unilateral transfers).

The sum of the trade balance and factor incomes gives the current account. This is also equal to the change in the external net asset position:

$$\Delta NFA_t \;=\; TB_{j,t} + NFI_{j,t} \;=\; Q_t B_{j,t+1} - Q_{t-1} B_{j,t}.$$

4 Calibration and productivity series

The North in the model represents the north and center of Italy. It includes Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombardia, Marche, Piemonte, Trentino-Alto Adige, Toscana, Umbria, Val d'Aosta, and Veneto. The South represents the south and island of Italy. It includes Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, and Sicilia. Figure 8 visualizes the two aggregated regions.

Figure 8: Northern and Central Regions (dark areas), and Southern and Island Regions (light areas). The classification refers to the official classification by the Italian statistical agency (ISTAT).

The regional aggregation that forms the North and the South follows the Italian statistical agency, ISTAT, which reports aggregate data for these two geographical areas. The first area is named "Centro-nord," while the second is named "Mezzogiorno." However, we will refer to them as North and South for simplicity. ISTAT also provides sectoral data from which we can construct the series for three major sectors: agriculture, manufacturing, and services. We

assume that the first two (agriculture and manufacturing) produce tradable goods, while the third (services) produces nontradable goods.

We calibrate the model annually and set the discount factor to $\beta = 0.96$, a value often used in macroeconomics. The utility parameter α is usually chosen to target the average working time observed in the data. This would be a suitable procedure if we knew the labor wedges $\tau^l_{i,j,t}$ and $\tilde{\tau}^l_{i,j,t}$. Unfortunately, we do not know the wedges, and to determine the wedges, we need to know α. Because of the indeterminacy, we pre-set $\alpha = 1.5$. This is not problematic because different values of α simply re-scale the labor wedges but do not change the relative values between regions, which is the focus of our paper.

For the calibration of the depreciation rate δ, we use data on consumption of fixed capital. This also requires the construction of empirical series for the stock of capital. We construct the capital series for each sector and each region using the perpetual inventory method. The detailed description is provided in Appendix A.1. Once we have the sectoral and regional measures of capital, we need to allocate them to the official and unofficial segments of each sector (tradable and nontradable sectors). This requires some imputation, described in Appendix B.

The calibration of the factor share parameters in the production of intermediate inputs, θ_T and θ_{NT}, requires measures of income shares (capital and labor) in each sector $i \in \{T, NT\}$. Unfortunately, we do not have data that allow us to construct measures of income shares for each sector (tradable and nontradable). Because of this, we impose the condition that $\theta_T = \theta_{NT} = \theta$. Determining the income share for the whole country requires several steps, which we describe in Appendix A.2.

Given the value of θ and the constructed capital series, we construct the productivity series in the official intermediate good sectors as Solow residuals:

$$z_{i,j,t} = \frac{X_{i,j,t}}{K^{\theta}_{i,j,t} L^{1-\theta}_{i,j,t}}.$$

When we compare the productivity of the South with the productivity of the North, the difference also captures differences in prices. This is especially important for the nontradable sector, where prices could be quite different between the two regions. We cannot separate the component from differences in prices and actual productivity because we only have price indices normalized to 1 in 2015 for all regions. This allows us to compute price changes over time in each region rather than compare their levels across regions. This point is important for our counterfactual exercises later in the paper.

Next, we calibrate the input share parameter in the production of final goods η. In equilibrium, we always have:

$$\eta = \frac{P_{T,j,t} M_{T,j,t}}{Y_{j,t}} = \frac{P_{T,j,t} M_{T,j,t}}{P_{T,j,t} M_{T,j,t} + P_{NT,j,t} M_{NT,j,t}}.$$

In other words, the parameter η represents the value share of tradable inputs in final production. We then use the average value of the share in the data to determine η. However, the average shares in the data for the two regions are not the same. Thus, we calibrate the common value η using the average of the two regions' shares.

Once we have the value of η, the productivity series in the final goods sector are constructed by inverting the production function:

$$A_{j,t} = \frac{Y_{j,t}}{M_{T,j,t}^{\eta} M_{NT,j,t}^{1-\eta}}.$$

The only remaining parameters to calibrate are those determining the endogenous wedges: ν for the unofficial production wedge and χ for the wedge on bond holdings. We set them to $\nu = \chi = 1$. Subsection 6.3 will conduct a sensitivity analysis with respect to these two parameters and show that our results are robust.

5 Measuring the wedges

Conditions (19)-(23) allow us to measure five wedges: $\tau_{i,j,t}^{l}$, $\tilde{\tau}_{i,j,t}^{l}$, $\tau_{i,j,t}^{k}$, $\tilde{\tau}_{i,j,t}^{k}$, and $\tau_{j,t}^{b}$. Following is a detailed description of each of them.

- The labor wedge in the official segment of sector i in region j is obtained by inverting Equation (19),

$$\tau_{i,j,t}^{l} = \frac{\frac{(1-\theta) X_{i,j,t} P_{i,j,t}}{C_{j,t}+G_{j,t}}}{\frac{\alpha L_{i,j,t}}{1-H_{j,t}-\widetilde{H}_{j,t}}} - 1.$$

  We compute the wedge using empirical counterparts for the following variables:

  $X_{i,j,t}P_{i,j,t}$: Official value added sector $i \in \{T, NT\}$, region $j \in \{N, S\}$.

  $C_{j,t}$: Consumption expenditures region $j \in \{N, S\}$.

  $G_{j,t}$: Government purchases region $j \in \{N, S\}$.

  $L_{i,j,t}$: Official hours in sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

  $H_{j,t} + \widetilde{H}_{j,t}$: Total hours of region $j \in \{N, S\}$.

- The unofficial labor wedge in sector i of region j is obtained by inverting Equation (20),

$$\tilde{\tau}_{i,j,t}^{l} = \frac{\frac{(1-\theta) \widetilde{X}_{i,j,t} P_{i,j,t}}{C_{j,t}+G_{j,t}}}{\frac{\alpha \widetilde{L}_{i,j,t}}{1-H_{j,t}-\widetilde{H}_{j,t}}} - 1.$$

  This is computed using empirical counterparts for the following variables:

$\widetilde{X}_{i,j,t}P_{i,j,t}$: Unofficial value added sector $i \in \{T, NT\}$, region $j \in \{N, S\}$.

$C_{j,t}$: Consumption expenditures region $j \in \{N, S\}$.

$G_{j,t}$: Government purchases region $j \in \{N, S\}$.

$\widetilde{L}_{i,j,t}$: Unofficial hours sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

$H_{j,t} + \widetilde{H}_{j,t}$: Total hours of region $j \in \{N, S\}$.

- The investment wedge in the official segment of sector i in region j is obtained by inverting Equation (21),

$$\tau^k_{j,t+1} = \frac{\frac{\beta\theta X_{i,j,t+1}P_{i,j,t+1}}{K_{i,j,t+1}}}{\frac{C_{j,t+1}+G_{j,t+1}}{C_{j,t}+G_{j,t}} - \beta(1-\delta)} - 1.$$

  The wedge is computed using empirical counterparts for the following variables:

  $X_{i,j,t+1}P_{i,j,t+1}$: Official value added sector $i \in \{T, NT\}$, region $j \in \{N, S\}$.

  $C_{j,t}$, $C_{j,t+1}$: Consumption expenditures region $j \in \{N, S\}$.

  $G_{j,t}$, $G_{j,t+1}$: Government purchases region $j \in \{N, S\}$.

  $K_{i,j,t+1}$: Official capital in sector $i \in \{T, NT\}$, region $j \in \{N, S\}$.

- The investment wedge in the unofficial segment of sector i of region j is obtained by inverting Equation (22),

$$\tau^k_{j,t+1} = \frac{\frac{\beta\theta \widetilde{X}_{i,j,t+1}P_{i,j,t+1}}{\widetilde{K}_{i,j,t+1}}}{\frac{C_{j,t+1}+G_{j,t+1}}{C_{j,t}+G_{j,t}} - \beta(1-\delta)} - 1,$$

  which we compute using empirical counterparts for the following variables:

  $\widetilde{X}_{i,j,t+1}P_{i,j,t+1}$: Unofficial value added sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

  $C_{j,t}$, $C_{j,t+1}$: Consumption expenditures region $j \in \{N, S\}$.

  $G_{j,t}$, $G_{j,t+1}$: Government purchases region $j \in \{N, S\}$.

  $\widetilde{K}_{i,j,t+1}$: Unofficial capital in sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

- The bond wedge is obtained by inverting equation (23),

$$\tau^b_{j,t+1} = \frac{\frac{C_{j,t+1}+G_{j,t+1}}{C_{j,t}+G_{j,t}}}{\beta \frac{1}{Q_t} \frac{P_{T,j,t+1}}{P_{T,j,t}}}.$$

  The empirical counterparts of the variables used to compute the wedge are:

  $C_{j,t}$, $C_{j,t+1}$: Consumption expenditures region $j \in \{N, S\}$.

$G_{j,t}$, $G_{j,t+1}$: Government purchases region $j \in \{N, S\}$.

$\frac{1}{Q_t}$: Real interest rate.

$P_{T,j,t+1}/P_{T,j,t}$: Growth of tradable price index in region $j \in \{N, S\}$.

The last wedge we need to compute is the productivity wedge in the unofficial segment of the economy, which is obtained by inverting the production function (2):

$$\tau^x_{i,j,t} = 1 - \frac{\widetilde{X}_{i,j,t}}{z_{i,j,t}\widetilde{K}^{\theta}_{i,j,t}\widetilde{L}^{1-\theta}_{i,j,t}}.$$

The computation of $\tau^x_{i,j,t}$ uses empirical counterparts for the following variables:

$\widetilde{X}_{i,j,t+1}$: Unofficial real value added in sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

$\widetilde{K}_{i,j,t+1}$: Unofficial capital in sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

$\widetilde{L}_{i,j,t}$: Unofficial hours in sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

$z_{i,j,t}$: Productivity in sector $i \in \{T, NT\}$ of region $j \in \{N, S\}$.

A detailed description of the empirical series listed above is provided in Appendix B.

5.1 Results

Figure 9 plots the labor wedges, $\tau^l_{i,j,t}$ and $\tilde{\tau}^l_{i,j,t}$, in the official and unofficial segments of the tradable and nontradable sectors. The labor wedge is higher in the South than in the North, except for the tradable official sector, where they are about the same. This suggests that part of the lower working hours observed in the South can result from stronger labor distortions in the South. The differences in the labor wedge between the South and the North are substantial in the nontradable sector, which is also the economy's largest sector.

Figure 10 plots the investment wedges $\tau^k_{i,j,t}$ and $\tilde{\tau}^k_{i,j,t}$. The wedges are quite volatile, with two visible spikes. The first is during the great financial crisis in 2008. The second spike is around the European debt crisis in 2011-2012, in which Italy was directly involved. These two periods were associated with declines in investment, which, in our model, are caused by higher investment wedges. One way to interpret the spikes is that the two crises increased the perceived risk leading to heavier discounting. Since our model does not have risk, this is captured by an increase in the investment wedge.

Comparing the two regions, we observe that the investment wedges are relatively high in the South. This could be another contributing factor to lower Southern income, in this case, due to lower investment and capital.

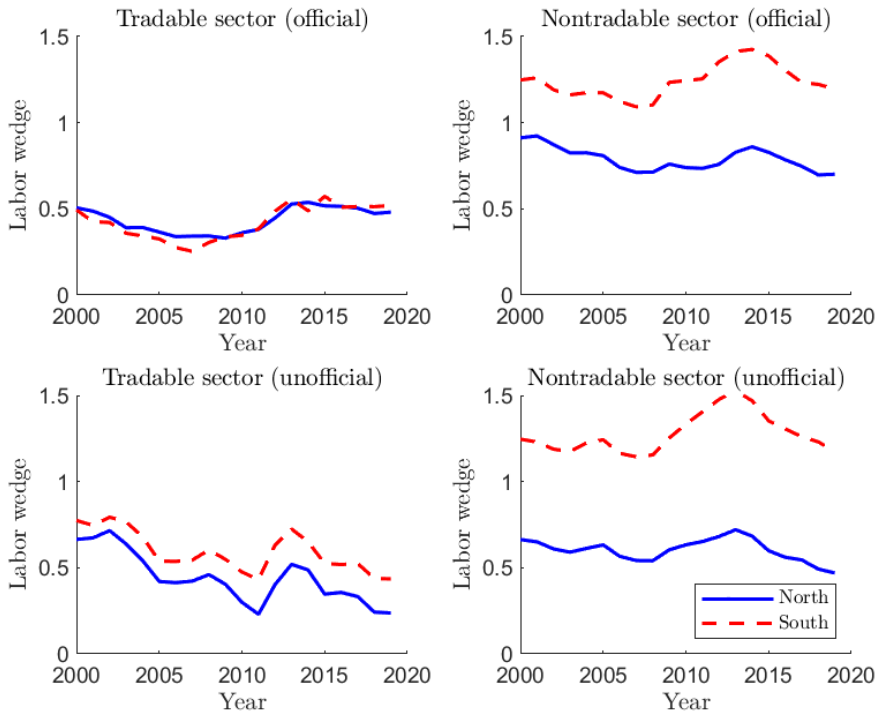

Figure 9: Labor wedge.

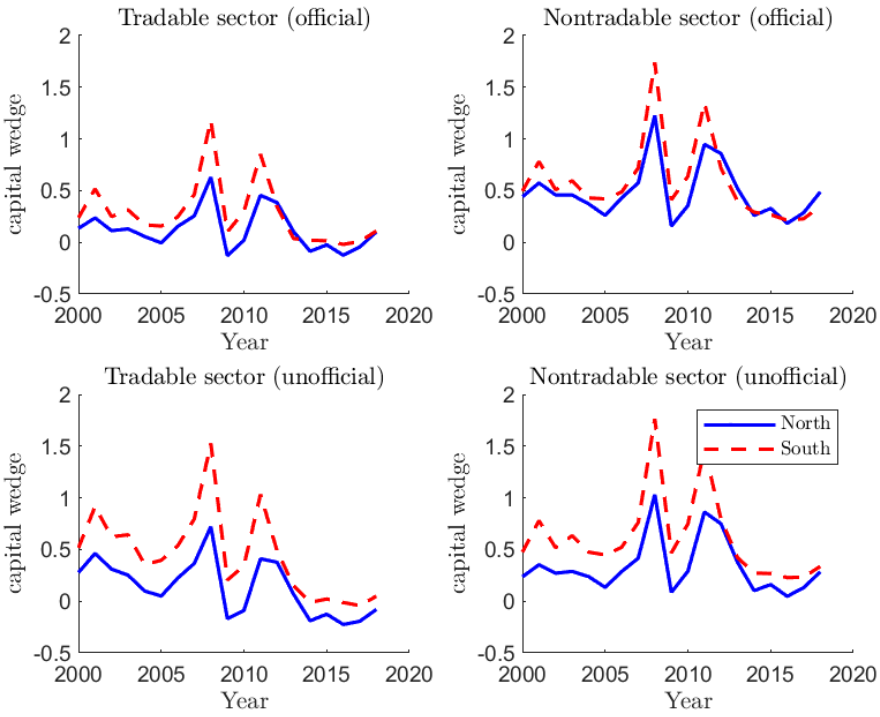

Figure 10: Investment wedge.

Figure 11 plots productivity in the official segments of the economy, $z_{i,j,t}$, and the unofficial segments, $(1-\tau^x_{i,j,t})z_{i,j,t}$. The difference between official and unofficial is driven by the productivity wedge $\tau^x_{i,j,t}$. When looking at the official segments of the tradable and nontradable sectors, we can see that productivity is lower in the South than in the North. However, as emphasized earlier, the difference results from differences in both price and actual productivity.[7] But independently of whether the differences come from productivity or prices, they provide another clue about the sources of regional income disparities: the South produces less in value not only because it utilizes less labor and accumulates less capital but also because capital and labor are less productive.

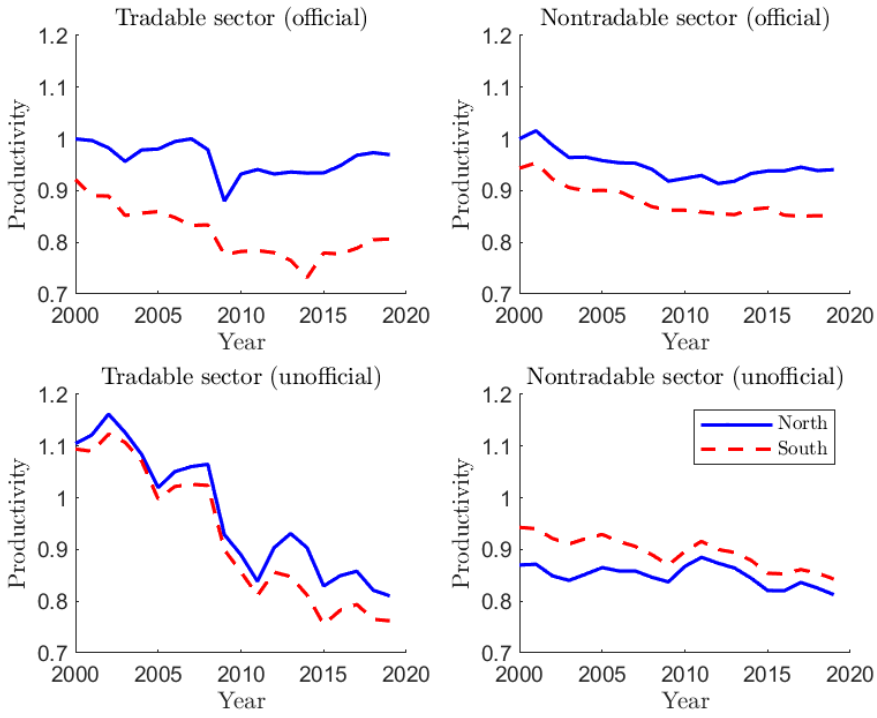

Figure 11: Productivity in the intermediate sectors.

The regional productivity differences are smaller in the unofficial segment of the tradable sector and even higher in the Southern unofficial segment of the nontradable sector (see the fourth panel of Figure 11). This could be relevant in explaining why a large fraction of employment in the South is unofficial.

The last graph presented in this section plots the productivity of final goods production. As shown in Figure 12, there are no large productivity differences between the North and the South in the production of final goods.

[7] We cannot separate the component coming from price and actual productivity differences because we only have price indices normalized to 1 in 2015 for all regions. Thus, we cannot compare their levels across regions.

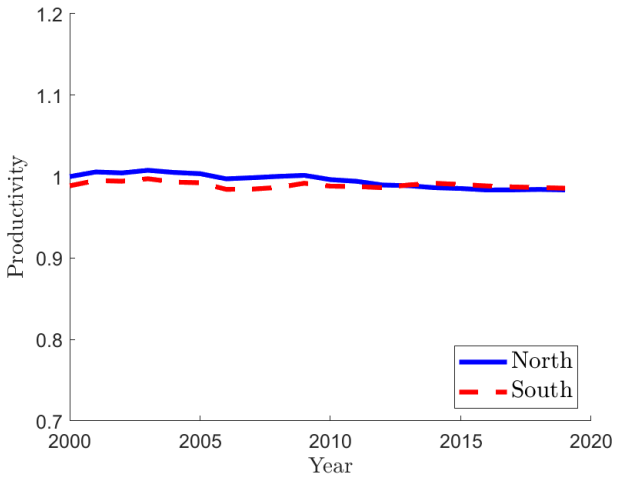

Figure 12: Productivity in the final goods sector.

6 Counterfactual analysis with steady states

The steady-state equilibrium for our model can be derived once all wedges and productivities are set to their average values. The productivity wedge in the unofficial segments of intermediate production, $\tau^x_{i,j,t}$, and the bond wedge, $\tau^b_{j,t}$, require special consideration since they are endogenous. Let us start with the productivity wedge, which we assumed to take the form:

$$\tau^x_{i,j,t} = \kappa_{i,j,t} + \nu \cdot \left(\frac{\widetilde{X}_{i,j,t}}{X_{i,j,t} + \widetilde{X}_{i,j,t}} \right).$$

Denote by $s_{i,j}$, without time subscript, the sample mean of the share of unofficial production in each sector/region. This is the average value of $\widetilde{X}_{i,j,t}/(X_{i,j,t} + \widetilde{X}_{i,j,t})$ in the data. Also, denote by $\tau^x_{i,j}$ (without time subscript) the average productivity wedge in the unofficial sector i of region j. The constant steady-state value of the exogenous component is $\kappa_{i,j} = \tau^x_{i,j} - \nu s_{i,j}$. This condition guarantees that if in the steady state, the unofficial share of production in each sector is equal to the average in the data, $s_{i,j}$, the steady-state productivity wedge is also equal to the average in the data.

The bond wedge is:

$$\tau^b_{j,t} = \zeta_{j,t} + \chi \cdot B_{j,t}.$$

Denote by B_j the average bond holding in region j. In equilibrium, we must have that $B_N \mu_N + B_S \mu_S = 0$. Also, in a steady state, the wedge must be the same in the two regions. In fact, in the steady state, condition (18) becomes $Q = \beta(1 - \tau^b_N)$ and $Q = \beta(1 - \tau^b_S)$. Since Q and β are the same for the two regions, τ^b_N must equal τ^b_S. Therefore, the constant value of the intercept is equal to $\zeta_j = \tau^b - \chi \cdot B_j$, where τ^b is the mean of the average wedges in the two regions and B_j is the average bond holding. Since we do not have data for B_j, we set it to zero.

6.1 Why is income lower in the South?

As we have seen in the first part of the paper, per capita income in the South is significantly lower than in the North. As is standard in national accounting, this can be related to differences in factor inputs (labor and capital) and productivity. Regarding factor inputs, we have already seen that per capita working hours in the South are significantly lower than in the North. In this subsection, we would like to answer why. Is it because there are distortions that directly affect the use of labor? Is it because the accumulation and usage of capital are distorted so that each worker has lower capital? Are there distortions that reduce the efficiency of capital and labor used in production, resulting in lower total factor productivity?

The wedge analysis provides us with a tool to answer these questions. To this end, we conduct counterfactual exercises starting from a baseline environment in which the two regions are characterized by their wedges. The baseline model replicates the data since the wedges are measured from it. Starting from this baseline environment, we change one of the wedges for the South and set it to the average measured in the data for the North. This allows us to answer how this wedge alone affects per capita production (domestic value added). We will do that for each wedge, one at a time.

Our counterfactual analysis will be based on steady-state comparisons, and the results are reported in Table 1. In the top section of the table, we start with the baseline model in which the North has its own (average) wedges, and the South has its own (average) wedges. We think of this case as representing the actual economy. As can be seen from the first line, the Southern output is only 57.4 percent of the Northern output. Although not reported, the gap in regional income, as opposed to output, is somewhat smaller because of net transfers from the North to the South.

The second row of Table 1 shows the steady-state values when we assign the same labor wedges to the South as in the North. With the new labor wedges, the income ratio increases by 3.7 percentage points, reducing the income gap. This shows that differences in labor market distortions contribute to regional income disparities but are not the leading cause.

The third row conducts a similar exercise, but only the investment wedges are changed this time. More specifically, the investment wedges for the South now take the average values measured in the data for the North (all the other wedges are kept at the average values measured earlier in the paper). This increases the output ratio by 6.1 percentage points, bigger than the increase generated by the labor wedge but still only a fraction of the total gap of 42.6 percent.

Productivity differences in intermediate production generate the largest increase in the output ratio. As seen in the fourth row of the table, by assigning to the South the same intermediate productivities as the North, the income ratio increases by 16.5 percentage points. This is about 40 percent of the output gap between the South and the North. We should emphasize again that differences in productivity could reflect differences in prices (at least in the nontradable

Table 1: Steady-state per capita output in units of tradable goods.

(a) Removing Southern wedge differences (one at a time)

	Output per-employee North	Output per-employee South	South/North ratio	Change South/North ratio
Baseline calibration	57,189	32,821	0.574	
Labor wedge South same as North	57,189	34,954	0.611	0.037
Investment wedge South same as North	57,189	36,331	0.635	0.061
Intermediate productivity South same as North	57,189	42,241	0.739	0.165
Final productivity South same as North	57,189	32,975	0.577	0.003
Absence of inter-regional transfers	54,545	37,605	0.689	0.116
Sum of changes in South/North ratio				0.382

(b) Adding Southern wedge differences (one at a time)

	Output per-employee North	Output per-employee South	South/North ratio	Change South/North ratio
Symmetric calibration	54,545	54,578	1.000	
South has its own labor wedge	54,545	51,876	0.951	-0.050
South has its own investment wedge	54,545	49,801	0.913	-0.088
South has its own intermediate productivity	54,545	43,536	0.798	-0.202
South has its own final productivity	54,545	54,375	0.997	-0.004
South receives net transfers from the North	57,189	49,748	0.870	-0.131
Sum of changes in South/North ratio				-0.475

sector) and actual productivity. Despite this, the exercise is still informative because differences in nontradable prices could reflect regional productivity gaps. It is plausible that the prices of tradable goods are lower in a poorer country or region (for example, because of some price-to-market strategy by firms), but they increase if the country or region becomes richer. Contrary to the intermediate production sector, assigning the same productivity in the final good sector to the South as in the North has a negligible impact on the output gap (see fifth row).

The impact of net fiscal transfers between regions is also sizable. As a byproduct of the baseline calibration, the steady-state equilibrium is characterized by a difference between domestic production and domestic absorption, that is, $Y_{j,t} - C_{j,t} - I_{j,t} - G_{j,t}$. We interpret the resulting difference as capturing the fiscal transfers between regions. The steady-state values for the South

and the North are, respectively, -6.56 percent and 2.08 percent of their output. We then ask how the steady-state equilibrium will change if we set these transfers to zero. This implies that the South loses fiscal transfers , which are 6.56 percent of the value of its domestic output. In comparison, the North will no longer pay transfers to the South, which are 2.08 percent of the North's output.[8]

The sixth row in the top section of Table 1 shows that removing inter-regional transfers increases the output ratio by 11.6 percentage points. This is a reduction in the output gap of about a quarter. The reason transfers from the North to the South lead to a reduction in Southern production and an increase in Northern production is that they have an income effect on labor supply. Thanks to the transfers from the North, the South can sustain higher consumption (higher standard of living), which reduces the value of working. The opposite happens in the North.

When we sum the changes induced by the five factors (labor wedge, investment wedge, intermediate productivity, final productivity, and inter-regional transfers), the income ratio increases by 38.2 percentage points. This differs from the overall output gap, which equals 42.6 percent, because interaction effects occur when two or more wedges change simultaneously. However, the sum of the independent effects is not that different from the total gap.

An alternative way to quantify the importance of the various wedges is to start from the benchmark case in which the South has exactly the same wedges, productivity, and zero transfers as the North. We can then ask how the output gap changes when the South is assigned its own wedge, one at a time. The results are in the bottom section of Table 1.

The first row reports per capita output in the benchmark calibration where the two regions have the same wedges, productivities, and fiscal transfers. Since the two regions are now symmetric, per capita output is the same. The remaining rows show how the ratio between the South's and the North's output changes when one of the wedges is changed for the South. The exercise provides a picture similar to the one shown in the top section of the table: the most important factor is the differential in intermediate productivity and then the inter-regional fiscal transfers. The sum of the independent effects is 47.5 percentage points, which is a bit larger than the output gap of 42.6 percent.

6.2 A summary measure

As a summary measure of the contribution of the various wedges, we take the averages of the contributions reported in the top and bottom sections of Table 1. The resulting numbers are

[8]These inter-regional transfers are smaller than those behind Figure 6. For the aggregation of the Southern regions, the average in the data is 11 percent, and for the aggregation of the Northern regions, the average is 8.5 percent. Therefore, our exercise can be interpreted as providing a lower estimate of the impact of eliminating inter-regional fiscal transfers.

shown in Table 2. The sum of the various contributions is 42.8 percent, which is very close to the gap generated by the baseline calibration of 42.6 percent.

Table 2: Summary contributions to output gap between Southern and Northern regions.

Labor market wedge	Investment wedge	Intermediate productivity	Final productivity	Inter-regional transfers	**Sum of all contributions**
4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**

To put it briefly, the lower per capita output of Southern Italy can be related to lower productivities, larger distortions in labor and capital markets, and inter-regional transfers. Productivity, though, seems to be the most important factor, followed by inter-regional fiscal transfers.

6.3 Sensitivity analysis

In calibrating the model, we set the parameter values of ν and χ to 1. The first parameter, ν, determines the endogenous productivity wedge between the official and unofficial segments of production. The second parameter, χ, determines the cost of holding bonds issued by the other region. However, the chosen values of these two parameters were not based on empirical observations. It is then important to show how the counterfactual results depend on these two parameters.

Table 3 reports the same summary statistics as in Table 2 but for different values of ν and χ. As can be seen, the parameter ν somewhat affects the results but only marginally. Even if we reduce the value of this parameter from 1.0 to 0.1, the contribution of intermediate productivity declines a little but remains the most important factor accounting for the income gap. The parameter χ, on the other hand, is completely irrelevant to the results.

Table 3: Summary contributions to output gap between Southern and Northern regions: Sensitivity to wedge parameters ν and χ.

	Labor market wedge	Investment wedge	Intermediate productivity	Final productivity	Inter-regional transfers	**Sum of all contributions**
$\nu = 2.0$, $\chi = 1.0$	4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**
$\nu = 1.0$, $\chi = 1.0$	4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**
$\nu = 0.5$, $\chi = 1.0$	4.4%	7.5%	18.3%	0.3%	12.3%	**42.8%**
$\nu = 0.1$, $\chi = 1.0$	4.2%	7.6%	15.3%	0.3%	12.3%	**39.8%**
$\nu = 1.0$, $\chi = 2.0$	4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**
$\nu = 1.0$, $\chi = 1.0$	4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**
$\nu = 1.0$, $\chi = 0.5$	4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**
$\nu = 1.0$, $\chi = 0.1$	4.3%	7.4%	18.4%	0.3%	12.3%	**42.8%**

7 Conclusion

This paper uses macroeconomic tools to investigate the possible sources of income disparities between Southern and Northern Italy. We apply the wedge analysis in a two-region model (North and South) with two sectors (tradable and nontradable) and two segments of the economy (official and unofficial). We find that the most important factor accounting for income disparities is productivity differences between North and South, followed by inter-regional fiscal transfers. Differential distortions in labor and capital markets also contribute to the disparity, but productivity differences and fiscal redistribution account for more than 70 percent of the output gap. The next step is to deepen the analysis to understand why the wedges (or distortions), especially those affecting productivity, are different between the North and the South. This should be the focus of future research.

References

Boeri, T., A. Ichino, E. Moretti, and J. Posch (2021): "Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces," *Journal of the European Economic Association*, 19, 3249–3292.

Boldrin, M. and F. Canova (2001): "Inequality and Convergence in Europe's Regions: Reconsidering European Regional Policies," *Economic Policy*, 16, 207–253.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007): "Business Cycle Accounting," *Econometrica*, 75, 781–836.

Chenery, H. B. (1962): "Development Policies for Southern Italy," *Quarterly Journal of Economics*, 76, 515–547.

Cheremukhin, A., M. Golosov, S. Guriev, and A. Tsyvinski (2015): "The Economy of People's Republic of China from 1953," Working Paper 21397, National Bureau of Economic Research.

——— (2017): "The Industrialization and Economic Development of Russia through the Lens of a Neoclassical Growth Model," *Review of Economic Studies*, 84, 613–649.

Clough, S. B. and C. Livi (1956): "Economic Growth in Italy: An Analysis of the Uneven Development of North and South," *Journal of Economic History*, 16, 334–349.

Daniele, V. and P. Malanima (2011): *Il divario nord-sud in Italia, 1861-2011*, vol. 273, Rubbettino Editore: Soveria Mannelli.

Eckaus, R. S. (1961): "The North-South Differential in Italian Economic Development," *Journal of Economic History*, 21, 285–317.

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, 499–528.

Hooghe, L., G. Marks, A. H. Schakel, S. Chapman-Osterkatz, S. Niedzwiecki, and S. Shair-Rosenfield (2016): *Measuring Regional Authority. Volume I. A Postfunctionalist Theory of Governance*, Oxford University Press.

Iuzzolino, G., G. Pellegrini, and G. Viesti (2011): "Convergence among Italian Regions, 1861-2011," Quaderni di storia economica (Economic History Working Papers) 22, Bank of Italy, Economic Research and International Relations Area.

——— (2013): "Regional Convergence," in *The Oxford Handbook of the Italian Economy Since Unification*, ed. by G. Toniolo, Oxford University Press, chap. 20, 571–598.

Rosés, J. R. and N. Wolf (2018): "Regional Economic Development in Europe, 1900–2010: A Description of the Patterns," in *The Economic Development of Europe's Regions*, ed. by J. R. Rosés and N. Wolf, Routledge, 3–41.

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

Zamagni, V. (1993): *The Economic History of Italy, 1860-1990*, Clarendon Press.

Appendix

A Calibration

This appendix describes the detailed steps for calibrating some of the model parameters. In particular, it describes the calibration of the depreciation rate and the factor shares in intermediate production.

A.1 Depreciation rate

To construct a series of total capital, we use the method of perpetual inventories. We need to jointly determine the series of capital stock $\overline{K}_{i,j,0}, ..., \overline{K}_{i,j,T}$ for each sector $i \in \{T, NT\}$ and region $j \in \{T, NT\}$, as well as the constant depreciation rate δ. We do that by solving the following system of equations:

$$\overline{K}_{i,j,t+1} = \overline{K}_{i,j,t} - \delta \overline{K}_{i,j,t} + \overline{I}_{i,j,t} \tag{27}$$

$$\frac{1}{T}\sum_{t=1}^{T} \overline{D}_t = \frac{1}{T}\sum_{t=1}^{T} \delta \overline{K}_t \tag{28}$$

$$\frac{\overline{K}_{i,j,0}}{\overline{Y}_{i,j,0}} = \frac{1}{5}\sum_{t=1}^{5} \frac{\overline{K}_{i,j,t}}{\overline{Y}_{i,j,t}} \tag{29}$$

The first equation is the law of motion for capital. The second equation imposes the condition that the average consumption of fixed capital at the national level, denoted by $\overline{D}_t$, is equal to the national depreciation calculated with the fixed rate δ. The third equation requires that the simple average of the capital-output ratio for the first five years is equal to the initial capital-output ratio.

We use data from 1970 to 2019. We measure investment $\overline{I}_{i,j,t}$ with gross fixed capital formation in constant LCUs. The variable $\overline{D}_t$ is the consumption of fixed capital in constant USD from the World Bank World Development Indicators. We convert it into LCUs using annual foreign exchange averages (Italian lira/dollar before 1999 and euro/dollar since 1999). The procedure returns a depreciation rate that is equal to $\delta = 0.052$.

A.2 Intermediate factor shares

1. From the first-order condition of labor, we have:

$$(1-\theta_i) z_{i,t} K_{i,j,t}^{\theta_i} L_{i,j,t}^{-\theta_i} P_{i,j,t} = (1+\tau_{i,j,t}^{l}) W_{i,j,t}.$$

Multiplying both sides by $L_{i,j,t}$ we obtain:

$$(1-\theta_i) X_{i,j,t} P_{i,j,t} = W_{i,j,t} L_{i,j,t} (1+\tau_{i,j,t}^{l}). \tag{30}$$

The term $W_{i,j,t}L_{i,j,t}(1+\tau^l_{i,j,t})$ is the observed labor income paid by firms. However, only $W_{i,j,t}L_{i,j,t}$ is the labor income received by households. A similar relation holds in the unofficial segment of the sector, that is,

$$(1-\theta_i)\tilde{X}_{i,j,t}P_{i,j,t} = (1+\tilde{\tau}^l_{i,j,t})\tilde{W}_{i,j,t}\tilde{L}_{i,j,t}. \tag{31}$$

We now aggregate over the two regions (North and South) and the two segments of the sector (official and unofficial). To aggregate quantities over the two regions, we must express them in the same units. Thus, we convert all quantities in terms of tradable goods. This is obtained by dividing region j quantities by region j tradable price $P_{T,j,t}$.[9]

Dividing equations (30) and (31) by $P_{T,j,t}$, adding them together and re-arranging we obtain:

$$1-\theta_i = \frac{\sum_{j=N,S}\left[(1+\tau^l_{i,j,t})W_{i,j,t}L_{i,j,t}/P_{T,j,t} + (1+\tilde{\tau}^l_{i,j,t})\tilde{W}_{i,j,t}\tilde{L}_{i,j,t}/P_{T,j,t}\right]}{\sum_{j=N,S}\left[X_{i,j,t}P_{i,j,t}/P_{T,j,t} + \tilde{X}_{i,j,t}P_{i,j,t}/P_{T,j,t}\right]}.$$

The numerator is the national labor income in sector $i \in \{T, NT\}$, and the denominator is the national value added, also in sector $i \in \{T, NT\}$. Thus, $1-\theta_i$ is the national labor share in the particular sector i. To calibrate θ_i, we use average shares for the national economy since we impose $theta_=\theta_{NT} = \theta$. Income share data are from the International Labor Organization (ILO).

B Data description

This appendix describes the data used in the wedge analysis, including some imputation choices made to construct the required empirical series.

Price of tradable goods. In the model $P_{T,j,t}$ is the price of tradable intermediates in region j, expressed in region j's final output. The Italian statistical agency, ISTAT, provides nominal price indices for tradable value added, $\mathcal{I}_{T,j,t}$, nontradable value added, $\mathcal{I}_{NT,j,t}$, and for total value added, $\mathcal{I}_{Y,j,t}$. The prices in the model, then, can be computed as $P_{T,j,t} = \mathcal{I}_{T,j,t}/\mathcal{I}_{Y,j,t}$ and $P_{NT,j,t} = \mathcal{I}_{NT,j,t}/\mathcal{I}_{Y,j,t}$. The price indices are normalized in each region and all equal 1 in 2015. Therefore, they cannot be used to derive measures of actual relative prices between regions.

Employment and unobserved value added. For employment, ISTAT reports numbers for "regular" and "irregular" occupations in each sector and each region. We interpret irregular occupations as employment in the unofficial or underground economy. ISTAT, however, does not provide data for regular and irregular value added for each region. It only reports estimates for the underground value added at the national level. This is what it calls "unmeasured economy activity." Also, the estimates for irregular value added at the national level are only available for the most recent years. Because of

[9] Remember that $P_{T,j,t}$ is the price of tradable inputs in units of final goods in region j. For example, if we take the wage $W_{i,j,t}$ in region j, which is expressed in units of the final good, and we divide by $P_{T,j,t}$, we obtain the wage expressed in units of tradable (intermediate) goods.

this, we need to make some imputations for quantifying unmeasured value added at the regional and sectoral levels.

We first construct a measure of irregular value added by sector for the national economy for the earlier years, 2000–2010. Given that the proportion of irregular-to-total employed mimics the ratio of underground-to-regular value added for the period available, 2011–2019, we splice the share of unmeasured value added backward using the rate of change in the share of irregular occupations. We back out the euro value by multiplying the ratio by the value of regular value added over 2000–2010. Finally, the imputation of the unmeasured value added in each sector for the two regions is obtained using the share of irregular workers in each region/sector.

The first step of the procedure computes:

$$y_{i,t-1} = \frac{y_{i,t}}{1+g_{i,t}} \qquad t = 2001, \ldots, 2011, \quad \forall i,$$

where $y_{i,t}$ is the unmeasured value added of the Italian economy as a fraction of total value added for each sector i, and $g_{i,t}$ is the rate of change of the ratio of irregular workers to total workers in the economy by sector. With that in hand, it is possible to back out the underground value added series for each sector by multiplying $y_{i,t}$ by the national aggregate, that is, $Y_{i,t} = y_{i,t} \times Y_t$. Finally,

$$Y_{i,j,t} = Y_{i,t} \times \frac{u_{i,j,t}}{u_{i,t}}, \qquad j = N, S, \qquad \forall i,$$

where $u_{i,j,t}$ is the number of irregular workers in sector i of region j and $u_{i,t}$ is the number of irregular workers in sector i for the whole national economy.

Official and unofficial capital. The previous appendix described how we constructed a measure of total sectoral capital in each region, $\overline{K}_{i,j,t}$. Given these series, we now need to allocate them to the official and unofficial segments of the economy so that we have empirical measures of $K_{i,j,t}$ and $\tilde{K}_{i,j,t}$. To do so, we start with the production in the official and unofficial segments of each sector:

$$X_{i,j,t} = z_{i,j,t} K_{i,j,t}^{\theta_i} L_{i,j,t}^{1-\theta_i}, \tag{32}$$

$$\widetilde{X}_{i,j,t} = (1-\tau_{i,j,t}^{x}) z_{i,j,t} \widetilde{K}_{i,j,t}^{\theta_i} \widetilde{L}_{i,j,t}^{1-\theta_i}. \tag{33}$$

Dividing the second equation by the first equation, we obtain:

$$\frac{\widetilde{X}_{i,j,t}}{X_{i,j,t}} = (1-\tau_{i,j,t}^{x}) \left(\frac{\widetilde{K}_{i,j,t}}{K_{i,j,t}}\right)^{\theta_i} \left(\frac{\widetilde{L}_{i,j,t}}{L_{i,j,t}}\right)^{1-\theta_i}. \tag{34}$$

If we had a measure of the relative efficiency of unofficial production –the term $1-\tilde{\tau}_{i,j,t}$– we could use measures of $\tilde{X}_{i,j,t}$, $X_{i,j,t}$, $\tilde{L}_{i,j,t}$, $L_{i,j,t}$, to compute $\tilde{K}_{i,j,t}/K_{i,j,t}$ from Equation (34). Once we have this ratio, we could use the total value of capital in sector i and region j, denoted by $\overline{K}_{i,j,t}$, to determine the individual allocations in the official and unofficial segments of the sector. Unfortunately, we do not have a direct measure of $1-\tilde{\tau}_{i,j,t}$. Hence, we have to use a proxy. Since $1-\tilde{\tau}_{i,j,t}$ is the relative productivity

between the unofficial and the official segments of the sector, as a proxy, we use the relative productivity of labor, that is,

$$1 - \tau^x_{i,j,t} = \frac{\frac{\tilde{X}_{i,j,t}}{\tilde{L}_{i,j,t}}}{\frac{X_{i,j,t}}{L_{i,j,t}}}. \tag{35}$$

Hours worked. In the model, agents work a proportion of their total endowment of one unit of time. We have data for the total number of working hours per year, region, and economic activity. We assign them to official and unofficial hours in proportion to the available data on regular and irregular occupations,

$$\text{Hours}_{i,j,t} = \overline{\text{Hours}}_{i,j,t} \times \frac{e_{i,j,t}}{\tilde{e}_{i,j,t} + e_{i,j,t}} \tag{36}$$

$$\widetilde{\text{Hours}}_{i,j,t} = \overline{\text{Hours}}_{i,j,t} \times \frac{\tilde{e}_{i,j,t}}{\tilde{e}_{i,j,t} + e_{i,j,t}} \tag{37}$$

where $\overline{\text{Hours}}_{i,j,t}$ are the total (official and unofficial) hours worked in a year by the whole labor force, and $e_{i,j,t}$ and $\tilde{e}_{i,j,t}$ represent the number of regular and irregular employed persons, respectively.

Next, we normalize the number of hours as a fraction of the total hours available in a year so that labor is between 0 and 1. The normalized measures of labor are:

$$L_{i,j,t} = \frac{\text{Hours}_{i,j,t}}{24 \times 365 \times e_{i,j,t}} \tag{38}$$

$$\widetilde{L}_{i,j,t} = \frac{\widetilde{\text{Hours}}_{i,j,t}}{24 \times 365 \times \tilde{e}_{i,j,t}} \tag{39}$$

C Additional data details

This appendix provides additional details about the empirical variables used in the analysis.

1. **Nominal value added, national and by region, and by economic sector:**

   Tipo aggregato: "Valore aggiunto"; Valutazione: "Prezzi correnti"; Correzione: "Dati grezzi"; Tipologia di prezzo: "Prezzi base"; Edizione: "Sep. 2022."

   Data are available from 1995 to 2021 and provided in millions of euros. We retrieved series for national, north, and south levels of aggregation, disaggregated for agriculture, manufacturing, and services. Agriculture and manufacturing define our tradable sector and services the nontradable sector.

2. **Real value added, national and by region, and by economic sector:**

   Tipo aggregato: "Valore aggiunto"; Valutazione: "Valori concatenati con anno di riferimento 2015"; Correzione: "Dati grezzi"; Tipologia di prezzo: "Prezzi base"; Edizione: "Sep. 2022."

Data are available from 1995 to 2021 and provided in millions of euros. We retrieved data for national, north, and south levels of aggregation, disaggregated for agriculture, manufacturing, and services sectors.

3. **Nominal consumption expenditures, national and by region:**

   Tipo aggregato: "Spesa per consumi finali sul territorio economico delle famiglie residenti e non residenti"; Valutazione: "Prezzi correnti"; Correzione: "Dati grezzi"; Edizione: "Dic. 2021."

   Data are available from 1995 to 2020 and provided in millions of euros. We retrieved data for national, north, and south levels of aggregation.

4. **Real consumption expenditures, national and by region:**

   Tipo aggregato: "Spesa per consumi finali sul territorio economico delle famiglie residenti e non residenti"; Valutazione: "Valori concatenati con anno di riferimento 2015"; Correzione: "Dati grezzi"; Edizione: "Dic. 2021."

   Data are available from 1995 to 2020 and provided in millions of euros. We retrieved data for national, north, and south levels of aggregation.

5. **Official employment, national and by region, and by economic sector:**

   Dataset: "Occupati (migliaia)"; Classe di età: "15 anni e più"; Tipologia di occupazione: "Regolare"; Edizione: "Dic. 2021."

   The dataset includes both employees and self-employed. Data are available from 2000 to 2019 and in raw count format. We retrieved data for the national, north, and south levels of aggregation, disaggregated for agriculture, manufacturing, and services.

6. **Unofficial employment, national and by region, and by economic sector:**

   Dataset: "Occupati (migliaia)"; Classe di età: "15 anni e più"; Tipologia di occupazione: "Irregolare"; Edizione: "Dic. 2021."

   Data are available from 2000 to 2019 and in raw count format. We retrieved data separately for agriculture, manufacturing, and services at the national, north, and south levels of aggregation.

7. **Employment rate, by region:**

   Aggregato: "Occupati (migliaia)"; Tipologia di occupazione: "Totale"; Posizione nella professione: "Totale"; Edizione: "Sep. 2022."

   Data for employed in thousands are available from ISTAT. Data are taken across employees and self-employed.

8. **Population data:**

   Dataset: "Popolazione residente ricostruita - Anni 2002-2019"; Popolazione al 1º gennaio; Sesso: "Totale"; Cittadinanza: "Totale."

   The dataset from ISTAT reconstructs the total population data (regardless of sex and citizenship) as of January 1st. Population data are available by age year. The working age population was constructed by summing each age in the 15-64 range bracket.

9. **Hours worked, national and by region, and by economic sector**

   Dataset: "Ore lavorate (migliaia)"; Classe di età: "15 anni e più"; Tipologia di occupazione: "Totale"; Edizione: "Dic. 2021".

   We retrieved data separately for agriculture, manufacturing, and services at the national, north, and south levels of aggregation.

10. **Gross and net nominal capital stock, national, and by economic sector:**

    Datasets: "Stock di attività non finanziarie (lordo) and Stock di attività non finanziarie (netto)"; "Attività non finanziarie: Totale capitale fisso per tipo di attività"; Valuazione: "Prezzi di sostituzione correnti"; Edizione: "Sep. 2022."

    This corresponds to a measure of the total stock of fixed capital in nonfinancial activities (gross or net) at current prices. Data are available at the national level only. They are available from 1995 to 2021, and by economic sector: agriculture, manufacturing, and services. The gross capital stock is the value of capital goods still in use and evaluated as if they were new capital goods, without accounting for their age or depreciation undergone through time. The net capital stock is the value of capital goods still in use and evaluated as if they were new capital goods, minus the *compounded* depreciation undergone through time.

11. **Gross and net real capital stock, national, and by economic sector:**

    Datasets: "Stock di attività non finanziarie (lordo) and Stock di attività non finanziarie (netto)"; "Attività non finanziarie: Totale capitale fisso per tipo di attività"; Valuazione: "Valori concatenati con anno di riferimento 2015"; Edizione: "Sep. 2022."

    This corresponds to the total stock of fixed capital in nonfinancial activities (gross or net) at chained prices (2015=100). Data are available at the national level only. They are available from 1995 to 2021, and by economic sector: agriculture, manufacturing, and services. The gross capital stock is the value of capital goods still in use and evaluated as if they were new capital goods, without accounting for their age or depreciation undergone through time. The net capital stock is the value of capital goods still in use and evaluated as if they were new capital goods, minus the *compounded* depreciation undergone through time.

12. **Unobserved economic activity, national, total and by sector:**

    ISTAT provides *ad hoc* reports for the unobserved economic activity (value added and units of full-time work). Data are available from 2011 to 2020 and disaggregated in different ways but not by region, as ISTAT discontinued such statistics in 2016. The data retrieved are in millions of euros for the total, and as a percentage of total unobserved activity when disaggregated by sector. In the latter case, we sum across all the segments of each sector and multiply it by the total yearly values in millions of euros to obtain the sectoral value in euros.

13. **Nominal interest rates on ten-year government securities:**

    > Series: "Long-Term Government Bond Yields: 10-year: Main (Including Benchmark) for Italy, Percent."

    The OECD compiled the series, which we retrieved through FRED, matching the average monthly rates retrieved from the International Financial Statistics of the IMF.

14. **Historical GDP, employment, and population data:**

    Data for GDP per capita and employment in thousands from 1871 are available in Felice (2019). The population data before 1952 are taken from the archives of historical data of ISTAT and made publicly available to researchers.

Table C.1: Italian regional descriptive statistics for major economic aggregates over the period 1995-2020

	Average values over the period 1995-2020				
	GDP per capita	*Consumption of final goods*†	*Disposable income*†	*Unemployment rate* (%)	*Irregularly employed*‡ (%)
Northern regions	*34373*	*22020*	*19607*	*5.87*	*9.55*
Piemonte	30373	21468	18835	7.78	8.75
Valle d'Aosta	39948	31279	19463	5.16	8.68
Liguria	31459	22358	19260	8.48	10.30
Lombardia	37713	21826	20474	5.33	10.50
Trentino Alto Adige	39648	27351	20653	4.01	8.32
Veneto	31896	20898	17788	5.53	8.66
Friuli-Venezia Giulia	30290	21664	18551	5.87	9.41
Emilia-Romagna	34585	22728	20676	5.07	9.19
Central regions	*31786*	*21279*	*17950*	*8.34*	*12.76*
Toscana	30373	21644	18295	6.59	10.69
Umbria	26784	20142	16983	7.93	12.20
Marche	26910	19960	16501	6.89	9.99
Lazio	34887	21587	18276	10.05	14.89
South & islands	*18997*	*17103*	*12285*	*16.72*	*18.57*
Abruzzo	24756	18796	14586	9.89	14.63
Molise	21466	18239	12980	11.38	13.34
Campania	18937	16026	12153	17.93	22.26
Puglia	18083	16373	12281	15.76	16.39
Basilicata	20448	17073	12251	13.21	12.02
Calabria	16994	17826	11364	17.49	21.42
Sicilia	18276	17550	11916	19.51	19.41
Sardegna	20562	18923	13063	14.56	12.72
Italy	*28484*	*20154*	*16713*	*9.67*	*12.77*

The values for GDP per capita, consumption of final goods, and disposable income are in euros. †: Value in per capita terms. ‡: Data are computed over the period 2000-2020.